УДК 004.85:539.1.074.8:681.7 **doi: 10.32620/aktt.2024.5.11**

А. С. МОСКАЛЕНКО, М. О. ЗАРЕЦЬКИЙ, М. О. ВИНОГРАДОВ, В. Ю. БАБИЧ

Сумський державний університет, Суми, Україна

МОДЕЛЬ ТА МЕТОД НАВЧАННЯ ДЕТЕКТОРА ОБ'ЄКТІВ НА АЕРОЗОБРАЖЕННІ З ОПТИМІЗАЦІЄЮ РОБАСТНОСТІ ТА КІЛЬКОСТІ ОБЧИСЛЕНЬ

Предметом *дослідження є нейромережеві детектори об'єктів, що широко застосовуються для аналізу аеровідеозображень. Зростає кількість завдань, що потребують обробки даних безпосередньо на борту безпілотного літального апарату, що обмежує доступні обчислювальні ресурси. Вразливість нейромереж до шуму, протиборчих атак та ін'єкцій помилок у ваги знижує їх функціональність. Актуальним* ***завданням*** *є розроблення моделі нейромережі, що забезпечує як обчислювальну ефективність, так і стійкість до збурень. У статті досліджується модель і метод забезпечення робастності нейромережевих детекторів об'єктів на аерозображеннях в умовах обмежених ресурсів.* ***Метою*** *є створення моделі, яка раціонально використовує бортові обчислювальні ресурси та забезпечує стійкість до збурень. Використано методи динамічних нейромереж, оптимізації робастності і резильєнтності. Отримано наступні* ***результати****. Розроблено детектор з екстрактором ознак на основі ViT-S/16, модифікованого гейтовими модулями для динамічного екзамену. Модель навчена на наборі даних RSOD і мета-навчена на результатах адаптації до різнотипних збурень. Було протестовано стійкість моделі до випадкових інверсій бітів у вагах (10 % ваг) та до протиборчих атак з амплітудою збурення до 3/255 (L∞ норма). Запропонована модель детектора з динамічним екзаменом та оптимізованою робастністю показала зниження кількості операцій з плаваючою крапкою на понад 20 % без втрати точності.* ***Висновки****. Розроблено модель детектора на основі екстрактора ознак ViT-S/16 та метод навчання детектора, що полягає у поєднанні функції втрат RetinaNet з функцією втрат гейтових блоків та у застосуванні мета-навчання на результатах адаптації до різнотипних синтетичних збурень. Тестування показало підвищення точності на 11,9 % в умовах впливу ін'єкції помилок і на 13,2% в умовах впливу протиборчих атак.*

Ключові слова: *детекція об'єктів; робастність; протиборчі атаки; ін'єкції несправностей; мета-навчання.*

1. Вступ

1.1. Мотивація дослідження

Зображення аеро фото- і відео- зйомки характеризуються значною варіативністю перспективного спотворення та масштабу, а також високим рівнем варіативності контексту та фону [1]. Це ставить високі вимоги до якості та обсягу даних, а також обчислювальних ресурсів. Ефективне використання ресурсів під час оброблення аерозображень важливе з погляду підвищення автономності бортової системи безпілотного апарату з обмеженими габаритами. Існуючі методи стиснення нейронних мереж для зменшення споживання ресурсів не забезпечують прийнятного рівня робастності до різних типів перешкод, що впливають на систему розпізнавання об'єктів на зображенні [2].

Стійкість до шуму, новизни в даних та пошкодження ваг в технології нейронних мереж для розпізнавання об'єктів на зображенні досягається шляхом введення певного рівня надлишковості для поглинання збурень [3, 4]. Дублювання нейронів або альтернативні нейронні шляхи вводиться для поглинання пошкодження ваг мережі [4]. Для підвищення робастності до шуму протиборчих атак використовуються знешумлюючі автоенкодери [5], додаються детектори збурення та викидів [6], використовуються методи ансамблювання [7], а також навчаються більші нейронні мережі на даних, що агументовані за допомогою збурень та інших технік. На даний момент, бракує досліджень, які пропонують ефективні методи забезпечення робастності в умовах обмежених ресурсів.

Моделі детекції об'єктів успішно розвиваються в напрямку вдосконалення архітектур та мікроархітектур для підвищення точності локалізації та класифікації об'єктів різних масштабів на аерозображенні [8]. Архітектури на основі згорткових мереж та візуальних трансформерів покращуються шляхом впровадження динамічних механізмів екзамену для покращення обчислювальної ефективності [9, 10]. Однак більшість експериментів проводяться переважно з використанням класифікації зображень, а не детекції об'єктів. Проте

ISSN 1727-7337 (print)
ISSN 2663-2217 (online)

детекція об'єктів є найбільш затребуваними для практичних застосувань. Крім того, існує недостатня кількість досліджень робастності динамічних нейронних мереж до різних видів збурень.

Існує прогалина в дослідженнях одночасного забезпечення робастності та зменшення обчислювальної складності моделей детекції об'єктів на аерозображеннях. Розроблення моделі та методу для забезпечення ефективної робастності детекторів об'єктів на аерозображеннях шляхом поєднання концепцій та технік з динамічних нейронних мереж, методів оптимізації робастності та резільєнтності нейронних мереж, є перспективною областю досліджень.

1.2. Сучасний стан

Для детекції об'єктів на зображеннях довгий час основними були одноетапні та двоетапні моделі згорткових нейронних мереж [8]. Архітектура згорткових мереж стає значно складнішою, коли враховується різномасштабність вигляду об'єктів та варіативність контексту навколишнього середовища на зображенні. Візуальні трансформери та їх модифікації забезпечують кращу інтеграцію локальної та глобальної контекстуальної інформації на зображеннях та демонструють високу узагальнюючу здатність для великих наборів даних [11]. Проте трансформери менш обчислювально ефективні, ніж згорткові мережі.

Для покращення обчислювальної ефективності нейронних мереж довгий час використовуються такі техніки, як квантування ваг, обрізання з'єднань або нейронів та дистиляція знань [12, 13]. Однак такі методи стиснення моделей можуть призвести до втрати точності та робастності. Один зі способів покращення адаптивності та обчислювальної ефективності полягає в застосуванні концепцій та методів з динамічних нейронних мереж [9, 10]. Два найпопулярніших підходи – це мережі з раннім виходом [14] та мережі з гейтовими модулями. У обох випадках обчислюються лише відповідні шари в залежності від контексту. Однак модель з гейтовими модулями пропонує більш загальну архітектуру, оскільки дозволяє вимикати будь-яку підмножину шарів в залежності від умов.

Нейронні мережі, включаючи згорткові мережі та візуальні трансформери, вразливі до протиборчих атак, даних поза навчальним розподілом та атак шляхом ін'єкції помилок у ваги мережі [15, 16]. Є деякі дослідження, які демонструють покращення робастності до збурень через стандартні методи динамічного екзамену [17, 18]. Але спрямована оптимізація динамічних нейронних мереж для збільшення їх робастності до збурень залишається слабо вивченою проблемою. Дослідження [19, 20] вивчають різні методи навчання, спрямовані на вдосконалення робастності до пошкоджень ваг нейронних мереж. Дослідження [21, 22] розглядають різні методи протиборчого навчання для підвищення робастності до шуму та протиборчих атак. Деякі дослідження вивчають різні методи попередньої обробки вхідних даних [23], постобробки результатів [24] та модифікації архітектур нейронних мереж [25] для підвищення їх стійкості до збурень. У праці [27] робиться перша спроба підвищення робастності динамічних нейронних мереж шляхом навчання в умовах впливу різнотипних збурень. Однак використання великої і попередньо навченої на великому обсязі даних нейронної мережі ViT-B/16 ускладнює розуміння наскільки даний підхід релевантний для менших нейронних мереж. Крім того не було досліджено впливу на робастність мережі до шуму в даних та вагах відсутність в процедурі точної настройки періодичної зміни задач. Таким чином ефективність одночасного застосування різних методів, зокрема для динамічних нейронних мереж, залишається не достатньо дослідженою.

1.3. Мета та підхід

Мета цього дослідження полягає в розробленні моделі та методу для забезпечення ефективної робастності детекторів об'єктів на зображеннях шляхом інтеграції концепцій та технік з динамічних нейронних мереж, методів оптимізації робастності та стійкості нейронних мереж.

