

ТЕХНОЛОГІЯ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ВІДЕОПОТОКУ ДЛЯ АВТОМАТИЧНОГО РОЗПІЗНАВАННЯ ЦІЛЕЙ СИСТЕМИ КЕРУВАННЯ ВОГНЕМ НА ОСНОВІ МАШИННОГО НАВЧАННЯ

Висоцька В. А. – д-р техн. наук, доцент, доцент кафедри «Інформаційні системи та мережі», Національний університет «Львівська політехніка», Львів, Україна.

Романчук Р. В. – аспірант кафедри «Інформаційні системи та мережі», Національний університет «Львівська політехніка», Львів, Україна.

АНОТАЦІЯ

Актуальність. Розпізнавання цілей є пріоритетом у військовій справі. Це завдання ускладнюється тим, що необхідно розпізнавати рухомі об'єкти, різний рельєф і ландшафт створюють перешкоди для розпізнавання. Бойові дії можуть відбуватися в різний час доби, відповідно, необхідно враховувати ракурс освітлення та загальне освітлення. Необхідно виявити об'єкт на відео шляхом сегментації відеокадрів, розпізнати та класифікувати його.

Мета – розроблення технології аналізу розроблення технології розпізнавання цілей в реальному масштабі часу, як складової системи управління вогнем, за рахунок використання штучного інтелекту, YOLO та машинного навчання.

Метод. У статті розроблено технологію аналізу відеопотоку для автоматичного розпізнавання цілей системи керування вогнем на основі машинного навчання. У роботі запропоновано розробку модуля розпізнавання цілей як складової системи управління вогнем в рамках запропонованої інформаційної технології з використанням штучного інтелекту. Сімейство моделей розпізнавання образів YOLOv8 використано для розробки модуля розпізнавання цілей. Методи, які використано під час дослідження сформованого датасету.

– Bounding Box: Noise – Up to 15% of pixels (обмежувальна рамка: додавання до зображення шуму «солі та перцю» – до 15% пікселів).

– Bounding Box: Blur – Up to 2.5px (обмежувальна рамка: додавання до зображення розмиття за Гаусом – до 2,5 пікселів).

– Cutout – 3 boxes with 10% size each (вирізати частину зображення – 3 коробки по 10% розміру кожна).

– Brightness – Between –25% and +25% (зміна яскравості зображення для збільшення стійкості моделі до зміни освітлення та налаштувань камери – від –25% до +25%).

– Rotation – Between –15° and +15° (обертання об'єкта зображення – за або проти годинникової стрілки на ступінь від –15° до +15°).

– Flip – Horizontal (перевертання об'єкта зображення горизонтально).

Результати. Дані зібрані з відкритих джерел, зокрема, з відеозаписів, розміщених у відкритих джерелах на платформі YouTube. Основним завданням попередньої обробки даних є класифікація трьох класів об'єктів на відео або в режимі реального часу – БТР, БМП і ТАНК. Набір даних формується за допомогою платформи Roboflow на основі інструментів маркування та, згодом, інструментів доповнення. Набір даних складається з 1193 унікальних зображень – приблизно порівну для кожного класу. Тренінг проводився з використанням ресурсів Google Colab. На навчання моделі пішло 100 epoch.

Висновки. Аналіз проводиться відповідно до показників mAP50 (середня точність 0,85), mAP50-95 (0,6), точності (0,89) і запам'ятовування (0,75). Великі втрати є через те, що під час дослідження не було враховано фон – навчання модуля на основі підтверджених даних (зображень) фону без техніки.

КЛЮЧОВІ СЛОВА: розпізнавання рухомих об'єктів, безпека, конфіденційність, YOLO, ідентифікація цілей, машинне навчання, БТР, БМП, ТАНК.

АБРЕВІАТУРА

БМП – бойова машина піхоти;

БПЛА – безпілотний літальний апарат;

БТР – бронетранспортер;

ATRU – automatic target recognition unit;

YOLO – you only live once (нейронна мережа).

НОМЕНКЛАТУРА

M – модуль спостереження для автоматичного розпізнавання та ідентифікації цілей системи керування вогнем;

I – множина вхідних даних у вигляді відеопотоку з відеокамери, $I = \{i_1, i_2, i_3, i_4\}$;

O – множина вихідних даних у вигляді розпізнавання та ідентифікації об'єктів військової техніки, $O = \{o_1, o_2, o_3\}$;

R – основні правила опрацювання вхідних даних відеопотоку, $R = \{r_1, r_2, r_3, r_4, r_5\}$;

U – параметри опрацювання вхідних даних відеопотоку, $U = \{u_1, u_2, u_3, u_4, u_5\}$;

N – нейронна мережа навчання розпізнавання, ідентифікації та класифікації об'єктів військової техніки типу ТАНК/БМП/БТР;

α – оператор аналізу та розкадровки вхідних даних відеопотоку;

β – оператор опрацювання зображень через сегментування та аналіз сегментованих об'єктів;

γ – оператор розпізнавання, ідентифікації та класифікації об'єктів військової техніки типу ТАНК/БМП/БТР;

μ – оператор розпізнавання потенційних будь-яких об'єктів на зображенні (будівлі, мостів, військової техніки тощо);

i_1 – множина даних із відеопотоку та зображень оригіналу;

i_2 – множина зображень військової техніки;

i_3 – множина даних фонів пейзажів місцевості;

i_4 – словники валідованих зображень військової техніки;

r_1 – правила розкадрування відеопотоку на зображення;

r_2 – правила сегментування зображення;

r_3 – правила аналізу сегментів зображення;

r_4 – правила машинного навчання нейронної мережі ідентифікації військової техніки;

r_5 – правила класифікації військової техніки;

u_1 – множина умов формування зображень з відеопотоку;

u_2 – множина вимог аналізу зображень, в тому числі фільтрування від шуму;

u_3 – множина умов опрацювання зображень;

u_4 – множина умов розпізнавання та ідентифікації об'єктів військової техніки;

u_5 – множина вимог класифікації розпізнаних об'єктів військової техніки;

χ – оператор розпізнавання потенційних об'єктів військової техніки на зображенні;

o_1 – множина розпізнаних всіх об'єктів на зображенні;

o_2 – множина розпізнаних всіх об'єктів військової техніки на зображенні;

o_3 – множина ідентифікованих всіх об'єктів військової техніки на зображенні;

ω – оператор ідентифікації розпізнаного об'єкту військової техніки на множині розкадрованих з відеопотоку зображеннях;

λ – оператор класифікації ідентифікованого об'єкту військової техніки на множині розкадрованих з відеопотоку зображеннях.

ВСТУП

Сьогодні провідні армії світу прагнуть підвищити можливості своїх основних зразків техніки та озброєння за рахунок модернізації існуючого парку, або розроблення новітніх зразків. Автоматичне розпізнавання цілей ATRU полягає у здатності алгоритму або пристрою розпізнавати цілі або об'єкти на основі даних отриманих від сенсорів, в тому числі від відеоспостереження, наприклад від БПЛА, таких як дрони, або від відео-реєстраторів на бойових машинах. З іншого боку у зв'язку зі збільшенням використання БПЛА для розвідки ворогом безпека та конфіденційність багатьох критично важливих місць можуть бути скомпрометовані. Тому вони також є законною ціллю для розпізнавання системою керування вогнем.

Інформаційна технологія розпізнавання цілей є основним складовим в крилатих ракетах та БПЛА, а

також для розроблення бойових роботів або роботів-саперів. Автоматичне розпізнавання цілей застосовують не лише у військових справах, але наприклад в організації пошуку людей/об'єктів (в морі, в зоні стихійних лих, пожеж тощо).

Задача автоматичного розпізнавання цілей в умовах бойових дій ускладнюється декількома чинниками, зокрема:

– можливий рух об'єкта, який розпізнається;

– рух об'єкта (бойової машини або БПЛА), звідки йде відеоспостереження та подальше розпізнавання цілей;

– різні погодні умови;

– різний рельєф місцевості та ландшафт, в тому числі лісові смуги;

– присутність інших об'єктів, які потенційно не є цілями (будівлі, підбиті/знищені бойові машини, частини споруд як мости тощо);

– освітлення;

– потенційно об'єкт, який розпізнається, не є ворогом;

– частина об'єкту, який розпізнається, сховано за перешкодами;

– кут спостереження для різних об'єктів різний (для БПЛА зверху вниз, для бойових машин не лише вперед/навколо, але вверх для наприклад БПЛА).

