

УДК 623.773.4, 004.85

DOI: <https://doi.org/10.33577/2312-4458.30.2024.14-23>С.А. Цибуля¹, В.Ю. Тимчук², А.М. Волокита³¹Національний університет оборони України, Київ²Національна академія сухопутних військ імені гетьмана Петра Сагайдачного, Львів³Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ

Article history: Received 12 February 2024; Revised 18 February 2024; Accepted 14 March 2024

ОЦІНЮВАННЯ СКЛАДНОСТІ НАБОРУ ВХІДНИХ ДАНИХ ДЛЯ СИСТЕМИ КОМП'ЮТЕРНОГО ЗОРУ

Стаття розглядає актуальну проблематику використання нейронних мереж у розробці систем комп'ютерного зору військового призначення, з акцентом на аналіз наборів вхідних даних (датасетів). Метою статті є розроблення способу оцінювання складності наборів даних, що базується на розмірах і якості вхідних зображень, та рівні й структурі їх анотації. Застосування попереднього оцінювання складності датасетів дозволить аналізувати вплив зазначених властивостей вхідних даних на тривалість тренування, розхід апаратних ресурсів та загальну ефективність моделей комп'ютерного зору, призначених для виявлення та ідентифікації замаскованих об'єктів.

Для досягнення поставленої мети у дослідженні були застосовані методики аналізу внутрішньої структури та властивостей датасетів, включно з аналізом якості зображень, розмірів, різноманітності об'єктів та складності анотацій. Отримані результати демонструють прямий зв'язок між складністю датасетів і ефективністю тренування моделей, виявляючи ключові параметри, що впливають на точність та швидкість навчання.

Особливу увагу у дослідженні приділено розробці універсального способу оцінювання складності датасетів, що дозволяє об'єднати різноманітні параметри в єдиний кількісний показник. Цей підхід не тільки сприяє глибшому аналізу впливу характеристик датасетів на процес навчання, але й дозволяє прогнозувати витрати обчислювальних ресурсів. Теоретичне значення роботи полягає у формуванні цілісного уявлення про фактори, що впливають на ефективність машинного навчання у сфері комп'ютерного зору для військового застосування. Практична значущість розкривається через можливість використання результатів для оптимізації розробки та тренування нейронних мереж, націлених на підвищення ефективності військових систем розпізнавання.

Результати дослідження надають внесок у розвиток теорії та практики створення систем комп'ютерного зору, пропонуючи інструментарій для ефективнішого вибору або створення датасетів з урахуванням вимог військово-технічних завдань. У подальших дослідженнях планується розширити використання запропонованого показника на різноманітних типах датасетів та архітектурах нейронних мереж, що має сприяти універсальності та масштабованості запропонованого підходу для оцінювання складності датасетів.

Ключові слова: машинне навчання, комп'ютерний зір, засіб маскування, приховування, зображення місцевості, індекс складності, датасет, набір даних, якість зображень, вхідні дані, визначення об'єктів, класифікація об'єктів, метаінформація, анотування, анотація.

Постановка проблеми

Одним зі способів оцінювання ефективності засобів маскування є проведення їх випробувань на місцевості. Для автоматизації цього процесу та виключення суб'єктивного фактора при роботі експертів-спостерігачів доцільно використовувати системи комп'ютерного зору, які здатні автоматично виявляти та ідентифікувати об'єкти, що замасковані та приховані на місцевості. На даний час найбільш ефективні

системи комп'ютерного зору створені на основі згорткових нейронних мереж, що належать до класу моделей машинного навчання з вчителем, але за останній час набирають популярності та показують гарні результати системи на основі архітектури Visual Transformers [1].

Одним із важливих аспектів при впровадженні систем комп'ютерного зору є забезпечення достатнього обсягу вихідних даних (так званий датасет, від англ. dataset) для їх навчання та тестування [2].

Сучасні алгоритми машинного навчання вимагають для адекватного навчання датасети, що містять десятки, а то й сотні тисяч зображень. Збір таких великих масивів даних вимагає певних технічних та людських ресурсів і пов'язаний з великими часовими та фінансовими витратами.

Крім обсягу даних критично важливою є наявність детальної та коректної метайнформації – розмітки (анотації, від англ. annotation) зображень за класами об'єктів. Процес анотування як ручний, так і напівавтоматичний, може бути ресурсомістким та схильним до помилок через людський фактор і обмеження наявних інструментів анотації. Окрім того, одним із проблемних питань створення високо-ефективних моделей комп'ютерного зору є їх залежність, значною мірою, від якості вхідних даних, що є важливою часткою складності датасету.

Під складністю набору вхідних даних для визначення об'єктів на зображеннях ми розуміємо емерджентні властивості датасету, мультифакторні характеристики даних, якісні та кількісні параметри зображень, різноманітність об'єктів і їх контекстуальну взаємодію, та структурні взаємозв'язки елементів анотації, що разом впливають на процес і ефективність навчання моделей машинного навчання та обчислювальні вимоги.

Складність датасету безпосередньо впливає на здатність моделі генералізувати навчання, точність розпізнавання об'єктів та швидкість тренування і перевірки. Водночас відсутність уніфікованих критеріїв або індексів для оцінювання складності датасетів ускладнює планування та розподіл ресурсів у проєктах розпізнавання об'єктів.

Зі збільшенням обсягів даних і складності їх наборів виникає потреба в інструментах для їх оцінювання та оптимізації використання ресурсів на етапі підготовки даних. Відсутність чітких підходів для оцінювання складності датасетів може призвести до нерационального використання часу та ресурсів, затрат на обробку даних та зниження ефективності розроблених моделей машинного навчання. Розроблення точних і надійних моделей вимагає глибокого розуміння впливу характеристик датасету на процес навчання.