Робастність характеризує здатність системи витримувати певний рівень збурень, зберігаючи функціональність без значного погіршення або втрати продуктивності. Під робастною ми розуміємо модель виявлення об'єктів, яка реалізує механізми поглинання та протистояння певному рівню та типу деструктивних збурень. Чим менше знижується продуктивність детектора під впливом збурення, тим більш робастним вважається детектор. Робастність можна визначати як залишкову функціональність після впливу екстремального деструктивного збурення. Ми будемо порівнювати робастність моделей на основі їх залишкового показника працездатності після впливу збурюючого фактора.

Ключові завдання полягають у наступному:

– аналіз існуючих рішень для забезпечення робастності та зменшення обчислювальної складності в системах детеції об'єктів на зображеннях;

– розроблення моделі динамічної нейронної мережі для детекції об'єктів на зображеннях в умовах

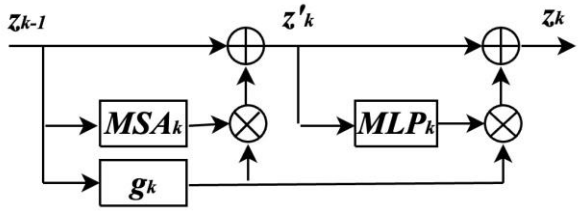
високої мінливості спостережень, впливу різних видів шуму та пошкодження ваг мережі;

– розроблення методу навчання для динамічного нейронного детектора об'єктів на зображеннях для забезпечення робастності до шуму, протиборчих атак та пошкодження ваг нейронної мережі;

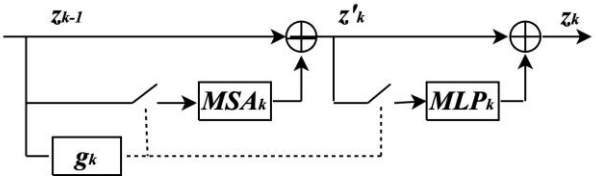
– дослідження залежності між продуктивністю розробленої моделі і методу навчання та певними гіперпараметрами системи штучного інтелекту.

2. Модель детекції об'єктів

Як основу екстрактора ознак було використано трансформер ViT-S/16, попередньо навчений на великому наборі даних [26]. Для покращення обчислювальної ефективності та адаптивності до контексту та збурень ми пропонуємо ввести гейтові модулі, які динамічно вимикають нерелевантні або скомпрометовані трансформені енкодери. Іншими словами, припускається, що лише релевантні ознаки будуть обчислені активованою підмножиною шарів. Запропонована архітектура глибокого екстрактора ознак на основі динамічної нейронної мережі показана на рис. 1. Кожен енкодер складається з блоку багатоголової самоуваги (MSA-блок) та блоку багатошарового персептрону (MLP-блок) зі зв'язками пропуску, а також гейтового модуля для активації або вимкнення k-го блоку в залежності від умов.



а



б

Рис. 1. Схематичне зображення структурної одиниці глибокого екстрактора знак на основі динамічного візуального трансформера:
а – режим навчання; б – режим екзамену

Залежність між вхідним тензором z_{k-1} та вихідним тензором z_k k-го блоку з відповідним з'єднанням пропуску та гейтовим модулем під час навчання можна визначити наступним чином:

$$z'_k = z_{k-1} + g_k(z_{k-1})\mathrm{MSA}_k(z_{k-1}); \quad (1)$$

$$z_k = z'_k + g_k(z_{k-1})\mathrm{MLP}_k(z'_k), \quad (2)$$

де g_k – гейтова функція, $g_k \in \{0,1\}$;

f_k – це функція обчислення ознак k-го структурного блоку візуального трансформера (багатоголова самоувага, багатошаровий персептрон).

Обчислювальна процедура для оновлення z_k під час екзамену має вигляд:

$$z_k = \begin{cases} z_{k-1}, & \text{if} g_k(z_{k-1}) = 0; \\ \tilde{z}_k + \mathrm{MLP}_k(\tilde{z}_k)^{\infty} \text{if } g_k(z_{k-1}) = 1; \end{cases} \quad (3)$$

де $\tilde{z}_k = \mathrm{MSA}(z_{k-1}) + z_{k-1}$.

Відповідно до функціонального призначення гейтового модуля та особливостей навчання багатошарових нейронних мереж, він повинен мати наступні властивості [9, 10]:

- низька обчислювальна складність порівняно з основним блоком, який активується або деактивується;

- стохастичність для запобігання переходу режиму в тривіальні рішення, тобто постійно виконувані блоки чи ніколи не виконувані блоки;

- здатність генерувати дискретні рішення та обчислювати градієнти для оптимізації параметрів гейтового модуля.

На рис. 2 показана структура гейтового модуля в режимах навчання та виведення. Додавання шуму Гумбеля

$$G = -\log\left(-\log(U)\right),$$

де U ~ Unif[0, 1], до нейронного виходу гейтового модуля дозволяє додати деяку стохастичність для уникнення тривіальних рішень в режимі екзамену. Використання Гумбель-Софтмакс методу забезпечує диференціювання гейтового модуля та можливість оптимізації його параметрів.

У виконаних експериментах модель гейтового модуля однакова для всіх блоків мережі. Для обчислення ознак для гейтового модуля використовується MLP з одним прихованим шаром з 24 нейронами та вихідним шаром з 2 нейронами. Використовується функція активації ReLU6, яка була описана вище. У випадку зі згортковою мережею функцію агрегації

ISSN 1727-7337 (print)
ISSN 2663-2217 (online)

(пулінгу) можна реалізувати як глобальний середнюючий пулінг. У випадку візуального трансформера послідовність токенів переформатовується в 2D сітку, схожу на проміжне представлення згорткової нейронної мережі, з подальшою згорткою (16 фільтрів з ядром 3x3) та Max Pool 2x2 [28].

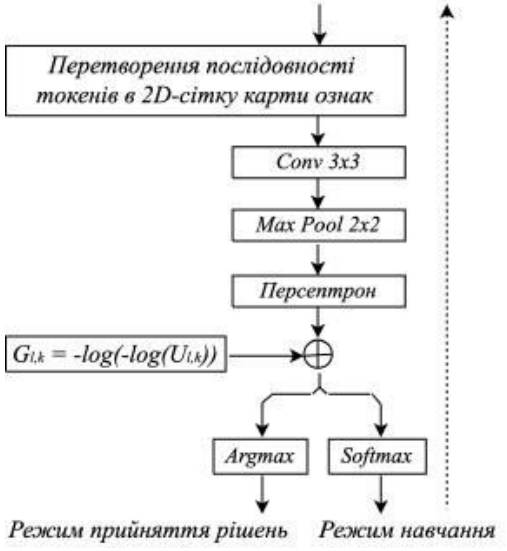


Рис. 2. Архітектура гейтового модуля

Ефективна адаптація універсального екстрактора ознак, не орієнтованого на конкретні задачі, до конкретного завдання детекції об'єктів на зображенні вимагає додавання специфічного пляшкового горла. Це пляшкове горло повинно виділяти ті ознаки, які найбільш зручні для кодування інформації про виявлені об'єкти різних розмірів. У [29], було запропоновано так званий Simple Feature Piramid Network, який формує 4 різномасштабних карт ознак (рис. 3).

Масштаб 1/32 формується за допомогою зменшення вдвічі через максимізуючий пулінг 2×2 (усереднюючий пулінг або згортка працюють аналогічно). Масштаб 1/16 просто використовує остаточну карту ознак ViT. Масштаб 1/8 (або 1/4) формується одним (або двома) шарами деконволюції з кроком=2. У випадку масштабу 1/4 перший шар деконволюції слідує за LayerNorm(LN) [29] та ReLU6 [30]. Потім для кожного рівня піраміди застосовується згортка 1×1 з LN для зменшення розмірності до 256, а потім згортка 3×3 також з LN, аналогічно до обробки на кожному рівні в FPN.

Пропонується використовувати функцію активації ReLU6 для підвищення робастності до збурень. ReLU6 зменшує область протиборчої атаки та можливість надмірної активації внаслідок пошкодження ваг за рахунок обмеження максимального значення.

Детектуюча головка, яка застосовується до кожної карти ознак, обчислює впевненість та обмежувальні рамки для виявлених об'єктів. Детектуюча головка складається з підмережі регресії обмежуючої рамки та класифікатора (рис. 4) [31].

