

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ОДЕСЬКА ПОЛІТЕХНІКА»
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

*Кваліфікаційна наукова
праця на правах рукопису*

ШВАНДТ МАКСИМ АЛЬБЕРТОВИЧ

УДК 004.932.2

ДИСЕРТАЦІЯ

**МОДЕЛІ ТА МЕТОДИ ДЕТЕКТУВАННЯ, ТРЕКІНГУ ТА НАСКРІЗНОЇ
ІДЕНТИФІКАЦІЇ ЛАБОРАТОРНИХ ТВАРИН У ВІДЕОДАНИХ**

Спеціальність: 122 – «Комп'ютерні науки»

Галузь знань: 12 – «Інформаційні технології»

Подається на здобуття наукового ступеня доктора філософії

Дисертація містить результати власних досліджень. Використання ідей,
результатів і текстів інших авторів мають посилання на відповідне джерело

_____ М.А. Швандт

Науковий керівник:

Волкова Наталія Павлівна, кандидат технічних наук, доцент

Одеса – 2026

АНОТАЦІЯ

Швандт М.А. Моделі та методи детектування, трекінгу та наскрізної ідентифікації лабораторних тварин у відеоданих. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 122 – Комп'ютерні науки. – Національний університет «Одеська політехніка», МОН України, Одеса, 2026.

У **вступі** обґрунтовано актуальність вирішення задач автоматизації моніторингу поведінки піддослідних об'єктів, зокрема, лабораторних тварин, у контрольованих умовах за відеоданими. Показана недосконалість існуючих моделей та методів комп'ютерного зору для задач детектування, трекінгу та наскрізної ідентифікації лабораторних тварин за умов нестаціонарного фону, оклюзій, злиття та розділення об'єктів. Визначено об'єкт, предмет, мету, задачі та методи дослідження; наведено наукову новизну; практичне значення одержаних результатів; висвітлено особистий внесок здобувача.

В **першому розділі** дисертаційної роботи проведено аналіз методів та підходів до детектування, трекінгу та збереження ідентичності лабораторних тварин у відеоданих, які використовуються в комп'ютерному зорі для контрольованих сцен. Розглянуто особливості систем відеомоніторингу поведінки лабораторних тварин та умови спостереження, що визначають вимоги до процедур детектування, трекінгу та збереження ідентичності об'єктів у часі за складних умов спостереження. Зазначено, що якість цих процедур впливає на точність формування траєкторій руху та подальшого обчислення поведінкових показників: швидкості, довжини шляху, часу перебування в заданих зонах, частоти відвідування зон, фактів контакту або зближення між об'єктами. Показано, що класичні методи (порогування за яскравістю та кольором, віднімання фону, зіставлення із шаблоном, відстеження за оптичним потоком, MeanShift та CamShift, трекінг ключових точок із фільтрами, тощо) характеризуються відносною алгоритмічною

простотою та невисокою обчислювальною складністю, однак є нестійкими до змін освітлення, деформацій форми, дрібних артефактів і перекриттів. Особливо це проявляється у відеозаписах, отриманих у контрольованому піддослідному середовищі, наприклад під час спостереження за тваринами в акваріумних умовах. У таких випадках базові підходи до детектування та локалізації, засновані лише на відніманні фону або різниці кадрів без просторової узгодженості, часто призводять до фрагментації об'єкта та збільшення кількості хибнопозитивних виявлень.

З огляду на зазначені обмеження класичних методів, в роботі розглянуто перехід до нейромережевого детектування як базового інструменту для локалізації лабораторних тварин у кадрі з подальшим трекінгом. Систематизовано основні сімейства нейромережевих детекторів та їхню логіку: двоетапні детектори (R-CNN, Faster R-CNN та похідні), одноетапні детектори (SSD та модифікації), а також детектори, що забезпечують відносну швидкодію, зокрема YOLO, CenterNet тощо. Визначено, що при побудові прикладного конвеєра для довготривалого відеоаналізу вирішальними є не лише точність, а й вимоги до обчислювальних ресурсів під час навчання, можливість навчання на локальному обладнанні без застосування хмарних ресурсів, стабільність навчання при обмеженій кількості зображень, а також час обробки одного кадру з урахуванням додаткової попередньої обробки.

Для формалізації вибору нейромережевого детектора запропоновано використовувати набір критеріїв, що враховує можливість локального навчання з прийнятним розміром навчального пакета (англ. batch size), придатність до виявлення малих об'єктів і маловиразних візуальних ознак, співвідношення «якість-швидкодія» з урахуванням часу детектування одного кадру та подальшого масштабування обробки на довгі відеозаписи, практична доступність реалізацій і простота інтеграції. Показано, що частина моделей, зокрема EfficientDet, окремі конфігурації SSD із ресурсомісткими магістральними мережами (англ. backbone) та Faster R-CNN-подібні рішення,

можуть бути малопридатними для локального навчання на звичайному обладнанні в умовах обмежених обчислювальних ресурсів через високі вимоги до пам'яті, зокрема при збільшенні розміру навчального пакета. Додатково встановлено, що не всі моделі забезпечують прийнятний час детекції, що обмежує їхнє практичне використання під час обробки довготривалих відеоданих. Це дозволило сформулювати вимоги до базового детектора для подальшої побудови конвеєра обробки відеоданих.

Результати аналізу наукових робіт з детектування, відстеження та ідентифікації об'єктів показали, що для цільового сценарію лабораторного відеоспостереження найбільш практичними є моделі, які забезпечують достатню швидкодію під час обробки окремого кадру та є придатними для навчання та використання на локальному обладнанні без істотних втрат точності при збереженні прийнятної оперативності. Зазначені вимоги визначають вибір базового детектора для подальшої побудови конвеєра обробки відеоданих і обґрунтовують ідею доповнення нейромережевого детектора методом суперпіксельної сегментації для виявлення та локалізації об'єктів, а також нейромережевими модулями детектування, сегментації екземплярів і повторної ідентифікації (англ. Re-ID) як складової наскрізної ідентифікації лабораторних тварин.

За результатами проведеного аналізу сформульовано мету та завдання дисертаційного дослідження.

У **другому розділі** дисертаційної роботи розв'язано задачу підвищення якості детектування та локалізації лабораторних тварин у довготривалих відеозаписах за умов нестаціонарного фону, артефактів, оклюзій, часткових перекриттів і злиття об'єктів, коли традиційні підходи на основі віднімання фону або різниці кадрів демонструють недостатню якість. Показано, що навіть у формально контрольованому лабораторному середовищі фон не є стаціонарним: у кадрі можуть з'являтися дрібні сторонні об'єкти, локальні неоднорідності, тіні, відблиски та інші артефакти, які за кольором, розміром або формою частково наближаються до об'єкта спостереження та можуть

спричиняти часткові перекриття його контуру, ускладнюючи виділення та локалізацію об'єкта. У результаті класичне віднімання фону з подальшою бінаризацією і фільтрацією за площею часто дає помилкові спрацьовування або, навпаки, розбиває об'єкт на фрагменти, що унеможлиблює стабільну локалізацію та подальший аналіз поведінкових метрик.

Для подолання вказаних обмежень у розділі запропоновано удосконалений метод детектування та локалізації лабораторних тварин у відеоданих на основі удосконаленого методу суперпиксельної сегментації SLIC. Метод призначений для автоматизованої обробки лабораторних відеозаписів, у яких ключовими джерелами похибок є нестационарний фон, артефакти та часткові оклюзії. Концепція запропонованого методу полягає в тому, щоб після формування різницевого зображення не виконувати грубе глобальне порогоування, а перейти до контекстно-залежного аналізу в межах сегментів (суперпикселів), отриманих за допомогою удосконаленого методу SLIC. Така постановка дозволяє відділяти локально узгоджені області об'єкта від фонових артефактів, які на різницевому зображенні можуть виглядати як дрібні або середні плями, подібні до частин тіла тварини.

Запропонований метод складається з наступних етапів:

1. Попередня обробка кадру та фонового зображення. Для зменшення впливу шумів, підвищення роздільної здатності та підсилення чіткості дрібних деталей об'єкта застосовуються двостороння фільтрація, регулювання яскравості й контрастності, медіанна фільтрація та, за потреби, масштабування зображення.

2. Формування різницевого зображення та первинне віднімання фону. Після попередньої обробки виконується віднімання фону, що дозволяє отримати карту змін і виділити кандидатів на рухомі об'єкти. Разом з тим показано, що саме на цьому кроці може з'являтися значна кількість хибних включень, зумовлених фоновими артефактами, локальними неоднорідностями, тінями, відблисками та дрібними сторонніми об'єктами, які зберігаються на різницевому зображенні та створюють хибні області.

3. Сегментація різницевого зображення на суперпикселі удосконаленим методом SLIC. Для подальшого контекстного аналізу різничеве зображення або пов'язане з ним представлення розбивається на сегменти на основі удосконаленого методу SLIC. Ключова відмінність удосконаленого методу SLIC полягає у модифікації метрики кластеризації, яка враховує просторові координати пікселів, ознаки LAB-моделі зображення, різничеве зображення, а також перевірку однорідності пікселів у локальному оточенні. Зокрема, центри кластерів описуються просторовими координатами та ознаками, сформованими на основі LAB-моделі зображення, різницевого зображення й перевірки однорідності пікселів у локальному оточенні. Відстань між пікселем і центром кластера обчислюється з урахуванням усіх зазначених ознак.

4. Контекстно-залежне пороговання в межах сегментів (суперпикселів) і відсікання артефактів. Показано, що просте перетворення різницевого зображення у бінарне та фільтрація плям за розміром є недостатньо точними, оскільки частина дрібних білих фрагментів може належати реальному об'єкту, але відокремлюється через залишкову неоднорідність після віднімання фону. Для усунення цього обмеження запропоновано спеціальну процедуру: для кожного сегмента обчислюється кількість пікселів, що перевищують заданий поріг; якщо ця кількість менша за встановлену частку (наприклад, менше ніж $1/3$ від усіх пікселів сегмента; конкретне співвідношення задається з урахуванням розміру зображення та очікуваного масштабу об'єкта), то значення пікселів сегмента зануляються. Такий механізм дозволяє відсікати сегменти, зумовлені фоновими артефактами, локальними неоднорідностями та частковими перекриттями, та водночас зменшити ризик втрати тонких частин контуру об'єкта.

5. Постобробка та відновлення цілісності об'єкта. Фрагментовані області об'єкта об'єднуються для покращення просторової когерентності, що дає більш стійку основу для подальшої локалізації у вигляді обмежувальної рамки або піксельної маски, трекінгу та формування траєкторій.

Експериментальна перевірка запропонованого методу на відеокадрах із бичками продемонструвала підвищення якості детектування та локалізації лабораторних тварин більш ніж на 6 % порівняно з базовим методом SLIC. Додатково показано, що модульність і інтерпретованість підходу дозволяють масштабувати його на інші лабораторні сценарії (зокрема, на поведінковий аналіз гризунів у коробчастих аренах) без залучення глибокого навчання чи ресурсомістких архітектур, що важливо для практичного застосування в умовах обмежених обчислювальних ресурсів.

Таким чином, сформульовано **перший та другий пункти наукової новизни**. **Удосконалено** метод суперпіксельної сегментації SLIC за рахунок використання багатопланового простору ознак, який базується на LAB-моделі зображення, враховує різницеве зображення та результати перевірки однорідності пікселів у локальному оточенні, що дозволило підвищити якість локалізації лабораторних тварин у відеоданих за умов нестаціонарного фону та наявності артефактів. **Удосконалено** метод детектування лабораторних тварин у відеоданих шляхом використання методу суперпіксельної сегментації SLIC і контекстної фільтрації, що дозволило підвищити якість детектування лабораторних тварин за умов нестаціонарного фону.

У **третьому розділі** розв'язано задачу підвищення якості детектування та сегментації лабораторних тварин на зображеннях за умов нестаціонарного фону, артефактів, часткових перекриттів, злиття та подальшого розділення об'єктів. За таких умов виникає необхідність у такій нейромережевій моделі, яка забезпечує високу якість локалізації на рівні пікселів, зберігає модульність для подальших розширень конвеєра «детектування–трекінг–повторна ідентифікація», допускає використання додаткових каналів ознак, сформованих на етапах попередньої обробки та призначених для підсилення ознак об'єкта в складних умовах сцени.

Обґрунтовано вибір базової нейромережової моделі Mask R-CNN як модульної двоетапної моделі з високою якістю сегментації. Визначено ключове обмеження типових реалізацій: орієнтація на звичайні 3-канальні

зображення та відсутність можливості безпосереднього використання додаткових штучно створених каналів, отриманих на етапах попередньої обробки, зокрема після віднімання фону, CLAHE або корекції тіней, які можуть суттєво підсилювати прояв об'єкта у складних умовах фону. Тому запропоновано модифікацію нейромережевої моделі Mask R-CNN, яка дозволяє інтегрувати багатоканальні дані без переходу до надмірно ресурсомістких магістралей типу ResNet101 зі збереженням прийнятної оперативності обробки.

Основою запропонованого рішення є двомагістральна нейромережева модель екстракції ознак: магістраль А для базових каналів зображення (основна інформація) та магістраль В для додаткових каналів ознак, сформованих на попередніх етапах конвеєра обробки. Реалізовано механізм злиття ознак $B \rightarrow A$ на ключових рівнях піраміди ознак, що дозволяє підсилювати представлення об'єкта додатковою інформацією саме там, де це найбільш корисно для детектування та сегментації за умов складного фону й часткових перекриттів. Обрано кероване SE-злиття (Squeeze-and-Excitation) після конкатенації ознак двох гілок, що забезпечує змішування ознак з урахуванням даних.

Порівняльний аналіз показав, що для запропонованої нейромережевої моделі Mask R-CNN на базі ResNet18 кількість епох навчання збільшено до 50. Отримано показники якості на рівні: $mAP50 = 0.87$, $Precision = 0.91$, $Recall = 0.91$ для обмежувальних рамок детекцій та $mAP50 = 0.9$, $Precision = 0.92$, $Recall = 0.92$ для масок сегментації за середнього часу виведення близько 105.5 мс, що підтверджує можливість досягнення високої якості детектування та сегментації на відносно компактній основі зі збереженням прийнятної оперативності обробки.

Сформульовано **третій пункт наукової новизни**: удосконалено нейромережеву модель Mask R-CNN за рахунок введення подвійної магістралі екстракції ознак із керованим SE-злиттям ($B \rightarrow A$), що дозволило підвищити якість детектування та сегментації лабораторних тварин у

відеоданих за умов часткових перекриттів, злиття та подальшого розділення об'єктів зі збереженням прийнятної оперативності.

У **четвертому розділі** розв'язано задачу наскрізної ідентифікації лабораторних тварин у довготривалих відеоспостереженнях за умов частих перекриттів, фрагментації траєкторій та незначних візуальних відмінностей між особинами.

Запропоновано концептуальну модель наскрізної ідентифікації лабораторних тварин у відеоданих без використання ручного маркування Mask-TAUDL, що інтегрує сегментацію екземплярів на основі Mask R-CNN із механізмом навчання асоціації треклетів на основі алгоритму TAUDL і може бути застосована як для однокамерного, так і для багатокамерного спостереження. На першому етапі сегментатор формує точні маски, після чого результати детектування, сегментації та трекінгу агрегуються у короткі треклети, що підвищує стійкість до короточасних оклюзій і фрагментації траєкторій. Для формування дескриптора використано двомагістральний екстрактор ознак twin-ResNet18: магістраль А працює з базовими каналами, магістраль В – з додатковими похідними каналами, отриманими на етапі попередньої обробки. Ознаки зливаються керовано через конкатенацію та SE-механізм, що дає змогу підсилювати інформативні компоненти й пригнічувати артефакти. Далі формується 256-вимірний $L2$ -нормалізований вектор вбудовування на основі ROI-ознак для повторної ідентифікації з урахуванням маски (масково-зважене агрегування), придатний для подальшого зіставлення треклетів, що підвищує дискримінативність представлення за рахунок фокусування на формі й текстурі тварини.

У концептуальній моделі Mask-TAUDL реалізовано поєднання одночасного навчання двох компонентів модуля TAUDL, що забезпечують навчання дискримінативних ознак треклетів у межах однієї сесії та асоціацію треклетів між різними сесіями запису без використання ручних ID-міток. Різні сесії запису інтерпретуються як «камери», що дозволяє навчати модель інваріантності до змін освітлення та середовища.

Дослідження показали, що поєднання сегментованих треклетів і ознак, отриманих двомагістральним екстрактором, підвищує стійкість асоціації та зменшує помилки перемикавання ідентичності в довготривалих лабораторних відеозаписах.

Сформульовано **четвертий пункт наукової новизни**: вперше запропоновано концептуальну модель наскрізної ідентифікації лабораторних тварин у відеоданих без використання ручного маркування Mask-TAUDL, яка інтегрує механізм отримання ознак на основі удосконаленої нейромережевої моделі Mask R-CNN із подвійною магістраллю екстракції ознак, масково-зважені ROI-ознаки та механізми асоціації треклетів без використання ручного маркування TAUDL, що забезпечує формування нормалізованих векторів вбудовувань і стабільне збереження ідентичності лабораторних тварин протягом відеопослідовності.

У **п'ятому розділі** розроблено, реалізовано та експериментально перевірено інтегрований конвеєр детектування, трекінгу та наскрізної ідентифікації лабораторних тварин у відеоданих, який покладено в основу спеціалізованої системи автоматизованої обробки лабораторних відеозаписів. Конвеєр поєднує попередню обробку кадрів, допоміжну сегментаційну гілку детектора на основі віднімання фону та удосконаленого методу сегментації SLIC, удосконалену нейромережеву модель Mask R-CNN із подвійною магістраллю ResNet18, арбітраж результатів детектування, трекінг, формування треклетів і подальшу асоціацію треклетів без використання ручного маркування на основі концептуальної моделі Mask-TAUDL.

Експериментальне тестування проведено на трьох тестових відеопослідовностях, що містили синтетичну контрольовану сцену та реальні лабораторні відеозаписи з мишами й бичками тривалістю 900, 3200 та 4320 кадрів. Оцінювання роботи конвеєра виконано за проксі-метриками збереження ідентичності, часової узгодженості локалізації, безперервності траєкторій і достатньої тривалості спостереження: PSWrate, median(tIoU), $P(tIoU < \theta)$, avg(gaps), median(T) та S(T).

Отримані результати підтверджують працездатність запропонованого інтегрованого конвеєра та його придатність до формування узгоджених у часі траєкторій із збереженням ідентичності об'єктів, а також можливість використання розробленого на його основі спеціалізованого застосунку для автоматизованого аналізу відеоспостережень за лабораторними тваринами.

Розроблені в дисертаційній роботі моделі, методи та інструментальні засоби впроваджено у лабораторні дослідження, що проводяться на базі біологічного факультету Одеського національного університету імені І. І. Мечникова, а також у навчальний процес Національного університету «Одеська політехніка» та Одеського національного університету імені І. І. Мечникова.

Ключові слова: комп'ютерний зір, обробка зображень, обробка відео, детектування об'єктів, трекінг об'єктів, сегментація зображень, неймережеві моделі, наскрізна ідентифікація.

СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Наукові праці, в яких опубліковані основні наукові результати дисертації:

1. Мороз В.В., Швандт М.А. Дослідження руху та поведінки лабораторних тварин методами детектування і трекінгу об'єктів // Вісник НТУ «ХПІ». Серія «Інформатика та моделювання». 2019. Вип. 13. № 1338. С. 66–76. DOI: 10.20998/2411-0558.2019.13.09. Видання включено до переліку наукових фахових видань України, категорія «Б».

2. Shvandt M. A., Moroz V. V. Overview of the detection and tracking methods of the lab animals // System Research & Information Technologies. 2022. No. 1. P. 124–148. DOI: 10.20535/SRIT.2308-8893.2022.1.10. Видання включено до переліку наукових фахових видань України, категорія «А» (Scopus).

3. Shvandt M. A., Moroz V. V. Overview of neural network object detection methods & models on the example of their use for lab animal observation // System Research & Information Technologies. 2025. No. 4. P. 71–103. DOI: 10.20535/SRIT.2308-8893.2025.4.05. Видання включено до переліку наукових фахових видань України, категорія «А» (Scopus).

4. Volkova N., Shvandt M. Segmentation-based approach for object detection // Proceedings of Odessa Polytechnic University. 2025. Issue 1(71). P. 145–156. DOI: 10.15276/opu.1.71.2025.17. Видання включено до переліку наукових фахових видань України, категорія «Б».

5. Volkova N., Shvandt M. Modification of the Mask R-CNN architecture for image detection and segmentation // Informatics and Mathematical Methods in Simulation. 2025. Vol. 15, No. 4. P. 496–506. DOI: 10.15276/imms.v15.no4.496. Видання включено до переліку наукових фахових видань України, категорія «Б».

6. Volkova N. P., Shvandt M. A. Unsupervised re-identification architecture based on segmented tracklets for animal behavior analysis // Herald of Advanced

Information Technology. 2025. Vol. 8, No. 4. P. 476–487. DOI: 10.15276/hait.08.2025.30. Видання включено до переліку наукових фахових видань України, категорія «А» (Scopus).

7. Shvandt M., Moroz V. General approaches to lab animal detection and tracking // Problems of Informatics and Modeling (PIM-2021): Proceedings/Abstracts of the XXI International Scientific and Technical Conference (September 09–14, 2021, Karolino-Buhaz, Ukraine).

URL: <https://repository.kpi.kharkov.ua/server/api/core/bitstreams/c079ac2d-ba5c-467a-b2f3-55d4e921588d/content> (дата звернення: 15.12.2025).

8. Volkova N., Shvandt M. Image preprocessing algorithm for object detection & tracking // Information Control Systems and Technologies (ICST-ODESA 2024): Proceedings of the XII International Scientific-Practical Conference (September 23–25, 2024, Odesa). 2024. P. 194–197.

URL: <https://icst-conf.com/2024.pdf> (дата звернення: 15.12.2025).

9. Shvandt M., Volkova N. Extended background subtraction algorithm for object detection & tracking // Project, Program, Portfolio Management (P3M-2024): The Proceedings of the International Research Conference (December 06–07, 2024, Odesa, Ukraine). 2024. P. 73–76. DOI: 10.5281/zenodo.15165174.

URL: <https://zenodo.org/records/15165174> (дата звернення: 15.02.2026).

10. Volkova N., Shvandt M. Segmentation-based extension for object detection // Modern Information Technology – 2025: Proceedings of the Fifteenth International Scientific Conference of Students and Young Scientists (May 15–16, 2025, Odesa, Ukraine). Odesa: Nauka i Tekhnika, 2025. DOI: 10.5281/zenodo.15521809.

URL: <https://doi.org/10.5281/zenodo.15521809> (дата звернення: 15.02.2026).

11. Volkova N., Shvandt M. Object detection network architecture for multichannel images // Information Control Systems and Technologies (ICST-ODESA 2025): Proceedings of the XIII International Scientific-Practical

Conference (September 24–26, 2025, Odesa). 2025. P. 166–169. DOI: 10.36059/978-966-397-531-3.

URL: <https://icst-conf.com/2025.pdf> (дата звернення: 15.12.2025).

12. Volkova N., Shvandt M. Improved approach to label-free identification using segmented tracklets for object behavior analysis // Informatics. Culture. Technique. 2025. No. 2. P. 279–282. DOI: 10.15276/ict.02.2025.42.

URL: <https://ict.od.ua/index.php/journal/en/article/view/53> (дата звернення: 15.12.2025).

SUMMARY

Shvandt, M.A. Models and methods for detection, tracking, and end-to-end identification of laboratory animals in video data. – Qualification scientific work in the form of manuscript.

Thesis for the PhD degree in specialty 122 Computer science. – Odesa Polytechnic National University, Ministry of Education and Science of Ukraine, Odesa, 2026.

The **introduction** substantiates the relevance of solving the problems of automating behavioral monitoring of experimental subjects, in particular laboratory animals, under controlled conditions using video data. It demonstrates the limitations of existing computer vision models and methods for the detection, tracking, and end-to-end identification of laboratory animals under conditions of a non-stationary background, occlusions, object merging, and object separation. The object, subject, aim, objectives, and methods of the research are defined; the scientific novelty and practical significance of the obtained results are presented; and the personal contribution of the applicant is outlined.

In the **first chapter** of the thesis, methods and approaches to the detection, tracking, and identity preservation of laboratory animals in video data, used in computer vision for controlled scenes, are analyzed. The chapter examines the specific features of video monitoring systems for laboratory animal behavior and the observation conditions that determine the requirements for detection, tracking, and object identity preservation over time under challenging observation conditions. It is noted that the quality of these procedures affects the accuracy of motion trajectory formation and the subsequent calculation of behavioral indicators, including speed, path length, time spent in predefined zones, frequency of zone visits, and instances of contact or proximity between objects. It is shown that classical methods, including brightness- and color-based thresholding, background subtraction, template matching, optical-flow-based tracking, MeanShift and CamShift, and keypoint tracking with filters, are characterized by

relative algorithmic simplicity and low computational complexity; however, they are unstable under changes in illumination, shape deformations, small artifacts, and occlusions. This is especially evident in video recordings obtained in a controlled experimental environment, for example during observations of animals under aquarium conditions. In such cases, basic approaches to detection and localization based only on background subtraction or frame differencing without spatial consistency often lead to object fragmentation and an increased number of false-positive detections.

Given the above limitations of classical methods, the thesis considers the transition to neural-network-based detection as a core tool for localizing laboratory animals in a frame with subsequent tracking. The main families of neural network detectors and their underlying principles are systematized: two-stage detectors, including R-CNN, Faster R-CNN, and their derivatives; one-stage detectors, including SSD and its modifications; and detectors that provide relatively high processing speed, in particular YOLO, CenterNet, and related models. It is determined that, when constructing an applied pipeline for long-duration video analysis, the decisive factors include not only accuracy, but also computational resource requirements during training, the possibility of training on local hardware without using cloud resources, training stability with a limited number of images, and the processing time per frame, taking into account additional preprocessing.

To formalize the selection of a neural network detector, it is proposed to use a set of criteria that takes into account the possibility of local training with an acceptable training batch size, suitability for detecting small objects and weakly expressed visual features, the quality–speed trade-off with regard to the detection time of a single frame and subsequent scaling of processing to long video recordings, the practical availability of implementations, and ease of integration. It is shown that some models, in particular EfficientDet, certain SSD configurations with resource-intensive backbone networks, and Faster R-CNN-like solutions, may be poorly suited for local training on standard hardware under limited computational resources due to high memory requirements, especially when

increasing the training batch size. It is additionally established that not all models provide an acceptable detection time, which limits their practical use in the processing of long-duration video data. This made it possible to formulate the requirements for the baseline detector for the subsequent construction of the video data processing pipeline.

The results of the analysis of scientific works on object detection, tracking, and identification showed that, for the target scenario of laboratory video monitoring, the most practical models are those that provide sufficient processing speed for individual frames and are suitable for training and deployment on local hardware without substantial loss of accuracy while maintaining acceptable processing efficiency. These requirements determine the choice of the baseline detector for the subsequent construction of the video data processing pipeline and justify the idea of supplementing the neural network detector with a superpixel segmentation method for object detection and localization, as well as with neural network modules for detection, instance segmentation, and re-identification (Re-ID) as a component of the end-to-end identification of laboratory animals.

Based on the results of the conducted analysis, the aim and objectives of the thesis research are formulated.

In the **second chapter** of the thesis, the problem of improving the quality of detection and localization of laboratory animals in long-duration video recordings is addressed under conditions of a non-stationary background, artifacts, occlusions, partial overlaps, and object merging, where traditional approaches based on background subtraction or frame differencing demonstrate insufficient quality. It is shown that, even in a formally controlled laboratory environment, the background is not stationary: small foreign objects, local inhomogeneities, shadows, specular reflections, and other artifacts may appear in the frame. In terms of color, size, or shape, these artifacts may partially resemble the object under observation and may cause partial occlusion of its contour, thereby complicating object extraction and localization. As a result, classical background subtraction followed by binarization and area-based filtering often produces false detections or, conversely, fragments

the object, which prevents stable localization and subsequent analysis of behavioral metrics.

To overcome these limitations, the chapter proposes an improved method for the detection and localization of laboratory animals in video data based on the improved SLIC superpixel segmentation method. The method is intended for the automated processing of laboratory video recordings in which the main sources of errors are a non-stationary background, artifacts, and partial occlusions. The concept of the proposed method is to avoid coarse global thresholding after the difference image has been formed and instead proceed to context-dependent analysis within segments, or superpixels, obtained using the improved SLIC method. This formulation makes it possible to separate locally consistent object regions from background artifacts, which may appear in the difference image as small or medium-sized blobs resembling parts of the animal's body.

The proposed method consists of the following stages:

1. Preprocessing of the frame and the background image. To reduce the influence of noise, increase resolution, and enhance the sharpness of fine object details, bilateral filtering, brightness and contrast adjustment, median filtering, and, if necessary, image upscaling are applied.

2. Construction of the difference image and initial background subtraction. After preprocessing, background subtraction is performed, which makes it possible to obtain a change map and identify candidates for moving objects. At the same time, it is shown that a considerable number of false inclusions may arise precisely at this stage, caused by background artifacts, local inhomogeneities, shadows, specular reflections, and small foreign objects that remain in the difference image and form spurious regions.

3. Segmentation of the difference image into superpixels using the improved SLIC method. For subsequent context-dependent analysis, the difference image or a related representation is partitioned into segments using the improved SLIC method. The key distinction of the improved SLIC method lies in the modification of the clustering metric, which takes into account the spatial coordinates of pixels,

features of the LAB image model, the difference image, and pixel homogeneity checks in the local neighborhood. In particular, the cluster centers are described by spatial coordinates and features formed on the basis of the LAB image model, the difference image, and pixel homogeneity checks in the local neighborhood. The distance between a pixel and a cluster center is computed by taking all the specified features into account.

4. Context-dependent thresholding within segments (superpixels) and artifact suppression. It is shown that simple conversion of the difference image into a binary image followed by blob-size filtering is insufficiently accurate, since some small white fragments may belong to the actual object but become separated due to residual inhomogeneity after background subtraction. To eliminate this limitation, a dedicated procedure is proposed: for each segment, the number of pixels exceeding a specified threshold is computed; if this number is below a predefined fraction, for example, less than one third of all pixels in the segment, with the exact ratio set with regard to the image size and the expected object scale, the pixel values within the segment are set to zero. This mechanism makes it possible to discard segments caused by background artifacts, local inhomogeneities, and partial occlusions, while simultaneously reducing the risk of losing thin parts of the object contour.

5. Post-processing and restoration of object integrity. Fragmented object regions are merged to improve spatial coherence, providing a more stable basis for subsequent localization in the form of a bounding box or pixel mask, tracking, and trajectory formation.

Experimental validation of the proposed method on video frames containing gobies demonstrated an improvement of more than 6% in the quality of detection and localization of laboratory animals compared with the baseline SLIC method. It is additionally shown that the modularity and interpretability of the approach make it possible to scale it to other laboratory scenarios, in particular to behavioral analysis of rodents in box-type arenas, without involving deep learning or

resource-intensive architectures, which is important for practical application under limited computational resources.

Thus, the **first** and **second points of scientific novelty** are formulated. The SLIC superpixel segmentation method **is improved** by using a multilayer feature space based on the LAB image model, taking into account the difference image and the results of pixel homogeneity checks in the local neighborhood, which made it possible to improve the quality of laboratory animal localization in video data under conditions of a non-stationary background and the presence of artifacts. The method for detecting laboratory animals in video data **is improved** by using SLIC superpixel segmentation and context-dependent filtering, which made it possible to improve the quality of laboratory animal detection under conditions of a non-stationary background.

In the **third chapter**, the problem of improving the quality of detection and segmentation of laboratory animals in images is addressed under conditions of a non-stationary background, artifacts, partial occlusions, object merging, and subsequent object separation. Under such conditions, there is a need for a neural network model that provides high-quality pixel-level localization, preserves modularity for further extensions of the “detection–tracking–re-identification” pipeline, and enables the use of additional feature channels formed at the preprocessing stages and intended to enhance object features under challenging scene conditions.

The choice of Mask R-CNN as the baseline neural network model is justified, since it is a modular two-stage model with high segmentation quality. A key limitation of typical implementations is identified: their orientation toward standard three-channel images and the lack of support for the direct use of additional artificially generated channels obtained at the preprocessing stages, in particular after background subtraction, CLAHE, or shadow correction, which can substantially enhance the visibility of the object under challenging background conditions. Therefore, a modification of the Mask R-CNN neural network model is proposed that enables the integration of multichannel data without switching to

excessively resource-intensive backbones such as ResNet101, while preserving acceptable processing efficiency.

The proposed solution is based on a dual-branch neural network model for feature extraction: branch A processes the base image channels, which contain the primary information, whereas branch B processes additional feature channels formed at the preceding stages of the processing pipeline. A $B \rightarrow A$ feature fusion mechanism is implemented at key levels of the feature pyramid, which makes it possible to enhance the object representation with additional information precisely where it is most useful for detection and segmentation under conditions of a complex background and partial occlusions. Controlled SE fusion (Squeeze-and-Excitation) after concatenation of the features from the two branches is selected, providing data-dependent feature mixing.

The comparative analysis showed that, for the proposed Mask R-CNN neural network model based on ResNet18, the number of training epochs was increased to 50. The following quality indicators were obtained: $mAP50 = 0.87$, Precision = 0.91, and Recall = 0.91 for detection bounding boxes, and $mAP50 = 0.90$, Precision = 0.92, and Recall = 0.92 for segmentation masks, with an average inference time of approximately 105.5 ms. This confirms the possibility of achieving high detection and segmentation quality on a relatively compact basis while preserving acceptable processing efficiency.

The **third point of scientific novelty** is formulated: the Mask R-CNN neural network model is improved by introducing a dual feature extraction backbone with controlled SE fusion ($B \rightarrow A$), which made it possible to improve the quality of detection and segmentation of laboratory animals in video data under conditions of partial occlusions, object merging, and subsequent object separation while preserving acceptable processing efficiency.

In the **fourth chapter**, the problem of end-to-end identification of laboratory animals in long-duration video observations is addressed under conditions of frequent occlusions, trajectory fragmentation, and minor visual differences between individuals.

A conceptual label-free model for the end-to-end identification of laboratory animals in video data, Mask-TAUDL, is proposed. It integrates Mask R-CNN-based instance segmentation with a tracklet association learning mechanism based on the TAUDL algorithm and can be applied to both single-camera and multi-camera observation scenarios. At the first stage, the segmenter generates accurate masks, after which the results of detection, segmentation, and tracking are aggregated into short tracklets, which increases robustness to short-term occlusions and trajectory fragmentation. For descriptor formation, a dual-branch feature extractor, twin-ResNet18, is used: branch A processes the base channels, whereas branch B processes additional derived channels obtained at the preprocessing stage. The features are fused in a controlled manner through concatenation and an SE mechanism, which makes it possible to enhance informative components and suppress artifacts. Then, a 256-dimensional $L2$ -normalized embedding vector is formed from mask-aware ROI features for re-identification using mask-weighted aggregation. This vector is suitable for subsequent tracklet matching and improves the discriminative capability of the representation by focusing on the animal's shape and texture.

The conceptual Mask-TAUDL model combines the simultaneous training of two components of the TAUDL module, which provide learning of discriminative tracklet features within a single session and tracklet association across different recording sessions without using manual ID labels. Different recording sessions are interpreted as “cameras,” which enables the model to learn invariance to changes in illumination and environment.

The studies showed that combining segmented tracklets with features obtained by the dual-branch extractor increases association robustness and reduces identity-switch errors in long-duration laboratory video recordings.

The **fourth point of scientific novelty** is formulated: for the first time, a conceptual label-free model for the end-to-end identification of laboratory animals in video data, Mask-TAUDL, is proposed. It integrates a feature extraction mechanism based on the improved Mask R-CNN neural network model with a

dual feature extraction backbone, mask-weighted ROI features, and label-free TAUDL tracklet association mechanisms, thereby enabling the formation of normalized embedding vectors and stable preservation of laboratory animal identity throughout the video sequence.

In the **fifth chapter**, an integrated pipeline for the detection, tracking, and end-to-end identification of laboratory animals in video data is developed, implemented, and experimentally validated. This pipeline forms the basis of a specialized system for the automated processing of laboratory video recordings. The pipeline combines frame preprocessing, an auxiliary segmentation branch of the detector based on background subtraction and the improved SLIC segmentation method, the improved Mask R-CNN neural network model with a dual ResNet18 backbone, detection result arbitration, tracking, tracklet formation, and subsequent label-free tracklet association based on the conceptual Mask-TAUDL model.

Experimental testing was conducted on three test video sequences, including a synthetic controlled scene and real laboratory video recordings of mice and gobies, with durations of 900, 3200, and 4320 frames. The pipeline performance was evaluated using proxy metrics of identity preservation, temporal localization consistency, trajectory continuity, and sufficient observation duration: PSWrate, median(tIoU), $P(tIoU < \theta)$, avg(gaps), median(T), and S(T).

The obtained results confirm the operability of the proposed integrated pipeline and its suitability for forming temporally consistent trajectories while preserving object identity, as well as the possibility of using the specialized application developed on its basis for the automated analysis of video observations of laboratory animals.

The models, methods, and software tools developed in the thesis have been implemented in laboratory studies conducted at the Faculty of Biology of Odesa I. I. Mechnikov National University, as well as in the educational process of Odesa Polytechnic National University and Odesa I. I. Mechnikov National University.

Keywords: computer vision, image processing, video processing, object detection, object tracking, image segmentation, neural network models, end-to-end identification.

List of publications of the applicant on the topic of the dissertation

1. Мороз В.В., Швандт М.А. Дослідження руху та поведінки лабораторних тварин методами детектування і трекінгу об'єктів // Вісник НТУ «ХПІ». Серія «Інформатика та моделювання». 2019. Вип. 13. № 1338. С. 66–76. DOI: 10.20998/2411-0558.2019.13.09.

2. Shvandt M. A., Moroz V. V. Overview of the detection and tracking methods of the lab animals // System Research & Information Technologies. 2022. No. 1. P. 124–148. DOI: 10.20535/SRIT.2308-8893.2022.1.10. (Scopus)

3. Shvandt M. A., Moroz V. V. Overview of neural network object detection methods & models on the example of their use for lab animal observation // System Research & Information Technologies. 2025. No. 4. P. 71–103. DOI: 10.20535/SRIT.2308-8893.2025.4.05. (Scopus)

4. Volkova N., Shvandt M. Segmentation-based approach for object detection // Proceedings of Odessa Polytechnic University. 2025. Issue 1(71). P. 145–156. DOI: 10.15276/opu.1.71.2025.17.

5. Volkova N., Shvandt M. Modification of the Mask R-CNN architecture for image detection and segmentation // Informatics and Mathematical Methods in Simulation. 2025. Vol. 15, No. 4. P. 496–506. DOI: 10.15276/imms.v15.no4.496.

6. Volkova N. P., Shvandt M. A. Unsupervised re-identification architecture based on segmented tracklets for animal behavior analysis // Herald of Advanced Information Technology. 2025. Vol. 8, No. 4. P. 476–487. DOI: 10.15276/hait.08.2025.30. (Scopus)

7. Shvandt M., Moroz V. General approaches to lab animal detection and tracking // Problems of Informatics and Modeling (PIM-2021): Proceedings/Abstracts of the XXI International Scientific and Technical Conference (September 09–14, 2021, Karolino-Buhaz, Ukraine).

URL: <https://repository.kpi.kharkov.ua/server/api/core/bitstreams/c079ac2d-ba5c-467a-b2f3-55d4e921588d/content> (access date: 15.12.2025).

8. Volkova N., Shvandt M. Image preprocessing algorithm for object detection & tracking // Information Control Systems and Technologies (ICST-ODESA 2024): Proceedings of the XII International Scientific-Practical Conference (September 23–25, 2024, Odesa). 2024. P. 194–197.

URL: <https://icst-conf.com/2024.pdf> (access date: 15.12.2025).

9. Shvandt M., Volkova N. Extended background subtraction algorithm for object detection & tracking // Project, Program, Portfolio Management (P3M-2024): The Proceedings of the International Research Conference (December 06–07, 2024, Odesa, Ukraine). 2024. P. 73–76. DOI: 10.5281/zenodo.15165174.

URL: <https://zenodo.org/records/15165174> (access date: 15.02.2026).

10. Volkova N., Shvandt M. Segmentation-based extension for object detection // Modern Information Technology – 2025: Proceedings of the Fifteenth International Scientific Conference of Students and Young Scientists (May 15–16, 2025, Odesa, Ukraine). Odesa: Nauka i Tekhnika, 2025. DOI: 10.5281/zenodo.15521809.

URL: <https://doi.org/10.5281/zenodo.15521809> (access date: 15.02.2026).

11. Volkova N., Shvandt M. Object detection network architecture for multichannel images // Information Control Systems and Technologies (ICST-ODESA 2025): Proceedings of the XIII International Scientific-Practical Conference (September 24–26, 2025, Odesa). 2025. P. 166–169. DOI: 10.36059/978-966-397-531-3.

URL: <https://icst-conf.com/2025.pdf> (access date: 15.12.2025).

12. Volkova N., Shvandt M. Improved approach to label-free identification using segmented tracklets for object behavior analysis // Informatics. Culture. Technique. 2025. No. 2. P. 279–282. DOI: 10.15276/ict.02.2025.42.

URL: <https://ict.od.ua/index.php/journal/en/article/view/53> (access date: 15.12.2025).

ЗМІСТ

Перелік умовних позначень	31
Вступ.....	32
1 Аналіз методів та підходів до детектування, трекінгу та збереження ідентичності лабораторних тварин у відеоданих.....	39
1.1 Аналіз систем відеомоніторингу поведінки лабораторних тварин та умов спостереження	39
1.1.1 Контрольовані відеосцени та постановки задач спостереження	41
1.1.2 Вимоги до алгоритмів детектування, трекінгу та збереження ідентичності об'єктів у складних умовах спостереження	43
1.2 Аналіз класичних підходів до детектування та трекінгу в контрольованих сценах	46
1.2.1 Попередня обробка відеоданих, моделювання фону та виділення об'єктів	46
1.2.2 Детектування об'єктів за ознаками та шаблонами	50
1.2.3 Класичні методи трекінгу об'єктів у відеоданих.....	55
1.2.4 Висновки щодо ефективності класичних методів детектування та трекінгу	59
1.3 Нейромережеві методи детектування та пов'язані підходи для задач відеоспостереження	64
1.3.1 Загальні принципи нейромережевого детектування об'єктів у відеоданих.....	65
1.3.2 Двоетапні та одноетапні детектори.....	67
1.3.3 Безанкерні підходи до детектування об'єктів та функції втрат	69
1.3.4 Магістральні мережі ResNet, MobileNet та піраміди ознак у нейромережевих детекторах	71
1.3.5 Сегментаційні архітектури для підвищення точності локалізації	73
1.3.6 Оптимізація нейромережевих детекторів для відеомоніторингу	74
1.3.7 Нейромережеві підходи до відстеження об'єктів у відеоданих.....	75

1.3.8 Висновки щодо нейромережових методів до детектування та трекінгу	76
1.4 Висновки до першого розділу	77
2 Методи детектування та локалізації лабораторних тварин у відеоданих на основі суперпиксельної сегментації SLIC та її удосконалення.....	79
2.1 Постановка задачі детектування та локалізації лабораторних тварин у відеоданих	80
2.2 Етапи попередньої обробки	85
2.2.1 Фільтрація шуму та нормалізація яскравості й контрасту	85
2.2.2 Віднімання фону та порогоування	87
2.3 Суперпиксельна сегментація SLIC.....	89
2.3.1 Базовий метод SLIC	89
2.3.2 Удосконалення методу SLIC.....	93
2.4 Результати детектування та локалізації лабораторних тварин із використанням удосконаленого методу SLIC.....	97
2.5 Висновки до другого розділу	101
3 Удосконалена нейромережева модель Mask R-CNN для детектування та сегментації лабораторних тварин у відеоданих	103
3.1 Постановка задачі нейромережевого детектування та сегментації лабораторних тварин.....	103
3.2 Базова нейромережева модель Mask R-CNN для детектування та сегментації лабораторних тварин	109
3.3 Удосконалення моделі Mask R-CNN для детектування та сегментації лабораторних тварин.....	115
3.3.1 Базовий екстрактор ознак ResNet.....	115
3.3.2 Подвійна магістраль екстракції ознак.....	119
3.3.3 Кероване злиття ознак в удосконаленій нейромережевій моделі Mask R-CNN.....	122
3.4 Навчання, налаштування та оцінювання удосконаленої нейромережевої моделі Mask R-CNN	127

3.5 Результати детектування та сегментації лабораторних тварин із використанням удосконаленої нейромережевої моделі Mask R-CNN	129
3.6 Висновки до третього розділу	134
4 Концептуальна модель Mask-TAUDL для наскрізної ідентифікації лабораторних тварин у відеоданих без ручного маркування	136
4.1 Формування сегментованих треклетів і ознак для повторної ідентифікації	136
4.1.1 Отримання треклетів із результатів детектування, сегментації та трекінгу	137
4.1.2 Масково-зважені ROI-ознаки та агрегація ознак у межах треклета	139
4.1.3 Формування вбудовувань та нормалізація	142
4.1.4 Дискримінація та асоціація треклетів без ручного маркування на основі принципів TAUDL	144
4.2 Функція оптимізації та інтеграція Mask-TAUDL з удосконаленою Mask R-CNN.....	148
4.3 Експериментальне дослідження концептуальної моделі Mask-TAUDL.	153
4.4 Висновки до четвертого розділу	157
5 Інтегрований конвеєр детектування, трекінгу та наскрізної ідентифікації лабораторних тварин у відеоданих.....	159
5.1 Структура інтегрованого конвеєра.....	159
5.2 Метрики оцінювання якості роботи інтегрованого конвеєра.....	163
5.3 Умови проведення тестування та використані набори даних	167
5.4 Результати застосування інтегрованого конвеєра для аналізу лабораторних відеоданих	168
5.5 Висновки до п'ятого розділу.....	175
Висновки	178
Список використаних джерел	182
Додаток А Список публікацій здобувача за темою дисертації	198

Додаток Б Акт впровадження результатів дисертаційної роботи у лабораторні дослідження біологічного факультету Одеського національного університету імені І. І. Мечникова	201
Додаток В Акт впровадження результатів дисертаційної роботи у навчальний процес Одеського національного університету імені І. І. Мечникова	202
Додаток Г Довідка про використання результатів дисертаційної роботи у науково-дослідницькій діяльності Національного університету «Одеська політехніка»	203
Додаток Д Довідка про впровадження результатів дисертаційної роботи у навчальному процесі Національного університету «Одеська політехніка» .	204

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

ID – Ідентифікатор

ROI – Region of Interest (область інтересу)

FPN – Feature Pyramid Network

RPN – Region Proposal Network

ROIAlign – операція вирівнювання області інтересу (ROI)

PyramidROIAlign – ROIAlign на рівнях піраміди ознак (FPN)

ВСТУП

Актуальність теми. Завдяки швидкому розвитку методів комп'ютерного зору та інформаційних технологій розширились можливості автоматизованої обробки відеоданих у біологічних та медико-біологічних експериментах. Це дозволяє проводити відеоспостереження для реєстрації активності та переміщень лабораторних тварин, їх взаємодій, реакції на зовнішні стимули та інших поведінкових метрик. Разом з тим, у практиці лабораторних досліджень досі широко використовуються ручна розмітка й аналіз відеозаписів, що є трудомістким, суб'єктивним та погано масштабованими на довготривалі експерименти й великі масиви даних. Тому виникає потреба у створенні інструментів, здатних забезпечити стійке детектування, трекінг і збереження ідентичності лабораторних тварин у відеоданих як бази для подальшого автоматизованого поведінкового аналізу. Разом з тим, що лабораторне середовище вважається контрольованим, реальні відеозаписи часто містять фактори, які є критичними для алгоритмів комп'ютерного зору: нестаціонарний фон (відблиски, шум, артефакти), а лабораторні тварини можуть частково перекриватися, зливатися, деформуватися в кадрі та швидко змінювати напрям руху. Такі фактори призводять до фрагментації об'єкта, збільшення кількості хибнопозитивних спрацьовувань або втрати цілісної локалізації, що унеможливорює коректне формування траєкторій і похідних показників поведінки.

