

УДК 004.3

Смолій Вікторія Миколаївна

доктор технічних наук, професор, професор кафедри інформаційних систем та технологій,
Національний університет біоресурсів і природокористування України

ORCID: <https://orcid.org/0000-0002-1268-7837>E-mail: vmsmolij@nubip.edu.ua**Смолій Натан Вікторович**

здобувач спеціальності «Інформаційні системи та технології», Національний технічний
університет України «Київський політехнічний інститут імені Ігоря Сікорського»

ORCID: <https://orcid.org/0009-0002-3763-6726>E-mail: hoibbitizukrainy@gmail.com**МОДУЛЬ МАШИННОГО ЗОРУ ДЛЯ ВИЯВЛЕННЯ ОБ'ЄКТІВ НА ЗОБРАЖЕННЯХ І
У ВІДЕОПОТОЦІ**

***Анотація.** Робота присвячена актуальній тематиці створення та застосування інтелектуальних методів комп'ютерного зору для автоматизованого виявлення й сегментації об'єктів захисного спорядження на зображеннях і у відеопотоці. Мета роботи полягає у визначенні об'єктів заданого типу на зображеннях, отриманих із відеокамери, а також розробці моделі детекції, яка зможе ефективно ідентифікувати та локалізувати ці об'єкти в різних умовах освітлення, масштабу та перспективи. В матеріалі статті досліджено процес підготовки даних, зокрема застосування методів аугментації для підвищення репрезентативності вибірки, а також виконано порівняльний аналіз параметрів, швидкодії та результатів роботи моделей SAM і YOLO. Наведено результати експериментальних досліджень, які свідчать про позитивний вплив збільшення обсягу та різноманітності датасету на збалансованість даних і узагальнювальну здатність моделей комп'ютерного зору. Запропоновані підходи окремого навчання моделей для задач сегментації та класифікації підтвердили свою ефективність у контексті автоматизованої обробки зображень. В роботі для реалізації завдання сегментації зображень було сформовано спеціалізований датасет для об'єктів захисного спорядження. На етапі підготовки до тренування моделі до цього набору даних було застосовано процедури аугментації, зокрема віддзеркалення, обертання, масштабування та зміну яскравості, що дозволило суттєво підвищити різноманітність навчальних прикладів. Збільшення розміру датасету забезпечило більш збалансоване представлення даних і підвищило узагальнювальну здатність моделі. Отримані в роботі результати підтверджують доцільність і ефективність запропонованого авторами для даної проблеми окремого навчання моделей для сегментації та класифікації зображень. До перспектив подальших розвідок слід віднести розширення датасету, оптимізацію обчислювальної складності моделей і дослідження їх застосування в режимі реального часу для відеоаналітики.*

***Ключові слова:** виявлення об'єктів, модель детекції, сегментація зображень, попереднє навчання моделі, процедури аугментації, донавчання, бінарні маски об'єктів.*

Вступ. Актуальність тематики зумовлена стрімким зростанням обсягів візуальних даних, що надходять із систем відеоспостереження, безпілотних платформ, автономних транспортних засобів та інтелектуальних інформаційних систем. Сучасні задачі аналізу зображень і відеопотоків потребують автоматизованих методів обробки, здатних працювати в режимі реального часу з високою точністю та надійністю. Методи глибокого навчання, зокрема нейронні мережі для детектування об'єктів, демонструють суттєву перевагу над класичними алгоритмами комп'ютерного зору в складних умовах освітлення, зашумлення та часткових перекриттів об'єктів. Практична значущість таких підходів полягає у їх широкому застосуванні в системах безпеки, моніторингу навколишнього середовища, медичній діагностиці та робототехніці. Отже, дослідження та вдосконалення методів виявлення об'єктів на основі глибокого навчання є важливим і своєчасним науково-практичним завданням.

Огляд літератури. Сучасний стан розвитку технологій штучного інтелекту значно розширює спектр завдань, які можуть бути вирішені за допомогою робототехніки [1], [2]. Найбільш необхідні напрямки автоматизації для галузі БПЛА включають такі завдання, як скорочення часу виконання місій, більш ефективно використання обчислювальних

потужностей та скорочення часу втручання людини в процеси управління БПЛА [3], [4], [5]. Патрулювання важкодоступних або небезпечних територій, пошук, ідентифікація та класифікація знайдених об'єктів на їх поверхні за допомогою традиційних методів – це складний процес з високими вимогами до часу та обчислювальних потужностей, який потребує вдосконалення завдяки впровадженню різних сучасних методів, таких як штучний інтелект [6], [7], [8].

Для сучасних безпілотних літальних апаратів велика різноманітність завдань, таких як: класифікація, спостереження, дослідження та класифікація, перенаправляється на певні автоматизовані підсистеми, що може підвищити продуктивність завдяки додатковим обчислювальним модулям, що залучаються. Організація та принципи роботи подібних систем наведені в [9], [10], [11]. Такий підхід є доцільним та поширеним завдяки наявності на ринку польотних контролерів [12], [13].

