

УДК 621.391

ЗАСТОСУВАННЯ НЕЙРОМЕРЕЖЕВИХ ТЕХНОЛОГІЙ ДЛЯ ВИЯВЛЕННЯ ПІДВОДНИХ БОЄПРИПАСІВ

В. І. СЛЮСАР

*Центральний науково-дослідний інститут озброєння та військової техніки Збройних Сил України
Україна, Київ*

Анотація. В статті обґрунтовані пропозиції щодо застосування нейронних мереж родини YOLO для виявлення підводних боєприпасів, що не спрацювали. При цьому використані попередньо навчені на датасеті MS COCO нейронні мережі YOLO3, YOLO4 та YOLO5. Доновчання нейромереж YOLO3 і YOLO4 здійснювалося на модифікованому датасеті підводного сміття Trash-ICRA19, кількість класів об'єктів у якому становила 13, з яких 2 були фіктивними. При цьому середньокласова точність детектування 13 класів об'єктів за допомогою YOLO4 у метриці mAP50 становила 75,2% або 88,9% з урахуванням фіктивних класів. Для тестування нейромереж використані зображення, отримані з відеозаписів процесу розмінування водойм за допомогою дистанційно керованих підводних апаратів (ДКПА). Запропоновано удосконалену схему нейронної мережі, що являє собою каскад з кількох послідовно з'єднаних YOLO-сегментів з багатопрхідною обробкою зображень, а також тензорно-матричний опис механізму уваги. Розроблені рекомендації щодо подальшого підвищення ефективності нейромережевого методу селекції підводних боєприпасів.

Ключові слова: дистанційно керований підводний апарат; ДКПА; нейромережа; YOLO3; YOLO4; YOLO5; Object Detection; багатопрхідна обробка; підводний об'єкт; обробка зображень; датасет

1. ВСТУП

Значний обсяг робіт з розмінування території України передбачає серед іншого очищення водойм, на дні яких можуть знаходитися боєприпаси, що не спрацювали. На практиці найбільш ефективно такі роботи проводяться з застосуванням дистанційно керованих підводних апаратів (ДКПА), які дозволяють вести підводну розвідку, виявляти небезпечні предмети та здійснювати маркування їх місцезнаходження у глобальній системі позиціонування. Зокрема, фахівці Державної служби з надзвичайних ситуацій застосовують на території Київської області при розмінуванні водойм ДКПА CHASING M2, оснащений відеокаме-

рою 4К та дротовою системою управління на відстані до 100 м [1, 2].

Разом з тим, можливість появи раптово виникаючих об'єктів в умовах низької прозорості води висуває жорсткі вимоги до швидкодії системи управління підводним апаратом. Якщо ним керує оператор, то час реакції буде досить затяжний. До цього ж слід додати порівняно велику постійну часу контуру керування підводного роботизованого комплексу, який не має можливості миттєво відпрацювати керівні сигнали через специфіку водного середовища. Особливо критичним випадком є наявність підводних течій, що можуть зносити апарат у бік вибухонебезпечного предмету та підвищують ризик інцидентів внаслідок ненавмисного зіткнення.

DOI: [10.20535/S0021347023030020](https://doi.org/10.20535/S0021347023030020)

© В. І. Слюсар, 2022

Через це важливим завданням є скорочення часу реакції роботизованого засобу підводної розвідки для підвищення швидкості реагування в ситуаціях небезпечного наближення до боєприпасів. З цією метою на борту дистанційно керованої платформи доцільно реалізувати нейромережову систему детектування об'єктів, що виявляла б небезпечні предмети та подавала екстрені сигнали управління на рушійну систему.

Метою статті є розробка пропозицій щодо застосування нейромережових технологій виявлення підводних вибухонебезпечних об'єктів.

2. ПІДГОТОВКА ДАНИХ

Для підтвердження можливості використання нейромереж в процесі підводної розвідки боєприпасів, що потребують утилізації, автором проведені експериментальні дослідження. В межах першого етапу вони базувалися на використанні нейронних мереж YOLO3 та YOLO4, попередньо навчених на датасеті MS COCO (Microsoft Common Objects in Context). Далі, згідно з методом трансферу навчання [3], здійснене додаткове навчання таких YOLO3 та YOLO4 за допомогою датасету, підготовленого для детектування об'єктів на зображеннях підводного сміття. Для цього відкинуто кілька вихідних шарів нейромережі, а в якості основи для формування датасету донавчання взятий датасет Trash-ICRA19 [4, 5], який на момент написання статті знаходився у вільному доступі за Інтернет-адресою <https://conservancy.umn.edu/handle/11299/214366>.

Відповідні зображення загальною кількістю 5700, що увійшли до датасету Trash-ICRA19, отримані з набору відеоданих J-EDI (JAMSTEC E-Library of Deep-sea Images) про різні види морського сміття. Вони містять покадрові зображення різних типів об'єктів, знятих відеокамерами ДКПА у реальному середовищі, і суттєво відрізняються за якістю, глибиною, прозорістю води та рівнем

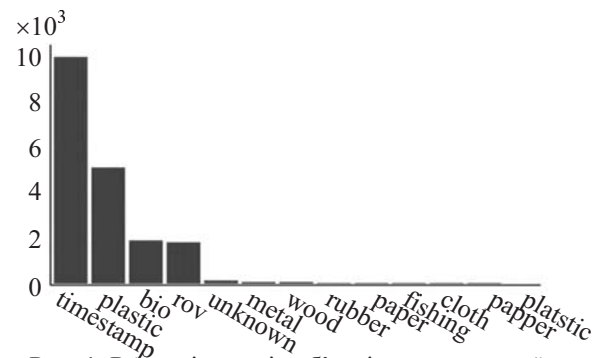


Рис. 1. Розподіл класів об'єктів тренувальної вибірки датасету за присутністю на зображеннях.

освітлення, кількістю об'єктів на сценах і параметрами використаних відеокамер.

Сформований з зазначених відеозаписів покадровий датасет доповнений до загальної кількості зображень 7666, з яких 6132 фотокадри сформували тренувальну вибірку, а решта 1534 — валідаційну. Формат зображень приведено до розміру 416×416 пікселів.

Усі екземпляри підводного сміття, різноманітні біологічні об'єкти, такі як рослини та тварини, а також ДКПА (Remotely Operated underwater Vehicle, ROV) у різних станах позначені обмежувальними рамками (боксами) та умовно поділені на 13 класів. У цьому полягала одна з головних відмінностей від підходу [1], де були розглянуті лише 3 класи об'єктів (plastic, ROV, bio). Крім цих трьох класів з [1], у модифікованому датасеті виділені ще 10 додаткових: cloth (тканина), fishing (риби), metal (металевий об'єкт), paper (папір), papper (альтернативний клас паперу), platstic (альтернативний клас пластика), rubber (гума), timestamp (службові дані на фотозображенні), unknown (невизначений об'єкт), wood (дерево).