Для автоматизованої обробки відеоданих у біологічних та медико-біологічних експериментах застосовують типовий конвеєр: попередня обробка, детектування, сегментація, асоціація детекцій у часі, формування треклетів, повторна ідентифікація. У цій роботі зазначені етапи розглядаються як складові процесу наскрізної ідентифікації лабораторних тварин, під якою розуміється збереження сталого ідентифікатора кожного об'єкта протягом відеопослідовності, зокрема після часткових перекриттів,

злиття, розділення об'єктів і фрагментації треків. При цьому методи такого конвеєра умовно можна поділити на два типи: класичні та нейромережеві.

Класичні методи (порогування, віднімання фону, оптичний потік, зіставлення із шаблонами, трекінг ключових точок тощо) відзначаються простотою та низькими вимогами до ресурсів, однак у сценах із нестаціонарним фоном і оклюзіями демонструють недостатню точність. Натомість сучасні нейромережеві детектори та сегментатори здатні забезпечити високу точність, проте потребують обґрунтованого вибору архітектури, стабільності навчання на обмежених вибірках, а також збереження оперативності при обробці довгих відеорядів на локальному обладнанні.

Таким чином, існує протиріччя, яке полягає в тому, що, з одного боку, для автоматизованого поведінкового моніторингу лабораторних тварин необхідні методи детектування, відстеження та збереження ідентичності об'єктів, які здатні забезпечити високу точність і стійкість у реальних лабораторних відеосценах. З іншого боку, існуючі методи комп'ютерного зору та машинного навчання, попри їхні переваги, не забезпечують у сукупності тих характеристик, які є необхідними для локальної обробки тривалих лабораторних відеоспостережень, а саме високої точності, стійкості до складних умов сцени та помірних вимог до обчислювальних ресурсів.

Тому актуальними є дослідження, спрямовані на розробку моделей та методів для вирішення науково-практичної задачі підвищення точності й стійкості детектування, трекінгу та наскрізної ідентифікації лабораторних тварин у відеоданих за умов нестаціонарного фону, оклюзій, злиття та розділення об'єктів, обмежених ресурсів і локальної обробки тривалих відеоспостережень.

Мета і задачі дослідження. Метою дослідження є підвищення якості та стійкості детектування, трекінгу та наскрізної ідентифікації лабораторних тварин у відеоданих за умов нестаціонарного фону, оклюзій, злиття та

розділення об'єктів шляхом розробки моделей та методів для конвеєра обробки відеоданих.

Для досягнення поставленої мети у роботі поставлено такі **задачі**:

- провести аналіз класичних і сучасних нейромережових підходів до детектування, трекінгу та наскрізної ідентифікації для конвеєра обробки відеоданих лабораторних тварин, визначити проблеми і обмеження їх використання та намітити шляхи їх вирішення;
- удосконалити метод детектування лабораторних тварин у відеоданих за рахунок використання процедури сегментації з метою підвищення якості їх детектування за умов нестационарного фону;
- удосконалити метод суперпиксельної сегментації SLIC для підвищення якості локалізації лабораторних тварин у відеоданих за умов нестационарного фону та наявності артефактів;
- удосконалити нейромережову модель Mask R-CNN шляхом введення подвійної магістралі екстракції ознак із керованим SE-злиттям для підвищення якості детектування та сегментації лабораторних тварин у відеоданих за умов часткового перекриття, злиття та подальшого розділення об'єктів при збереженні прийнятної оперативності;
- розробити концептуальну модель наскрізної ідентифікації лабораторних тварин у відеоданих без використання ручного маркування Mask-TAUDL для формування векторів вбудовувань, асоціації треклетів та збереження ідентичності об'єктів протягом відеопослідовності;
- реалізувати інтегрований конвеєр детектування, трекінгу та наскрізної ідентифікації лабораторних тварин у відеоданих на основі розроблених моделей і методів, провести його експериментальне дослідження та апробацію.

Об'єктом дослідження є процеси автоматизованого детектування, трекінгу та наскрізної ідентифікації лабораторних тварин у відеоданих.

Предметом дослідження є моделі та методи детектування, трекінгу та наскрізної ідентифікації лабораторних тварин у відеоданих за умов нестаціонарного фону, оклюзій, злиття та розділення об'єктів.

Методи дослідження. Для досягнення мети дисертаційної роботи застосовано методи цифрової обробки зображень і відео (фільтрація, підсилення контрасту, формування різницевих представлень, постобробка масок), методи сегментації (метод суперпиксельної сегментації SLIC та її модифікацію, контекстно-залежне порогування в межах сегментів), методи машинного і глибокого навчання (нейромережеві детектори, сегментація екземплярів, модифікацію нейромережевої моделі Mask R-CNN), методи побудови та злиття ознак (багатоканальні представлення, механізми керованого злиття ознак із використанням підходу стиснення та стимуляції SE (Squeeze-and-Excitation)), методи наскрізної ідентифікації лабораторних тварин у відеоданих, що охоплюють навчання без використання ручного маркування, формування векторів ознак, повторну ідентифікацію об'єктів (Re-ID) та асоціацію треклетів, а також методи експериментального оцінювання якості за метриками Accuracy, Precision, Recall, mAP50; для Re-ID – mAP та Rank-k.

Наукова новизна одержаних результатів полягає у вирішенні науково-технічної задачі підвищення стійкості й якості детектування, сегментації, трекінгу та наскрізної ідентифікації лабораторних тварин у складних умовах спостереження, що забезпечує основу для подальшої автоматизації поведінкового аналізу. Основні наукові результати, що виносяться на захист, полягають у наступному:

- удосконалено метод суперпиксельної сегментації SLIC за рахунок використання багатопланового простору ознак, який базується на LAB-моделі зображення, враховує різницеве зображення та результати перевірки однорідності пікселів у локальному оточенні, що дозволило підвищити якість локалізації лабораторних тварин у відеоданих за умов нестаціонарного фону та наявності артефактів;

– удосконалено метод детектування лабораторних тварин у відеоданих шляхом використання методу суперпiксельної сегментації SLIC i контекстної фiльтрації, що дозволило підвищити якість детектування лабораторних тварин за умов нестаціонарного фону;

– удосконалено нейромережеву модель Mask R-CNN за рахунок введення подвійної магістралі екстракції ознак із керованим SE-злиттям ($B \rightarrow A$), що дозволило підвищити якість детектування та сегментації лабораторних тварин у відеоданих за умов часткових перекриттів, злиття та подальшого розділення об'єктів зі збереженням прийнятної оперативності;

– вперше запропоновано концептуальну модель наскрізної ідентифікації лабораторних тварин у відеоданих без використання ручного маркування Mask-TAUDL, яка інтегрує механізм отримання ознак на основі удосконаленої нейромережевої моделі Mask R-CNN із подвійною магістраллю екстракції ознак, масково-зважені ROI-ознаки та механізми асоціації треклетів без використання ручного маркування TAUDL, що забезпечує формування нормалізованих векторів вбудовувань і стабільне збереження ідентичності лабораторних тварин протягом відеопослідовності.

Практичне значення одержаних результатів полягає у створенні прикладних рішень на основі розроблених моделей та методів детектування, трекінгу та наскрізної ідентифікації для конвеєра автоматизованої обробки відеоданих лабораторних тварин, придатного для використання у довготривалих спостереженнях за піддослідними лабораторними тваринами за умов нестаціонарного фону, оклюзій, злиття та розділення об'єктів. Удосконалені в роботі методи та моделі, а саме метод детектування лабораторних тварин у відеоданих шляхом використання методу суперпiксельної сегментації SLIC, удосконалений метод суперпiксельної сегментації SLIC, удосконалена нейромережева модель Mask R-CNN та розроблена концептуальна модель наскрізної ідентифікації лабораторних тварин у відеоданих без використання ручного маркування Mask-TAUDL дозволили підвищити якість детектування, локалізації та сегментації

лабораторних тварин і забезпечити стабільність збереження їх ідентичності. Це напряду впливає на якість відновлення траєкторій рухів лабораторних тварин і подальше обчислення поведінкових метрик. На базі запропонованих моделей та методів розроблено спеціалізований застосунок на основі інтегрованого конвеєра детектування, трекінгу та наскрізної ідентифікації лабораторних тварин у відеоданих, який використовується у лабораторних дослідженнях, що проводяться на базі біологічного факультету Одеського національного університету імені І. І. Мечникова.

Дисертацію виконано згідно тематичних планів НДР Національного університету «Одеська політехніка» за період 2024 – 2025 рр. в рамках держбюджетної теми № 264-73 – «Методи та технології моделювання і забезпечення якості складних програмних систем у динамічних середовищах» (№ держреєстрації 0125U003655).

Результати роботи впроваджено у навчальний процес Національного університету «Одеська політехніка» та Одеського національного університету імені І. І. Мечникова.

Особистий внесок здобувача. У роботах, виконаних у співавторстві, здобувачу особисто належать постановка та формалізація прикладної задачі відеоспостереження за лабораторними тваринами; обґрунтування вимог до базового детектора з урахуванням локального навчання та обмежень обчислювальної потужності обладнання; розробка методу детектування лабораторних тварин із використанням удосконаленого методу суперпіксельної сегментації SLIC і контекстної фільтрації [1-4, 7-10]; удосконалення нейромережевої моделі Mask R-CNN за рахунок введення подвійної магістралі екстракції ознак із керованим SE-злиттям ($B \rightarrow A$) [5, 11]; розробка концептуальної моделі наскрізної ідентифікації лабораторних тварин у відеоданих без використання ручного маркування Mask-TAUDL [6, 12]; проведення експериментальних досліджень, підбір метрик оцінювання та аналіз отриманих результатів.

Публікації. За матеріалами дисертаційної роботи опубліковано 12 наукових праць, в тому числі 6 статей в наукових виданнях згідно зі списком наукових фахових видань МОН України, 3 з яких включено до міжнародної наукометричної бази Scopus, та 6 матеріалів конференцій.

Апробація результатів дисертації. Основні положення і практичні результати дисертаційної роботи доповідалися та одержали схвалення на таких конференціях: XXI Міжнародна науково-технічна конференція «Проблеми інформатики і моделювання» (ПІМ-2021) (09-14 вересня 2021 р., Кароліно-Бугаз), XII Міжнародна науково-практична конференція «Інформаційні управляючі системи і технології» (ІУСТ-Одеса-2024) (23-25 вересня 2024 р., м. Одеса), IX Міжнародна науково-практична конференція «Project, Program, Portfolio Management» (РЗМ-2024) (06-07 грудня 2024 р., м. Одеса), XV Міжнародна наукова конференція студентів та молодих вчених «Сучасні інформаційні технології – 2025» (Modern Information Technology - 2025) (15-16 травня 2025 р., м. Одеса), XIII International Scientific-Practical Conference «Information Control Systems and Technologies» (ICST-ODESA 2025) (м. Одеса, 2025 р.), XI Міжнародна наукова конференція «Інформатика. Культура. Техніка» (м. Одеса, 16-17 жовтня, 2025 р.).

Структура і обсяг роботи. Дисертація складається зі вступу, п'яти розділів, висновків, списку використаних джерел і додатків. Повний обсяг дисертації складає 205 сторінок, в тому числі: 150 сторінок основного тексту, 22 рисунки, 11 таблиць, список використаних джерел зі 139 найменувань і 5 додатків.

1 АНАЛІЗ МЕТОДІВ ТА ПІДХОДІВ ДО ДЕТЕКТУВАННЯ, ТРЕКІНГУ ТА ЗБЕРЕЖЕННЯ ІДЕНТИЧНОСТІ ЛАБОРАТОРНИХ ТВАРИН У ВІДЕОДАНИХ

1.1 Аналіз систем відеомоніторингу поведінки лабораторних тварин та умов спостереження

Задачі дослідження поведінки піддослідних тварин у лабораторних умовах традиційно спираються на спостереження за рухом та взаємодіями об'єктів у часі. Відеореєстрація є одним із найзручніших способів отримання первинних даних, однак практична цінність таких даних проявляється лише тоді, коли їх можна перетворити на кількісні показники: траєкторії, характеристики активності, час перебування в певних ділянках середовища, факти контакту або зближення між об'єктами тощо. Саме тому в системах автоматизованого моніторингу поведінки ключовими стають процедури виявлення об'єктів у кадрі та їх відстеження у відеопотоці з формуванням траєкторій і збереженням їхньої ідентичності у часі [1]. Прикладами таких систем є пропріетарні комерційні платформи на кшталт EthoVision, а також прикладні фреймворки для збору та онлайн-обробки потоків даних, що зменшують поріг інтеграції відеоаналізу в експеримент [2, 3].

У загальному випадку функціонування подібних систем можна описати як послідовність етапів (рис. 1.1): отримання відеоряду (ключові елементи: камера, роздільна здатність, частота кадрів, кут огляду); попередня обробка (нормалізація, фільтрація шуму, корекція освітлення та контрасту); детектування або сегментація об'єктів інтересу (локалізація у вигляді обмежувальних рамок, масок або інших представлень); відстеження (трекінг) об'єктів (асоціація детекцій між кадрами та побудова траєкторій); обчислення поведінкових метрик.

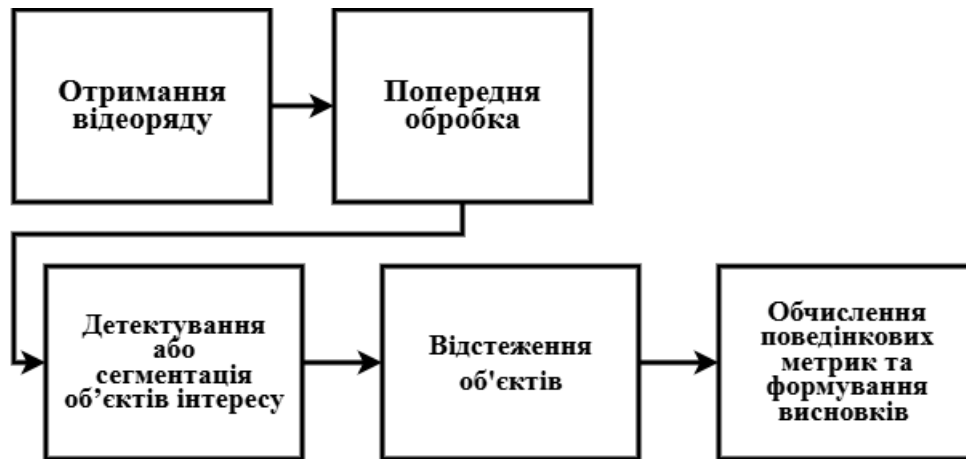


Рисунок 1.1 – Загальна схема систем автоматизованого спостереження

В той же час у сучасних системах окремі етапи цього конвеєра можуть деталізуватися або доповнюватися проміжними рівнями представлення, що підвищують точність аналізу поведінки та стійкість до складних умов спостереження.

Так, у сучасних підходах дедалі частіше використовується оцінювання пози без маркування як проміжний рівень представлення, який забезпечує більш інформативний опис руху та пози тварини порівняно з використанням центроїдів або обмежувальних рамок (bounding box) [4-7]. Зокрема, показано ефективність DeepLabCut для безмаркерного відстеження частин тіла, а також розвинуто підходи до оцінки поз у випадку багатьох тварин та збереження їхньої ідентичності у складних взаємодіях [8, 9]. У ряді сценаріїв до конвеєра між етапами трекінгу та оцінювання поведінки додають етап збереження ідентичності (Re-ID) для довготривалих спостережень або для ситуацій частих перекриттів і злиттів об'єктів [8, 10]. Практичну значущість цієї складової добре ілюструють системи, що орієнтовані на відстеження багатьох тварин із підтриманням їхньої ідентичності в групах: від idtracker.ai до TRex, які прямо націлені на задачі колективної поведінки, оклюзій та взаємодій [11-15].

За результатами відеомоніторингу зазвичай визначають такі типові метрики:

- характеристики траєкторії (довжина шляху, швидкість, прискорення, зміни напрямку руху);
- часові показники (час перебування у визначених зонах, частота відвідування зон, латентність до входу/виходу);
- показники взаємодії (контакт, зближення, одночасне перебування в зоні, переслідування тощо);
- агреговані показники активності у вибраних часових інтервалах (добові цикли, реакції на зовнішній вплив, зміни між серіями експериментів).

Перелік таких показників визначається конкретним протоколом експерименту, однак у всіх випадках їх обчислення критично залежить від якості детектування, трекінгу та збереження ідентичності об'єктів у часі, оскільки похибки локалізації, пропуски або перемикання ідентичності безпосередньо спотворюють поведінкові висновки. У межах сучасних конвеєрних підходів ці метрики дедалі частіше обчислюються не лише з траєкторій, а й шляхом автоматичної класифікації поведінкових актів на основі треків або позових ознак. Для цього застосовують як інтерактивне навчання класифікаторів (JAABA), так і повні пакети з ML-класифікацією (SimBA), а також системи розпізнавання соціальної поведінки тварин або об'єктів спостереження із поєднанням оцінки пози та класифікації поведінки (MARS) [16-18].

1.1.1 Контрольовані відеосцени та постановки задач спостереження

Характерною особливістю лабораторних досліджень є те, що об'єкти спостереження розміщуються у спеціально підготовленому середовищі, яке має спростити реєстрацію поведінки та зменшити вплив зовнішніх факторів. При цьому постановка задачі зазвичай формулюється через набір подій або характеристик, які необхідно визначити: траєкторії руху, переміщення між визначеними зонами, час перебування в окремих частинах арени, частоту

контактів між об'єктами тощо. У практиці досліджень на гризунах такі постановки часто підтримуються стандартними інструментами відеотрекінгу, що дозволяють систематизувати фенотипування поведінки в експериментах [2].

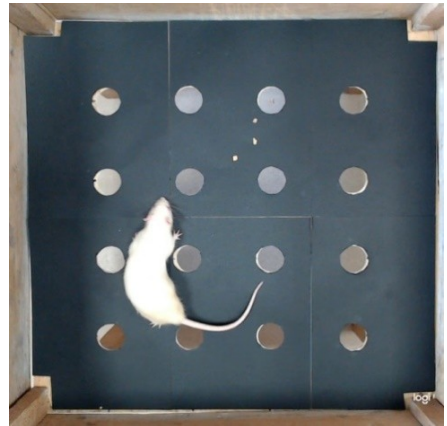
У роботах, присвячених огляду методів детектування та трекінгу, як наочні приклади контрольованих сцен наведено, зокрема, експерименти з гризунами у тестових аренах зі структурованою поверхнею та експерименти з рибами в акваріумі з камерою, встановленою зверху (рис. 1.2) [1]. Для першого випадку задачі можуть включати фіксацію переміщень між характерними точками, оцінювання часу перебування у центральній зоні та реєстрацію моментів контакту між кількома тваринами. Для другого випадку постановка задачі формально подібна: відстеження траєкторій і реєстрація контактів, але з урахуванням специфіки водного середовища та фонового шуму, що з'являється у відео [8, 10]. У випадку високопродуктивних експериментів на водних моделях, зокрема личинок зебрафіша (данио-рерио), застосовуються спеціалізовані рішення для автоматичного трекінгу і категоризації рухових маневрів, орієнтовані на масове опрацювання відеоданих [19-22].

Незважаючи на контрольованість сцени, такі експерименти не є простими для автоматизованого аналізу. На відеозаписах, отриманих на практиці, часто виникають перекриття об'єктів, особливо за одночасної присутності кількох тварин, може відбуватись зміна видимого контуру та масштабу об'єкта, а також виникати складності, пов'язані з фоновими ефектами, такими як нерівномірне освітлення, відблиски, шум, локальні неоднорідності поверхні. У випадку, коли контрольована сцена представлена акваріумом, додатковими чинниками є оптичні артефакти водного середовища та нестаціонарність фону. Саме тому навіть у контрольованих умовах до алгоритмів висуваються вимоги, близькі до вимог задач відстеження декількох тварин: забезпечення коректної роботи в умовах

взаємодій, оклюзій, часткового злиття та розділення об'єктів, а також збереження ідентичності індивідів у часі [9, 11, 12].



а)



б)



в)



г)

Рисунок 1.2 – Приклади контрольованих сцен: а) тестова арена для спостереження за мишами; б) тестова арена для спостереження за щурами; в) акваріум для спостереження за бичками; г) сцена спостереження за курами на фермі [23]

1.1.2 Вимоги до алгоритмів детектування, трекінгу та збереження ідентичності об'єктів у складних умовах спостереження

Ефективність детектування й трекінгу істотно залежить від того, які візуальні ознаки використовуються як основа для відокремлення об'єкта від фону, а також від стабільності самого фону у часі (рис. 1.3). У загальному

випадку на складність задачі впливають такі чинники, як неповнота візуальної інформації при проєкції тривимірного об'єкта на площину кадру, часткові або повні оклюзії, зміна форми й масштабу об'єкта, складні та важко прогнозовані траєкторії, а також зміни освітлення і шум у зображенні [1]. Додатково, у сценах з багатьма тваринами зростає роль коректного розв'язання задачі збереження ідентичності об'єктів; саме цю проблему цілеспрямовано розглядають і вирішують сучасні системи відстеження багатьох тварин зі збереженням ідентичності індивідів у групах [11, 12].

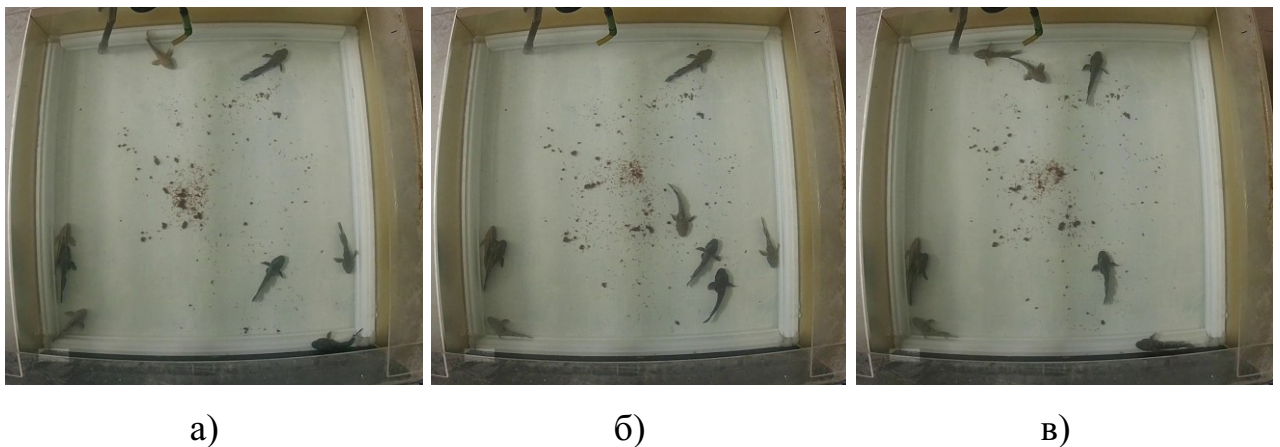


Рисунок 1.3 – Приклад поступової зміни заднього фону контрольованого середовища: а)-в) послідовні кадри відеоданих

З прикладної точки зору, у системах моніторингу поведінки тварин виділяють кілька груп вимог до алгоритмів детектування, трекінгу та збереження ідентичності об'єктів, а саме:

1) вимога до точності локалізації, достатньої для обчислення траєкторних і зональних показників, тобто помилка координат не повинна спотворювати реальні поведінкові відмінності;

2) вимога стійкості до змін фону, освітлення та шуму, оскільки похибки накопичуються протягом тривалих серій;

3) вимога збереження ідентичності об'єктів у часі, тобто мінімізація перемикань ідентичності, оскільки саме з цієї відповідності формується інтерпретована траєкторія кожної тварини.

4) вимога прийнятної обчислювальної складності алгоритму для оперативної обробки відео в лабораторних умовах [8, 10].

Виконання цих вимог сприяло появі інструментів «швидкого трекінгу» для лабораторних експериментів, наприклад ToxTrac, а також універсальних фреймворків для збору й обробки даних у реальному часі [2, 3].

Окремо слід зазначити, що для детектування у відеопотоці існують різні підходи: в одних підходах повне виявлення об'єкта здійснюється лише на першому кадрі з подальшим супроводом, що зменшує обчислювальні витрати, але підвищує ризик накопичення помилки, в інших – детекція виконується на кожному кадрі з подальшою асоціацією міжкадрових відповідностей, що потенційно підвищує надійність, але збільшує вимоги до швидкодії [4]. У задачах довготривалого поведінкового моніторингу вибір між цими підходами є принциповим, оскільки помилки трекінгу мають тенденцію накопичуватися і без корекції призводять до зниження точності оцінювання та порушення збереження ідентичності об'єктів у часі.

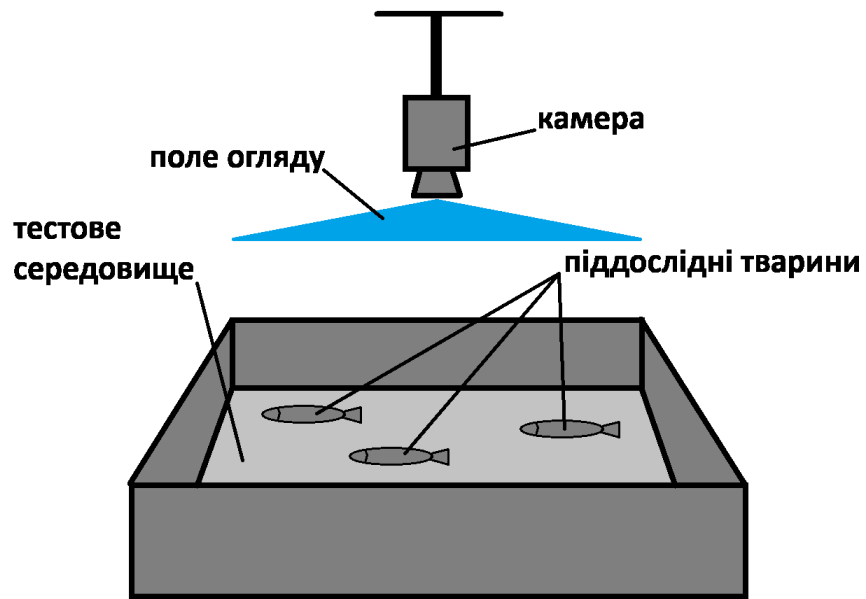
Таким чином, навіть у контрольованих лабораторних сценах задача автоматизованого відеомоніторингу поведінки тварин потребує методів, які одночасно забезпечують: достатню якість локалізації, стійкість до змін фону та освітлення, надійне відстеження із збереженням ідентичності об'єктів у часі та прийнятну обчислювальну складність для практичного застосування. З урахуванням цих вимог у наступних підрозділах розглядаються підходи до детектування та трекінгу, що історично використовувалися в комп'ютерному зорі, а також сучасні нейромережеві моделі детектування та інструменти аналізу руху об'єктів, які можуть бути інтегровані в конвеєр лабораторного спостереження [4-8, 10, 24-31].

1.2 Аналіз класичних підходів до детектування та трекінгу в контрольованих сценах

1.2.1 Попередня обробка відеоданих, моделювання фону та виділення об'єктів

У задачах відеомоніторингу поведінки лабораторних тварин якість кінцевого результату значною мірою залежить не лише від характеристик детектора або трекера, а й від етапів попередньої обробки та виділення об'єктів інтересу з відеоряду. У контрольованих сценах (камера розташована зверху, геометрія арени, клітки або іншого контрольованого середовища є відносно сталою, межі поля спостереження відомі) ці етапи часто реалізуються як модуль нормалізації зображення та відділення переднього плану (англ. foreground) від фону (англ. background). Саме така декомпозиція (рис. 1.4) дозволяє спростити подальше детектування й зменшити кількість хибних спрацювань, зумовлених змінами освітлення, шумом або незначними фоновими змінами [32, 33].

Попередня обробка включає перетворення кольорового зображення (RGB) у зображення в градаціях сірого або іншу колірну модель (HSV, YCbCr), нормалізацію яскравості та контрасту, пригнічення шуму та підсилення структурних елементів (контурів, локальних екстремумів) [34]. Вибір представлення зображення залежить від умов зйомки: наприклад, при суттєвих змінах освітлення або наявності відблисків корисним може бути перехід до представлень, де яскравісна складова відокремлена від кольорової, або використання стійких нормалізаційних перетворень [32]. На практиці ці операції реалізуються типовими засобами бібліотек комп'ютерного зору, зокрема фільтрацією, гістограмною нормалізацією, перетворенням колірних просторів тощо, і становлять важливу частину прикладних конвеєрів, побудованих на базі OpenCV та інших подібних інструментів [35-40].



а)



б)

Рисунок 1.4 – Приклад організації відеомоніторингу поведінки лабораторних тварин: а – схематичне зображення піддослідного середовища на прикладі акваріуму; б – реальне піддослідне середовище

Для подальшого виділення об'єкта після етапу попередньої обробки у вигляді окремої області або маски застосовують підходи, які умовно можна поділити на:

- 1) порогові та контурні методи;
- 2) методи на основі різниці кадрів і моделі фону;
- 3) методи сегментації з подальшою постобробкою.

У найпростішому випадку, за сталих умов освітлення та фону застосовується порогова сегментація: пікселі відносять до об'єкта, якщо їх інтенсивність або колір відрізняється від фону більше ніж на заданий поріг. Після цього виконують виділення контурів або зв'язних компонент, визначаючи геометричні характеристики об'єкта, такі як площа, периметр, центроїд, обмежувальна рамка. Ускладнення реальних умов, зокрема наявність тіней, локальних перепадів освітленості та неоднорідності поверхні, швидко знижує стабільність такого підходу, однак навіть за таких умов він може бути ефективним як частина каскаду, наприклад для попереднього грубого відділення областей інтересу перед застосуванням більш складних алгоритмів [32, 41].

Більш стійкою альтернативою є моделювання фону та виділення переднього плану як відхилення від фонові моделі. У лабораторних відеозаписах фон часто є квазістаціонарним, однак у ньому присутні певні коливання, зокрема мерехтіння освітлення, слабкі рухи підстилки, шум матриці, відблиски та коливання води в акваріумних контрольованих сценах. Тому практично важливими стають адаптивні моделі, що оновлюють оцінку фону з часом. У класичному варіанті використовують різницю з поточним або усередненим фоном, після чого виконують пороговання та морфологічну постобробку, зокрема усунення дрібних шумових компонент, «закриття» розривів контуру та стабілізацію маски) [33, 42]. Узагальнений огляд традиційних і сучасних підходів до моделювання заднього фону та детектування переднього плану подано в роботі [33], де підкреслюється, що

ключовими джерелами помилок є зміни освітлення, динамічний фон, тіні та періодичні фонові рухи.

Особливу роль у системах із верхнім розташуванням камери відіграє локальна корекція освітленості. На практиці навіть у контрольованому лабораторному середовищі освітлення може змінюватися з часом під час довготривалих серій запису або бути нерівномірним у просторі, коли краї арени є темнішими чи світлішими. У таких випадках фонове відділення корисно поєднувати з нормалізацією, наприклад, через оцінку повільно змінної компоненти освітленості, і вже на нормалізованих кадрах виконувати відокремлення переднього фону. Такий підхід дозволяє зменшити кількість хибних детекцій, стабілізувати форму маски та підвищити надійність подальшої асоціації об'єктів у трекінгу [32, 33].

У випадку, коли об'єкт віділено у вигляді маски або області, стає актуальним питання точної локалізації та підготовки ознак для подальшого трекінгу. Для цього використовують як методи виділення границь (для знаходження контуру), так і детектори локальних структур. Наприклад, для оцінки контурів широко застосовують детектор границь Кенні [42], а для виявлення плямоподібних (англ. blob-like) структур – детектори на основі LoG (Laplacian of Gaussian) [43]. У практичних сценах ці методи рідко використовуються в чистому вигляді. Частіше вони є частиною послідовного конвеєра, де спочатку формується приблизна область об'єкта, потім уточнюється контур і границя, а на завершальному етапі виконується постобробка для зменшення фрагментації. У задачах із частими контактами та перекриттями, наприклад за одночасної присутності кількох тварин в одній зоні, важливим є розділення злитих областей, для чого можуть застосовуватися сегментаційні перетворення, зокрема підхід водорозділу (англ. Watershed) [44] або методи, що спираються на контурні чи ієрархічні подання сегментації [35].

Окремим практичним аспектом є оперативність обробки: у контрольованому лабораторному середовищі часто потрібна або обробка

майже в реальному часі, або швидка пакетна обробка великих обсягів записів. Тому класичні схеми, що поєднують попередню обробку, віднімання фону та подальший аналіз простої геометрії об'єкта, й досі широко застосовуються як базовий рівень, на який накладаються складніші методи. З точки зору реалізації ці підходи добре підтримуються типовими інструментами OpenCV, що робить їх зручними для інтеграції у прикладні системи відеоспостереження [35-37]. Разом із тим, ефективність таких методів є обмеженою за умов нестаціонарного фону, змін освітлення та частих оклюзій. Саме тому в сучасних системах усе частіше переходять до комбінованих рішень, де сегментація та виділення переднього плану використовуються як допоміжний модуль для підвищення стійкості детектування та трекінгу [33, 42].

На основі розглянутого можна зазначити, що класичні підходи попередньої обробки та моделювання фону забезпечують базу для детектування в контрольованих лабораторних сценах завдяки простоті, швидкодії та зрозумілості налаштувань. Водночас за ускладнення сцени, зокрема за наявності динамічного фону, оклюзій та/або неоднорідного освітлення, зростає роль поєднання цих методів із більш інформативними ознаками та алгоритмами трекінгу, які здатні компенсувати помилки сегментації й забезпечити стійке збереження об'єктів у часі.

1.2.2 Детектування об'єктів за ознаками та шаблонами

Після етапів попередньої обробки та виділення переднього плану (п. 1.2.1) у класичних конвеєрах комп'ютерного зору виконується етап детектування об'єктів, який зазвичай зводиться до пошуку характерних структур у зображенні – контурів, кутів, локальних максимумів і мінімумів, прямоподібних структур або до зіставлення з наперед заданим шаблоном. Такі підходи застосовуються в задачах виявлення та локалізації об'єктів у контрольованих сценах завдяки відносній простоті реалізації та низькій

обчислювальній складності, що робить їх придатними для обробки відеопотоку [10, 35-37]. У лабораторному моніторингу тварин вони застосовуються насамперед як (а) базовий рівень детектування при стабільному фоні, (б) допоміжний модуль для ініціалізації трекінгу, або (в) механізм формування кандидатних областей (ROI) перед більш складною обробкою [32, 45].

Контурно-орієнтоване детектування у класичних підходах базується на припущенні, що на границі об'єкта відбувається явно виражена зміна інтенсивності пікселів, яку можна виявити за допомогою градієнтних операторів. На відміну від порогової обробки за інтенсивністю або простого віднімання заднього фону, контурні методи націлені не лише на відокремлення переднього плану, а й на уточнення геометрії об'єкта, зокрема його меж, силуету та форми, що надалі використовується як основа для локалізації (обмежувальна рамка, контур або маска) та для побудови ознак трекінгу. Тому на практиці контурні методи рідко виступають самостійним детектором: частіше вони працюють як уточнювальний етап у конвеєрі, де після попереднього грубого виділення області об'єкта виконується деталізація його меж [32, 33, 46].

Типовим представником класичних методів детектування границь є алгоритм Кенні, який дозволяє сформувати карту границь із подальшим виділенням зв'язних компонент та апроксимацією границь [42]. Водночас, у задачах підготовки ознак для міжкадрового зіставлення та трекінгу доцільно доповнювати контурні ознаки детекторами локальних структур. Зокрема, детектор Гарріса забезпечує виявлення кутових точок, які є відносно стійкими до малих змін зображення та можуть використовуватися як опорні точки для подальшого відстеження [35]. У прикладних схемах ці точки служать або для уточнення положення об'єкта, або як набір надійних ознак, що підтримують трекінг у моменти локального погіршення маски чи контрасту. Близькою за призначенням є процедура відбору «добрих точок для трекінгу» (детектор Ши-Томасі, англ. Shi-Tomasi), яка часто

використовується як більш практичний варіант ініціалізації набору опорних точок у відео [36].

Однак у лабораторних відео контурний і кутовий сигнал може бути неповним через тіні, відблиски, нерівномірну освітленість, зернистість або часткові оклюзії. Це призводить до розривів контуру, втрати частини опорних точок або появи хибних структур на фоні. Тому такі ознаки зазвичай комбінують з постобробкою, зокрема морфологічним замиканням дрібних розривів, фільтрацією шумових компонент, згладжуванням меж та з додатковими критеріями відбору областей і точок за геометричними та статистичними ознаками, що дає змогу стабілізувати локалізацію в часі [32, 33]. У підсумку контурно-орієнтований підхід у поєднанні з детекторами локальних структур, зокрема детекторами Гарріса та Ші-Томасі, забезпечує підвищення точності локалізації при відносно невеликій обчислювальній складності. Водночас якість контурних і кутових ознак суттєво залежить від стабільності освітлення та контрасту, тому на практиці їх поєднують з іншими стратегіями детектування і трекінгу в задачах поведінкового моніторингу тварин [32].

Ще одним важливим класом класичних підходів є методи на основі ключових точок і дескрипторів (ознакові методи), які дають змогу встановлювати відповідності між фрагментами зображення в сусідніх кадрах навіть за наявності помірних змін масштабу, повороту та освітлення. Базовим і найвідомішим представником цього напрямку є метод SIFT (Scale-Invariant Feature Transform), у якому поєднано пошук характерних ключових точок у масштабному просторі та побудову інваріантних дескрипторів для подальшого зіставлення [47]. У практичних системах комп'ютерного зору, окрім SIFT, широко застосовують альтернативні методи, які відрізняються компромісом між стабільністю роботи й обчислювальною складністю: метод SURF пропонує прискорене виділення ознак і дескрипторів (орієнтація на швидкість при збереженні робастності) [48]; метод KAZE використовує нелінійний масштабний простір, що краще зберігає межі об'єктів і є

корисним у сценах із вираженою структурою [49]; метод ORB є ефективним варіантом із бінарними дескрипторами, орієнтованим на швидкодію та застосування у реальному часі [50]. Ці методи реалізовані та добре підтримуються в поширених бібліотеках комп'ютерного зору, що спрощує їх інтеграцію у прикладні конвеєри [51], а їх використання розглядається у межах загальних підходів до формування та зіставлення ознак у цифровій обробці зображень [32, 52].

У задачах відеомоніторингу поведінки тварин розглянуті підходи зазвичай застосовують не як самостійний детектор тварини, а як інструмент для ініціалізації відповідностей між кадрами; повторної локалізації об'єкта після короточасної втрати; підтримки трекінгу у сценаріях з частковими оклюзіями або зниженням контрасту [32]. Водночас у лабораторних відео з верхньої камери, особливо під час спостереження за гризунами, можливості ключових точок часто обмежені: тварина може мати слабо виражену текстуру, а в кадрі домінують однорідні ділянки та плавні переходи інтенсивності, через що кількість стабільних ключових точок зменшується, а їх повторюваність між кадрами погіршується. Тому методи на основі ключових точок і дескрипторів у чистому вигляді рідко забезпечують достатню стабільність роботи для повного циклу детектування, однак залишаються корисними як допоміжний механізм, наприклад, для стабілізації траєкторії в моменти деформації маски, часткового перекриття або тимчасових змін освітлення [32].

У контрольованих сценах із відносно передбачуваним виглядом об'єкта застосовують також метод зіставлення із шаблоном (template matching). Суть підходу полягає у пошуку в кадрі такої позиції та, у розширених варіантах, масштабу та/або повороту, де кореляційна міра між шаблоном і підвінком зображення є максимальною. Підходи зіставлення із шаблоном широко представлені як у теоретичних оглядах, так і в прикладних роботах [53, 54], а також у різних практичних реалізаціях [55]. У задачах лабораторного моніторингу цей метод може працювати задовільно для ситуацій, де форма

та/або вигляд тварини не змінюється радикально, а фон є достатньо стабільним. Однак у реальних записах положення тіла змінюється, відбуваються деформації силуету, а також змінюються умови освітлення – усе це знижує надійність жорсткого шаблону. Як наслідок, зіставлення із шаблоном часто використовується як складова комбінованого конвеєра, наприклад, для уточнення позиції після попереднього грубого виділення області об'єкта або для повторної локалізації після короточасної втрати [54, 56-59].

Практичний досвід застосування класичних методів у відеозадачах показує, що їх сильною стороною є швидкодія і простота реалізації, а слабкою – обмежена стійкість до варіативності вигляду об'єкта та до складних умов зйомки. У контексті поведінкового моніторингу тварин це проявляється у двох типових випадках. Перший – це помилкове «прилипання» до фону, коли через тінь та/або відблиск межа об'єкта стає нечіткою і алгоритм частково включає фон до маски. Другий – втрата об'єкта або сплутування під час контакту, коли кілька тварин зливаються в одну область або контури перекриваються. Саме ці проблеми надалі стимулювали розвиток більш стійких підходів, зокрема методів трекінгу з використанням рухових моделей, оптичного потоку та кореляційних фільтрів, які частково компенсують нестабільність детекції на окремих кадрах [60, 61]. Таким чином, контурні підходи, методи ключових точок і дескрипторів (ознакові) та зіставлення із шаблоном формують базовий інструментарій класичного детектування, який може бути ефективним у контрольованих лабораторних сценах за умови стабільного фону та помірної варіативності вигляду об'єкта. Водночас, при наявності оклюзій, значних деформацій силуету та змін освітлення їх застосування потребує або складної системи евристик та постобробки, або поєднання з алгоритмами трекінгу, здатними підтримувати траєкторію навіть при тимчасовому падінні якості детекції.

Отже, контурні підходи, методи ключових точок і дескрипторів та зіставлення із шаблоном формують базовий інструментарій класичного

детектування, який може бути ефективним у контрольованих лабораторних сценах за стабільного фону та помірної варіативності вигляду об'єкта [10, 16-18, 62]. Їхні основні обмеження пов'язані з чутливістю до змін освітлення, деформацій силуету та оклюзій, що призводить до фрагментації області об'єкта, появи хибних структур або втрати стійких відповідностей між кадрами [32, 33]. У таких умовах практичну цінність мають комбіновані схеми, де детектування підтримується механізмами міжкадрового зіставлення та відстеження, здатними зменшувати накопичення помилок у часі [60, 61].

1.2.3 Класичні методи трекінгу об'єктів у відеоданих

Класичні підходи до трекінгу в комп'ютерному зорі формувалися як відповідь на практичну потребу підтримувати траєкторію об'єкта у часі за умов, коли детектування на кожному кадрі є нестабільним або надто ресурсомістким. У задачах лабораторного відеомоніторингу тварин ця логіка особливо важлива: об'єкт спостереження, тобто тварина, може змінювати позу та силует, перетинатися з іншими тваринами, частково перекриватися предметами сцени або тимчасово знижувати контраст відносно фону. Тому трекінг часто реалізують як окремий модуль, який використовує міжкадрові відповідності та моделі руху для стабілізації локалізації [32, 60].

Одним із базових підходів є оцінювання оптичного потоку – поля векторів переміщення пікселів або локальних областей між двома послідовними кадрами. Класичним представником є метод Лукаса-Канаде (англ. Lucas-Kanade), що формулює задачу як локальну мінімізацію різниці яскравостей за припущенням малих переміщень і приблизної сталої освітленості [62]. Для практичного використання у відео цей підхід часто застосовують у пірамідальному варіанті, який дозволяє працювати з більшими зміщеннями за рахунок багаторівневого представлення зображення [63]. Оптичний потік у трекінгу реалізується як супровід набору опорних

(ключових) точок (feature tracking) або як оцінка переміщення всередині області об'єкта [64]; типові реалізації широко доступні в OpenCV [65]. Важливою передумовою коректної роботи є стабільність фотометричних умов та відсутність значних неафінних деформацій на локальному фрагменті. Тому на практиці оптичний потік чутливий до зміни освітлення, до шуму, до розмиття руху, а також до ситуацій, коли локальна структура на об'єкті недостатньо виражена, тобто кількість придатних опорних точок є малою [56, 61, 63, 66].

У типових відеозадачах поширеною конкретизацією оптичного потоку є трекер Канаде-Лукаса-Томасі (англ. Kanade-Lucas-Tomasi), у якому спочатку обирають надійні точки для відстеження на основі критерію Shi-Tomasi або близьких до нього критеріїв, а далі кожну точку супроводжують методом Лукаса-Канаде (англ. Lucas-Kanade) [37, 63]. Такий підхід є швидким і добре масштабується на великі відеопослідовності, однак має суттєве обмеження: він забезпечує трекінг точок, а не об'єкта як цілісної сутності. Тому для отримання траєкторії об'єкта часто застосовують агрегування переміщень точок, наприклад, шляхом оцінювання афінного перетворення або зміщення центроїда області, і при цьому трекінг стає вразливим до втрати частини точок через оклюзії, зміну текстури чи деформацію силуету [32, 37].

Окремий напрям класичних методів представлений ядерними (англ. kernel-based) методами трекінгу, у яких об'єкт описується статистикою ознак у вікні, часто через гістограми кольору або яскравості, а положення оновлюється як задача пошуку максимуму подібності, наприклад в алгоритмі зміщення середнього (англ. mean-shift). Класичним прикладом цього підходу є трекінг на основі ядра, де оптимізація виконується у просторі положень вікна [66]. Такі методи привабливі простотою та швидкодією, але мають характерні обмеження: за появи фонових ділянок, подібних за кольором чи текстурою, або при суттєвій зміні вигляду об'єкта вони схильні до зсуву

вікна, а також вони гірше працюють у випадках часткових оклюзій та різких змін масштабу [32, 67].