Доступні опції передбачають велику різноманітність роз'ємів для зовнішніх пристроїв для різних цілей: обмін даними GPS, фрагментами даних телеметрії та іншими типами повідомлень, визначених через MavLink. Такі дані можна кластеризувати в набори даних для різних цілей: пріоритетне черги завдань та процедури обробки даних [14], [15]. Системи комп'ютерного зору реалізують різні алгоритми машинного навчання та складні методи обробки даних для завдань обробки зображень та локалізації. Складність структури системи комп'ютерного зору може змінюватися залежно від типу завдання та набору використовуваних технологій [16], [17].

Завдання дослідження певних територій та виявлення й ідентифікації певних об'єктів залежить від взаємодії з нейронною мережею, детальна робота, архітектура та навчання якої обговорюються в [20]. Для цієї конкретної нейронної мережі необхідно розробити інструменти для створення бортового журналу польотів та створення звітів про виявлені об'єкти. Для взаємодії з моделлю було обрано API Ultralytics. Крім того, важливо зазначити, що для обробки зображень було використано бібліотеку OpenCV завдяки її здатності легко переходити від обробки тестових даних з медіафайлів до даних камери в режимі реального часу [25].

Робота присвячена розробці та дослідженню методів виявлення об'єктів на зображеннях і у відеопотоці з використанням сучасних технологій глибокого навчання. Основна увага приділяється створенню, навчанню та практичному застосуванню моделей детектування об'єктів на основі нейронних мереж, що забезпечують високу точність і швидкість обробки візуальних даних [18].

Мета статті полягає в визначенні об'єктів заданого типу на зображеннях, отриманих із відеокамери, а також розробці моделі детекції, яка зможе ефективно ідентифікувати та локалізувати ці об'єкти в різних умовах освітлення, масштабу та перспективи.

Матеріали і методи дослідження. Для досягнення поставленої мети було використано попередньо натренований детектор, проведено його донавчання (fine-tuning) на спеціалізованій вибірці даних і адаптовано модель до конкретних умов задачі [24, 26]. Отримані результати моделювання та навчання підлягають оцінюванню на тестовій вибірці, що дозволяє проаналізувати точність, повноту та швидкодію моделі, а також визначити можливі напрямки її подальшої оптимізації та вдосконалення.

У межах дослідження було застосовано модель YOLOv11s для сегментації зображень, що містять об'єкти захисного спорядження. Проведено порівняльний аналіз якості результатів між підходами з попереднім навчанням моделі та zero-shot learning. Тестування моделей здійснювалося у середовищі Ultralytics із використанням мови програмування Python 3.11.

Для реалізації завдання сегментації зображень було сформовано спеціалізований датасет, що складався із 150 початкових зображень, на яких представлені об'єкти захисного спорядження. На етапі підготовки до тренування моделі до цього набору даних було застосовано процедури аугментації, зокрема віддзеркалення, обертання, масштабування та зміну яскравості, що дозволило суттєво підвищити різноманітність навчальних прикладів.

Донавчений сегментатор на основі моделі YOLO було порівняно з моделлю SegmentAnythinModel (SAM). Архітектури моделей наведені в табл. 1.

Таблиця 1 – Архітектури моделей SAM та YOLO

SAM	YOLOv11s
Backbone: Vision Transformer (ViT-B) Розмір: Base (B) Кількість шарів ViT: 12 Розмір вектору ознак (hidden size): 768 Кількість голів у self-attention: 12 Кількість параметрів: ~63 млн Тип сегментаційного голови: Mask decoder із трансформерною архітектурою	Single-stage, anchor-free детектор з підтримкою сегментації Кількість шарів: 113 Кількість параметрів: 2,868,664 (~2.87 млн) GFLOPs: ~9.7 GFLOPs

Архітектура SAM базується на Vision Transformer (ViT-B) з трансформерним декодером масок, має значно більшу кількість параметрів (~63 млн) і орієнтована на високоточну сегментацію завдяки моделюванню глобальних залежностей. YOLOv11s є одностадійним anchor-free детектором із підтримкою сегментації, який містить суттєво менше параметрів (~2,87 млн) та характеризується низькою обчислювальною складністю (~9,7 GFLOPs). Таким чином, SAM забезпечує вищу гнучкість і якість сегментації ціною значних обчислювальних витрат, тоді як YOLOv11s оптимізований для швидкодії та застосувань у реальному часі.

Результати обговорення. В спеціалізованому датасеті для вирішення завдання сегментації розподіл об'єктів на зображеннях по класах має вигляд, наведений на рис. 1.