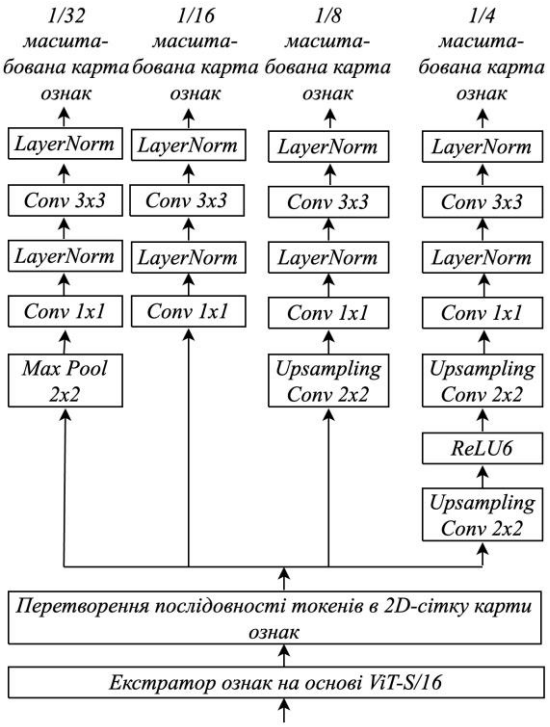


Рис. 3. Архітектура Simple Feature Piramid Network для візуального трансформера

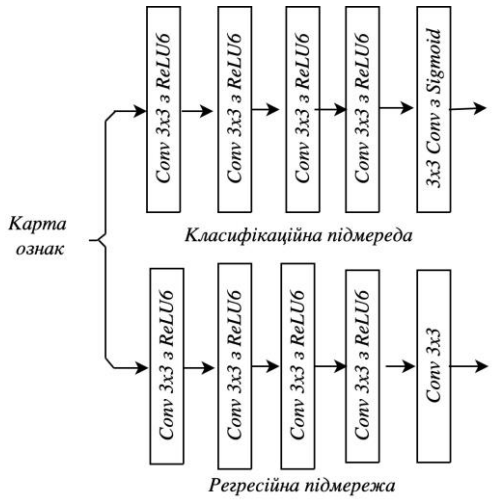


Рис. 4. RetinaNet-подібна детектуюча головка

Пропонується побудувати архітектуру одноетапної детекції, схожу на RetinaNet, для покращення продуктивності. Для кожної комірки карти ознак формується 9 якірних рамок, кожна з різним розміром та відношенням сторін [8].

Кожну цільову рамку порівнюють з якірними рамками на кожному етапі навчання. Якщо ступінь перетину (Intersection of Union, IoU) між якірною

рамкою та цільовою рамкою більший за 0,5, то відповідна якірна рамка призначається цільовій рамці. Якщо IoU менше 0,4, то якірна рамка вважається фоновою рамкою. У всіх інших випадках якірна рамка буде ігноруватися під час навчання. Класифікаційна підмережа навчається з урахуванням отриманих призначень (клас об'єкту або фон). Підмережа регресії навчається з урахуванням координат вибраної якірної рамки. Помилка обчислюється відносно якірної рамки, а не цільової рамки.

3. Метод навчання

Візуальні трансформери можуть бути попередньо навчені за допомогою будь-якого ефективного методу. На даний момент найпотужніші методи переднього навчання без вчителя – це самодистиляція без міток (DINO) [32] та DINOv2 [33].

Точна підгонка екстрактора ознак та тренування мережі піраміди ознак з детектуючою головкою передбачає мінімізацію композитної функції втрат

$$L = \lambda_{loc} L_{loc} + \lambda_{cls} L_{cls} + \lambda_{usage} L_{usage}, \qquad (4)$$

де L_{loc} – функція втрат для регресії координат виявлених об'єктів на зображенні;

L_{cls} – функція втрат для класифікації виявлених об'єктів на зображені;

L_{usage} – функція втрата, що характеризує відхилення бажаного динамічного ступеня стиснення нейронної мережі від реального;

$\lambda_{loc}\lambda_{cls}\lambda_{usage}$ – коефіцієнти балансу між різними компонентами комбінованої функції втрат.

Функція втрат для регресії координат обмежувальної рамки обчислюється за формулою:

$$L_{loc} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j\in\{x,y,w,h\}} \mathrm{smooth}_{L1}\left(P_{i,j} - T_{i,j}\right), \quad (5)$$

де $P_{i,j}$ – прогнозовані значення різниці між координатами і розмірів якірної і цільової комірки;

$T_{i,j}$ – реальні значення різниці координат і розмірів якірної і цільової комірки;

$$smoot\square_{L1}(x) = \begin{cases} 0{,}5x^2, & якщо\ |x| < 1; \\ |x| - 0{,}5, & якщо\ |x| \geq 1; \end{cases} \qquad (6)$$

Функція втрат для класифікації обчислюються за формулою функції фокальних втрат

$$L_{cls} = -\left(1-p_t\right)^{\gamma} \log\left(p_t\right), \qquad (7)$$

де p_t – ймовірність прогнозування і-го класу;

γ – параметр фокусування.

Функція (7) є вдосконаленою функцією крос-ентропії. Відмінність полягає в додаванні параметра $\gamma \in (0, +\infty)$, який вирішує проблему незбалансованих класів. Під час навчання більшість об'єктів, оброблених класифікатором, є фоном, який є окремим класом. Тому може виникнути проблема, коли нейромережа навчається краще виявляти фон, ніж інші об'єкти. Додатковий параметр вирішує цю проблему шляхом зменшення значення помилки для легко класифікованих класів об'єктів.

Функція L_{usage}, яка обчислює відхилення бажаного ступеню динамічного стиснення нейромережі від реального. Ця функція залежить від кількості активованих блоків основної моделі на основі рішень, прийнятих гейтовими модулями. Функція L_{usage} обчислюється аналогічно до [9, 10]:

$$L_{usage} = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{1}{n}\sum_{i=1}^{n} g_{k,i} - t_k\right)^2, \qquad (8)$$

де $g_{k,i}$ – вихід k-го модуля для і-го екземпляру навчальних даних;

t_k – приблизним цільовим рівнем виконання кожного блоку нейромережі на міні-партії даних, $t_k \in (0,1)$ ($t_k = 0{,}5$ за замовчуванням);

K – це кількість гейтових модулів, які контролюють активацію K блоків нейромереж;

n – це розмір навчальної міні-партії.

Для підвищення робастності до шуму та пошкодження ваг алгоритм навчання детектора об'єктів базується на таких принципах:

– одночасне навчання ваг основної мережі та ваг гейтових модулів;

– навчання спочатку проводиться на основному наборі даних в звичайних умовах, а потім в умовах епізодичного навчання з невеликою кількістю прикладів адаптації до кожного типу синтетичних збурень;

– генерація синтетичних збурень даних або ваг згідно зі сценарієм білого ящику;

– узагальнення досвіду під час адаптації до збурень повинно базуватися на мета-оновленні з використанням алгоритму MAML, REPTIL або вагового усереднення.

Забезпечення робастності мережі детекції об'єктів в умовах обмежених ресурсів полягає в оптимізації основної мережі та гейтових модулів таким чином, щоб в режимі екзамену обчислювалися лише ті компоненти нейромережі, які можуть забезпечити найточніший прогноз в умовах збурень.

Нехай $\tau_i \mid \left\{i = \overline{1, N}\right\}$ це набір реалізацій збурень, які стосуються системи детекції об'єктів на зображеннях [33]. Як збурення τ_i розглядаються шум

ISSN 1727-7337 (print)
ISSN 2663-2217 (online)

протиборчих атак та пошкодження ваг мережі. Нехай $D = \{D^{tr}; D^{val}\}$ це набір даних, на якому модель була навчена виконувати основне завдання у відомих умовах. Нехай адаптація моделі здійснюється за результатами вирішення K задач навчання на невеликій кількості прикладів (few-shot learning) з набору даних $D = \{D_k^{tr}; D_k^{val} | k = \overline{1,K}\}$, що з D. Навчальні підвибірки D_k^{tr} використовуються на етапі точної підгонки, а тестовий набір D_k^{val} використовується на етапі мета-оновлення. Також задано набір параметрів $\Psi = \langle \Theta, \phi \rangle$, та W, де Θ – параметри базового екстрактору ознак системи; ϕ – параметри гейтових модулів; W – параметри піраміди ознак та детектуючої головки.

Мета-навчання на основі градієнтів передбачає знаходження таких значень параметрів Ψ^*, які забезпечать мінімізацію очікуваної функції втрат L на наборі реалізацій різних типів збурень τ_i під час адаптації на $D_{\tau i}$

$$\Psi^* = \arg \min_{\Psi} \mathop{E}_{\tau \sim p(\tau)} \left[L_{\tau_i} \left(U_{\tau_i} \left(\Psi, W_{\tau_i}, D_{\tau_i} \right) \right) \right], \quad (9)$$

де U – оператор, який поєднує вплив збурення та адаптацію за T кроків, який відображає поточний стан Ψ у новий стан Ψ.