Слід вказати, що альтернативні класи паперу (papper) та пластика (platstic) додані штучно з метою дослідження впливу на працездатність нейромережі незбалансованості наявного датасету по кількості зображень у класах. Зазначені два класи не містили жодного із зразків зображень. Інші класи представ-

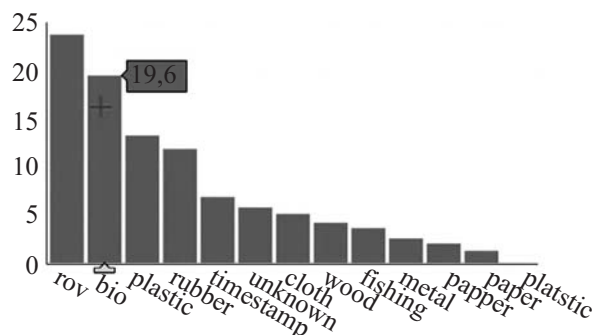


Рис. 2. Розподіл наявних у тренувальному датасеті об'єктів за процентом простору, який вони займають на зображеннях.

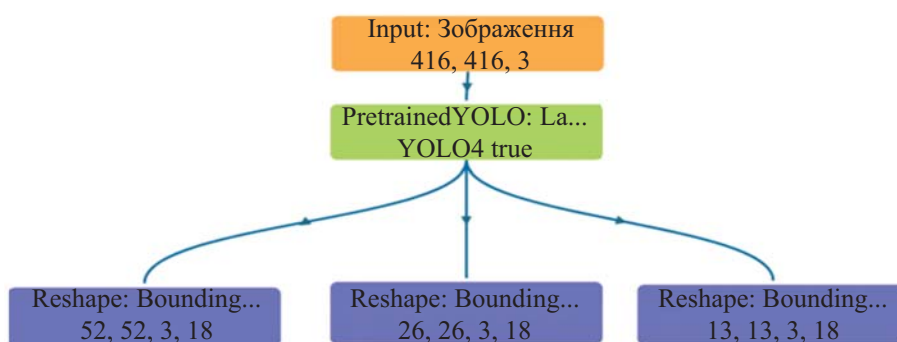


Рис. 3. Схема застосування попередньо навченої YOLO4 для вирішення завдання Object Detection.

лені досить мізерною кількістю прикладів. Ілюстрація загального розподілу об'єктів за класами у тренувальній вибірці модифікованого датасету представлена на рис. 1. Більше всього на зображеннях міститься службових відміток та пластику.

Характеризуючи використаний в роботі датасет, слід вказати, що досить нерівномірним було не тільки наповнення його класів зображеннями, а й розподіл наявних на них об'єктів за процентом простору, який вони займають (рис. 2). Як видно, найбільшими за кількістю зайнятих пікселів є апарати ROV та біо-об'єкти.

3. ВИКОРИСТАННЯ НЕЙРОМЕРЕЖІ YOLO4

Загальна схема застосування попередньо навченої YOLO4 для вирішення завдання Object Detection наведена на рис. 3. Ця нейронна мережа має 61640962 параметри, з яких можливо проводити тренування по 61588354

параметрах. Розмірність вихідних тензорів шарів Reshape 52×52×3×18, 26×26×3×18, 13×13×3×18 зумовлена трьома розмірами боксів (52×52, 26×26, 13×13) та триколіровим представленням RGB зображень. Величина параметру 18 у розмірності вихідних тензорів пояснюється тим, що до кількості класів 13 мають бути додані 4 координати бокса і мітка класу, тобто 5 додаткових величин.

Та обставина, що фото деяких з об'єктів зовсім не представлені у датасеті, хоча й позначені у загальному переліку класів, помітно вплинула на показники точності ідентифікації класів об'єктів. Однак, не зважаючи на внесені викривлення, для YOLO3 після 25 епох донавчання з оптимізатором Adam усереднена по усіх класах точність за метрикою mAP50 становила 75% (рис. 4). Схожу величину цього показника (75,2%) продемонструвала YOLO4 після 34 епох навчання. Відповідний розподіл

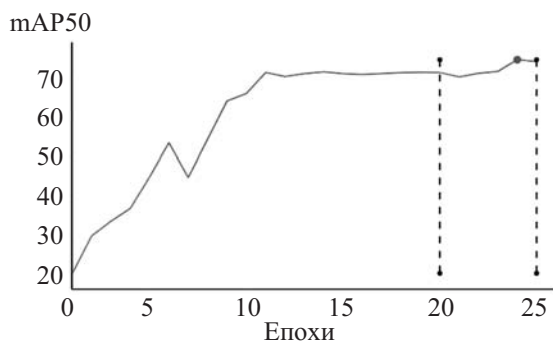


Рис. 4. Середньокласова точність за метрикою mAP50 неймережі YOLO3 після 25 епох.

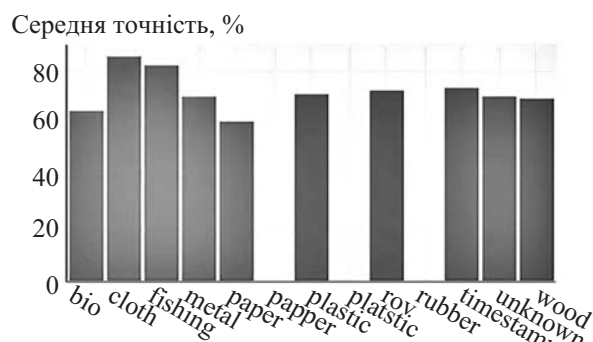


Рис. 5. Розподіл точності визначення координат об'єктів за класами в YOLO4.

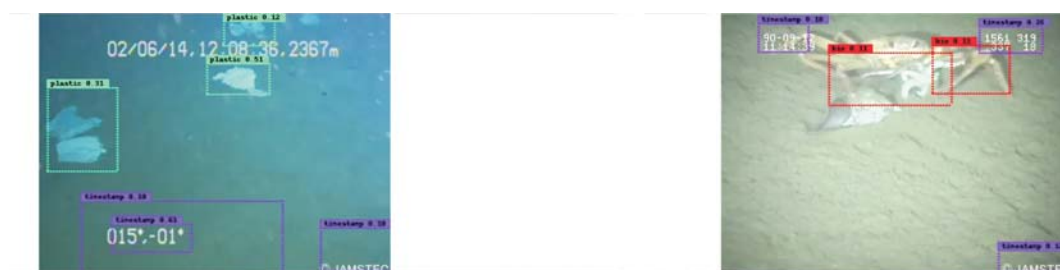


Рис. 6. Приклади детектування підводних об'єктів.