Як бачимо, умови освітлення, різні розміри об'єктів, рухомий фон та різноманітні фонові контрасти значно впливають на якість, оперативність та швидкість розпізнавання об'єктів з відеоспостереження [1–2]. Необхідно не лише виявити об'єкт на відео, сегментувавши відео-кадри, а і розпізнати та класифікувати його (наприклад, дрон чи птах, машина чи будівля тощо), і це все зазвичай під час руху як об'єкта спостереження, так і об'єкта-спостерігача при несприятливих умовах у реальному часі. Виявлення літаючих об'єктів або об'єктів, які рухаються в несприятливих умовах, на відео відрізняється від стандартного виявлення об'єктів, оскільки розмір нерухомого/рухомого/літаючого і/або частково захованого об'єкта за іншим об'єктом постійно змінюється в кадрах залежно від його відстані та руху об'єкта-спостерігача. Він має такі проблеми, як низька роздільна здатність, зміни освітлення через день і ніч і нестабільний фон, різні погодні умови. Також точність розпізнавання залежить від якості камери спостереження, вибір якої під час бойових дій не контрольованим процесом. Складність спостереження збільшується при розпізнаванні від 2D (попереду перед камерою) до 3D (зверху з дна під різними кутами на різних висотах) рухомих об'єктів з врахуванням масштабування та пропорцій. Аналогічно зменшує точність розпізнавання об'єктів, які між собою візуально подібні та відрізняються малими ознаками або їх відсутності при спостереженні в різних ракурсах або при частково схованому корпусі за іншими природними об'єктами або будівлями (наприклад, деякі модифікації танків серії Т). Отже, виявлення та

розпізнавання нерухомих/рухомих/ літаючих/ рухливих і/або частково захованих об'єктів за іншими об'єктами в стані нерухомоті/рухомоті спостерігача в різних погодних умовах, ландшафтах, освітленні та на різній висоті мають великий обсяг спостереження та високу мобільність. Існує сильна потреба в таких програмах у реальному світі через відмінності їх розміру в межах одного типу об'єкта та просторової роздільної здатності датчика.

Метою дослідження є розроблення технології розпізнавання цілей в реальному масштабі часу, як складової системи управління вогнем, за рахунок використання штучного інтелекту. Для досягнення мети були поставлені такі завдання:

– визначити ефективний підхід інтелектуального аналізу відеопотоку для автоматичного розпізнавання цілей системи керування вогнем на основі машинного навчання;

– провести експериментальне дослідження методів та інструментів інтелектуального аналізу відеопотоку для автоматичного розпізнавання цілей системи керування вогнем на основі машинного навчання;

– розробити метод розрахунку вірогідності розпізнавання цілей системи керування вогнем;

– здійснити аналіз результатів експериментальної апробації запропонованої технології аналізу відеопотоку для автоматичного розпізнавання цілей системи керування вогнем на основі машинного навчання.

Об'єкт дослідження – процеси аналізу відеопотоку для автоматичного розпізнавання цілей системи керування вогнем. Предмет дослідження – методи та засоби аналізу відеопотоку для автоматичного розпізнавання цілей системи керування вогнем.

1 ПОСТАНОВКА ПРОБЛЕМИ

За останні роки під час повномасштабної війни в Україні з поступовим вдосконаленням технології керування дронами зображення та відео дистанційного зондування БПЛА стали важливим джерелом оперативних даних. Паралельно оснащення бойових машин відео-регістраторами для відео-спостереження з елементами штучного інтелекту та машинного навчання для розпізнавання об'єктів реальному часі дозволить збільшити рівень безпеки учасникам бойових дій при відповідному своєчасному реагуванні на результати розпізнавання цілей.

Відео-кадри → сегментація відеокадру → виявлення потенційних об'єктів → виявлення об'єктів в русі → розпізнавання об'єктів як потенційно небезпечних → класифікація об'єктів → ідентифікація об'єкту.

Модуль спостереження для автоматичного розпізнавання та ідентифікації цілей системи керування вогнем M подано імітаційною моделлю через кортеж:

$$M = \langle I, O, R, U, N, \alpha, \beta, \gamma \rangle.$$

Основними процесами моделі спостереження для автоматичного розпізнавання та ідентифікації цілей системи керування вогнем є «Опрацювання відеопотоку», «Опрацювання зображень», «Машинне навчання» та «Класифікація об'єктів».

Процес «Опрацювання відеопотоку» опишемо суперпозицією:

$$C_{AU} = \mu \circ \beta \circ \alpha, C_{AU} = \mu(\beta(\alpha(i_1, i_2, i_3), r_1, u_1), u_2).$$

Процес «Опрацювання зображень» опишемо суперпозицією:

$$C_{CU} = \chi \circ \beta \circ \alpha, \text{ тобто} \\ C_{CU} = \chi(\beta(\alpha(C_{AU}, i_2, i_3, i_4), o_1, r_2, u_3), r_3).$$

Процес «Машинне навчання» опишемо як:

$$C_{UL} = \omega \circ \gamma \circ \beta \circ \alpha, \\ C_{UL} = \omega(\gamma(\beta(\alpha(C_{CU}, i_1), o_2, i_4), u_4), r_4).$$

Процес «Класифікація об'єктів» опишемо як:

$$C_{US} = \lambda \circ \gamma \circ \beta \circ \alpha, \\ C_{US} = \lambda(\gamma(\beta(\alpha(C_{US}, i_1), o_3, i_4), u_5), r_5).$$

Аналіз здійснюється за допомогою класифікації чи кластеризації, що сегменти за певними критеріями. Хоч збір інформації й відбувається автоматично, проте все ще необхідна реалізація таких досліджень згідно розпізнавання, ідентифікації та класифікації об'єктів в несприятливих умовах в русі та при поганій якості зображення, та відповідна обробка результатів. Також суттєво впливає на результати дослідження ефективність опрацювання відповідного фону та об'єктів на ньому (наприклад зелене на фоні зеленому або частина об'єкта сховане за іншим об'єктом). Ще одним з найважливіших критеріїв подібної технології є можливість збору даних в залежності від періоду доби та пори року, та їх періодичність із-за зміни фону із-за результатів ведення активних бойових дій в певній місцевості.

2 АНАЛІЗ ЛІТЕРАТУРНИХ ДЖЕРЕЛ

Сьогодні зазвичай застосовують нейронні мережі та глибоке навчання для таких завдань, як сегментація зображень [3–5], виявлення об'єктів [6–8] і класифікація зображень [9–11]. Більшість застосовуваних на даний момент глибоких нейронних мережевих моделей, таких як PSPNET [4], U-NET [5], RESNET [10], VGG [11], розроблені на основі зібраних вручну наборів даних зображень (не відео) при сприятливих умовах, таких як MS-COCO [12], VOC2012 [13], VOC2007 [14].

Існує два загальні сценарії застосування методів виявлення об'єктів дистанційного зондування дронами

або на основі відео-спостереження з машини, зокрема опрацювання даних передбачається:

– після польоту/поїздки за допомогою стаціонарних комп'ютерів (потрібна висока точність виявлення та ідентифікації).

– в режимі реального часу під час польоту/поїздки, коли вбудований комп'ютер на дроні або в машині відповідно синхронно опрацьовує дані відео у режимі реального часу. Параметри моделі повинні бути в межах певного масштабу, щоб відповідати вимогам до роботи вбудованого обладнання. Після виконання робочих умов точність виявлення методу також повинна бути якомога вищою.

Тому застосовані методи виявлення об'єктів на основі нейронної мережі повинні відповідати різним вимогам для кожного сценарію.

Таким чином, методи нейронної мережі для виявлення об'єктів на відео дистанційного зондування безпілотників або при відео-спостереженні в бойових машинах повинні мати можливість адаптуватися до конкретних характеристик цих даних. Вони повинні бути розроблені відповідно до вимог опрацювання даних після польоту/поїздки, що може забезпечити високу точність і швидкість відкликання, або вони повинні бути розроблені як моделі з параметрами меншого масштабу, які можна розгортати у вбудованих апаратних середовищах для опрацювання в реальному часі на дрони/машини. У цій роботі ми пропонуємо застосування нейронної мережі на основі архітектури YOLOv8 для автоматичного розпізнавання об'єктів як потенційних цілей системи керування вогнем.