Актуальність цієї проблеми посилюється зі стрімким розвитком технологій машинного навчання та комп'ютерного зору, які знаходять застосування в широкому спектрі військової справи, від автономних транспортних засобів, медичного діагностування поранених до автоматизованих систем виявлення та ураження противника.

Отже, створення способу оцінювання складності датасетів відкриває можливість для значного покращення процесу планування та реалізації проєктів

в області комп'ютерного зору, забезпечуючи оптимальне використання ресурсів та підвищення точності моделей. Отримані результати можуть сприяти підвищенню ефективності оцінювання, шляхом автоматизації їх тестування в ході розроблення, засобів маскування військових об'єктів у різних географічних умовах і, як наслідок, видкому створенню ефективних засобів маскування.

Аналіз останніх досліджень і публікацій

Питанням оцінювання наборів даних для роботи систем комп'ютерного зору знайшла своє відображення у наукових роботах різних дослідників. Важливість високоякісних розмічених даних підкреслюється у роботі [3]. Відсутність яких є ключовою перешкодою для створення, одного із потенційно найважливішого застосування штучного інтелекту у військовій сфері – систем автоматизованого розпізнавання та ураження цілей.

Стаття [4] надає огляд метрик складності задач класифікації та алгоритмів їх розрахунку, які аналізують перекриття між класами об'єктів, відношення між ознаками й класами, лінійну сепарабельність та інші властивості, що впливають на здатність моделей ефективно класифікувати дані.

Основна увага авторів статті [5] приділена аналізу ефективності детектора об'єктів на основі нейронної мережі Faster R-CNN при використанні стислих вхідних даних з відеокамер автоматизованої водійської системи. Для оцінки якості вхідних даних були використані традиційні метрики об'єктивної оцінки якості зображень (IQA англ. image quality assessment), такі як повна референсна IQA (FR-IQA), часткова референсна IQA (RR-IQA), безреференсна IQA (Blind-IQA), а також сучасні методи глибокого аналізу зображень, зокрема інформаційно-ваговий індекс структурної схожості (IW-SSIM) і адаптований алгоритм BRISQUE для безреференсної оцінки. Дослідження виявило вплив якості зображення на складність вхідних даних і точність моделей виявлення об'єктів. Результати досліджень виявили кореляцію між рівнями стиснення зображень і точністю детектора об'єктів.

У статті «Spectral metric for dataset complexity assessment» автори вказують на необхідності розробки нових підходів щодо оцінювання складності датасетів через недосконалість наявних метрик, що не враховують великі обсяги даних, які використовуються для тренування глибоких нейронних мереж [6]. Пропонується кумулятивний спектральний градієнт як метод оцінювання датасетів за допомогою спектральної кластеризації вхідних даних.

Аналіз наукових робіт демонструє, що попри прогрес в цій галузі існує актуальна потреба в розробці більш ефективних способів оцінювання

датасетів, що використовуються в системах комп'ютерного зору військового призначення і працюють з даними низької якості.

Формулювання мети статті

Метою статті є розроблення способу оцінювання складності наборів даних, що базується на розмірах і якості вхідних зображень, та рівні й структурі їх анотації.

Виклад основного матеріалу дослідження

Існує багато факторів, які визначають складність датасету. В першу чергу вона залежить від якості зображень, що присутні у датасеті. Низька якість зображень може значно ускладнити ідентифікацію об'єктів як на етапі анотування, так і під час навчання моделі. На якість зображення впливають: розмитість, шум, погане освітлення, низька роздільна здатність та інші фактори.

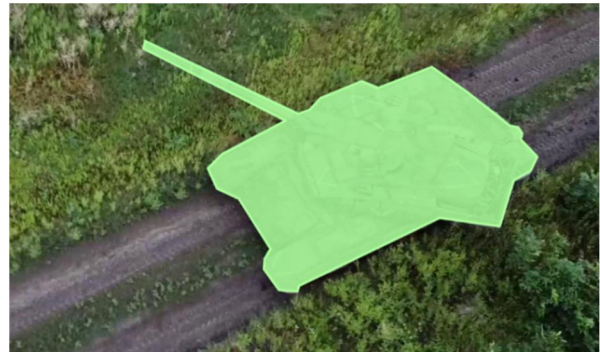
Велика різноманітність об'єктів на зображенні, включаючи різні форми, розміри, кольори та текстури, вказують на більшу складність датасету. Цей аспект охоплює деталізацію об'єктів, наявність та розпізнаваність характерних ознак. Датасети з високим рівнем деталізації об'єктів зазвичай є менш складними для розпізнавання. Вплив на складність також має фон та контекст зображення, в якому розташовані об'єкти. Датасети з об'єктами на складному фоні або в різноманітних контекстах можуть вимагати більш глибокого навчання моделі для ефективного розпізнавання. На складність впливає різноманітність умов фотографування, включаючи різне освітлення, кути зору та відстань до об'єктів, що призводить до варіативності позиціонування об'єктів.

Основними завданнями комп'ютерного зору є визначення та класифікації об'єктів на зображеннях (рис. 1) не залежно від їх якості. Це стосується як роботи із фотографіями, так і з відео, кадри якого теж по своїй суті є статичним зображенням. Ці завдання служать різній меті та мають ключові відмінності. Так, метою класифікації є визначення приналежності цілого зображення до певного класу чи категорії на основі його вмісту.