точності визначення координат об'єктів за класами наведено на рис. 5.

4. ЕКСПЕРИМЕНТАЛЬНА ПЕРЕВІРКА ВИКОРИСТАННЯ YOLO4

Приклад детектування підводних об'єктів наведено на рис. 6. При цьому обрано фіксований розмір батчу 4, щоб мінімізувати вимоги до апаратних ресурсів Google CoLab, на базі якого проводилося навчання (відеокарта Tesla V100-SXM2-16GB). Оскільки, як відомо, розмір батча здійснює помітний вплив на точність класифікації об'єктів, в подальшому доцільно провести додаткові дослідження для батчів 8, 16, 64. Цілком ймовірно, що оптимізація величини цього параметру дозволить підвищити точність зазначених нейронних мереж YOLO за метрикою mAP50.

Слід зауважити, що усереднений за усіма класами показник точності mAP50 не в усіх випадках відображує кращий результат. Оскільки у модифікованому датасеті задіяні два фіктивних класи (без відповідних об'єктів), то до-

сягнуту середньокласову величину точності слід скорегувати, помноживши її на коефіцієнт $13/(13 - 2) = 13/11$. При цьому отримаємо скориговану точність $75,2\% \times 13/11 = 88,87\%$, що є досить непоганим результатом.

Для підтвердження ефективності використання донавичених в описаний вище спосіб нейронних мереж при виявленні підводних боєприпасів, що не спрацювали, в якості тестових зображень використано кадри з відеозаписів розмінування у 2022 році водойм у Київській області [1, 2]. Необхідно зазначити, що підводні боєприпаси не входили до переліку об'єктів, на яких тренувалася неймережа YOLO. Однак концепція навчання без попереднього досвіду ZSL (Zero-Shot Learning) [6, 7] дозволяє використовувати здобутий неймережею «досвід» з розпізнавання підводного сміття для виявлення підводних боєприпасів без наявності тренувальних прикладів таких об'єктів. Це здійснюється за допомогою атрибутивного простору, що містить спільні описові характеристики (атрибути) об'єктів.

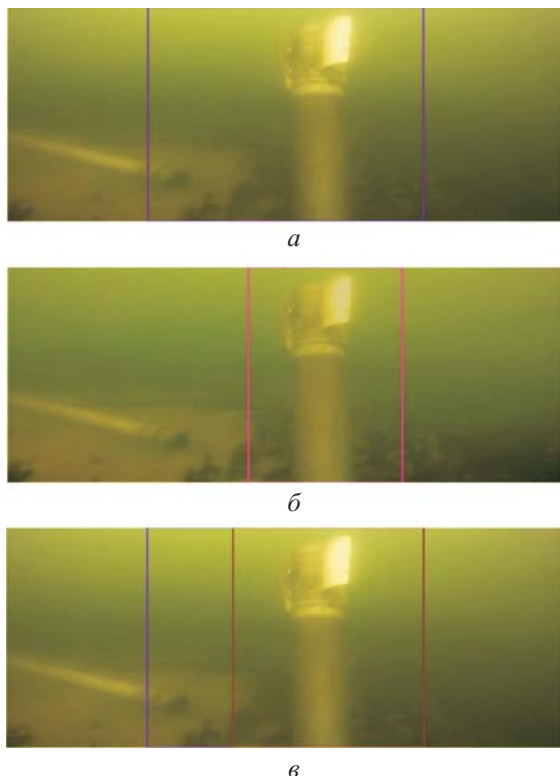


Рис. 7. Виявлений YOLO4 боєприпас реактивної системи залпового вогню (РСЗВ) без попередньої фільтрації зображення (а), після неї (б), та після 2-го проходу (в).



Рис. 9. Приклад сприйняття боєприпасу в YOLO4 як пластика.

Так, неймережа, яка вже навчилася розпізнавати певні форми, матеріали або кольори у підводному смітті, може використовувати ці атрибути для виявлення підводних боєприпасів. Концепція ZSL по суті полягає в тому, що неймережа вивчає зв'язки між класами і атрибутами, що дозволяє їй узагальнювати і вважати новий об'єкт одним з тих об'єктів, на яких вона вчилася. Zero-Shot Learning має широке застосування в області комп'ютерного зору і оброб-

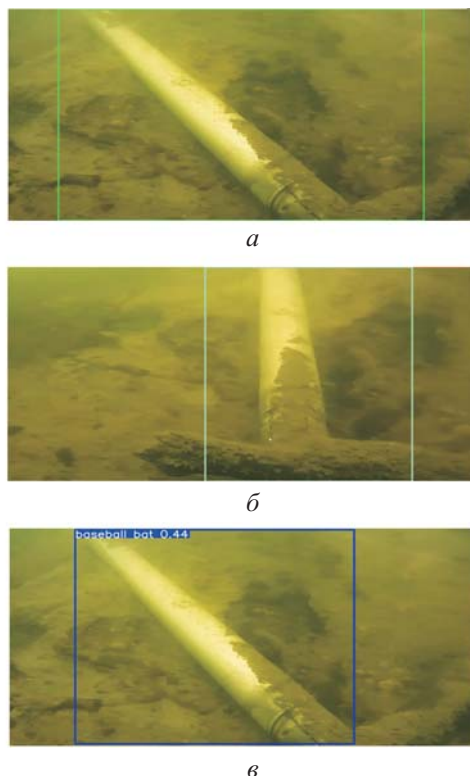


Рис. 8. Локалізація боєприпасу в YOLO4 при одному (а) та іншому (б) ракурсі, та YOLO5 XL з меншим за розміром боксом (в).

ки природних мов. Цей підхід може бути використаний для розпізнавання об'єктів, класифікації тексту, виявлення аномалій і вирішення багатьох інших завдань, де інформація про нові класи об'єктів обмежена або недоступна.

Результати, отримані за допомогою YOLO4 в рамках реалізації методу ZSL наведені на рис. 7, 8.

Найчастіше вибухонебезпечні об'єкти визначалися як пластик (рис. 9), метал, ROV або невідомий об'єкт. Попередня фільтрація зображення шляхом перебалансування його контрасту та яскравості дозволяла підвищити точність локалізації боєприпасу (рис. 7б).

