



ЗАСТОСУВАННЯ КОМБІНАЦІЇ СІМЕЙСТВА YOLOV12 ТА СЕГМЕНТАЦІЙНИХ МОДЕЛЕЙ ДЛЯ АВТОМАТИЗОВАНОГО ВИЗНАЧЕННЯ ВОДНИХ ВИДІВ БІОРЕСУРСІВ

Досліджено особливості застосування сучасних методів глибокого навчання для автоматизованого виявлення, розпізнавання та сегментації водних видів біоресурсів. Проаналізовано сучасні архітектури нейронних мереж для завдань комп'ютерного зору та оброблення зображень. Застосовано комбінацію моделей комп'ютерного зору YOLOv12 (англ. *You Only Look Once* 12) для високошвидкісної детекції об'єктів та Mask R-CNN (англ. *Mask Region-based Convolutional Neural Network*) для точної сегментації водних видів біоресурсів, інтегруючи обидві архітектури в єдиний конвеєр для підвищення їх точності та стійкості. Розроблено конвеєр дій щодо оброблення зображень, що містить етапи попереднього оброблення зображень, детекції об'єктів, сегментації водної біомаси та подальшого аналізу об'єктів. Проведено попереднє оброблення даних, що містило нормалізацію кольору, зменшення шуму та геометричні аугментації, такі як обертання, масштабування та випадкове обрізання, щоб покращити моделі для виявлення риби. Проведено навчання та тестування моделей на великих анотованих наборах зображень. Наведено результати оцінювання продуктивності моделей за допомогою метрик precision, recall, F1-score, mean average precision (mAP) та intersection-over-union (IoU). Встановлено, що інтегрований метод оброблення зображень з використанням комбінованої моделі перевищує продуктивність окремих моделей детекції об'єктів або сегментації водної біомаси. З'ясовано, що використання аугментацій та попереднього оброблення зображень істотно підвищує узагальнювальну здатність моделей та зменшує кількість помилкових спрацьовувань. Введено метод визначення морфометричних показників на основі сегментованих об'єктів. Встановлено, що отримані морфометричні показники дають можливість надійно оцінювати довжину, ширину та площу риб. Проведено результати оцінювання впливу розміру навчальної бази водної біомаси, якості даних та стратегій аугментації на точність роботи моделей, показано, що ретельно спроектовані конвеєри даних істотно покращують стабільність детекції об'єктів та точність сегментації водної біомаси. Розроблено програмний модуль автоматизованого аналізу зображень, який забезпечує визначення розмірних характеристик об'єктів та можливість подальшого використання результатів для оцінювання водної біомаси. Наведено результати експериментальної перевірки комбінованої моделі аналізу зображення, які підтверджують ефективність запропонованого підходу та можливість його використання в автоматизованих системах моніторингу та обліку рибних ресурсів.

Ключові слова: подивитися можна тільки раз; конволюційна нейронна мережа на основі маскових областей; автоматичне розпізнавання об'єктів; глибоке навчання моделі; нейронні мережі; детекція видів риб.

Вступ / Introduction

Рациональне використання та збереження водних видів біоресурсів є одним із ключових завдань сучасного природокористування. У контексті зростання антропогенного навантаження на водні екосистеми, інтенсивного промислового вилову риби, зміни клімату та деградації середовища існування виникає потреба у створенні ефективних інструментів моніторингу та обліку рибних ресурсів. Традиційні методи оцінювання стану іхтіофауни, що базуються на польових спостереженнях, тралових обловах або ручному аналізі фото та відео, є трудомісткими, потребують значних часових та фінансових витрат, а також залежать від суб'єктивного фактору дослідника. З огляду на це, особливої актуальності набуває застосування методів комп'ютерного зору та штучного інтелекту для автоматизованого аналізу водних

біоресурсів.

Сучасний розвиток глибокого навчання (англ. *Deep Learning*) моделей значно розширив можливості автоматичного розпізнавання об'єктів на зображеннях і відео. Особливу роль у цьому відіграють згорткові нейронні мережі CNN (англ. *Convolutional Neural Networks*), які демонструють високу точність у завданнях класифікації, детекції об'єктів та сегментації водної біомаси. У сфері комп'ютерного зору для завдань локалізації об'єктів значного поширення набули алгоритми сімейства YOLO (англ. *You Only Look Once*), які забезпечують високу швидкість роботи та придатні для оброблення відеопотоку зображень в режимі реального часу. Водночас, для завдань точного виділення контурів об'єктів на зображенні активно використовують моделі інстанс-сегментації водної біомаси, зокрема – архітектури типу Mask R-CNN та їх сучасні модифікації.

© Copyright 2026 by the author(s)

Служала Олег Тарасович, аспірант, кафедра комп'ютерних наук.

Email: sluzhala.oleh@nltu.lviv.ua; <https://orcid.org/0009-0005-3655-5221>

Овсяк Олександр Володимирович, д-р техн. наук, професор, кафедра комп'ютерних наук.

Email: ovsjak@nltu.edu.ua; <https://orcid.org/0000-0003-2620-1938>

Цитування за ДСТУ: Служала О. Т., Овсяк О. В. Застосування комбінації сімейства YOLOV12 та сегментаційних моделей для автоматизованого визначення водних видів біоресурсів. Науковий вісник НЛТУ України. 2026, т. 36, № 3. С. 132–138.