З метою підвищення стійкості та зменшення дрейфу були розвинуті методи кореляційних фільтрів (англ. correlation filters), де трекінг формулюється як навчання фільтра, що дає максимальний відгук у позиції об'єкта. Важливою перевагою є можливість ефективної реалізації через перехід до частотної області, що забезпечує високу швидкодію. Класичним прикладом є підхід до адаптивних кореляційних фільтрів для трекінгу у відео [67]. Подальші розвідки цього напрямку спрямовані на врахування зміни масштабу та на підвищення стійкості за складних деформацій, зокрема через багатомасштабні механізми оцінювання [68]. Попри практичну ефективність, кореляційні фільтри також мають уразливість до тривалих оклюзій і до суттєвих змін вигляду об'єкта, оскільки в процесі онлайн-оновлення модель може адаптуватися до фону і накопичувати помилку, що проявляється як дрейф [68, 69].

У практичних сценаріях відеоаналізу широко використовують інтерактивні, тобто ініціалізовані користувачем, трекери одиночного об'єкта, які стартують після ручного задання початкової області (ROI) на першому кадрі й надалі підтримують її положення у часі. Такі трекери привабливі як швидкий інструмент для отримання траєкторії без ресурсомісткого детектування на кожному кадрі, а також як базовий компонент у напівручних або контрольованих лабораторних постановках. У прикладних реалізаціях (зокрема в екосистемі OpenCV) представлені різні варіанти таких трекерів, що відрізняються набором використовуваних ознак та способом оновлення моделі об'єкта в процесі супроводу [52, 69, 70].

До типових інтерактивних трекерів належать, зокрема, BOOSTING-трекер і MIL-трекер, які реалізують ідею онлайн-оновлення моделі об'єкта на основі поточних спостережень. Перевагою цієї групи є здатність певною мірою адаптуватися до змін вигляду об'єкта, зокрема його пози, масштабу та освітлення, однак основним ризиком залишається накопичення помилок і

поступовий зсув моделі на фон або на сусідні структури, особливо за оклюзій та різких змін руху. Зокрема, підхід багатоекземплярного навчання в онлайн-режимі (англ. online multiple instance learning) лежить в основі MIL-трекінгу [70] і пояснює його робастність у частині випадків, але при складних взаємодіях об'єктів дрейф залишається типовою проблемою [71]. Іншу підгрупу інтерактивних методів становлять трекери на основі кореляційних фільтрів, наприклад, KCF-трекер, MOSSE-трекер, де задача супроводу формулюється через пошук максимуму кореляційного відгуку для поточного шаблону або моделі. Їх ключова перевага – висока швидкодія, однак за умов тривалих оклюзій або схожості фону з об'єктом кореляційні трекери також схильні до деградації через оновлення моделі й помилкове переучування на фон [68]. Більш ресурсомісткі інтерактивні трекери, наприклад CSRT, намагаються підвищити стабільність за рахунок багатоканальних ознак і оцінювання надійності, що корисно у випадках деформації силуету або нестабільної текстури, однак такі рішення зазвичай потребують більших обчислювальних ресурсів [52, 70].

Окремо виділяється клас схем відстеження-навчання-детектування (англ. Tracking-learning-detection, TLD), у яких поєднуються трекінг, механізм навчання, тобто оновлення моделі об'єкта, та детектування для повторного захоплення. Мета таких підходів – зменшити критичний недолік багатьох інтерактивних трекерів: відсутність надійного механізму повернення об'єкта після втрати. Для контролю коректності супроводу використовують, зокрема, критерій похибки прямого-зворотного проходу (англ. forward-backward error) як індикатор потенційного збою трекінгу [72], а загальна TLD-архітектура поєднує спостереження у часі з повторним детектуванням для відновлення локалізації [73]. На практиці TLD-подібні схеми можуть бути корисними там, де об'єкт тимчасово зникає або змінює вигляд, однак їхня робастність істотно залежить від якості детектора та коректності оновлення моделі: за несприятливих умов помилки

накопичуються і проявляються як поступова деградація або хибне переприв'язування [72-74].

Практичні обмеження інтерактивних (і загалом класичних) трекерів узагальнено у роботах з порівняння алгоритмів на бенчмарках та оглядах трекінгу [60, 61]. Для лабораторних сцен із тваринами найбільш критичними чинниками є: дрейф через помилки оновлення моделі; оклюзії та злиття у сценах із кількома об'єктами; зміни масштабу та форми внаслідок деформації силуету; фотометричні зміни, зокрема, освітлення, відблиски, шум і розмиття руху; а також деградація якості в довготривалих серіях через накопичення похибок [60, 61]. Додатковим ускладненням у задачах поведінкового моніторингу тварин є висока візуальна подібність індивідів, часті контакти та перекриття, що робить навіть короточасні збої трекінгу суттєвими для коректного формування траєкторій, збереження ідентичності об'єктів у часі та подальшого обчислення поведінкових метрик [32].

1.2.4 Висновки щодо ефективності класичних методів детектування та трекінгу

За результатами узагальнення й експериментального тестування традиційних підходів у роботі [74] можна зробити висновок, що жоден класичний метод не є універсальним для задач відеомоніторингу лабораторних тварин: ефективність суттєво визначається умовами сцени (статичність фону, контраст «об'єкт-фон», наявність перекриттів і злиття об'єктів, кількість об'єктів, характер руху) та вимогами до рівня автоматизації.

Порівняння показує, що інтерактивні трекери, зокрема реалізації сімейств BOOSTING, MIL, KCF, MEDIANFLOW, MOSSE і CSRT та TLD можуть демонструвати прийнятну якість у контрольованих фрагментах і для окремих сценаріїв, а їхньою ключовою перевагою є відсутність потреби в «порожньому» задньому фоні або його оцінюванні [74]. Водночас у задачі

тривалого трекінгу тварин вони виявляються нестабільними на довгих серіях, оскільки потребують ручної ініціалізації об'єкта й є чутливими до перекриттів, злиття, накопичення помилок та перемикання ідентичності. Зокрема, TLD-подібні схеми орієнтовані на повторну локалізацію об'єкта, але на практиці можуть генерувати надлишкові хибні спрацювання і втрачати об'єкт інтересу за складної динаміки сцени [72-74].

Для статичних лабораторних постановок (фіксована камера, відносно стабільний фон) найбільш практичним компромісом є віднімання фону (за наявності еталонного фону або з його оцінюванням), оскільки воно забезпечує автоматичне виділення об'єктів за невисокої обчислювальної складності [33, 74]. У таких умовах суттєво зростає роль постобробки масок та областей: очищення шуму, усунення фрагментації, розділення злитих компонент, адже саме якість первинної сегментації безпосередньо впливає на подальшу стійкість трекінгу та коректність метрик поведінки [74].

Окрему групу становлять методи, що додають інформацію про рух: підходи трекінгу на основі ядра (зокрема MeanShift-подібних схем), оптичний потік та точковий трекінг і фільтрація. Вони не завжди забезпечують повноцінне формування траєкторії без додаткових процедур, однак можуть бути корисними як допоміжні компоненти - для уточнення переміщення, стабілізації траєкторії або оцінювання напрямку/швидкості, тобто як джерело додаткових ознак, що підсилюють базовий детектор або трекер [63, 67, 74]. Разом із тим їхні обмеження є типовими для лабораторних відео: погіршення якості при повільному русі, складність у випадках злиття об'єктів, чутливість до оклюзій та шуму відеоряду [74].

Щодо детектування (табличне порівняння в [74]), практичні сценарії демонструють, що у простих умовах (наприклад, «темна підлога – світлий об'єкт») можуть працювати й прості ознакові або порогові рішення, проте зі зростанням складності середовища підвищується потреба в комбінуванні кількох технік (сегментація/контури/ознакові підходи) для отримання достатньо стабільного виділення об'єкта [32, 74]. Саме тому в прикладному

сенсі доцільною є побудова конвеєра, де детектування забезпечує надійне виділення об'єкта, а трекінг – узгодження у часі з мінімізацією дрейфу та перемикань ідентичності.

У результаті аналізу та порівняння встановлено, що у задачах лабораторного відеомоніторингу вибір методу детектування і трекінгу визначається умовами сцени (стабільність фону, контраст «об'єкт-фон», рівень шуму, частота перекриттів і злиттів) та вимогами до автоматизації й швидкодії [33, 74]. Показано, що застосування одного класичного методу без урахування цих факторів не забезпечує достатньої стійкості: зростає ризик фрагментації області об'єкта, дрейфу трекера та накопичення помилок на довгих відеопослідовностях [60, 61]. Встановлено, що методи оптичного потоку і точкового супроводу можуть давати прийнятні результати лише за наявності достатньої кількості стабільних опорних структур та помірних переміщень; за умов оклюзій, розмиття або слабкої текстури ефективність різко знижується [56, 61, 63, 64, 74]. Кореляційні фільтри та інтерактивні трекери забезпечують високу швидкодію, однак є чутливими до тривалих перекриттів і схожості об'єкта з фоном, що призводить до деградації моделі та втрати цілі [52, 68]. Визначено, що схеми типу TLD можуть зменшувати наслідки короткочасної втрати об'єкта за рахунок повторної локалізації та контролю відмов, проте якість істотно залежить від коректності оновлення моделі й детектора; за несприятливих умов помилки накопичуються і проявляються як нестабільність супроводу [73, 74].

Таким чином, обґрунтовано доцільність комбінованого підходу, у якому первинне виділення об'єкта підтримується сегментаційною та морфологічною постобробкою, а модуль трекінгу використовує міжкадрове узгодження та рухові ознаки для зменшення дрейфу і підвищення стійкості локалізації та зниження ризику втрати ідентичності об'єктів у часі. У таблиці 1.1 наведено порівняння класичних підходів до детектування об'єктів [33, 63, 74].

Таблиця 1.1 – Порівняння класичних підходів до детектування об'єктів

Метод	Переваги	Недоліки/особливості
Сегментація зображення	– Зручний у застосуванні, коли об'єкти інтересу істотно відрізняються від фону за певним параметром. – Може використовуватися як доповнення до інших методів для виділення основних частин зображення.	– Низька універсальність. – Потребує дуже точного налаштування параметрів для кожного конкретного випадку. – Чутлива до умов освітлення.
Зіставлення з шаблоном (Template matching)	– Корисне для аналізу сцени за статичної камери та коли об'єкти виглядають майже однаково. – Може застосовуватись для детектування типових виробів на конвеєрі.	– Може давати збої при повороті, зміні масштабу або часткових оклюзіях. – Потребує додаткових кроків пошуку за наявності поворотів або зміни масштабу.
Методи ключових точок і дескрипторів (feature-based)	– Відносна інваріантність до невеликих поворотів, зміни масштабу та змін освітлення сцени. – Придатні для грубого пошуку об'єкта.	– Потребує компактного представлення характерних ознак. – Не дають змоги безпосередньо задавати об'єкт як екземпляр певного класу. – Можуть давати хибні спрацювання при динамічній зміні форми об'єкта.
Категоріальне розпізнавання (HOG / Хаар)	– HOG добре підкреслює об'єкти з великою кількістю деталей. – Підхід Хаара більш придатний для задач детекції та розпізнавання облич.	– Якщо об'єкт здебільшого однорідний і без суттєвих деталей, HOG переважно підсвічує лише контури. – Обидва підходи потребують попередньо навченого класифікатора. – Якість роботи погіршується, коли об'єкти схильні змінювати форму.

У таблиці 1.2, відповідно, наведено особливості, переваги та недоліки, виявлені на основі огляду методів трекінгу об'єктів.

Таблиця 1.2 – Порівняння класичних підходів до трекінгу об'єктів

Метод	Переваги	Недоліки/особливості
1	2	3
Трекінг за кольором (Color-based tracking)	– Автоматичне виявлення об'єкта. – Проста реалізація. – Висока швидкодія. – Добре працює з об'єктами, що змінюють форму.	– Не обробляє оклюзії та злиття об'єктів. – Потребує додаткових алгоритмів для відстеження центроїда. – Потребує ручного налаштування порогів за кольором. – Нестабільний, якщо кольори об'єктів занадто схожі.
Віднімання фону (Background subtraction)	– Подібні переваги, як і в трекінгу за кольором. – Достатньо одного порогового значення замість діапазону.	– Потребує наявності зображення фону. – Інші проблеми аналогічні трекінгу за кольором.
Оцінювання фону (Background estimation)	– Корисне, якщо «порожнього» фону немає. – Може застосовуватись як доповнення до віднімання фону.	– Об'єкти, що рухаються мало або повільно, можуть класифікуватись як фон. – Ефективне лише за статичності фону в цілому.
Трекінг на основі ядра (MeanShift і CAMShift)	– Досить висока швидкодія. – У CAMShift область ROI може адаптувати розмір у процесі.	– У MeanShift ROI має фіксований розмір. – CAMShift залежить від заданого діапазону кольорів. – CAMShift чутливий до освітлення та змін форми об'єкта.
Оптичний потік (Optical flow)	– Добре працює за статичного фону. – Об'єкт може визначатися автоматично за рухом. – Працює прийнятно при зміні форми об'єкта.	– При оклюзії потрібне повторне виявлення об'єкта. – Об'єкт губиться, коли його рух сповільнюється. – Не вирішує проблему злиття об'єктів.
Точкове відстеження (фільтр Калмана)	– Добра якість при трекінгу одиночних об'єктів на шумних зображеннях.	– Складні обчислення та реалізація. – Погано працюють зі злиттям та оклюзіями.

Продовження таблиці 1.2

1	2	3
Інтерактивні трекиери BOOSTING, MIL, KCF	– Добра швидкодія для BOOSTING та MIL). – KCF забезпечує найвищу швидкодію. – Придатні для нестатичного фону, зокрема за рухомої камери.	– KCF частіше втрачає об'єкт, ніж MIL та BOOSTING. – Погано працюють зі злиттям та оклюзіями.
Інтерактивний трекер (TLD)	– Може обробляти оклюзії та злиття об'єктів. – Працює з нестатичним фоном, зокрема за рухомої камери.	– Дає забагато хибних спрацювань. – Схильний втрачати об'єкт інтересу.
Інтерактивні трекиери MEDIANFLOW, MOSSE, CSRT	– CSRT демонструє добру швидкодію на високих FPS. – CSRT стійкий до змін освітлення, масштабу та пози або повороту. – Придатні для нестатичного фону, зокрема за рухомої камери.	– MEDIANFLOW добре працює лише за передбачуваних і малих переміщень без оклюзій. – Методи загалом краще працюють при передбачуваному русі об'єкта. – Методи не обробляють оклюзії.

Таким чином, результати порівняння підтверджують, що ефективність класичних методів детектування та трекінгу істотно залежить від умов сцени, характеру руху об'єктів і вимог до рівня автоматизації. Для лабораторних відеоданих Найбільш суттєвими обмеженнями для задач лабораторного відеомоніторингу залишаються нестационарність фону, зміни освітлення, оклюзії, злиття об'єктів, дрейф трекера та накопичення помилок на довгих відеопослідовностях.

1.3 Нейромережеві методи детектування та пов'язані підходи для задач відеоспостереження

Перехід від традиційних алгоритмів комп'ютерного зору до нейромережевих моделей детектування об'єктів зумовлений насамперед потребою підвищити стійкість до складних і мінливих умов зйомки. На

практиці навіть у контрольованих лабораторних сценах виникають ефекти, які складно надійно описати фіксованими правилами або пороговими критеріями: нерівномірне освітлення, тіні й відблиски, зміни контрасту, часткові перекриття та деформації силуету. Нейромережеві детектори, навчені на достатньо різноманітних даних, дозволяють формувати більш стійкі ознакові представлення та підтримувати відносно стабільну якість локалізації в умовах, де ефективність класичних методів значно знижується [74, 75].

Окремою перевагою сучасних архітектур є можливість поєднати в одному конвеєрі кілька задач – локалізацію, класифікацію та, за потреби, сегментацію з формуванням масок, що є критично важливим у сценаріях із нестационарним фоном, частими контактами та перекриттям між об'єктами. При цьому практичний вибір моделі завжди є компромісом між якістю та швидкодією: двоетапні детектори часто демонструють вищу якість локалізації, тоді як одноетапні та «полегшені» архітектури забезпечують оперативність і можливість використання в системах реального часу [75-77].

1.3.1 Загальні принципи нейромережевого детектування об'єктів у відеоданих

Нейромережеве детектування об'єктів зазвичай формулюється як задача одночасного визначення класу та просторової локалізації об'єкта на зображенні. Найпоширенішим представленням локалізації є обмежувальна рамка (англ. bounding box), для якої модель повертає координати та оцінку впевненості. У відеоспостереженні цей рівень є базовим: результати детектування, отримані на окремих кадрах, надалі використовуються для асоціації у часі, побудови траєкторій і обчислення поведінкових метрик [75, 77].

Ключова перевага нейромережевих детекторів полягає в тому, що вони формують ознаки автоматично під час навчання, а не покладаються на

вручну задані правила чи пороги. Це підвищує стійкість до змін освітлення, контрасту, форми та часткових перекриттів, однак у прикладних задачах якість істотно залежить від відповідності навчальних даних реальним умовам зйомки. Тому практично важливими є точне налаштування (англ. *fine-tuning*), підбір архітектури та параметрів навчання так, щоб модель зберігала стабільність у складних кадрах із нечіткими межами та нестационарним фоном [75].

Для оцінювання якості детекторів застосовують метрики, що враховують точність локалізації, зокрема міру перетину над об'єднанням IoU (англ. *Intersection over Union*) та похідні показники на кшталт метрики середньої точності mAP (англ. *mean Average Precision*). У реальних системах паралельно контролюють швидкодію, зокрема час прогнозування моделі та кількість оброблених кадрів за секунду, а також ресурсні вимоги, оскільки на практиці завжди існує компроміс між якістю та оперативністю. Часто вибір починається з порівняння попередньо навчених моделей з наборів моделей (англ. *model zoo*) та подальшої адаптації під конкретну задачу [75, 77].

Особливість відеоданих полягає в тому, що помилки детектора мають накопичувальний ефект: пропуск об'єкта або зсув обмежувальної рамки на одному кадрі може спричинити розрив траєкторії, помилки асоціації між кадрами або зростання похибки метрик поведінки. Тому для відеоспостереження важлива не лише середня якість детекцій, а й стабільність у часі – здатність відновлювати детекцію після короткочасної втрати та мінімізувати хибні спрацювання. Це досягається як вибором архітектури, наприклад, використанням багатомасштабних ознак або наявністю сегментаційної гілки, так і постобробкою та інтеграцією з трекінгом, що є важливим для збереження ідентичності об'єктів у відеопослідовності [74, 75].

1.3.2 Двоетапні та одноетапні детектори

У сучасному комп'ютерному зорі нейромережеві детектори об'єктів найчастіше поділяють на двоетапні (англ. two-stage) та одноетапні (англ. one-stage). Такий поділ відображає різницю в організації пошуку об'єкта: двоетапні моделі спочатку формують кандидатні області (англ. region proposals), а вже потім уточнюють координати та клас; одноетапні – виконують локалізацію й класифікацію без окремого етапу формування кандидатних областей, за один прохід мережею. У прикладних відеосценах, зокрема контрольованих, цей вибір зазвичай визначається компромісом «якість-оперативність» та характером помилок, які є найбільш критичними: пропуски, хибні спрацювання, нестабільність за перекриттів [24, 75-76].

Двоетапні детектори історично асоціюються з вищою точністю локалізації та меншою кількістю хибних спрацювань у складних випадках. Класичною базовою лінійкою є сімейство R-CNN, де перехід до Faster R-CNN зробив підхід практичнішим завдяки використанню мережі пропозицій регіону (англ. Region Proposal Network, RPN) для генерації кандидатних областей без зовнішніх алгоритмів [4]. Двоетапна схема добре узгоджується з задачами, де об'єкт має складні межі, складний або нестаціонарний фон чи часткові перекриття, оскільки механізм пропозицій і подальшого уточнення дозволяє детальніше перевіряти кандидатні області. Водночас за це супроводжується зростанням обчислювальної складності: збільшується кількість операцій і час інференсу, що може бути критичним для систем, які мають працювати онлайн [4, 75-76, 78-84].

Окремим напрямом розвитку двоетапних моделей є поєднання детектування з сегментацією екземплярів (англ. instance segmentation), коли разом із обмежувальною рамкою формується маска об'єкта. Найвідомішим представником є Mask R-CNN, у якому до Faster R-CNN додається гілка прогнозування маски для кожного ROI [5]. Для практичних задач моніторингу це важливо, оскільки маска може бути інформативнішою за

обмежувальну рамку у ситуаціях, коли потрібна точніша локалізація контуру, а також коли необхідно отримувати стабільніші ознаки для подальшого трекінгу або аналізу взаємодій. Недоліком залишається те, що сегментація екземплярів, як правило, підвищує обчислювальні витрати та збільшує складність моделі, тому її використання частіше обґрунтовують випадками зі складним фоном, перекриттями або високими вимогами до точності [5, 75, 76].

Одноетапні детектори будуються за принципом щільного прогнозування: модель одразу на вихідних картах ознак оцінює ймовірність наявності об'єкта та параметри локалізації. Популярним представником цього класу є SSD, де детекція виконується на кількох масштабах карт ознак для підвищення якості на об'єктах різного розміру [6]. Найбільш відомою лінійкою одноетапних детекторів є моделі сімейства YOLO, де основний акцент зроблено на швидкодії та придатності для роботи в реальному часі [25, 26]. Для прикладних відеосистем це часто означає меншу затримку та простішу інтеграцію в потокову обробку, проте в складних умовах (малі об'єкти, часткові перекриття, низький контраст) одноетапні моделі можуть поступатися двоетапним за стабільністю локалізації та якістю класифікації [24, 76, 85-88].

На практиці для задач відеомоніторингу важливо враховувати не лише усереднену якість детекцій, а й характер помилок. Якщо критично мінімізувати пропуски та нестабільність локалізації в неоднозначних кадрах (перекриття, контакт, нечітка межа), доцільно орієнтуватися на двоетапні підходи або моделі з масками. Якщо ж ключовим є оперативність обробки та робота в реальному часі, то одноетапні детектори, особливо оптимізовані варіанти, є більш практичними. Таким чином, вибір архітектурного сімейства формує основу для подальшого підбору конкретних моделей і механізмів, які розглядатимуться в наступних підпунктах [75, 76].

1.3.3 Безанкерні підходи до детектування об'єктів та функції втрат

Класичні детектори тривалий час спиралися на схему засновану на анкерах, де модель прогнозує зсуви відносно наперед заданих «якорів» – анкерів, тобто наборів прямокутників різних масштабів і пропорцій. Практичний недолік цього підходу полягає в необхідності підбирати гіперпараметри анкерів під конкретний клас об'єктів і масштаб у кадрі, а також у зростанні кількості кандидатів, що ускладнює оптимізацію та постобробку. Як розвиток цієї ідеології з'явилися безанкерні (англ. Anchor-free) підходи, у яких локалізація виконується без явно заданих якорів: об'єкт задається або через точки (центр/кути), або через щільні регресійні оцінки координат. Для прикладних відеозадач це важливо тим, що зменшується залежність від ручного налаштування та спрощується адаптація до специфічних сцен [24, 76].

Однією з найвідоміших безанкерних ідеологій є підхід «об'єкти як точки» (англ. Objects as Points), у якому детектування зводиться до знаходження центру об'єкта на тепловій мапі (англ. Heatmap) та регресії параметрів прямокутника відносно цього центру [27]. Подальшим розвитком є CenterNet у варіанті на основі ключових точок, де об'єкт може представлятися як набір ключових точок із додатковими обмеженнями узгодженості, що дозволяє стабілізувати локалізацію та зменшувати кількість хибних кандидатів [28]. Поруч із цим існують підходи на базі ключових точок, де об'єкт задається парою кутів (CornerNet) або комбінаціями крайніх точок і центру; такі моделі спираються на ідею, що геометрія обмежувальної рамки може бути відновлена з набору характерних точок, а не з анкерів [29]. У прикладному сенсі подібні схеми можуть бути корисними, коли об'єкт має відносно компактну форму, але в кадрі присутні перекриття або зашумлені межі, через які анкерний підхід продукує надлишкові кандидати [75, 76, 89-91].

Ще один поширений безанкерний підхід – FCOS, де на кожній позиції карти ознак прогнозуються відстані до меж обмежувальної рамки (ліворуч, праворуч, вгору, вниз) та додаткова оцінка центровості (англ. *centerness*), яка пригнічує низькоякісні передбачення на периферії об'єкта [7]. На відміну від представлення на основі ключових точок, FCOS спирається на регресію відстаней до меж об'єкта і добре масштабується до різних розмірів об'єктів у поєднанні з багатомасштабними ознаками. Для задач відеоспостереження перевага полягає в зменшенні складності налаштування, оскільки відсутня потреба в підборі анкерів, а також у більш однорідній поведінці моделі на різних сценах, якщо навчальні дані покривають типові варіації умов [76].

Ефективність безанкерних детекторів значною мірою залежить від вибору функцій втрат і способу балансування простих і складних прикладів. У задачах детектування більшість позицій на картах ознак відповідають фону, через що стандартна крос-ентропія може призводити до домінування фонових прикладів у градієнтах. Для розв'язання цієї проблеми застосовують фокусну функцію втрат (англ. *Focal Loss*), яка зменшує внесок добре класифікованих прикладів і збільшує вагу складних, що підвищує здатність моделі виявляти об'єкти на фоні великої кількості негативних позицій [30, 91]. У практичних сценаріях це важливо, оскільки саме помилки на «складних» кадрах (низький контраст, часткові перекриття, нетипові пози) найчастіше впливають на стабільність подальшого трекінгу та накопичення похибки у відеоаналізі [75, 92].

Отже, безанкерні та представлені на основі ключових точок детектори можна розглядати як природний розвиток одноетапної парадигми, спрямований на зменшення залежності від ручного підбору анкерів і на підвищення ефективності навчання при великій частці фону. У прикладних задачах вибір між представленням на основі ключових точок (CenterNet/CornerNet-подібні) та регресійними схемами (FCOS) зазвичай визначається вимогами до стабільності локалізації, масштабною варіативністю об'єктів та обчислювальними обмеженнями, тоді як роль

функцій втрат (зокрема Focal Loss) є критичною для отримання стабільної поведінки детектора на реальних відеоданих [7, 27-30, 76, 91].

1.3.4 Магістральні мережі ResNet, MobileNet та піраміди ознак у нейромережевих детекторах

Незалежно від того, чи є детектор двоетапним або одноетапним, його якість значною мірою визначається тим, які ознаки формуються всередині мережі. Типова архітектура складається з двох частин: екстрактора ознак, або магістральної мережі (англ. backbone), що перетворює вхідне зображення на ієрархію карт ознак, та «голови» детектора (англ. head), що виконує класифікацію та регресію координат, а інколи й масок. Саме екстрактор задає базову інформативність представлення: якщо ознаки слабо розрізняють об'єкт і фон або погано зберігають структуру на потрібних масштабах, то навіть «сильна голова» детектора не забезпечить стабільної локалізації [75, 76].

Одним з найпоширеніших екстракторів ознак у прикладних детекторах є ResNet, у якому введено залишкові (англ. residual) з'єднання для стабільного навчання глибоких мереж і формування багаторівневих ознак [31]. Практично це означає, що ResNet-екстрактор часто дає хороший баланс між якістю і керованістю навчання, а також широко підтримується в реалізаціях детекторів, зокрема в двоетапних підходах. Водночас у задачах, де критичною є швидкодія або обмежені ресурси, наприклад під час обробки потокового відео на відносно слабкому графічному або центральному процесорі, застосовують «легкі» екстрактори на кшталт MobileNet, побудовані на глибинно-роздільних (англ. depthwise-separable) згортках і оптимізовані під ефективне прогнозування моделі [93]. У таких випадках досягається зменшення затримки й енергоспоживання, але часто ціною певного падіння точності або стійкості на складних кадрах, що потрібно враховувати при виборі моделі для відеомоніторингу [75, 76, 93-98].

Важливим чинником для детектування у реальних сценах є масштабна варіативність: об'єкти можуть мати різний розмір залежно від відстані до камери, положення в кадрі або постановки експерименту. Якщо детектор використовує ознаки лише одного рівня, то малий об'єкт може бути втрачений через значне зменшення роздільної здатності на глибших шарах, а великий – через недостатній контекст на ранніх шарах. Саме тому широко застосовують підхід пірамідної мережі ознак (англ. Feature Pyramid Network, FPN), який об'єднує ознаки з кількох рівнів і формує піраміду представлень, придатну для детектування на різних масштабах [98, 99]. У прикладному сенсі FPN підвищує стабільність детектора щодо розміру об'єкта і часто є компонентом за замовчуванням у сучасних архітектурах, особливо коли потрібна надійніша робота на малих або частково перекритих об'єктах [75, 76, 99, 100].

З практичної точки зору, комбінація екстрактора ознак і FPN є одним із основних механізмів керування компромісом «якість-оперативність». Використання потужного екстрактора (ResNet) і повної пірамідальної структури може забезпечити кращу якість локалізації, але збільшує обчислювальні витрати; натомість легкі екстрактори, зокрема MobileNet, у поєднанні з компактними варіантами пірамідальних ознак зменшують затримку й роблять систему придатнішою для оперативної обробки, але можуть бути чутливішими до складного фону або невиразних меж. Для задач відеоспостереження це особливо важливо, оскільки стабільність детектування в часі залежить не лише від високого середнього значення mAP, а й від того, наскільки добре модель стабільно локалізує об'єкт у кадрах з погіршеними умовами: низьким контрастом, перекриттями або частковою втратою видимості [75, 76].

1.3.5 Сегментаційні архітектури для підвищення точності локалізації

У задачах, де об'єкт має нечіткі межі, складний фон або часткові перекриття, представлення лише у вигляді обмежувальної рамки часто є недостатнім. Тому доцільно застосовувати підходи з піксельним представленням об'єкта - семантичну сегментацію, оскільки маска точніше відображає форму, краще відокремлює об'єкт від фону та дає інформативніші ознаки для подальших етапів, зокрема стабілізації трекінгу або аналізу взаємодій [5, 75, 76]. Навіть якщо у фінальній системі використовується обмежувальна рамка, маска може слугувати допоміжним інструментом для уточнення локалізації та фільтрації хибних або нестабільних результатів детектування.

Базовою моделлю сучасної сегментації є FCN, де замість «класу для всього зображення» виконується щільне прогнозування по пікселях із відновленням роздільної здатності через підвищення роздільної здатності (англ. upsampling) або розгортки (англ. deconvolution) [100, 101]. Для кращого збереження меж використовують архітектури з поверненням просторових деталей із ранніх рівнів: найвідомішою є U-Net зі схемою кодувальник-декодувальник та пропускними з'єднаннями, що поєднують глобальний контекст і локальні ознаки [101]. Для випадків об'ємних даних запропоновано V-Net, яка розвиває принцип повністю згорткової мережі для 3D-представлень [102]. Окремо виділяють Stacked Hourglass, де «стиснення-розгортання» повторюється кілька разів для уточнення просторових карт; хоча ці моделі часто застосовують у задачах оцінювання пози та ключових точок, їхня ідея багаторазового інтегрування контексту на різних масштабах є корисною і для сегментаційних задач [103]. Загалом сегментаційні архітектури FCN, U-Net, V-Net, а також просторові моделі уточнення, зокрема Hourglass, підвищують якість локалізації завдяки переходу від обмежувальної рамки до масок або карт ознак, хоча зазвичай це

супроводжується зростанням обчислювальних витрат і ресурсних вимог [104-109].

1.3.6 Оптимізація нейромережевих детекторів для відеомоніторингу

У системах відеоспостереження вибір нейромережевого детектора визначається не лише показниками якості, зокрема mAP та IoU, а й можливістю оперативної обробки відеопотоку (наприклад, у кадрах за секунду) за наявних обчислювальних ресурсів. Тому на практиці часто застосовують підходи, спрямовані на зменшення обчислювальних витрат без критичної втрати якості: використання компактних екстракторів ознак, зокрема MobileNet, зниження роздільної здатності вхідного зображення, скорочення кількості рівнів багатомасштабних ознак та підбір конфігурації моделі під характерний масштаб об'єктів у кадрі [76, 77, 93, 98, 99]. У прикладному сенсі це означає, що замість універсальної «найточнішої» моделі доцільно обирати варіант, який забезпечує стабільні результати детектування і, відповідно, надійні траєкторії за прийнятного часу обробки кадру.

Окремий клас рішень базується на масштабуванні архітектури та керованому доборі складності моделі. Так, у EfficientNet запропоновано підхід збалансованого масштабування глибини, ширини та роздільної здатності, що дозволяє підвищувати якість детекцій за порівнянних ресурсів [10]. На цьому принципі побудовано й EfficientDet, де узгоджено масштабуються екстрактор ознак і детекторна частина з багатомасштабним об'єднанням ознак [8]. Практична перевага таких підходів полягає у наявності лінійки моделей різної складності – від швидких до більш точних, що полегшує вибір під конкретні апаратні обмеження та вимоги до оперативності [8, 10, 75, 76].

Крім підбору архітектури, застосовують методи підготовки моделі до розгортання: спрощення структури, скорочення обчислень та спеціалізацію під задані ресурси. Наприклад, у підході «один раз для всіх» Once-for-All формується «універсальна» мережа, з якої можна отримувати підмережі різної складності під задані обмеження (швидкодія, пам'ять) зі збереженням прийнятної якості [110]. У підсумку оптимізація нейромережевих детекторів для відеозадач зводиться до керування співвідношенням між якістю детектування та прийнятною оперативністю обробки: від вибору архітектури й параметрів багатомасштабних ознак до масштабування та спеціалізації моделі під конкретну апаратну платформу [8, 10, 76, 77, 93, 99, 109, 110].

1.3.7 Нейромережеві підходи до відстеження об'єктів у відеоданих

У сучасних відеосистемах відстеження дедалі частіше реалізують не як окремий супровід об'єкта за локальними ознаками, а як конвеєр, у якому нейромережевий детектор формує результати детектування для кожного кадру, а далі виконується міжкадрове зіставлення цих результатів, тобто асоціація, та формування і підтримання траєкторій. Така схема є більш стійкою до короточасних втрат видимості, але її якість сильно залежить від стабільності результатів детектування, зокрема пропусків та хибних спрацювань, і від того, наскільки надійно працює механізм зіставлення у випадках часткових перекриттів і схожих об'єктів у сцені [60, 61]. Порівняно з класичними трекерами, що схильні до дрейфу при накопиченні похибок, детекційно-асоціативний підхід дозволяє регулярно оновлювати локалізацію об'єкта на основі поточного кадру, зменшуючи ризик поступового зміщення траєкторії [60].

Для відстеження одного об'єкта у кадрі широко розвивалися нейромережеві трекери на основі сіамських архітектур, де зіставляються ознаки еталонного фрагмента, тобто цілі, та поточного кадру. Характерним прикладом є Fully-Convolutional Siamese Networks for Object Tracking, у яких

повністю згорткова схема формує карту подібності й дозволяє ефективно знаходити ціль у наступних кадрах [111]. Практично такі підходи корисні тим, що краще узагальнюють на зміну масштабу, повороти або часткові зміни вигляду, однак проблеми оклюзій і тимчасового зникнення об'єкта часто потребують додаткових механізмів перевірки коректності відстеження та повторного захоплення. Саме тому поряд із нейромережевими трекерами в прикладних рішеннях зберігається потреба в механізмах виявлення відмов (наприклад, через критерії на кшталт forward-backward error) і в архітектурних схемах, де відстеження поєднується з навчанням моделі об'єкта та повторним детектуванням, як у підходах tracking-learning-detection (TLD) [72-74].

У практичних відеозадачах ключовим обмеженням для будь-якого трекера, зокрема нейромережевого, залишаються типові чинники деградації: перекриття та злиття об'єктів, зміни масштабу й форми, розмиття руху, зміни освітлення та дрейф при довгих серіях. Саме ці проблеми систематизовано в оглядових роботах і бенчмарках з оцінювання трекінгових алгоритмів, де показано, що «сильний» трекер у середньому може суттєво втрачати якість у ситуаціях, критичних для прикладної інтерпретації траєкторій [60, 61]. Тому на практиці доцільно поєднувати нейромережеві детектори з механізмами надійної асоціації та контролю помилок відстеження, а також (за потреби) використовувати додаткові ознаки об'єкта, які підвищують стійкість збереження ідентичності об'єктів у часі [60, 61, 75, 76, 111].

1.3.8 Висновки щодо нейромережових методів до детектування та трекінгу

Проведений аналіз нейромережових методів детектування та трекінгу показав, що практичне використання таких підходів у відеомоніторингу визначається не лише якістю детектування, а й сукупністю прикладних чинників: вимогами до обсягу навчальних даних, стабільністю навчання,

оперативністю обробки кадрів і стійкістю результатів у варіативних умовах сцени [75, 76]. Важливими при цьому є вибір між двоетапними й одноетапними детекторами, а також використання багатомасштабних ознак, які дають змогу працювати з об'єктами різного розміру та різного рівня деталізації [4, 25, 75, 76, 98, 99].

Аналіз двоетапних, одноетапних, безанкерних, сегментаційних і трекінгових нейромережових підходів показав, що жодна група моделей не є універсальною. Двоетапні моделі та архітектури з масками доцільні за складних меж, перекриттів і нестационарного фону [4, 5, 75, 76]; одноетапні й компактні моделі є практичнішими для оперативної обробки [6, 25, 26, 75, 76]; безанкерні підходи та багатомасштабні ознаки зменшують залежність від ручного налаштування й підвищують стабільність локалізації об'єктів різного розміру [7, 27-30, 98, 99]; нейромережові трекери та детекційно-асоціативні схеми потребують додаткових механізмів контролю помилок і повторної ідентифікації для збереження ідентичності об'єктів у часі [60, 61, 74-76, 111-113].

Отже, результати аналізу показують, що нейромережові методи доцільно розглядати як основу для підвищення якості детектування та сегментації об'єктів у складних умовах відеоспостереження. Їхня здатність формувати інформативні ознаки, працювати з багатомасштабними представленнями та підтримувати сегментаційне уточнення локалізації визначає їхню роль у подальшій побудові конвеєра обробки відеоданих, який має поєднувати детектування, трекінг і повторну ідентифікацію для збереження ідентичності об'єктів у часі.

1.4 Висновки до першого розділу

У першому розділі дисертації проведено ґрунтовний аналіз сучасного стану проблеми автоматизованого відеомоніторингу лабораторних тварин у контрольованих умовах спостереження. Досліджено особливості систем

відеоаналізу поведінки тварин, умови формування відеоданих, вимоги до процедур детектування, трекінгу та збереження ідентичності об'єктів у часі. Підкреслено, що автоматизоване виявлення й відстеження лабораторних тварин є основою для формування траєкторій руху та подальшого обчислення поведінкових показників, необхідних для кількісного аналізу поведінки в біологічних і лабораторних дослідженнях.

Встановлено, що класичні методи комп'ютерного зору мають переваги відносної алгоритмічної простоти, однак не є універсальними для задач довготривалого відеомоніторингу тварин. Їх ефективність суттєво залежить від умов сцени: статичності або нестационарності фону, контрасту «об'єкт-фон», перекриттів і злиття об'єктів, кількості об'єктів, характеру руху та вимог до рівня автоматизації. Основними причинами помилок є фрагментація області об'єкта після віднімання фону або порогової обробки, поява хибнопозитивних фрагментів, дрейф трекера, накопичення похибок у часі та перемикання ідентичності під час контактів або перекриттів.

Обґрунтовано доцільність використання нейромережевих методів детектування як базового інструменту локалізації лабораторних тварин у кадрі. Показано, що практичне застосування таких методів у довготривалому відеоаналізі потребує врахування не лише якості детектування, а й можливості локального навчання, обмежень обчислювальних ресурсів, стабільності навчання на невеликих вибірках і часу обробки одного кадру.

Таким чином, визначено напрям подальших досліджень, який дозволяє підвищити якість детектування та сегментації, забезпечити прийнятну оперативність обробки відеоданих, а також підвищити стійкість трекінгу та наскрізної ідентифікації лабораторних тварин у контрольованих відеосценах за рахунок розроблення інтегрованого конвеєра на основі нейромережевого детектування із сегментаційним уточненням локалізації та механізмами повторної ідентифікації, що забезпечують стійке відстеження і збереження ідентичності об'єктів за умов нестационарного фону, часткових перекриттів, злиття та розділення об'єктів.

2 МЕТОДИ ДЕТЕКТУВАННЯ ТА ЛОКАЛІЗАЦІЇ ЛАБОРАТОРНИХ ТВАРИН У ВІДЕОДАНИХ НА ОСНОВІ СУПЕРПІКСЕЛЬНОЇ СЕГМЕНТАЦІЇ SLIC ТА ЇЇ УДОСКОНАЛЕННЯ

Як було зазначено у розділі 1, процедури детектування та сегментації є одними з ключових у системах автоматизованого відеомоніторингу лабораторних тварин, оскільки саме вони формують первинне просторове представлення об'єкта, яке надалі використовується для трекінгу, формування траєкторій і збереження ідентичності об'єктів у часі [74, 75, 114]. У контрольованих лабораторних відеосценах стаціонарна камера та фіксована область спостереження поєднуються з локальними артефактами, зокрема тінями, відблисками, повільними змінами освітлення, шумом і нестационарністю фону. За таких умов нестабільність детектування на окремих кадрах, зокрема пропуски, фрагментація маски або нестійка локалізація, призводить до розривів траєкторій, збільшення кількості треклетів і ускладнення подальшої повторної ідентифікації, що безпосередньо впливає на точність похідних поведінкових метрик та практичну придатність системи [74, 75, 115].

У цьому розділі розроблено методи детектування та локалізації лабораторних тварин у відеоданих на основі суперпиксельної сегментації SLIC та її удосконаленні. Використання SLIC дає змогу перейти від нестійкого піксельного пороговання після віднімання фону до аналізу локально узгоджених областей – суперпикселів. Запропоноване удосконалення методу SLIC передбачає використання багатопланового простору ознак, який враховує LAB-представлення та представлення різницевого зображення, а також перевірку однорідності пікселів у локальному оточенні. Це спрямовано на підвищення точності локалізації та детектування лабораторних тварин за умов нестационарного фону й наявності артефактів.

2.1 Постановка задачі детектування та локалізації лабораторних тварин у відеоданих

Задача автоматизованого аналізу відеоданих у лабораторних умовах, як правило, розглядається у вигляді наскрізного конвеєра, що включає послідовні етапи: попередня обробка кадрів, детектування та сегментація, трекінг, формування треклетів, повторна ідентифікація та обчислення показників рухів. У межах цього розділу основну увагу зосереджено на ранніх етапах такого конвеєра – детектуванні та локалізації лабораторних тварин, оскільки помилки на цьому рівні безпосередньо призводять до фрагментації маски, пропусків об'єкта, розривів траєкторій, збільшення кількості треклетів і погіршення подальшого збереження ідентичності об'єктів у часі [74, 75, 114, 115].

У цій роботі об'єктом детектування є окрема лабораторна тварина в кадрі. Детектування трактується як встановлення факту наявності об'єкта та визначення його просторового положення в кадрі. Результатом детектування є локалізований об'єкт, представлений обмежувальною рамкою, а за наявності сегментаційного представлення – також маскою. Локалізація означає визначення координат області, у якій знаходиться об'єкт, у вигляді обмежувальної рамки та/або піксельної маски. ID у межах роботи розглядається не як клас або тип об'єкта, а як ідентичність конкретної лабораторної тварини у часі, що має зберігатися між кадрами та треклетами [115]. Під треклетом розуміється неперервний фрагмент траєкторії, який утворюється між моментами невизначеності (перекриття, злиття, пропуск детектування) і надалі може бути з'єднаний із іншими фрагментами за критеріями повторної ідентифікації. Загальну схему такого конвеєра наведено на рис. 2.1.

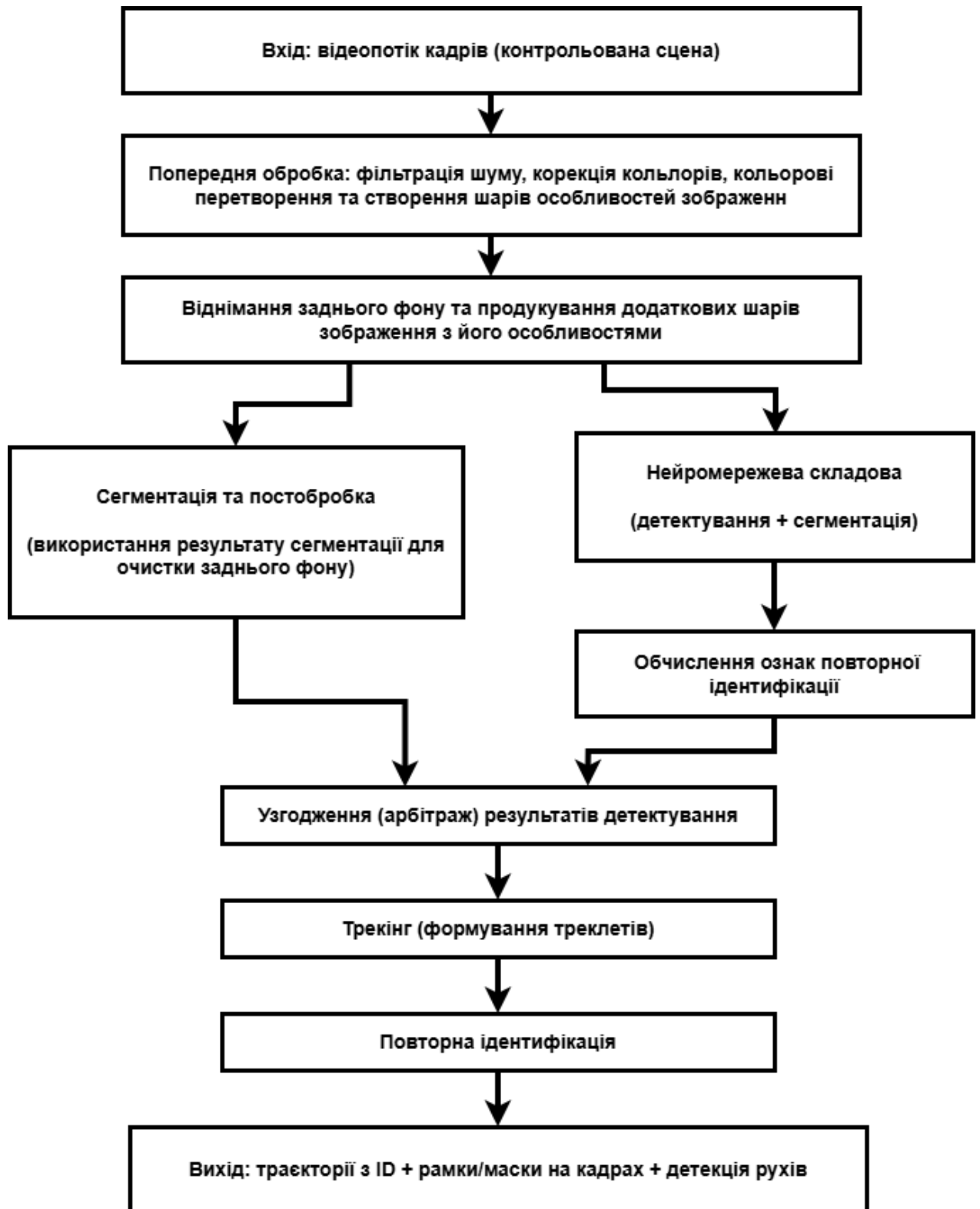


Рисунок 2.1 – Загальна схема конвеєра автоматизованого детектування, трекінгу та повторної ідентифікації лабораторних тварин у відеоданих

У межах цього розділу основну увагу зосереджено на лівій частині конвеєра, пов'язаній із попередньою обробкою кадрів, відніманням фону,

суперпіксельною сегментацією SLIC та формуванням допоміжних масок і обмежувальних рамок. Нейромережева складова конвеєра, а також етапи трекінгу, формування треклетів і повторної ідентифікації розглядається в наступних розділах.

