

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 неймережевого детектора методом суперпиксельної сегментації для виявлення та локалізації об'єктів, а також неймережевими 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 нестаціонарний фон, артефакти та часткові оклюзії. Концепція запропонованого методу полягає в тому, щоб після формування різницевого зображення не виконувати грубе глобальне пороговування, 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 суперпікселі 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 порогування в межах сегментів (суперпікселів) і відсікання артефактів. Показано, що просте перетворення різницевого зображення у бінарне та фільтрація плям за розміром ϵ

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 нейромережевої 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 A працює з базовими каналами, магістраль B – 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, нейромережеві 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 «ХП». 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→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).