Citation APA: Sluzhala, O. T., & Ovsyak, O. V. (2026). Application of a combination of YOLOV12 family and segmentation models for the automated identification of aquatic species. *Scientific Bulletin of UNFU*, 36(3), 132–138. <https://doi.org/10.36930/40360314>

Застосування алгоритмів глибинного навчання моделей у галузі виявлення водних видів біоресурсів особливо актуальне для контролю вилову та вилову згідно зі законодавством. Однак, складність природного середовища – нерівномірне освітлення, шум зображення, розмиття внаслідок руху, зміна масштабу та ракурсу знімання, часткове перекриття об'єктів, складний фон – значно ускладнює завдання точного їх розпізнавання. За таких умов моделі тільки детекції об'єктів не завжди забезпечують достатню точність, оскільки прямокутні bounding box-області часто містять значну частину фону або перекриваються між собою.

Одним із перспективних напрямів підвищення точності розпізнавання водних видів біоресурсів є комбінування моделей детекції об'єктів із моделями інстанс-сегментації водної біомаси. У такому методі швидка модель детекції об'єктів (наприклад, YOLO) використовує для попереднього визначення потенційних областей розташування риб, після чого сегментаційна модель (наприклад, сучасні модифікації Mask R-CNN або інші архітектури instance segmentation) уточнює контури об'єкта та дає змогу точніше виділити його форму. Такий гібридний метод дає змогу одночасно досягати високої швидкості роботи системи та високої точності локалізації об'єктів.

Попри значну кількість досліджень у сфері застосування глибинного навчання моделей для аналізу зображень, питання ефективного поєднання швидких моделей детекції об'єктів із сучасними моделями сегментації водної біомаси для завдань автоматизованого обліку водних видів біоресурсів досі залишається недостатньо дослідженим. У більшості наявних робіт використовують або тільки моделі детекції об'єктів, або тільки сегментаційні архітектури нейронних мереж, що обмежує можливості системи. Окрім цього, багато досліджень базуються на застарілих версіях алгоритмів, тоді як нові покоління моделей комп'ютерного зору демонструють значно вищу ефективність розпізнавання образів.

Отже, актуальність цього дослідження полягає у необхідності розроблення та дослідження комбінованого методу для автоматичного розпізнавання риб на зображенні, який поєднає сучасні моделі детекції об'єктів та інстанс-сегментації водної біомаси, що дає змогу підвищити точність і надійність ідентифікації водних видів біоресурсів.

Об'єкт дослідження – автоматизоване розпізнавання риб на зображенні для визначення та обліку водних видів біоресурсів.

Предмет дослідження – методи та алгоритми комп'ютерного зору для автоматизованого визначення водних видів біоресурсів, які забезпечать підвищення точності визначення контурів об'єктів.

Мета роботи – розробити комбіновану модель комп'ютерного зору для автоматичного виявлення та сегментації риб на зображенні, що дасть змогу поєднати сучасні алгоритми детекції об'єктів та інстанс-сегментації водної біомаси.

Для досягнення зазначеної мети визначено такі основні завдання дослідження:

- проаналізувати сучасні методи комп'ютерного зору для завдань детекції об'єктів та сегментації водної біомаси, що дасть змогу визначити переваги, недоліки та області ефективного застосування наявних підходів;
- дослідити архітектурні особливості сучасних моделей детекції об'єктів сімейства YOLO та моделей інстанс-

сегментації водної біомаси, що дасть змогу обґрунтувати вибір моделей для побудови системи автоматизованого розпізнавання та сегментації водних видів біоресурсів;

- розробити комбінований алгоритм розпізнавання риб, що поєднає детекцію об'єктів і сегментацію, що забезпечить підвищення точності визначення контурів об'єктів;
- реалізувати програмну модель системи аналізу зображень для автоматичного виявлення водних видів біоресурсів, що дасть змогу автоматизувати процес оброблення зображень;

- провести експериментальні дослідження ефективності запропонованого методу, що дасть змогу оцінити точність детекції об'єктів та сегментації водної біомаси;
- виконати порівняльний аналіз результатів роботи системи з використанням різних архітектур нейронних мереж, що дасть змогу визначити найефективніші моделі та параметри їх навчання для завдань розпізнавання риб.

Аналіз останніх досліджень та публікацій. У роботі [5] досліджено основи глибинного навчання моделей та показано, що багатопланові нейронні мережі здатні формувати ієрархічні наведення даних, що дає змогу автоматично виділяти інформативні ознаки зображень без використання ручного проєктування дескрипторів. Автори зазначають, що використання глибинних нейронних мереж значно підвищує ефективність завдань комп'ютерного зору, зокрема – класифікації, детекції об'єктів та сегментації водної біомаси.

У дослідженні [17] проведено аналіз розвитку методів детекції об'єктів за останні двадцять років та показано, що сучасні алгоритми детекції об'єктів базуються переважно на згорткових нейронних мережах. Автори виявили, що найефективнішими є одноетапні та двоетапні методи детекції об'єктів, які забезпечують високу точність локалізації об'єктів на зображенні.

У роботі [12] запропоновано алгоритм YOLO, у якому завдання детекції об'єктів розглядають як єдине завдання регресії координат обмежувальних прямокутників та ймовірностей належності об'єкта до певного класу. Автори показали, що такий підхід дає змогу виконувати детекцію об'єктів у реальному часі, що є важливим для аналізу відеопотоку.