Рис. 1. Зображення танка Т-90 (з мережі Internet)

Завдання визначення об'єктів полягає в локалізації та в подальшій класифікації одного або кількох об'єктів на зображенні. Це включає не тільки визначення того, що зображено на зображенні, але й розташування об'єкта за допомогою обмежувальної рамки. Цей процес є більш складним і вимагає аналізу деталей зображення та розуміння його контексту. Широко застосовуються два основних підходи виділення об'єктів на зображеннях: семантична сегментація об'єктів та визначення об'єктів за допомогою прямокутних обмежувальних рамок (баундінг-боксів, від англ. bounding box) (рис. 2).



а



б



в

Рис. 2. Сегментація за допомогою:

а) маски; б) полігонів; в) рамки

Контурна сегментація об'єктів дозволяє точно окреслити форму об'єкта, відділяючи його силует від фону за допомогою полігонів. Це є ключовим для задач, де необхідна детальна інформація про геометрію та межі об'єкта. З іншого боку, виявлення об'єктів з

використанням рамок застосовується для визначення місцезнаходження та приблизних розмірів об'єкта на зображенні. Цей метод є особливо корисним у задачах, де потрібно швидко ідентифікувати та класифікувати різноманітні об'єкти без необхідності в детальному аналізі їхньої форми. Зазвичай баундінг-бокси – це прямокутні рамки, але в деяких завданнях можуть бути у вигляді паралелограмів. Вони поділяються на ті, що завжди вирівнюються відповідно до основних осей координатної системи (AABB від англ. axis-aligned bounding boxes), та рамки, що обертаються відповідно до орієнтації об'єкта на зображенні (OBB від англ. oriented bounding boxes). AABB можуть бути неефективними для об'єктів неправильної форми або об'єктів, які мають певну орієнтацію, оскільки вони можуть займати більше простору, ніж необхідно, але вони є простими в обчисленні та широко використовуються для швидкої перевірки перетину між об'єктами.

Відповідно до класифікації та визначення об'єктів потребують різні підходи до створення датасетів для навчання нейронних мереж. Так, для завдання класифікації достатньо відсортувати масив вхідних зображень по класах, що найкраще описують зміст цілого зображення, і на основі характеристик зображень визначити структуру нейронної мережі. Більш складні дії необхідно здійснити для завдання визначення об'єктів: виконати анотування зображень датасету, розмітити їх обмежувальними рамками для кожного об'єкта, що знаходиться на зображенні, та визначити класи об'єктів.

Об'єкти з маскувальним покриттям зазвичай зливаються з фоном, що ускладнює їх виявлення [7]. Окрім високої внутрішньої подібності між замаскованими об'єктами та їхнім фоном ці об'єкти зазвичай різноманітні за масштабом та нечіткі на вигляд. Тому в проєктах роботи зі зображеннями із замаскованими об'єктами необхідно виконувати їх сегментацію для надання чіткого розуміння моделі, де знаходиться, і що з себе являє цей об'єкт.

При анотуванні замаскованого об'єкта ОВТ на зображенні об'єкти поділяються на наступні класи:

замаскований об'єкт – невідомий об'єкт ОВТ знаходиться повністю під маскувальним покриттям;

частково замаскований – об'єкт не повністю вкритий маскувальним покриттям, тип ОВТ не відомий;

незамаскований – об'єкт повністю або частково вкритий маскувальним покриттям, але яке не заважає визначити конкретний тип ОВТ.

У випадках часткового знаходження ОВТ під маскувальним покриттям виконується анотування шляхом визначення двох об'єктів на зображенні: засобу маскування та об'єкта ОВТ.

Анотування даних впливає на складність датасету. Точність і консистентність анотацій в датасеті мають критичне значення на подальше навчання моделей.

Збалансованість у розподілі класів, різноманітність між класами та однорідність всередині класів допомагають знизити складність датасету та покращити навчання моделі. Наявність перекриття об'єктів на зображенні ускладнює їх анотування, а під час роботи моделі їх ідентифікацію та класифікацію. Якісна та коректна анотація даних відіграє вирішальну роль у розробці та точності моделей машинного навчання. Однак її процес створення пов'язаний з певними складностями [8]:

Вартість анотування даних. Анотування даних виконується вручну або автоматично. Ручний процес вимагає значних зусиль, часу та людських ресурсів, що призводить до збільшення фінансових витрат.

Точність анотування. Людські помилки під час процесу анотування можуть призвести до низької якості даних, що безпосередньо впливає на продуктивність і прогнози моделей. Вони залежать від досвіду та старанності особи, яка працює з даними.

Масштабованість. Зі збільшенням обсягу даних процес анотування стає більш складним і трудомістким.

Конфіденційність даних та безпека. Не всі типи даних можливо віддавати на оутсорсінг стороннім особам для анотування. Дані військового призначення, а особливо результати їх обробки, можуть мати певні ступені обмеження доступу.

Керування даними. Робота з різними типами даних, як-от текст, зображення, аудіо та відео, вимагає різноманітні техніки анотування і досвіду виконавців. Координація та керування процесом анотування для різних типів даних може бути складним і ресурсомістким.

На ринку представлено багато комерційного та відкритого програмного забезпечення для анотування зображень. Вони підтримують різні структури для зберігання анотованих даних, які зазвичай знаходяться в окремих файлах і містять інформацію про зображення, таку як ім'я, розмір, шлях, а також інформацію про об'єкти анотації: координати рамок, класи, ідентифікатори тощо. Найпопулярніші формати це: XML, який зберігає структуровані дані у вигляді тегів та атрибутів; JSON, що зберігає дані у вигляді пар ключ-значення; CSV – дані знаходяться у вигляді рядків і стовпців, розділених комами або іншими символами. Також існують інші формати даних на основі XML та JSON, які можуть містити додаткову інформацію, а також самі зображення (Pascal VOC XML, CONLL2003, Brush labels to NumPy тощо). Підготовка персоналу до роботи з цим програмним забезпеченням вимагає проведення з ним навчальних курсів та інших додаткових фінансових витрат.