На сьогодні запропоновано численні методи виявлення об'єктів на основі нейронних мереж, зокрема за допомогою серії YOLO [15–22]. На відміну від двоетапних методів, одноетапний метод поєднує визначення розташування об'єкта та класифікацію в один крок, досягаючи виявлення об'єкта в реальному часі як на робочому столі, так і на вбудованому обладнанні. Ці методи не тільки досягають хороших результатів ідентифікації, але також пропонують низку вдосконалень у таких областях, як методи збільшення даних навчання, методи навчання мережі, функції втрат, функції активації та структури мережеских моделей.

В [18] запропоновано YOLOX, модель нейронної мережі з одноетапним виявленням об'єктів. В [19] автори запропонували модель нейронної мережі з одноетапним виявленням цілі. В роботі [20] автори дослідили оптимальну швидкість і точність виявлення об'єктів на основі YOLOv4. Автори в [23] описали CSPDarkNet як магістральну структуру мережі, покращуючи здатність до навчання згорткових нейронних мереж, дозволяючи мережі підтримувати точність вилучення карти ознак. У роботі [24] автори запропонували метод CrowdDet на основі нейронної мережі для виявлення щільних і взаємно закритих цілей на зображеннях. У шийній частині мережі введено модуль SPPF і модуль PAFPN [25]. Автор все © Висоцька В. А., Романчук Р. В., 2024
DOI 10.15588/1607-3274-2024-3-7

ще використовують CSPDarkNet [23] як магістральну мережу, але вводять SiLU як функцію активації, яка вирішує проблему градієнтної дисперсії, коли вхід функції ReLU негативний, а вихід дорівнює 0 [26–27].

Існує багато досліджень на основі різних версій YOLO, але досі залишається найбільш важливою проблема при ідентифікації об'єктів ефективне виявлення малих об'єктів та точність класифікації різних рухливих об'єктів при різних умовах навколишнього середовища (наприклад об'єкт в різних градаціях зеленого кольору забарвлення на фоні також різного спектру зеленого кольору) в відео потоці різної якості зображення при подальшому розкадровці та сегментуванні відповідних зображень для ідентифікації та класифікації.

Зображення дистанційного зондування часто мають великі розміри, складний фон і значну присутність дрібних об'єктів. Запропоноване рішення зосереджено на оптимізації точного виявлення малих об'єктів на відстані та об'єктів в русі при різних умовах навколишнього середовища. Перевага мереж серії YOLO полягає у використанні багаторівневих головок виявлення, що дозволяє виявляти об'єкти різного розміру з різних рівнів векторів ознак. Наш підхід головним чином зосереджений на виявленні маленьких об'єктів, а також рухливих об'єктів за допомогою векторів ознак із нижніх шарів, які мають вищу просторову роздільну здатність. Для досягнення цієї мети ми використовуємо модуль машинного навчання для оптимізації їхніх семантичних характеристик. Головки виявлення можуть отримувати вектори ознак як з високою просторовою роздільною здатністю, так і з точною семантичною інформацією, підвищуючи таким чином загальну точність ідентифікації.

Під час глибокого навчання методи вилучення ознак SIFT (масштабно-інваріантне перетворення ознак) і HOG (гістограми орієнтованих градієнтів) виконували це завдання, застосовуючи деякі алгоритми машинного навчання поверх класифікатора. Частина методів глибокого навчання застосовують для кольорових зображень, а інші – до ІЧ-зображень. Опрацювання ІЧ-зображень у режимі реального часу є простішою, оскільки вимагає менших обсягів пам'яті та обчислювальної потужності. Також не впливає на різні умови освітлення. Але зібрати навчальний датасет для деяких предметних областей (наприклад військові техніки під час війни) практично не можливо. Алгоритми сімейства YOLO, засновані на архітектурі CNN, є широко використовуваними і добре відомими алгоритмами для вирішення проблем виявлення об'єктів. YOLO v4 і YOLO v5 – це в основному вживані моделі. YOLO v4, будучи модифікованою версією YOLO3, використовує межетапну часткову мережу (CSPNet) у Darknet, створюючи нову магістраль екстрактора функцій під назвою CSPDarknet53. Щоб підвищити ефективність алгоритму, YOLOv4 використовує мішок

безкоштовних і мішок спеціальних пропозицій. Повна втрата IOU (CIOU), регуляризація блокування випадання та багато підходів до розширення. Активація Mish, Diou-NMS і модифіковані мережі агрегації шляхів включені в пакет спеціальностей. Але YOLOv5 відрізняється від попередніх версій. Тут замість Darknet використовується PyTorch. Він використовує CSPDarknet53 як структурну підтримку. Ця магістраль усуває надлишкову інформацію про градієнт, яку можна побачити у великих магістралях, і включає зміну градієнта в карти функцій, що пришвидшує висновок, покращує точність і зменшує розмір моделі за рахунок зменшення кількості параметрів. Він посилює потік інформації, використовуючи мережу агрегації шляхів (PANet), що призводить до трьох різних виходів карти функцій для багатомасштабного прогнозування. Це покращує здатність моделі ефективно передбачати дрібні та великі предмети. Зображення надсилається до PANet для об'єднання функцій після введення в CSPDarknet53 для видалення функцій. Швидкість обробки YOLO v4 і v5 коливається від 45 до 150 кадрів в секунду. Однак, на відміну від швидшого R-CNN, він має меншу похибку запам'ятовування та вищу локалізацію. Оскільки кожна сітка може запропонувати лише дві обмежувальні рамки, вона також має проблеми з виявленням предметів поблизу та дрібних об'єктів. Останнім доповненням до сімейства засобів виявлення об'єктів YOLO є модель YOLO v8. Це найшвидший і найточніший на сьогодні детектор об'єктів у реальному часі. Всі інструменти на основі YOLOv8 перевершують попередні детектори об'єктів за швидкістю та точністю.

3 МАТЕРІАЛИ ТА МЕТОДИ

У роботі обговорюється гібридний підхід із використанням CNN-LSTM для покращення продуктивності виявлення військової техніки рашки з рухомим фоном і різними відстанями, а також її ТАНК/БТР/БМП. Основні внески статті.

1. Збір зображень із відкритих джерел за платформи ютуб.

2. Гібридна модель CNN-LSTM з гіперпараметричним налаштуванням за допомогою байєсівської оптимізації для виявлення об'єктів.

3. Детальний аналіз моделі YOLOV8 на різних діапазонах зображень і визначення їх точності з певним значенням довіри.

Алгоритми виявлення об'єктів у глибокому навчанні в основному поділяються на регіональні та регресійні. Основним завданням виявлення об'єктів військової техніки рашки є виявлення об'єкта в кадрі, де присутні об'єкти різних цільових класів; отже, класифікація об'єктів є необхідною умовою для виявлення об'єктів за допомогою обмежувальної рамки.

1. Алгоритми на основі регресії використовуються переважно для виявлення об'єктів у реальному часі. Це одноетапні структури, засновані на глобальній © Висоцька В. А., Романчук Р. В., 2024
DOI 10.15588/1607-3274-2024-3-7

регресії/класифікації, які прямо відображають пікселі зображення на координати обмежувальної рамки, зменшуючи витрати часу. Однією з найшвидших моделей розпізнавання об'єктів є YOLO, яка може аналізувати кадри зі швидкістю до 150 FPS для невеликих мереж. Хоча YOLO не є найточнішою моделлю з точки зору середньої середньої точності (mAP), вона показала себе достатньо добре під час навчання.

У частині виявлення об'єктів військової техніки проводився експеримент з автоматичного розпізнавання та ідентифікації цілей системи керування вогнем. Це було реалізовано за допомогою різних моделей виявлення об'єктів. Експеримент зосереджений на реалізації алгоритму YOLO v8 для порівняння з іншими версіями.

2. Алгоритми на основі регіональних пропозицій. Після появи відповідності CNN запропонована регіональна згорткова нейронна мережа (R-CNN). У порівнянні з моделями без глибоких CNN, R-CNN значно підвищив ефективність виявлення щодо середньої середньої точності (mAP). Він має кілька недоліків, включаючи дороге навчання з точки зору грошей і часу, і, найгірше, високу затримку (час виявлення). Спираючись на роботу R-CNN, швидкий R-CNN підвищує точність, прискорюючи навчання та тестування. Швидкий R-CNN різко скорочує час навчання та тестування, однак регіональні пропозиції все ще створюються з використанням традиційних методів, які потребують багато часу для попередньої обробки. Швидкий R-CNN пропонується як вирішення проблеми вузького місця з пропозиціями регіонів, які роблять пропозиції регіонів через нейронну мережу. Швидкий/швидкий R-CNN та інші детектори об'єктів, які використовують мережі регіональних пропозицій, продемонстрували передову продуктивність у багатьох тестах. Однак вони не завжди успішно знаходять дрібні предмети. Поточні підходи гірші щодо повторюваності та узагальнення, коли обставини реального світу постійно змінюються, оскільки вони залежать від конкретних даних зображення.