Для подальшого опису задачі детектування та локалізації формалізуємо результат обробки окремого кадру. Нехай вхідні дані задаються відеопослідовністю кадрів $\{I_t\}_{t=1}^T$. Результат детектування на кадрі t подається множиною детекцій:

$$D_t = \{d_t^{(i)}\}_{i=1}^{N_t}, \quad (2.1)$$

де N_t – кількість детекцій на кадрі.

Кожна детекція $d_t^{(i)}$ містить обмежувальну рамку $b_t^{(i)}$, оцінку впевненості $s_t^{(i)}$, наявності – маску $m_t^{(i)}$, а також ознаковий вектор $f_t^{(i)}$, який може використовуватись для повторної ідентифікації на подальших етапах:

$$d_t^{(i)} = (b_t^{(i)}, m_t^{(i)}, s_t^{(i)}, f_t^{(i)}). \quad (2.2)$$

При перекриттях, злиттях або пропусках детектування траєкторія може перериватися, утворюючи треклети; їх подальше з'єднання та збереження ідентичності об'єктів у часі виконується на основі ознак повторної ідентифікації, що розглядається в наступних розділах роботи [115, 116]. Таким чином, вимоги до детектування визначаються не лише точністю локалізації на окремому кадрі, а й стабільністю у часі та зменшенням кількості пропусків детектування, оскільки це безпосередньо впливає на кількість треклетів і надійність збереження ідентичності [75, 114, 115].

Для задачі відеомоніторингу лабораторних тварин у контрольованій сцені приймаються такі припущення:

- камера є стаціонарною;
- область спостереження лабораторних тварин є фіксованою;
- фон є переважно статичним, проте можливі повільні зміни освітлення, тіні, відблиски та локальні артефакти;
- кількість лабораторних тварин у кадрі є обмеженою і може змінюватися;
- допускаються часткові перекриття, короткочасні злиття тварин і фрагментація області об’єкта [34, 74, 75].

За таких умов метод детектування та локалізації має зменшувати вплив локальних артефактів, забезпечувати відтворюваність результатів у довгих серіях кадрів і формувати просторове представлення лабораторної тварини, придатне для подальшого трекінгу та збереження ідентичності.

Для оцінювання точності локалізації об’єкта обмежувальною рамкою використовується коефіцієнт перетину-об’єднання Intersection over Union (IoU):

$$\text{IoU}(b, \hat{b}) = \frac{|b \cap \hat{b}|}{|b \cup \hat{b}|}, \quad (2.3)$$

де b – еталонна рамка, визначена вручну

$\hat{b}$ – рамка, отримана методом детектування.

На основі істинно-позитивних, хибнопозитивних і хибнонегативних рішень обчислюються точність (Precision) і відгук (Recall), та $F1$ -міра:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (2.4)$$

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (2.5)$$

$$F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recal}}{\text{Precision} + \text{Recal}}, \quad (2.6)$$

де TP (істинно-позитивні) – кількість випадків, коли об’єкт детектовано і локалізовано коректно (задовольняє поріг IoU);

FP (хибно-позитивні) – кількість хибних детектувань (детекція там, де об’єкта немає, або локалізація не проходить поріг IoU);

FN (хибно-негативні) – кількість пропусків (об’єкт присутній в еталоні, але детектування відсутнє або не проходить поріг IoU);

TN (істинно-негативні) – кількість випадків, коли детектування коректно відсутнє (фон без об’єкта і без детекцій).

Схему запропонованого методу детектування лабораторних тварин у відеоданих на основі суперпiксельної сегментації SLIC наведено на рисунку 2.2.

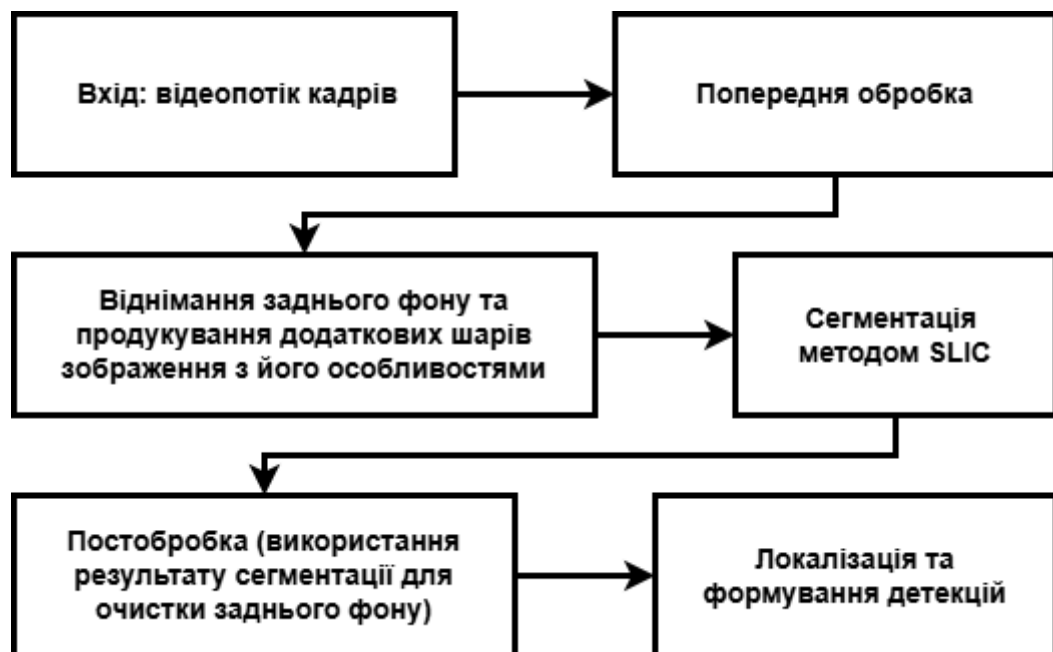


Рисунок 2.2 – Загальна схема методу детектування лабораторних тварин у відеоданих на основі суперпiксельної сегментації SLIC

Метод охоплює етапи попередньої обробки, віднімання фону, сегментації на основі SLIC, контекстної фільтрації в межах суперпикселів та формування об'єктної області у вигляді маски або обмежувальної рамки [34, 105, 117, 118]. Його призначення полягає у підвищенні стабільності локалізації лабораторних тварин за рахунок переходу від піксельного пороговання до аналізу локально узгоджених областей. Метод є відносно простим з обчислювальної точки зору, придатним для потокової обробки та керованим за параметрами, що важливо для контрольованих сцен із фоновими артефактами, тінями, відблисками та повільними змінами освітлення [34, 75]. Подальші підрозділи описують етапи попередньої обробки, віднімання фону, використання базового методу SLIC та його удосконалення для підвищення точності локалізації й детектування за умов нестационарного фону та артефактів.

2.2 Етапи попередньої обробки

2.2.1 Фільтрація шуму та нормалізація яскравості й контрасту

Стійкість детектування в контрольованій лабораторній відеосцені значною мірою визначається якістю первинної підготовки кадрів. Навіть за стаціонарної камери у відеоданих присутні шум сенсора, артефакти стиснення, локальні зміни яскравості, тіні та відблиски. Це може призводити до появи розрізнених активних пікселів після віднімання фону та до фрагментації об'єкта на етапі сегментації. Тому перед подальшими етапами виконується фільтрація шуму та нормалізація яскравості й контрасту, які зменшують вплив завад і підвищують відтворюваність детектування в довгих серіях кадрів [32, 34].

На першому етапі виконується медіанна фільтрація, яка дає змогу приглушити імпульсний шум без суттєвого розмивання меж об'єкта. Нехай $\Omega(x, y)$ – околиця пікселя (x, y) . Тоді:

$$I_{med}(x, y) = \text{median}\{I(u, v) | (u, v) \in \Omega(x, y)\}. \quad (2.7)$$

Додатково застосовується двостороння фільтрація, яка враховує просторову близькість пікселів та подібність їхніх інтенсивностей [119] і визначається як:

$$I_{bf}(x, y) = \frac{1}{W(x, y)} \sum_{(u, v) \in \Omega(x, y)} \left(G_s(x - u, y - v) G_r(I(x, y) - I(u, v)) I(u, v) \right), \quad (2.8)$$

де

$$G_s(x - u, y - v) = \exp\left(-\frac{(x - u)^2 + (y - v)^2}{2\sigma_s^2}\right), \quad (2.9)$$

$$G_r(I(x, y) - I(u, v)) = \exp\left(-\frac{(I(x, y) - I(u, v))^2}{2\sigma_r^2}\right), \quad (2.10)$$

а нормувальний множник:

$$W(x, y) = \sum_{(u, v) \in \Omega(x, y)} \left(G_s(x - u, y - v) G_r(I(x, y) - I(u, v)) \right). \quad (2.11)$$

Для стабілізації яскравості (нормалізації контрасту) в тривалих серіях кадрів використовується вирівнювання гістограми. Нехай $r \in \{0, \dots, L - 1\}$ – рівень яскравості, $p(k)$ – оцінка ймовірності рівня k . Тоді кумулятивна функція:

$$F(r) = \sum_{k=0}^r p(k), \quad (2.12)$$

а перетворення інтенсивності:

$$s = (L - 1) \cdot F(r). \quad (2.13)$$

Додатково використовуються перетворення інтенсивності з насиченням та (опціонально) вирівнювання тіней [120]. У контрольованих лабораторних відеосценах тіні можуть змінювати видимий контур об'єкта та створювати помилкові активності у різницевому зображенні. Тому за потреби вирівнювання тіней застосовується як підготовчий крок перед відніманням фону. Його призначення – зменшити вплив повільних локальних перепадів освітлення та відокремити «фонові» затемнення від реального руху об'єкта. Практично цей крок може реалізовуватися як компенсація поля освітленості (оцінка повільної компоненти яскравості та її нормалізація), що робить фон більш однорідним. Це зменшує кількість хибних спрацьовувань після порогування та стабілізує подальшу сегментацію.

У разі потреби, зокрема за малої роздільної здатності кадру або слабкої виразності меж об'єкта, може застосовуватися масштабування зображення як додаткова операція, що підвищує чіткість контурів та стабільність подальшої сегментації [32, 34].

2.2.2 Віднімання фону та порогування

За стаціонарної камери базовим кроком виділення областей активності є віднімання фону, яке формує різницеве зображення між поточним кадром і фоновим зображенням [33, 105]. Нехай $B_t(x, y)$ – фонове зображення на кадрі t . Тоді різницеве зображення визначається як:

$$\Delta_t(x, y) = |I_t(x, y) - B_t(x, y)|. \quad (2.14)$$

Після цього формується бінарна маска активності порогуванням:

$$M_t(x, y) = 1\{\Delta_t(x, y) > \tau\}, \quad (2.15)$$

де τ – порогове значення,

$1\{\cdot\}$ – індикаторна функція (дорівнює 1, якщо умова істинна, і 0 – інакше).

У довготривалих лабораторних відеосеріях фон може повільно змінюватися через зміни освітлення, тіні, відблиски або появу фонових артефактів. Практичний недолік порогоування на рівні пікселів полягає в тому, що навіть поодинокі шумові пікселі або локальні зміни освітлення можуть формувати хибні активності. Тому доцільним є перехід від піксельної маски M_t до прийняття рішень на рівні областей (сегментів), що узгоджується з наступним етапом суперпіксельної сегментації [34, 105, 117].

Для зменшення хибних спрацьовувань, спричинених локальними артефактами у різницевому зображенні, застосовується контекстна фільтрація в межах сегментів: рішення про належність області до об'єкта приймається не за окремими пікселями, а за часткою активних пікселів у сегментованій області [34]. Нехай $\{S_j\}$ – множина сегментів кадру, надалі суперпікселів, а P^{S_j} – кількість пікселів у сегменті S_j . Кількість активних пікселів у сегменті визначається як:

$$A_j = \sum_{p \in S_j} 1\{\Delta_t(p) > \tau\}. \quad (2.16)$$

Сегмент вважається таким, що належить до області об'єкта, якщо частка активних пікселів у ньому не менша за заданий поріг α . У роботі використовується правило $A_j \geq \alpha P^{S_j}$, $\alpha = \frac{1}{3}$. Це правило має дві практично важливі властивості. По-перше, воно пригнічує поодинокі активні пікселі: якщо активність не підтримується значною частиною сегмента, сегмент відкидається як шумовий. По-друге, воно узгоджує віднімання фону з

подальшою сегментацією: попередня маска формується як сукупність локально узгоджених сегментів, а не як неструктурований набір пікселів, що підвищує стабільність детектування в серіях кадрів із тінями, відблисками та оклюзіями [34, 105].

Отримані на цьому етапі представлення, а саме: нормалізоване зображення, різницеве зображення, маска активності та очищена сегментна маска, надалі використовуються як складові багат шарового простору ознак для удосконаленого методу SLIC, що розглядається в підрозділі 2.3.

2.3 Суперпіксельна сегментація SLIC

2.3.1 Базовий метод SLIC

У запропонованому методі детектування сегментація використовується як засіб уточнення локалізації та підвищення стійкості детектування: замість прийняття рішень на рівні окремих пікселів формується структуроване подання на рівні областей, з яких надалі отримуються сегментаційна маска та обмежувальна рамка об'єкта. Тому під час вибору алгоритму сегментації ключовими є не лише якість меж, а й стабільність детектування та співвідношення між якістю та оперативністю обробки в умовах контрольованої лабораторної відеосцени, зокрема за наявності тіней, відблисків, слабких перепадів контрасту та часткових перекриттів.

Для обґрунтування вибору базового методу було проведено попереднє порівняння кількох методів сегментації в однакових умовах і в єдиній постановці задачі детектування: метод графової сегментації (Felzenszwalb) [121], метод пошуку у просторі ознак (Quickshift) [122], методу водорозділу (Watershed) та метод SLIC (суперпіксельна сегментація) [118]. Оцінювання виконувалось на наборі з 8 зображень, загальна кількість еталонних об'єктів – 38, поріг коректної детекції задавався як $\text{IoU} \geq 0.5$. Для кожного методу вимірювалися: середній час сегментації (мс/кадр), показники Precision,

Recall, $F1$, а також середній IoU та Boundary $F1$ як характеристика якості меж.

Результати тестувань показали, що метод пошуку у просторі ознак забезпечує прийнятні показники якості (Recall $\approx 81.84\%$, $F1 \approx 73.85\%$, IoU ≈ 0.660), однак має найгіршу швидкодiю: у середньому 5441 мс/кадр (~ 0.18 кадр/с). Така обчислювальна вартість робить його малопридатним як проміжний крок у конвеєрі детектування. Метод графової сегментації працює значно швидше за пошук у просторі ознак (1487 мс/кадр, ~ 0.67 кадр/с), але поступається за якістю меж і середнім IoU (Boundary $F1 \approx 0.686$; IoU ≈ 0.643) та має нижчий Precision ($\approx 68.53\%$) при близькому $F1 \approx 73.72\%$. Метод водорозділу є найшвидшим серед протестованих (663 мс/кадр, ~ 1.51 кадр/с), але дає найнижчу повноту детектування (Recall $\approx 44.05\%$), що призводить до значної кількості пропусків: лише 21 коректна детекція з 38. Попри відносно непоганий середній IoU (≈ 0.688), низький відгук істотно погіршує придатність методу для трекінгу, оскільки пропуски детектування породжують розриви траєкторій і збільшують кількість треклетів.

Метод SLIC (Simple Linear Iterative Clustering) демонструє найкраще співвідношення між якістю та оперативністю обробки: середній час 871 мс/кадр (~ 1.15 кадр/с) при високих показниках якості детектування (Precision $\approx 78.52\%$, Recall $\approx 68.64\%$, $F1 \approx 73.25\%$), а також найвищих серед розглянутих значеннях Boundary $F1 \approx 0.735$ і середнього IoU ≈ 0.716 . За кількістю коректних детекцій метод SLIC забезпечує 30 з 38, що істотно краще за метод водорозділу (21 з 38) при прийнятній швидкодії. Отже, за результатами попереднього порівняння методів сегментації встановлено, що метод SLIC є найбільш доцільним базовим методом сегментації для подальшого використання в конвеєрі детектування: він забезпечує одночасно достатню якість локалізації (за IoU та якістю меж) і швидкодiю, необхідну для стабільної роботи детектування в часовій послідовності кадрів. Саме тому метод SLIC обрано як базовий метод сегментації для подальшого

удосконалення та використання в методі детектування і локалізації лабораторних тварин у відеоданих.

Метод SLIC [118] формує суперпікселі шляхом локальної кластеризації пікселів у просторі ознак «колір-координати». Зазвичай для цього використовується колірна LAB-модель зображення, а кожен центр кластера описується вектором:

$$C_k = [l_k, a_k, b_k, x_k, y_k]^T, \quad (2.17)$$

де l_k, a_k, b_k - кольорова компонента;

x_k, y_k - просторова компонента;

k – кількість суперпікселів.

Нехай N – кількість пікселів зображення, тоді середній просторовий крок між центрами кластерів визначається як:

$$S = \sqrt{\frac{N}{k}}, \quad (2.18)$$

Для кожного пікселя $i = (x_i, y_i)$ та центра кластера C_k обчислюються дві складові відстані:

– відстань у просторі кольору:

$$d_c = \sqrt{(l_i - l_j)^2 + (a_i - a_j)^2 + (b_i - b_j)^2}, \quad (2.19)$$

– просторова відстань:

$$d_s = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}, \quad (2.20)$$

Інтегральна відстань у SLIC задається як зважена комбінація d_c та d_s .
У практиці часто використовують нормовану форму:

$$D = \sqrt{\left(\frac{d_c}{m}\right)^2 + \left(\frac{d_s}{S}\right)^2}, \quad (2.21)$$

де m – параметр компактності (чим більший m , тим компактніші суперпікселі; чим менший – тим сильніше суперпікселі прилягають до меж об'єктів).

Після обчислення відстаней пікселі тимчасово призначаються найближчим кластерам (у межах локального вікна навколо центра), далі центри кластерів переобчислюються як середні значення ознак пікселів, що увійшли до кластера. Процес повторюється ітераційно до стабілізації або до досягнення заданої кількості ітерацій [118]. В алгоритмі 2.1. наведено загальний псевдокод алгоритму. Окремим важливим етапом базового SLIC є постобробка зв'язності, оскільки після ітерацій кластеризації можуть виникати дрібні ізольовані області; їх доцільно приєднувати до сусідніх суперпікселів для забезпечення топологічної коректності сегментації [118].

Алгоритм 2.1 – Базовий алгоритм суперпіксельної сегментації SLIC

/ Ініціалізація */*

Ініціалізувати центри кластерів $C_k = [l_k, a_k, b_k, x_k, y_k]^T$ шляхом вибірки пікселів на регулярних кроках сітки S .

Перемістити центри кластерів у положення з найнижчим градієнтом у околиці 3×3 .

Встановити мітку $l(i) = -1$ для кожного пікселя i .

Встановити відстань $d(i) = \infty$ для кожного пікселя i .

повторити:

/ Присвоєння */*

for (для) кожного центру кластера C_k зробити

for (для) кожного пікселя I у $2S \times 2S$ регіоні навкруги C_k зробити

Обчисліть відстань D між C_k та i .

Якщо $D < d(i)$ тоді

```

        встановити  $d(i) = D$ 
        встановити  $l(i) = k$ 
    end if
end for
end for
/* Оновити */
Обчисліть нові центри кластерів.
Обчисліть залишкову похибку  $E$ .
допоки  $E \leq \text{threshold}$ 

```

Разом із тим базовий метод SLIC спирається переважно на LAB-модель зображення, тому за слабкого контрасту, тіней, відблисків і локальних артефактів може формувати суперпікселі, які недостатньо точно відповідають межам області лабораторної тварини. У задачі детектування доцільно враховувати додаткові представлення зображення, які підсилюють відмінність між об'єктом і фоном, зокрема різницеве зображення після віднімання фону та результати попередньої нормалізації кадру. Це дає змогу формувати більш інформативний простір ознак для кластеризації пікселів і зменшувати вплив випадкових шумових або фонових фрагментів. Тому в підрозділі 2.3.2 розглядається удосконалення методу SLIC на основі багат шарового простору ознак і перевірки локальної однорідності.

2.3.2 Удосконалення методу SLIC

Базовий метод SLIC було удосконалено шляхом переходу до багат шарового простору ознак, в якому кластеризація виконується не лише за трьома кольірними компонентами LAB-моделі зображення, а й за набором з n шарів ознак, сформованих на етапах попередньої обробки та віднімання фону [34, 105, 123]. Такий підхід дає змогу враховувати як кольірно-просторову близькість пікселів, так і додаткові ознаки, що характеризують відмінність між областю лабораторної тварини та фоном.

Нехай для кожного пікселя сформовано n шарів ознак. Тоді центр кластера задається як:

$$C_k = [l_{k1}, \dots, l_{km}, x_k, y_k]^T, \quad m = \overline{3, n}, \quad (2.22)$$

де l_{km} - кольорова компонента у k -му кластері;

x_k, y_k - просторова компонента;

n – кількість шарів ознак.

Таким чином, замість стандартної тривимірної колірної частини (LAB) використовується n -вимірний простір ознак, що дозволяє враховувати не лише колірну близькість пікселів, а й додаткові представлення, важливі для відокремлення області лабораторної тварини від фону, наприклад, різницеве зображення після віднімання фону, нормалізоване зображення в градаціях сірого, контрастні представлення тощо. Відстань у просторі ознак в удосконаленому методі SLIC визначається як евклідова відстань у n -вимірному просторі шарів:

$$d_c = \sqrt{\sum_{m=1}^n (l_{j,m} - l_j)^2}, \quad (2.23)$$

де $l_{j,m}$ – значення m -го шару для пікселя j .

Просторова відстань d_s залишається стандартною. Для узгодженого врахування ознакової та просторової складових використовується нормована відстань:

$$D' = \sqrt{\left(\frac{d_c}{N_c}\right)^2 + \left(\frac{d_s}{S}\right)^2}, \quad (2.24)$$

де N_c – коефіцієнт нормування ознакової відстані.

Така форма дає змогу більш гнучко налаштовувати баланс між подібністю за ознаками та просторовою компактністю суперпікселів, що особливо важливо в задачах детектування лабораторних тварин за умов нестационарного фону й локальних артефактів [34].

Другою складовою удосконалення є додаткова перевірка локальної однорідності пікселів на етапі обчислення відстані між пікселем і центром кластера. Її призначення полягає в тому, щоб зменшити вплив пікселів, які різко відрізняються від свого локального оточення, зокрема шумових пікселів, крайових викидів і локальних артефактів. Для кожного пікселя-кандидата i :

- обирається локальна околиця радіуса r ;
- для кожного шару ознак обчислюється середнє значення в околиці;
- формується середній локальний вектор ознак:

$$\mu_i = [\mu_{i,1}, \mu_{i,2}, \dots, \mu_{i,n}]^T, \quad (2.25)$$

- обчислюється відхилення пікселя від локального середнього:

$$d_{loc(i)} = \sqrt{\sum_{m=1..n} (l_{i,m} - \mu_{i,m})^2}. \quad (2.26)$$

Якщо $d_{loc(i)} > T_h$, де T_h – поріг локальної однорідності, піксель пропускається на етапі призначення кластеру (тобто не бере участі в поточному порівнянні з центром C_k). Така перевірка зменшує вплив шуму та локальних артефактів, покращує однорідність суперпікселів і сприяє більш точному формуванню об'єктних областей для подальшого детектування та локалізації [34].

Оскільки в запропонованому методі сегментація використовується як етап детектування, після побудови суперпікселів доцільно виконувати додаткову постобробку:

- об'єднання дрібних суміжних областей (плям), які з високою ймовірністю належать одному об'єкту;
- відкидання дуже малих сегментів, що мають шумовий характер;
- формування об'єктних областей у вигляді масок та/або обмежувальних рамок для подальшого трекінгу.

Такі операції не змінюють суті методу SLIC, але підвищують практичну придатність сегментації для задачі детектування та локалізації лабораторних тварин у відеопослідовностях [34, 105]. В алгоритмі 2.2 наведено псевдокод удосконаленого методу SLIC.

Алгоритм 2.2 – Псевдокод удосконаленого методу сегментації суперпікселів SLIC

Вхід:

$\{L^1, L^2, \dots, L^n\}$ – набір ознакових шарів зображення

k – бажана кількість суперпікселів

N_c, N_s – коефіцієнти нормування відстаней

T – кількість ітерацій

r – радіус локальної околиці для перевірки однорідності

T_h – поріг локальної однорідності

Вихід:

M – мітки суперпікселів

1. Сформувати n -шарове представлення зображення.

2. Обчислити $S = \sqrt{\frac{N}{k}}$.

3. Ініціалізувати центри $C_k = [l_{k1}, \dots, l_{km}, x_k, y_k]^T$, $m = \overline{3, n}$ на регулярній сітці.

4. Ініціалізувати $D_{\min(i)} = \infty$, $M_i = -1$.

5. Для $t = 1, \dots, T$:

5.1. Для кожного центра C_k :

– розглядати пікселі у вікні $2S \times 2S$;

– для кожного пікселя i у вікні:

а) обчислити локальне середнє μ_i в -околиці (по всіх n шарах);

б) обчислити $d_{loc(i)}$;

- с) якщо $d_{loc(i)} > T_h$:
 пропустити піксель i ;
 інакше обчислити:
- d_c (у -вимірному просторі ознак),
 - d_s (просторова відстань),
 - $D' = \sqrt{\left(\frac{d_c}{N_c}\right)^2 + \left(\frac{d_s}{S}\right)^2}$
- якщо $D' < D_{min(i)}$:
 $D_{min(i)} = D'$;
 $M_i = k$.

5.2. Переобчислити центри C_k як середні значення ознак і координат

5.3. Скинути $D_{min(i)} = \infty$.

6. Виконати постобробку зв'язності суперпікселів.
 Повернути M .

Запропоноване удосконалення методу SLIC спрямоване не лише на покращення візуальної якості суперпікселів, а передусім на підвищення точності локалізації та детектування лабораторних тварин у відеоданих. Багатошаровий простір ознак дає змогу врахувати LAB-модель зображення, різницеве зображення та результати попередньої обробки, а перевірка однорідності пікселів в локальному оточенні пікселя зменшує чутливість до шуму й локальних артефактів. У результаті підвищується якість формування об'єктних областей, що створює більш стійку основу для подальшого трекінгу та збереження ідентичності об'єктів у часі [34, 105].

2.4 Результати детектування та локалізації лабораторних тварин із використанням удосконаленого методу SLIC

Проведено експериментальне дослідження удосконаленого методу SLIC у задачі детектування та локалізації лабораторних тварин у відеоданих (рис. 2.3) на наборі з 75 зображень (кадри, вибрані з відеопослідовності). Для тестування зображення (включаючи зображення порожнього фону) були зменшені на 65% від початкового розміру 1784×1520 . Початкову сітку суперпікселів було задано так, щоб містити комірки сітки розміром у 80

пікселів. Як для базового, так і для удосконаленого SLIC кількість ітерацій кластеризації становила 10 ітерацій, розмір околиці перевірки локальної однорідності – 7, а відповідний поріг кольорової відмінності – 130. На етапі постобробки було виконано об'єднання близько розташованих фрагментів об'єктної області, оскільки такі фрагменти можуть відповідати частинам однієї лабораторної тварини, розділеним локальними перешкодами або елементами фону. Окрім стандартних шарів LAB-моделі зображення, як було використано різницеве зображення.

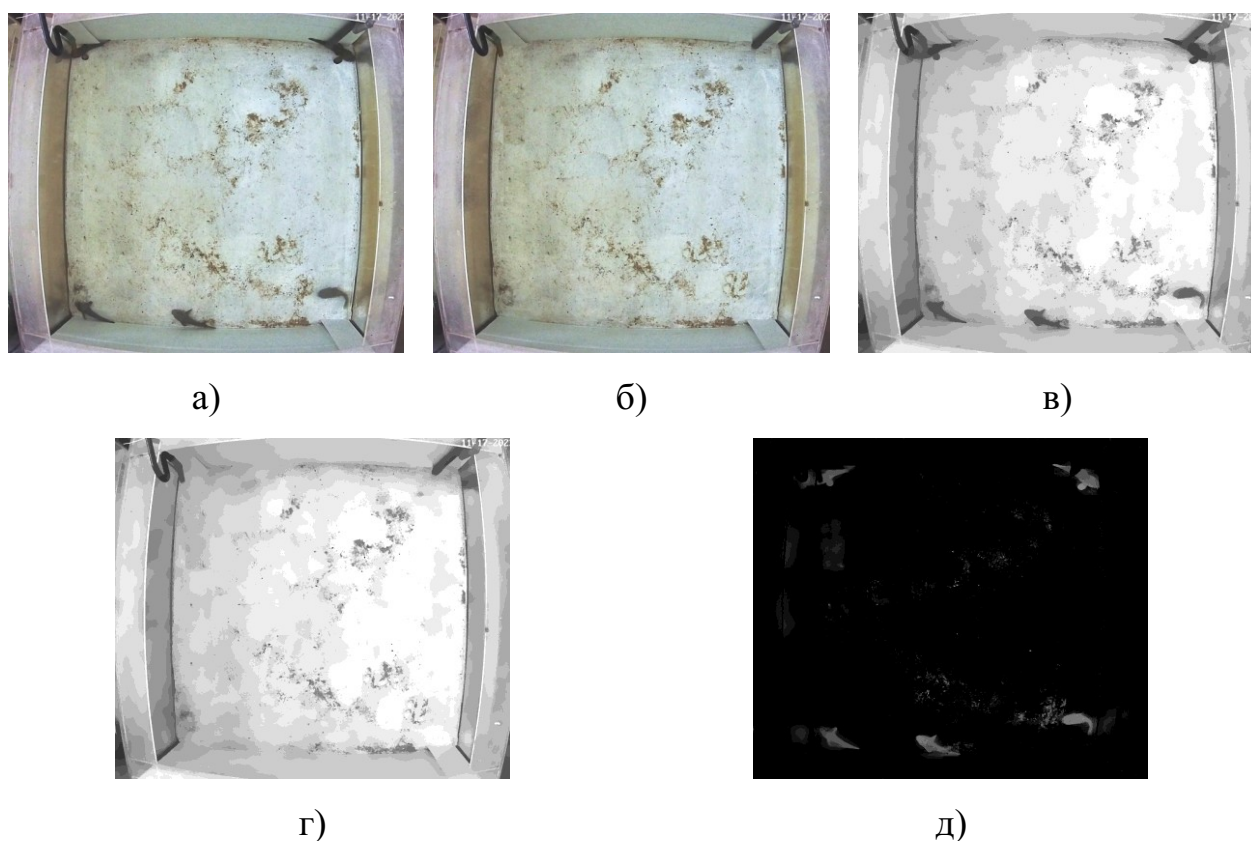


Рисунок 2.3 – Приклади етапів попередньої обробки: а) оригінальний кадр відео; б) зображення фону; в) кадр після фільтрації, корекції освітлення та перетворення у градації сірого; г) фон після фільтрації, корекції освітлення та перетворення у градації сірого; д) результат віднімання фону

У таблицях 2.1 та 2.2 наведено матриці плутанини для базового SLIC та удосконаленого методу SLIC.

Таблиця 2.1 – Матриця плутанини для детектування за допомогою базового методу SLIC

Еталонні маркери (відмічені вручну), %	Результати, отримані за допомогою базового SLIC, %	
	Лабораторна тварина	Фон
Лабораторна тварина	83.008	16.992
Фон	0.322	99.678

У матрицях помилок, відсоток пікселів, що належать до області, що є об'єктом, та сегментовані як пікселі об'єкта, позначено TP (істинно позитивний). TN (істинно негативний) – відсоток пікселів, що належать до фону та сегментовані як пікселі фону. FP (хибнопозитивний) – відсоток пікселів, що належать до фону, але сегментовані як об'єкт. FN (хибнонегативний) – відсоток пікселів, що належать до області, що є об'єктом, але сегментовані як пікселі фону.

Таблиця 2.2 – Матриця плутанини для детектування за допомогою удосконаленого методу SLIC

Еталонні маркери (відмічені вручну), %	Результати, отримані за допомогою модифікованого SLIC, %	
	Лабораторна тварина	Фон
Лабораторна тварина	91.245	8.755
Фон	0.509	99.491

Крім того, було розраховано загальну кількість коректних детектувань відносно загальної кількості еталонних об'єктів на всьому наборі зображень. Результати наведено в таблиці 2.3. На рис. 2.3 наведено приклади проміжних етапів попередньої обробки кадру та формування різницевого зображення.

Таблиця 2.3 – Порівняння кількості коректних детектувань для базового та удосконаленого SLIC

Метод	Коректні детектування	Еталонні об'єкти
Базовий SLIC	316	389
Удосконалений SLIC	340	

На рис. 2.4 показано результат суперпіксельної сегментації у вигляді сітки суперпікселів, накладеної на різницеве зображення.

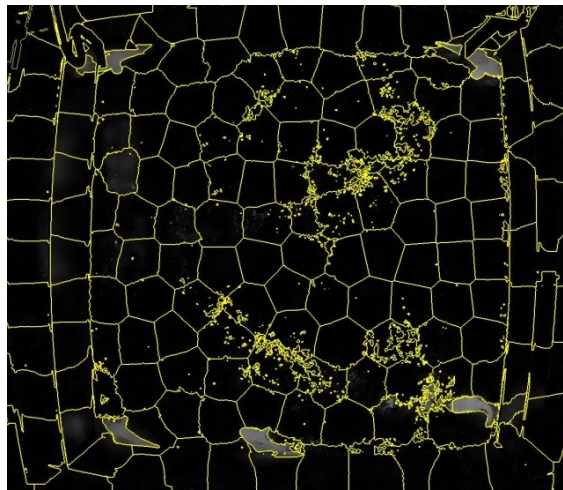
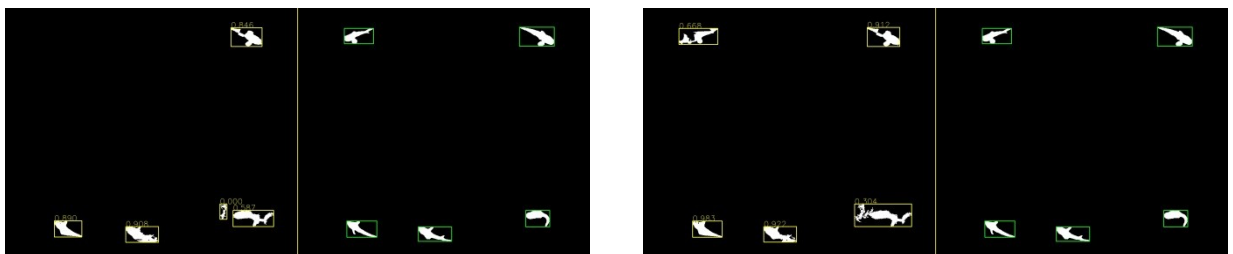


Рисунок 2.4 – Результат суперпіксельної сегментації, накладений на різницеве зображення

На рис. 2.5 наведено приклад візуального порівняння результатів детектування з використанням базового та удосконаленого SLIC.



а)

б)

Рисунок 2.5 – Порівняння результатів детектування: а) базовий SLIC; б) удосконалений SLIC

Удосконалений метод SLIC дозволив підвищити якість детектування на ~6,8% на заданому наборі зображень, що є важливим для коректної ідентифікації об'єктів на кожному кадрі.

2.5 Висновки до другого розділу

У другому розділі розроблено та обґрунтовано методи детектування й локалізації лабораторних тварин у відеоданих на основі суперпіксельної сегментації SLIC та її удосконалення. Показано, що використання SLIC дає змогу перейти від нестійкого піксельного порогування після віднімання фону до аналізу локально узгоджених областей, що є важливим за умов нестаціонарного фону, тіней, відблисків і локальних артефактів.

Обґрунтовано вибір SLIC як базового методу сегментації для задачі детектування та локалізації лабораторних тварин. Порівняння з іншими методами сегментації показало, що SLIC забезпечує доцільне співвідношення між якістю формування меж, точністю локалізації та оперативністю обробки, а також є зручним для подальшого параметричного налаштування під умови контрольованої лабораторної відеосцени.

Удосконалення методу SLIC полягає у використанні багат шарового простору ознак, який базується на LAB-моделі зображення, представленні різницевого зображення та результатах попередньої обробки кадру. Додатково введено перевірку однорідності пікселів у локальному оточенні, що зменшує вплив шуму, фонових артефактів і локальних неоднорідностей під час формування суперпікселів.

Експериментальна перевірка показала, що порівняно з базовим SLIC удосконалений метод забезпечив підвищення частки правильно сегментованих об'єктних пікселів і дозволив підвищити якість детектування більш ніж на 6 %. Отримані результати підтверджують доцільність

використання удосконаленого SLIC і контекстної фільтрації як складових методу детектування та локалізації лабораторних тварин у відеоданих.

Таким чином, результати другого розділу створюють основу для подальшого формування стабільних просторових представлень лабораторних тварин у вигляді масок та обмежувальних рамок, що надалі використовуються у трекінгу, формуванні треклетів і збереженні ідентичності об'єктів у відеопослідовності.

3 УДОСКОНАЛЕНА НЕЙРОМЕРЕЖЕВА МОДЕЛЬ MASK R-CNN ДЛЯ ДЕТЕКТУВАННЯ ТА СЕГМЕНТАЦІЇ ЛАБОРАТОРНИХ ТВАРИН У ВІДЕОДАНИХ

3.1 Постановка задачі нейромережевого детектування та сегментації лабораторних тварин

У другому розділі було розроблено та обґрунтовано методи детектування й локалізації лабораторних тварин у відеоданих на основі суперпіксельної сегментації SLIC та її удосконалення, що дало змогу сформувати стабільні просторові представлення лабораторних тварин у вигляді масок та обмежувальних рамок для подальшого використання у трекінгу, формуванні треклетів і збереженні ідентичності об'єктів у відеопослідовності. Разом з тим для задач подальшого стійкого трекінгу та аналізу поведінки такого рівня представлення не завжди достатньо, особливо у випадках часткових перекриттів, злиття та подальшого розділення об'єктів, деформацій силуету, неоднорідного фону та змін освітлення. За таких умов доцільним є використання нейромережевого детектування з одночасною сегментацією об'єкта, яке формує не лише обмежувальну рамку, а й піксельну маску об'єкта, тобто більш інформативне просторове представлення [5, 76, 124]. Відповідно, на загальній схемі конвеєра, наведений у розділі 2 (рис. 2.1), нейромережева складова розглядається як основне джерело детекцій, обмежувальних рамок та масок екземплярів лабораторних тварин.

Для формалізації задачі нейромережевого детектування та сегментації лабораторних тварин уточнимо використання основних понять у межах цього розділу. Об'єкт розглядається як окремий екземпляр (інстанс) лабораторної тварини, а не як клас; результатом детектування є обмежувальна рамка та оцінка достовірності, а результатом сегментації – піксельна маска відповідного екземпляра. На цьому етапі формується просторовий опис

об'єкта в межах окремого кадру без гарантованого збереження його ID у часовій послідовності; підтримання відповідності об'єктів у часі та збереження їх ідентичності забезпечуються на подальших етапах трекінгу, формування треклетів і повторної ідентифікації.

Нехай задано відеопослідовність кадрів $\{I_t\}_{t=1}^T$, де I_t – зображення кадру в момент часу t , T – кількість кадрів у відеопослідовності. У розроблюваному підході на вхід нейромережевої моделі подається не лише базове зображення кадру, а й додаткові інформаційні шари (канали), сформовані на попередніх етапах обробки та призначені для підсилення ознак об'єкта за умов складного фону. Тоді вхідне представлення кадру можна подати як

$$X_t = \Phi(I_t, A_t), \quad (3.1)$$

де A_t – набір додаткових шарів (похідних карт, масок або допоміжних представлень кадру);

Φ – оператор формування багатоканального вхідного тензора зображення.

Результат роботи нейромережевого модуля детектування і сегментації для кадру t задається множиною екземплярів

$$\mathcal{D}_t = \{d_t^{(i)}\}_{i=1}^{N_t}, \quad (3.2)$$

де N_t – кількість детектованих об'єктів у кадрі, а

$$d_t^{(i)} = (b_t^{(i)}, m_t^{(i)}, s_t^{(i)}), \quad (3.3)$$

де $b_t^{(i)}$ – обмежувальна рамка i -го об'єкта;

$m_t^{(i)}$ – піксельна маска відповідного екземпляра;

$s_t^{(i)}$ – оцінка достовірності детектування.

У цьому розділі розглядається задача побудови удосконаленої нейромережевої моделі Mask R-CNN, яка забезпечує:

- детектування лабораторних тварин у кадрі з формуванням обмежувальних рамок;
- сегментацію детектованих екземплярів із побудовою піксельних масок;
- підвищення якості детектування та сегментації за рахунок використання додаткових інформаційних шарів і модифікації екстрактора ознак;
- узгодження із загальним конвеєром обробки відеоданих, у якому результати нейромережевої моделі використовуються як основні для подальшого покадрового зіставлення, формування треклетів і повторної ідентифікації, а результати методів детектування й локалізації, розроблених у розділі 2, використовуються для підсилення ознак, резервної локалізації та перевірки геометричної правдоподібності детекцій.

Оскільки нейромережева модель інтегрується в загальний конвеєр обробки відеоданих, її результати можуть додатково перевірятися за критеріями достовірності та просторової правдоподібності. Це дає змогу відсіювати низьковпевнені або геометрично неправдоподібні детекції та, за потреби, використовувати результати методів детектування й локалізації на основі суперпіксельної сегментації SLIC як резервне джерело локалізації.

Нехай D_t^{NN} – множина нейромережевих детектувань на кадрі t . Для кожного детектування $d \in D_t^{\text{NN}}$ відомі обмежувальна рамка $b(d)$ та оцінка впевненості $s(d)$. Тоді формується відфільтрована множина детектувань

$$\tilde{D}_t^{\text{NN}} = \{d \in D_t^{\text{NN}} | s(d) \geq s_{\min} \wedge g(b(d)) = 1\}, \quad (3.4)$$

де $s_{\min}$ – поріг впевненості (перевірка 1);

$g(b)$ – індикатор правдоподібності геометрії рамки (перевірка 2).

Нехай рамка b задається координатами $(x_{\min}, y_{\min}, x_{\max}, y_{\max})$. Її ширина та висота визначаються як

$$w = x_{\max} - x_{\min}, \quad h = y_{\max} - y_{\min}. \quad (3.5)$$

Площа рамки та співвідношення сторін обчислюються відповідно як

$$A(b) = w \cdot h, \quad \rho(b) = \frac{w}{h}. \quad (3.6)$$

Індикатор геометричної правдоподібності визначається як індикаторна функція, що набуває значення 1, якщо умова істинна, і 0 – якщо умова хибна:

$$g(b) = 1\{A_{\min} \leq A(b) \leq A_{\max} \wedge \rho_{\min} \leq \rho(b) \leq \rho_{\max}\}, \quad (3.7)$$

де $A_{\min}, A_{\max}$ – допустимі межі площі рамки;

$\rho_{\min}, \rho_{\max}$ – допустимі межі співвідношення сторін для контрольованої сцени (задано апіорно або за статистикою стабільних детектувань/траєкторій).

Тоді правило арбітражу для формування вихідної множини детектувань D_t має вигляд:

$$D_t = \begin{cases} \tilde{D}_t^{\text{NN}}, & \text{if } |\tilde{D}_t^{\text{NN}}| > 0 \\ D_t^{\text{SLIC}}, & \text{if } |\tilde{D}_t^{\text{NN}}| = 0 \end{cases} \quad (3.8)$$

де D_t^{SLIC} – множина детектувань, отримана методами детектування та сегментації на основі суперпіксельної сегментації SLIC [116, 117].

Таким чином, методи детектування та сегментації на основі суперпіксельної сегментації SLIC використовуються не лише у випадку

нульового виходу нейромережі, а й у кадрах, де нейромережеві детектування є низьковпевненими або геометрично неправдоподібними.

Трекінг як процес складається з етапів: покадрове зіставлення, треклети, перекриття, злиття та/або розділення. Після арбітражу множина D_t надходить на вхід трекінгу, який реалізується як покадрове зіставлення детектувань з активними траєкторіями. Для кожної активної траєкторії τ_j підтримується поточний стан, а саме остання локалізація, прогноз положення та накопичені ознаки. На кадрі t для кожної траєкторії формується прогнозована рамка $\hat{b}_{t|t-1}^{(i)}$ на основі попередніх положень. Далі для кожної пари $(\tau_j, d_t^{(i)})$ обчислюються критерії узгодження:

- геометричний: перекриття рамок $\text{IoU}(\hat{b}_{t|t-1}^{(i)}, b_t^{(i)})$ та/або відстань між центрами;
- ознаковий: схожість зовнішнього вигляду (якщо доступні ознаки повторної ідентифікації) через норму різниці $\|\hat{f}^{(j)} - f_t^{(i)}\|_2$ [115].

Узагальнено вартості зіставлення можна задати, наприклад, як:

$$c_{j,i} = \alpha \left(1 - \text{IoU}(\hat{b}_{t|t-1}^{(i)}, b_t^{(i)}) \right) + (1 - \alpha) \|\hat{f}^{(j)} - f_t^{(i)}\|_2, \quad (3.9)$$

де $0 \leq \alpha \leq 10$ визначає баланс між геометричним і ознаковим критеріями.

Після обчислення матриці вартостей виконується призначення «траєкторія $\leftrightarrow$ детектування» з порогоми допустимості, наприклад, мінімальний IoU та/або максимальна допустима вартість $c_{j,i}$. Призначені траєкторії оновлюються локалізацією та (за наявності) ознаками; непризначені детектування ініціалізують нові траєкторії (після підтвердження на кількох кадрах для зменшення впливу хибних детектувань), а траєкторії без призначення переводяться у стан «пропуск» і завершуються після перевищення допустимої кількості пропусків.

При перекриттях і злиттях об'єктів покадрове зіставлення стає неоднозначним: детектор може сформувати одну обмежувальну рамку та піксельну маску для двох об'єктів або нестабільні сегментаційні маски. У таких випадках траєкторії доцільно переводити у режим формування треклетів, тобто фіксувати неперервні фрагменти до моменту невизначеності, а після розділення – виконувати з'єднання треклетів за ознаками повторної ідентифікації, що дозволяє відновити правильну відповідність ID і зменшити перемикування ідентичностей [115]. Отже, у межах цього розділу трекінг, формування треклетів і повторна ідентифікація розглядаються як подальші етапи конвеєра, для яких результати нейромережевого детектування та сегментації формують вхідні дані.