Існує кілька підходів до оцінювання складності датасетів, кожен з яких має свої особливості та область застосування. Один з найпростіших підходів

полягає у використанні статистичних мір для аналізу анотацій датасету. Це включає аналіз розміру та площі обмежувальних рамок, визначення їх розподілу, кількості масок на зображення та взаємного перекриття. Ці показники допомагають зрозуміти, наскільки складним є процес анотування та сегментування об'єктів у датасеті. Детальний аналіз розподілу даних в анотації до датасету може виявити важливі особливості, такі як наявність викидів, перекося у розподілі класів або складність меж класів. Для аналізу складності датасету можна також використовувати метрики, які дозволяють оцінити різноманітність форм, розмірів, текстур та інших характеристик об'єктів як в межах одного класу, так і між різними класами. Висока різноманітність зазвичай вказує на більшу складність датасету та більш трудомісткий процес навчання моделі, яка має навчитися розрізняти більш тонкі відмінності між об'єктами.

Оцінювання, засноване на моделі машинного навчання, полягає у використанні попередньо тренуваних моделей для визначення складності датасету. Такий аналіз може включати тренування моделі на датасеті та аналіз її поведінки, включаючи точність на валідаційному наборі, швидкість збіжності під час тренування та чутливість до змін у гіперпараметрах.

Одним із перспективних способів є використання спеціалізованих індексів складності, які об'єднують кілька мір в одне числове значення. Індеси можуть враховувати різноманітні аспекти, такі як якість зображення, структуру анотації, різноманітність класів, баланс між класами, розміри об'єктів, ступінь перекриття між ним тощо. Індекс допомагає оцінити, наскільки складно буде моделі навчитися з даного набору даних та які потенційні виклики можуть виникнути під час тренування.

Створення універсального виразу для визначення індексу складності датасету для моделі визначення об'єктів на зображеннях є складним завданням через велику кількість змінних факторів, які потрібно враховувати. Автори пропонують інтегральний показник, що використовує ряд параметрів, кожен з яких враховує певні аспекти складності даних:

- кількість вхідних зображень (розмір датасету);
- розміри вхідних зображень;
- якість зображень;
- структуру анотації об'єктів на зображеннях.

Визначення інтегрального показника складності датасету виконується за виразом

$$C = k_N \times k_S \times k_Q \times k_A, \quad (1)$$

де C – показник складності датасету;

k_N – коефіцієнт розміру датасету;

k_S – коефіцієнт розміру вхідних зображень;

k_Q – коефіцієнт якості зображень;

k_A – коефіцієнт структурної складності анотації.

Коефіцієнт розміру датасету k_N , залежить від кількості вхідних зображень (розмір датасету). Для нормалізації значень кількості зображень датасетів різних розмірів і перетворення їх у коефіцієнт, який можна використовувати у формулі обрахунку складності, одним із підходів може бути використання логістичної функції

$$k_N = \frac{L}{1 + e^{-k(N - N_0)}}, \quad (2)$$

де N – кількості вхідних зображень, шт.;

L – максимальне значення коефіцієнта k_N , яке функція може прийняти. Для діапазону $[0;1]$ дорівнює 1;

k – коефіцієнт нахилу;

N_0 – значення N , при якому $k_N(N) = L/2$.

Значення k визначаються емпірично залежно від датасету, що оцінюється. Приклад розрахунку значення k_N , при $k=0,001$, наведено у таблиці 1.

Таблиця 1

Значення коефіцієнта розміру датасету

Кількість зображень, шт.	1000	1250	4100	10000
k_N	0,62	0,65	0,89	0,99

Коефіцієнт розміру вхідних зображень k_S визначається як логарифмічне перетворення добутку розмірів зображення по осях X та Y відповідно

$$k_S = \sum_{i=1}^N \log_{10}(W_i \times H_i) / N, \quad (3)$$

де W_i, H_i – ширина та висота зображення, px (pixel's).

Логарифмічне перетворення дозволяє зменшити розкид значень і стиснути ці значення, особливо коли вони мають широкий діапазон, роблячи їх більш порівнянними (табл. 2).

Таблиця 2

Значення коефіцієнта розміру вхідних зображень

Розмір зображень, $W_i \times H_i$, px.	240x240	704x512	854x480	1368x898
k_S	4,76	5,56	5,61	6,09

Для оцінки якості зображення необхідно використовувати алгоритми, які не використовують еталонні зображення для порівняння (NR-IQA, від англ. No-reference image quality assessment). Популярні алгоритми в цьому класі: BIQI (від англ. Blind image quality index), BLINDS (від англ. Blind image integrity notator using DCT statistics), DIIVINE (від англ. Distortion identification-based image verity and integrity evaluation), PIQE (від англ. Perceptual image quality evaluator) та BRISQUE (від англ. Blind/referenceless image spatial quality evaluator), кожен з яких має свої унікальні підходи та характеристики [9]. Алгоритм BRISQUE базується на аналізі статистичних властивостей

зображення, таких як яскравість, контрастність, рівень шуму тощо. Також враховуються умови знімання, такі як освітлення, рівень розмиття, низька різкість, що може свідчити про проблеми з фокусуванням або рухом камери. BRISQUE відрізняється високою точністю оцінки якості та простотою реалізації порівняно з іншими алгоритмами. Низькі значення зазвичай вказують на високу якість зображення, тоді як вищі значення свідчать про нижчу якість.