У роботі [14] запропоновано алгоритм Faster R-CNN, який використовує мережу Region Proposal Network для генерування областей-кандидатів. У дослідженні показано, що двоетапні методи детекції об'єктів забезпечують високу точність виявлення об'єктів, однак потребують більших обчислювальних ресурсів порівняно з одноетапними методами.

У дослідженні [3] розроблено архітектуру Mask R-CNN, яка дає змогу виконувати інстанс-сегментацію об'єктів шляхом додавання окремої гілки сегментації зображень до мережі Faster R-CNN. Автори показали, що використання інстанс-сегментації зображень дає змогу отримувати точні контури об'єктів, що є важливим для визначення морфометричних характеристик об'єктів.

У дослідженні [8] автори проаналізували застосування повністю згорткових мереж FCN (англ. *Fully Convolutional Networks*) для семантичної сегментації зображень та дослідили, що FCN дають можливість виконувати сегментацію піксель за пікселем, що істотно підвищує точність визначення меж об'єктів у зображеннях з високою деталізацією.

У роботі [9] автори провели аналіз методів сегментації зображень із застосуванням глибинного навчання моделей і встановили, що сучасні моделі, включно з U-Net та її модифікаціями, дають можливість досягти ви-

сокої точності сегментації зображень навіть за обмеженої кількості навчальних даних та складних умов зображень.

У роботі [2] автори дослідили теоретичні основи глибинного навчання моделей та підкреслили значення правильного налаштування архітектури нейронних мереж і параметрів навчання для досягнення оптимальної продуктивності моделей у завданнях класифікації та розпізнавання об'єктів.

Проведений аналіз наукових джерел показав, що наявні методи комп'ютерного зору забезпечують високу точність детекції об'єктів та сегментації водної біомаси, однак більшість досліджень розглядають ці завдання окремо. Недостатньо дослідженим залишається питання поєднання методів детекції об'єктів та інстанс-сегментації водної біомаси для одночасного розпізнавання риб та визначення їх морфометричних характеристик. Отже, актуальним є розроблення комбінованого методу, що поєднає сучасні алгоритми детекції об'єктів та сегментації зображень для підвищення точності автоматизованого аналізу рибних ресурсів.

Матеріали та методи дослідження. У дослідженні використано методи комп'ютерного зору та глибинного навчання моделей для автоматичного виявлення та сегментації риб на зображеннях. Методологічною основою роботи є комбінований метод комп'ютерного зору, який поєднує алгоритм детекції об'єктів на основі згорткових нейронних мереж із моделлю інстанс-сегментації водної біомаси. Така архітектура методу дає змогу поєднати високу швидкість оброблення зображення зі значною точністю визначення контурів об'єктів.

Джерела похибок та обмеження дослідження. Під час аналізу зображень існує низка факторів, які можуть впливати на точність роботи системи. До основних джерел похибок належать: нерівномірне освітлення, шум зображення, розмиття внаслідок руху, зміна масштабу та ракурсу знімання, часткове перекриття об'єктів, складний фон. Додатковою проблемою є значна варіативність форм і розмірів риб, що може ускладнювати їх розпізнавання водних видів біоресурсів.

Ще одним потенційним джерелом похибок є неточності у процесі анотації навчальних даних. Оскільки розмітку зображень виконують вручну, можливі помилки у визначенні меж об'єктів. Для зменшення впливу цього фактора використовувалася перевірка анотацій декількома експертами.

Окрім цього, на результати дослідження може впливати обмеженість навчального набору даних. Для підвищення узагальнювальної здатності моделі використовувалися методи аугментації даних та крос-валідації.

Застосування зазначених методів дає змогу мінімізувати вплив основних джерел похибок та забезпечити відтворюваність результатів дослідження. Описані матеріали та методи дають можливість повністю повторити експериментальні дослідження та перевірити отримані результати.

Результати дослідження та їх обговорення / Research results and their discussion

У межах дослідження проведено експериментальну перевірку ефективності запропонованого методу до автоматичного виявлення та сегментації риб на фото. Основною метою експериментальної частини роботи є визначення точності детекції об'єктів, якості сегментації контурів риб та оцінювання продуктивності комбінованої моделі, що поєднає алгоритм детекції об'єктів на основі нейронної мережі YOLOv12 та сегментаційну модель Mask R-CNN. Результати дослідження дають змогу оцінити ефективність запропонованого методу та порівняти його з альтернативними методами, що використовують у завданнях аналізу зображень.

Архітектура запропонованої системи комп'ютерного зору базується на каскадному поєднанні двох нейронних мереж, коли вони працюють одна за одною, як на конвеєрі (рис. 1). Перша мережа виконує детекцію об'єктів на зображенні та визначає координати прямокутних областей, що містять потенційні об'єкти. Друга мережа здійснює сегментацію цих областей та формує бінарну маску об'єкта.