Формально задача навчання нейромережевої моделі полягає у знаходженні параметрів нейромережевої моделі θ , що мінімізують функціонал втрат на навчальній вибірці:

$$\theta^* = \operatorname{argmin}_{\theta} \frac{1}{M} \sum_{j=1}^M L(X^{(i)}, Y^{(i)}, \theta), \quad (3.10)$$

де M – кількість навчальних прикладів;

$X^{(i)}$ – вхідне багатоканальне зображення;

$Y^{(i)}$ – еталонна розмітка (рамки та маски екземплярів);

L – сукупна функція втрат моделі, яка має комбінований характер і включає втрати класифікації, регресії рамки та побудови маски [79]:

$$L = L_{cls} + L_{box} + L_{mask}. \quad (3.11)$$

При цьому припущення, в межах яких розробляється і досліджується удосконалена нейромережева модель, залишаються відповідно такими ж, як і в розділі 2:

- камера є стаціонарною або квазістаціонарною без істотних зсувів сцени між кадрами;
- сцена спостереження є обмеженою та контрольованою;
- освітлення може змінюватися, але не призводить до повної втрати контрасту об'єкта;
- у кадрі може бути одна або кілька лабораторних тварин; можливі контакти, часткові перекриття та короткочасні злиття областей;
- об'єкти можуть бути візуально подібними, тому лише геометричних ознак часто недостатньо для стійкого розрізнення ID у часі;
- у цьому розділі основна увага приділяється якості детектування та сегментації, результати яких створюють основу для подальшого трекінгу, формування треклетів, повторної ідентифікації та наскрізного збереження ID лабораторних тварин у відеопослідовності.

3.2 Базова нейромережева модель Mask R-CNN для детектування та сегментації лабораторних тварин

У задачах аналізу відеосцен, де результатом має бути не лише локалізація об'єкта у вигляді обмежувальної рамки, а й його піксельне представлення, доцільно застосовувати нейромережеві моделі, що реалізують детектування та сегментацію екземплярів у єдиному конвеєрі. На відміну від «чистого» детектування, сегментація екземплярів дає можливість точніше відокремлювати об'єкт від фону, коректніше описувати форму та краще розділяти об'єкти, що контактують або частково перекриваються. Для прикладних задач подальшого трекінгу це важливо, оскільки маска може бути джерелом додаткових стабільних просторових ознак (контур, площа,

компактність, локальні ознаки всередині об'єкта), що зменшує чутливість системи до локальних завад, неоднорідного освітлення та деформацій силуету.

Сучасні підходи, що поєднують детектування і сегментацію, умовно можна поділити на дві групи: двоетапні регіон-орієнтовані моделі, у яких спочатку формуються кандидатні області, а потім виконується уточнення та побудова маски в межах цих областей; та одноетапні щільні моделі, у яких детектування й маски прогнозуються безпосередньо з карт ознак за один прохід. Загальні тенденції розвитку детектування, зокрема баланс «швидкодія-точність» і розширення задач до сегментації, узагальнені в оглядових роботах [24]. У межах даної роботи вибір базової моделі визначається не лише показниками точності на абстрактних бенчмарках, а й відповідністю вимогам наскрізного конвеєра: якістю масок, стійкістю до перекриттів, можливістю модифікації екстрактора ознак та подальшим використанням отриманих ознак для трекінгу, формування треклетів і повторної ідентифікації.

Як базову нейромережеву модель для детектування та сегментації лабораторних тварин обрано Mask R-CNN, що є розвитком Faster R-CNN і додає окрему гілку прогнозування маски для кожної області інтересу [4, 5]. Двоетапна організація моделі – формування пропозицій регіонів із подальшим уточненням локалізації та сегментацією кожної області є принципово важливою для задач, у яких маска виступає не побічним результатом, а ключовим елементом подальшого аналізу. По-перше, Mask R-CNN містить механізм RoIAlign, введений для точнішого узгодження ознак в областях інтересу та зменшення похибок квантування під час вибірки ознак; це позитивно впливає на якість меж масок екземплярів [5]. По-друге, у цій моделі природно розділяються задачі класифікації, регресії рамки й побудови маски, що дає можливість цілеспрямовано модифікувати саме ті компоненти, які є ключовими для якості сегментації, не порушуючи загальну логіку двоетапного детектування [5, 99].

Як альтернатива двоетапним підходам сьогодні активно використовуються одноетапні сегментаційні моделі сімейства YOLO (YOLO-seg), у яких модель повертає обмежувальні рамки та маски екземплярів за один прохід. У практичних реалізаціях це оформлено як окремий режим задачі сегментації, де вихід моделі включає рамки, оцінки достовірності та маски екземплярів [125, 126]. Перевага таких моделей полягає у високій продуктивності та зручності застосування в сценаріях з обмеженнями на час обробки кадру. Разом з тим, у даній роботі вимога надвисокої швидкодії не є визначальною: нейромережева модель використовується як джерело максимально інформативних і стабільних результатів для наступних етапів конвеєра: покадрове зіставлення; формування треклетів; підтримання ID та повторної ідентифікації. Саме тому доцільність вибору моделі Mask R-CNN слід розглядати не лише за значеннями метрик якості, а й за відповідністю моделі способу використання масок і ознак у побудованому конвеєрі.

У межах задачі детектування та сегментації лабораторних тварин можна виокремити три практичні аспекти, за якими Mask R-CNN є доцільною базовою моделлю порівняно з YOLO-seg:

– якість та стабільність масок екземплярів як вхідних ознак для подальших етапів конвеєра. У запропонованому в роботі конвеєрі маска використовується не лише як кінцевий результат сегментації, а як основа для формування стабільних просторових ознак і подальшого трекінгу. Двоетапна схема з побудовою маски в межах ROI та використанням RoIAlign дає змогу більш керовано підвищувати якість меж масок і зменшує ризик систематичних похибок на контурах [5]. Для YOLO-seg маска є інтегрованою складовою одноетапного виходу, і хоча сучасні реалізації демонструють високі показники, можливості цілеспрямованого налаштування механізмів формування маски під конкретні артефакти сцени зазвичай більш обмежені, оскільки нейромережева модель оптимізується під загальну швидкодію [125-127];

– модульність і придатність до удосконалення екстрактора ознак. Модель Mask R-CNN побудована за модульним принципом: екстрактор ознак, модуль формування пропозицій регіонів, механізм вирівнювання ROI-ознак, а також окремі голови для класифікації, регресії рамки та прогнозування маски є чітко розділеними компонентами [5]. Це створює природні точки для подальших дослідницьких модифікацій – зокрема, на рівні екстрактора ознак або на рівні використання ROI-ознак для формування додаткових представлень об'єкта. Важливо, що в Mask R-CNN обробка ведеться на рівні областей інтересу, тому ознаки для кожного екземпляра формуються з просторово локалізованого фрагмента сцени, що зменшує вплив фону та спрощує узгодження додаткових гілок із задачами детектування та сегментації екземплярів [4, 5]. На відміну від цього, YOLO-seg реалізує одноетапне щільне прогнозування, у якому детектування та маски тісно пов'язані в єдиній вихідній голові моделі, оптимізованій під швидкодію [125, 126]. Така організація є ефективною для прикладних сценаріїв із жорсткими часовими обмеженнями, однак за потреби глибших архітектурних змін часто вимагає узгодження форматів виходів і внутрішніх представлень у декількох вузлах моделі одночасно. У результаті підвищується чутливість до налаштувань навчання та ризик деградації якості масок при зміні екстрактора чи структури вихідних голів [125-127];

– обробка складних випадків, пов'язаних із контактами, частковими перекриттями та злиттям об'єктів. У сценах із частковими перекриттями та візуально подібними об'єктами критичною стає здатність моделі коректно розділяти екземпляри й відновлювати межі навіть при неповній видимості. У двоетапних моделях існує додатковий «ступінь свободи» у вигляді регіональних пропозицій і ROI-обробки, що на практиці часто полегшує підвищення якості сегментації в складних сценах через налаштування екстрактора та масштабів ознак, зокрема за рахунок піраміди ознак [4, 5, 84]. Це узгоджується з метою цього розділу: підвищити якість детектування й

сегментації лабораторних тварин саме в ситуаціях, критичних для подальшого трекінгу, формування треклетів і повторної ідентифікації.

Модель Mask R-CNN належить до класу двоетапних регіон-орієнтованих моделей і поєднує в одному конвеєрі детектування об'єктів та побудову масок екземплярів [5]. Її побудова ґрунтується на ідеї, що для якісної сегментації екземплярів доцільно спочатку сформувати множину кандидатних областей, які з високою ймовірністю містять об'єкт, а вже далі виконати уточнення локалізації та прогнозування маски в межах кожної з таких областей. Як розвиток Faster R-CNN, Mask R-CNN зберігає основні компоненти двоетапного детектора (екстракцію ознак, модуль пропозицій регіонів і гілки уточнення), але додає паралельну гілку сегментації, яка формує бінарну маску екземпляра для кожного детектованого б'єкта [4, 5].

Базова нейромережева модель Mask R-CNN реалізує послідовний конвеєр обробки зображення. Спочатку вхідне зображення перетворюється екстрактором ознак на набір карт ознак. У класичній постановці роль екстрактора виконує згорткова нейромережа типу ResNet, а для підвищення якості детектування об'єктів різних масштабів часто застосовується піраміда ознак (FPN), яка формує багаторівневі представлення з різною просторовою роздільною здатністю [31, 99]. Наявність багатомасштабних ознак є важливою для сегментації екземплярів, оскільки об'єкти можуть істотно відрізнятися за розміром у кадрі, а також мати структури, які потребують високої деталізації меж.

Далі використовується модуль формування пропозицій регіонів (RPN), який на основі карт ознак генерує множину кандидатних областей. Кожна область описується обмежувальною рамкою та оцінкою достовірності. У Faster R-CNN і Mask R-CNN цей етап виконує роль «грубої фільтрації»: замість перебору всіх можливих позицій і масштабів у кадрі аналіз обмежується відносно малою множиною кандидатів, що зменшує складність задачі та підвищує точність подальшого уточнення [4]. Після цього застосовується відбір кращих регіонів із урахуванням оцінки достовірності та

усунення дублювань, і саме ці регіони використовуються як області інтересу для наступних гілок мережі.

Ключовим елементом моделі Mask R-CNN є механізм вирівнювання ознак області інтересу – RoIAlign. На відміну від попередніх варіантів вирівнювання, RoIAlign спрямований на зменшення похибок, пов'язаних із квантуванням координат області інтересу під час вибірки ознак, що особливо впливає на точність побудови маски та її меж [5]. Після вирівнювання кожна область інтересу приводиться до фіксованого розміру ознакового тензора, що робить можливим застосування стандартних під-мереж (англ. subnetworks) для різних областей незалежно від їхнього початкового розміру в кадрі.

Після отримання ROI-ознак у Mask R-CNN паралельно працюють дві основні гілки. Перша гілка відповідає за детектування у вигляді: (1) класифікації об'єкта (або належності до класу) та (2) уточнення обмежувальної рамки через регресію координат [4, 5]. Друга гілка – маскова – формує бінарну маску екземпляра для кожної ROI. Важливо, що маскова гілка прогнозує маску в узгодженій системі координат ROI, а не для всього кадру. Це дозволяє будувати маску з орієнтацією на локалізовану область і фокусувати модель на деталях меж у межах конкретного об'єкта [5]. На практиці така структура є зручною, оскільки результати детектування (рамка) і сегментації (маска) виходять узгодженими за об'єктом і можуть розглядатися як єдиний опис екземпляра.

Сукупна оптимізація Mask R-CNN задається комбінацією втрат, які відповідають різним гілкам моделі: втрати класифікації, втрати регресії рамки та втрати побудови маски [5]. Загальну структуру базової моделі Mask R-CNN показано на рис. 3.1.

Такий підхід дозволяє навчати модель у багатозадачному режимі, де параметри екстрактора ознак формуються під одночасні вимоги точного детектування та якісної сегментації. Розділення гілок на рівні ROI дає змогу зберігати спеціалізацію: регресія рамки оптимізується під геометричну

локалізацію, а маскова голова оптимізується під піксельне відтворення форми об'єкта.

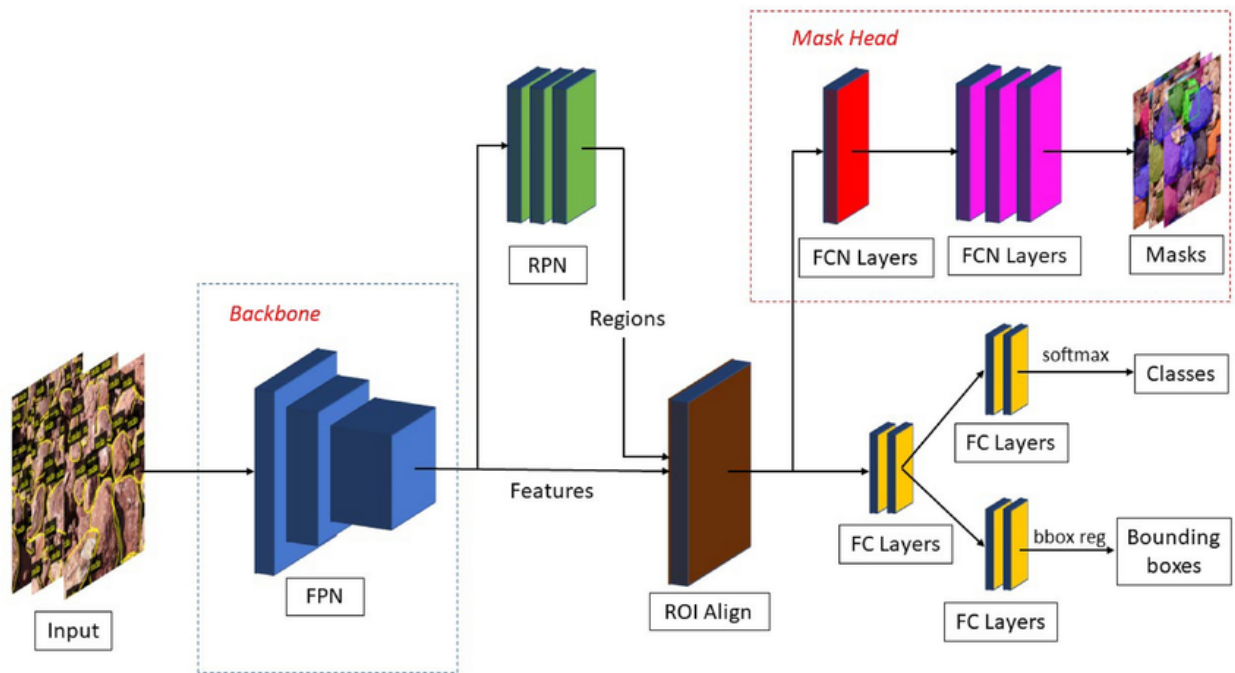


Рисунок 3.1 – Загальна структура базової моделі Mask R-CNN [5, 129]

Таким чином, базова нейромережева модель Mask R-CNN поєднує багатомасштабну екстракцію ознак, генерацію кандидатних областей, точне вирівнювання ROI-ознак та паралельне прогнозування рамки і маски екземпляру. Така організація є придатною основою для подальших модифікацій, спрямованих на підвищення якості детектування й сегментації у специфічних умовах сцени та для використання додаткових інформаційних шарів.

3.3 Удосконалення моделі Mask R-CNN для детектування та сегментації лабораторних тварин

3.3.1 Базовий екстрактор ознак ResNet

У двоетапних моделях детектування та сегментації, зокрема в Mask R-CNN, якість кінцевого результату значною мірою визначається

властивостями магістральної мережі, що виконує екстракцію ознак із вхідного зображення. Саме на цьому етапі формується багаторівневе представлення сцени, яке надалі використовується модулем генерації пропозицій регіонів, механізмом вирівнювання областей інтересу та окремими гілками класифікації, регресії рамки і сегментації. Отже, екстрактор ознак повинен одночасно забезпечувати достатню дискримінативність ознак, збереження просторової структури об'єктів та стійкість до локальних варіацій фону, освітлення і масштабу [5, 30, 98-99].

Одним зі способів підвищення якості згорткових мереж є збільшення їх глибини, оскільки глибші моделі можуть формувати складніші ієрархії ознак. Водночас зі зростанням кількості шарів виникає проблема деградації якості: збільшення глибини не завжди приводить до покращення результатів і може ускладнювати оптимізацію моделі через нестійке проходження градієнтів [30]. Для усунення цієї проблеми в моделі ResNet запропоновано підхід залишкового навчання, який полягає у тому, що мережа навчається не безпосередньому відображенню $H(x)$, а залишковій функції

$$F(x) = H(x) - x. \quad (3.12)$$

Тоді вихід блока обчислюється як

$$y = F(x, W) + x, \quad (3.13)$$

де x – вхідний тензор ознак;

$F(x, W)$ – залишкова функція, реалізована послідовністю згорткових шарів із параметрами W ;

y – вихід блока [30].

Такий підхід означає, що мережа навчається коригувати вхідне представлення, а не заново відтворювати повне відображення. Якщо

оптимальним виявляється близьке до тотожного перетворення, залишкова функція може набувати малих значень, а основний сигнал проходить через додатковий зв'язок без суттєвих втрат. Саме це покращує проходження градієнтів у глибокій мережі та робить навчання стабільнішим. У випадку, коли розмірності вхідного та вихідного тензорів не збігаються, тотожний зв'язок замінюється проєкційним перетворенням:

$$y = F(x, W) + W_s x, \quad (3.14)$$

де W_s – параметри лінійного перетворення, яке узгоджує розмірності ознак [30].

Базовим елементом моделі ResNet є залишковий блок, що складається з кількох згорткових шарів, між якими виконуються нормалізація та нелінійні перетворення, а також існує прямий обхідний зв'язок, який додає вхід до виходу згорткової частини. У найпростішому варіанті для ResNet18 і ResNet34 використовується так званий базовий блок (англ. basic block), який містить дві послідовні згортки 3×3 . Нехай x_l – вхід l -го блока, тоді його вихід можна подати у вигляді:

$$x_{l+1} = \sigma(F(x_l, W_l) + x_l), \quad (3.15)$$

де $\sigma(p)$ – нелінійна функція активації;

$F(x_l, W_l)$ – результат проходження згорткових операцій, нормалізації та активацій у межах блока.

За такого формування вихідний тензор одночасно містить як нові локальні ознаки, так і інформацію, що безпосередньо передається з попереднього рівня [30]. На рис. 3.2 показано приклад залишкового блока мережі ResNet.

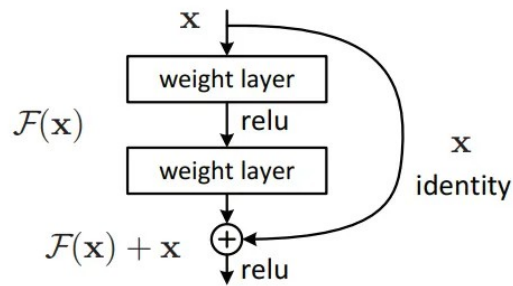


Рисунок 3.2 – Залишковий блок мережі ResNet [93]

Загальна структура ResNet будується як послідовність груп залишкових блоків, між якими поступово зменшується просторова роздільна здатність карт ознак і збільшується кількість каналів (рис. 3.3).

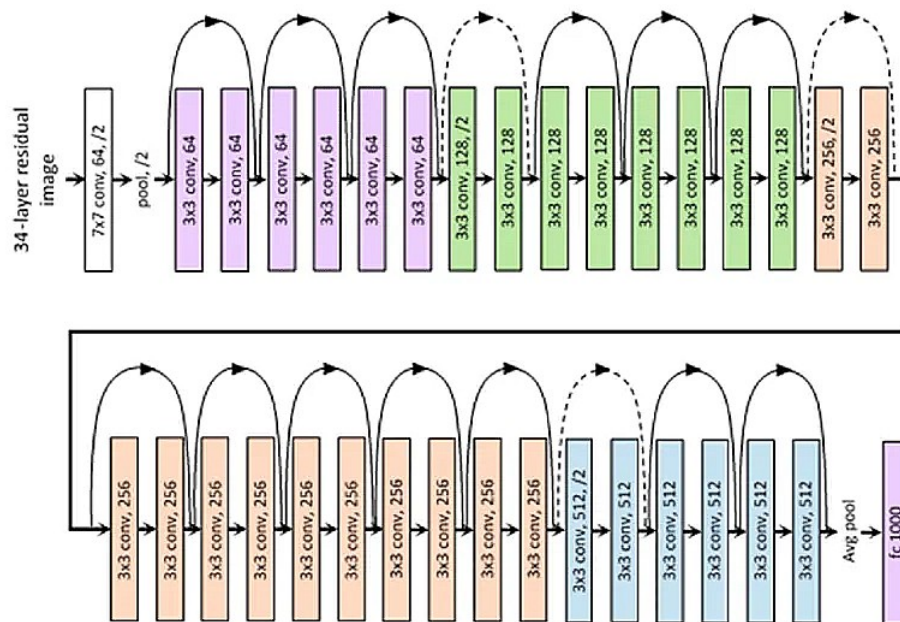


Рисунок 3.3 – Загальна структура мережі ResNet [93]

На ранніх етапах мережа виділяє низькорівневі ознаки (контури, локальні текстури, прості структури), а на глибших рівнях – більш абстрактні й семантично насичені представлення об'єктів. Для задач детектування така ієрархія є принципово важливою, оскільки локалізація об'єкта спирається одночасно і на точні просторові деталі, і на високорівневий контекст. Саме

тому в сучасних детекторах ознаки різних рівнів додатково узгоджуються в межах пірамід ознак [98-99].

У дослідженні як базову магістральну мережу обрано ResNet18. Ця нейромережева модель є однією з найкомпактніших у сімействі ResNet, але при цьому зберігає ключові властивості залишкового навчання та формування ієрархічних ознак [30]. Порівняно з глибшими варіантами, зокрема ResNet50 або ResNet101, ResNet18 має меншу обчислювальну складність, менший обсяг параметрів і є зручнішою основою для подальшого удосконалення моделі. Це важливо для запропонованої постановки задачі, де планується використання не одного, а двох паралельних потоків екстракції ознак, а отже надмірно складна магістральна мережа могла б істотно збільшити обчислювальні витрати та ускладнити навчання моделі в цілому.

Крім того, ResNet18 є достатньо глибокою для формування семантично інформативних ознак і водночас досить достатньо компактною для експериментального дослідження різних схем злиття ознак і підключення додаткових гілок. Цей баланс між глибиною, виразною здатністю та обчислювальною помірністю робить її доцільним вибором як базового екстрактора ознак в удосконаленій моделі Mask R-CNN. Таким чином, модель ResNet є ефективною основою для екстракції ознак у задачах детектування та сегментації завдяки використанню залишкових зв'язків, які полегшують навчання глибоких моделей і покращують стабільність передачі ознак між рівнями мережі. Використання ResNet18 як магістральної мережі забезпечує достатню якість ознак при прийнятній складності та створює підґрунтя для побудови подвійної магістралі екстракції ознак, що розглядається в наступному підрозділі.

3.3.2 Подвійна магістраль екстракції ознак

Базова конфігурація Mask R-CNN передбачає використання одного екстрактора ознак, який формує карти ознак з вхідного зображення та

передає їх у подальші модулі: генерацію пропозицій регіонів, вирівнювання ROI-ознак і голови детектування/сегментації [5]. У класичній постановці вхідним сигналом для такого екстрактора є триканальне зображення, а внутрішні представлення мережі оптимізуються під найбільш загальні ознаки об'єктів. Проте в прикладних відеосценах із складним фоном, нерівномірним освітленням і частими частковими перекриттями окремих екземплярів лабораторних тварин одного лише стандартного вхідного подання може бути недостатньо для формування стійких ознак, особливо якщо об'єкти мають слабо виражені межі або є візуально подібними. У такій ситуації доцільним є використання додаткових інформаційних шарів, які підсилюють відокремлення об'єкта від фону або акцентують специфічні властивості сцени (контрастні, різницеві, сегментаційні чи інші похідні представлення), а також побудова схеми екстракції ознак, здатної окремо інтерпретувати різні типи інформації [124, 128].

З цією метою пропонується подвійна магістральна схема екстракції ознак, побудована на основі двох паралельних мереж ResNet18. Її призначення полягає у розділенні обробки інформації на два потоки: перший потік орієнтується на базове вхідне подання (зображення), тоді як другий потік орієнтується на додаткові шари, сформовані на етапі попередньої обробки або отримані як проміжні результати інших модулів конвеєра. Таким чином, екстракція ознак виконується не в одному просторі, а в двох взаємодоповнювальних підпросторах, після чого ознаки поєднуються для формування узгодженого багатоканального представлення, придатного для детектування та сегментації [124].

Нехай I – базове зображення кадру, а A – набір додаткових шарів (карти ознак, різницеві шари, допоміжні представлення). Тоді пропонована подвійна магістраль задається двома операторами екстракції ознак:

$$F_A = \psi_A(I), \quad F_B = \psi_B(I), \quad (3.16)$$

де $\psi_A(p), \psi_B(p)$ – екстрактори ознак типу ResNet18;

F_A, F_B – відповідні багаторівневі карти ознак.

У термінах реалізації обидва екстрактори мають однакову базову структуру ResNet18, однак відрізняються вхідним сигналом (типом вхідних каналів) і, відповідно, розподілом ознак, які формуються на ранніх рівнях мережі. Це важливо, оскільки ранні рівні ResNet особливо чутливі до статистики вхідних даних і фактично виконують «первинне кодування» сигналу в простір ознак.

Основна ідея подвійної схеми полягає в тому, що базове зображення і додаткові шари часто несуть різну за природою інформацію. Наприклад, зображення містить повну інтенсивнісну та текстурну структуру сцени, тоді як додаткові шари можуть містити підсилену інформацію про контраст, межі, рух або відокремлення фону. Якщо об'єднати такі канали «на вході» в одну багатоканальну матрицю і подавати в один екстрактор, то мережа змушена одночасно узгоджувати різноманітні статистики каналів у спільному потоці фільтрів. Подвійна схема, навпаки, дозволяє навчати первинні фільтри кожної гілки відповідно до типу сигналу та формувати ознаки, оптимальні для кожного підпростору представлення. Далі ці ознаки можуть бути поєднані на більш високому рівні, де мережа вже оперує узагальненими просторовими структурами, і різниця статистик вхідних шарів не є настільки критичною.

У задачі сегментації екземплярів та детектування в складних сценах така стратегія має практичні переваги. По-перше, формується додатковий канал інформації про межі та структуру об'єкта, що підсилює відокремлення екземпляра від фону і підвищує стійкість у випадках часткових перекриттів. По-друге, зменшується ризик ситуацій, коли нейромережа «втрачає» об'єкт через тимчасове зниження контрасту або локальні освітлювальні артефакти, оскільки додаткові шари можуть містити стабільніші підказки про присутність об'єкта. По-третє, подвійна магістраль створює умови для більш гнучкого керування внеском кожного підпростору ознак, оскільки на етапі

поєднання ознак можна регулювати їхні ваги, правила злиття та умови пріоритетності.

З позиції подальшої інтеграції в Mask R-CNN подвійна магістраль розглядається як заміна стандартного екстрактора ознак. Її вихідні карти ознак повинні бути узгоджені за розмірами та масштабами, щоб залишатися сумісними з механізмами піраміди ознак (якщо вона застосовується), генерацією пропозицій регіонів і вирівнюванням ROI-ознак. Саме тому важливим етапом є узгодження ознак двох гілок і їхнє злиття у єдине представлення.

3.3.3 Кероване злиття ознак в удосконаленій нейромережевій моделі Mask R-CNN

Ключовим етапом побудови подвійної магістралі є злиття ознак двох паралельних гілок у таке спільне представлення, яке: зберігає просторову структуру ознак; не руйнує багатомасштабну організацію, необхідну для детектування та сегментації екземплярів; є сумісним із подальшими компонентами Mask R-CNN (модулем формування пропозицій регіонів, механізмом вирівнювання ROI-ознак і головами детектування/сегментації) [4, 5, 98, 99].

Для опису механізму злиття в межах подвійної магістралі введемо такі позначення: А-гілка відповідає за екстракцію ознак із базового зображення кадру, а В-гілка – за екстракцію ознак із додаткових інформаційних шарів, сформованих на попередніх етапах обробки. Позначення $B \rightarrow A$ у цьому підрозділі використовується для опису керованого впливу допоміжної В-гілки на представлення основної А-гілки: ознаки В-гілки не формують окремий вихід моделі, а застосовуються для уточнення та підсилення ознак А-гілки через механізм керованого SE-злиття.

У даній роботі для поєднання інформації використано кероване SE-злиття, тобто конкатенацію ознак двох гілок із подальшим канальним перезважуванням за механізмом стискання-активація (англ. Squeeze-and-Excitation) [104, 124]. Такий підхід дозволяє виконати змішування ознак з урахуванням даних: після конкатенації SE-модуль навчається визначати, які канали об'єднаного представлення є більш корисними на конкретному етапі мережі та для конкретного зображення, замість простого додавання ознак або їх фіксованого поєднання [124].

Нехай на певному рівні l дві гілки формують карти ознак F_A^l та F_B^l однакової просторової роздільної здатності:

$$H(F_A^l) = H(F_B^l), \quad W(F_A^l) = W(F_B^l), \quad (3.17)$$

Тоді на першому кроці виконується конкатенація по канальному виміру:

$$U^l = \text{Concat}(F_A^l, F_B^l), \quad (3.18)$$

де $U^l \in \mathbb{R}^{H \times W \times C}$ – об'єднана карта ознак (із сумарною кількістю каналів C).

Далі до U^l застосовується механізм керованого SE-злиття, що складається з трьох стандартних операцій: стиснення (squeeze), активація (excitation) та перезважування (scale) [104]. На етапі стиснення для кожного каналу формується глобальна характеристика (глобальний контекст) за просторовим усередненням:

$$z_c = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W U_{i,j,c}^l, \quad z \in \mathbb{R}^C. \quad (3.19)$$

Далі етап активації перетворює вектор z у набір ваг каналів $s \in \mathbb{R}^C$ через двошаровий перцептрон із «вузьким місцем» (коефіцієнт стиснення r , ReLU-активація δ та сигмоїдна нелінійність σ):

$$s = \sigma(W_2 \delta(W_1 z)), \quad W_1 \in \mathbb{R}^{\frac{C}{r} \times C}, \quad W_2 \in \mathbb{R}^{C \times \frac{C}{r}}. \quad (3.20)$$

Завершальною операцією є перезважування каналів об'єднаної карти ознак:

$$\tilde{U}_{i,j,c}^l = s_c \cdot U_{i,j,c}^l. \quad (3.21)$$

У результаті операції керованого SE-злиття формується карта $\tilde{U}^l$, у якій канали автоматично підсилюються або пригнічуються залежно від їхньої корисності для поточного кадру та поточного рівня ознак [104, 124]. Це є принциповою відмінністю від простого додавання $F_A^l + F_B^l$: кероване SE-злиття дозволяє зменшувати внесок шумових або погано масштабованих допоміжних ознак, зберігаючи при цьому просторову структуру і не вносячи значних обчислювальних витрат (SE-блок додає порівняно малу кількість параметрів) [104, 124].

У запропонованій схемі злиття використовується як кероване злиття «B→A»: інформація допоміжної гілки впливає на оновлення представлення основної гілки через спільне SE-перезважування каналів після конкатенації [124]. Це дозволяє підсилювати ті компоненти ознак, які відповідають об'єкту на складному фоні або при часткових перекриттях, і водночас пригнічувати фонові домішки. Важливо, що злиття може застосовуватися на кількох рівнях магістралі (на вході стадії або всередині стадії), що забезпечує багатомасштабне поєднання ознак і узгоджується з подальшим використанням піраміди ознак [98, 99, 104, 124].

На рис. 3.4 наведено фрагмент конвеєра удосконаленої неймережевої моделі Mask R-CNN від входу моделі до передачі результатів екстрактора ознак до модуля FPN.

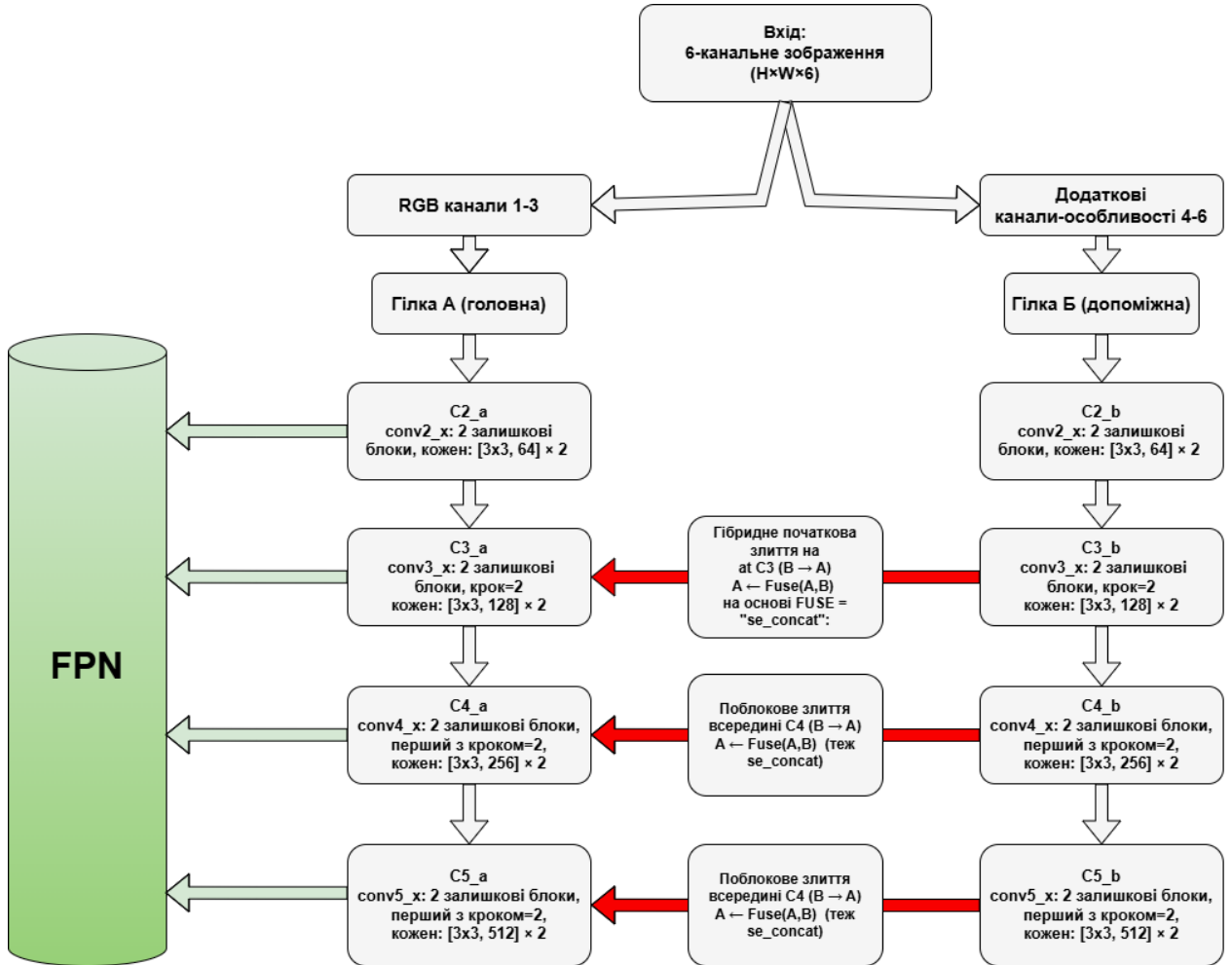


Рисунок 3.4 – Схема подвійної магістралі екстракції ознак із керованим SE-ЗЛИТТЯМ

Після отримання $\tilde{U}^l$ для забезпечення сумісності з подальшими блоками Mask R-CNN (зокрема з FPN-з'єднаннями та входом RPN) може застосовуватися канална проєкція (згортка 1×1) для приведення кількості каналів до потрібного значення:

$$F_{\text{fuse}}^l = \phi(\tilde{U}^l), \quad (3.22)$$

де $\phi(p)$ – оператор канального перетворення, який виконує змішування каналів і нормалізацію розмірності [104].

Сформовані об'єднані карти $\{F_{\text{fuse}}^l\}$ далі використовуються стандартним конвеєром Mask R-CNN: на їх основі модуль RPN формує пропозиції регіонів, RoIAlign формує ROI-ознаки, після чого паралельні голови виконують детектування (класифікація і регресія рамок) та сегментацію екземплярів [4, 5]. Таким чином, удосконалення локалізується на рівні екстракції та злиття ознак, тоді як подальша логіка двоетапного детектування і прогнозування масок зберігається.

У табл. 3.1 наведено структуру подвійної магістралі екстракції ознак із зазначенням основних етапів, просторового кроку, точок злиття $B \rightarrow A$ та виходу карт ознак до FPN.

Таблиця 3.1 – Структура подвійної магістралі екстракції ознак

Етап	Просторовий крок	А-гілка (основна)	В-гілка (додаткова)	Точки злиття ($B \rightarrow A$)	Вихід для FPN
1	2	3	4	5	6
Вхід	$1 \times$	Розділення 6-канального входу: А – канали 1-3, В – канали 4-6	те саме	–	–
C1 (вхідний)	$2 \times$	conv $7 \times 7/2$, BN, ReLU $\rightarrow$ 64 кан.; maxpool/2	те саме	–	–
C2 (шар 1)	$4 \times$	ResNet-18 шар 1 $\rightarrow$ 64 кан.	64 кан.	–	$C2(A) \rightarrow$ передається до FPN через 1×1 -згортку
C3 (шар 2)	$8 \times$	128 кан.	128 кан.	Злиття на вході стадії C3: $B \rightarrow A$ (SE-concat)	$C3(A') \rightarrow$ передається до FPN

Продовження таблиці 3.1

1	2	3	4	5	6
C4 (шар 3)	16×	256 кан.	256 кан.	Поблокове злиття всередині C4 після залишкових блоків	C4(A') → передається до FPN
C5 (шар 4)	32×	512 кан.	512 кан.	Поблокове злиття всередині C5 після залишкових блоків	C5(A') → передається до FPN

Отже, кероване SE-злиття забезпечує кероване, адаптивне до даних поєднання ознак двох гілок із мінімальним приростом обчислювальної складності та збереженням просторової структури ознак. Це створює передумови для підвищення якості детектування та сегментації екземплярів в кадрах зі складним фоном і частковими перекриттями, що було перевірено експериментально [124].

3.4 Навчання, налаштування та оцінювання удосконаленої нейромережевої моделі Mask R-CNN

Експериментальні дослідження модифікованої нейромережевої моделі виконувалися з метою (1) перевірки працездатності запропонованої схеми екстракції та злиття ознак у складі Mask R-CNN, (2) кількісного порівняння з базовими конфігураціями Mask R-CNN із різними екстракторами ознак та (3) оцінювання компромісу «якість детектування та сегментації – оперативність». При цьому протокол експериментів побудовано таким чином, щоб забезпечити відтворюваність: фіксувалися параметри навчання, обсяг вибірок, режим аугментації, а також умови виконання інференсу (включно з підходом до вимірювання часу обробки кадру) [5, 124].

Для навчання й оцінювання моделі використано набір зображень лабораторних тварин у форматі JPG, сформований із відеозаписів контрольованих сцен спостереження. Базовий набір було розділено на 340 тренувальних та 13 валідаційних зображень. Для збільшення різноманітності навчальних прикладів застосовано аугментацію, зокрема доповнення шляхом обертання зображень, що дозволяє підвищити стійкість моделі до зміни орієнтації об'єкта та частково компенсувати обмежений обсяг вихідної вибірки [124]. Загалом обсяг навчальної вибірки у результаті аугментації було збільшено у 4 рази.

Удосконалення моделі Mask R-CNN полягає у використанні розширеного багатоканального представлення кадру, подвійної магістралі екстракції ознак та керованого SE-злиття. Окрім базового зображення, у модель подаються додаткові інформаційні шари, зокрема зображення в градаціях сірого, різницеві зображення та представлення з корекцією тіней, які формуються на етапах попередньої обробки та описані в розділі 2. Такий підхід дозволяє подати в мережу не лише «сирі» інтенсивності, а й похідні сигнали, що підсилюють відокремлення об'єкта від фону й стабілізують контурні ознаки в складних сценах [124].

Усі сесії навчання та тестування виконувалися на локально доступному обладнанні з метою коректного порівняння продуктивності та можливостей застосування в однакових умовах. Експерименти проводилися на системі з процесором Intel Core i9-13980HX, 64 ГБ оперативної пам'яті та відеокартою NVIDIA GeForce RTX 4090 [124]. З огляду на обмеження пам'яті графічного процесора, розмір пакета навчання (batch size) для різних конфігурацій доводилося підбирати таким чином, щоб забезпечити стабільне навчання моделі.

Як базове порівняння було навчено кілька варіантів Mask R-CNN із різними екстракторами ознак (ResNet101, ResNet50, ResNet34, ResNet18). Для цих базових конфігурацій кількість епох навчання було встановлено на рівні 30, що відповідає моменту, коли приріст якості перестає бути суттєвим (за

обраними метриками на валідаційній вибірці) [124]. Для удосконаленої нейромережевої моделі Mask R-CNN із подвійною магістраллю та керованим SE-злиттям ознак кількість епох було збільшено до 50, оскільки ускладнення структури екстрактора ознак і поява додаткового механізму злиття вимагають більш тривалого навчання для стабілізації ваг та узгодження внеску обох гілок [104, 124].

Навчання проводилося за стандартною для Mask R-CNN багатокomпонентною цільовою функцією, яка включає складові, пов'язані з класифікацією, регресією параметрів обмежувальної рамки та прогнозуванням маски екземпляра [5]. У підрозділі 3.5 ці конфігурації порівнюються між собою за показниками якості та часовими витратами.

Оцінювання якості детектування виконувалося з використанням метрики $mAP50$, тобто середньої точності при пороговому значенні перекриття 0.5, що узгоджується з практикою оцінювання моделей детектування на поширених бенчмарках [106, 107]. Для оцінювання якості сегментації додатково застосовано показники, що обчислюються на основі матриці неточностей: точність (Accuracy), точність позитивних передбачень (Precision) та відгук (Recall). Окремо оцінювалася оперативність моделей шляхом вимірювання середнього часу виконання інференсу (виведення) на кадрі. Щоб зменшити вплив разових накладних витрат, час вимірювався без урахування часу ініціалізаційного запуску, тобто перший прогін моделі не включався в усереднення [8]. Такий протокол є важливим для коректного порівняння конфігурацій з різними магістралями, оскільки дозволяє зіставляти їх у режимі, близькому до практичного застосування.

3.5 Результати детектування та сегментації лабораторних тварин із використанням удосконаленої нейромережевої моделі Mask R-CNN

Експериментальне дослідження було спрямоване на перевірку ефективності запропонованого нейромережевого підходу до детектування та

сегментації екземплярів, а також на порівняння базових конфігурацій Mask R-CNN з різними магістралями екстракції ознак із модифікованим варіантом на основі подвійної магістралі ResNet18 зі керованим SE-злиттям ознак. Оцінювання виконувалося в однакових умовах для всіх конфігурацій: навчання проводилося на одній і тій самій навчальній вибірці, тестування – на валідаційній частині, а оперативність визначалась як середній час виведення без урахування першого запуску (розігріву) моделі [8].

На першому етапі було навчено та протестовано базові варіанти Mask R-CNN з магістралями ResNet101, ResNet50, ResNet34 та ResNet18. Через обмеження пам'яті графічного процесора розмір пакета навчання відрізнявся залежно від конфігурації (чим більша магістраль, тим менший batch size), тоді як кількість епох для базових конфігурацій фіксувалась на рівні 30 – тобто на значенні, коли приріст точності переставав бути суттєвим [8]. Зведені результати для базових конфігурацій наведено в табл. 3.2.

Таблиця 3.2 – Навчання та оцінювання ефективності базової нейромережевої моделі Mask R-CNN

Метрика	Значення			
Нейромережева модель	Mask R-CNN	Mask R-CNN	Mask R-CNN	Mask R-CNN
Основа (екстрактор ознак)	ResNet18	ResNet34	ResNet50	ResNet101
Розмір пакета навчання	4	6	4	2
Тренувальні епохи (од.)	30	30	30	30
mAP50, рамка	0.85	0.86	0.93	0.96
Precision, рамка	0.86	0.89	0.97	0.98
Recall, рамка	0.9	0.89	0.93	0.95
mAP50, маска	0.83	0.83	0.92	0.91
Precision, маска	0.85	0.87	0.96	0.93
Recall, маска	0.89	0.87	0.91	0.9
Середній час виведення, мс (без ініціалізаційного запуску)	104.9	91.1	101.8	102.5
Кількість параметрів екстрактора, млн. (прибл.)	11.2	21.3	23.6	42.7
Кількість параметрів повної моделі, млн. (прибл.)	31.5	41.7	44.7	63.7

Як видно з табл. 3.2, найбільші значення mAP50 для масок серед базових конфігурацій показали варіанти з ResNet101 та ResNet50 (0.91 та 0.92 відповідно), що відповідає очікуваному ефекту від більш потужного

екстрактора ознак. Водночас різниця між цими двома варіантами за часом виведення є незначною (102.5 мс та 101.8 мс), тобто в даних умовах обмеження швидкодії визначається не лише глибиною магістралі, а й сумарною структурою моделі та процедурою постобробки результатів [124].

Варіанти з меншою магістраллю (ResNet34 та ResNet18) демонструють суттєве зменшення якості за mAP50 (до 0.83), однак є потенційно привабливими як основа для модифікації через нижчу параметричну складність. При цьому ResNet34 показує найменший час виведення (91.1 мс), але поступається за точністю сегментації та детектування, тоді як ResNet18 має mAP50 0.83 та час виведення 104.9 мс, тобто не дає переваги по швидкодії, але при цьому зберігає найменшу складність серед розглянутих магістралей [124]. З практичної точки зору це означає, що просте «спрощення магістралі» не гарантує одночасного виграшу і в якості, і в оперативності; саме тому модифікація повинна бути спрямована на якісніше використання вхідної інформації та на підсилення ознак, релевантних для об'єкта.

Для удосконаленої нейромережевої моделі Mask R-CNN з подвійною магістраллю екстрактора ознак і керованим SE-злиттям кількість епох навчання було збільшено до 50, що відповідає необхідності стабілізувати параметри додаткових блоків і узгодити внесок двох гілок екстракції ознак [104, 124]. Результати наведено в табл. 3.3.