Псевдокод мета-навчального алгоритму для підвищення стійкості системи штучного інтелекту показаний на рис. 5. Алгоритм стохастичного градієнтного спуску (SGD) в операторі U за T кроків виконує мета-оновлення градієнту параметрів Ψ. Для спрощення обчислень та збільшення стабільності параметри, що мета-оптимізуються, можуть бути оновлені за допомогою алгоритму REPTIL [34]. Тип збурюючого впливу не змінюється протягом одного кроку мета-адаптації. Однак кожен крок мета-адаптації починається з вибору типу збурюючого впливу, за яким слідує генерація n реалізацій збурюючого впливу з подальшою вкладеною петлею адаптації для кожного з них.

Як показано на рис. 5, формування атак на зразки вхідних даних реалізується за допомогою функції Adversarial_perturbation ().

Запропоновано використовувати протиборчі атаки типу «білого ящику» для мета-навчання, наприклад, можна використовувати атаки FGSM або атаки PGD [24]. Для тестування запропоновано алгоритми протиборчих атак типу “чорного ящику”. Прикладом може бути атака на основі алгоритму пошуку стратегії еволюції коваріантної матриці (CMA-ES) [35]. Інтенсивність впливу збурення обмежується L_∞-norm or L_0-norm. У цьому випадку, якщо зображення нормалізується шляхом ділення яскравості пікселя на 255, то вказана інтенсивність збурень також ділиться на 255.

Вхідні дані: $p(\tau)$ – розподіл реалізацій збурень;
α, β – гіперпараметри кроку навчання;
T – кількість кроків адаптації;
D – набір даних.

Попереднє навчання параметрів Ψ на даних D
Доки не виконана задану кількість епох виконувати:
  Вибрати тип збурення з множини
  { інєкція помилок у ваги мережі, протиборча атака ухилення}
  Сформувати вибірку реалізацій збурення
  $\tau_i \sim p(\tau), i = \overline{1,n}$
  Для i=1,2,…,n виконувати:
    Копіювати поточні значення параметрів:
    $\dot{\Psi}_{\tau_i}, \dot{W}_{\tau_i} \leftarrow copy(\Psi, W_{base})$
    Вибірка навчальних $D_{\tau_i}^{tr}$ та валідаційних $D_{\tau_i}^{val}$
    зразків з набору D
    Якщо тип збурення є інєкцією помилок у ваги мережі:
      $\dot{\Psi}_{\tau_i} \leftarrow Fault\_injection(\dot{\Psi}_{\tau_i}, \dot{W}_{\tau_i}, D_{\tau_i}^{tr})$
    Якщо тип збурення є протиборчою атакою ухилення:
      $D_{\tau_i}^{tr}, D_{\tau_i}^{val} \leftarrow Adversarial\_perturbation(D_{\tau_i}^{tr}, D_{\tau_i}^{val})$
    $\Psi_{\tau_i} \leftarrow SGD_{\Psi}(L_{\tau_i}(\dot{\Psi}_{\tau_i}, \dot{W}_{\tau_i}, D_{\tau_i}^{tr}), T, \alpha)$
    $\Psi \leftarrow \Psi + \beta(\Psi_{\tau_i} - \Psi)$

Рис. 5. Псевдокод мета-навчання для оптимізації робастності мережі детекції об’єктів

Функція Fault_injection() генерує несправності для впливу на тензори нейромережі [36]. Запропоновано вибрати найважчий тип несправності для поглинання, що включає в себе генерацію інверсії випадково вибраного біту (введення перепаду бітів) у ваговому коефіцієнті моделі. Під час навчання запропоновано пошкодити найбільш чутливі ваги. Для визначення таких ваг, тестові набори даних повинні бути проходити через мережу, і обчислювати градієнти, які потім можна відсортувати за їх абсолютними значеннями. Для топ k градієнтів ваг, один біт інвертується на випадковій позиції. Запропоновано генерувати пошкодження випадкових ваг для тестових цілей.

4. Експерименти

Сучасні бортові системи безпілотних літальних апаратів дедалі частіше оснащуються допоміжними комп'ютерами для підвищення їхньої автономності в складних умовах. Популярними одноплатними комп'ютерами є Orange Pi 5 та Raspberry Pi 5 з продуктивністю процесорів у діапазоні від 0,1 до 0,2 TOPS. Вхідні дані для такої нейронної мережі надає циф-

рова камера, підключена до одноплатного комп'ютера через інтерфейс MIPI CSI, а комунікація з приймачем команд і польотним контролером здійснюється через інтерфейси UART.

Екстрактор ознак ViT-S/16, що складає основну частину детектора, має складність що не перевищує 40 GFLOPs, яка в результаті стиснення динамічними механізмами може становити менше 30 GFLOPs разом з пірамідою ознак і детектуючими головками. Тобто швидкість оброблення кадрів може становити від 3 до 6 кадрів за секунду. Однак це не є проблемою, якщо детектор поєднується зі швидкими трекерами типу ByteTrack чи звичайним фільтром Калмана. Для побудови динамічного обчислювального графу мережі буде використано фреймворк PyTorch.

Попереднє навчання екстрактора ознак проводиться методом самодистиляції без міток (DINO) на наборі даних ImageNet-1k з роздільною здатністю 512x512 пікселів. Під час точного налаштування використовується кроковий коефіцієнт навчання, починаючи з 0,1 та зменшуючи його в 10 разів після 150 та 250 епох. Remote Sensing Object Detection (RSOD) - це набір даних анотованих аерозображень з роздільною здатністю 512x512, який використовується для моделювання основного завдання у навчанні з учителем [37].

Функція втрат (8) має такі коефіцієнти компонентів: $\lambda_{\text{loc}} = 0{,}5,\ \lambda_{\text{cls}} = 0{,}5,\ \lambda_{\text{usage}} = 2.$ Запропоновано оцінити обчислювальну складність моделі в режимі екзамену як середні значення FLOPs на тестовому наборі даних. Це впливає на параметр t_k, який також можна назвати динамічним коефіцієнтом стиснення. Цей параметр є приблизним цільовим рівнем виконання кожного блоку нейромережі на міні-партії даних під час фази навчання.

Усереднене значення середньої Precision (mAP) для всіх категорій детекції використовується як метрика оцінювання детектора об'єктів на тестових аерозображеннях. Середня точність кожного класу обчислюється як площа під кривою Precision-Recall. Далі mAP визначається як mAP@0,5, що представляє середню Precision, коли поріг IoU встановлено на 0,5. Враховуючи елементи випадковості, запропоновано використовувати середні значення при оцінці mAP та FLOPs. Для цього генерується та застосовується 100 екземплярів певного типу перешкод до однієї й тієї ж моделі або набору даних.

Таблиця 1 показує результати тестування моделі ViT-S/16, навченої на наборі imagenet-1k та наборі аерозображень RSOD в умовах впливу ін'єкції пошкодження ваг (відсоток змінених ваг встановлюється за допомогою частки пошкоджень, fault_rate = 0,1). Навчання на аерозображеннях проводиться окремо для кожного значення динамічного коефіцієнта стиснення для порівняння результатів та вибору найбільш компромісного варіанту.

Таблиця 2 показує результати тестування моделі ViT-S/16, попередньо навченої на ImageNet-1k та наборі аерозображень RSOD, на тестових зображеннях, спотворених протиборчою атакою (амплітуда збурення 3/255 згідно з L_∞). Навчання на зображеннях RSOD проводиться окремо для кожного значення динамічного коефіцієнта стиснення для порівняння результатів та вибору оптимального варіанту. mAP також використовується як метрика оцінювання детектора об'єктів на зображенні.