COLOR	CLASS NAME	COUNT
Orange	gloves	98
Purple	helmet	326
Red	mask	32
Blue	safety_vest	244
Green	suit	19

Рисунок 1 – Розподіл об'єктів за кількістю по зображеннях датасету

Найбільшу кількість прикладів має клас helmet (326), що свідчить про його домінування у навчальній вибірці. Значну представленість також має клас safety_vest (244), тоді як класи gloves (98) і mask (32) представлені помірно та слабо відповідно. Найменшу кількість зразків зафіксовано для класу suit (19), що вказує на суттєвий дисбаланс даних. Така нерівномірність розподілу класів може впливати на якість навчання моделі та потребує застосування методів балансування або зважування класів. Детальне описання аугментації та розподіл зображень між тренувальним, тестовим та валідаційним датасетами наведено на рис. 2.

Загальний обсяг вибірки становить 390 зображень, що відображено разом із прикладами анованих кадрів. Датасет поділено на навчальну, валідаційну та тестову підвибірки у співвідношенні 92 %, 6 % та 2 % відповідно. Навчальна вибірка містить 360 зображень і використовується для безпосереднього налаштування параметрів моделі. Валідаційна вибірка складається з 23 зображень і призначена для контролю якості навчання та підбору гіперпараметрів. Тестова вибірка включає 7 зображень і застосовується для остаточної оцінки узагальнювальної здатності моделі.

На рис. 3 відображено динаміку якості навчання моделі в процесі тренування за показниками Validation Mean IoU та функції втрат.

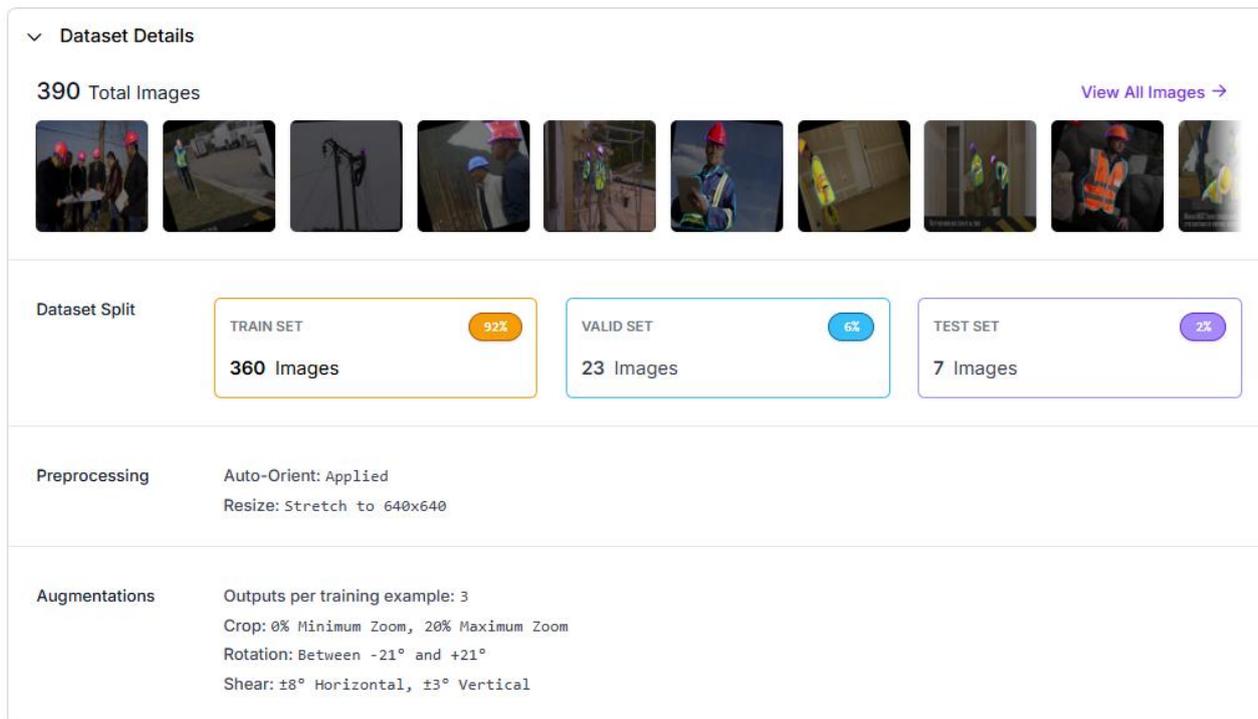


Рисунок 2 – Характеристика кроків аугментації даних та розподіл зображень між розділами датасету