Таблиця 3.3 – Навчання та оцінювання ефективності удосконаленої моделі Mask R-CNN із подвійною магістраллю ResNet18

Метрика	Значення
1	2
Розмір пакета навчання	4
Тренувальні епохи (од.)	50
mAP50, рамка	0.87
Precision (Точність), рамка	0.91
Recall (Відгук), рамка	0.91
mAP50, маска	0.9

Продовження таблиці 3.3

1	2
Precision (Точність), маска	0.92
Recall (Відгук), маска	0.92
Середній час виведення, мс (без ініціалізаційного запуску)	105.5
Кількість параметрів екстрактора, млн. (прибл.)	24.4
Кількість параметрів повної моделі, млн. (прибл.)	44.7
Кількість параметрів ядра екстрактора без злиття, млн. (прибл.)	22.4

Порівняння табл. 3.2 та табл. 3.3 показує, що удосконалена модель забезпечує суттєвий приріст якості відносно базової конфігурації з ResNet18. Для рамок mAP50 зростає з 0.85 до 0.87, Precision – з 0.86 до 0.91, Recall – з 0.90 до 0.91. Для масок приріст є ще виразнішим: mAP50 зростає з 0.83 до 0.90, Precision – з 0.85 до 0.92, Recall – з 0.89 до 0.92 [8]. Отже, модифікація дозволяє істотно підвищити насамперед якість сегментації екземплярів при використанні ResNet18 як базової магістралі, що є важливим для подальших етапів трекінгу та підтримання ID.

Одночасно слід відзначити, що за часом виведення модифікована модель (105.5 мс) залишається на рівні базового варіанту з ResNet18 (104.9 мс) і близька до конфігурацій з ResNet50 та ResNet101 (≈ 102 мс) [124]. Це означає, що в даних умовах основний ефект модифікації проявляється саме у підвищенні якості детектування та сегментації, тоді як виграш у швидкодії не є ключовим фактором. З практичної точки зору такий результат є прийнятним, оскільки метою модифікації було наблизити якість «легшого» екстрактора до рівня більш важких конфігурацій, зберігаючи при цьому можливість подальшої інтеграції додаткових каналів/ознак і подальших модулів конвеєра [124]. На рис. 3.5 показано приклад детектування лабораторних тварин (бичків), включаючи у випадку «злиття» 2 бичків в 1 об'єкт при близькому розташуванні.

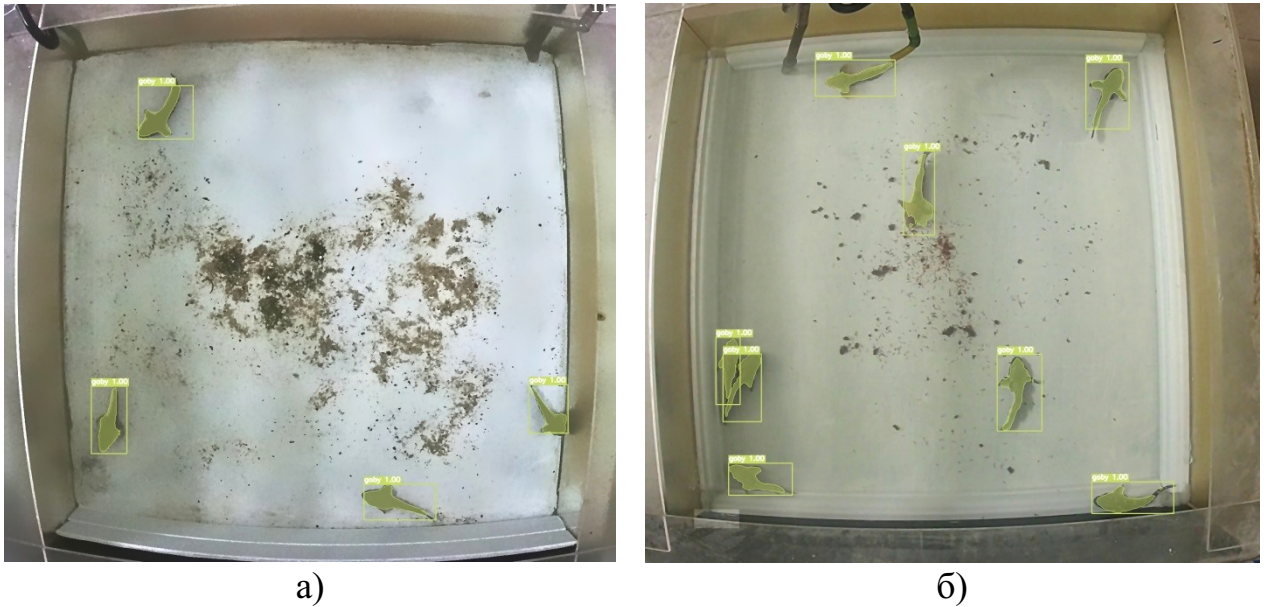


Рисунок 3.5 – Візуалізація роботи удосконаленої моделі Mask R-CNN із подвійною магістраллю екстракції ознак і керованим SE-злиттям ознак на кадрах двох різних відео а) та б).

Порівняльний аналіз показує також, що запропонована нейромережева модель забезпечує якість прогнозу масок, дуже близьку до базових конфігурацій з ResNet50/ResNet101 за mAP50 (0.90 проти 0.91 та 0.92 відповідно), при цьому зберігає перевагу в параметричній складності перед ResNet101 та допускає роботу з багатоканальними зображеннями, де додаткові канали можуть відображати попередньо обчислені характеристики сцени [124]. Останнє є суттєвим для розв'язання задач детектування і сегментації в умовах нечіткого фону та часткових перекриттів: розширене подання кадру дозволяє підсилювати інформативні компоненти (зокрема ті, що підкреслюють об'єкт відносно фону), а механізм керованого SE-злиття забезпечує адаптивне каналне перезважування, що сприяє зменшенню впливу фонових домішок [104, 124].

Отже, експериментальні результати підтверджують доцільність удосконалення Mask R-CNN шляхом введення подвійної магістралі з адаптивним злиттям ознак: при використанні «легкого» базового екстрактора досягається суттєве покращення якості детектування й сегментації

екземплярів, а отримані показники наближаються до рівня більш важких базових конфігурацій. Це створює підґрунтя для подальшого розвитку підходу в частині інтеграції модулів, що використовують маски та ROI-ознаки для підтримання відповідності об'єктів у часі та підвищення стійкості трекінгу в складних сценах [124].

3.6 Висновки до третього розділу

У третьому розділі розроблено та обґрунтовано удосконалену нейромережеву модель Mask R-CNN для детектування та сегментації екземплярів лабораторних тварин у відеоданих. Показано, що для подальшого трекінгу, формування треклетів і повторної ідентифікації важливим є не лише формування обмежувальних рамок, а й отримання піксельних масок екземплярів, які забезпечують більш інформативне просторове представлення об'єкта.

Обґрунтовано вибір Mask R-CNN як базової моделі нейромережевого детектування та сегментації. Її перевагами для поставленої задачі є двоетапна регіон-орієнтована структура, наявність механізму RoIAlign, окрема гілка прогнозування маски та можливість цілеспрямованої модифікації екстрактора ознак без порушення загальної логіки детектування й сегментації.

Для підвищення якості ознакового представлення запропоновано удосконалення екстрактора ознак на основі подвійної магістралі ResNet18. Перша гілка обробляє базове зображення кадру, а друга – додаткові інформаційні шари, сформовані на попередніх етапах обробки. Такий підхід дозволяє окремо інтерпретувати різномірні типи інформації та формувати більш стійке багатоканальне представлення об'єкта в умовах складного фону, змін освітлення та часткових перекриттів.

Розроблено механізм керованого злиття SE-concat ознак двох гілок за схемою $B \rightarrow A$. У межах цього підходу ознаки допоміжної B-гілки

використовуються для уточнення та підсилення представлення основної А-гілки. Кероване SE-злиття забезпечує адаптивне каналне перезважування об'єднаних ознак, що дозволяє підсилювати інформативні канали та зменшувати вплив шумових або фонових компонентів.

Проведено навчання та експериментальне оцінювання базових конфігурацій Mask R-CNN із магістралями ResNet18, ResNet34, ResNet50 і ResNet101, а також удосконаленої моделі з подвійною магістраллю ResNet18 та керованим SE-злиттям ознак. Результати показали, що запропонована модель забезпечує суттєве покращення якості порівняно з базовою Mask R-CNN на основі ResNet18. Для масок mAP50 зросла з 0.83 до 0.90, Precision – з 0.85 до 0.92, Recall – з 0.89 до 0.92. Для рамок також отримано покращення: mAP50 зросла з 0.85 до 0.87, Precision – з 0.86 до 0.91, Recall – з 0.90 до 0.91.

Встановлено, що удосконалена нейромережева модель за якістю сегментації екземплярів наближається до базових конфігурацій з більш складними магістралями ResNet50 та ResNet101, водночас зберігаючи прийнятний час виведення на кадрі. Це підтверджує доцільність використання подвійної магістралі ResNet18 із керованим SE-злиттям як компромісного рішення між якістю детектування/сегментації та обчислювальною складністю.

Отже, результати третього розділу підтверджують ефективність удосконалення нейромережевої моделі Mask R-CNN шляхом введення подвійної магістралі екстракції ознак і керованого SE-злиття. Отримані детекції, обмежувальні рамки, маски екземпляра та ознакові представлення створюють основу для подальшого трекінгу, формування треклетів і збереження ідентичності лабораторних тварин у відеопослідовності.

4 КОНЦЕПТУАЛЬНА МОДЕЛЬ MASK-TAUDL ДЛЯ НАСКРІЗНОЇ ІДЕНТИФІКАЦІЇ ЛАБОРАТОРНИХ ТВАРИН У ВІДЕОДАНИХ БЕЗ РУЧНОГО МАРКУВАННЯ

4.1 Формування сегментованих треклетів і ознак для повторної ідентифікації

У попередніх розділах було сформовано основу для побудови наскрізного конвеєра відеоаналізу лабораторних тварин. У розділі 2 розроблено методи детектування та локалізації на основі суперпіксельної сегментації SLIC та її удосконалення, що забезпечують формування допоміжних масок, обмежувальних рамок і ознакових шарів за умов нестаціонарного фону та локальних артефактів. У розділі 3 розроблено удосконалену нейромережеву модель Mask R-CNN із подвійною магістраллю екстракції ознак та керованим SE-злиттям, яка забезпечує формування основних детекцій, масок екземплярів лабораторних тварин та ROI-представлень для подальшого трекінгу, формування треклетів і повторної ідентифікації.

Разом з тим, коректне детектування та сегментація об'єктів в окремих кадрах ще не гарантують збереження сталої ідентичності лабораторних тварин упродовж усієї відеопослідовності. У практичних записах із частими перекриттями, дотиками об'єктів, злиттям і подальшим розділенням, а також зі змінами їх видимого контуру трекінг нерідко призводить до фрагментації траєкторій і перемикання ідентичностей. Для задач аналізу поведінки це є критичним, оскільки навіть короткочасне порушення відповідності між об'єктом та його траєкторією спотворює кількісні метрики: час перебування в зоні, довжину шляху, кількість і тривалість контактів та інші показники взаємодії між особинами [84, 97, 127, 129].

Практично зручним представленням для подальшого відновлення ідентичності є треклети – короткі стійкі фрагменти траєкторій, які

формуються між подіями невизначеності: перекриттями, пропусками детектування, злиттям або розділенням об'єктів. Перехід від покадрової асоціації до треклетного рівня дозволяє агрегувати інформацію про об'єкт на певному проміжку часу, підвищити стабільність ознак зовнішнього вигляду та зменшити вплив одиничних помилок детектування або сегментації на подальше зіставлення [23, 86, 87, 96, 128, 130].

Отже, у цьому розділі запропоновано концептуальну модель Mask-TAUDL для наскрізної ідентифікації лабораторних тварин у відеоданих без використання ручного маркування. Її призначення полягає у формуванні нормалізованих векторів вбудовувань на основі масково-зважених ROI-ознак, отриманих із результатів сегментації екземплярів об'єктів, та подальшій асоціації треклетів для збереження сталої ідентичності об'єктів у відеопослідовності. При цьому джерелом слабкого нагляду виступає структура самих треклетів: кадри в межах одного треклета розглядаються як такі, що належать одній особині, а подальше зіставлення треклетів виконується за подібністю сформованих вбудовувань без явних ID-міток [115, 116, 131-137].

4.1.1 Отримання треклетів із результатів детектування, сегментації та трекінгу

Після виконання детектування та сегментації екземплярів лабораторних тварин кожен кадр відеопослідовності містить набір локалізованих об'єктів, поданих у вигляді обмежувальних рамок, масок екземплярів та оцінок достовірності. Саме ці результати, сформовані на попередніх етапах конвеєра, є вхідними даними для покадрового трекінгу та подальшого формування треклетів. У прикладних відеосценах із частими перекриттями, дотиками й тимчасовими частковими зникненнями об'єкта трекінг рідко формує одну безперервну траєкторію на весь запис. Замість цього зазвичай виникають фрагменти треків, розділені моментами

невизначеності (взаємні перекриття, злиття силуетів, різка зміна видимого контуру, короткі відмови детектування). Для таких умов доцільно переходити до рівня треклетів – коротких надійних ділянок треку, в межах яких відповідність об'єкта в часі вважається достовірною і може бути використана для накопичення ознак зовнішнього вигляду [115, 116].

Нехай для кожного кадру t після детектування та сегментації екземплярів об'єктів отримано множину детектованих екземплярів об'єктів у вигляді:

- обмежувальної рамки b_t^k ,
- бінарної маски m_t^k ,
- оцінки достовірності s_t^k ,

де k – номер екземпляра у кадрі.

Трекінг виконує покадрове зіставлення цих екземплярів та формує послідовності, для яких збережено часову узгодженість. Треклетом вважається відрізок такої послідовності на інтервалі кадрів $[t_s, t_e]$, в межах якого не зафіксовано подій, що істотно підвищують ризик помилки відповідності: тривалого перекриття, злиття екземплярів, пропуску детектування або різкої зміни видимої форми об'єкта. Тоді треклет можна подати як множину:

$$\tau_i = \{(b_t, m_t s_t)\}_{t=t_s}^{t_e}. \quad (4.1)$$

Практична цінність треклетів полягає в тому, що вони дозволяють перейти від «миттєвих» покадрових ознак до часово агрегованих ознак, стійкіших до шуму та локальних артефактів. Навіть якщо в окремих кадрах маска або рамка частково спотворені, наприклад через тінь, локальний відблеск або короткочасне перекриття, агрегація ознак у межах треклета дає більш стабільний опис об'єкта, який зручніше використовувати для повторної ідентифікації [115, 119]. Це особливо важливо для лабораторних тварин із високою візуальною подібністю, коли індивідуальні відмінності

між особинами слабо виражені, а фонові домішки можуть переважати у звичайних ознаках, отриманих лише з обмежувальної рамки.

4.1.2 Масково-зважені ROI-ознаки та агрегація ознак у межах треклета

Ефективність повторної ідентифікації на рівні треклетів визначається насамперед тим, які ознаки використовуються для опису лабораторної тварини. Якщо формувати ознаки лише за обмежувальною рамкою, то до опису неминуче потрапляють фонові пікселі, тіні, відбиття та інші артефакти сцени. Це підвищує варіативність ознак однієї й тієї самої особини в часі та зменшує роздільність ознак різних особин, особливо коли вони візуально подібні. Тому в запропонованій моделі Mask-TAUDL результати сегментації екземплярів розглядаються не лише як засіб геометричної локалізації, а і як основа для зменшення впливу фону під час формування ознак повторної ідентифікації [115].

Нехай для кадру t та детектування k відомі: обмежувальна рамка b_t^k і маска екземпляра m_t^k . Після вирівнювання області інтересу за допомогою операції ROIAlign формується тензор ознак H_t^k фіксованого розміру, який містить каналні ознаки об'єкта в межах рамки. Маска m_t^k приводиться до того самого просторового розміру, що й ROI-представлення, утворюючи $\hat{m}_t^k$. Використання маски можна подати як просторове пригнічення фонові складової у межах ROI:

$$H_t'^k(i, j, c) = H_t^k(i, j, c) \cdot \hat{m}_t^k(i, j), \quad (4.2)$$

де c – номер каналу;

(i, j) – координати у просторовій решітці ROI.

Таким чином, подані просторові позиції, що відповідають фону, мають нульовий або зменшений внесок у подальше формування дескриптора, тоді як ознаки всередині маски визначають основний зміст представлення об'єкта, а отже, маска екземпляру виконує функцію очищення ознак від фонових домішок [115].

Для отримання компактного векторного опису об'єкта просторово зважений тензор $H_t'^k$ може бути агрегований у дескриптор шляхом усереднення по простору:

$$r_t^k(c) = \frac{\sum_i \sum_j H_t'^k(i, j, c)}{\sum_i \sum_j \hat{m}_t^k(i, j) + \varepsilon}, \quad (4.3)$$

де ε – малий додатний коефіцієнт для уникнення ділення на нуль.

Отриманий вектор r_t^k описує об'єкт переважно за його формою, контуром і текстурними характеристиками в межах маски та в меншій мірі залежить від навколишнього фону. Така інтерпретація підкреслює роль інстанс-сегментації в підсистемі повторної ідентифікації: маска екземпляра дозволяє зосередитися на релевантній частині зображення та зменшити вплив випадкових візуальних домішок. У моделі Mask-TAUDL це є принциповим, оскільки маска екземпляра використовується як механізм пригнічення фонових домішок, що підвищує роздільну здатність ознак для повторної ідентифікації [115].

У запропонованій моделі процес формування ознак описується через ROI-представлення, отримані з удосконаленої Mask R-CNN. Нехай F_t – тензор ознак, отриманий із сегментатора для кадру t . Для k -го об'єкта з рамкою b_t^k формується ROI-представлення

$$R_t^k = \text{ROIAlign}(F_t, b_t^k), \quad (4.4)$$

після чого з нього будується дескриптор кадру r_t^k , що описує зовнішній вигляд об'єкта на цьому кадрі.

У такій постановці результати сегментації екземпляру забезпечують точнішу локалізацію об'єкта, стабільніше виділення області інтересу та зменшення впливу фону, тоді як самі ознаки надалі агрегуються в межах треклета [115].

Наступним кроком є агрегація ознак у межах треклета. Оскільки повторна ідентифікація виконується не для окремих кадрів, а для треклетів, ознаки мають бути узагальнені в часі. Нехай деякий треклет τ_i охоплює кадри $t_s, \dots, t_e$, і для кожного кадру отримано вектор r_t . Тоді треклетний дескриптор $\bar{r}_i$ визначається як середнє:

$$\bar{r}_i = \frac{1}{N} \sum_{t=t_s}^{t_e} r_t, \quad N = t_e - t_s + 1. \quad (4.5)$$

Часова агрегація зменшує вплив одиничних похибок: навіть якщо на окремому кадрі маска частково спотворена через короткочасне перекриття, відблиск або зміну освітлення, усереднений треклетний опис зберігає стабільність. Крім того, у межах треклета можна застосовувати вагову агрегацію з урахуванням достовірності s_t : кадри з низькою достовірністю мають менший внесок у дескриптор. Така ідея є природною, оскільки помилкові або нестійкі детектування не повинні домінувати в ознаковому поданні об'єкта.

Таким чином, виходом етапу формування ознак є набір агрегованих треклетних дескрипторів $\{\bar{r}_i\}$, побудованих на основі маско-зважених ROI-ознак і узагальнених у часі в межах відповідних треклетів. Саме ці дескриптори виступають основою для подальшого формування компактних вбудовувань та оцінювання якості пошуку за подібністю без ручного маркування [96, 116, 128]. З практичної точки зору агрегування ROI-ознак у

межах треклета створює більш стійке представлення лабораторної тварини, оскільки зменшує вплив локальних похибок окремих кадрів і часткових перекриттів. Це безпосередньо підвищує якість асоціації треклетів у режимі без ручного маркування, оскільки подальші вбудовування стають менш чутливими до фону і більше відображають характеристики самої особини, що є необхідною умовою для надійного відновлення ідентичності у сценах з високою візуальною подібністю особин та частими взаємодіями.

4.1.3 Формування вбудовувань та нормалізація

Після формування масково-зважених ознак і їх часової агрегації в межах треклета необхідно перейти до такого компактного подання, яке зручно використовувати для порівняння треклетів і подальшої асоціації. Для цього треклетний дескриптор перетворюється у вбудовування (ембединг) фіксованої розмірності. У моделі Mask-TAUDL вбудовування розглядається як вектор ознак, який повинен бути достатньо інваріантним до локальних змін зовнішнього вигляду та умов зйомки, але водночас роздільним для різних особин. Саме на такі вектори надалі накладаються цілі навчання без учителя за принципом асоціації треклетів [96, 115].

Нехай $\bar{r}_i$ – агрегований дескриптор треклета τ_i , отриманий за (4.5). Далі застосовується перетворення в простір фіксованої розмірності D за допомогою модуля повторної ідентифікації (лінійного або малогабаритного нелінійного відображення):

$$e_i = W\bar{r}_i + b, \quad e_i \in \mathbb{R}^D. \quad (4.6)$$

У проведених дослідженнях як типовий розмір простору ознак використовується $D = 256$, що є достатнім для формування дискримінативного подання при помірних обчислювальних витратах [115]. Вектор e_i ще не є безпосередньо придатним для порівняння, оскільки його

масштаб може змінюватися залежно від довжини треклета, якості детектувань і внутрішньої статистики ознак. Тому застосовується $L2$ -нормалізація:

$$f_i = \frac{e_i}{\|e_i\|_2}. \quad (4.7)$$

Після нормалізації всі вбудовування мають однакову довжину, а порівняння векторів стає більш стабільним. На практиці це дозволяє використовувати як міру подібності косинусну близькість:

$$\text{sim}(f_i, f_j) = f_i^T f_j, \quad (4.8)$$

або еквівалентну їй евклідову відстань між нормалізованими векторами:

$$d(f_i, f_j) = \|f_i - f_j\|_2. \quad (4.9)$$

Умовою «близькості» треклетів може бути або перевищення косинусної подібності заданого порогу, або належність до множини найближчих сусідів у просторі вбудовувань.

Оскільки вбудовування формується з агрегованих ознак треклета, воно узагальнює інформацію про зовнішній вигляд об'єкта на часовому проміжку та є менш чутливим до одиничних похибок. Додатково, у випадку коротких треклетів доцільно враховувати їх надійність (довжину, середню достовірність детектувань) як ваговий коефіцієнт під час побудови множин навчальних пар або під час формування міждтреклетних зв'язків у процесі навчання без ручного маркування. Це дозволяє зменшувати вплив треклетів, сформованих у складних кадрах (наприклад, при сильних перекриттях), на структуру простору ознак.

Таким чином, виходом етапу формування ознак є набір нормалізованих треклетних вбудовувань $\{f_i\}$, які будуються з масково-зважених ROI-ознак, агрегуються в часі та приводяться до єдиного масштабу $L2$ -нормалізацією. Ці вектори виступають базовими об'єктами оптимізації в навчанні без ручного маркування TAUDL: модель повинна зменшувати відстань між вбудовуваннями треклетів однієї особи та збільшувати відстань між вбудовуваннями різних осіб, використовуючи як джерело нагляду структуру треклетів і механізми асоціації без ручного маркування ID [116].

4.1.4 Дискримінація та асоціація треклетів без ручного маркування на основі принципів TAUDL

В умовах однієї камери базовим механізмом навчання без ручного маркування ознак для повторної ідентифікації є дискримінація треклетів у межах однієї сцени або сеансу спостереження, що відповідає принципу Per-Camera Tracklet Discrimination (PCTD) [116]. Ідея полягає в тому, що треклет уже містить слабкий нагляд: усі кадри всередині одного треклета належать одній і тій самій особі (принаймні на короткому інтервалі, де трекінг є надійним). Це дозволяє організувати навчання без ручних ID-міток, використовуючи сам факт належності кадрів до треклета як навчальний сигнал.

На першій ітерації кожному треклету τ_i призначається псевдомітка y_i , яка трактується як «клас треклета». Далі мережа навчається так, щоб вбудовування кадрів, що належать одному треклету, групувалися разом, а вбудовування різних треклетів розділялися. Практично це реалізується як задача багатокласової класифікації: для набору вбудовувань, сформованих із кадрів треклета (або з його агрегованого представлення), обчислюється розподіл імовірностей за псевдокласами, і оптимізація підсилює правильний псевдоклас відповідного треклета [116]. Така постановка дає два ефекти: усередині треклета ознаки стають більш компактними (менше

«розмазування» через шум); різні треклети відштовхуються один від одного, що підвищує роздільну здатність ознакового простору.

Окремо важливо, що в розглянутій системі вбудовування формуються з масково-зважених ROI-ознак (п. 4.1.2), тобто фон пригнічується за маскою екземпляра. Це підвищує якість PCTD саме в лабораторних сценах, де тіні, відбиття та структура підкладки можуть формувати помилкові ідентифікаційні сигнали при використанні лише обмежувальної рамки [115]. Внаслідок цього навчання в просторі ознак орієнтується на стабільні характеристики лабораторної тварини, а не на випадкові фрагменти фону.

Разом з тим, потрібно враховувати принципове обмеження: через перекриття й розриви трекінгу одна й та сама особина може породжувати кілька різних треклетів, які на старті отримують різні псевдомітки. Отже, PCTD формує корисну початкову структуру ознакового простору, але сам по собі не гарантує об'єднання фрагментів траєкторії однієї особини в єдиний ID на довгих інтервалах. Тому наступним кроком вводиться механізм асоціації треклетів за принципом ССТА, адаптований до однокамерного випадку як міжфрагментна або міжсесійна асоціація [116].

У класичній постановці TAUDL механізм ССТА (Cross-Camera Tracklet Association – асоціація треклетів між камерами) призначений для узгодження треклетів, які належать одному й тому самому об'єкту, але спостерігаються в різних камерах, тобто за різних ракурсів та умов зйомки) [116]. Для однокамерної лабораторної сцени фізичного переходу між камерами немає, однак у довготривалих експериментах запис часто природно поділяється на окремі сесії або кліпи, що можуть відрізнятися часом зйомки, умовами освітлення, дрейфом фону та якістю кадрів. Саме такі сесії доцільно розглядати як умовні домени, між якими потрібно зберігати сталі ID лабораторних тварин. Тому ССТА інтерпретується як міжсесійна асоціація треклетів, тобто встановлення відповідностей між треклетами різних сесій, які з високою ймовірністю належать одній і тій самій особині [115, 116].

Таким чином, PCTD та CCTA в межах дослідження доцільно розглядати як загальну основу TAUDL, адаптовану до однокамерного сценарію. На практиці це означає, що після детектування та сегментації екземплярів об'єктів модулем Mask R-CNN результати трекінгу агрегуються у короткі треклети, для яких формуються вбудовування зовнішнього вигляду, а подальша асоціація виконується не між різними фізичними камерами, а між фрагментами спостереження в межах однієї камери та різних сесій/кліпів довготривалого запису. Отже, на відміну від базової багатокамерної постановки TAUDL, у поточній реалізації ключовою задачею є не міжкамерне перенесення ідентичності, а збереження сталої відповідності об'єктів у часі за умов перекриттів, злиття, розділення та фрагментації траєкторій.

Нехай для кожного треклета отримано $L2$ -нормалізоване вбудовування f_i . Як вже зазначено, подібність між двома треклетами доцільно оцінювати косинусною близькістю, яка для нормалізованих векторів дорівнює їх скалярному добутку (4.8). За потреби подібність можна еквівалентно задавати відстанню між векторами (4.9). Далі міжсесійні кандидати відповідності формуються за принципом взаємних найближчих сусідів. Нехай є дві сесії A і B . Для кожного треклета $i \in A$ визначається найкращий партнер у сесії B :

$$j^*(i) = \arg \max_{j \in B} s(i, j). \quad (4.10)$$

Після цього перевіряється взаємність: чи є i також найкращим партнером для знайденого $j^*(i)$ у сесії A :

$$i^*(j) = \arg \max_{i \in A} s(i, j). \quad (4.11)$$

Пара $(i, j^*(i))$ приймається як надійна міжсесійна відповідність, якщо виконується взаємність та порогова умова подібності θ :

$$i^*(j^*(i)) = i, \quad s(i, j^*(i)) \geq \theta. \quad (4.12)$$

Таким чином формується множина міжсесійних відповідностей:

$$\mathcal{M} = \{(i, j) \mid i \in A, \quad j \in B, \quad j = j^*(i), \quad i = i^*(j), \quad s(i, j) \geq \theta\}. \quad (4.13)$$

Окрім порогу θ , на практиці доцільно застосовувати прості фільтри: не використовувати надто короткі треклети; виключати треклети з низькою середньою достовірністю детектувань; відкидати зв'язки, що суперечать часовій структурі, наприклад, якщо це очевидно конкуруючі треклети, які не можуть належати одному ID у межах однієї сесії. Це зменшує кількість хибних асоціацій у режимі без використання ручного маркування.

Після побудови множини $\mathcal{M}$ вводиться асоціативна складова оптимізації, яка зменшує відстань між вбудовуваннями треклетів у парі. У простому вигляді це можна записати як мінімізацію квадрату $L2$ -відстані між векторами:

$$L_{\text{CSTA}} = \sum_{(i,j) \in \mathcal{M}} \|f_i - f_j\|_2^2. \quad (4.14)$$

Еквівалентно, можна максимізувати косинусну близькість (мінімізуючи $1 - s(i, j)$):

$$L_{\text{CSTA}} = \sum_{(i,j) \in \mathcal{M}} (1 - f_i^T f_j). \quad (4.15)$$

У підсумку, для однокамерного випадку CSTA працює як міжсесійний механізм узгодження фрагментів траєкторій: він вводить додаткові зв'язки

між треклетами, які є найбільш правдоподібними відповідностями між сесіями, і зменшує ефект фрагментації треків у довготривалих записах. У TAUDL це поєднується з PCTD: внутрішньосесійна дискримінація формує роздільний простір ознак, а міжсесійна асоціація додає контрольоване наближення знайдених відповідностей [115, 116].

4.2 Функція оптимізації та інтеграція MASK-TAUDL з удосконаленою Mask R-CNN

Модель MASK-TAUDL інтегрується з удосконаленою Mask R-CNN на рівні ROI-ознак, сформованих після вирівнювання області інтересу. Для кожного екземпляра об'єкта використовується тензор ROI-ознак H та відповідна маска екземпляра m , за якою формується масково-зважене представлення. Після просторової агрегації ознак отримується вектор r , який подається на модуль повторної ідентифікації, який формує вектор вбудовування e та його нормалізовану версію f (п. 4.1.3). У загальному вигляді:

$$f = \frac{e}{\|e\|_2}, \quad e = Wr + b. \quad (4.16)$$

Далі f використовується для навчання без ручного маркування за принципом TAUDL: поєднуються дискримінація треклетів у межах сесії (PCTD) та міжсесійна асоціація (адаптований CSTA для однієї камери) [96, 115]. У цій роботі сесії або кліпи однієї камери трактуються як умовні домени, між якими потрібно узгодити ID лабораторних тварин.

Сумарна цільова функція задається як зважена сума двох складових:

$$L = L_{\text{PCTD}} + \lambda L_{\text{CSTA}}. \quad (4.17)$$

- де L_{PSTD} – формує роздільний простір ознак всередині сесії;
 L_{CSTA} – зближує пари треклетів, що вважаються відповідними між різними сесіями;
 λ – параметр, що визначає внесок асоціативної складової [116].

Нехай у межах однієї сесії згенеровано множину треклетів τ . На старті кожному треклету τ_i призначається унікальна псевдомітка y_i (псевдоклас треклета). Для кожного треклета використовується агреговане вбудовування f_i (або множина вбудовувань по кадрах треклета). Дискримінаційна складова PSTD записується як крос-ентропія для класифікації за псевдомітками:

$$L_{\text{PSTD}} = - \sum_{i \in \tau} \log p(y_i | f_i). \quad (4.18)$$

Ймовірність $p(y|f)$ задається softmax-класифікатором:

$$p(k|f) = \frac{\exp(w_k^T f)}{\sum_{l=1}^K \exp(w_l^T f)}, \quad (4.19)$$

- де K – кількість треклетів (псевдокласів) у поточній сесії,
 w_k – ваги класифікаційної голови для k -го псевдокласу [116].

Таким чином, PSTD «стискає» ознаки всередині треклета та розводить різні треклети, формуючи роздільний простір вбудовувань.

У розглянутій постановці CSTA реалізується як міжсесійна асоціація: з двох різних сесій A і B формується множина пар $\mathcal{M}$ взаємних найближчих сусідів. Для цих пар вводиться асоціативна ціль, яка зближує відповідні вбудовування. Узгоджений запис складової CSTA для однієї камери (міжсесійна асоціація) для використання разом із PSTD-класифікацією має вигляд:

$$L_{\text{CSTA}} = - \sum_{(i,j) \in \mathcal{M}} [\log p(y_j | f_i) + \log p(y_i | f_j)], \quad (4.20)$$

де y_i і y_j – псевдомітки треклетів у відповідних сесіях.

Це означає, що якщо треклети з різних сесій є взаємно найближчими, модель повинна підвищити узгодженість їхніх вбудовувань так, щоб вони підтримували один одного в класифікаційній постановці [116]. Альтернативно (еквівалентно за змістом) можна використовувати метрику зближення вбудовувань, наприклад:

$$L_{\text{CSTA}} = \sum_{(i,j) \in \mathcal{M}} \|f_i - f_j\|_2^2. \quad (4.21)$$

Таке представлення зручне коли акцент робиться саме на метричному узгодженні. Навчання з урахуванням інтеграції з Mask R-CNN організовується ітеративно:

- 1) після детектування, сегментації та трекінгу формуються треклети, для яких обчислюються треклетні вбудовування f_i (з масково-зважених ROI-ознак);
- 2) виконується навчання за L_{RSTD} для стабілізації простору ознак усередині сесій;
- 3) періодично оновлюється множина $\mathcal{M}$ (взаємні найближчі сусіди між сесіями) та додається L_{CSTA} , який «склеює» фрагменти однієї особини між сесіями;
- 4) отриманий простір вбудовувань використовується для асоціації треклетів у єдині довготривалі траєкторії з узгодженим ID.

Ключовим для цієї інтеграції є те, що модуль повторної ідентифікації працює з ROI-ознаками та масками екземплярів, отриманими з удосконаленої Mask R-CNN, тобто використовує вже наявні проміжні представлення моделі. Масково-зважене формування r зменшує фонові домішки та

підвищує стабільність вбудовувань у сценах із перекриттями та змінним фоном [96, 126, 136, 137]. Рисунок 4.1 відображає концептуальну модель Mask-TAUDL, яка інтегрує нейромережеву модель Mask R-CNN з моделлю TAUDL для формування масково-зважених вбудовувань та асоціації треклетів без використання ручного маркування.

Сумарна функція втрат для інтегрованої нейромережевої моделі Mask R-CNN та Mask-TAUDL складається з декількох компонент:

1) Втрати Mask R-CNN, пов'язані з детектуванням та сегментацією екземплярів об'єктів: у Mask R-CNN загальна втрата складається з втрат регіонної пропозиції (RPN) та втрат голів другого етапу – класифікації, регресії рамки та побудови маски [14]:

$$L_{\text{MaskRCNN}} = L_{\text{RPNclass}} + L_{\text{RPNbox}} + L_{\text{ROIclass}} + L_{\text{ROIbox}} + L_{\text{mask}}, \quad (4.22)$$

де L_{RPNclass} – класифікація «об'єкт/фон» для опорних рамок;

L_{RPNbox} – регресія рамок у RPN;

L_{ROIclass} – класифікація класу екземпляра на рівні ROI;

L_{ROIbox} – регресія обмежувальної рамки на рівні ROI;

L_{mask} – піксельна втрата для маски екземпляра (зазвичай бінарна крос ентропія по масці для правильного класу) [5].

2) Повна інтегрована втрата: як зазначено, TAUDL складається з PCTD та CSTA (4.17) [116]. Тоді загальна інтегрована функція витрат у випадку внутрішньосесійної дискримінації та міжсесійної (міжкамерної) асоціації має вигляд:

$$L_{\text{total}} = L_{\text{MaskRCNN}} + \beta L_{\text{TAUDL}} = L_{\text{MaskRCNN}} + \beta(L_{\text{PCTD}} + \lambda L_{\text{CSTA}}), \quad (4.23)$$

де β задає баланс між якістю детектування і сегментації, та стабільністю ознак Re-ID;

λ – баланс всередині TAUDL між дискримінацією треклетів і їх міжсесійною асоціацією [116].

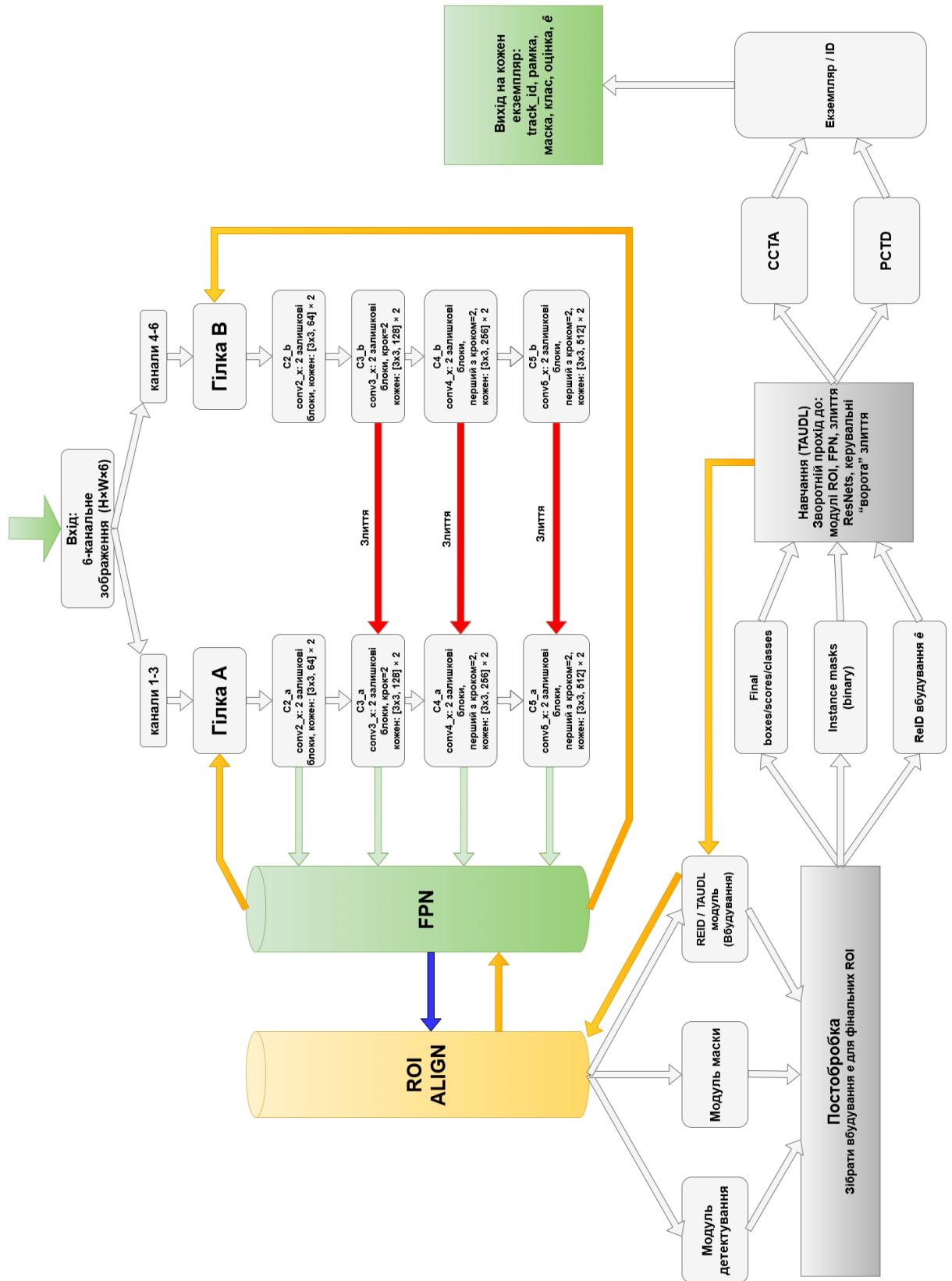


Рисунок 4.1 – Концептуальна модель MASK-TAUDL

Таким чином, L_{MaskRCNN} обчислюється на кадрах із розміченими локалізаціями та масками екземплярів об'єктів [5]. Частина L_{TAUDL} обчислюється на треклетах, сформованих трекінгом: із ROI-ознак та масок екземплярів формуються вбудовування, після чого оптимізуються L_{RSTD} і L_{CSTA} у режимі без використання ручного маркування [96, 115]. Водночас у поточній реалізації для однокамерного випадку основний акцент зроблено на однокамерній треклетно-орієнтованій складовій Re-ID, тоді як міжсесійний компонент CSTA розглядається як адаптований механізм узгодження треклетів між різними фрагментами довготривалого запису та концептуальне розширення базової однокамерної схеми.

4.3 Експериментальне дослідження концептуальної моделі Mask-TAUDL

Експериментальні дослідження виконано для перевірки працездатності концептуальної моделі Mask-TAUDL у монокулярній (однокамерній) постановці та для оцінювання впливу різних екстракторів ознак на якість повторної ідентифікації без ручного маркування. У випробуваннях використовувався обчислювальний стенд із процесором Intel Core i9-13980HX, 64 ГБ оперативної пам'яті та відеокартою NVIDIA GeForce RTX 4090 (16 ГБ). Тестовий датасет включав 41 тестове та 8 валідаційних зображень разом із відповідними треклетами, отриманими за результатами детектування, сегментації екземплярів об'єктів та трекінгу.

Навчання відбувалось за модульною схемою з почерговою оптимізацією: спочатку виконується тренування удосконаленої Mask R-CNN як детектора та сегментатора екземплярів об'єктів, після чого на основі побудованих треклетів виконується навчання модуля повторної ідентифікації без ручного маркування за принципами TAUDL. Такий порядок є доцільним, оскільки якість треклетів, зокрема рівень їх фрагментації, наявність фонових

домішок та помилки при перекриттях, безпосередньо визначає якість псевдонагляду для ідентифікації без використання ручного маркування.

У монокулярній постановці важливим є коректне тлумачення складових TAUDL, які у класичному формулюванні орієнтовані на багатокамерні дані. Дискримінація треклетів на основі камери (PCTD) у цьому випадку застосовується до кожного сеансу або фрагмента запису як до окремого домену: кожен видобутий треклет у межах одного сеансу розглядається як псевдоклас, а модель навчається розрізняти ці псевдокласи, зменшуючи вплив фрагментації та шумних асоціацій. Міжкамерна асоціація треклетів (CSTA) у монокулярних лабораторних даних інтерпретується як міжсесійна асоціація: камери трактуються як різні записи тієї самої арени за змінених умов освітлення, оптичних перешкод або параметрів експерименту; далі знаходяться пари взаємних найближчих сусідів між треклетами різних сеансів і їхні вбудовування узгоджуються у просторі ознак.

Під час інтеграції модуля повторної ідентифікації з детектором-сегментатором використовується сумарна функція втрат, яка поєднує складові Mask R-CNN – втрати класифікації та регресії рамок на рівні RPN і голови детектора, а також втрату маски сегментації – з додатковою втратою асоціації треклетів TAUDL. У частині TAUDL застосовується пакетно-складна триплетна постановка: для якоря (r) визначається найскладніша позитивна та негативна дистанції

$$d_r^+ = \max_{s \in P(r)} d_{rs}, \quad d_r^- = \min_{s \in N(r)} d_{rs}, \quad (4.24)$$

а внесок якоря у втрату задається

$$l_r = \max\{0, m + d_r^+ - d_r^-\}, \quad (4.25)$$

де m – гіперпараметр запасу (margin).

Сумарна втрата TAUDL обчислюється як середнє по множині допустимих якорів.

Для оцінювання якості повторної ідентифікації використано метрики mean Average Precision (mAP) та Rank- k , що відповідає точності за першими k результатами або кумулятивній характеристиці відповідностей. Метрика Rank-1 відображає частку випадків, коли правильна відповідність має найбільшу схожість і посідає перше місце в ранжуванні, а Rank-5 та Rank-10 – частку випадків, коли правильна відповідність потрапляє відповідно до перших 5 або 10 позицій. Метрика mAP узагальнює якість ранжування по всій множині запитів і є більш чутливою до порядкування відповідностей, а не лише до факту потрапляння правильної відповіді до перших k позицій ранжування. У таблиці 4.1 наведено результати порівняння точності TAUDL при використанні різних екстракторів ознак: стандартних ResNet та варіанта з використанням ваг, отриманих під час тренування детектора.

Таблиця 4.1 – Точність MASK-TAUDL з різними екстракторами ознак (mAP, Rank- k)

Екстрактор ознак (backbone)	Ініціалізація	Розмір пакета	К-сть епох	mAP	Rank-1	Rank-5	Rank-10
ResNet-18	випадкова	10	100	0.31	0.50	0.75	0.875
ResNet-34	випадкова	10	100	0.30	0.375	0.625	0.875
ResNet-50	випадкова	8	100	0.37	0.375	0.75	0.875
Подвійний ResNet-18 (об'єднаний), ваги від детектора	ваги з тренування детектора	8	120	0.33	0.625	0.875	1.00

Отримані значення показують, що збільшення ємності екстрактора ознак, зокрема використання ResNet50, може підвищувати mAP, однак не гарантує зростання Rank-1 у малих та складних наборах даних із візуально схожими об'єктами. Натомість використання екстрактора, узгодженого з детектуванням і сегментацією екземплярів об'єктів, дає більш стабільне

ранжування найближчих відповідностей: у наведених випробуваннях досягнуто $\text{Rank-1} = 0.625$ та $\text{Rank-10} = 1.0$ при $\text{mAP} = 0.33$. Практично це означає, що для задач довготривалого відстеження з підтриманням ID критичним є не лише узагальнена якість ранжування за mAP, а й надійність першого вибору за Rank-1, оскільки саме перші відповідності визначають правильність асоціації треклетів і зменшують кількість перемикань ідентичності. На рис. 4.2 показано приклад повторної ідентифікації лабораторних тварин.