Таблиця 1

Залежність усереднених значень mAP та FLOPs під час тестування в умовах впливу інєкції пошкодження ваг від методу попереднього навчання та заданого коефіцієнта динамічного стиснення

Коефіцієнт стиснення	Попередньо навчена модель в звичайних умовах		Попередньо навчена модель під впливом протибочої атаки		Мета-навчена модель	
	FLOPs (10^9)	mAP	FLOPs (10^9)	mAP	FLOPs (10^9)	mAP
1,0	39,3	0,83	39,3	0,86	39,3	0,93
0,8	35,1	0,83	34,8	0,87	33,8	0,94
0,6	**32,6**	**0,84**	**31,7**	**0,87**	**31,0**	**0,94**
0,4	28,9	0,82	28,0	0,81	27,6	0,90
0,2	26,5	0,71	26,1	0,72	25,2	0,82

Аналіз Таблиці 1 показує, що попереднє навчання в умовах впливу пошкодження ваг з оптимальним коефіцієнтом стиснення мережі дозволяє збільшити mAP під час тестування в умовах впливу ін'єкції пошкодження ваг на 3,5 %. У цьому випадку мета-навчання на основі результатів адаптації до збурень забезпечує збільшення mAP на 11,9 % під час тестування в умовах впливу ін'єкції пошкодження ваг.

Аналіз Таблиці 2 показує, що попереднє навчання в умовах впливу протиборчих атак з оптимальним коефіцієнтом стиснення мережі дозволяє збільшити mAP під час тестування в умовах впливу протиборчої атаки на 3,6 %. А мета-навчання на результатах адаптації до перешкод забезпечує збільшення mAP під час тестування в умовах впливу протиборчих атак на 13,2 %.

Аналіз Таблиці 1 та Таблиці 2 показує, що зменшення коефіцієнта динамічного стиснення (тобто збільшення стиснення) спочатку приводить до покращення валідаційної метрики. Однак подальше змен-

ISSN 1727-7337 (print)
ISSN 2663-2217 (online)

шення коефіцієнта динамічного стиснення призводить до зменшення значення валідаційної метрики. Без навчання або мета-навчання в умовах збурень зменшення коефіцієнта стиснення завжди призводить до зменшення значення валідаційної метрики під час тестування в умовах збурень. Однак навчання або мета-навчання в умовах збурень дозволяє навіть зі зменшеним коефіцієнтом стиснення трохи збільшити валідаційну метрику моделі. Тим не менш, після досягнення певного порогового значення коефіцієнта стиснення подальше його зменшення призводить до зменешення значення валідаційної метрики.

Таблиця 2

Залежність усереднених значень mAP та FLOPs під час тестування в умовах впливу протиборчої атаки від методу попереднього навчання та заданого коефіцієнта динамічного стиснення

Коефіцієнт стиснення	Попередньо навчена модель в звичайних умовах		Попередньо навчена модель під впливом протибочої атаки		Мета-навчена модель	
	FLOPs (10^9)	mAP	FLOPs (10^9)	mAP	FLOPs (10^9)	mAP
1,0	39,3	0,82	39,3	0,85	39,3	0,93
0,8	36,4	0,82	36,1	0,86	35,8	0,93
0,6	**34,3**	**0,83**	**32,2**	**0,86**	**31,2**	**0,94**
0,4	31,2	0,79	29,2	0,81	28,1	0,89
0,2	27,1	0,70	26,9	0,71	25,9	0,81

Тестування оптимальної моделі ViT-S/16 на наборі даних RSOD з двома різними типами перешкод показало, що mAP досягла значення 0,94, а усереднена обчислювальна складність становить не більше 34,3 GFLOPs. Таким чином, зменшення FLOPs на 20 % навіть привело до збільшення mAP на 1 % в умовах динамічного стиснення порівняно з випадком без стиснення (динамічний коефіцієнт стиснення дорівнює 1,0).

5. Обговорення

Результати з Таблиці 1 та Таблиці 2 підтверджують зменшення FLOPs великої мережі під час екзамену до рівня менших нейромереж за рахунок застосування динамічних механізмів до екстрактора ознак. Для впровадження динамічного екстрактора ознак до енкодерів мережі додаються гейт модулі (2), які додають 1,76 % фіксованих накладних витрат на операції з плаваючою комою. З певним ступенем стиснення обсяг обчислень в режимі екзамену зменшується, а середнє значення точності детекції об'єктів (mAP) в умовах впливу перешкод практично не змінюється або навіть збільшується.

У [38] мережа Yolo5l на наборі даних RSOD без впливу пертурбацій забезпечує mAP у розмірі 82,47 та потребує 88 GFLOPs для режиму екзамену на одному зображенні. В цьому випадку запропонований детектор під впливом ін'єкції пошкодження ваг при ступені стиснення 0,6 потребує лише 32,6 GFLOPs, але забезпечує mAP, що становить 0,94. Таким чином, отримані результати в умовах впливу ін'єкції перешкод перевершують відомі результати, отримані в нормальних умовах без пошкодження ваг. Крім того, якщо допустити зниження mAP на 5 %, можна зекономити 10% FLOPs.

Існування оптимального стиснення, яке економить ресурси та навіть трохи збільшує mAP, може бути пояснене тим, що вразливі та несуттєві частини нейромережі адаптивно вимикаються. Схожа здатність механізму раннього виходу збільшувати стійкість до несправностей та стійкість до атак розглядалася в [17, 18].

Аналіз Таблиці 1 та Таблиці 2 показав, що навчання в умовах збурень значно збільшує mAP під час тестування в умовах впливу збурення порівняно з попереднім навчанням в звичайних умовах. Так само, мета-навчання на результатах адаптації до збурень дозволяє збільшити mAP під час тестування в умовах впливу збурень порівняно з простим попереднім навчанням в умовах збурень. Це може бути пояснене збільшеною ефективністю гейтових модулів. Ймовірно, з належним навчанням гейтові модулі краще розпізнають пертурбації, які негативно впливають на продуктивність кодувальника.

6. Висновки

Вперше розроблено модель детекції об'єктів на аерозображеннях з динамічним екзаменом та оптимізованою робастністю. Ця модель складається з екстрактора ознак на основі візуального трансформера, гейтових модулів та спрощеної мережі піраміди ознак, а також детектуючої головки RetinaNet. Гейтові модулі навчені вимикати трансформерні кодувальники, які не достатньо релевантні для оброблення певних вхідних даних та збурень. Запропонована модель дозволила зменшити FLOPs на більш ніж на 20 % без зменшення значень валідаційних метрик.

Вперше розроблено метод навчання для детекції об'єктів, який поєднує функцію втрат RetinaNet з функцією втрат гейтових блоків та застосовує мета-навчання на результатах адаптації до синтетичних збурень. Аналіз експериментальних результатів показав, що попереднє навчання на анотованих даних під впливом ін'єкції пошкодження ваг з оптимальним

коєфіцієнтом стиснення дозволило збільшити mAP на 3,5 %, тоді як мета-навчання на результатах адаптації до збурень привело до збільшення mAP на 11,9 %. Крім того, попереднє навчання на анотованих даних в умовах впливу протиборчої атаки з оптимальним ступенем стиснення покращило mAP на 3,6 %, а мета-навчання на результатах адаптації до збурень забезпечило збільшення mAP на 13,2 %.

Дана модель, навчена за запропонованим методом під впливом введення несправностей, продемонструвала mAP, що на 13,9 % більшу ніж в Yolo5l, яка тестується без впливу ін'єкції пошкодження ваг. Крім того, запропонована модель вимагала в 2,7 разів менше FLOPs, ніж Yolo5l, що свідчить про вищу обчислювальну ефективність порівняно з іншими популярними підходами.

Майбутні дослідження будуть присвячені удосконаленню детектора об'єктів за рахунок аналізу просторово-часової інформації та цілеспрямованого керування польотом для збору додаткових даних для зняття невизначеності в певних ділянках простору.

Внесок авторів: розроблення концептуальних положень та методології дослідження, формулювання висновків – **Альона Москаленко**; огляд та аналіз джерел, розроблення програмного забезпечення для моделювання – **Микола Зарецький**; розроблення математичних моделей, аналіз результатів дослідження – **Максим Виноградов**; розроблення алгоритму навчання – **Владислав Бабич**.

Конфлікт інтересів

Автори заявляють, що не мають конфлікту інтересів щодо цього дослідження, будь то фінансовий, особистий, авторський або інший, що могло б вплинути на дослідження та його результати, представлені в цій статті.

Фінансування

Робота виконана в рамках проекту «Виконання завдань перспективного плану розвитку наукового напряму "Технічні науки" Сумського державного університету» (№ держреєстрації 0121U112684), що фінансується Міністерством освіти і науки України.

Наявність даних

Рукопис має пов'язані дані в репозиторії даних: набори даних RSOD, доступний в [38] та може бути знайдений тут https://github.com/RSIA-LIESMARS-WHU/RSOD-Dataset- , доступний на 7 жовтня 2024 року. Набір даних ILSVRC2012 можна знайти тут http://image-net.org/challenges/LSVRC/2012/index, доступний на 7 жовтня 2024 року.