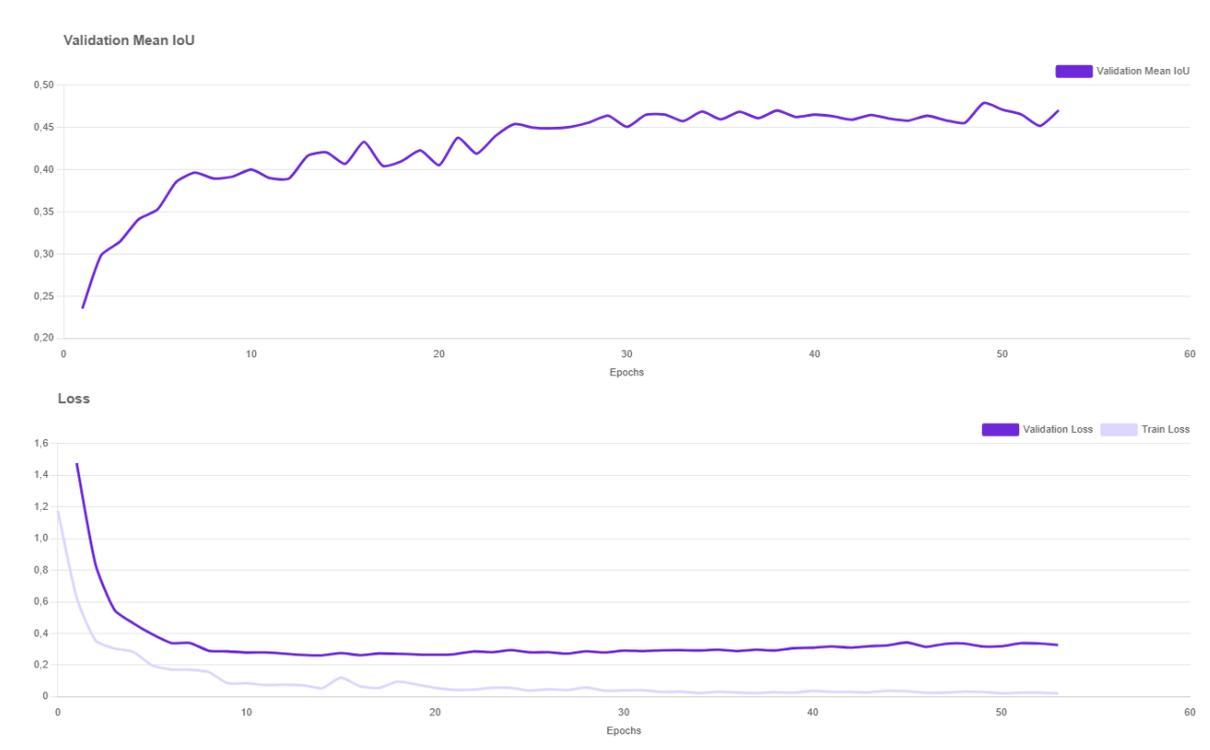


Рисунок 3 – Функції оцінки якості створеної моделі

Значення Validation Mean IoU демонструє стале зростання на початкових епохах із подальшим виходом на плато на рівні приблизно 0,46–0,48, що свідчить про досягнення моделлю стабільної якості сегментації. Крива функції втрат для навчальної вибірки різко зменшується на перших епохах і надалі повільно спадає, що вказує на ефективну оптимізацію параметрів моделі. Водночас значення Validation Loss зменшується до певного мінімуму, після

чого спостерігаються незначне поступове зростання, що може свідчити про необхідність передчасної зупинки навчання. Така поведінка кривих свідчить про відсутність суттєвого перенавчання та коректну узагальнювальну здатність моделі на валідаційних даних.

Після збалансованого представлення даних і підвищення узагальнювальної здатності моделі, були оброблені зображення із сформованого датасету, зокрема деякі подано на рис. 4.



Рисунок 4 – Приклад зображення із сформованого датасету

На рис. 4.а наведено зображення працівника на будівельному майданчику з анотацією елементів засобів індивідуального захисту, що ілюструє приклад детекції захисного шолома, жилета та рукавиць у виробничому середовищі. На рис. 4.б представлено сцену з групою людей у громадському просторі, на якій виконано розмітку обличчя та захисних масок, що демонструє здатність датасету охоплювати різні умови освітлення, ракурси та щільність об'єктів. Наведені приклади демонструють репрезентативність сформованого датасету та його придатність для навчання моделей комп'ютерного зору з метою розпізнавання засобів індивідуального захисту в реальних умовах. Ненавчена модель виявилася малоприсадоною для прямого порівняння з донавченою, оскільки попереднє тренування було орієнтоване на розпізнавання базових категорій об'єктів. Для оцінювання ефективності роботи моделі в межах створеного датасету було проаналізовано графік зміни метрики IoU (Intersection over Union) протягом процесу навчання, що дало змогу відстежити динаміку покращення якості сегментації. Оскільки донавчання виконувалося у віддаленому середовищі, застосовувалися рекомендовані параметри, визначені розробниками сервісу [25]. Серед них ключову роль для задачі сегментації відіграють такі гіперпараметри: оптимізатор Stochastic Gradient Descent, швидкість навчання (learning rate) – 0.01, коефіцієнт моменту (momentum) – 0.937 та регуляризаційний параметр weight decay – 0.0005, які забезпечують стабільність і збіжність процесу оптимізації моделі.