Для перевірки працездатності алгоритму BRISQUE на оригінальне зображення танка (рис. 1) було додано різні типи шуму (рис. 3), що можуть виникати в процесі використання БПЛА через помилки в процесі передавання даних, несправності датчиків камери, або інших технічних недоліків чи роботи засобів РЕБ: «сіль та перець» (salt-and-pepper noise), гауссівський шум та перешкоди в аналоговий телевізійний сигнал.



а



б



в

Рис. 3. Зображення з різним типом шуму

Шум типу «сіль та перець» – це вид перешкод, який часто зустрічається на цифрових зображеннях. Цей шум характеризується раптовими перепадами яскравості, коли окремі пікселі на зображенні отримують максимальне (зазвичай біле, «сіль») або мінімальне (зазвичай чорне, «перець») можливе значення. Візуально це створює ефект випадково розкиданих світлих і темних плям на зображенні. Для симулювання ефекту поганого телевізійного сигналу використовується комбінація шумів, різних ефектів розмиття, а також вертикальні та горизонтальні смуги (рис. 3в). Результати розрахунків, що виконані згідно з роботою [10], наведено у таблиці 3.

Відповідно значення коефіцієнта якості зображень датасету визначається за виразом

$$k_Q = \frac{1}{N} \sum_{i=1}^N (\text{BRISQUE}_i). \quad (4)$$

Таблиця 3

Значення параметрів якості зображення

Тип шуму	Оригінальне зображення танка (рис. 1)	Salt-and-pepper 0,05%	Gaussian	TV noise
BRISQUE	33,74	85,17	155,85	46,18

Визначення структурної складності анотації виконаємо за допомогою мультиплікативного об'єднання всіх параметрів анотації:

$$k_A = \sqrt[8]{D_i \times K_i \times (Q_i + 1) \times nr \times a \times ar \times p \times par}, \quad (5)$$

де D – середня кількість об'єктів на зображеннях у датасеті, шт.;

K – середня кількість різних класів об'єктів на зображеннях, шт.;

O – середній ступінь перекриття об'єктів (оцінка від 0 до 1, де 1 вказує на повне перекриття). Оскільки він може дорівнювати 0, то додаємо 1;

nr – середня кількість точок, які формують полігон навколо анотованих об'єктів датасету. Вказує на складність форми об'єкта, де більша кількість точок зазвичай означає детальнішу або складнішу форму, шт.;

a – середня площа анотованих об'єктів. Вказує на розмір об'єкта в межах зображення, px^2 ;

ar – середнє співвідношення площі анотованих об'єктів до загальної площі зображень. Цей параметр допомагає оцінити, яку частину зображення займає анотований об'єкт;

p – середній периметр полігонів, що обмежують анотовані об'єкти. Довжина замкненої лінії контуру об'єкта, px .;

par – середнє відношення периметра полігонів до їх площі. Цей показник вказує на складність форми об'єкта, високі значення можуть свідчити про більш складну або неправильну форму.

$$D = \frac{1}{N} \sum_{i=1}^N (d_i), \quad (6)$$

де d – кількість об'єктів на зображенні (приймаємо завжди >0), шт. Аналогічно визначаються середні значення інших параметрів анотації.

У відкритому доступі є набори даних для виявлення замаскованих об'єктів: CAMO [7], CAMO++ [11], COD10K [12], UW-RS [13], NC4K [14] тощо. Але більшість з них створені для виявлення біологічних об'єктів флори та фауни, а датасети з військовими об'єктами зазвичай у відкритий доступ не викладаються [15, 16], за рідкісним виключенням, як COD1K [17]. Попри це їх можливо використати для навчання та тестування систем військового призначення [18].

Результати аналізу анотації об'єктів наведено в таблиці 4. Також у таблиці наведено значення параметрів для датасету COD1K, який має набір анотованих зображень військовослужбовців у індивідуальному маскувальному одязі (рис. 4).

Таблиця 4

Значення параметрів структурної складності анотації

Параметр	Рис. 2а	Рис. 4	CAMO	COD1K
N	1	1	1250	1000
D	1	1	1,01	1
K	1	1	1,01	1
O	0	0	0,01	0
np	32	280	596,11	186
a	20394,4	20525	76719,8	7952,84
ar	0,049	0,05	0,16	0,02
p	648,02	768,7	1456,78	464,68
par	0,032	0,04	0306	0,13
k_A	5,34	7,38	12,62	6,02



Рис. 4. Зображення військовослужбовця з датасету COD1K та його сегментація [17]

Розрахунок інтегрального показника складності для датасетів CAMO, NC4K, COD1K наведено у табл. 5.

Складові параметри інтегрального показника використовуються для оцінки деталізації, складності та відносного розміру анотованих об'єктів у датасеті, що може бути корисним для розуміння якості анотацій та потенційних викликів при тренуванні моделей машинного навчання.

Кожний з параметрів впливає на апаратні ресурси, що використовуються при навчанні моделей. Для обліку яких використовуються обчислювальні блоки, які використовуються для виконання обчислень у комп'ютерних системах, таких як центральні процесори (CPU), графічні процесори (GPU), спеціалізовані тензорні процесори (TPU) тощо. Також вони враховують використання оперативної пам'яті EOM та пам'яті GPU.