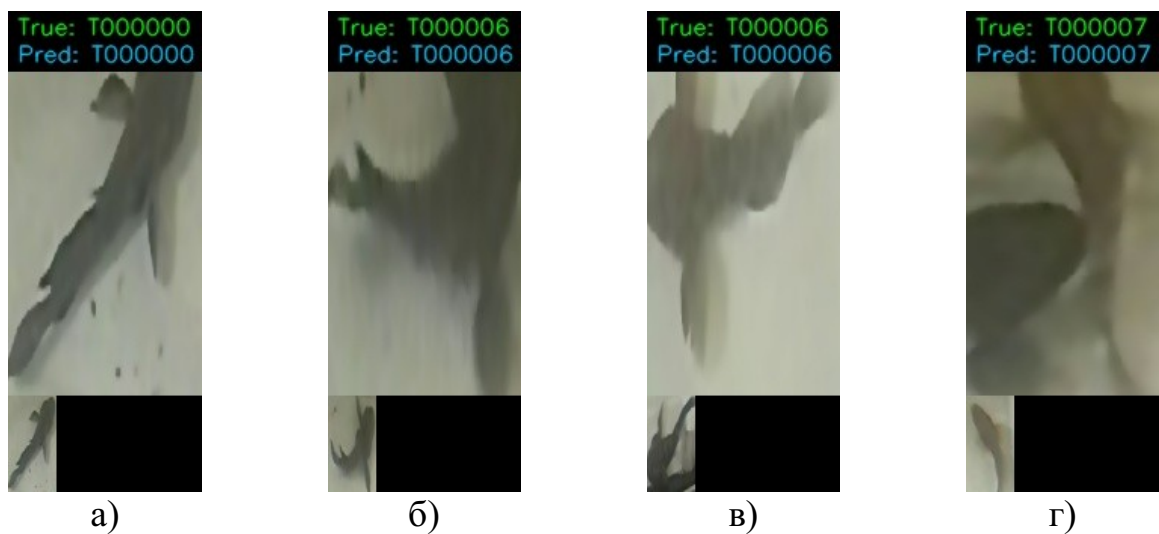


Рисунок 4.2 – Приклади повторної ідентифікації лабораторних тварин: а)-г) – зіставлення істинного та передбаченого ID об'єкта між кадрами

За результатами проведених експериментів встановлено, що поєднання детектування та сегментації екземплярів об'єктів із подальшим навчанням асоціації треклетів у режимі без ручного маркування дозволяє сформувати вбудовування, придатні для підтримання сталої ідентичності в монокулярних лабораторних записах за наявності перекриттів, фонових завад і змін умов зйомки. Ключовими чинниками, що визначають підсумкову якість, є: якість треклетів, сформованих на основі масок екземплярів; коректна постановка PCTD/ССТА для сеансів як доменів у межах однієї камери; узгодженість

екстрактора ознак із задачею детектування та сегментації, що підвищує стійкість ознак до фону та оптичних спотворень.

4.4 Висновки до четвертого розділу

У розділі розроблено та досліджено завершальний компонент наскрізного конвеєра – модуль повторної ідентифікації об'єктів на основі треклетів і навчання без ручного маркування. Показано, що перехід від покадрового зіставлення до треклетного представлення дає змогу накопичувати ознаки зовнішнього вигляду на надійних фрагментах траєкторій і зменшує чутливість до короточасних похибок детектування та перекриттів об'єктів. Для підвищення дискримінативності ознак запропоновано використовувати масково-зважені ROI-ознаки, що пригнічують внесок фону та оптичних артефактів і забезпечують більш стабільне формування вбудовувань для подальшої асоціації треклетів [126, 137].

У запропонованій моделі інтеграція з удосконаленою Mask R-CNN відбувається на рівні ROI-ознак, з яких з урахуванням масок екземплярів формуються нормалізовані вектори вбудовувань для подальшої асоціації треклетів. У випадку однієї камери постановка відповідає треклетно-орієнтованому навчанню без ручного маркування, заснованому на принципах TAUDL. При цьому PCTD використовується для дискримінації треклетів у межах сеансу, а ССТА інтерпретується як адаптований міжсесійний механізм узгодження треклетів між різними фрагментами довготривалого запису [96, 116, 126, 137].

Встановлено, що складові TAUDL у монокулярній постановці можуть бути узгоджені з задачею збереження ідентичності лабораторних тварин у часі: PCTD формує роздільний простір вбудовувань, а міжсесійна асоціація за принципом ССТА забезпечує контрольоване наближення відповідних треклетів у просторі ознак. Сформульована інтегрована функція оптимізації

поєднує складові Mask R-CNN для детектування та сегментації з втратами TAUDL, що забезпечує узгоджене формування ознакового простору для асоціації треклетів [96, 126, 137].

Результати проведених експериментів підтвердили працездатність запропонованої моделі та показали, що вибір екстрактора ознак впливає як на узагальнену якість ранжування за mAP, так і на надійність першої відповідності за Rank-1, яка є критичною для запобігання перемиканням ідентичності. Застосування треклетного представлення разом із масково-зваженими ознаками та асоціацією без ручного маркування дозволяє знизити наслідки фрагментації треків і підвищити стійкість збереження ID у сценах із перекриттями та змінними умовами зйомки. Отримані результати створюють методичну основу для практичного використання інтегрованого конвеєра відеоаналізу в довготривалих лабораторних експериментах.

5 ІНТЕГРОВАНІЙ КОНВЕЄР ДЕТЕКТУВАННЯ, ТРЕКІНГУ ТА НАСКРІЗНОЇ ІДЕНТИФІКАЦІЇ ЛАБОРАТОРНИХ ТВАРИН У ВІДЕОДАНИХ

5.1 Структура інтегрованого конвеєра

Метою цього розділу є розробка та експериментальна перевірка працездатності і ефективності повного інтегрованого конвеєра детектування, трекінгу та наскрізної ідентифікації лабораторних тварин у відеоданих. Конвеєр призначений для перетворення вхідного відеопотоку на набір узгоджених у часі траєкторій лабораторних тварин з узгодженими ідентифікаторами у часі.

Узагальнена логіка конвеєра відображає взаємозв'язок основних етапів обробки відеоданих (рис. 5.1). Для кожного кадру відеопослідовності виконується попередня обробка з метою підвищення стабільності локалізації та зменшення впливу шумів, тіней, локальних неоднорідностей і дрібних артефактів сцени. Після цього формується множина кандидатів об'єктів із двох джерел:

1) удосконаленої нейромережевої моделі Mask R-CNN [79] із подвійною магістраллю екстракції ознак і керованим SE-злиттям ($B \rightarrow A$) [99, 104], яка формує обмежувальну рамку, маску екземпляра та оцінку достовірності детекції;

2) допоміжної сегментаційної гілки, заснованої на відніманні фону та суперпіксельній сегментації, яка формує додаткові кандидати локалізації у випадках слабких, нестабільних або відсутніх нейромережевих детекцій.

Нейромережева модель Mask R-CNN використовує багатоканальне представлення кадру. Його основою є попередньо оброблене кольорове зображення, доповнене каналами, що формуються з проміжних результатів допоміжної сегментаційної гілки, зокрема з ознак яскравості та різницевого

представлення сцени. Така побудова дозволяє поєднати фотометричну інформацію кадру з ознаками відокремлення переднього плану від фону, що підвищує стійкість локалізації в умовах нестационарного фону, часткових перекриттів, відблисків, тіней та локальних артефактів.

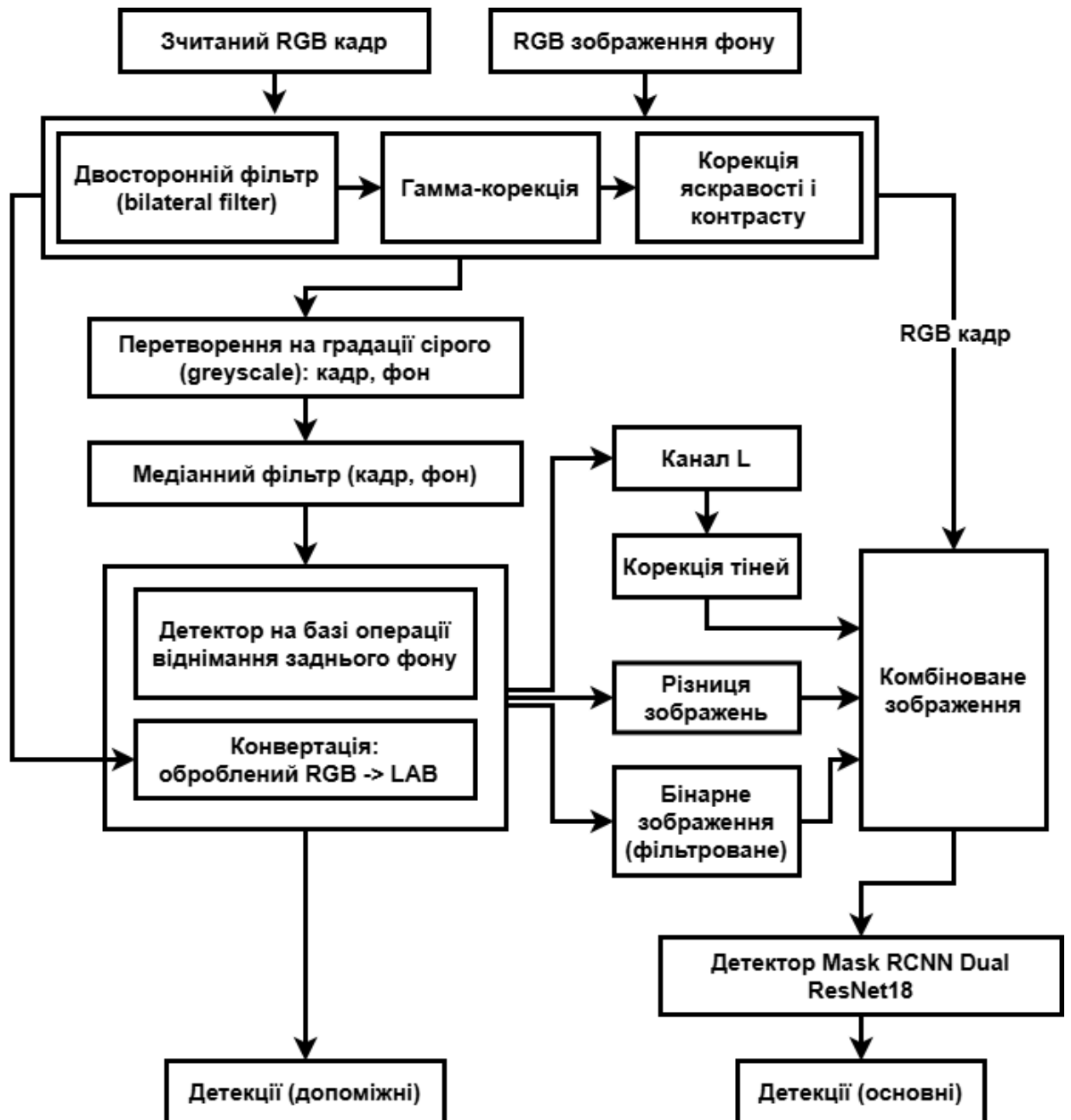


Рисунок 5.1 – Етапи обробки вхідних зображень у межах інтегрованого конвеєра

Після отримання результатів від двох джерел виконується арбітраж, який формує єдину множину екземплярів об'єктів для поточного кадру. У межах арбітражу нейромережеві детекції розглядаються як основне джерело локалізації, тоді як допоміжна сегментаційна гілка використовується для просторово-часової підтримки, підтвердження або резервного формування кандидатів. Якщо нейромережевий детектор формує достатньо надійну локалізацію, саме вона становить основу фінального результату. Якщо ж нейромережеві детекції є слабкими, сумнівними або відсутніми, допоміжні кандидати, отримані з різницевого зображення та суперпіксельної сегментації, можуть використовуватися для уточнення або заміни локалізації. Таким чином, арбітраж виконує функцію узгодження двох джерел детектування та підвищує загальну стійкість конвеєра в умовах складного фону й нестабільних кадрів. Загальну схему інтегрованого конвеєра представлено на рис. 5.2.

Фінальна множина екземплярів об'єктів після арбітражу надходить на вхід модуля трекінгу. Трекінг реалізується як покадрова асоціація поточних екземплярів з активними об'єктами, що вже підтримуються в системі, і ґрунтується насамперед на просторовій та масковій узгодженості між сусідніми кадрами. Для кожного об'єкта підтримуються попередні та поточні параметри локалізації, зокрема обмежувальні рамки, маски та ідентифікатори. Такий підхід дозволяє формувати стійкі часові фрагменти траєкторій, коректно обробляти короткі пропуски та враховувати не лише положення об'єкта, а й його форму.

Для забезпечення збереження ідентичності при перекриттях і розривах видимості у складі конвеєра застосовується модель Mask-TAUDL, яка виконує повторну ідентифікацію без ручного маркування. Вона формує вектори вбудовувань треклетів на основі масково-зважених ROI-ознак та виконує асоціацію фрагментів траєкторій.



Рисунок 5.2 – Загальна схема інтегрованого конвеєра детектування, трекінгу та наскрізної ідентифікації лабораторних тварин

Такий підхід узгоджується з постановкою задачі повторної ідентифікації треклетів, у межах якої відновлення відповідності між фрагментами траєкторій здійснюється з використанням ознак зовнішнього вигляду та просторово-часової узгодженості, що дає змогу підвищити надійність збереження ідентичності об'єктів у складних сценах [130]. Результатом роботи конвеєра є набір траєкторій, де кожен об'єкт описується послідовністю кадр t , обмежувальна рамка, маска та ID.

У межах даного розділу конвеєр тестується повністю. Вхідними даними є відеоматеріали контрольованих сцен лабораторного спостереження. Оскільки для тестових відео відсутня повна покадрова розмітка з істинними ідентифікаторами на кожному кадрі, оцінювання якості роботи конвеєра виконується за допомогою проксі-метрик, що характеризують стабільність локалізації в часі, безперервність траєкторій, збереження ідентичності між сусідніми кадрами та практичну придатність повного інтегрованого рішення.

5.2 Метрики оцінювання якості роботи інтегрованого конвеєра

Вихідні дані конвеєра для кожного кадру t мають вигляд множини екземплярів об'єктів:

$$D(t) = \{(ID_i, bbox_i(t), mask_i(t), score_i(t))\}_i, i = \overline{1, n_t}, \quad (5.1)$$

де $mask_i(t)$ є бінарною маскою екземпляра (формат Mask R-CNN [86]) у координатах кадру;

$bbox_i(t)$ – обмежувальна рамка детекції;

$score_i(t)$ – оцінка детекції відповідно;

ID_i – ідентифікатор траєкторії об'єкта.

Для оцінювання якості та стабільності трекінгу за відсутності повної еталонної розмітки у роботі використовуються чотири проксі-показники, які

відповідають ключовим вимогам до інтегрованого конвеєра: збереженню ідентичності, часовій узгодженості локалізації, безперервності траєкторій та достатній тривалості спостереження. Вибір таких показників узгоджується з класичними підходами до оцінювання багатооб'єктного трекінгу, у яких враховуються помилки ідентичності, фрагментація траєкторій та якість асоціації об'єктів у часі [138], а також із практикою оцінювання алгоритмів трекінгу в межах показників типу MOTChallenge [139]:

1) PSW_{rate} – проксі-метрика перемикань ID між кадрами. У класичних постановках задач трекінгу багатьох об'єктів (англ. Multi-Object Tracking) помилки ідентичності аналізуються як перемикання ідентифікатора траєкторії. За відсутності еталонних даних використовується проксі-оцінка: якщо об'єкт на кадрі t має очевидне просторове продовження на кадрі $t + 1$ за перекриттям масок, але отримує інший ID, така ситуація трактується як псевдоперемикання ID (англ. pseudo switch). Така логіка узгоджується з практикою асоціації траєкторій у трекінгу з використанням глибоких метрик [130].

Нехай між множинами екземплярів $D(t)$ і $D(t + 1)$ виконується зіставлення за максимальною подібністю масок, наприклад допомогою Угорського алгоритму або жадібного алгоритму, за метрикою:

$$IoU(m_a, m_b) = \frac{|m_a \cap m_b|}{|m_a \cup m_b|}. \quad (5.2)$$

До набору відповідностей $M_{t \rightarrow t+1}$ включаються лише пари з $IoU \geq \tau$, де τ – поріг (у експериментах $\tau = \langle \tau \rangle$). Тоді

$$PSW(t) = \sum_{(a,b) \in M_{t \rightarrow t+1}} [ID_a(t) \neq ID_b(t + 1)], \quad (5.3)$$

а нормований показник:

$$PSW_{rate} = \frac{\sum_t PSW(t)}{\sum_t |M_{t \rightarrow t+1}|}. \quad (5.4)$$

Менше значення PSW_{rate} відповідає кращому збереженню ідентичності об'єктів у часі.

2) $tIoU$ – часова узгодженість локалізації маски (англ. temporal consistency). Друга метрика оцінює стабільність локалізації в часі для одного й того ж ID. Для кожної траєкторії розраховується IoU між масками на сусідніх кадрах:

$$tIoU(ID, t) = IoU(mask_{ID}(t), mask_{ID}(t + 1)). \quad (5.5)$$

Звітуються агреговані статистики:

- $\widetilde{tIoU}$ – медіана по всіх валідних кроках;
- $P_{unstable}$ – частка кроків, де $tIoU < \theta$ (поріг нестабільності, $\theta = \langle \theta \rangle$).

Використання масок екземплярів як основи для оцінювання часової узгодженості є природним у задачі сегментації екземплярів, де Mask R-CNN повертає маски об'єктів [79], а стабільність локалізації є важливою умовою коректного трекінгу.

3) $FRAG^*$ – проксі-метрика розривів (фрагментації) траєкторій. Для оцінювання стійкості відстеження використовується проксі-метрика фрагментації, що рахується як кількість пропусків у присутності треку. Для кожного ID формується відсортований список кадрів, на яких присутній цей ID:

$$F(ID) = \{t_1, t_2, \dots, t_k\}. \quad (5.6)$$

Тоді кількість розривів визначається як:

$$gaps(ID) = \sum_{j=1}^{k-1} \{t_{j+1} - t_j > 1\}, \quad (5.7)$$

а загальна характеристика:

- $\overline{gaps}$ – середня кількість розривів на трек,
- P_{nogap} – частка треків без розривів.

Інтерпретаційно $FRAG^*$ характеризує здатність конвеєра зберігати траєкторії без частих переривань, що є критичним для подальшого формування треклетів та їх асоціації за ознаками [131].

4) $S(T)$ та медіана довжини треку – вимоги до тривалості спостереження. Для відповіді на питання про те, яка кількість кадрів або яка тривалість відеофрагмента є достатньою для отримання стійких траєкторій, оцінюється розподіл довжин треків. Для кожного ID визначається:

$$L(ID) = |F(ID)| \quad (\text{довжина в кадрах}), \quad (5.8)$$

$$T(ID) = \frac{L(ID)}{FPS_{in}} \quad (\text{довжина в секундах}). \quad (5.9)$$

Звітуються:

- $\tilde{L}$ – медіана довжини треку,
- Частка треків довших за задані пороги $T_0 \in \{2, 5, 10, 20\}$ секунд (або відповідні значення у кадрах):

$$S(T_0) = P(T(ID) \geq T_0). \quad (5.10)$$

Ці показники дозволяють пов'язати якість відстеження з реальними сценаріями поведінкового моніторингу: якщо значна частка треків коротша за необхідний поріг, то отримання стабільних показників поведінки або руху може потребувати більшої тривалості спостереження чи корекції налаштувань трекінгу та повторної ідентифікації.

5.3 Умови проведення тестування та використані набори даних

Для проведення експериментів було використано обладнання, доступне на загальному ринку, без залучення спеціалізованих серверних обчислювальних систем. Основні обчислення виконувалися на ноутбучі з процесором Intel Core i9-13980HX, 64 ГБ оперативної пам'яті та відеокартою NVIDIA GeForce RTX 4090 Laptop GPU з 16 ГБ відеопам'яті. Така конфігурація дозволяє оцінити працездатність запропонованого конвеєра в умовах локальної обробки відеоданих, що є важливим для практичного використання в лабораторних дослідженнях.

Як тестові дані використано три відеопослідовності. Перше відео тривалістю 2 хв сформовано на основі набору синтетичних зображень рухомих об'єктів, що імітують рух лабораторних тварин в умовному контрольованому середовищі. Дві інші відеопослідовності є записами поведінки реальних лабораторних тварин – мишей та бичків. Відеозаписи були надані Біологічним факультетом Одеського національного університету імені І. І. Мечникова та отримані в межах відповідних лабораторних досліджень.

Для синтетичного відео набір даних для навчання удосконаленої Mask R-CNN із подвійною магістраллю ResNet18 містив 90 тренувальних та 10 тестових зображень. Набір даних для навчання модуля повторної ідентифікації Mask-TAUDL містив 13 послідовностей із 152 зображень сумарно.

Для тестування на відео з бичками було сформовано датасет із 340 зображень. Його основою стали 85 зображень, отриманих безпосередньо як кадри з декількох відеозаписів бичків. Для розширення навчальної вибірки застосовано аугментацію шляхом обертання наявних зображень на 90°, 180° та 270°, що дозволило збільшити набір до 340 зображень. Набір тестових зображень містив 12 зображень. Набір даних для навчання модуля Mask-TAUDL містив 19 послідовностей із 30 зображень.

Для відео з мишами було сформовано окремий набір даних. Набір даних для навчання детектора містив 280 зображень, а набір для валідації – 48 зображень. Відповідний набір даних для навчання модуля Mask -TAUDL містив 15 послідовностей із 30 зображень.

Таким чином, використані тестові дані охоплюють як синтетичну контрольовану послідовність, так і реальні лабораторні відеозаписи з різними типами об'єктів і умовами сцени. Це дає змогу перевірити працездатність інтегрованого конвеєра на рівні детектування та сегментації екземплярів об'єктів, формування треклетів і подальшої наскрізної ідентифікації без ручного маркування на основі моделі Mask-TAUDL.

5.4 Результати застосування інтегрованого конвеєра для аналізу лабораторних відеоданих

Експерименти виконуються шляхом запуску повного інтегрованого конвеєра на тестових відеопослідовностях контрольованих сцен. Для кожного кадру обчислюються результати детектування та сегментації екземплярів об'єктів – обмежувальна рамка, маска екземпляра та оцінка достовірності. Далі виконується арбітраж детекцій, після чого здійснюється трекінг і формування треклетів. На наступному етапі застосовується модель Mask-TAUDL для обчислення векторів вбудовувань і асоціації треклетів, що забезпечує збереження глобальних ID за умов короткочасних переривань видимості та перекриттів.

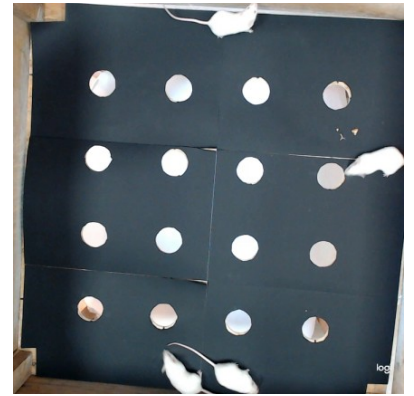
На наступних рисунках наведено приклади роботи проміжних етапів обробки зображень у межах інтегрованого конвеєра. Рисунок 5.3 демонструє підготовку зображень для допоміжної гілки детектування на основі віднімання фону. На цьому етапі виконується корекція кольору, фільтрація шуму та перетворення зображення у відповідне представлення для подальшого виділення об'єктів.



а)



б)



в)



г)



д)



е)

Рисунок 5.3 – Підготовка зображень для допоміжної гілки детектування на основі віднімання фону: а, б – зображення порожнього фону контрольованого середовища; в, г – кадри з лабораторними тваринами; д, е – попередньо оброблені кадри з лабораторними тваринами

На рисунку 5.4 показано результати роботи детектора на основі операції віднімання фону та подальшого очищення різницевого зображення з використанням результатів сегментації модифікованим методом SLIC, включно з етапом локалізації об'єктів після обробки різницевого зображення, на якому для виділених областей формуються обмежувальні рамки та умовні центри детектованих об'єктів.

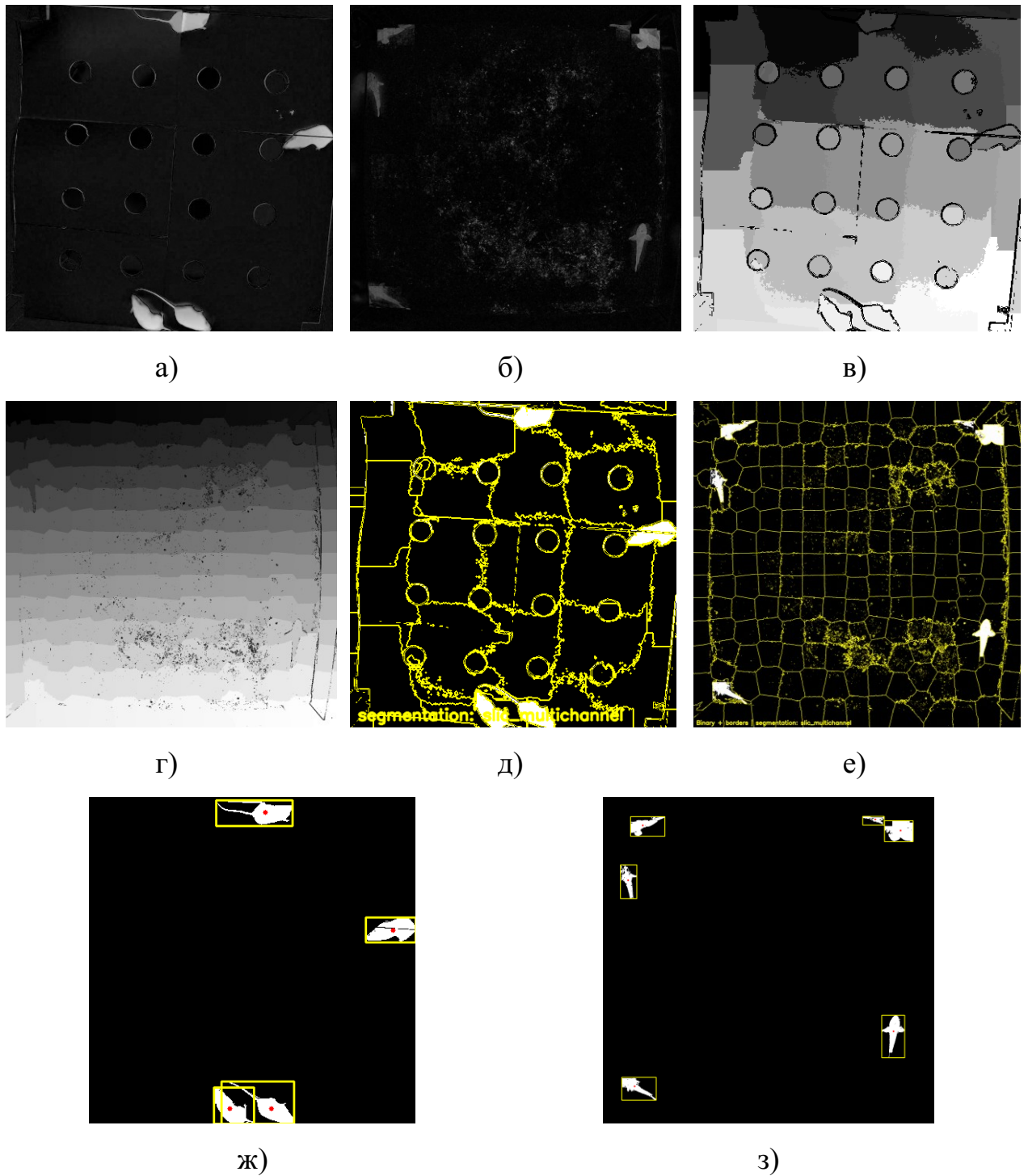
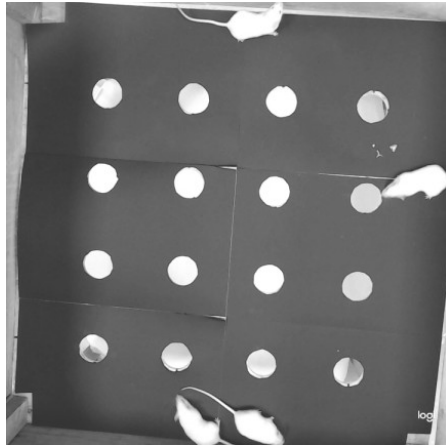
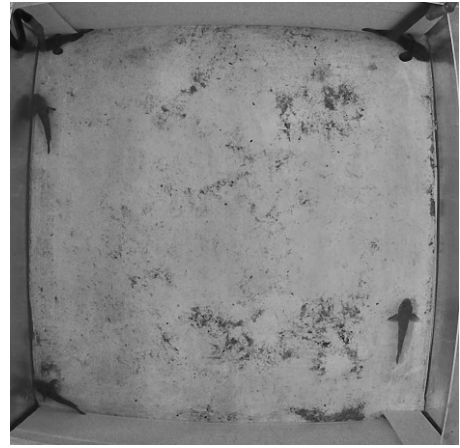


Рисунок 5.4 – Результати роботи допоміжної гілки детектування на основі віднімання фону та удосконаленого методу SLIC: а, б – різницеві зображення для кадрів з мишами та бичками відповідно; в, г – суперпідкселі, отримані удосконаленим методом SLIC; д, е – межі суперпідкселів, накладені на різницеві зображення; ж, з – обмежувальні рамки та умовні центри об'єктів для кадрів з мишами та бичками відповідно

На рис. 5.5 показано шар L після корекції тіней, який є складовою 6-канального зображення, що подається на вхід детектора на основі удосконаленої Mask R-CNN із подвійною магістраллю ResNet18.



а)



б)



в)



г)

Рисунок 5.5 – Робота детектора на основі удосконаленої Mask R-CNN із подвійною магістраллю ResNet18: а, б – шар L для кадрів з мишами та бичками відповідно; в, г – результати детектування для кадрів з мишами та бичками відповідно

Рисунок 5.6 демонструє результати арбітражу детекцій та збереження ідентифікаторів об'єктів між кадрами протягом певного проміжку часу.

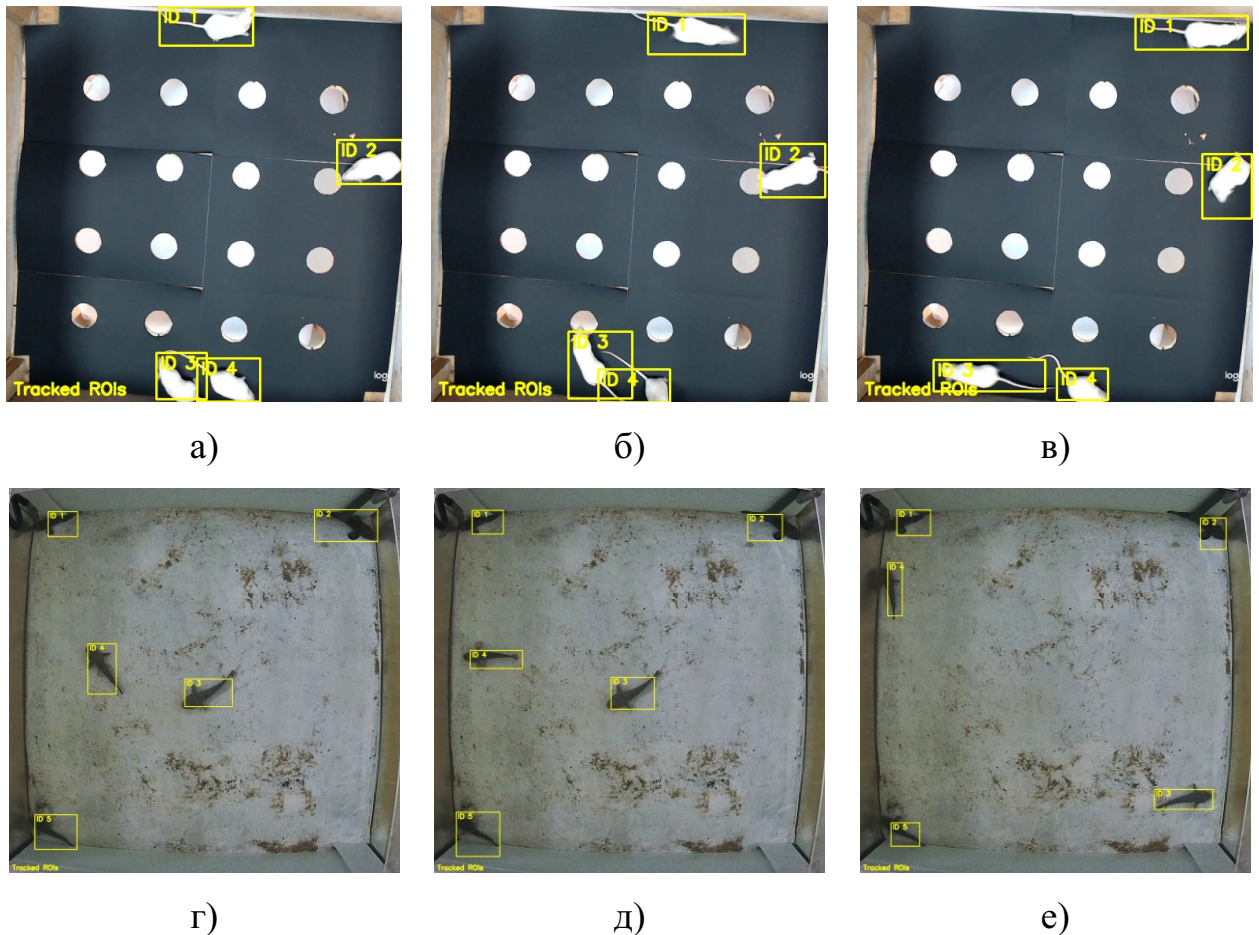


Рисунок 5.6 – Збереження ідентифікаторів об’єктів між кадрами протягом певного проміжку часу: а, б, в – трекінг мишей; г, д, е – трекінг бичків

Налаштування конвеєра включають: поріг прийняття детекцій $score_thr$, параметри асоціації трекінгу max_age , поріг IoU для матчингу між кадрами в метриках $\tau = \langle \tau \rangle$, поріг нестабільності $\theta = \langle \theta \rangle$, а також набір порогів T_0 для функції виживання $S(T_0)$. Підсумкові значення метрик (агреговано по всіх тестових відеопослідовностей) наведено в табл. 5.1.

Таблиця 5.1 – Показники стабільності трекінгу та збереження ідентичності в інтегрованому конвеєрі

Метрика	Відео 1	Відео 2	Відео 3
1	2	3	4
Кількість об’єктів (треків)	5	5	4
Кількість кадрів	900	3200	4320

Продовження таблиці 5.1

1	2	3	4
Проксі-перемикання ID PSW_{rate}	0.002	0	0.01
Медіана часової узгодженості масок $\text{median}(tIoU)$	0.83	0.81	0.85
Частка нестабільних кроків $P(tIoU < \theta)$	0.01	0.07	0.05
Середня кількість розривів на трек $\text{avg}(gaps)$	8	0	61
Медіана довжини треку $\text{median}(T)$ (сек.)	28.09	0.04	166.89

Інтерпретація метрик PSW_{rate} та $FRAG^*$ пов'язана з оцінюванням надійності асоціації та збереження ідентичності об'єктів у часі, що є ключовим для задач багатооб'єктного трекінгу. Застосування ознак зовнішнього вигляду та механізмів асоціації треклетів є типовим способом зменшення помилок ідентичності у сучасних трекерах [130], а використання треклетного підходу узгоджується з ідеєю повторної ідентифікації фрагментів траєкторій [131]. Метрика $tIoU$ відображає стабільність локалізації масок екземплярів, що є природним критерієм узгодженості при сегментації екземплярів на основі Mask R-CNN [79].

Для оцінювання впливу тривалості спостереження на стабільність метрик додатково аналізується залежність показників від довжини вікна спостереження. Відео розбивається на фрагменти тривалістю $T_0 \in \{2, 5, 10, 20\}$ секунд, після чого метрики розраховуються для кожної тривалості окремо. Така оцінка є практично важливою, оскільки у реальних сценаріях поведінкового моніторингу стабільні висновки про траєкторії та взаємодії об'єктів потребують достатньо довгих безперервних відрізків спостереження. Відповідні заміри наведено у табл. 5.2.

Очікувано, найбільш складними ситуаціями для конвеєра є часткові або повні перекриття об'єктів, різкі зміни форми маски через нечіткий фон чи погіршення якості кадру, а також короточасні пропуски детектування, що призводять до розривів траєкторій. У таких випадках зростають PSW_{rate} та $\text{avg}(gaps)$, а також збільшується частка нестабільних кроків $P(tIoU < \theta)$. З

іншого боку, використання масок екземплярів у трекінгу дозволяє точніше враховувати форму об'єкта, ніж лише рамки [79], а застосування асоціації з урахуванням ознак зовнішнього вигляду та механізмів асоціації треклетів є стандартним підходом для зменшення помилок ідентичності [130], [131].

Таблиця 5.2 – Залежність показників роботи конвеєра від тривалості вікна спостереження

Відео	Тривалість вікна	PSW_{rate}	$avg(gaps)$	$S(T)$	$median(T)$	$median(tIoU)$
1	2 с	0.003	0.45	0.62	1.95	0.83
	5 с	0.002	1.17	0.37	4.73	0.83
	10 с	0.002	2.53	0.13	9.46	0.83
	20 с	0.003	3.9	0.1	0.83	0.83
2	2 с	0	0	0.12	0.07	0.81
	5 с	0	0	0.04	0.06	0.81
	10 с	0	0	0.01	0.05	0.81
	20 с	0	0	0.002	0.06	0.81
3	2 с	0.01	0.65	0.73	1.95	0.84
	5 с	0.01	1.71	0.52	4.83	0.84
	10 с	0.01	3.33	0.33	9.37	0.84
	20 с	0.01	6.72	0.17	18.69	0.84

Отже, у розділі виконано експериментальну перевірку працездатності повного інтегрованого конвеєра детектування-трекінгу-реідентифікації об'єктів у контрольованих сценах. Конвеєр реалізує послідовність обробки «попередня обробка → сегментація екземплярів на основі модифікованої Mask R-CNN → арбітраж результатів → трекінг і формування треклетів → re-ID без ручного маркування (Mask-TAUDL) і об'єднання треклетів у глобальні траєкторії», що забезпечує отримання узгоджених у часі траєкторій із стабільними ідентифікаторами.

Якість роботи конвеєра оцінювалась за проксі-метриками трекінгу, орієнтованими на ключові практичні вимоги: збереження ідентичності (PSW_{rate}), часова узгодженість локалізації $median(tIoU)$ та частка нестабільних кроків), безперервність траєкторій ($FRAG^*$), а також вимоги до тривалості спостереження ($S(T)$, медіана довжини треку). Отримані експериментальні значення (табл. 5.1-5.2) дозволяють сформулювати рекомендації щодо мінімальної тривалості спостереження T^* (у секундах і кадрах), починаючи з якої показники стабілізуються та стають репрезентативними для реальних сценаріїв моніторингу, а також підтверджують практичну придатність запропонованого конвеєра за критеріями стабільності трекінгу та оперативності обробки відеоданих.

Таким чином, результати експериментальної перевірки демонструють працездатність інтегрованого конвеєра, який поєднує попередню обробку, детектування та сегментацію екземплярів об'єктів, арбітраж результатів, трекінг, формування треклетів і подальшу наскрізну ідентифікацію без використання ручного маркування на основі Mask-TAUDL. Оцінювання за проксі-метриками дозволяє характеризувати стабільність трекінгу, збереження ідентичності, безперервність траєкторій та достатню тривалість спостереження для практичних сценаріїв поведінкового моніторингу.

5.5 Висновки до п'ятого розділу

У розділі виконано експериментальну перевірку працездатності створеного конвеєра детектування, трекінгу та наскрізної ідентифікації лабораторних тварин у відеоданих на основі розроблених в роботі моделей на методів. Конвеєр поєднує попередню обробку кадрів, допоміжну сегментаційну гілку на основі віднімання фону та удосконаленого методу SLIC, удосконалену Mask R-CNN із подвійною магістраллю ResNet18, арбітраж результатів детектування, трекінг, формування треклетів і

подальшу асоціацію треклетів без використання ручного маркування на основі концептуальної моделі Mask-TAUDL.

Показано, що використання двох джерел локалізації, а саме, нейромережевого детектора-сегментатора та допоміжної сегментаційної гілки, дозволяє підвищити стійкість конвеєра в умовах нестаціонарного фону, відблисків, тіней, локальних артефактів і нестабільних детекцій. Арбітраж результатів забезпечує формування єдиної множини екземплярів об'єктів для подальшого трекінгу, а використання масок екземплярів дає змогу враховувати не лише положення, а й форму лабораторної тварини у кадрі.

Для оцінювання роботи інтегрованого конвеєра сформовано систему проксі-метрик, орієнтованих на ключові практичні вимоги до відеомоніторингу: збереження ідентичності, часову узгодженість локалізації, безперервність траєкторій і достатню тривалість спостереження. Такі метрики дозволяють оцінювати якість роботи конвеєра навіть за відсутності повної покадрової еталонної розмітки з істинними ID об'єктів.

Найбільш складними для конвеєра є ситуації часткових або повних перекриттів, короткочасних пропусків детектування, різких змін форми маски та погіршення якості кадру. У таких випадках зростає ризик розривів траєкторій і перемикань ідентичності. Застосування треклетного подання, масково-зважених ознак і асоціації треклетів на основі Mask-TAUDL надало основу для зменшення наслідків фрагментації траєкторій і підтримання сталих ID у довготривалих лабораторних відеозаписах.

Отримані результати підтверджують практичну придатність запропонованого інтегрованого конвеєра для обробки відеоданих лабораторного спостереження в умовах локальних обчислювальних ресурсів. Розроблений конвеєр може бути використаний як основа для подальшого автоматизованого аналізу поведінки лабораторних тварин, зокрема для отримання траєкторій, оцінювання активності, аналізу контактів і формування кількісних поведінкових показників.

На основі створеного інтегрованого конвеєра розроблено спеціалізований застосунок для автоматизованої обробки відеозаписів з лабораторними тваринами. Він реалізує основні етапи запропонованого конвеєра; попередню обробку кадрів, детектування, сегментацію, трекінг, формування траєкторій і збереження ідентичності об'єктів. Розроблене програмне рішення використовується у лабораторних дослідженнях, що проводяться на базі біологічного факультету Одеського національного університету імені І. І. Мечникова.

ВИСНОВКИ

У дисертаційній роботі розв'язано науково-практичну задачу підвищення точності й стійкості детектування, трекінгу та наскрізної ідентифікації лабораторних тварин у відеоданих за умов нестаціонарного фону, оклюзій, злиття та розділення об'єктів, обмежених ресурсів і локальної обробки тривалих відеоспостережень. Запропоновані моделі, методи та інтегрований конвеєр автоматизованої обробки відеоданих забезпечили підвищення якості локалізації об'єктів, формування узгоджених у часі траєкторій і підтримання наскрізної ідентифікації лабораторних тварин у складних умовах відеоспостереження.

Проведено аналіз методів та підходів до детектування, трекінгу та збереження ідентичності лабораторних тварин у відеоданих. Показано, що класичні методи комп'ютерного зору, зокрема порогування, віднімання фону, зіставлення із шаблонами, оптичний потік, MeanShift, CamShift і трекінг ключових точок, є відносно простими за та характеризуються низькими вимогами до обчислювальних ресурсів, однак недостатньо стійкими до змін освітлення, нестаціонарного фону, артефактів, часткових перекриттів і злиття об'єктів. Встановлено, що нейромережеві методи детектування й сегментації забезпечують вищу якість локалізації, але потребують обґрунтованого вибору архітектури з урахуванням локального навчання, обмежених обчислювальних ресурсів і прийнятної оперативності обробки довготривалих відеозаписів.

Обґрунтовано вимоги до конвеєра автоматизованої обробки відеоданих лабораторних тварин. Визначено, що для практичного використання у задачах поведінкового моніторингу конвеєр має забезпечувати не лише локалізацію об'єктів у кадрі, а й формування узгоджених у часі траєкторій, збереження ідентичності лабораторних тварин після оклюзій, злиття та розділення об'єктів, а також можливість роботи на локальному обладнанні без використання ресурсомістких хмарних обчислень. Це дало змогу

сформувати загальну логіку подальшого конвеєра: попередня обробка кадрів, детектування, сегментація екземплярів, трекінг, формування треклетів і повторна ідентифікація.

Удосконалено метод суперпиксельної сегментації SLIC для задачі локалізації лабораторних тварин у відеоданих. Запропоноване удосконалення полягає у використанні багат шарового простору ознак, що враховує LAB-представлення, різницеve зображення та перевірки однорідності пікселів у локальному оточенні. Це дозволило підвищити дозволу підвищити якість локалізації лабораторних тварин у відеоданих за умов нестационарного фону, наявності артефактів, відблисків, тіней і часткових перекриттів.

Удосконалено метод детектування лабораторних тварин у відеоданих шляхом використання методу суперпиксельної сегментації SLIC і контекстної фільтрації в межах сегментів. Запропонований метод забезпечує перехід від грубого глобального порогоування різницевого зображення до аналізу локально узгоджених областей на рівні суперпикселів. Це забезпечує відсікання фонових артефактів і зменшує ризик втрати тонких частин контуру об'єкта. За результатами експерименту встановлено, що запропонований метод забезпечує підвищення якості детектування та локалізації лабораторних тварин більш ніж на 6 % порівняно з базовим методом SLIC.

Удосконалено нейромережеву модель Mask R-CNN для детектування та сегментації лабораторних тварин у відеоданих за умов складного фону, часткових перекриттів, злиття та подальшого розділення об'єктів. Запропоноване удосконалення базується на введенні подвійної магістралі екстракції ознак із використанням ResNet18: одна гілка обробляє базові канали зображення, а інша – додаткові канали ознак, сформовані шляхом використання методу суперпиксельної сегментації SLIC і контекстної фільтрації в межах сегментів. Для інтеграції ознак використано кероване SE-злиття, що дозволяє підсилювати інформативні канали та приглушувати шумові або менш корисні компоненти. Отримані результати дозволили

підвищити якість детектування та сегментації при збереженні прийнятної оперативності обробки: для запропонованої моделі отримано $mAP50 = 0.87$ для обмежувальних рамок і $mAP50 = 0.90$ для масок сегментації за середнього часу виведення близько 105.5 мс.

Розроблено концептуальну модель наскрізної ідентифікації лабораторних тварин у відеоданих без використання ручного маркування Mask-TAUDL. Запропонована модель поєднує результати детектування та сегментації екземплярів із формуванням сегментованих треклетів, масково-зваженим вилученням ROI-ознак, побудовою нормалізованих векторів вбудовувань і асоціацією треклетів без ручних ID-міток. Використання масок екземплярів дозволяє зменшити вплив фону під час формування ознак, а треклетне подання підвищує стійкість до короточасних оклюзій і фрагментації траєкторій.

На основі розробленої концептуальної моделі Mask-TAUDL реалізовано й експериментально перевірено модуль наскрізної ідентифікації лабораторних тварин у відеоданих без використання ручного маркування. Показано, що використання подвійного екстрактора ознак ResNet18 з ініціалізацією вагами, отриманими під час навчання сегментації, забезпечує збалансовану якість повторної ідентифікації за компактності моделі. Отримані значення $mAP = 0.33$, $Rank-1 = 0.625$, $Rank-5 = 0.875$ і $Rank-10 = 1$ підтверджують працездатність запропонованого підходу для асоціації сегментованих треклетів, відновлення відповідності між фрагментами траєкторій і забезпечення наскрізної ідентифікації лабораторних тварин у довготривалих відеопослідовностях.