Функція втрат у моделі сегментації формується як поєднання трьох складових компонентів. Box loss реалізовано у вигляді CIoU Loss (Complete IoU), який відповідає за точність регресії обмежувальних прямокутників. Classification loss базується на Binary Cross-Entropy (BCE) Loss і використовується для класифікації об'єктів за їх типами. Mask loss також побудовано на BCE Loss, що забезпечує коректну бінаризацію масок об'єктів на зображенні. Додатково для підвищення стабільності та точності сегментації застосовувався Dice loss, який покращує збалансованість між передбаченими та реальними областями об'єктів.

Zero-shot модель була застосована для розв'язання задачі сегментації зображень з метою виявлення елементів захисного будівельного спорядження, таких як жилети, маски та рукавиці. Робота SAM ґрунтується на принципі promptable segmentation, тобто модель отримує підказку (prompt) від користувача і виконує точну сегментацію обраного об'єкта на зображенні.

Архітектурно SAM побудована на Vision Transformer (ViT), який відповідає за витягнення ознак, та масковому предикторі, що генерує бінарні маски об'єктів. Результати сегментації, отримані за допомогою SAM, характеризуються високою точністю відтворення контурів і форм об'єктів, а також універсальністю — модель ефективно працює з будь-якими типами об'єктів без потреби у додатковому донавчанні.

Результати роботи моделі zero-shot підходу за наперед заданими точками та порівняння з YOLO наведені на рис. 5, лівий та правий стовбці відповідно.