3. Класифікація нерухомого/рухомого/літаючого і/або частково захованого об'єкта військової техніки за іншим об'єктом при несприятливих умовах у реальному часі. Виявлення літаючих об'єктів або об'єктів, які рухаються в несприятливих умовах, на відео відрізняється від стандартного виявлення об'єктів, оскільки розмір нерухомого/рухомого/літаючого і/або частково захованого об'єкта за іншим об'єктом постійно змінюється в кадрах залежно від його відстані та руху об'єкта-спостерігача. Він має такі проблеми, як низька роздільна здатність, зміни освітлення через день і ніч і нестабільний фон, різні погодні умови. Також точність розпізнавання залежить від якості камери спостереження, вибір якої під час бойових дій не контролюється процесом. Складність спостереження збільшується при розпізнаванні 2D

(попереду перед камерою) до 3D (зверху з дна під різними кутами на різних висотах) рухомих об'єктів з врахуванням масштабування та пропорцій. Аналогічно зменшує точність розпізнавання об'єктів, які між собою візуально подібні та відрізняються малими ознаками або їх відсутності при спостереженні в різних ракурсах або при частково схованому корпусі за іншими природними об'єктами або будівлями.

4. Методи виявлення об'єктів на основі глибокого навчання для різних задач, які включають виявлення смуги, інтелектуальні системи транспортних засобів, виявлення рухомих об'єктів, в тому числі військової техніки тощо.

Аналіз здійснюється за допомогою класифікації чи кластеризації, що сегменти за певними критеріями. Хоч збір інформації й відбувається автоматично, проте все ще необхідна реалізація таких досліджень згідно розпізнавання, ідентифікації та класифікації об'єктів в несприятливих умовах в русі та при поганій якості зображення, та відповідна обробка результатів. Також суттєво впливає на результати дослідження ефективність опрацювання відповідного фону та об'єктів на ньому (наприклад зелене на фоні зеленому або частина об'єкта сховане за іншим об'єктом). Ще одним з найважливіших критеріїв подібної технології є можливість збору даних в залежності від періоду доби та пори року, та їх періодичність із-за зміни фону із-за результатів ведення активних бойових дій в певній місцевості.

4 ЕКСПЕРИМЕНТИ

В роботі авторами пропонується розробка модуля розпізнавання цілей, як складової системи управління вогнем в рамках запропонованої інформаційної технології за рахунок використання штучного інтелекту.

Для розробки модуля розпізнавання цілей використано сімейство моделей для розпізнавання моделей YOLOv8. Це найновіша версія популярного продукту компанії Ultralytics для виявлення об'єктів та сегментації зображень у реальному часі. YOLOv8 поставляється в комплекті з такими попередньо підготовленими моделями:

– Моделі класифікації зображень, що попередньо навчаються на базі даних ImageNet з роздільною здатністю зображення 224.

– Контрольні точки сегментації екземпляра, навчені на наборі даних сегментації COCO з роздільною здатністю зображення 640.

– Контрольні точки виявлення об'єктів, що навчаються на основі набору даних COCO detection з роздільною здатністю зображення 640.

– Виведення виконується зі швидкістю майже 105 кадрів в секунду на графічному процесорі середньостатистичного сучасного ноутбука, а адвельдіка модель працює в середньому зі швидкістю 17 кадрів за секунду.

YOLOv8 використовує фреймворк PyTorch – фреймворк для розроблення глибоких нейронних мереж від компанії Facebook. YOLOv8 має ряд переваг у порівнянні з іншими засобами, зокрема:

– Підтримка більшістю сервісів надання обчислювальних потужностей.

– Велика кількість способів застосування і використання моделей.

– Великий рівень точності, що підтверджено тестам на наборах даних COCO та Roboflow 100.

Як зазначалось вище, YOLOv8 досягає високої точності на наборі даних COCO. Наприклад, модель YOLOv8m – досягає 50,2% mAP при вимірюванні на COCO. Під час порівняння з Roboflow 100, набором даних, який конкретно оцінює продуктивність моделі в різних областях, YOLOv8 отримав значно кращий результат, ніж YOLOv5. Крім того, YOLOv8 надає для розробників значний перелік функцій. На відміну від інших моделей, у яких завдання розподіляються між багатьма різними файлами Python, YOLOv8 постачається з CLI-інтерфейсом, який робить навчання моделі більш зрозумілим.

Для навчання використовувався сервіс Google Colaboratory, більш відомий як “Colab”. “Colab” – це хмарна версія Jupyter Notebook. Щоб використовувати Colab, не потрібно встановлювати та виконувати або оновлювати апаратне забезпечення комп'ютера, щоб відповідати вимогам Python до інтенсивного навантаження CPU/GPU. Крім того, Colab надає безкоштовний доступ до обчислювальної інфраструктури, як-от сховище, оперативна пам'ять, обчислювальні потужності, графічні процесори (GPU) і тензорні процесори (TPU).

Методи, які використано під час дослідження сформованого датасету.

– Flip – Horizontal (перевертання об'єкта зображення горизонтально).

– Rotation – Between -15° and $+15^\circ$ (обертання об'єкта зображення – за або проти годинникової стрілки на ступінь від -15° до $+15^\circ$).

– Brightness – Between -25% and $+25\%$ (зміна яскравості зображення для збільшення стійкості моделі до зміни освітлення та налаштувань камери – від -25% до $+25\%$).

– Cutout – 3 boxes with 10% size each (вирізати частину зображення – 3 коробки по 10% розміру кожна).

– Bounding Box: Blur – Up to 2.5px (обмежувальна рамка: додавання до зображення розмиття за Гаусом – до 2,5 пікселів).

– Bounding Box: Noise – Up to 15% of pixels (обмежувальна рамка: додавання до зображення шуму «солі та перцю» – до 15% пікселів).

Останні два пункти застосовують для розширення рівня обмежувальної рамки при формуванні/генеруванні нових навчальних даних, лише змінюючи вміст обмежувальних рамок вихідного зображення. Збільшення зображення – це

процес збільшення розміру набору даних шляхом маніпулювання наявними навчальними даними. Збільшення зображення допомагає моделі краще узагальнити для широкого спектру контекстів. Наприклад, можна змінити яскравість або темність об'єкта відносно його фону. Або, можливо, розмийте об'єкт відносно його фону для завдань, які часто включають зйомку об'єктів, що швидко рухаються. Модифікації лише обмежувальної рамки призводять до системних покращень, особливо для моделей, які були невеликими наборами даних (декілька тисяч фото). Можна також змінити кольори лише об'єктів на OCR-зображенні, запровадити розмиття лише для рухомих об'єктів, як-от військової техніки в різних спектрах зеленого, обертати об'єкти, як-от об'єкти у зоні спостереження згори, і перевернути орієнтацію об'єктів, щоб створити ефект дзеркального відображення, подібний до тих, що присутні в більшості ситуацій з камерою.

5 РЕЗУЛЬТАТИ

Перед розробленням модуля автоматичного розпізнавання цілей системи керування вогнем організовано робочий простір та визначені правила доступу до сховища даних та опрацювання відповідних даних з нього. Для навчання використовувався сервіс Google Colaboratory, більш відомий як "Colab". "Colab" – це хмарна версія Jupyter Notebook. Основною задачею препроцесінгу даних є класифікація трьох класів об'єктів на відео або у режимі реального часу – БТР, БМП, ТАНК. Далі проводилось створення та наповнення відповідного набору даних. Дані наповнювалися з відкритих джерел, зокрема з відеокліпів, розміщених у відкритих джерелах на YouTube платформі (відео з просування техніки рашки в перші дні війни по території України та з військових парадів російської техніки). Цей процес включав в себе пошук зображень та відео з вищеназваними об'єктами та розміткою відповідних об'єктів. Датасет сформований за допомогою Roboflow platform на основі засобів розмітки та в подальшому засобів аугментації. Набір даних складається з 1193 унікальних зображень – приблизно порівну для кожного класу. Після застосування попереднього опрацювання зображень та методів аргументації, набір даних має наступний вигляд (Таблиця 1):

Таблиця 1 – Розподіл даних в наборі даних

Тип набору	Абсолютне значення	Відносне значення
Train Set	2490	87%
Valid Set	225	8%
Test Set	138	5%

```
train: ../train/images  
val: ../valid/images  
test: ../test/images  
nc: 3  
names: ['bmp', 'btr', 'tank']
```

6 ОБГОВОРЕННЯ

Навчання проводилось із застосування ресурсів Google Colab. Було взято 100 епох для навчання моделі. Статистичні результати навчання нейронної мережі зображені на Рис. 1–2. Аналіз провадився за метриками mAP50 (mean Average Precision), mAP50-95, precision та recall (Рис. 1).