Таблиця 5

Значення інтегрального показника складності

Параметр	Значення параметрів для датасету		
	CAMO	NC4K	COD1K
k_N	0,65	0,89	0,62
k_S	5,74	5,59	5,61
k_Q	24,82	22,5	21,24
k_A	12,62	12,5	6,02
C	1168,66	1399,25	444,74

Навчання датасетів CAMO, COD1K, COD10K, UW-RS, NC4K та результати моделювання у середовищі Google Colaboratory [19], дозволили визначити зв'язок між параметрами інтегрального показника складності датасету та витратою обчислювальних блоків (рис. 5):

1. Коефіцієнт розміру датасету показує, що вплив розміру датасету може становити від 20% до 40%. Великі датасети значно покращують здатність моделі узагальнювати на нових даних, але також збільшують вимоги до обчислювальних ресурсів і часу на тренування.

2. Коефіцієнт розміру вхідних зображень показує, що вплив розміру вхідних зображень може бути в межах від 20% до 30%. Більші зображення збільшують обчислювальні вимоги, особливо до пам'яті, але можуть покращити якість розпізнавання завдяки більшій кількості деталей.

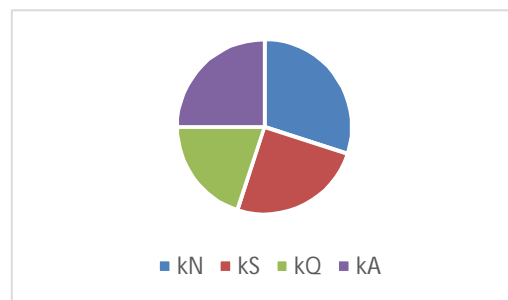


Рис. 5. Вплив параметрів інтегрального показника складності датасету на витрату часу та обчислювальних блоків

2. Коефіцієнт розміру вхідних зображень показує, що вплив розміру вхідних зображень може бути в

межах від 20% до 30%. Більші зображення збільшують обчислювальні вимоги, особливо до пам'яті, але можуть покращити якість розпізнавання завдяки більшій кількості деталей.

3. Коефіцієнт якості зображень показує, що вплив якості зображень може коливатися від 10% до 25%. Якість зображень має важливе значення для точності моделей, особливо в задачах, де важливі тонкі деталі, такі як текстури засобів маскування.

4. Коефіцієнт структурної складності анотації показує, що вплив структури анотації може становити від 15% до 35%. Складність анотацій визначає, наскільки складною буде задача для моделі, включаючи потребу в більш складних архітектурах для їх обробки.

Ці оцінки є дуже узагальненими та можуть суттєво змінюватися залежно від конкретних датасетів. Важливо зауважити, що ефективність моделі та вимоги до обчислювальних ресурсів залежать не лише від датасету, але й від вибраної архітектури моделі, оптимізації під час тренування та способів попередньої обробки даних.

В ході моделювання було підтверджене припущення про кореляцію між значенням інтегрального показника та часом навчання і витратою обчислювальних блоків – чим більше збільшення показника, тим більше час і розхід ресурсів (рис. 6).

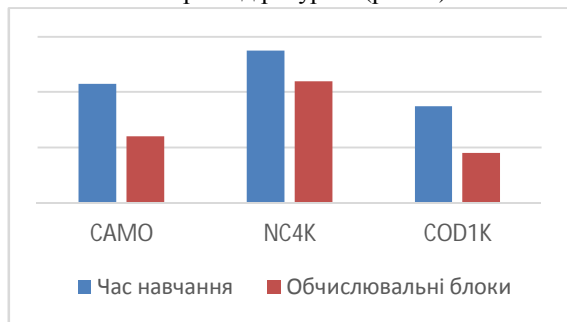


Рис. 6. Зв'язок між інтегральним показником складності датасету та витратою часу та обчислювальних блоків

Важливим напрямом подальших досліджень є проведення тестування індексу складності на різних датасетах, що дозволить перевірити його надійність, універсальність і корисність. Дослідження показують необхідність подальших наукових робіт у розвитку метрик для оцінювання складності датасетів. У контексті створення ефективних засобів маскування, у ході дослідження визначена необхідність у зборі високоякісних розмічених наборів даних. Це спонукає до розроблення нових підходів щодо автоматизації процесу анотації даних, які мінімізують витрати часу та зусиль при підготовці високоякісних датасетів.

Висновки

У статті авторами досліджено вплив різних параметрів датасетів на їх складність, від якої залежить ефективність тренування моделей машинного навчання.

З урахуванням положень теорії складності обчислень можна виділити такі основні складові складності датасету [20]:

1. Статистична складність. Яка залежить від розміру датасету (кількості зображень та анотацій), кількості та розподілу різних класів об'єктів на зображеннях, їх репрезентативності та різноманітності.

2. Структурна складність. Залежить від організації даних: тип та формат файлів зображень, формат та структура зберігання анотацій.

3. Логічна складність. Залежить від складності виявлення та отримання інформації з анотацій та зображень, складності розроблення алгоритмів для оброблення та аналізу зображень та анотацій.

4. Інформаційна складність. Враховує характеристики зображень: розміри (ширина та висота), кількість каналів кольорів, наявність артефактів, спотворень або зашумленість; рівень освітлення та контрастність.

Дослідження підтвердило кореляцію між інтегральним показником складності датасету та часом навчання та ресурсами ЕОМ: збільшення показника веде до збільшення ресурсних витрат. Зокрема, було виявлено, що існує тісний зв'язок між розмірами датасетів, розмірами вхідних зображень, їх якістю, структурою анотацій та витратами обчислювальних блоків і часу на тренування. Ці параметри мають бути враховані для ефективного планування фінансування та розподілу ресурсів при проектуванні систем комп'ютерного зору. Відсутність уніфікованих підходів для оцінювання складності датасетів може призвести до нераціонального використання часу та зниження ефективності моделей, що підкреслює актуальність проведених досліджень.