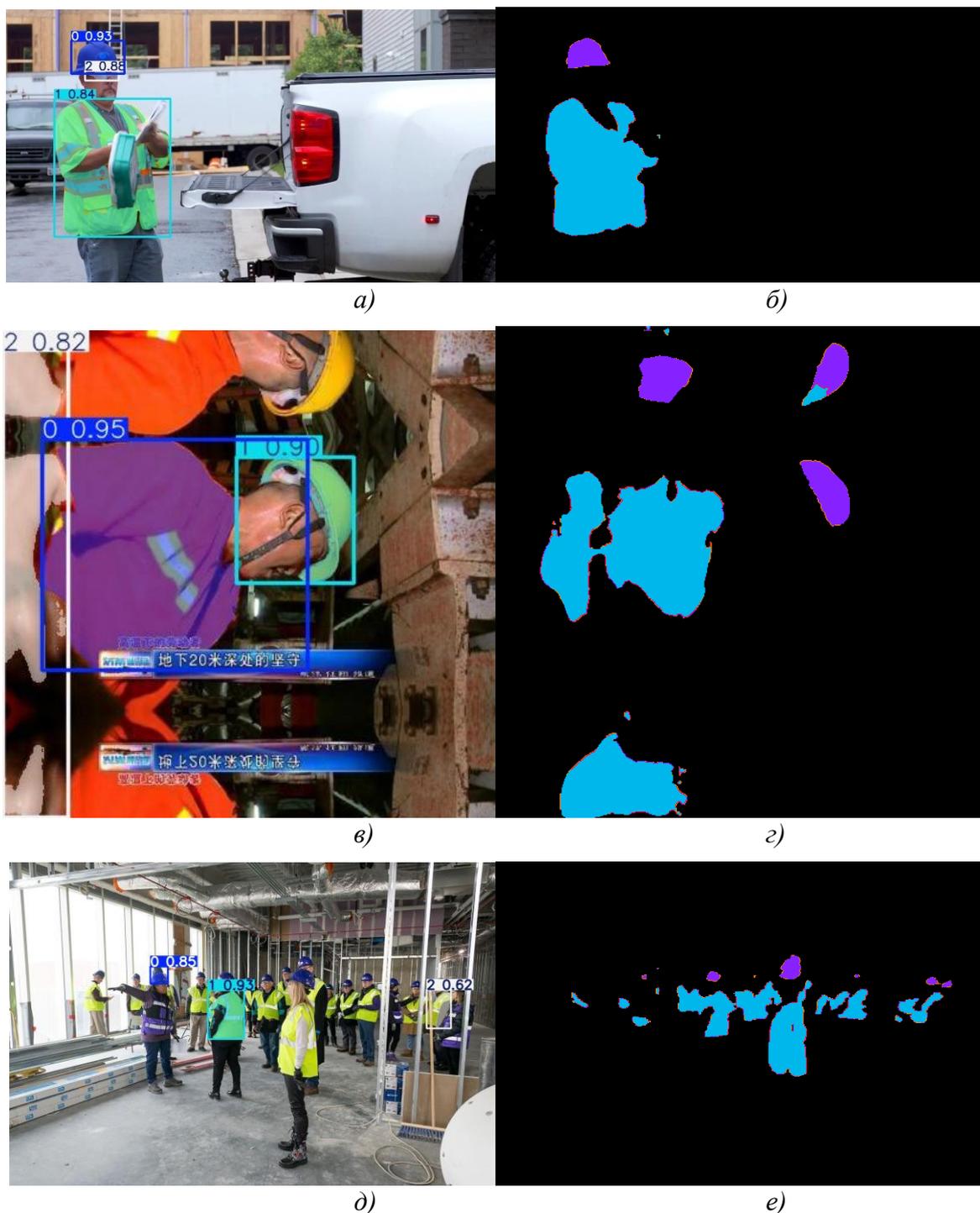


Рисунок 5 – Результати роботи моделі zero-shot підходу та порівняння з YOLO

Лівий стовбець (рис.5, а, в, д) ілюструє детекцію об'єктів із використанням обмежувальних рамок та ймовірнісних оцінок, що дозволяє ідентифікувати наявність засобів індивідуального захисту та локалізувати їх у межах зображення. Другий стовбець (рис.5, б, г, е) відображає піксельну сегментацію, де об'єкти класифіковано на рівні кожного пікселя, що забезпечує більш точне відокремлення форми людини та елементів екіпірування від фону. На відміну від підходу з рамками, сегментація дозволяє коректно враховувати складні контури та часткові перекриття

об'єктів. Водночас детекція з обмежувальними рамками є обчислювально простішою та більш придатною для застосувань у режимі реального часу.

Під час навчання моделі було застосовано спеціальний параметр, характерний для задач сегментації – `mask_ratio = 4`, який визначає ступінь наближення форми сегментованої області до графічних примітивів. У результаті значення метрики IoU (перетину з розміченими даними) досягло приблизно 40%, що свідчить про задовільну точність моделі для поставленого завдання.

У процесі попередньої обробки даних застосовано автоматичну орієнтацію зображень та масштабування до розміру 640×640 пікселів зі збереженням пропорцій шляхом розтягування. Для підвищення різноманітності навчальних даних використано методи аугментації, які генерують по три вихідні зразки на одне навчальне зображення. Зокрема, реалізовано випадкове кадрування, обертання в діапазоні від -21° до $+21^\circ$ та зсуви за горизонталлю і вертикаллю, що сприяє підвищенню стійкості моделі до змін масштабу та положення об'єктів.

Порівняння швидкостей роботи моделей SAM та YOLO демонструють суттєву різницю у швидкості обробки між моделлю сегментації та моделлю детекції об'єктів. Для сегментації зображень розміром 1024×1024 час інференсу становить приблизно 5–6,3 с на одне зображення у разі повної сегментації зображення та 120 мс у разі сегментації за наперед визначеними точками, що є обчислювально затратним і значно перевищує час обробки зображення моделлю YOLO з додатковими вимогами до знаходження точок, що відповідають приблизному знаходженню об'єкта. Натомість для детекції об'єктів на зображеннях меншого розміру (384×640 – 640×640) час інференсу коливається в межах 43–88 мс, що на два порядки швидше. Попередня та післяобробка в обох випадках займають незначну частку загального часу й не є критичними для продуктивності. Таким чином, сегментація забезпечує детальніший результат ціною значного зниження швидкодії, тоді як детекція є суттєво ефективнішою з точки зору часу обробки. Отримані результати свідчать, що для задач реального часу доцільніше застосовувати детекційні моделі, тоді як сегментаційні підходи більше підходять для офлайн-аналізу. Результати сегментації демонструють високу точність країв і об'єктів різної форми.

Висновки. Було сформовано спеціалізований датасет для сегментації зображень об'єктів захисного спорядження, що забезпечив якісну основу для навчання моделей комп'ютерного зору. Застосування процедур аугментації, таких як віддзеркалення, обертання, масштабування та зміна яскравості, дозволило суттєво розширити обсяг і різноманітність даних, підвищивши їх репрезентативність. Детально порівняні параметри, швидкості та результати роботи моделей SAM та YOLO. Отримані результати свідчать, що збільшення обсягу датасету позитивно вплинуло на збалансованість даних і узагальнювальну здатність моделей. Проведене порівняння підтвердило ефективність використання підходу окремого навчання моделей для задач сегментації та класифікації. Таким чином, запропонований підхід може бути використаний як ефективне рішення для автоматизованої обробки зображень у завданнях виявлення об'єктів на зображеннях і у відеопотоці.

До перспектив подальших розвідок слід віднести розширення датасету, оптимізацію обчислювальної складності моделей і дослідження їх застосування в режимі реального часу для відеоаналітики.