AP (Average precision) – це популярна метрика вимірювання точності детекторів таких об'єктів, як Faster R-CNN, SSD і т.д. Average Precision обчислює середню точність для recall в діапазоні від 0 до 1. Це показник точності моделі, що враховує лише "легкі" виявлення. mAP50-95: середнє значення середньої точності, розраховане при різних порогових значеннях IoU, що варіюються від 0,50 до 0,95. Воно дає повне уявлення про продуктивність моделі на різних рівнях складності виявлення.

Precision вимірює, наскільки точні ваші прогнози. Наприклад, який відсоток ваших прогнозів коректний. Recall вимірює, наскільки добре ви знаходите усі позитивні зразки. Наприклад, ми можемо знаходити 75% від усіх можливих позитивних випадків у наших кращих прогнозах. Як ми можемо бачити з рис. 1, метрика Precision дає більший коливання на початку та стає більш схожим на mAP50 наприкінці, коли кількість випробувань збільшується. mAP50-95 має погані значення (0,5–0,6) наприкінці, коли кількість проб збільшується. Метрика Recall має відносно сталі значення в межах 0,85–0,75 після половини проведених дослідів (епох). Це не є досить добрим результатом та потребує подальшого дослідження та навчання моделі на більшому датасеті актуальних даних. Метрика Precision дає трохи кращі результати – в межах 0,85–0,89.

Metrics

Model accuracy measured on validation set

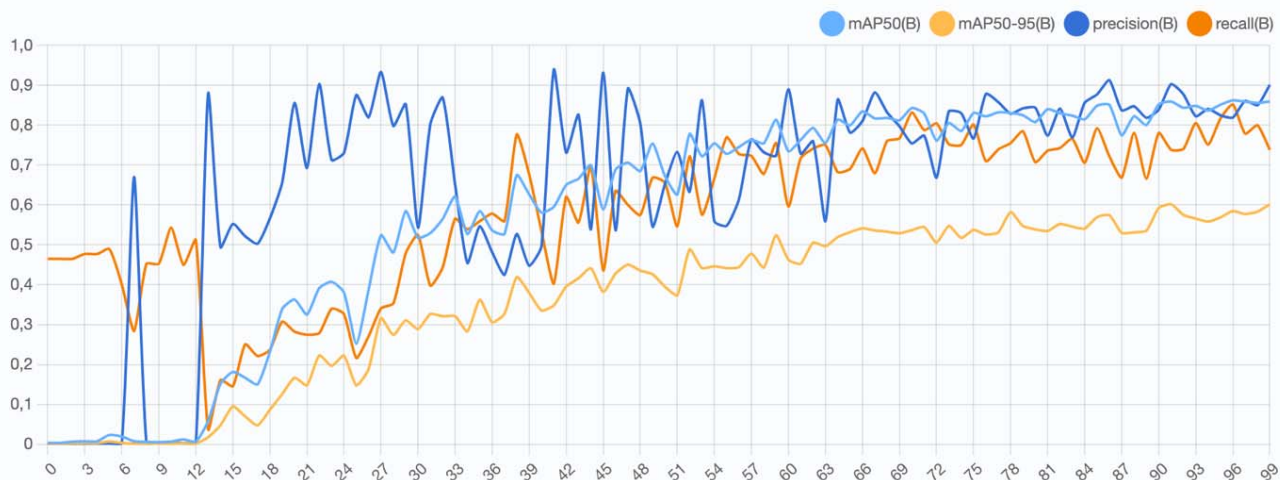


Рисунок 1 – Графік точності моделі

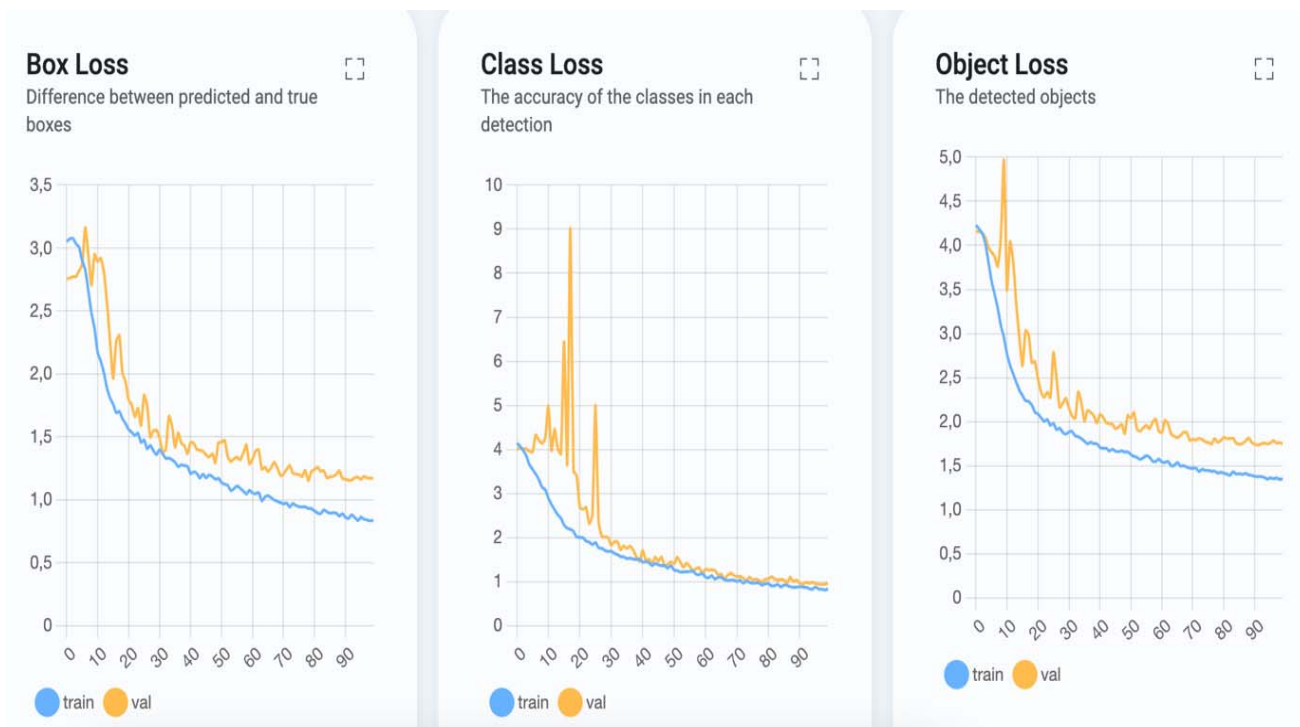


Рисунок 2 – Графіки втрат моделі

На рисунках 3–5 зображені приклади роботи системи. Великі втрати є через те, що під час дослідження не було враховано фон – навчання модуля на основі підтверджених даних (зображень)

фону без техніки. Це буде наступний крок. Також необхідно розширити класифікацію для об'єктів військової техніки – що саме Т-64, Е-72 чи Т-90.



Рисунок 3 – Приклад розпізнавання об'єкту класу БМП



Рисунок 4 – Приклад розпізнавання об'єкту класу ТАНК



Рисунок 5 – Приклад розпізнавання об'єкту класу БМП

Великі втрати присутні із-за того, що в дослідженні не враховується фон – навчання модуля на з врахування валідованих даних (зображень) фону без техніки. Це буде наступним кроком. Також необхідно розширити класифікацію для об'єктів військової техніки – який саме так Т-64, Е-72 чи Т-90.