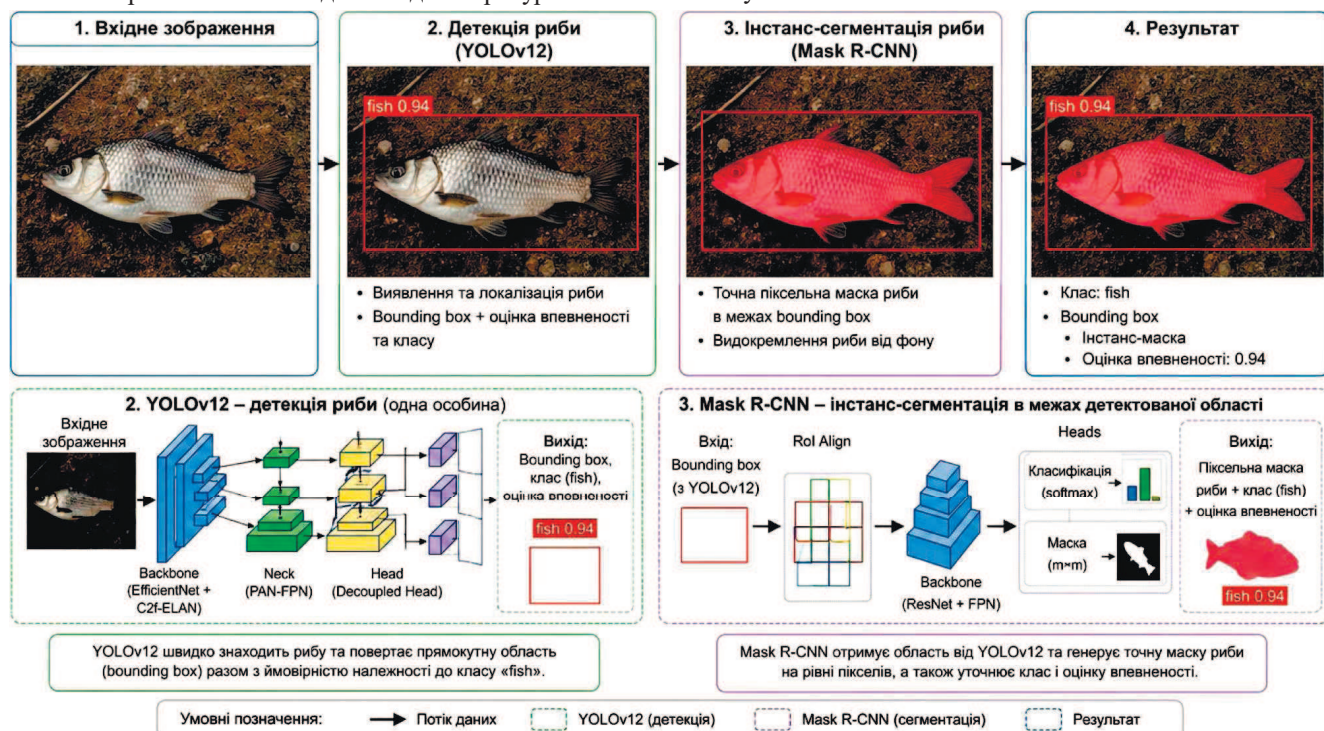


Рис. 1. Архітектура комбінованої моделі комп'ютерного зору / Architecture of the combined model of computer vision

На першому етапі використовується YOLOv12, який виконує швидке виявлення риби на вхідному зображенні. Модель аналізує сцену в один прохід та формує прямокутну область інтересу (bounding box), що локалізує об'єкт, разом із оцінкою впевненості та ймовірним класом. Завдяки високій швидкодії та ефективній архітектурі, цей етап забезпечує первинну фільтрацію області зображення, значно зменшуючи обчислювальне навантаження для наступних процедур оброблення.

На другому етапі для уточнення просторових характеристик об'єкта застосовується Mask R-CNN, який отримує на вхід область, визначену попереднім детектором. Використовуючи механізм вирівнювання ознак (RoI Align) та багаторівневу згорткову архітектуру, модель виконує піксельну класифікацію всередині знайденої області, генеруючи точну інстанс-маску риби. Це дає змогу не тільки відокремити об'єкт від фону, але й коректно обробляти складні контури та часткові перекриття, що є критично важливим для аналізу біоресурсів у реальних умовах зйомки.

Комбінування детекції об'єктів та їх інстанс-сегментації в єдиному конвеєрі забезпечує баланс між швидкістю та точністю оброблення зображень. Використання YOLOv12 для попередньої локалізації дає змогу обмежити область пошуку, тоді як Mask R-CNN забезпечує високоточне виділення об'єкта на рівні пікселів. Формально цей процес можна описати системою рівнянь.

На першому етапі виконується детекція об'єктів на вхідному зображенні I_k із використанням моделі детекції об'єктів, а саме:

$$R_k = D(I_k), k = \overline{1, K},$$

де: R_k – обмежувальна рамка (англ. *Bounding Box*), що відповідає знайденому k -му об'єкту; $D(\cdot)$ – функція детекції об'єктів, реалізована на основі моделі YOLO; K – кількість виявлених об'єктів на зображенні.

На другому етапі для кожної знайденої області виконується сегментація, що дає змогу виділити точний контур об'єкта. Для цього із вхідного зображення вирізають фрагмент, обмежений рамкою, після чого застосовують сегментаційну модель:

$$M_k = M(I_{R_k}), k = \overline{1, K},$$

де: M_k – бінарна маска k -го об'єкта, яка визначає належність кожного пікселя до об'єкта або фону; I_{R_k} – підзображення k -го об'єкта, що відповідає області R_k ; $M(\cdot)$ – функція сегментації водної біомаси, реалізована на основі моделі Mask R-CNN.

Отже, загальний процес оброблення зображення є композицією двох функціональних блоків детекції об'єктів, сегментації та аналізу водної біомаси, що забезпечує підвищення точності визначення об'єктів порівняно з використанням кожного з методів окремо. Використання моделі детекції об'єктів дає змогу локалізувати потенційні об'єкти інтересу, тоді як сегментаційна модель уточнює їх просторову структуру. Запропонований метод забезпечує ефективне оброблення зображень за умов шуму, змінного освітлення та складного фону.