Список використаних джерел

1. Schunfelder, P., Stebel, F., Andreou, N., & Kunig, M. (2024). Deep learning-based text detection and recognition on architectural floor plans. *Automation in Construction*, 157, Article 105156. <https://doi.org/10.1016/j.autcon.2023.105156>.
2. Giakoumoglou, N., Pechlivani, E. M., & Tzovaras, D. (2023). Generate-paste-blend-detect: Synthetic dataset for object detection in the agriculture domain. *Smart Agricultural Technology*, 5, Article 100258. <https://doi.org/10.1016/j.atech.2023.100258>.
3. Ashourpour, M., Azizpour, G., & Johansen, K. (2024). Real-time defect and object detection in assembly line: A case for in-line quality inspection. In *Lecture Notes in Mechanical Engineering* (pp. 99–106). Springer. https://doi.org/10.1007/978-3-031-38241-3_12.

4. Azevedo, P., & Santos, V. (2024). Comparative analysis of multiple YOLO-based target detectors and trackers for ADAS in edge devices. *Robotics and Autonomous Systems*, 171, Article 104558. <https://doi.org/10.1016/j.robot.2023.104558>.
5. Sanjai Siddharthan, M., Aravind, S., & Sountharajan, S. (2024). Real-time road hazard classification using object detection with deep learning. In *Lecture Notes in Networks and Systems* (Vol. 789, pp. 479–492). Springer. https://doi.org/10.1007/978-981-99-6586-1_33.
6. Wei, Z., Zhang, Y., Wang, X., Zhou, J., Dou, F., & Xia, Y. (2024). A YOLOv8-based approach for steel plate surface defect detection. *Metallurgija*, 63(1), 28–30.
7. Wu, F., Zhang, Y., Wang, L., Hu, Q., Fan, S., & Cai, W. (2023). A deep learning-based lightweight model for the detection of marine fishes. *Journal of Marine Science and Engineering*, 11(11), Article 2156. <https://doi.org/10.3390/jmse11112156>.
8. Zhang, G., Tang, Y., Tang, H., Li, W., & Wang, L. (2023). A global lightweight deep learning model for express package detection. *Journal of Intelligent & Fuzzy Systems*, 45(6), 12013–12025. <https://doi.org/10.3233/JIFS-232874>.
9. Wang, J., Dai, H., Chen, T., Liu, H., Zhang, X., Zhong, Q., & Lu, R. (2023). Toward surface defect detection in electronics manufacturing by an accurate and lightweight YOLO-style object detector. *Scientific Reports*, 13, Article 33804. <https://doi.org/10.1038/s41598-023-33804-w>.
10. Li, A., Zhang, Z., Sun, S., Feng, M., & Wu, C. (2023). MultiNet-GS: Structured road perception model based on multi-task convolutional neural network. *Electronics*, 12(19), Article 3994. <https://doi.org/10.3390/electronics12193994>.
11. Han, L., Ma, C., Liu, Y., Jia, J., & Sun, J. (2023). SC-YOLOv8: A security check model for the inspection of prohibited items in X-ray images. *Electronics*, 12(20), Article 4208. <https://doi.org/10.3390/electronics12204208>.
12. Mao, J., Wang, L., Wang, N., Hu, Y., & Sheng, W. (2023). A novel method of human identification based on dental impression image. *Pattern Recognition*, 144, Article 109864. <https://doi.org/10.1016/j.patcog.2023.109864>.
13. Kara, E., Zhang, G., Williams, J. J., Ferrandez-Quinto, G., Rhoden, L. J., Kim, M., Kutz, J. N., & Rahman, A. (2023). Deep learning based object tracking in walking droplet and granular intruder experiments. *Journal of Real-Time Image Processing*, 20, 269–311. <https://doi.org/10.1007/s11554-023-01341-4>.
14. Zhou, S., Zhong, M., Chai, X., Zhang, N., Zhang, Y., Sun, Q., & Sun, T. (2024). Framework of rod-like crops sorting based on multi-object oriented detection and analysis. *Computers and Electronics in Agriculture*, 216, Article 108516. <https://doi.org/10.1016/j.compag.2023.108516>.
15. Shan, P., Yang, R., Xiao, H., Zhang, L., Liu, Y., Fu, Q., & Zhao, Y. (2023). UAVPNet: A balanced and enhanced UAV object detection and pose recognition network. *Measurement*, 222, Article 113654. <https://doi.org/10.1016/j.measurement.2023.113654>.
16. Talaat, F. M., & ZainEldin, H. (2023). An improved fire detection approach based on YOLOv8 for smart cities. *Neural Computing and Applications*, 35, 20939–20954. <https://doi.org/10.1007/s00521-023-08809-1>.
17. Liu, S., Fan, Q., Zhao, C., & Li, S. (2023). RTAD: A real-time animal object detection model based on a large selective kernel and channel pruning. *Information*, 14(10), Article 535. <https://doi.org/10.3390/info14100535>.
18. Smolii, V. M., Smolii, N. V., Kovalenko, O. Y., & Shvydenko, M. Z. (2025). Channel extractor for UAV PPM signals. *CEUR Workshop Proceedings*, 3917, 226–236. <https://ceur-ws.org/Vol-3917/>.
19. Su, Y., Tan, W., Dong, Y., Xu, W., Huang, P., Zhang, J., & Zhang, D. (2024). Enhancing concealed object detection in active millimeter wave images using wavelet transform. *Signal Processing*, 216, Article 109303. <https://doi.org/10.1016/j.sigpro.2023.109303>.
20. Liu, C., Wang, K., Li, Q., Zhao, F., Zhao, K., & Ma, H. (2024). Powerful-IoU: More straightforward and faster bounding box regression loss with a nonmonotonic focusing mechanism. *Neural Networks*, 170, 276–284. <https://doi.org/10.1016/j.neunet.2023.11.041>.