ВИСНОВКИ

Вирішено актуальне завдання в моделюванні процесів аналізу відеопотоку для автоматичного розпізнавання цілей системи керування вогнем на основі машинного навчання. Розпізнавання цілей є пріоритетом у військовій справі. Це завдання ускладнюється тим, що треба розпізнавати рухомі об'єкти, різний рельєф і ландшафт створюють перешкоди для розпізнавання. Бойові дії відбуваються в різний час доби, відповідно, необхідно враховувати ракурс освітлення та загальне освітлення. Необхідно виявити об'єкт на відео шляхом сегментації відеокадрів, розпізнати та класифікувати його.

Наукова новизна отриманих результатів полягає у тому, що вперше запропоновано метод розпізнавання на класифікації для об'єктів військової техніки – який саме так Т-64, Е-72 чи Т-90.

Практична цінність полягає у тому, що вона є структурною складовою процесу аналізу відеопотоку для автоматичного розпізнавання цілей системи керування вогнем на основі машинного навчання.

В результаті проведених експериментів реалізовано та проаналізовано функціонування пайплайн із збору та опрацювання зображень відеоконтенту для автоматичного розпізнавання цілей

системи керування вогнем на основі машинного навчання.

Тренінг проводився з використанням ресурсів Google Colab. На навчання моделі пішло 100 епох. Аналіз проводився відповідно до mAP50 (середня точність), mAP50-95, точності та показників запам'ятовування. Запропонований спосіб може бути використаний для ідентифікації об'єктів військового характеру та розпізнавання цілей для створення (модернізації) сучасних систем управління вогнем сучасної військової техніки. метрика

Precision дає більший коливання на початку та стає більш схожим на mAP50 наприкінці, коли кількість випробувань збільшується. mAP50-95 має погані значення (0,5–0,6) наприкінці, коли кількість проб збільшується.

Метрика Recall має відносно стає значення в межах 0,85–0,75 після половини проведених дослідів (epoch). Це не є досить добрим результатом та потребує подальшого дослідження та навчання моделі на більшому датасеті актуальних даних.

Метрика Precision дає трохи кращі результати – в межах 0,85–0,89. Великі втрати є через те, що під час дослідження не було враховано фон – навчання модуля на основі підтверджених даних (зображень) фону без техніки. Це буде наступний крок. Також необхідно розширити класифікацію для об'єктів військової техніки – що саме Т-64, Е-72 чи Т-90.

Тому перспективи подальших досліджень полягають в дослідженні відопотоку з врахуванням навчання моделі на фонових зображень території України з відповідними природними ландшафтами.

ПОДЯКИ

Дослідження виконано за підтримки Міністерства освіти і науки України за проектом № 0124U000925 «Методи та засоби ідентифікації бойових машин на основі технологій глибинного навчання для автоматизованого керування цілерозподілом».

ЛІТЕРАТУРА

1. An improved deep learning method for flying object detection and recognition / [S. S. Aote, N. Wankhade, A. Pardhi et al.] // *Signal, Image and Video Processing*. – 2023. – Vol. 18(1). – P. 143–152. DOI: 10.1007/s11760-023-02703-y.
2. Zhang Z. Drone-YOLO: An Efficient Neural Network Method for Target Detection in Drone Images / Z. Zhang // *Drones*. – 2023. – Vol. 7 (2023). – Art. 526. DOI:10.3390/drones7080526.
3. Mask R-CNN / [K. He, G. Gkioxari, P. Dollár, R. Girshick] // *Computer Vision : the IEEE International Conference, Venice, Italy, 22–29 October 2017 : proceedings*. – Venice: IEEE, 2017. – P. 2961–2969. – Access mode: https://openaccess.thecvf.com/content_ICCV_2017/papers/He_Mask_R-CNN_ICCV_2017_paper.pdf.
4. Pyramid scene parsing network / [H. Zhao, J. Shi, X. Qi et al.] // *Computer Vision and Pattern Recognition : the IEEE Conference, Honolulu, HI, USA, 21–26 July 2017 : proceedings*. – Honolulu: IEEE, 2017. – P. 2881–2890. Access mode: https://openaccess.thecvf.com/content_cvpr_2017/papers/Zhao_Pyramid_Scene_Parsing_CVPR_2017_paper.pdf.
5. Ronneberger O. U-net: Convolutional networks for biomedical image segmentation / O. Ronneberger, P. Fischer, T. Brox // *Lecture Notes in Computer Science*. – 2015. – Vol. 9351. – P. 234–241. DOI: 10.1007/978-3-319-24574-4_28.
6. Faster R-CNN: Towards real-time object detection with region proposal networks / [S. Ren, K. He, R. Girshick, J. Sun] // *Advances in neural information processing systems*. – 2015. – Vol. 28. – P. 1–9. Access mode: https://proceedings.neurips.cc/paper_files/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf.
7. Girshick R. Fast R-CNN / R. Girshick // *Computer Vision : the IEEE International Conference, Santiago, Chile, 7–13 December 2015 : proceedings*. – Chile: IEEE, 2015. – P. 1440–1448. – Access mode: https://openaccess.thecvf.com/content_iccv_2015/papers/Girshick_Fast_R-CNN_ICCV_2015_paper.pdf.
8. Rich feature hierarchies for accurate object detection and semantic segmentation / [R. Girshick, J. Donahue, T. Darrell, J. Malik] // *Computer Vision and Pattern Recognition : the IEEE Conference, Columbus, OH, USA, 23–28 June 2014 : proceedings*. – Columbus: IEEE, 2014. – P. 580–587. Access mode: https://openaccess.thecvf.com/content_cvpr_2014/papers/Girshick_Rich_Feature_Hierarchies_2014_CVPR_paper.pdf.
9. Densely connected convolutional networks / [G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger] // *Computer Vision and Pattern Recognition : the IEEE Conference, Honolulu, HI, USA, 21–26 July 2017 : proceedings*. – Honolulu: IEEE, 2017. – P. 4700–4708. Access mode: https://openaccess.thecvf.com/content_cvpr_2017/papers/Huang_Densely_Connected_Convolutional_CVPR_2017_paper.pdf.
10. Deep residual learning for image recognition / [K. He, X. Zhang, S. Ren, J. Sun] // *Computer Vision and Pattern Recognition : the IEEE Conference, Las Vegas, NV, USA, 27–30 June 2016 : proceedings*. – Las Vegas: IEEE, 2016. – P. 770–778. Access mode: https://openaccess.thecvf.com/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf.
11. Simonyan K. Very deep convolutional networks for large-scale image recognition / K. Simonyan, A. Zisserman // *arXiv*. – © Висоцька В. А., Романчук Р. В., 2024
DOI 10.15588/1607-3274-2024-3-7
- Access mode: <https://arxiv.org/abs/1409.1556>. DOI: 10.48550/arXiv.1409.1556.
12. Microsoft COCO: Common Objects in Context / [T. Lin, M. Maire, S. J. Belongie et al.] // *Lecture Notes in Computer Science*. – 2014. – Vol. 8693. – P. 740–755. DOI: 10.1007/978-3-319-10602-1_48.
13. The PASCAL Visual Object Classes Challenge Results / [M. Everingham, L. Van Gool, C. K. I. Williams et al.] // *Pascal-Network.org*. – Access mode: <http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html>.
14. The PASCAL Visual Object Classes Challenge Results / [M. Everingham, L. Van Gool, C. K. I. Williams et al.] // *Pascal-Network.org*. – Access mode: <http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html>.
15. Jocher G. YOLO by Ultralytics / G. Jocher, A. Chaurasia, J. Qiu // *GitHub*. – Access mode: <https://github.com/ultralytics/ultralytics/blob/main/CITATION.cff>.
16. Wang C. Y. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors / C. Y. Wang, A. Bochkovskiy, H. Y. M. Liao // *Computer Vision and Pattern Recognition : the IEEE/CVF Conference, Vancouver, BC, Canada, 18–22 June 2023 : proceedings*. – Vancouver: IEEE, 2023. – P. 7464–7475. Access mode: https://openaccess.thecvf.com/content/CVPR2023/papers/Wang_YOLOv7_Trainable_Bag-of-Freebies_Sets_New_State-of-the-Art_for_Real-Time_Object_Detectors_CVPR_2023_paper.pdf.
17. YOLOv6: A single-stage object detection framework for industrial applications / [C. Li, L. Li, H. Jiang et al.] // *arXiv*. – Access mode: <https://arxiv.org/abs/2209.02976>. DOI: 10.48550/arXiv.2209.02976.
18. Yolox: Exceeding yolo series in 2021 / [Z. Ge, S. Liu, F. Wang et al.] // *arXiv*. – Access mode: <https://arxiv.org/abs/2107.08430>. DOI: 10.48550/arXiv.2107.08430.
19. Ultralytics/Yolov5: V6.0 – YOLOv5n 'Nano' Models, Roboflow Integration, TensorFlow Export, OpenCV DNN Support / [G. Jocher, A. Stoken, A. Chaurasia et al.] // *Zenodo*. – 2021. – Access mode: <https://zenodo.org/record/5563715>, <https://github.com/ultralytics/yolov5/releases>.
20. Bochkovskiy A. Yolov4: Optimal speed and accuracy of object detection / A. Bochkovskiy, C. Y. Wang, H. Y. M. Liao // *arXiv*. – Access mode: <https://arxiv.org/abs/2004.10934>. DOI: 10.48550/arXiv.2004.10934
21. Redmon J. YOLO9000: Better, faster, stronger / J. Redmon, A. Farhadi // *Computer Vision and Pattern Recognition : the IEEE Conference, Honolulu, HI, USA, 21–26 July 2017 : proceedings*. – Honolulu: IEEE, 2017. – P. 7263–7271. Access mode: https://openaccess.thecvf.com/content_cvpr_2017/papers/Redmon_YOLO9000_Better_Faster_CVPR_2017_paper.pdf.
22. You only look once: Unified, real-time object detection / [J. Redmon, S. Divvala, R. Girshick, A. Farhadi] // *Computer Vision and Pattern Recognition : the IEEE Conference, Las Vegas, NV, USA, 27–30 June 2016 : proceedings*. – Las Vegas: IEEE, 2016. – P. 779–788. Access mode: https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Redmon_You_Only_Look_CVPR_2016_paper.pdf.
23. CSPNet: A new backbone that can enhance learning capability of CNN / [C. Y. Wang, H. Y. M. Liao, Y. H. Wu, P. Y. Chen, J.W. Hsieh, I.H. Yeh] // *Computer Vision and Pattern Recognition Workshops: the IEEE/CVF Conference, Seattle, WA, USA, 14–19 June 2020 : proceedings*. – Seattle: IEEE, 2020. – P. 390–391. Access mode: https://openaccess.thecvf.com/content_CVPRW_2020/papers/w28/Wang_CSPNet_A_New_Backbone_That_Can_Enhance_Learning_Capability_of_CVPRW_2020_paper.pdf.
24. Detection in crowded scenes: One proposal, multiple predictions / [X. Chu, A. Zheng, X. Zhang, J. Sun] // *Computer*