5. ВИКОРИСТАННЯ НЕЙРОМЕРЕЖИ YOLO5

Другий етап досліджень полягав у перевірці ефективності неймереж родини YOLO5 (табл. 1), які попередньо навчені на датасеті

Таблиця 1. Характеристики нейромереж родини YOLO5

Назва	Кількість шарів	Кількість параметрів	Об'єм структури, МБайт
YOLO5 Small (S)	213	7225885	14
YOLO5 Medium (M)	290	21172173	40,7
YOLO5 Large (L)	367	46533693	89,2
YOLO5 Extra Large (XL)	444	86705005	166

MS COCO, безпосередньо для виявлення та класифікації підводних боєприпасів без попереднього донавчання на датасеті зображень підводного сміття. Фактично мала місце більш жорстка за умовами процедура ZSL. Отримані результати засвідчили, що мережі родини YOLO5 більш ретельно налаштовуються на конкретні класи об'єктів. Ця обставина не дозволяє їх так само легко масштабувати на різні середовища та типи об'єктів, як це можливо у випадку YOLO4.

Зокрема, нейромережі YOLO5, для яких не проводилося додаткове навчання після попереднього тренування на датасеті MS COCO, не змогли зафіксувати боєприпаси на зображеннях (рис. 7–9). В свою чергу, YOLO5 XL ідентифікувала боєприпас на рис. 8а як бейсбольну біту з імовірністю 0,44 (рис. 8в). Однак, головне при цьому виявити сам об'єкт. До того ж, YOLO5 XL продемонструвала більш точну порівняно з YOLO4 локалізацію боєприпасу на рис. 8а меншим за розміром боксом (рис. 8в).

Як видно з рис. 10а, YOLO5S краще локалізує водолаза як персону порівняно з розглянутою вище YOLO4, яка детектує його як біооб'єкт з більшим за розмірами боксом. Пролілюстрований на рис. 10а результат отриманий шляхом подання зображення спочатку на вхід YOLO4, а потім з її виходу — на YOLO5S.

Суттєво, що при безпосередній подачі фотознімку на вхід YOLO5S, розмір сформованого нею боксу лишається тим же, однак точність детектування водолаза при цьому зростає до 88%. Перехід до більш складного варіанту

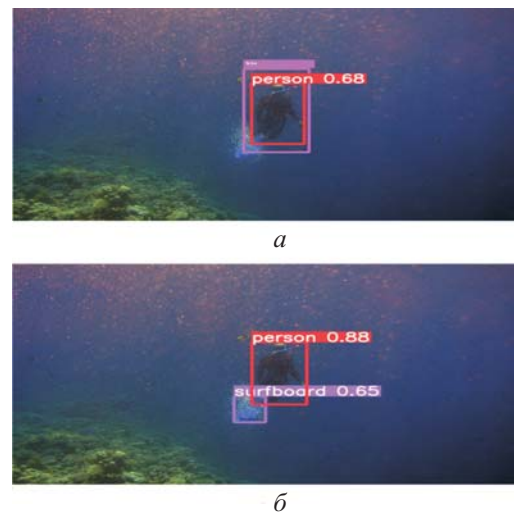


Рис. 10. Виявлення водолаза на зображенні за допомогою YOLO4 (бокс «біо») та YOLO5S (бокс «person») (а) і за допомогою YOLO5L (б).

YOLO5L дозволяє не тільки поліпшити точність, а й додатково детектувати окремі його складові (рис. 10б).

Таким чином, перехід при виявленні підводних боєприпасів до застосування мереж YOLO5, які додатково навчені на відповідному датасеті, дозволяє більш точно локалізувати місцезнаходження боєприпасів та підвищити ймовірність їх ідентифікації.

6. ЗАСТОСУВАННЯ ПОВТОРНОЇ ОБРОБКИ

В процесі експериментальних досліджень виявлено, що повторна обробка раніше розпізнаних зображень в деяких випадках підвищує якість детектування вибухонебезпечних об'єктів. Для прикладу на рис. 11а наведено результати

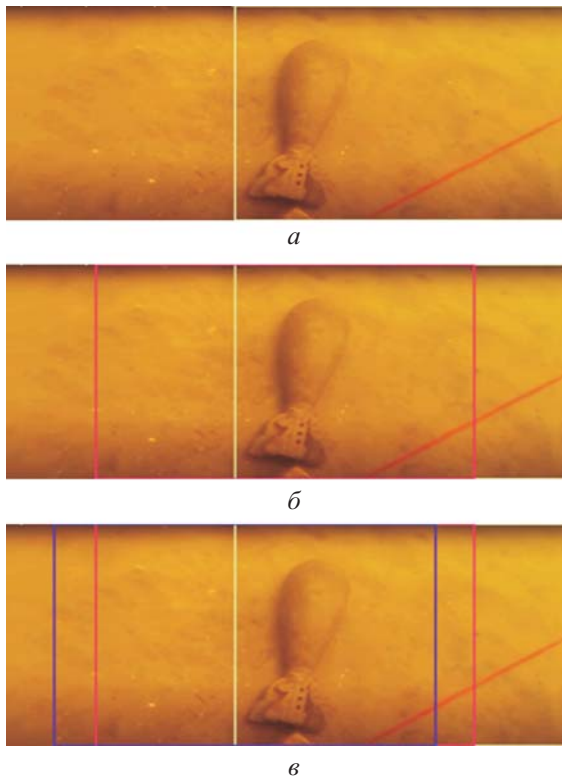


Рис. 11. Розмір боксу при використанні 1-й (а), 2-й (б), 3-й (в) проходів.

тати детектування міни на дні водойми шляхом одноразової обробки фотознімку нейромережою YOLO4, після повторного виконання детектування об'єктів по фотознімку з маркерним боксом, накладеним після першого проходу через нейромережу (рис. 11б), та після трикратного проходження одної й тої ж нейромережі відповідно (рис. 11в).

Як видно з рис. 11, повторна обробка фотознімку дозволяє сформувати точніше локалізований за розмірами бокс як перетин усіх накладених боксів. Аналогічний результат отриманий стосовно боеприпасу, зображеного на рис. 7.

7. ЗАПРОПОНОВАНА СТРУКТУРА НЕЙРОННОЇ МЕРЕЖІ

Виявлений ефект дозволяє запропонувати для практичного використання при детектуванні підводних боеприпасів поліпшену схему нейронної мережі, що являє собою каскад з

кількох послідовно з'єднаних YOLO-сегментів. Варіант запропонованої схеми для розглянутого випадку трикратного детектування об'єктів по одному й тому ж зображенню наведено на рис. 12.