21. Xu, W., Liu, C., Wang, G., Zhao, Y., Yu, J., Muhammad, A., & Li, D. (2024). Behavioral response of fish under ammonia nitrogen stress based on machine vision. *Engineering Applications of Artificial Intelligence*, 128, Article 107442. <https://doi.org/10.1016/j.engappai.2023.107442>.
22. Dimauro, G., Barbaro, N., Camporeale, M. G., Fiore, V., Gelardi, M., & Scalera, M. (2024). DeepCilia: Automated, deep-learning-based engine for precise ciliary beat frequency estimation. *Biomedical Signal Processing and Control*, 90, Article 105808. <https://doi.org/10.1016/j.bspc.2023.105808>.
23. Zhao, X., & Song, Y. (2023). Improved ship detection with YOLOv8 enhanced with MobileViT and GSConv. *Electronics*, 12(22), Article 4666. <https://doi.org/10.3390/electronics12224666>.
24. Smolii, V. M., Smolii, N. V., & Sayapin, S. P. (2024). Search and classification of objects in the zone of reservoirs and coastal zones. *CEUR Workshop Proceedings*, 3666, 37–51. <https://ceur-ws.org/Vol-3666/paper04.pdf>.
25. Ultralytics. Train settings. In *Ultralytics Docs*. <https://docs.ultralytics.com/modes/train/#train-settings>.
26. Tang, J., Xie, N., Li, K., Liang, Y., & Shen, X. (2024). Trajectory tracking control for fixed-wing UAV based on DDPG. *Journal of Aerospace Engineering*, 37(2), Article 04024011. <https://doi.org/10.1061/JAEEZ.ASENG-5286>.

Smolij Viktorija

Doctor of Technical Sciences, Professor, Professor of the Department of Information Systems and Technologies,

National University of Life and Environmental Sciences of Ukraine

ORCID: <https://orcid.org/0000-0002-1268-7837>

E-mail: vmsmolij@nubip.edu.ua

Smolij Natan

Postgraduate Student of the specialty "Information Systems and Technologies",

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

ORCID: <https://orcid.org/0009-0002-3763-6726>

E-mail: hoibbitizukrainy@gmail.com

MACHINE VISION MODULE FOR OBJECT DETECTION IN IMAGES AND VIDEO STREAMS

Abstract. *This work addresses the topical issue of developing and applying intelligent computer vision methods for the automated detection and segmentation of protective equipment objects in images and video streams. The aim of the work is to identify objects of a specified type in images captured by a video camera, as well as to develop a detection model capable of effectively identifying and localising these objects under various lighting conditions, scales and perspectives. The article examines the data preparation process, in particular the application of augmentation methods to improve the representativeness of the sample, and performs a comparative analysis of the parameters, performance and results of the SAM and YOLO models. The results of experimental studies are presented, demonstrating the positive impact of increasing the volume and diversity of the dataset on data balance and the generalisation ability of computer vision models. The proposed approaches to training separate models for segmentation and classification tasks have proven their effectiveness in the context of automated image processing. In this work, a specialised dataset of protective equipment items was created to address the image segmentation task. During the model pre-training phase, data augmentation techniques—including mirroring, rotation, scaling and brightness adjustment—were applied to this dataset, which significantly increased the diversity of the training examples. Increasing the size of the dataset ensured a more balanced representation of the data and improved the model's generalisation ability. The results obtained in this work confirm the feasibility and effectiveness of the authors' proposed approach to this problem, namely the separate training of models for image segmentation and classification. Prospects for further research include expanding the dataset, optimising the computational complexity of the models, and investigating their application in real-time for video analytics.*

Keywords: *Object Detection, Detection Model, Image Segmentation, Model Pre-Training, Augmentation Procedures, Fine-Tuning, Binary Object Masks.*