Теоретична значущість роботи полягає у визначенні ключових факторів, які впливають на складність датасетів та їх вплив на процес тренування моделей глибокого навчання, що використовуються в системах комп'ютерного зору військового призначення. Практичне значення роботи полягає у можливості застосування індексу складності датасету для оптимізації використання обчислювальних та людських ресурсів під час розробки систем розпізнавання об'єктів, запобігаючи надмірному витрачання на обробку занадто складних або недостатньо інформативних датасетів. Оцінка та оптимізація складності датасетів на основі індексу може сприяти підвищенню точності та ефективності моделей машинного навчання, оскільки дозволяє більш цілеспрямовано підходити до вибору та підготовки тренувальних даних, спрощуючи планування та адаптацію процесу тренування. Індекс спонукає до більш ретельного підходу до створення та анотування датасетів, підвищуючи їх загальну якість. Розроблення індексу складності датасету, таким чином, робить значний внесок у науку та практику розробки систем комп'ютерного зору, надаючи розробникам дієвий

інструмент для підвищення ефективності та якості своїх проєктів.

Загалом результати дослідження демонструють важливість комплексного підходу до оцінювання та підготовки датасетів для використання в задачах комп'ютерного зору, що має значне теоретичне та практичне значення для розвитку цієї галузі науки.

Список літератури

1. Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020. DOI: <https://doi.org/10.48550/arXiv.2010.11929>.
2. Санжаров В.В., Фролов В.А., Волобой А.Г., Галактионов В.А., Павлов Д.С. Система генерації наборів зображень для задач комп'ютерного зору на основі фотореалістичного рендеринга // Препринти ІПМ ім. М.В. Келдиша. 2020. № 80. С. 29. DOI: <http://doi.org/10.20948/prepr-2020-80>.
3. Hartnett G. S., Zhang L. A., O'Connell C., Lohn A. J., Aguirre J. Empirical evaluation of physical adversarial patch attacks against overhead object detection models. 2022. DOI: <https://doi.org/10.48550/arXiv.2206.12725>.
4. Lorena A.C., Garcia L.P., Lehmann J., Souto M.C., Ho T.K. How complex is your classification problem? A survey on measuring classification complexity // ACM Computing Surveys (CSUR). 2019. Vol. 52, Issue 5. Pp. 1-34. DOI: <https://doi.org/10.1145/3347711>.
5. Gummadi D., Chan P. H., Wang H. et al. Correlating traditional image quality metrics and DNN-based object detection: a case study with compressed camera data. TechRxiv, 22 November 2023. DOI: <http://dx.doi.org/10.36227/techrxiv.24566371.v1>.
6. Branchaud-Charron F., Achkar A., Jodoin P.-M. Spectral Metric for Dataset Complexity Assessment // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA, 2019. Pp. 3210-3219. DOI: <https://doi.org/10.1109/CVPR.2019.00333>.
7. Le T.-N., Nguyen T. V., Nie Z., Tran M.-T., Sugimoto A. Anabranch network for camouflaged object segmentation // Computer Vision and Image Understanding. 2019. Vol. 184. Pp. 45-56. DOI: <https://doi.org/10.1016/j.cviu.2019.04.006>.
8. The A to Z of Data Annotation. URL: <https://uk.shaip.com/blog/the-a-to-z-of-data-annotation/> (дата звернення 21.02.2024).
9. Ошаровська О.В. Оцінювання показників якості ТВ зображень. ОНАЗ ім. О.С. Попова. Цифрові Технології. 2016. № 19. С. 91-102.
10. Ricardo Ocampo. Automatic image quality assessment in python. URL: <https://towardsdatascience.com/automatic-image-quality-assessment-in-python-391a6be52c11> (дата звернення 29.02.2024).
11. Le T.-N., Cao Y., Nguyen T.-C. et al. Camouflaged instance segmentation in-the-wild: Dataset, method, and benchmark suite // IEEE Transactions on Image Processing. 2021. Vol. 31. Pp. 287-300. DOI: <https://doi.org/10.1109/TIP.2021.3130490>.
12. Fan D.-P., Ji G.-P., Cheng M.-M., Shao L. Concealed Object Detection // IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 44, No. 10. 1 Oct. 2022. Pp 6024-6042. DOI: <https://doi.org/10.1109/TPAMI.2021.3085766>.
13. Dong F., Zhu J. Underwater Camouflage Object Detection Dataset. arXiv preprint arXiv:2303.12946, 2023. DOI: <https://doi.org/10.48550/arXiv.2303.12946>.
14. Lv Y., Zhang J., Dai Y., Li A., Liu B., Barnes N., Fan D.P. Simultaneously localize, segment and rank the camouflaged

objects // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021. Pp. 11591-11601. DOI: <https://doi.org/10.48550/arXiv.2103.04011>.

15. Yi X., Wu J., Ma B., Ou Y., Liu L. LGA-RCNN: Loss-Guided Attention for Object Detection. arXiv preprint. 2021. DOI: <https://doi.org/10.48550/arXiv.2104.13763>.

16. Liu M. A military reconnaissance network for small-scale open-scene camouflaged people detection // Expert Systems. Vol. 40, Issue 10. e13444, 2023. DOI: <https://doi.org/10.1111/exsy.13444>.

17. Zheng Y., Zhang X., Wang F., Cao T., Sun M., Wang X. Detection of people with camouflage pattern via dense deconvolution network // IEEE Signal Processing Letters. Vol. 26, No. 1. Jan. 2019. Pp. 29-33. DOI: <https://doi.org/10.1109/LSP.2018.2825959>.