Реалізовано інтегрований конвеєр детектування, трекінгу та наскрізної ідентифікації лабораторних тварин у відеоданих. Конвеєр об'єднує попередню обробку кадрів, допоміжну сегментаційну гілку на основі віднімання фону та удосконаленого SLIC, нейромережеве детектування й сегментацію екземплярів на основі удосконаленої Mask R-CNN, арбітраж результатів детектування, трекінг, формування треклетів і подальшу

асоціацію треклетів за допомогою концептуальної моделі Mask-TAUDL. Така інтеграція дозволяє перейти від локальної обробки окремих кадрів до формування узгоджених у часі траєкторій із підтриманням сталих ідентифікаторів лабораторних тварин.

Для оцінювання роботи інтегрованого конвеєра сформовано систему проксі-метрик, орієнтованих на практичні вимоги лабораторного відеомоніторингу: збереження ідентичності, часову узгодженість локалізації, безперервність траєкторій і достатню тривалість спостереження. Такий підхід дозволяє оцінювати якість роботи конвеєра навіть за відсутності повної покадрової еталонної розмітки з істинними ідентифікаторами об'єктів. Результати тестування підтвердили працездатність конвеєра та його придатність для аналізу довготривалих лабораторних відеоданих у реальних умовах спостереження.

Практичне значення одержаних результатів полягає у створенні прикладних рішень для автоматизованої обробки відеоданих спостережень за лабораторними тваринами, які можуть бути використані для формування траєкторій, оцінювання активності, аналізу контактів і побудови кількісних поведінкових показників. Дисертаційну роботу виконано згідно з тематичними планами науково-дослідних робіт Національного університету «Одеська політехніка» за період 2025–2026 рр. у межах держбюджетної теми № 264-73 «Методи та технології моделювання і забезпечення якості складних програмних систем у динамічних середовищах». На основі запропонованих моделей і методів розроблено спеціалізований застосунок для автоматизованої обробки відеозаписів, який використовується у лабораторних дослідженнях на базі біологічного факультету Одеського національного університету імені І. І. Мечникова. Результати дисертаційної роботи також впроваджено у навчальний процес Національного університету «Одеська політехніка» та Одеського національного університету імені І. І. Мечникова.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Smith A.R. Color Gamut Transform Pairs // SIGGRAPH '78: Proceedings of the 5th annual conference on Computer graphics and interactive techniques. 1978. P. 12–19. DOI: 10.1145/800248.807361.
2. Spink A.J., Tegelenbosch R.A.J., Buma M.O.S., Noldus L.P.J.J. The EthoVision video tracking system: a tool for behavioral phenotyping of transgenic mice // Physiology & Behavior. 2001. Vol. 73. No. 5. P. 731–744. DOI: 10.1016/S0031-9384(01)00530-3.
3. Lopes G. et al. Bonsai: an event-based framework for processing and controlling data streams // Frontiers in Neuroinformatics. 2015. Vol. 9. Art. 7. DOI: 10.3389/fninf.2015.00007.
4. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks: препринт arXiv:1506.01497. 2015. Режим доступу: <https://arxiv.org/abs/1506.01497>. – Дата звернення: 8 травня 2026 р.
5. He K., Gkioxari G., Dollár P., Girshick R. Mask R-CNN: препринт arXiv:1703.06870. 2017. Режим доступу: <https://arxiv.org/abs/1703.06870>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1703.06870.
6. Liu W. et al. SSD: Single Shot MultiBox Detector: препринт arXiv:1512.02325. 2015. Режим доступу: <https://arxiv.org/abs/1512.02325>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1512.02325.
7. Tian Z., Shen C., Chen H., He T. FCOS: Fully Convolutional One-Stage Object Detection: препринт arXiv:1904.01355. 2019. Режим доступу: <https://arxiv.org/abs/1904.01355>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1904.01355.
8. Mathis A. et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning // Nature Neuroscience. 2018. Vol. 21. No. 9. P. 1281–1289. DOI: 10.1038/s41593-018-0209-y.

9. Lauer J. et al. Multi-animal pose estimation, identification and tracking with DeepLabCut // *Nature Methods*. 2022. Vol. 19. No. 4. P. 496–504. DOI: 10.1038/s41592-022-01443-0.
10. Pereira T.D. et al. Fast animal pose estimation using deep neural networks // *Nature Methods*. 2019. Vol. 16. No. 1. P. 117–125. DOI: 10.1038/s41592-018-0234-5.
11. Romero-Ferrero F. et al. idtracker.ai: tracking all individuals in small or large collectives of unmarked animals // *Nature Methods*. 2019. Vol. 16. No. 2. P. 179–182. DOI: 10.1038/s41592-018-0295-5.
12. Walter T., Couzin I.D. TRex: a fast multi-animal tracking system with markerless identification, and 2D estimation of posture and visual fields // *eLife*. 2021. Vol. 10. e64000. DOI: 10.7554/eLife.64000.
13. Chen S. et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs: препринт arXiv:1606.00915. 2016. Режим доступа: <https://arxiv.org/abs/1606.00915>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1606.00915.
14. Cheng Z., Wu Y., Xu Z., Lukasiewicz T., Wang W. Segmentation is All You Need: препринт arXiv:1904.13300. 2019. Режим доступа: <https://arxiv.org/abs/1904.13300>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1904.13300.
15. Pereira T.D. et al. SLEAP: A deep learning system for multi-animal pose tracking // *Nature Methods*. 2022. Vol. 19. No. 4. P. 486–495. DOI: 10.1038/s41592-022-01426-1.
16. Kabra M., Robie A.A., Rivera-Alba M., Branson S., Branson K. JAABA: interactive machine learning for automatic annotation of animal behavior // *Nature Methods*. 2013. Vol. 10. No. 1. P. 64–67. DOI: 10.1038/nmeth.2281.
17. Nilsson S.R.O. et al. Simple Behavioral Analysis (SimBA): an open source toolkit for computer classification of complex social behaviors in experimental animals: препринт bioRxiv. 2020. DOI: 10.1101/2020.04.19.049452.

18. Segalin C. et al. The Mouse Action Recognition System (MARS) software pipeline for automated analysis of social behaviors in mice // eLife. 2021. Vol. 10. e63720. DOI: 10.7554/eLife.63720.
19. Mirat O., Sternberg J.R., Severi K.E., Wyart C. ZebraZoom: an automated program for high-throughput behavioral analysis and categorization // Frontiers in Neural Circuits. 2013. Vol. 7. Art. 107. DOI: 10.3389/fncir.2013.00107.
20. Zhao X., Yan S., Gao Q. An Algorithm for Tracking Multiple Fish Based on Biological Water Quality Monitoring // IEEE Access. 2019. Vol. 7. P. 15018–15026. DOI: 10.1109/ACCESS.2019.2895072.
21. Delcourt J., Denoel M., Ylief M., Poncin P. Video multitasking of fish behaviour: A synthesis and future perspectives // Fish and Fisheries. 2013. Vol. 14. No. 2. P. 186–204. DOI: 10.1111/j.1467-2979.2012.00462.x.
22. Sevillano L. Convolutional Neural Network for Fish Detection and Tracking in UUV Videos: магістер. дис. Trondheim: NTNU, 2019. 74 p.
23. NESTLER Horizon Project (Distributor). NESTLER Poultry Video Dataset [Електронний ресурс]. 2025. Режим доступу: <https://zenodo.org/records/15063494>. DOI: 10.5281/zenodo.15063494.
24. Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: Unified, Real-Time Object Detection: препринт arXiv:1506.02640. 2015. Режим доступу: <https://arxiv.org/abs/1506.02640>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1506.02640.
25. Redmon J., Farhadi A. YOLO9000: Better, Faster, Stronger: препринт arXiv:1612.08242. 2016. Режим доступу: <https://arxiv.org/abs/1612.08242>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1612.08242.
26. Zhou X., Wang D., Krähenbühl P. Objects as Points: препринт arXiv:1904.07850. 2019. Режим доступу: <https://arxiv.org/abs/1904.07850>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1904.07850.
27. Duan K., Bai S., Xie L., Qi H., Huang Q., Tian Q. CenterNet: Keypoint Triplets for Object Detection: препринт arXiv:1904.08189. 2019. Режим

доступу: <https://arxiv.org/abs/1904.08189>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1904.08189.

28. Law H., Deng J. CornerNet: Detecting Objects as Paired Keypoints: препринт arXiv:1808.01244. 2018. Режим доступу: <https://arxiv.org/abs/1808.01244>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1808.01244.

29. Lin T.-Y., Goyal P., Girshick R., He K., Dollár P. Focal Loss for Dense Object Detection: препринт arXiv:1708.02002. 2017. Режим доступу: <https://arxiv.org/abs/1708.02002>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1708.02002.

30. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition: препринт arXiv:1512.03385. 2015. Режим доступу: <https://arxiv.org/abs/1512.03385>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1512.03385.

31. Pratt W.K. Digital Image Processing: 4th ed. Hoboken: Wiley-Interscience, 2007. 807 p.

32. Bouwmans T. Traditional and Recent Approaches in Background Modeling for Foreground Detection: An Overview // Computer Science Review. 2014. Vol. 11–12. P. 31–66. DOI: 10.1016/j.cosrev.2014.04.001.

33. Arbelaez P., Maire M., Fowlkes C., Malik J. Contour Detection and Hierarchical Image Segmentation // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2011. Vol. 33. No. 5. P. 898–916. DOI: 10.1109/TPAMI.2010.161.

34. Volkova N., Shvandt M. Segmentation-based approach for object detection // Proceedings of Odessa Polytechnic University. 2025. Issue 1(71). P. 145–156. DOI: 10.15276/opu.1.71.2025.17.

35. Harris C., Stephens M. A Combined Corner and Edge Detector // Proceedings of the Fourth Alvey Vision Conference. 1988. P. 147–151.

36. Shi J., Tomasi C. Good Features to Track // CVPR 1994: Proceedings. 1994. P. 593–600.

37. Lim D., Ang A., Seng K. Real-time Tracking using OpenCV (C++/Python) // ICT4M 2015: Proceedings. 2015. p. DOI: 10.1109/ICT4M.2015.7407498.
38. Shijie J. et al. Using RGB Color Space for Object Detection and Tracking // ICIST 2016: Proceedings. 2016. p. DOI: 10.1109/ICIST.2016.7483331.
39. Culjak M., Abram D., Pribanic T., Dzapo H., Cifrek M. A Brief Introduction to OpenCV // MIPRO 2012: Proceedings of the 35th International Convention. 2012. P. 1725–1730.
40. Viola P., Jones M. Rapid Object Detection using a Boosted Cascade of Simple Features // CVPR 2001: Proceedings. 2001. 8 p.
41. Mallick S. Simple Background Estimation in Videos using OpenCV (C++/Python) [Электронный ресурс]. Режим доступа: <https://learnopencv.com/simple-background-estimation-in-videos-using-opencv-c-python/>
42. Canny J. A Computational Approach to Edge Detection // IEEE Transactions on Pattern Analysis and Machine Intelligence. 1986. Vol. 8. No. 6. P. 679–698. DOI: 10.1109/TPAMI.1986.4767851.
43. Kong H., Akakin H.C., Sarma S.E. A Generalized Laplacian of Gaussian Filter for Blob Detection and Its Applications // IEEE Transactions on Cybernetics. 2013. Vol. 43. No. 6. P. 1719–1733. DOI: 10.1109/TSMCB.2012.2228639.
44. Beucher S. The Watershed Transformation Applied to Image Segmentation // Scanning Microscopy. 2000. Vol. 6. 26 p.
45. Opromolla M., Fasano C., Rufino G. A Template Matching based Approach for Autonomous Detection and Tracking of Space Debris // Acta Astronautica. 2015. Vol. 108. P. 48–59. DOI: 10.1016/j.actaastro.2014.12.013.
46. Dalal N., Triggs B. Histograms of Oriented Gradients for Human Detection // CVPR 2005: Proceedings. 2005. 9 p. DOI: 10.1109/CVPR.2005.177.

47. Lowe D.G. Distinctive Image Features from Scale-Invariant Keypoints // International Journal of Computer Vision. 2004. 28 p. DOI: 10.1023/B:VISI.0000029664.99615.94.
48. Bay H., Tuytelaars T., Van Gool L. SURF: Speeded Up Robust Features // Computer Vision – ECCV 2006: Proceedings. Lecture Notes in Computer Science. 2006. Vol. 3951. P. 404–417. DOI: 10.1007/11744023_32.
49. Alcantarilla P.F., Bartoli A., Davison A.J. KAZE Features // Computer Vision – ECCV 2012: Proceedings. Lecture Notes in Computer Science. 2012. Vol. 7577. P. 214–227. DOI: 10.1007/978-3-642-33783-3_16.
50. Rublee E., Rabaud V., Konolige K., Bradski G. ORB: an efficient alternative to SIFT or SURF // ICCV 2011: Proceedings of the IEEE International Conference on Computer Vision. 2011. P. 2564–2571. DOI: 10.1109/ICCV.2011.6126544.
51. OpenCV 4.5.2. Documentation [Електронний ресурс]. Режим доступу: <https://docs.opencv.org/4.5.2/>
52. Lowe D.G. Object recognition from local scale-invariant features // ICCV 1999: Proceedings. 1999. Vol. 2. P. 1150–1157.
53. Brunelli R. Template Matching Techniques in Computer Vision. Theory and Practice. Chichester: Wiley, 2009. 339 p.
54. Khedher A.S., Alkababji A.M., Hadi O. Improving the Reliability of Object Recognition Based On Template Matching // AL Rafdain Engineering Journal. 2015. P. 81–88.
55. OpenCV. Template Matching: tutorial [Електронний ресурс]. Режим доступу: https://docs.opencv.org/3.4/de/da9/tutorial_template_matching.html
56. Danelljan M., Häger G., Khan F.S., Felsberg M. Learning Spatially Regularized Correlation Filters for Visual Tracking: препринт arXiv:1608.05571. 2016. Режим доступу: <https://arxiv.org/abs/1608.05571>. – Дата звернення: 8 травня 2026 р.

57. Fiaz M., Mahmood A. Real-time Visual Object Tracking: A Multi- scale Approach // Journal of Visual Communication and Image Representation. 2017. Vol. 48. P. 170–180. DOI: 10.1016/j.jvcir.2017.08.004.
58. Rilman I., Brown D. Retinal Vessel Tracking by using Correlation Filters and Template Matching // ICoDSA 2017: Proceedings. 2017. 6 p. DOI: 10.1109/ICoDSA.2017.8275525.
59. Wu Y., Lim J., Yang M.-H. Online Object Tracking: A Benchmark // CVPR 2013: Proceedings. 2013. 14 p. DOI: 10.1007/s11263-013-0620-5.
60. Shvandt M., Moroz V. General approaches to lab animal detection and tracking // Problems of Informatics and Modeling (PIM-2021): Proceedings/Abstracts of the XXI International Scientific and Technical Conference (September 09–14, 2021, Karolino-Buhaz, Ukraine).
61. Godoy Y. OpenCV By Example. Birmingham: Packt Publishing, 2016. 294 p.
62. Lucas B., Kanade T. An iterative image registration technique with an application to stereo vision // IJCAI '81: Proceedings. 1981. P. 674–679.
63. He A., Fu Y. Optical Flow Tracking with Pyramid Lucas-Kanade Method // CSAE 2018: Proceedings. 2018. 6 p. DOI: 10.1145/3207677.3278027.
64. Radgui A., Demonceaux C., Mouaddib E., Aboutajdine D., Rziza M. An adapted Lucas–Kanade’s method for optical flow estimation in catadioptric images // OMNIVIS 2008: Proceedings. Marseille, 2008. 12 p.
65. OpenCV. Optical Flow (Feature Tracking): tutorial [Електронний ресурс]. Режим доступу: https://docs.opencv.org/3.4/d4/dee/tutorial_optical_flow.html
66. Comaniciu D., Ramesh V., Meer P. Kernel-based Object Tracking // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2003. Vol. 25. No. 5. P. 564–577. DOI: 10.1109/TPAMI.2003.1195991.
67. Bolme D.S., Beveridge J.R., Draper B.A., Lui Y.M. Visual Object Tracking using Adaptive Correlation Filters // CVPR 2010: Proceedings. 2010. 10 p. DOI: 10.1109/CVPR.2010.5539960.

68. Danelljan M. et al. Accurate Scale Estimation for Robust Visual Tracking // BMVC 2014: Proceedings. 2014. P. 1–11. DOI: 10.5244/C.28.65.

69. Mallick S. Object Tracking using OpenCV (C++/Python) [Электронный ресурс]. Режим доступа: <https://learnopencv.com/object-tracking-using-opencv-cpp-python/>

70. Babenko B., Yang M.-H., Belongie S. Robust Object Tracking with Online Multiple Instance Learning // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2011. Vol. 33. No. 8. P. 1619–1632. DOI: 10.1109/TPAMI.2010.226.

71. Kalal Z., Mikolajczyk K., Matas J. Forward-Backward Error: Automatic Detection of Tracking Failures // ICPR 2010: Proceedings. Istanbul, 2010. 4 p. DOI: 10.1109/ICPR.2010.675.

72. Kalal Z., Mikolajczyk K., Matas J. Tracking-Learning-Detection // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2012. Vol. 34, No. 7. P. 1409–1422. DOI: 10.1109/TPAMI.2011.239.

73. Zhou S., Peng Y., Gong K., Shu L. An Improved TLD Tracking Algorithm for Fast-moving Object // CSECE 2018: Proceedings (Advances in Computer Science Research). 2018. Vol. 80. P. 69–73.

74. Shvandt M. A., Moroz V. V. Overview of the detection and tracking methods of the lab animals // System Research & Information Technologies. 2022. No. 1. P. 124–148. DOI: 10.20535/SRIT.2308-8893.2022.1.10.

75. Shvandt M. A., Moroz V. V. Overview of neural network object detection methods & models on the example of their use for lab animal observation // System Research & Information Technologies. 2025. No. 4. P. 71–103. DOI: 10.20535/SRIT.2308-8893.2025.4.05.

76. TensorFlow 2 Detection Model Zoo [Электронный ресурс]. Режим доступа: https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tf2_detection_zoo.md

77. Lu X., Li B., Yue Y., Li Q., Yan J. Grid R-CNN: препринт arXiv:1811.12030. 2018. Режим доступа: <https://arxiv.org/abs/1811.12030>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1811.12030.

78. Girshick R. Fast R-CNN: препринт arXiv:1504.08083. 2015. Режим доступа: <https://arxiv.org/abs/1504.08083>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1504.08083.

79. Girshick R. et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation: препринт arXiv:1311.2524. 2013. Режим доступа: <https://arxiv.org/abs/1311.2524>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1311.2524.

80. He K., Zhang X., Ren S., Sun J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition: препринт arXiv:1406.4729. 2014. Режим доступа: <https://arxiv.org/abs/1406.4729>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1406.4729.

81. Dai J. et al. R-FCN: Object Detection via Region-based Fully Convolutional Networks: препринт arXiv:1605.06409. 2016. Режим доступа: <https://arxiv.org/abs/1605.06409>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1605.06409.

82. Liu L., Ouyang W., Wang X., Fieguth P., Chen J., Liu X., Pietikäinen M. Deep Learning for Generic Object Detection: A Survey // International Journal of Computer Vision. 2020. Vol. 128, No. 2. P. 261–318. DOI: 10.1007/s11263-019-01247-4.

83. Callaris E. Faster R-CNN Object Detection [Електронний ресурс]. Режим доступа: <https://towardsdatascience.com/faster-rcnn-object-detection-f865e5ed7fc4>

84. Mohamed H.E.-D. et al. MSR-YOLO: Method to Enhance Fish Detection and Tracking // ANT 2020: The 11th International Conference on Ambient Systems, Networks and Technologies. Warsaw, 2020.

85. Kundu R. YOLO: Algorithm for Object Detection Explained [+Examples] [Електронний ресурс]. Режим доступу: <https://www.v7labs.com/blog/yolo-object-detection>

86. Tsang S.-H. Brief Review: YOLOv5 for Object Detection [Електронний ресурс]. Режим доступу: <https://sh-tsang.medium.com/brief-review-yolov5-for-object-detection-84cc6c6a0e3a>

87. Bochkovskiy A., Wang C.-Y., Liao H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection: препринт arXiv:2004.10934. 2020. Режим доступу: <https://arxiv.org/abs/2004.10934>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.2004.10934.

88. Trivedi S. CenterNet: Objects as Points – A Comprehensive Guide [Електронний ресурс]. Режим доступу: <https://medium.com/visionwizard/centernet-objects-as-points-a-comprehensive-guide-2ed9993c48bc>

89. Huang L., Yang Y., Deng Y., Yu Y. DenseBox: Unifying Landmark Localization with End to End Object Detection: препринт arXiv:1509.04874. 2015. Режим доступу: <https://arxiv.org/abs/1509.04874>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1509.04874.

90. Wang D., Shelhamer E., Darrell T. Deep Layer Aggregation: препринт arXiv:1707.06484. 2017. Режим доступу: <https://arxiv.org/abs/1707.06484>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1707.06484.

91. Trivedi S. Understanding Focal Loss – A Quick Read [Електронний ресурс]. Режим доступу: <https://medium.com/visionwizard/understanding-focal-loss-a-quick-read-b914422913e7>

92. Howard A.G. et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications: препринт arXiv:1704.04861. 2017. Режим доступу: <https://arxiv.org/abs/1704.04861>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1704.04861.

93. Bangar S. Resnet Architecture Explained [Електронний ресурс]. Режим доступу: <https://medium.com/@siddheshb008/resnet-architecture-explained-47309ea9283d>
94. Szegedy C. et al. Going Deeper with Convolutions: препринт arXiv:1409.4842. 2014. Режим доступу: <https://arxiv.org/abs/1409.4842>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1409.4842.
95. Krizhevsky A., Sutskever I., Hinton G.E. ImageNet Classification with Deep Convolutional Neural Networks // Communications of the ACM. 2017. Vol. 60. No. 6. P. 84–90. DOI: 10.1145/3065386.
96. Simonyan K., Zisserman A. Very Deep Convolutional Neural Networks for Large-Scale Image Recognition: препринт arXiv:1409.1556. 2014. Режим доступу: <https://arxiv.org/abs/1409.1556>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1409.1556.
97. Zeiler M.D., Fergus R. Visualizing and Understanding Convolutional Neural Networks: препринт arXiv:1311.2901. 2013. Режим доступу: <https://arxiv.org/abs/1311.2901>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1311.2901.
98. Lin T.-Y. et al. Feature Pyramid Networks for Object Detection: препринт arXiv:1612.03144. 2016. Режим доступу: <https://arxiv.org/abs/1612.03144>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1612.03144.
99. Singh P. Feature Pyramid Networks (FPN): Feature Pyramids for Object Detection [Електронний ресурс]. Режим доступу: <https://medium.com/analytics-vidhya/fpn-feature-pyramid-networks-77d8be41817c>
100. Long J., Shelhamer E., Darrell T. Fully Convolutional Networks for Semantic Segmentation: препринт arXiv:1411.4038. 2014. Режим доступу: <https://arxiv.org/abs/1411.4038>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1411.4038.
101. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation: препринт arXiv:1505.04597. 2015. Режим

доступу: <https://arxiv.org/abs/1505.04597>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1505.04597.

102. Milletari F., Navab N., Ahmadi S.-A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation: препринт arXiv:1606.04797. 2016. Режим доступу: <https://arxiv.org/abs/1606.04797>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1606.04797.

103. Newell A., Yang K., Deng J. Stacked Hourglass Networks for Human Pose Estimation: препринт arXiv:1603.06937. 2016. Режим доступу: <https://arxiv.org/abs/1603.06937>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1603.06937.

104. Hu J., Shen L., Albanie S., Sun G., Wu E. Squeeze-and-Excitation Networks: препринт arXiv:1709.01507. 2017. Режим доступу: <https://arxiv.org/abs/1709.01507>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1709.01507.

105. Shvandt M., Volkova N. Extended background subtraction algorithm for object detection & tracking // Project, Program, Portfolio Management (P3M-2024): The Proceedings of the International Research Conference (December 06–07, 2024, Odesa, Ukraine). 2024. P. 73. DOI: 10.5281/zenodo.15165174.

106. Lin T.-Y. et al. Microsoft COCO: Common Objects in Context: препринт arXiv:1405.0312. 2014. Режим доступу: <https://arxiv.org/abs/1405.0312>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1405.0312.

107. Wood L., Chollet F. Efficient Graph-Friendly COCO Metric Computation for Train-Time Model Evaluation: препринт arXiv:2207.12120. 2022. Режим доступу: <https://arxiv.org/abs/2207.12120>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.2207.12120.

108. Calleris A. Hourglass Network [Електронний ресурс]. Режим доступу: <https://medium.com/@calleris.enrico/hourglass-network-6e74cdb9ce2f>

109. Cai H., Gan C., Wang T., Zhang Z., Han S. Once-for-All: Train One Network and Specialize it for Efficient Deployment: препринт arXiv:1908.09791.

2020. Режим доступу: <https://arxiv.org/abs/1908.09791>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1908.09791.

110. Zhang Y. et al. Object Detection in 20 Years: A Survey: препринт arXiv:1905.05055. 2020. Режим доступу: <https://arxiv.org/abs/1905.05055>. – Дата звернення: 8 травня 2026 р. DOI: 10.48550/arXiv.1905.05055.

111. Bertinetto L., Valmadre J., Henriques J.F., Vedaldi A., Torr P.H.S. Fully-Convolutional Siamese Networks for Object Tracking // ECCV 2016: Proceedings. 2016. 11 p. DOI: 10.48550/arXiv.1606.09549.

112. Li S., Yang L., Tan H., Lan L., Lin Y. Self-supervised re-identification for online joint multi-object tracking // Knowledge and Information Systems. 2025. Advance online publication. DOI: 10.1007/s10115-024-02237-w.

113. Liu H., Jie Z., Karlekar J., Zhao Y., Chai D., Lim J.H., Pranata S. Video-based person re-identification with accumulative motion context // IEEE Transactions on Circuits and Systems for Video Technology. 2018. Vol. 28, No. 10. P. 2788–2802. DOI: 10.1109/TCSVT.2017.2715499.

114. Мороз В.В., Швандт М.А. Дослідження руху та поведінки лабораторних тварин методами детектування і трекінгу об'єктів // Вісник НТУ «ХПІ». Серія «Інформатика та моделювання». 2019. Вип. 13. № 1338. С. 93–103. DOI: 10.20998/2411-0558.2019.13.09.

115. Volkova N. P., Shvandt M. A. Unsupervised re-identification architecture based on segmented tracklets for animal behavior analysis // Herald of Advanced Information Technology. 2025. Vol. 8, No. 4. P. 476–487. DOI: 10.15276/hait.08.2025.30.

116. Li M., Zhu X., Gong S. Unsupervised Person Re-identification by Deep Learning Tracklet Association // European Conference on Computer Vision (ECCV 2018): Proceedings. 2018. DOI: 10.1007/978-3-030-01225-0_45.

117. Volkova N., Shvandt M. Segmentation-based extension for object detection // Modern Information Technology – 2025: Proceedings of the Fifteenth International Scientific Conference of Students and Young Scientists (May 15–16,

2025, Odesa, Ukraine). Odesa: Nauka i Tekhnika, 2025. DOI: 10.5281/zenodo.15521809.

118. Achanta R., Shaji A., Smith K., Lucchi A., Fua P., Süsstrunk S. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2012. Vol. 34, No. 11. P. 2274–2282. DOI: 10.1109/TPAMI.2012.120.

119. Tomasi C., Manduchi R. Bilateral Filtering for Gray and Color Images // Computer Vision: Proceedings of the Sixth International Conference on Computer Vision (ICCV'98). Bombay, 1998. P. 839–846. DOI: 10.1109/ICCV.1998.710815.

120. Sanin A., Sanderson C., Lovell B.C. Shadow detection: a survey and comparative evaluation of recent methods // Pattern Recognition. 2012. Vol. 45, No. 4. P. 1684–1695. DOI: 10.1016/j.patcog.2011.10.001.

121. Felzenszwalb P.F., Huttenlocher D.P. Efficient Graph-Based Image Segmentation // International Journal of Computer Vision. 2004. Vol. 59, No. 2. P. 167–181. DOI: 10.1023/B:VISI.0000022288.19776.77.

122. Vedaldi A., Soatto S. Quick Shift and Kernel Methods for Mode Seeking // Computer Vision – ECCV 2008: Proceedings. Lecture Notes in Computer Science. 2008. Vol. 5305. P. 705–718. DOI: 10.1007/978-3-540-88693-8_52.

123. Volkova N., Shvandt M. Image preprocessing algorithm for object detection & tracking // Information Control Systems and Technologies (ICST-ODESA 2024): Proceedings of the XII International Scientific-Practical Conference (September 23–25, 2024, Odesa). Odesa, 2024. P. 194–197.

124. Volkova N., Shvandt M. Modification of the Mask R-CNN architecture for image detection and segmentation // Informatics and Mathematical Methods in Simulation. 2025. Vol. 15, No. 4. P. 496–506. DOI: 10.15276/imms.v15.no4.496.

125. Ultralytics. Instance Segmentation (Segment task) Documentation [Електронний ресурс]. Режим доступу: <https://docs.ultralytics.com/tasks/segment/>

126. Ultralytics. Explore Ultralytics YOLOv8 (Models / YOLOv8) [Электронный ресурс]. Режим доступа: <https://docs.ultralytics.com/models/yolov8/>
127. Sapkota R., Ahmed D., Karkee M. Comparing YOLOv8 and Mask R-CNN for object segmentation in complex orchard environments: препринт arXiv:2312.07935. 2023. Режим доступа: <https://arxiv.org/abs/2312.07935>. – Дата звернення: 8 травня 2026 р.
128. Shrivastava S., Bhattacharjee S., Deb D. Segmentation of mine overburden dump particles from images using Mask R CNN // Scientific Reports. 2023. Vol. 13, No. 1. Article ID 2046. DOI: 10.1038/s41598-023-28586-0.
129. Thai T.T. et al. Mask R-CNN vs. YOLOv8 on Phenomics Stomatal dataset // Frontiers in Plant Science. 2024. [Электронный ресурс]. Режим доступа: <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2024.1414849/full>
130. Wojke N., Bewley A., Paulus D. Simple online and realtime tracking with a deep association metric // 2017 IEEE International Conference on Image Processing (ICIP). 2017. P. 3645–3649. DOI: 10.1109/ICIP.2017.8296962.
131. Babaee M., Athar A., Rigoll G. Multiple people tracking using hierarchical deep tracklet re-identification: препринт arXiv:1811.04091. 2018. DOI: 10.48550/arXiv.1811.04091.
132. Liu Q., Chen D., Chu Q., Yuan L., Liu B., Zhang L., Yu N. Online multi-object tracking with unsupervised re-identification learning and occlusion estimation // Neurocomputing. 2022. Vol. 483. P. 333–347. DOI: 10.1016/j.neucom.2022.01.008.
133. Yang J., Wang S., Jiang T. Progressive unsupervised video person re-identification with accumulative motion and tracklet spatial–temporal correlation // Future Generation Computer Systems. 2023. Vol. 142. P. 90–100. DOI: 10.1016/j.future.2022.12.023.
134. Wu S., Sun Y., Luo J., Zhang C., Huang Q. Tracklet self-supervised learning for unsupervised person re-identification // Proceedings of the AAAI

Conference on Artificial Intelligence. 2020. Vol. 34, No. 7. P. 12362–12369. DOI: 10.1609/aaai.v34i07.6921.

135. Zang X., Li G., Gao W., Shu X. Exploiting robust unsupervised video person re-identification // IET Image Processing. 2022. Vol. 16, No. 3. P. 729–741. DOI: 10.1049/ipr2.12380.

136. Volkova N., Shvandt M. Object detection network architecture for multichannel images // Information Control Systems and Technologies (ICST-ODESA 2025): Proceedings of the XIII International Scientific-Practical Conference. Liha-Pres (Lviv–Torun). 2025. P. 166–169. DOI: 10.36059/978-966-397-531-3.

137. Volkova N., Shvandt M. Improved approach to label-free identification using segmented tracklets for object behavior analysis // Informatics. Culture. Technique. 2025. No. 2. P. 279–282. DOI: 10.15276/ict.02.2025.42.

138. Bernardin K., Stiefelwagen R. Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics // EURASIP Journal on Image and Video Processing. 2008. Article ID 246309. DOI: 10.1155/2008/246309.

139. Dendorfer P., Osep A., Milan A., Schindler K., Cremers D., Reid I., Roth S., Leal-Taixé L. MOTChallenge: A Benchmark for Single-Camera Multiple Target Tracking // International Journal of Computer Vision. 2021. Vol. 129. No. 4. P. 845–881. DOI: 10.1007/s11263-020-01393-0.

ДОДАТОК А

СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Мороз В.В., Швандт М.А. Дослідження руху та поведінки лабораторних тварин методами детектування і трекінгу об'єктів // Вісник НТУ «ХПІ». Серія «Інформатика та моделювання». 2019. Вип. 13. № 1338. С. 66–76. DOI: 10.20998/2411-0558.2019.13.09. Видання включено до переліку наукових фахових видань України, категорія «Б».

<https://pim.khpi.edu.ua/article/view/2411-0558.2019.13.09/179328>

2. Shvandt M. A., Moroz V. V. Overview of the detection and tracking methods of the lab animals // System Research & Information Technologies. 2022. No. 1. P. 124–148. DOI: 10.20535/SRIT.2308-8893.2022.1.10. Видання включено до переліку наукових фахових видань України, категорія «А» (Scopus).

<https://journal.iasa.kpi.ua/article/view/237497/255908>

3. Shvandt M. A., Moroz V. V. Overview of neural network object detection methods & models on the example of their use for lab animal observation // System Research & Information Technologies. 2025. No. 4. P. 71–103. DOI: 10.20535/SRIT.2308-8893.2025.4.05. Видання включено до переліку наукових фахових видань України, категорія «А» (Scopus).

<https://journal.iasa.kpi.ua/article/view/351422/338441>

4. Volkova N., Shvandt M. Segmentation-based approach for object detection // Proceedings of Odessa Polytechnic University. 2025. Issue 1(71). P. 145–156. DOI: 10.15276/opu.1.71.2025.17. Видання включено до переліку наукових фахових видань України, категорія «Б».

<https://pratsi.op.edu.ua/index.php/pratsi/article/view/516/488>

5. Volkova N., Shvandt M. Modification of the Mask R-CNN architecture for image detection and segmentation // Informatics and Mathematical Methods in Simulation. 2025. Vol. 15, No. 4. P. 496–506. DOI: 10.15276/imms.v15.no4.496.

Видання включено до переліку наукових фахових видань України, категорія «Б».

[https://immm.op.edu.ua/files/archive/n4_v15_2025/2025_4\(4\).pdf](https://immm.op.edu.ua/files/archive/n4_v15_2025/2025_4(4).pdf)

6. Volkova N. P., Shvandt M. A. Unsupervised re-identification architecture based on segmented tracklets for animal behavior analysis // Herald of Advanced Information Technology. 2025. Vol. 8, No. 4. P. 476–487. DOI: 10.15276/hait.08.2025.30. Видання включено до переліку наукових фахових видань України, категорія «А» (Scopus).

<https://hait.od.ua/index.php/journal/article/view/230/202>

7. Shvandt M., Moroz V. General approaches to lab animal detection and tracking // Problems of Informatics and Modeling (PIM-2021): Proceedings/Abstracts of the XXI International Scientific and Technical Conference (September 09–14, 2021, Karolino-Buhaz, Ukraine).

<https://repository.kpi.kharkov.ua/server/api/core/bitstreams/c079ac2d-ba5c-467a-b2f3-55d4e921588d/content>

8. Volkova N., Shvandt M. Image preprocessing algorithm for object detection & tracking // Information Control Systems and Technologies (ICST-ODESA 2024): Proceedings of the XII International Scientific-Practical Conference (September 23–25, 2024, Odesa). 2024. P. 194–197.

<https://icst-conf.com/2024.pdf>

9. Shvandt M., Volkova N. Extended background subtraction algorithm for object detection & tracking // Project, Program, Portfolio Management (P3M-2024): The Proceedings of the International Research Conference (December 06–07, 2024, Odesa, Ukraine). 2024. P. 73–76. DOI: 10.5281/zenodo.15165174.

<https://zenodo.org/records/15165174>

10. Volkova N., Shvandt M. Segmentation-based extension for object detection // Modern Information Technology – 2025: Proceedings of the Fifteenth International Scientific Conference of Students and Young Scientists (May 15–16, 2025, Odesa, Ukraine). Odesa: Nauka i Tekhnika, 2025. DOI: 10.5281/zenodo.15521809.

<https://doi.org/10.5281/zenodo.15521809>

11. Volkova N., Shvandt M. Object detection network architecture for multichannel images // Information Control Systems and Technologies (ICST-ODESA 2025): Proceedings of the XIII International Scientific-Practical Conference (September 24–26, 2025, Odesa). 2025. P. 166–169. DOI: 10.36059/978-966-397-531-3.

<https://icst-conf.com/2025.pdf>

12. Volkova N., Shvandt M. Improved approach to label-free identification using segmented tracklets for object behavior analysis // Informatics. Culture. Technique. 2025. No. 2. P. 279–282. DOI: 10.15276/ict.02.2025.42.

<https://ict.od.ua/index.php/journal/en/article/view/53>

ДОДАТОК Б

**АКТ ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ ДИСЕРТАЦІЙНОЇ РОБОТИ У
ЛАБОРАТОРНІ ДОСЛІДЖЕННЯ БІОЛОГІЧНОГО ФАКУЛЬТЕТУ
ОДЕСЬКОГО НАЦІОНАЛЬНОГО УНІВЕРСИТЕТУ ІМЕНІ І. І.
МЕЧНИКОВА**

ЗАТВЕРДЖУЮ

Проректор з наукової роботи
Одеського національного


університету
імені І. І. Мечникова
Гавриш В.О.
05 2026 р.

ЗАТВЕРДЖУЮ

В.о. ректора
Одеського національного


університету
імені І. І. Мечникова
Насолава М.І.
«05» 2026 р.

АКТ

про впровадження у лабораторні дослідження результатів дисертаційної роботи за темою «Моделі та методи детектування, трекінгу та наскрізної ідентифікації лабораторних тварин у відеоданих» на здобуття наукового ступеня доктора філософії за спеціальністю 122 – Комп'ютерні науки Швандта Максима Альбертовича

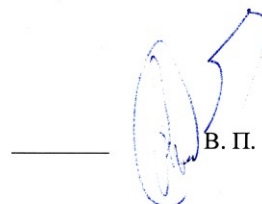
Комісія у складі заступника декана з наукової роботи біологічного факультету Одеського національного університету імені І. І. Мечникова, кандидата біологічних наук, доцента кафедри молекулярної біології, біохімії і генетики Чернадчук Сніжани Сергіївни та завідувача кафедри зоології, гідробіології та загальної екології, доктора біологічних наук, професора Стойловського Володимира Петровича склали цей акт про те, що в Одеському національному університеті імені І. І. Мечникова у лабораторних дослідженнях, що проводяться на базі біологічного факультету, використовувались оригінальні матеріали, одержані у дисертаційній роботі Швандта М. А., а саме: спеціалізований конвеєр детектування, трекінгу та наскрізної ідентифікації лабораторних тварин у відеоданих для автоматизованого моніторингу їхньої поведінки.

Заступник декана з
наукової роботи
біологічного факультету,
к. б. н., доцент



С. С. Чернадчук

Завідувач кафедри
зоології,
гідробіології та загальної
екології,
д. б. н., професор



В. П. Стойловський

ДОДАТОК В
АКТ ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ ДИСЕРТАЦІЙНОЇ РОБОТИ У
НАВЧАЛЬНИЙ ПРОЦЕС ОДЕСЬКОГО НАЦІОНАЛЬНОГО
УНІВЕРСИТЕТУ ІМЕНІ І. І. МЕЧНИКОВА



ЗАТВЕРДЖУЮ

В.о. ректора

Одеського національного університету
імені І.І. Мечникова

Ніколаєва М.І.

05

2026 р.

АКТ

про впровадження в навчальний процес результатів дисертаційної роботи за темою «Моделі та методи детектування, трекінгу та наскрізної ідентифікації лабораторних тварин у відеоданих» на здобуття наукового ступеня доктора філософії за спеціальністю 122 - Комп'ютерні науки Швандта Максима Альбертовича

Комісія у складі заступник декана з наукової роботи біологічного факультету Одеського національного університету імені І. І. Мечникова, кандидата біологічних наук, доцента кафедри молекулярної біології, біохімії і генетики Чернадчук Сніжани Сергіївни та завідувача кафедри зоології, гідробіології та загальної екології доктора біологічних наук, професора Стойловського Володимира Петровича склали цей акт про те, що в Одеському національному університеті імені І.І. Мечникова, при викладанні дисциплін «Основи іхтіології» та «Спеціальна іхтіологія», а також в дипломному проектуванні здобувачів спеціальностей 091 «Біологія» використовувались оригінальні матеріали, які отримано у дисертаційній роботі Швандта М.А., а саме: розроблені моделі та методи комп'ютерного зору що є складовими спеціалізованого конвеєру детектування, трекінгу та ідентифікації лабораторних тварин для автоматизованого моніторингу їхньої поведінки.

Заступник декана з
наукової роботи
біологічного факультету,
к. б. н., доцент

С. С. Чернадчук

Завідувач кафедри
зоології,
гідробіології та загальної
екології,
д. б. н., професор

В. П. Стойловський

ДОДАТОК Г

**ДОВІДКА ПРО ВИКОРИСТАННЯ РЕЗУЛЬТАТІВ ДИСЕРТАЦІЙНОЇ
РОБОТИ У НАУКОВО-ДОСЛІДНИЦЬКІЙ ДІЯЛЬНОСТІ
НАЦІОНАЛЬНОГО УНІВЕРСИТЕТУ «ОДЕСЬКА ПОЛІТЕХНІКА»**

ДОВІДКА

про використання результатів дисертаційної роботи

Швандта Максима Альбертовича

«Моделі та методи детектування, трекінгу та наскрізної ідентифікації лабораторних тварин у відеоданих»

у науково-дослідницькій діяльності

Національного університету «Одеська Політехніка»

Довідка видана стосовно того, що у науково-дослідницькій діяльності Національного університету «Одеська Політехніка» використані наступні наукові результати, отримані у дисертаційній роботі Швандта Максима Альбертовича:

– нейромережева модель Mask R-CNN за рахунок введення подвійної магістралі екстракції ознак із керованим SE-злиттям ($B \rightarrow A$) для підвищення якості детектування та сегментації лабораторних тварин у відеоданих за умов часткового перекриття і розділення лабораторних тварин при збереженні прийнятної оперативності.

Дисертація є складовою частиною НДР № 264-73 «Методи та технології моделювання і забезпечення якості складних програмних систем у динамічних середовищах» (№ держреєстрації 0125U003655).

Проректор з наукової та
науково-педагогічної роботи
д.т.н., професор



Керівник НДР № 264-73
к.т.н., доцент

Дмитро ДМИТРИШИН

Наталія КОМЛЕВА

ДОДАТОК Д
ДОВІДКА ПРО ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ ДИСЕРТАЦІЙНОЇ
РОБОТИ У НАВЧАЛЬНОМУ ПРОЦЕСІ НАЦІОНАЛЬНОГО
УНІВЕРСИТЕТУ «ОДЕСЬКА ПОЛІТЕХНІКА»

«ЗАТВЕРДЖУЮ»

Перший проректор, проректор з
науково-педагогічної та виховної
роботи Національного університету

«Одеська політехніка»,

професор Сергій Пестеренко



04 2026 р.

ДОВІДКА

про впровадження результатів дисертації доктора філософії

ШВАНДТА Максима Альбертовича

**«Моделі та методи детектування, трекінгу та наскрізної ідентифікації лабораторних
тварин у відеоданих»**

у навчальному процесі Національного університету «Одеська політехніка»

Чинна довідка видана стосовно того, що в навчальних курсах бакалаврської підготовки «Основи нейронних мереж» та «Методи побудови штучного інтелекту», які викладаються студентам за спеціальностями 113 «Прикладна математика» і 124 - «Системний аналіз та наука про дані» на кафедрі прикладної математики та інформаційних технологій Інституту комп'ютерних систем Національного університету «Одеська політехніка», використовуються наукові результати, одержані в дисертаційній роботі Швандта М.А.

Основні наукові та практичні результати, що знайшли застосування:

1. Метод суперпиксельної сегментації лабораторних тварин у відеоданих, удосконалений за рахунок використання багатопланового простору ознак, який враховує LAB-представлення та представлення різницевого зображення, а також перевірок однорідності пікселів у локальному оточенні, що дозволяє підвищити точність локалізації об'єктів за умов нерівномірного освітлення, оптичних артефактів та складного фону.

—тема «Системи штучного інтелекту. Передчасна обробка сигналів та зображень» - у межах дисципліни «Методи побудови штучного інтелекту»;

— тема «Сегментація зображень. Ідентифікація об'єктів на зображенні» - у межах дисципліни «Методи побудови штучного інтелекту».

2. Метод детектування об'єктів у відеоданих, що поєднує метод суперпиксельної сегментації SLIC і контекстну фільтрацію результатів та забезпечує підвищення точності детектування об'єктів у складних сценах.

– тема «Сегментація зображень. Ідентифікація об'єктів на зображенні» - у межах дисципліни «Методи побудови штучного інтелекту»;

– тема «Класифікація зображень» - у межах дисципліни «Методи побудови штучного інтелекту».

3. Удосконалена нейромережева модель Mask R-CNN із подвійною магістраллю екстракції ознак і керованим SE-злиттям ($B \rightarrow A$), що дозволяє підвищити якість детектування та сегментації об'єктів у відеоданих за умов часткового перекриття, злиття та подальшого розділення об'єктів при збереженні прийнятної оперативності.

– тема «Структура штучної нейронної мережі. Функція активації штучного нейрона» – у межах дисципліни «Основи нейронних мереж»;

– тема «Методи та алгоритми навчання штучних нейронних мереж» - у межах дисципліни «Основи нейронних мереж»;

– тема «Штучні нейронні мережі. Застосування нейронних мереж» - у межах дисципліни «Методи побудови штучного інтелекту».

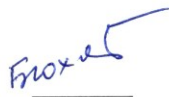
4. Концептуальна модель наскрізної ідентифікації об'єктів у відеоданих без використання ручного маркування на основі сегментованих треклетів, що поєднує маско-зважені ознаки, нейромережеве формування компактних ембедингів та механізми асоціації треклетів без ручного маркування.

– тема «Самонавчання нейронних мереж» - у межах дисципліни «Основи нейронних мереж»;

– тема «Методи та алгоритми навчання штучних нейронних мереж» - у межах дисципліни «Основи нейронних мереж»;

5. Окремі результати дисертаційної роботи Швандта М.А. використовувались при виконанні кваліфікаційних робіт здобувачів спеціальності 124 «Системний аналіз та наука про дані», зокрема оригінальні матеріали щодо розроблених моделей і методів комп'ютерного зору, які є складовими спеціалізованого конвеєра детектування, трекінгу та наскрізної ідентифікації об'єктів у відеоданих.

Заступник завідувача кафедри
ПМІТ
з наукової роботи, к. е. н.,
доцент



Біруте ЮХИМЕНКО

Директор
Інституту комп'ютерних
систем,
д.т.н., професор



Світлана АНТОЩУК