У вказаній схемі блок Boxing по завершенні кожного з проходів зображення формує координати і мітки маркерних боксів (Bounding Boxes), Image+boxing — накладає сформовані бокси на початкове зображення, а Boxing Concatenation формує перетин множини усіх боксів для подальшого накладання поліпшеного за розмірами остаточного боксу на вихідне зображення. Остаточний за розмірами Bounding Box формується як перетин усіх боксів, згенерованих попередньо за кілька проходів, наприклад за методом non-maximum suppression [8]. Вибір кольору фінального боксу має здійснюватися з урахуванням ефекта оклюзії у відповідному регіоні зображення, де розташований боеприпас.

До речі, запропонована схема може поєднувати у собі не тільки однотипні YOLO-мережі. Зазначений факт поліпшення детектування об'єктів спрацьовує в контексті рис. 10а, й у випадку комбінації YOLO4 та, наприклад, нейромережі YOLO5S. В подальшому доцільно дослідити комбінації й інших типів мереж.

7.1. Використання механізму уваги

Для підвищення якості класифікації виявлених об'єктів до структури нейронних мереж, що мають донавчатися, слід додати один чи кілька сегментів механізмів уваги. Як відомо, в контексті машинного навчання та штучного інтелекту механізм уваги (attention mechanism) — це структура, яка дозволяє нейромрежам зосереджувати свій «погляд» на певних аспектах даних, ігноруючи інші [9]. У контексті обробки зображень, механізми уваги використовуються для того, щоб модель могла зосередити свою увагу на певних частинах зображення, які є найважливішими для виконання конкретного завдання.

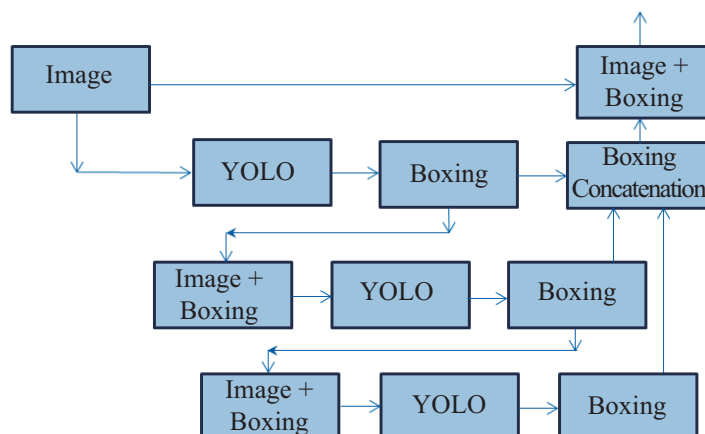


Рис. 12. Пропонована схема багатопрхідної обробки зображень.

Наприклад, при виконанні завдання розпізнавання підводних боєприпасів, модель нейронної мережі з механізмом уваги може зосередити свою чутливість на областях зображення, де знаходяться об’єкти, які потрібно розпізнати, ігноруючи менш важливі зони. Це корисно при роботі з великими зображеннями, де важливі деталі можуть займати лише невелику частину сцени. Такий механізм виявився ефективним та використовується в багатьох моделях машинного навчання, включаючи моделі трансформерів, які є основою потужних моделей мовного моделювання GPT-3, GPT-4 та ін.

Для математичного опису механізму уваги в рамках тензорно-матричної теорії нейромереж з обробки кольорових зображень слід обрати запропонований автором проникаючий добуток матриць [10]. По суті, він являє собою блоковий добуток Адамара сукупності блок-матриць різних розмірностей з однаковою розмірністю блоків.

Згідно з визначенням [10], проникаючим торцевим добутком $r \times g$ -матриці \mathbf{C} і m -вимірного тензора \mathbf{D} , розгорнутого в блокову матрицю, що містить $r \times g$ -блоки ($\mathbf{D} = [D_m]$, $m > 1$), є матриця вида:

$$\mathbf{C} \blacktriangleright \mathbf{D} = [\mathbf{C} \circ D_m], \quad (1)$$

де \blacktriangleright — символ проникаючого торцевого добутку, $\mathbf{C} \circ D_m$ — добуток Адамара.

Якщо тензор \mathbf{D} записаний як блок-рядок, то відповідно отримаємо:

$$\begin{aligned} \mathbf{C} \blacktriangleright \mathbf{D} &= [\mathbf{C} \circ D_m] = \\ &= [\mathbf{C} \circ D_1 \ \mathbf{C} \circ D_2 \ \dots \ \mathbf{C} \circ D_m \ \dots]. \end{aligned} \quad (2)$$

При цьому для кожного з кольорових каналів RGB-зображень може формуватися однакова за просторовим розподілом пікселів матрична маска \mathbf{C} , кожен з вагових коефіцієнтів якої має перемножатися за принципом добутку Адамара на блок матриці пікселів \mathbf{D} відповідного кольору (див. внизу с. 774).

В зазначеному контексті матриця фільтрувальної маски \mathbf{C} визначає локалізацію зон уваги по виходу відповідного механізму. При цьому окремо взятий блок матриці \mathbf{D} відповідає пікселям кадру зображення одного з трьох RGB-каналів нейронної мережі.

В загальному випадку кількість блоків в матриці \mathbf{D} може бути довільною, якщо замість каналів RGB розглядати вихідні тензори згорткових шарів Conv2D. У разі застосування тривимірних згорткових шарів Conv3D блоковий характер може мати і матриця масок \mathbf{C} , при цьому фіксований розподіл зон уваги буде задіяний не для усіх блоків матриці даних \mathbf{D} , а для певних їх груп. Відповідні групи блоків

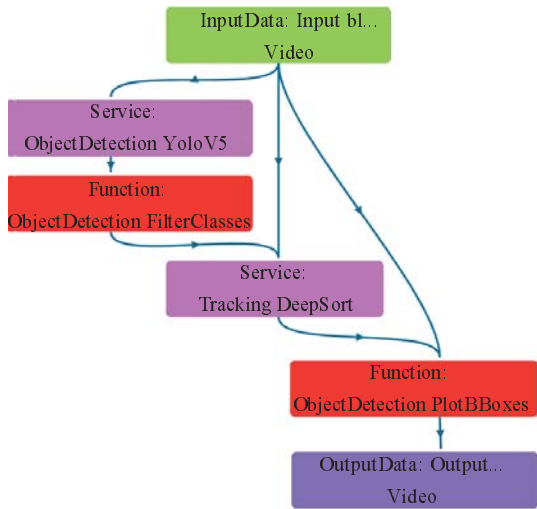


Рис. 13. Схема відеотрекінгу для виявлення підводних боєприпасів на основі попередньо навченої нейромережі YOLO5.