Використання штучного інтелекту

Автори використовували технології штучного інтелекту для надання перевірених результатів. Написання тексту статті було здійснено без використання технологій штучного інтелекту.

Подяки

Автори висловлюють подяку Міністерству освіти і науки України за підтримку Лабораторії Інтелектуальних Систем в рамках дослідницького проекту «Виконання завдань перспективного плану розвитку наукового напряму "Технічні науки" Сумського державного університету» (№ держреєстрації 0121U112684)".

Всі автори прочитали та погодилися з опублікованою версією рукопису.

Література

1. *Polina, A. M. Aerial object detection analysis: Challenges and preliminary results [Electronic resource] / Agnes Maria Polina, Hari Suparwito, & Rosalia Arum Kumalasanti // E3S Web of Conferences. – 2024. – Vol. 475. – Article No. 02017. DOI: 10.1051/e3sconf/202447502017.*

2. *Deep Neural Networks Compression: a comparative survey and choice recommendations [Electronic resource] / Giosué Cataldo Marinó, & et al. // Neurocomputing. – 2022. DOI: 10.1016/j.neucom.2022.11.072.*

3. *A survey on fault-tolerant methodologies for deep neural networks [Text] / Rizwan Syed, & et al. // Pomiary automatyka robotyka. – 2023. – Vol. 27, iss. 2. – P. 89–98. DOI: 10.14313/par_248/89.*

4. *Diversity supporting robustness: enhancing adversarial robustness via differentiated ensemble predictions [Electronic resource] / Xi Chen, & et al. // Computers & security. – 2024. – Vol. 142. – Article No. 103861. – DOI: 10.1016/j.cose.2024.103861.*

5. *On the limitations of denoising strategies as adversarial defenses [Electronic resource] / Z. Niu, & et al. // ArXiv. – 2020. DOI: 10.48550/arXiv.2012.09384.*

6. *Lust, J. Efficient Detection of Adversarial, Out-of-distribution and Other Misclassified Samples [Electronic resource] / Julia Lust, & Alexandru P. Condurache // Neurocomputing. – 2021. DOI: 10.1016/j.neucom.2021.05.102.*

7. *Adversarial Robust Aerial Image Recognition Based on Reactive-Proactive Defense Framework with Deep Ensembles [Electronic resource] / Zihao Lu, & et al. // Remote Sensing. – 2023. – Vol. 15, iss. 19. – Article No. 4660. DOI: 10.3390/rs15194660.*

8. *A survey of modern deep learning based object detection models [Electronic resource] / Syed Sahil*

ISSN 1727-7337 (print)
ISSN 2663-2217 (online)

Abbas Zaidi, & et al. // Digital signal processing. – 2022. – Vol. 126. – Article No. 103514. DOI: 10.1016/j.dsp.2022.103514.

9. Han Y., Huang G., Song S., Yang L., Wang H., Wang Y. Dynamic Neural Networks: A Survey. arXiv, 2021. DOI: 10.48550/ arXiv.2102.04906.

10. AdaViT: adaptive vision transformers for efficient image recognition [Electronic resource] / Lingchen Meng, & et al. // 2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR), New Orleans, LA, USA, 18 – 24.06.2022. – 2022. DOI: 10.1109/cvpr52688.2022.01199.

11. Paul, S. Vision transformers are robust learners [Text] / Sayak Paul, & Pin-Yu Chen // Proceedings of the AAAI conference on artificial intelligence. – 2022. – Vol. 36, iss2. – P. 2071–2081. DOI: 10.1609/aaai.v36i2.20103.

12. Kim, J. PQK: Model compression via pruning, quantization, and knowledge distillation [Electronic resource] / J. Kim, S. Chang, & N. Kwak // ArXiv. – 2021. DOI: 10.48550/arXiv.2106.14681.

13. Ps and Qs: quantization-aware pruning for efficient low latency neural network inference [Electronic resource] / Benjamin Hawks, & et al. // Frontiers in artificial intelligence. – 2021. – Vol. 4. DOI: 10.3389/frai.2021.676564.

14. Haque, M. Dynamic neural network is all you need: understanding the robustness of dynamic mechanisms in neural networks [Electronic resource] / Mirazul Haque, & Wei Yang // 2023 IEEE/CVF international conference on computer vision workshops (ICCVW), Paris, France, 2 – 6.10.2023. –2023. DOI: 10.1109/iccvw60793.2023.00163.

15. Resilience and resilient systems of artificial intelligence: taxonomy, models and methods [Electronic resource] / Viacheslav Moskalenko, & et al. // Algorithms. – 2023. – Vol. 16, iss. 3. – Article No. 165. DOI: 10.3390/a16030165.

16. Moskalenko, V. Neural network based image classifier resilient to destructive perturbation influences – architecture and training method [Text] / Viacheslav Moskalenko, & Alona Moskalenko // Radioelectronic and computer systems. – 2022. – No 3. – P. 95–109. DOI: 10.32620/reks.2022.3.07.

17. Aegis: mitigating targeted bit-flip attacks against deep neural networks [Electronic resource] / J. Wang, & et al. // ArXiv. – 2023. DOI: 10.48550/ arXiv.2302.13520.

18. Haque, M. Dynamic neural network is all you need: understanding the robustness of dynamic mechanisms in neural networks [Electronic resource] / Mirazul Haque, & Wei Yang // 2023 IEEE/CVF international conference on computer vision workshops (ICCVW), Paris, France, 2 – 6.10.2023. – 2023. DOI: 10.1109/iccvw60793.2023.00163.

19. Transient-Fault-Aware design and training to enhance dnns reliability with zero-overhead [Electronic resource] / Niccolo Cavagnero, & et al. // 2022 IEEE 28th international symposium on on-line testing and robust system design (IOLTS), Torino, Italy, 12–14.09. 2022. – 2022. DOI: 10.1109/iolts56730.2022.9897813.

20. One-bit flip is all you need: when bit-flip attack meets model training [Electronic resource] / Jianshuo Dong, & et al. // 2023 IEEE/CVF international conference on computer vision (ICCV), Paris, France, 1 – 6.10.2023. – 2023. DOI: 10.1109/iccv51070.2023.00432.

21. Cao, H. Adversarial training for better robustness [Electronic resource] / Houze Cao, & Meng Xue // Lecture notes of the institute for computer sciences, social informatics and telecommunications engineering. – Cham, 2023. – P. 75–84. DOI: 10.1007/978-3-031-35982-8_6.

22. Recent advances in adversarial training for adversarial robustness [Electronic resource] / Tao Bai, & et al. // Thirtieth international joint conference on artificial intelligence {IJCAI-21}, Montreal, Canada, 19–27.08.2021. – California, 2021. DOI: 10.24963/ ijcai.2021/591.

23. Athalye, A. Obfuscated gradients give a false sense of security: circumventing defenses to adversarial examples [Electronic resource] / A. Athalye, N. Carlini, & D. Wagner // ArXiv. – 2018. DOI: 10.48550/arXiv. 1802.00420.

24. Mitigating adversarial attacks for deep neural networks by input deformation and augmentation [Electronic resource] / Pengfei Qiu, & et al. // 2020 25th asia and south pacific design automation conference (ASP-DAC), Beijing, China, 13 – 16.01.2020. – 2020. DOI: 10.1109/asp-dac47756.2020.9045107.

25. Challenging the adversarial robustness of dnns based on error-correcting output codes [Text] / Bowen Zhang, & et al. // Security and communication networks. – 2020. – Vol. 2020. – P. 1–11. DOI: 10.1155/2020/8882494.

26. Maurício, J. Comparing vision transformers and convolutional neural networks for image classification: a literature review [Electronic resource] / José Maurício, Inês Domingues, & Jorge Bernardino // Applied sciences. – 2023. – Vol. 13, iss. 9. – Article No. 5521. DOI: 10.3390/app13095521.

27. Moskalenko, V. Object detection with affordable robustness for UAV aerial imagery: model and providing method [Text] / Viacheslav Moskalenko, Andrii Korobov, & Yulia Moskalenko // Radioelectronic and computer systems. – 2024. – No. 3. – P. 55–66. DOI: 10.32620/reks.2024.3.04.

28. Wang, L. Aerial image object detection with vision transformer detector (vitdet) [Electronic resource] / Liya Wang, & Alex Tien // IGARSS 2023 - 2023 IEEE

international geoscience and remote sensing symposium, Pasadena, CA, USA, 16–21.07.2023. – 2023. DOI: 10.1109/igarss52108.2023.10282836.