- Vision and Pattern Recognition : the IEEE/CVF Conference, Seattle, WA, USA, 14–19 June 2020 : proceedings. – Seattle: IEEE, 2020. – P. 12214–12223. Access mode: https://openaccess.thecvf.com/content_CVPR_2020/papers/ChuDetection_in_Crowded_Scenes_One_Proposal_Multiple_Predictions_CVPR_2020_paper.pdf
25. Path aggregation network for instance segmentation / [S. Liu, L. Qi, H. Qin et al.] // Computer Vision and Pattern Recognition : the IEEE Conference, Salt Lake City, UT, USA, 18–23 June 2018 : proceedings. – Salt Lake City: IEEE, 2018. – P. 8759–8768. Access mode: https://openaccess.thecvf.com/content_cvpr_2018/papers/Liu_Path_Aggregation_Network_CVPR_2018_paper.pdf
26. Glorot X. Deep sparse rectifier neural networks / X. Glorot, A. Bordes, Y. Bengi // Artificial Intelligence and Statistics (Proceedings of Machine Learning Research): the Fourteenth International Conference, Ft. Lauderdale, FL, USA, 11–13 April 2011 : proceedings. – Ft. Lauderdale: MLResearchPress, 2011. – P. 313–326. Access mode: <https://proceedings.mlr.press/v15/glorot11a/glorot11a.pdf>
27. Eilfwing S. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning / S. Eilfwing, E. Uchibe, K. Doya // Neural Netw. – 2018. – Vol. 107 (2018). – P. 3–11. DOI: 10.1016/j.neunet.2017.12.012.
- Accepted 30.06.2024.
Received 29.08.2024.

UDC 004.9

INTELLIGENT VIDEO ANALYSIS TECHNOLOGY FOR AUTOMATIC FIRE CONTROL TARGET RECOGNITION BASED ON MACHINE LEARNING

Vysotska V. – PhD, Associate Professor of Information Systems and Networks Department, Lviv Polytechnic National University, Lviv, Ukraine.

Romanchuk R. – Post-graduate student of Information Systems and Networks Department, Lviv Polytechnic National University, Lviv, Ukraine.

ABSTRACT

Context. Target recognition is a priority in military affairs. This task is complicated by the fact that it is necessary to recognize moving objects, different terrain and landscape create obstacles for recognition. Combat actions can take place at different times of the day, accordingly, it is necessary to take into account the perspective of lighting and general lighting. It is necessary to detect the object in the video by segmenting the video frames, recognize and classify.

Objective of the study is to develop a technology for the analysis of the development of a technology for recognizing targets in real time as a component of the fire control system, due to the use of artificial intelligence, YOLO and machine learning.

Method. The article develops a video stream analysis technology for automatic target recognition of the fire control system based on machine learning. The paper proposes the development of a target recognition module as a component of the fire control system within the framework of the proposed information technology using artificial intelligence. The YOLOv8 pattern recognition model family was used to develop the target recognition module. The methods used during the study of the formed dataset.

- Bounding Box: Noise – Up to 15% of pixels (limiting frame: adding salt and pepper noise to the image – up to 15% of pixels).
- Bounding Box: Blur – Up to 2.5px (bounding box: adding Gaussian blur to the image – up to 2.5 pixels).
- Cutout – 3 boxes with 10% size each (cut out a part of the image – 3 boxes of 10% size each).
- Brightness Between –25% and +25% (changing the brightness of the image to increase the resistance of the model to changes in lighting and camera settings – from –25% to +25%).
- Rotation – Between –15° and +15° (rotation of the image object – clockwise or counterclockwise by degrees from –15° to +15°).
- Flip – Horizontal (flip the image object horizontally).

Results. The data is collected from open sources, in particular, from videos posted in open sources on the YouTube platform. The main task of data preprocessing is the classification of three classes of objects on video or in real time – APC, BMP and TANK. The dataset is formed using the Roboflow platform based on the labeling tools and subsequently the augmentation tools. The dataset consists of 1193 unique images – approximately equally for each class. The training was conducted using Google Colab resources. It took 100 epochs to train the model.

Conclusions. Analysis is performed according to mAP50 (average precision as 0.85), mAP50-95 (0.6), precision (0.89) and recall (0.75). Big losses are due to the fact that the background was not taken into account during the research – training the module on the basis of confirmed data (images) of the background without technology

KEYWORDS: moving object recognition, security, privacy, YOLO, target identification, machine learning, APC, BMP, TANK.

REFERENCES

1. Aote S. S., Wankhade N., Pardhi A., Misra N., Agrawal H., Potnurwar A. An improved deep learning method for flying object detection and recognition, *Signal, Image and Video Processing*, 2023, Vol. 18(1), pp. 143–152. DOI:10.1007/s11760-023-02703-y.
2. Zhang Z. Drone-YOLO: An Efficient Neural Network Method for Target Detection in Drone Images, *Drones*, 2023, Vol. 7 (2023), Art. 526. DOI:10.3390/drones7080526.
3. He K., Gkioxari G., Dollár P., Girshick R. Mask R-CNN, *Computer Vision : the IEEE International Conference, Venice, Italy, 22–29 October 2017 : proceedings*. Venice, IEEE, 2017, pp. 2961–2969. Access mode: https://openaccess.thecvf.com/content_ICCV_2017/papers/He_Mask_R-CNN_ICCV_2017_paper.pdf.
4. Zhao H., Shi J., Qi X., Wang X., Jia J. Pyramid scene parsing network, *Computer Vision and Pattern Recognition : the IEEE Conference, Honolulu, HI, USA, 21–26 July 2017 : proceedings*. Honolulu, IEEE, 2017, pp. 2881–2890. Access mode: https://openaccess.thecvf.com/content_cvpr_2017/papers/Zhao_Pyramid_Scene_Parsing_CVPR_2017_paper.pdf.
5. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation, *Lecture Notes in Computer Science*, 2015, Vol. 9351, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28.

6. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks, *Advances in neural information processing systems*, 2015, Vol. 28, pp. 1–9. Access mode: https://proceedings.neurips.cc/paper_files/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf.
7. Girshick R. Fast R-CNN, *Computer Vision : the IEEE International Conference, Santiago, Chile, 7–13 December 2015 : proceedings*. Chile, IEEE, 2015, pp. 1440–1448. Access mode: https://openaccess.thecvf.com/content_iccv_2015/papers/Girshick_Fast_R-CNN_ICCV_2015_paper.pdf.
8. Girshick R., Donahue J., Darrell, T. Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation, *Computer Vision and Pattern Recognition : the IEEE Conference, Columbus, OH, USA, 23–28 June 2014 : proceedings*. Columbus, IEEE, 2014, pp. 580–587. Access mode: https://openaccess.thecvf.com/content_cvpr_2014/papers/Girshick_Rich_Feature_Hierarchies_2014_CVPR_paper.pdf.
9. Huang G., Liu Z., Van Der Maaten L., Weinberger K. Q. Densely connected convolutional networks, *Computer Vision and Pattern Recognition : the IEEE Conference, Honolulu, HI, USA, 21–26 July 2017 : proceedings*. Honolulu, IEEE, 2017, pp. 4700–4708. Access mode: https://openaccess.thecvf.com/content_cvpr_2017/papers/Huang_Densely_Connected_Convolutional_CVPR_2017_paper.pdf.
10. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition, *Computer Vision and Pattern Recognition : the IEEE Conference, Las Vegas, NV, USA, 27–30 June 2016 : proceedings*. Las Vegas: IEEE, 2016, pp. 770–778. Access mode: https://openaccess.thecvf.com/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf.
11. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition, *arXiv*. Access mode: <https://arxiv.org/abs/1409.1556>. DOI: 10.48550/arXiv.1409.1556.
12. Lin T., Maire M., Belongie S. J., Bourdev L. D., Girshick R. B., Hays J., Perona P., Ramanan D., Doll'ar P., Zitnick C. L. Microsoft COCO: Common Objects in Context, *Lecture Notes in Computer Science*, 2014, Vol. 8693, pp. 740–755. DOI: 10.1007/978-3-319-10602-1_48.
13. Everingham M., Van Gool L., Williams C. K. I., Winn J., Zisserman A. The PASCAL Visual Object Classes Challenge Results, *Pascal-Network.org*. Access mode: <http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html>.
14. Everingham M., Van Gool L., Williams C. K. I., Winn J., Zisserman A. The PASCAL Visual Object Classes Challenge Results, *Pascal-Network.org*. Access mode: <http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html>.
15. Jocher G., Chaurasia A., Qiu J. YOLO by Ultralytics, *GitHub*. Access mode: <https://github.com/ultralytics/ultralytics/blob/main/CITATION.cff>.
16. Wang C. Y., Bochkovskiy A., Liao H. Y. M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, *Computer Vision and Pattern Recognition : the IEEE/CVF Conference, Vancouver, BC, Canada, 18–22 June 2023 : proceedings*. Vancouver: IEEE, 2023, pp. 7464–7475. Access mode: https://openaccess.thecvf.com/content/CVPR2023/papers/Wang_YOLOv7_Trainable_Bag-of-Freebies_Sets_New_State-of-the-Art_for_Real-Time_Object_Detectors_CVPR_2023_paper.pdf.
17. Li C., Li L., Jiang H., Weng K., Geng Y., Li L., Ke Z., Li Q., Cheng M., Nie W., Li Y., Zhang B., Liang Y., Zhou L., Xu X., Chu X., Wei X., Wei X. YOLOv6: A single-stage object detection framework for industrial applications, *arXiv*. Access mode: <https://arxiv.org/abs/2209.02976>. DOI: 10.48550/arXiv.2209.02976.
18. Ge Z., Liu S., Wang F., Li Z., Sun J. Yolox: Exceeding yolo series in 2021, *arXiv*. – Access mode: <https://arxiv.org/abs/2107.08430>. DOI: 10.48550/arXiv.2107.08430.
19. Jocher G., Stoken A., Chaurasia A., Borovec J., NanoCode012, TaoXie, Kwon Y., Michael K., Changyu L., Fang J., Abhiram V., Laughing, Tkianai, yxNONG, Skalski P., Hogan A., Nadar J., Imyhxy, Mammana L., Wang Alex, Fati C., Montes D., Hajek J., Diaconu L., Minh M. T., Marc, Albinxavi, Fatih, Oleg, Wanghaoyang Ultralytics/Yolov5: V6.0 – YOLOv5n 'Nano' Models, Roboflow Integration, TensorFlow Export, OpenCV DNN Support, *Zenodo*, 2021. Access mode: <https://zenodo.org/record/5563715>, <https://github.com/ultralytics/yolov5/releases>.
20. Bochkovskiy A., Wang C. Y., Liao H. Y. M. Yolov4: Optimal speed and accuracy of object detection, *arXiv*. Access mode: <https://arxiv.org/abs/2004.10934>. DOI: 10.48550/arXiv.2004.10934
21. Redmon J., Farhadi A. YOLO9000: Better, faster, stronger, // *Computer Vision and Pattern Recognition : the IEEE Conference, Honolulu, HI, USA, 21–26 July 2017 : proceedings*. Honolulu, IEEE, 2017, pp. 7263–7271. Access mode: https://openaccess.thecvf.com/content_cvpr_2017/papers/Redmon_YOLO9000_Better_Faster_CVPR_2017_paper.pdf.
22. Redmon J., Divvala S., Girshick R., Farhadi A. You only look once: Unified, real-time object detection, *Computer Vision and Pattern Recognition : the IEEE Conference, Las Vegas, NV, USA, 27–30 June 2016 : proceedings*. Las Vegas, IEEE, 2016, pp. 779–788. Access mode: https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Redmon_You_Only_Look_CVPR_2016_paper.pdf.
23. Wang C. Y., Liao H. Y. M., Wu Y. H., Chen P. Y., Hsieh J. W., Yeh I. H. CSPNet: A new backbone that can enhance learning capability of CNN, *Computer Vision and Pattern Recognition Workshops: the IEEE/CVF Conference, Seattle, WA, USA, 14–19 June 2020 : proceedings*. Seattle, IEEE, 2020, pp. 390–391. Access mode: https://openaccess.thecvf.com/content_CVPRW_2020/papers/w28/Wang_CSPNet_A_New_Backbone_That_Can_Enhance_Learning_Capability_of_CVPRW_2020_paper.pdf.
24. Chu X., Zheng A., Zhang X., Sun J. Detection in crowded scenes: One proposal, multiple predictions, *Computer Vision and Pattern Recognition : the IEEE/CVF Conference, Seattle, WA, USA, 14–19 June 2020 : proceedings*. Seattle, IEEE, 2020, pp. 12214–12223. Access mode: https://openaccess.thecvf.com/content_CVPR_2020/papers/Chu_Detection_in_Crowded_Scenes_One_Proposal_Multiple_Predictions_CVPR_2020_paper.pdf.
25. Liu S., Qi L., Qin H., Shi J., Jia J. Path aggregation network for instance segmentation, *Computer Vision and Pattern Recognition : the IEEE Conference, Salt Lake City, UT, USA, 18–23 June 2018 : proceedings*. Salt Lake City: IEEE, 2018, pp. 8759–8768. Access mode: https://openaccess.thecvf.com/content_cvpr_2018/papers/Liu_Path_Aggregation_Network_CVPR_2018_paper.pdf.
26. Glorot X., Bordes A., Bengi Y. Deep sparse rectifier neural networks, *Artificial Intelligence and Statistics (Proceedings of Machine Learning Research): the Fourteenth International Conference, Ft. Lauderdale, FL, USA, 11–13 April 2011 : proceedings*. Ft. Lauderdale, MLResearchPress, 2011, pp. 313–326. Access mode: <https://proceedings.mlr.press/v15/glorot11a/glorot11a.pdf>.
27. Elfving S., Uchibe E., Doya K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning, *Neural Netw.*, 2018, Vol. 107 (2018), pp. 3–11. DOI: 10.1016/j.neunet.2017.12.012.