Формування експериментального набору даних.

Для проведення експериментів сформовано набір зображень, отриманих із мережі Інтернет. Фотоматеріали містили зображення різних видів риб.

Загальний обсяг сформованого набору даних становив понад кілька сотень зображень. Кожне зображення

анотоване вручну з використанням спеціалізованого програмного забезпечення для розмітки даних. Для кожного об'єкта визначали координати прямокутної області (англ. *Bounding Box*), що охоплює рибу, а також створювали піксельну маску, яка описує точну форму об'єкта. Отриманий набір даних розділено на навчальну, валідаційну та тестову підмножини.

Результати детекції об'єктів. На першому етапі експерименту проведено навчання моделі YOLOv12 для завдання детекції видів риб на зображеннях. Після завершення процесу навчання модель була протестована на незалежному тестовому наборі даних в розмірі 300 зображень (табл. 1). Оцінювання якості детекції об'єктів здійснювалося за допомогою метрик precision, recall та mean Average Precision. Отримані результати показали, що модель YOLOv12 здатна ефективно визначати місцезнаходження риб на зображеннях навіть за складних умов.

Табл. 1. Точність детекції видів риб різними моделями комп'ютерного зору / Accuracy of fish detection by different models of computer vision

Модель	Precision	Recall	mAP@0,5	mAP@0,5:0,95	Середня тривалість оброблення кадру, мс
Faster R-CNN	0,87	0,83	0,86	0,61	85
YOLOv7	0,91	0,88	0,90	0,66	18
YOLOv8	0,93	0,90	0,92	0,69	16
YOLOv12	0,95	0,92	0,94	0,72	14

Середнє значення точності детекції об'єктів (англ. *Precision*) визначається як

$$\text{Precision} = \frac{TP}{TP + FP},$$

де: TP – кількість правильно визначених об'єктів; FP – кількість хибнопозитивних результатів. Повнота (англ. *Recall*) визначається як

$$\text{Recall} = \frac{TP}{TP + FN},$$

де FN – кількість пропущених об'єктів.

За результатами тестування модель продемонструвала високі значення цих показників. Значення precision перевищувало 0,9 (див. табл. 1), що свідчить про незначну кількість хибнопозитивних результатів. Показник recall також був високим, що підтверджує здатність моделі знаходити більшість об'єктів на зображенні.

Для комплексного оцінювання ефективності детекції об'єктів використано метрику mean Average Precision, яка визначається як середнє значення площі під кривою precision – recall. Високе значення цієї метрики свідчить про ефективність використання архітектури YOLOv12 для завдання детекції видів риб.

Результати сегментації водної біомаси. Другим етапом дослідження була перевірка ефективності моделі Mask R-CNN для завдання сегментації риб. Сегментаційна модель отримувала на вхід області зображення, які були визначені моделлю детекції об'єктів (табл. 2).

Унаслідок оброблення зображення для кожної його області модель формує бінарну маску, яка описує форму об'єкта. Цей процес у комп'ютерному зорі називається сегментацією зображення (зокрема, пооб'єктною або семантичною сегментацією). Якість сегментації водної біомаси оцінювалася за допомогою метрики *IoU* (англ. *Intersection over Union*), яка визначає ступінь перекриття між передбаченою маскою та еталонною маскою.

Табл. 2. Порівняння якості сегментації риб різними архітектурами нейронних мереж / Comparison of the quality of fish segmentation by different architectures of neural networks

Модель сегментації водної біомаси	IoU	Dice coefficient	Pixel accuracy	Середня тривалість оброблення області зображення, мс
U-Net	0,79	0,84	0,90	21
DeepLabV3+	0,81	0,86	0,91	24
Mask R-CNN	0,84	0,88	0,93	28

Метрика IoU визначається як

$$IoU = \frac{Area(M_{pred} \cap M_{gt})}{Area(M_{pred} \cup M_{gt})},$$

де: M_{pred} – передбачена маска об'єкта; M_{gt} – еталонна маска.

Результати експериментів показали, що модель Mask R-CNN здатна точно визначати контури риб навіть у випадках часткового перекриття об'єктів або складного фону. Середнє значення IoU перевищувало 0,8 (див. табл. 2), що свідчить про високу точність сегментації водної біомаси.

Результати роботи комбінованої моделі. Після окремого тестування моделей детекції об'єктів та сегментації водної біомаси проведено експеримент із використанням їхньої комбінованої архітектури цих моделей, у якій модель YOLOv12 використовують для попереднього виявлення об'єктів, а модель Mask R-CNN – для уточнення їх контурів (табл. 3).

Табл. 3.. Порівняння методів аналізу зображень риб / Comparison of methods for the analysis of fish images

Метод	Precision	Recall	IoU	FPS (кадр/с)
Детекція (YOLOv12)	0,95	0,92	–	72
Сегментація (Mask R-CNN)	0,90	0,88	0,84	24
Комбінована модель (YOLOv12 + Mask R-CNN)	0,96	0,94	0,87	38

Результати показали, що використання каскадної структури дає змогу підвищити точність визначення форми об'єктів та зменшити кількість помилкових сегментацій. Оскільки сегментаційна модель обробляє тільки обмежену кількість областей, які попередньо були визначені моделлю детекції об'єктів, обчислювальне навантаження системи залишається відносно невеликим.