29. Exploring plain vision transformer backbones for object detection [Text] / Yanghao Li, et al. // Lecture notes in computer science. – Cham, 2022. – P. 280–296. DOI: 10.1007/978-3-031-20077-9_17.

30. Parking lot occupancy detection with improved mobilenetv3 [Text] / Yusufbek Yuldashev, & et al. // Sensors. – 2023. – Vol. 23, iss. 17. – Article No. 7642. DOI: 10.3390/s23177642.

31. AI-based object detection latest trends in remote sensing, multimedia and agriculture applications [Electronic resource] / Saqib Ali Nawaz, & et al. // Frontiers in plant science. – 2022. – Vol. 13. DOI: 10.3389/fpls.2022.1041514.

32. Emerging properties in self-supervised vision transformers [Electronic resource] / Mathilde Caron, & et al. // 2021 IEEE/CVF international conference on computer vision (ICCV), Montreal, QC, Canada, 10–17.10.2021. – 2021. DOI: 10.1109/iccv48922.2021.00951.

33. Development of self-supervised learning with dinov2-distilled models for parasite classification in screening [Electronic resource] / Natchapon Pinetsuksai, & et al. // 2023 15th international conference on information technology and electrical engineering (ICITEE), Chiang Mai, Thailand, 26–27.10.2023. – 2023. DOI: 10.1109/icitee59582.2023.10317719.

34. Optimisation of deep neural network model using Reptile meta learning approach [Electronic resource] / Uday Kulkarni, & et al. // Cognitive computation and systems. – 2023. DOI: 10.1049/ccs2.12096.

35. Kotyan, S. Adversarial robustness assessment: Why in evaluation both L0 and L∞ attacks are necessary [Electronic resource] / Shashank Kotyan, & Danilo Vasconcellos Vargas // Plos one. – 2022. – Vol. 17, iss. 4. – Article No. e0265723. DOI: 10.1371/journal.pone.0265723.

36. Li, G. TensorFI: a configurable fault injector for tensorflow applications [Electronic resource] / Guanpeng Li, Karthik Pattabiraman, & Nathan DeBardeleben // 2018 IEEE international symposium on software reliability engineering workshops (ISSREW), Memphis, TN, 15–18.10.2018. – 2018. DOI: 10.1109/issrew.2018.00024.

37. A method for detecting lightweight optical remote sensing images using improved yolov5n [Text] / ChangMan Zou, & et al. // Journal of multimedia information system. – 2023. – Vol. 10, iss. 3. – P. 215–226. DOI: 10.33851/jmis.2023.10.3.215.

38. Xie, T. YOLO-RS: a more accurate and faster object detection method for remote sensing images [Electronic resource] / Tianyi Xie, Wen Han, & Sheng Xu // Remote sensing. – 2023. – Vol. 15, iss. 15. – Article No. 3863. DOI: 10.3390/rs15153863.

References

1. Polina, A. M., Suparwito, H., & Kumalasanti, R. A. Aerial object detection analysis: Challenges and preliminary results. *E3S Web of Conferences*, 2024, vol. 475, article no. 02017. DOI: 10.1051/e3sconf/202447502017.

2. Marinó, G. C., Petrini, A., Malchiodi, D., & Frasca, M. Deep neural networks compression: A comparative survey and choice recommendations. *Neurocomputing*, 2023, vol. 520, pp. 152–170. DOI: 10.1016/j.neucom.2022.11.072.

3. Syed, R., Ulbricht, M., Piotrowski, K., & Krstic, M. A Survey on Fault-Tolerant Methodologies for Deep Neural Networks. *Pomiary Automatyka Robotyka*, 2023, vol. 27, iss. 2, pp. 89-98. DOI: 10.14313/par_248/89.

4. Chen, X., Huang, W., Peng, Z., Guo, W., & Zhang, F. Diversity supporting robustness: Enhancing adversarial robustness via differentiated ensemble predictions. *Computers & Security*, 2024, vol. 142, article no. 103861. DOI: 10.1016/j.cose.2024.103861.

5. Niu, Z., Chen, Z., Li, L., Yang, Y., Li, B., & Yi, J. On the Limitations of Denoising Strategies as Adversarial Defenses. *arXiv*, 2020. DOI: 10.48550/ARXIV.2012.09384.

6. Lust, J., & Condurache, A. P. Efficient detection of adversarial, out-of-distribution and other misclassified samples. *Neurocomputing*, 2022, vol. 470, pp. 335–343. DOI: 10.1016/j.neucom.2021.05.102.

7. Lu, Z., Sun, H., Ji, K., & Kuang, G. Adversarial Robust Aerial Image Recognition Based on Reactive-Proactive Defense Framework with Deep Ensembles. *Remote Sensing*, 2023, vol. 15, iss. 19, article no. 4660. DOI: 10.3390/rs15194660.

8. Zaidi, S. S. A., Ansari, M. S., Aslam, A., Kanwal, N., Asghar, M., & Lee, B. A survey of modern deep learning based object detection models. *Digital Signal Processing*, 2022, vol. 126, article no. 103514. DOI: 10.1016/j.dsp.2022.103514.

9. Han, Y., Huang, G., Song, S., Yang, L., Wang, H., & Wang, Y. Dynamic Neural Networks: A Survey. *arXiv*, 2021. DOI: 10.48550/ARXIV.2102.04906.

10. Meng, L., Li, H., Chen, B.-C., Lan, S., Wu, Z., Jiang, Y.-G., & Lim, S.-N. AdaViT: Adaptive Vision Transformers for Efficient Image Recognition. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, article no. 01199. DOI: 10.1109/cvpr52688.2022.01199.

11. Paul, S., & Chen, P.-Y. Vision Transformers Are Robust Learners. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, vol. 36, iss. 2, pp. 2071-2081. DOI: 10.1609/aaai.v36i2.20103.

ISSN 1727-7337 (print)
ISSN 2663-2217 (online)

12. Kim, J., Chang, S., & Kwak, N. PQK: Model Compression via Pruning, Quantization, and Knowledge Distillation. *arXiv*, 2021. DOI: 10.48550/ARXIV.2106.14681.

13. Hawks, B., Duarte, J., Fraser, N. J., Pappalardo, A., Tran, N., & Umuroglu, Y. Ps and Qs: Quantization-Aware Pruning for Efficient Low Latency Neural Network Inference. *Frontiers in Artificial Intelligence*, 2021, vol. 4. DOI: 10.3389/frai.2021.676564.

14. Haque, M., & Yang, W. Dynamic Neural Network is All You Need: Understanding the Robustness of Dynamic Mechanisms in Neural Networks. *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2023, pp. 1489–1498. DOI: 10.1109/iccvw60793.2023.00163.

15. Moskalenko, V., Kharchenko, V., Moskalenko, A., & Kuzikov, B. Resilience and Resilient Systems of Artificial Intelligence: Taxonomy, Models and Methods. *Algorithms*, 2023, vol. 16, iss. 3, article no. 165. DOI: 10.3390/a16030165.

16. Moskalenko, V., & Moskalenko, A. Neural network based image classifier resilient to destructive perturbation influences – architecture and training method. *Radioelectronic and Computer Systems*, 2022, no. 3, pp. 95-109. DOI: 10.32620/reks.2022.3.07.

17. Wang, J., Zhang, Z., Wang, M., Qiu, H., Zhang, T., Li, Q., Li, Z., Wei, T., & Zhang, C. Aegis: Mitigating Targeted Bit-flip Attacks against Deep Neural Networks. *arXiv*, 2023. DOI: 10.48550/ARXIV.2302.13520.

18. Haque, M., & Yang, W. Dynamic Neural Network is All You Need: Understanding the Robustness of Dynamic Mechanisms in Neural Networks. *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2023, pp. 1489-1498. DOI: 10.1109/iccvw60793.2023.00163.

19. Cavagnero, N., Santos, F. D., Ciccone, M., Averta, G., Tommasi, T., & Rech, P. Transient-Fault-Aware Design and Training to Enhance DNNs Reliability with Zero-Overhead. *IEEE 28th International Symposium on On-Line Testing and Robust System Design (IOLTS)*, 2022. DOI: 10.1109/iolts56730.2022.9897813.

20. Dong, J., Qiu, H., Li, Y., Zhang, T., Li, Y., Lai, Z., Zhang, C., & Xia, S.-T. One-bit Flip is All You Need: When Bit-flip Attack Meets Model Training. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 4665-4675. DOI: 10.1109/iccv51070.2023.00432.