відповідатимуть одному значенню індексу додаткової розмірності тензора даних по виходу шару Conv3D.

Додаткову свободу у розподілі зон локалізації уваги надає підхід, що може бути описаний за допомогою проникаючого прямого (кронекерівського) добутку матриць [11]. Сутність його зводиться до поширення операції (1) на випадок, коли тензори \mathbf{C} і \mathbf{D} представлені у вигляді блокових матриць з блоками однакової розмірності:

$$\mathbf{C} \llbracket \mathbf{D} = [C_{ij} \llbracket \mathbf{D}] = [C_{ij} \circ D_{mr}]. \quad (3)$$

$$\mathbf{C} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} b_{111} & b_{121} & b_{131} & b_{112} & b_{122} & b_{132} & b_{113} & b_{123} & b_{133} \\ b_{211} & b_{221} & b_{231} & b_{212} & b_{222} & b_{232} & b_{213} & b_{223} & b_{233} \\ b_{311} & b_{321} & b_{331} & b_{312} & b_{322} & b_{332} & b_{313} & b_{323} & b_{333} \\ b_{411} & b_{421} & b_{431} & b_{412} & b_{422} & b_{432} & b_{413} & b_{423} & b_{433} \end{bmatrix},$$

$$\mathbf{C} \llbracket \mathbf{D} =$$

$$\begin{bmatrix} a_{11} \cdot b_{111} & a_{12} \cdot b_{121} & a_{13} \cdot b_{131} & a_{11} \cdot b_{112} & a_{12} \cdot b_{122} & a_{13} \cdot b_{132} & a_{11} \cdot b_{113} & a_{12} \cdot b_{123} & a_{13} \cdot b_{133} \\ a_{21} \cdot b_{211} & a_{22} \cdot b_{221} & a_{23} \cdot b_{231} & a_{21} \cdot b_{212} & a_{22} \cdot b_{222} & a_{23} \cdot b_{232} & a_{21} \cdot b_{213} & a_{22} \cdot b_{223} & a_{23} \cdot b_{233} \\ a_{31} \cdot b_{311} & a_{32} \cdot b_{321} & a_{33} \cdot b_{331} & a_{31} \cdot b_{312} & a_{32} \cdot b_{322} & a_{33} \cdot b_{332} & a_{31} \cdot b_{313} & a_{32} \cdot b_{323} & a_{33} \cdot b_{333} \\ a_{41} \cdot b_{411} & a_{42} \cdot b_{421} & a_{43} \cdot b_{431} & a_{41} \cdot b_{412} & a_{42} \cdot b_{422} & a_{43} \cdot b_{432} & a_{41} \cdot b_{413} & a_{42} \cdot b_{423} & a_{43} \cdot b_{433} \end{bmatrix}$$

Стосовно завдання обробки зображень підводних боєприпасів, такий варіант матричного добутку дозволяє отримати поелементний добуток кожного блока матриці масок уваги \mathbf{C} на всі відповідні йому блоки матриці даних \mathbf{D} :

$$\begin{bmatrix} C_{11} \\ C_{21} \\ \vdots \\ C_{P1} \end{bmatrix} \llbracket \begin{bmatrix} D_{11} & D_{12} & \cdots & D_{1G} \\ D_{21} & D_{22} & \cdots & D_{2G} \\ \vdots & \vdots & \ddots & \vdots \\ D_{P1} & D_{P2} & \cdots & D_{PG} \end{bmatrix} =$$

$$\begin{bmatrix} C_{11} \circ D_{11} & C_{11} \circ D_{12} & \cdots & C_{11} \circ D_{1G} \\ C_{11} \circ D_{21} & C_{11} \circ D_{22} & \cdots & C_{11} \circ D_{2G} \\ \vdots & \vdots & \ddots & \vdots \\ C_{P1} \circ D_{P1} & C_{P1} \circ D_{P2} & \cdots & C_{P1} \circ D_{PG} \end{bmatrix}.$$

Слід зазначити, що при використанні попередньо навчених версій нейромережі YOLO, механізми уваги можуть бути підключені до її вихідних каналів, що відповідають різним за розмірами маркерним боксам. Такий підхід запропонований в [12].

7.2. Використання масштабування

На додаток до використання механізму уваги у вихідних каскадах нейромережі, на її вході може бути задіяно сегмент з масштабування зображення [13]. Таке масштабування може здійснюватися на вході першого каскаду

багатокаскадної нейромережі або реалізовуватися поступово, з покаскадним нарощуванням коефіцієнту масштабування.

В ідеалі уся багатокаскадна схема має бути реалізована на самому підводному дроні, однак при обмежених обчислювальних ресурсах частина каскадів може бути винесена на обладнання оператора. При цьому має бути забезпечений відповідний канал зв'язку для міжкаскадної передачі даних.

7.3. Обробка відеопотоку даних

Описані варіанти детектування боєприпасів нескладно поширити на відеопотік даних. Для цього на додаток до нейромережі, що забезпечує детектування об'єктів, слід задіяти додатковий механізм алгоритмічного відеотрекінгу, наприклад, за алгоритмом DeepSort (рис. 13). В результаті забезпечується детектування небезпечних предметів по відео у реальному масштабі часу.

Застосування розглянутих технологій виявлення підводних боєприпасів дозволяє знизити навантаження на канали передачі даних від безекіпажних платформ, оскільки замість масивних за обсягом зображень реальних підводних сцен можуть передаватися лише дані доповненої реальності у вигляді координат маркерних рамок (Bounding Boxes) [14]. Такий підхід дозволяє суттєво знизити трафік і прискорити проходження даних виявлення підводних боєприпасів від багатьох сенсорів.

Його втіленням є відкритий протокол отримання даних від сенсорів SAPIENT (Sensing for Asset Protection using Integrated Electronic Network Technology) [15]. Цей протокол розроблений на замовлення міністерства оборони Великої Британії і базується на XML-повідомленнях або двоїчних повідомленнях на основі Protobuf [16].

Протокол SAPIENT отримав широке впровадження в системах боротьби з безпілотними літальними апаратами, зокрема міністерство оборони Великої Британії вже прийняло

SAPIENT в якості стандарту функціональної сумісності для багатосенсорних систем протидії дронам. Тому його масштабування на підводний домен, зокрема акустичні засоби зв'язку, буде цілком доречним і забезпечить мультидоменну єдність архітектури передачі даних.

8. ВИСНОВКИ

Підводячи підсумок, слід вказати, що завдання детектування за допомогою нейромереж підводних боєприпасів, які не спрацювали, є досить актуальним і може бути вирішено з ефективністю, що не поступається спроможностям людини, та потенційно може їх перевершити. Для цього потрібно особливу увагу приділити формуванню якісного датасету, вибору оптимальної структури задіяних нейромереж та стратегії їх донавчання.