Продуктивність системи оцінювалася за швидкістю оброблення зображень. Експериментальні результати показали, що запропонований метод дає змогу обробляти зображення з частотою декількох десятків кадрів за секунду за використання сучасного графічного процесора. Це свідчить про можливість застосування запропонованої системи для аналізу зображень в режимі реального часу.

Аналіз результатів дослідження. Візуальний аналіз результатів роботи моделі підтвердив ефективність запропонованого методу. На більшості тестових зображень система коректно визначала місцезнаходження риб та формувала точні сегментаційні маски (рис. 2). У випадках складного фону або часткового перекриття об'єктів, використання сегментаційної моделі давало змогу значно покращити якість виділення об'єкта порівняно з використанням тільки моделі детекції об'єктів.

Отримані результати дослідження демонструють, що поєднання швидких алгоритмів детекції об'єктів з точними методами сегментації водної біомаси дає змогу створювати ефективні системи автоматичного аналізу

зображень. Запропонований комбінований метод комп'ютерного зору забезпечує високу точність визначення об'єктів та може використовуватися для подальших завдань аналізу водних видів біоресурсів, зокрема – контролю вилову відповідно до чинного законодавства.

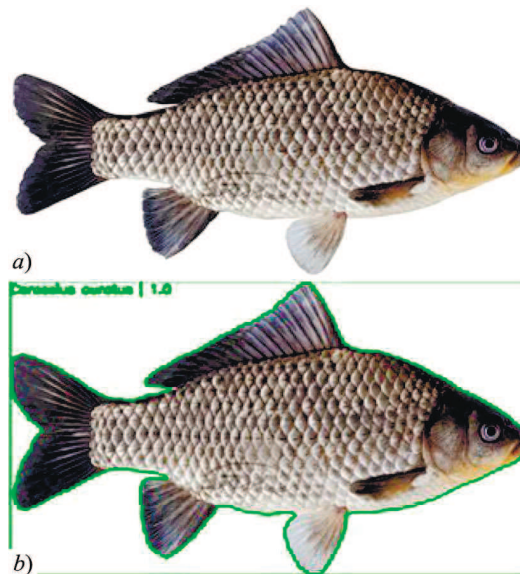


Рис. 2. Приклад проаналізованого зображення за допомогою комбінованого методу (до та після) / An example of an image analysed using a combined method (before and after)

Отримані результати дослідження свідчать про ефективність використання сучасних методів комп'ютерного зору для автоматизованого аналізу водних видів біоресурсів. Поєднання алгоритмів глибокого навчання з класичними методами виділення ознак дає змогу підвищити точність і стабільність розпізнавання водних об'єктів на зображенні. Використана модель YOLOv12 продемонструвала високу швидкість оброблення кадрів та достатню точність виявлення об'єктів навіть за умов складного фону та різної освітленості.

Вибір моделі YOLOv12 зумовлений її архітектурними перевагами. Ця модель належить до сучасного покоління одноетапних детекторів об'єктів, які виконують їх локалізацію та класифікацію в межах одного проходу нейронної мережі. Така модель забезпечує високу швидкість оброблення зображень у реальному часі. Окрім цього, моделі характерна покращена архітектура backbone та head-частини мережі, що підвищує точність виявлення малих і частково перекритих об'єктів.

Для підвищення точності визначення контурів риб застосовано сегментаційну модифікацію моделі YOLOv12. На відміну від класичних детекторів, які визначають тільки обмежувальні прямокутники об'єктів, сегментаційна модель формує піксельну маску кожного об'єкта. Це дає змогу точніше визначати форму риби, відокремлювати її від фону та виконувати подальший аналіз морфологічних характеристик. Така модель є особливо важливою для завдань біологічної ідентифікації видів риб, оскільки багато видів риб мають подібні пропорції тіла, але відрізняються деталями форми плавців або контурів тіла.

Додатково у роботі застосовано методи виділення локальних ознак – ORB (англ. *Oriented FAST and Rotated BRIEF*), SIFT (англ. *Scale-Invariant Feature Transform*) та SURF (англ. *Speeded-Up Robust Features*). Використання цих методів дає змогу аналізувати ключові точки зображення та формувати дескриптори, які ха-

рактизують локальні особливості об'єкта. У поєднанні з результатами сегментації водної біомаси це забезпечує надійніше розпізнавання видів риб навіть за умов часткового перекриття або змін положення об'єкта у кадрі.

Аналіз результатів дослідження методів аналізу зображення показав, що комбінований метод комп'ютерного зору забезпечує вищу точність ідентифікації риби на зображенні порівняно з використанням тільки одного методу. Сегментація об'єктів на зображенні дає змогу точно виділити область об'єкта, а алгоритми локальних ознак виконують детальніший аналіз внутрішньої структури зображення. Отже, формується багаторівнева система аналізу зображення, яка поєднує глобальні та локальні характеристики об'єкта.

Водночас, проведене дослідження комбінованої моделі аналізу зображення має певні обмеження. Точність роботи моделі значною мірою залежить від якості навчального набору даних та різноманітності зображень у ньому. У разі недостатньої кількості прикладів певного виду риб модель може демонструвати нижчу точність класифікації водних об'єктів. Додатковим фактором, що впливає на результати їх класифікації, є якість зображення, наявність шуму та зміни освітлення.

Перспективним напрямом подальших етапів дослідження є розширення навчального набору даних, використання складніших ансамблевих моделей та інтеграція методів відстеження водних об'єктів на фото у системах контролю вилову риб.