21. Cao, H., & Xue, M. Adversarial Training for Better Robustness. *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Springer Nature Switzerland*, 2023, pp. 75-84. DOI: 10.1007/978-3-031-35982-8_6.

22. Bai, T., Luo, J., Zhao, J., Wen, B., & Wang, Q. Recent Advances in Adversarial Training for Adversarial Robustness. *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21)*, 2021, pp. 4312-4321. DOI: 10.24963/ijcai.2021/591.

23. Athalye, A., Carlini, N., & Wagner, D. Obfuscated Gradients Give a False Sense of Security: Circumventing Defenses to Adversarial Examples. *arXiv*, 2018. DOI: 10.48550/ARXIV.1802.00420.

24. Qiu, P., Wang, Q., Wang, D., Lyu, Y., Lu, Z., & Qu, G. Mitigating Adversarial Attacks for Deep Neural Networks by Input Deformation and Augmentation. *2020 25th Asia and South Pacific Design Automation Conference (ASP-DAC)*, 2020, pp. 157-162. DOI: 10.1109/asp-dac47756.2020.9045107.

25. Zhang, B., Tondi, B., Lv, X., & Barni, M. Challenging the Adversarial Robustness of DNNs Based on Error-Correcting Output Codes. *Security and Communication Networks*, 2020, vol. 2020, pp. 1–11. DOI: 10.1155/2020/8882494.

26. Maurício, J., Domingues, I., & Bernardino, J. Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: *A Literature Review. Applied Sciences*, 2023, vol. 13, iss. 9, article no. 5521. DOI: 10.3390/app13095521.

27. Moskalenko, V., Korobov, A., & Moskalenko, Y. Object detection with affordable robustness for UAV aerial imagery: model and providing method. *Radioelectronic and Computer Systems*, 2024, no. 3, pp. 55–66. DOI: 10.32620/reks.2024.3.04.

28. Wang, L., & Tien, A. Aerial Image Object Detection with Vision Transformer Detector (ViTDet). *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, 2023. DOI: 10.1109/igarss52108.2023.10282836.

29. Li, Y., Mao, H., Girshick, R., & He, K. Exploring Plain Vision Transformer Backbones for Object Detection. *Lecture Notes in Computer Science, Springer Nature Switzerland*, 2022, pp. 280–296. DOI: 10.1007/978-3-031-20077-9_17.

30. Yuldashev, Y., Mukhiddinov, M., Abdusalomov, A. B., Nasimov, R., & Cho, J. Parking Lot Occupancy Detection with Improved MobileNetV3. *Sensors*, 2023, vol. 23, iss. 17, article no. 7642. DOI: 10.3390/s23177642.

31. Nawaz, S. A., Li, J., Bhatti, U. A., Shoukat, M. U., & Ahmad, R. M. AI-based object detection latest trends in remote sensing, multimedia and agriculture applications. *Frontiers in Plant Science*, 2022, vol. 13. DOI: 10.3389/fpls.2022.1041514.

32. Caron, M., Touvron, H., Misra, I., Jegou, H., Mairal, J., Bojanowski, P., & Joulin, A. Emerging Properties in Self-Supervised Vision Transformers. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, article no. 00951. DOI: 10.1109/iccv48922.2021.00951.

33. Pinetsuksai, N., Kittichai, V., Jomtarak, R., Jaksukam, K., Tongloy, T., Boonsang, S., & Chuwongin, S.

Development of Self-Supervised Learning with Dinov2-Distilled Models for Parasite Classification in Screening. *2023 15th International Conference on Information Technology and Electrical Engineering (ICITEE)*, 2023, pp. 323-328. DOI: 10.1109/icitee59582.2023.10317719.

34. Kulkarni, U., S M M., Hallyal, R., Sulibhavi, P., G S. V., Guggari, S., & Shanbhag, A. R. Optimisation of deep neural network model using Reptile meta learning approach. *Cognitive Computation and Systems, Institution of Engineering and Technology (IET)*, 2023. DOI: 10.1049/ccs2.12096.

35. Kotyan, S., & Vargas, D. V. Adversarial robustness assessment: Why in evaluation both L0 and L∞ attacks are necessary. *PLOS ONE*, 2022, vol. 17, iss. 4, article no. e0265723. DOI: 10.1371/journal.pone.0265723.

36. Li, G., Pattabiraman, K., & DeBardeleben, N. TensorFI: A Configurable Fault Injector for TensorFlow Applications. *2018 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)*, 2018. DOI: 10.1109/issrew.2018.00024.

37. Zhao, Y., Sun, H., & Wang, S. A Method for Detecting Lightweight Optical Remote Sensing Images Using Improved Yolov5n. *Journal of Multimedia Information System*, 2023, vol. 10, iss. 3, pp. 215-226. DOI: 10.33851/jmis.2023.10.3.215.

38. Zou, C., Jeon, W.-S., Rhee, S.-Y., & Cai, M. YOLO-RS: A More Accurate and Faster Object Detection Method for Remote Sensing Images. In Remote Sensing. *Remote Sensing*, 2023, vol. 15, iss. 15, article no. 3863. DOI: 10.3390/rs15153863.

Надійшла до редакції 10.08.2024 розглянута на редколегії 15.10.2024

MODEL AND TRAINING METHOD FOR AERIAL IMAGE OBJECT DETECTOR WITH OPTIMIZATION OF BOTH ROBUSTNESS AND COMPUTATIONAL EFFICIENCY

Alona Moskalenko, Mykola Zaretskyi, Maksym Vynohradov, Vladyslav Babych

The **subject** of research is Neural network-based object detectors, which are widely used for video image analysis. An increasing number of tasks now demand data processing directly at the source, which limits the available computational resources. However, the vulnerability of neural networks to noise, adversarial attacks, and weight error injections significantly diminishes their robustness and overall effectiveness. The relevant **task** is to develop models that provide both computational efficiency and robustness against perturbations. This paper investigates a model and method for enhancing the robustness of neural network detectors under limited resources. The **objective** is to design a model that allocates resources optimally while maintaining stability. To achieve this, the study employs techniques such as dynamic neural networks, robustness optimization, and resilience strategies. The following **results** were obtained. A detector with a feature extractor based on ViT-S/16, modified with gate modules for dynamic examination was developed. The model was trained on the RSOD dataset and meta-learned on the adaptation results to various perturbations. The model's resistance to random bit inversions in weights (10 % of weights) and to adversarial attacks with perturbation amplitudes up to 3/255 (L∞ norm) was tested. **Conclusion**. The proposed detector model incorporating dynamic examination and optimized robustness, reduced floating-point operations by over 20 % without loss of accuracy. A novel method for training the detector was developed, combining the RetinaNet loss function with the loss function of gate blocks and applying meta-learning on the adaptation results for various types of synthetic perturbations. Testing demonstrated an increase in accuracy by 11.9 % under the influence of error injection and by 13.2 % under the influence of adversarial attacks.

Keywords: object detection; robustness; adversarial attacks; fault injections; meta-learning.

Москаленко Альона Сергіївна – канд. техн. наук, доц., доц. каф. комп'ютерних наук, Сумський державний університет, Суми, Україна.

Зарецький Микола Олександрович – PhD, старший викладач каф. комп'ютерних наук, Сумський державний університет, Суми, Україна.

Виноградов Максим Олексійович – асп. каф. комп'ютерних наук, Сумський державний університет, Суми, Україна.

Владислав Бабич Юрійовіч – асп. каф. комп'ютерних наук, Сумський державний університет, Суми, Україна.

Alona Moskalenko – PhD, Associate Professor at the Computer Science Department of Sumy State University, Sumy, Ukraine,
e-mail: a.moskalenko@cs.sumdu.edu.ua, ORCID: 0000-0003-3443-3990, Scopus Author ID: 57148522500.

Mykola Zaretskyi – PhD, Senior Lecturer at the Computer Sciences Department of Sumy State University, Sumy, Ukraine, e-mail: n.zaretskij@gmail.com, ORCID: 0000-0001-9117-5604, Scopus Author ID: 57213687285.

Maksym Vynohradov – PhD Student of the Computer Science Department of Sumy State University, Sumy, Ukraine, e-mail: vinogradov.max97@gmail.com, ORCID: 0009-0009-1234-922X.

Vladyslav Babych – PhD Student of the Computer Science Department of Sumy State University, Sumy, Ukraine, e-mail: vladuslav09@gmail.com, ORCID: 0009-0009-2521-0841.