З метою розширення динамічного діапазону візуальних сцен, що підлягають аналізу, бажаним є перехід до застосування 10-бітної технології формування і обробки відеопотоків [17] та багатоспектральних відеокамер. Крім того, залежно від якості освітлення та прозорості води доцільно проводити попередню обробку і фільтрацію зображень підводних сцен з метою максимального наближення їх спектральної палітри до показників, що забезпечують штатне функціонування нейромережі за параметрами точності локалізації та класифікації об'єктів.

При підготовці датасету слід досягти збалансованості його наповнення для усіх класів боєприпасів, що мають розрізнятися. В найпростішому випадку це може бути визначення типу боєприпасу, наприклад: снаряд ствольної артилерії, реактивний снаряд (ракета системи залпового вогню), протитанкова граната, міна, авіабомба. З накопиченням необхідних відеоданих в подальшому доцільно ставити завдання класифікації не тільки типу боєприпасу, а й його марки.

Серед інших напрямів подальших досліджень слід вказати запровадження формування 3D-боксів [18] з метою визначення вектора орієнтації підводних об'єктів, поєднання обробки у нейромережі вихідних даних стереокамер, лідара та ультразвукового сонара, що забезпечить не тільки підвищення точності виявлення, а й визначення відстані до боєприпасів та їх геометричних розмірів.

Все це дозволить більш чітко планувати заходи щодо підйому, транспортування та знешкодження підводних боєприпасів. В перспективі усі зазначені роботи мають бути покладені на дистанційно керованого робота з маніпуляторами, керування яким здійснюється за допомогою екзоскелету з копіювачем рухів оператора [19].

БІБЛІОГРАФІЧНИЙ СПИСОК

- Espresso.tv, "Піротехніки розмінують водойми на Київщині," *Youtube канал Espresso.tv*, 2022. <https://www.youtube.com/watch?v=TxL8MQhBWnU>
- 5-й Канал, "Атрибути 'русского міра': як сапери розмінують село Горенка під Києвом," *Youtube канал 5.tv*, 2022. <https://www.youtube.com/watch?v=Jd4nWc4apTQ>
- V. Slyusar *et al.*, "Improving the model of object detection on aerial photographs and video in unmanned aerial systems," *Eastern-European J. Enterp. Technol.*, vol. 1, no. 9(115), pp. 24–34, 2022, doi: [10.15587/1729-4061.2022.252876](https://doi.org/10.15587/1729-4061.2022.252876)
- M. S. Fulton, J. Hong, J. Sattar, "Trash-ICRA19: A Bounding Box Labeled Dataset of Underwater Trash," 2020. doi: [10.13020/x0qn-y082](https://doi.org/10.13020/x0qn-y082)
- M. Fulton, J. Hong, M. J. Islam, J. Sattar, "Robotic detection of marine litter using deep visual detection models," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 5752–5758, doi: [10.1109/ICRA.2019.8793975](https://doi.org/10.1109/ICRA.2019.8793975)
- C. H. Lampert, H. Nickisch, S. Harmeling, "Attribute-based classification for zero-shot visual object categorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 453–465, 2014, doi: [10.1109/TPAMI.2013.140](https://doi.org/10.1109/TPAMI.2013.140)
- K. Jang *et al.*, "Simulation to scaled city: zero-shot policy transfer for traffic control via autonomous vehicles," in *Proceedings of the 10th ACM/IEEE International Conference on Cyber-Physical Systems*, 2019, pp. 291–300, doi: [10.1145/3302509.3313784](https://doi.org/10.1145/3302509.3313784)
- N. O. Salscheider, "FeatureNMS: Non-maximum suppression by learning feature embeddings," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 7848–7854, doi: [10.1109/ICPR48806.2021.9412930](https://doi.org/10.1109/ICPR48806.2021.9412930)
- D. Bahdanau, K. Cho, Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2015. doi: [10.48550/arXiv.1409.0473](https://doi.org/10.48550/arXiv.1409.0473)
- V. I. Slyusar, "A family of face products of matrices and its properties," *Cybern. Syst. Anal.*, vol. 35, no. 3, pp. 379–384, 1999, doi: [10.1007/BF02733426](https://doi.org/10.1007/BF02733426)
- V. Slyusar *et al.*, "Improvement of the object recognition model on aerophots using deep convolutional neural network," *Eastern-European J. Enterp. Technol.*, vol. 5, no. 2 (113), pp. 6–21, 2021, doi: [10.15587/1729-4061.2021.243094](https://doi.org/10.15587/1729-4061.2021.243094)
- K. Liu, L. Peng, S. Tang, "Underwater object detection using TC-YOLO with attention mechanisms," *Sensors*, vol. 23, no. 5, p. 2567, 2023, doi: [10.3390/s23052567](https://doi.org/10.3390/s23052567)
- V. I. Slyusar, "Архітектурно-математичні основи удосконалення нейронних мереж з класифікації зображень," *Штучний інтелект*, Т. 27, № 1, с. 245–258, 2022, doi: [10.15407/jai2022.01.245](https://doi.org/10.15407/jai2022.01.245)
- V. Slyusar, "The role of Artificial Intelligence in cross-platform tailoring of AR data," in *VIII International Scientific and Practical Conference*, 2020, doi: [10.13140/RG.2.2.22122.13760](https://doi.org/10.13140/RG.2.2.22122.13760)
- M. O. D. UK, "SAPIENT Interface Control Document. DSTL/PUB145591, 01-Feb-2023," 2023.
- M. O. D. UK, "SAPIENT autonomous sensor system. Last updated 20 April 2023," 2023.
- N. Barman, N. Khan, and M. G. Martini, "Analysis of Spatial and Temporal Information Variation for 10-Bit and 8-Bit Video Sequences," *2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, pp. 1–6, IEEE, 2019, doi: [10.1109/CAMAD.2019.8858486](https://doi.org/10.1109/CAMAD.2019.8858486)
- A. Mauri, R. Khemmar, B. Decoux, M. Haddad, R. Bouteau, "Real-time 3D multi-object detection and localization based on deep learning for road and railway smart mobility," *J. Imaging*, vol. 7, no. 8, p. 145, 2021, doi: [10.3390/jimaging7080145](https://doi.org/10.3390/jimaging7080145)
- V. I. Slyusar, "Концепция виртуализации поля боя 2050 года," *Озброєння та військова техніка*, № 3, с. 111–112, 2021.