Обговорення результатів дослідження. У роботі [6] автори провели аналіз особливостей застосування методів детекції об'єктів на основі глибинного навчання моделей в аквакультури та проаналізували сучасні підходи до автоматизації моніторингу видів риб. Автори встановили, що використання згорткових нейронних мереж дає змогу ефективно вирішувати завдання підрахунку, класифікації та відстеження водних об'єктів, забезпечуючи високу точність і швидкість оброблення даних.

У дослідженні [7] автори розробили підхід до детекції видів риб за складних підводних умов із використанням глибинного навчання моделей та дослідили вплив шумів, освітлення та фону на якість розпізнавання водних видів біоресурсів. Автори встановили, що застосування нейронних мереж дає змогу досягти високої точності детекції об'єктів навіть за умов низької якості зображень і складного середовища.

У роботі [10] автори дослідили використання алгоритмів сімейства YOLO для виявлення риб у підводному середовищі та проаналізували ефективність цих моделей у завданнях реального часу. Автори показали, що YOLO-засновані підходи забезпечують високу швидкість оброблення зображень зі збереженням достатнього рівня точності детекції об'єктів.

У роботі [16] автори дослідили застосування глибинного навчання моделей для виявлення риб у завданнях, пов'язаних із гідроенергетикою, та проаналізували можливості використання нейронних мереж для автоматизованого моніторингу водних ресурсів. Автори зазначили, що застосування глибинних моделей дає змогу підвищити ефективність аналізу відеоданих за реальних умов експлуатації.

У дослідженні [4] автори розробили підхід до виявлення та класифікації риб у помірних кліматичних умовах із використанням глибинного навчання моделей та проаналізували ефективність моделей для різних типів

зображень. Автори встановили, що використання нейронних мереж дає змогу досягти високої точності класифікації навіть за значної варіативності об'єктів.

У роботі [15] автори дослідили використання методів самонавчання (self-supervised learning) для підрахунку риб на зображеннях та проаналізували ефективність такого підходу для роботи з великими обсягами неанотованих даних. Автори показали, що використання самонавчання дає змогу зменшити залежність від розмічених даних і підвищити масштабованість систем аналізу зображень.

Унаслідок обговорення результатів дослідження встановлено, що сучасні методи глибинного навчання моделей забезпечують високий рівень точності у завданнях детекції об'єктів та сегментації водної біомаси, а також демонструють ефективність під час оброблення складних зображень. Водночас, з'ясовано, що у наявних дослідженнях переважно розглядають окремі аспекти завдання, зокрема – детекція, сегментація, або класифікація об'єктів, тоді як питання їх комплексного поєднання в єдину систему досліджено недостатньо. Окрім цього, встановлено, що наявні рішення не повною мірою враховують можливості інтеграції різних архітектур нейронних мереж у межах єдиного конвеєра оброблення даних. Тому, результати проведеного аналізу методів аналізу зображення підтверджують доцільність виконання цього дослідження, спрямованого на розроблення комбінованого методу, який відрізняється від наявних тіснішою інтеграцією моделей детекції об'єктів та сегментації водної біомаси і забезпечує підвищення ефективності автоматизованого аналізу зображень.

Отже, внаслідок виконаної роботи можна сформулювати такі наукову новизну та практичну значущість результатів дослідження.

Наукова новизна отриманих результатів дослідження – отримав подальший розвиток метод автоматичного розпізнавання риб на зображенні, який, на відміну від наявних методів, поєднує сучасні моделі детекції об'єктів та інстанс-сегментації водної біомаси, що дає змогу підвищити точність і надійність ідентифікації водних видів біоресурсів.

Практична значущість результатів дослідження – запропонований метод можна застосувати для автоматизації процесів моніторингу водних видів біоресурсів, контролю промислового вилову різних видів риб, що поєднує моделі детекції об'єктів та сегментації водної біомаси, а також використати під час створення інтелектуальних систем аналізу зображень та розроблення автоматизованих систем обліку рибних ресурсів.

Висновки / Conclusion

Розроблено комбіновану модель комп'ютерного зору для автоматичного виявлення та сегментації риб на зображенні, що дало змогу поєднати сучасні алгоритми детекції об'єктів та інстанс-сегментації водної біомаси. За результатами проведеного дослідження можна зробити такі основні висновки.

1. Проаналізовано сучасні методи комп'ютерного зору для завдань детекції об'єктів та сегментації водної біомаси. Встановлено, що підходи, засновані винятково на детекції або сегментації даних, мають обмеження щодо точності або швидкодії, тоді як їх комбінування є перспективним напрямом підвищення ефективності аналізу зображень.

2. Досліджено архітектурні особливості моделей сімейства YOLO та моделей інстанс-сегментації. Обґрунтовано доцільність використання швидких детекторів для первинної локалізації об'єктів і сегментаційних моделей задля уточнення їхніх контурів, що стало основою вибору архітектури системи.
3. Розроблено комбінований метод розпізнавання риб, який поєднує детекцію об'єктів та сегментацію. Запропонований підхід забезпечує підвищення точності визначення контурів об'єктів за рахунок послідовного використання двох типів моделей.
4. Реалізовано програмну модель системи аналізу зображень для автоматичного виявлення водних видів біоресурсів. Розроблена система забезпечує автоматизацію процесу оброблення зображень та може бути використана в завданнях моніторингу водної біомаси.
5. Проведено етапи експериментального дослідження ефективності запропонованого методу. Отримані результати підтвердили високу точність детекції об'єктів і сегментації водної біомаси, а також стійкість алгоритму до змін умов знімання водних об'єктів.
6. Виконано порівняльний аналіз різних архітектур нейронних мереж. Визначено, що комбінування моделей детекції та сегментації забезпечує кращі результати порівняно з використанням окремих підходів, а також дає можливість визначити допустимі параметри навчання для завдань розпізнавання риб.