Подано до редакції 12.07.2022

Після доопрацювання 12.11.2022

Прийнята до публікації 08.12.2022

INFORMATION ON THE ARTICLE

APPLICATION OF NEURAL NETWORK TECHNOLOGIES FOR UNDERWATER MUNITIONS DETECTION

Vadym I. Slyusar, ORCID: 0000-0002-2912-3149, swadim@ukr.net, Central Scientific Research Institute of Armament and Military Equipment of Armed Forces of Ukraine, Ukraine

In the article, the substantiated proposals for the use of YOLO family neural networks to detect the underwater undetonated munitions are proposed. At the same time, the YOLO3, YOLO4 and YOLO5 neural networks previously trained on the MS COCO dataset are used. The retraining of YOLO3 and YOLO4 neural networks is carried out on the modified Trash-ICRA19 underwater trash dataset, with the number of object classes equal to 13 and 2 of them are fictitious. The average class detection accuracy of 13 object classes using YOLO4 in the mAP50 metric is equal to 75.2% or 88.9% taking into account fictitious classes. The images obtained from video recordings of the demining reservoirs process with the help of remotely operated underwater vehicles (ROV) are used to test neural networks. The improved neural network as a cascade of several serially connected YOLO-segments with multi-pass image processing and tensor-matrix description of the attention mechanism are proposed. The recommendations for further increasing the efficiency of the neural network method of underwater munition selection are developed.

Keywords: remotely operated underwater vehicle; ROV; neural network; YOLO3; YOLO4; YOLO5; object detection; multi-pass processing

REFERENCES

- Espresso.tv, "Pyrotechnicians demine reservoirs in the Kyiv region," *Youtube channel Espresso.tv*, 2022. <https://www.youtube.com/watch?v=TxL8MQhBWnU>.
- 5th Channel, "Attributes of the 'Russian world': how sappers demine the village of Gorenka near Kyiv," *Youtube channel 5.tv*, 2022. <https://www.youtube.com/watch?v=Id4nWc4apTQ>.
- V. Slyusar et al., "Improving the model of object detection on aerial photographs and video in unmanned aerial systems," *Eastern-European J. Enterp. Technol.*, vol. 1, no. 9(115), pp. 24–34, 2022, doi: [10.15587/1729-4061.2022.252876](https://doi.org/10.15587/1729-4061.2022.252876).
- M. S. Fulton, J. Hong, J. Sattar, "Trash-ICRA19: A Bounding Box Labeled Dataset of Underwater Trash," 2020. doi: [10.13020/x0qn-y082](https://doi.org/10.13020/x0qn-y082).
- M. Fulton, J. Hong, M. J. Islam, J. Sattar, "Robotic detection of marine litter using deep visual detection models," *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 5752–5758, doi: [10.1109/ICRA.2019.8793975](https://doi.org/10.1109/ICRA.2019.8793975).
- C. H. Lampert, H. Nickisch, S. Harmeling, "Attribute-based classification for zero-shot visual object categorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 453–465, 2014, doi: [10.1109/TPAMI.2013.140](https://doi.org/10.1109/TPAMI.2013.140).
- K. Jang et al., "Simulation to scaled city: zero-shot policy transfer for traffic control via autonomous vehicles," *Proceedings of the 10th ACM/IEEE International Conference on Cyber-Physical Systems*, 2019, pp. 291–300, doi: [10.1145/3302509.3313784](https://doi.org/10.1145/3302509.3313784).
- N. O. Salscheider, "FeatureNMS: Non-maximum suppression by learning feature embeddings," *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 7848–7854, doi: [10.1109/ICPR48806.2021.9412930](https://doi.org/10.1109/ICPR48806.2021.9412930).
- D. Bahdanau, K. Cho, Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2015. doi: [10.48550/arXiv.1409.0473](https://doi.org/10.48550/arXiv.1409.0473).
- V. I. Slyusar, "A family of face products of matrices and its properties," *Cybern. Syst. Anal.*, vol. 35, no. 3, pp. 379–384, 1999, doi: [10.1007/BF02733426](https://doi.org/10.1007/BF02733426).
- V. Slyusar et al., "Improvement of the object recognition model on aerophotos using deep convolutional neural network," *Eastern-European J. Enterp. Technol.*, vol. 5, no. 2(113), pp. 6–21, 2021, doi: [10.15587/1729-4061.2021.243094](https://doi.org/10.15587/1729-4061.2021.243094).
- K. Liu, L. Peng, S. Tang, "Underwater object detection using TC-YOLO with attention mechanisms," *Sensors*, vol. 23, no. 5, p. 2567, 2023, doi: [10.3390/s23052567](https://doi.org/10.3390/s23052567).
- V. Slyusar, "Architectural and mathematical fundamentals of improvement neural networks for classification of images," *Artif. Intell.*, vol. 27, no. jai2022.27(1), pp. 245–258, 2022, doi: [10.15407/jai2022.01.245](https://doi.org/10.15407/jai2022.01.245).
- V. Slyusar, "The role of Artificial Intelligence in cross-platform tailoring of AR data," *VIII International Scientific and Practical Conference*, 2020, doi: [10.13140/RG.2.2.22122.13760](https://doi.org/10.13140/RG.2.2.22122.13760).
- M. O. D. UK, "SAPIENT Interface Control Document. DSTL/PUB145591, 01-Feb-2023," 2023. uri: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1144352/SAPIENT_Interface_Control_Document_v7_FINAL_fixed2.pdf.
- M. O. D. UK, "SAPIENT autonomous sensor system. Last updated 20 April 2023," 2023. uri: <https://www.gov.uk/guidance/sapient-autonomous-sensor-system>.
- N. Barman, N. Khan, M. G. Martini, "Analysis of spatial and temporal information variation for 10-bit and 8-bit video sequences," *2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, 2019, pp. 1–6, doi: [10.1109/CAMAD.2019.8858486](https://doi.org/10.1109/CAMAD.2019.8858486).
- A. Mauri, R. Khemmar, B. Decoux, M. Haddad, R. Boutheau, "Real-time 3D multi-object detection and localization based on deep learning for road and railway smart mobility," *J. Imaging*, vol. 7, no. 8, p. 145, 2021, doi: [10.3390/jimaging7080145](https://doi.org/10.3390/jimaging7080145).
- V. I. Slyusar, "2050 battlefield virtualization concept," *Weapons Mil. Equip.*, no. 3, pp. 111–112, 2021, doi: [10.34169/2414-0651.2021.3\(31\).111-112](https://doi.org/10.34169/2414-0651.2021.3(31).111-112).

Received July 12, 2022

Revised November 12, 2022

Accepted December 8, 2022