References

1. Cui, S., Zhou, Y., Wang, Y., & Zhai, L. (2020). Fish detection using deep learning. *Applied Computational Intelligence and Soft Computing*, article ID 3738108, 1–10. <https://doi.org/10.1155/2020/3738108>
2. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT Press. URL: <https://www.deeplearningbook.org/>
3. He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969. <https://doi.org/10.1109/ICCV.2017.322>
4. Knausgård, K. M., Wiklund, A., Sørdaalen, T. K., Halvorsen, K., Kleiven, A. R., Jiao, L., & Goodwin, M. (2020). Temperate fish detection and classification: A deep learning approach. *arXiv preprint*, 1–12. <https://doi.org/10.48550/arXiv.2005.07518>
5. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
6. Liu, H., Ma, X., Yu, Y., Wang, L., & Hao, L. (2023). Application of deep learning-based object detection techniques in fish aquaculture: A review. *Journal of Marine Science and Engineering*, 11(4), 1–23. <https://doi.org/10.3390/jmse11040867>
7. Liu, M., Jiang, W., Hou, M., Qi, Z., Li, R., & Zhang, C. (2023). A deep learning approach for object detection of rockfish in challenging underwater environments. *Frontiers in Marine Science*, 10, 1–14. <https://doi.org/10.3389/fmars.2023.1242041>
8. Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440. <https://doi.org/10.1109/CVPR.2015.7298965>
9. Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtamavaz, N., & Terzopoulos, D. (2022). Image segmentation using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7), 3523–3542. <https://doi.org/10.1109/TPAMI.2021.3059968>
10. Ouis, M. Y., & Akhloufi, M. (2024). YOLO-based fish detection in underwater environments. *Environmental Sciences Proceedings*, 29(1), 1–8. <https://doi.org/10.3390/ECRS2023-16315>
11. Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. *arXiv preprint*, 1–6. <https://doi.org/10.48550/arXiv.1804.02767>
12. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788. <https://doi.org/10.1109/CVPR.2016.91>
13. Reis, D., Kupec, J., Hong, J., & Daoudi, A. (2023). Real-time flying object detection with YOLOv8. *arXiv preprint*, 1–9. <https://doi.org/10.48550/arXiv.2305.09972>
14. Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137–1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
15. Tarling, P., Cantor, M., Clapés, A., & Escalera, S. (2021). Deep learning with self-supervision to count fish in underwater images. *arXiv preprint*, 1–10. <https://doi.org/10.48550/arXiv.2104.14964>
16. Xu, W., & Matzner, S. (2018). Underwater fish detection using deep learning for water power applications. *arXiv preprint*, 1–9. <https://doi.org/10.48550/arXiv.1811.01494>
17. Zou, Z., Shi, Z., Guo, Y., & Ye, J. (2023). Object detection in 20 years: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3), 1–20. <https://doi.org/10.48550/arXiv.1905.05055>

O. T. Sluzhala, O. V. Ovsyak

Ukrainian National Forestry University, Lviv, Ukraine

APPLICATION OF A COMBINATION OF YOLOV12 FAMILY AND SEGMENTATION MODELS FOR THE AUTOMATED IDENTIFICATION OF AQUATIC SPECIES

This study investigates the application of modern deep learning approaches for automated detection, recognition, and segmentation of aquatic species. The research analyzes state-of-the-art convolutional neural network architectures for computer vision tasks involving object localization and instance segmentation. The proposed approach employs a combination of YOLOv12 for high-speed object detection and Mask R-CNN for precise instance segmentation and integrates both architectures into a unified processing pipeline to improve accuracy, robustness, and computational efficiency. Moreover, the approach introduces a structured workflow that includes data preprocessing, model training, inference, and post-processing stages. The method applies data preprocessing techniques such as color normalization, noise reduction, contrast enhancement, and geometric augmentations, for example rotation, scaling, and random cropping, to improve model generalization and detection performance. The models are trained and evaluated on large-scale annotated image datasets, and the evaluation uses precision, recall, F1-score, mean average precision (mAP), and intersection-over-union (IoU) metrics. The results show that the integrated approach outperforms standalone detection or segmentation models. Moreover, combining detection and segmentation significantly improves boundary delineation and reduces false positives. The analysis also evaluates the effects of training dataset size, annotation quality, and augmentation strategies on model generalization and shows that carefully designed data pipelines significantly improve detection stability and segmentation fidelity. The developed software module implements the full processing pipeline and enables automated extraction of quantitative characteristics from images. Consequently, the proposed methodology provides a scalable and reproducible solution for automated monitoring and supports real-time or near real-time analysis. The benefits of this study include reduced manual effort, improved data consistency, and the possibility of integration with monitoring and catch control systems. As a result, the methodology lays the foundation for further development of intelligent systems for automated fish detection, classification, and size estimation in practical applications.

Keywords: YOLO; Mask R CNN; automatic object recognition; deep learning; neural networks; species detection.