18. Тимчук В., Литвин В., Перегуда О. Декомпозиція машини глибинного навчання на основі наборів спеціалізованих датасетів для зменшення часу обробки просторової інформації. Військово-технічний збірник. 2023. № 28. С. 60–68. DOI: <https://doi.org/10.33577/2312-4458.28.2023.60-68>.

19. Google Colab. URL: <https://colab.research.google.com/> (дата звернення 29.02.2024).

20. Махровська Н. А., Погромська Г. С. Алгоритми і структури даних: навчально-методичний посібник. Миколаїв: Миколаївський національний університет імені В.О. Сухомлинського, 2019. 279 с.

References

1. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X. et al. (2020), "An image is worth 16x16 words: Transformers for image recognition at scale". *arXiv preprint arXiv:2010.11929*. DOI: <https://doi.org/10.48550/arXiv.2010.11929>.
2. Sanzharov, V.V., Frolov, V.A., Voloboy, A.G., Galaktionov, V.A. and Pavlov, D.S. (2020), "Sistema generatsii naborov izobrazheniy dlya zadach kompyuternogo zreniya na osnove fotorealisticnogo renderinga" [A system for generating sets of images for computer vision tasks based on photorealistic rendering]: *Preprinty IPM im. M. V. Keldysha*, (80), p. 29. DOI: <http://doi.org/10.20948/prepr-2020-80>. [in Russian].
3. Hartnett, G.S., Zhang, L.A., O'Connell, C., Lohn, A.J. and Aguirre, J. (2022), "Empirical evaluation of physical adversarial patch attacks against overhead object detection models". DOI: <https://doi.org/10.48550/arXiv.2206.12725>.
4. Lorena, A.C., Garcia, L.P., Lehmann, J., Souto, M.C. and Ho, T.K. (2019), "How complex is your classification problem? A survey on measuring classification complexity". *ACM Computing Surveys (CSUR)*, 52(5), pp.1-34. DOI: <https://doi.org/10.1145/3347711>.
5. Gummadi, D., Chan, P.H., Wang, H. et al. (2023), "Correlating traditional image quality metrics and DNN-based object detection: a case study with compressed camera data". *TechRxiv*. DOI: <http://dx.doi.org/10.36227/techrxiv.24566371.v1>.
6. Branchaud-Charron, F., Achkar, A. and Jodoin, P.-M. (2019), "Spectral Metric for Dataset Complexity Assessment". *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, pp. 3210-3219. DOI: <https://doi.org/10.1109/CVPR.2019.00333>.
7. Le, T.-N., Nguyen, T.V., Nie, Z., Tran, M.-T. and Sugimoto, A. (2019), "Anabranch network for camouflaged object segmentation". *Computer Vision and Image Understanding*, 184, pp.45-56. DOI: <https://doi.org/10.1016/j.cviu.2019.04.006>.

8. “The A to Z of Data Annotation”. URL: <https://uk.shaip.com/blog/the-a-to-z-of-data-annotation/>. [Accessed 21 February 2024].
9. Osharovska, O.V. (2016). “Otsinyuvannya pokaznykiv yakosti tv zobrazen” [Evaluation of TV brightness indicators image]. *Tsifrovi Tehnologiyi*. ONAZ im. O.S. Popova, (19), pp.91-102. [in Ukrainian].
10. Ocampo, R. (2024), “Automatic image quality assessment in python”. URL: <https://towardsdatascience.com/automatic-image-quality-assessment-in-python-91a6be52c11> [Accessed 29 February 2024].
11. Le, T.-N., Cao, Y., Nguyen, T.-C. et al. (2021), “Camouflaged instance segmentation in-the-wild: Dataset, method, and benchmark suite”. *IEEE Transactions on Image Processing*, 31, pp.287-300. DOI: <https://doi.org/10.1109/TIP.2021.3130490>.
12. Fan, D.-P., Ji, G.-P., Cheng, M.-M. and Shao, L. (2022), “Concealed Object Detection”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10), pp.6024-6042. DOI: <https://doi.org/10.1109/TPAMI.2021.3085766>.
13. Dong, F. and Zhu, J. (2023), Underwater Camouflage Object Detection Dataset. *arXiv preprint arXiv:2303.12946*. DOI: <https://doi.org/10.48550/arXiv.2303.12946>.
14. Lv, Y., Zhang, J., Dai, Y., Li, A., Liu, B., Barnes, N. and Fan, D.P. (2021), Simultaneously localize, segment and rank the camouflaged objects. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.11591-11601. DOI: <https://doi.org/10.48550/arXiv.2103.04011>.
15. Xi, X., Wu, J., Ma, B., Ou, Y. and Liu, L. (2021), “LGA-RCNN: Loss-Guided Attention for Object Detection”. *arXiv preprint arXiv: 2104.13763*. DOI: <https://doi.org/10.48550/arXiv.2104.13763>.
16. Liu, M. (2023), “A military reconnaissance network for small-scale open-scene camouflaged people detection”. *Expert Systems*, 40(10), e13444. DOI: <https://doi.org/10.1111/exsy.13444>.
17. Zheng, Y., Zhang, X., Wang, F., Cao, T., Sun, M. and Wang, X. (2019), “Detection of people with camouflage pattern via dense deconvolution network”. *IEEE Signal Processing Letters*, 26(1), pp.29-33. DOI: <https://doi.org/10.1109/LSP.2018.2825959>.
18. Timchuk, V., Litvin, V., & Pereguda, O. (2023), “Dekompozitsiya mashini glibinnogo navchannya na osnovi naboriv spetsializovanih datasetiv dlya zmenshennya chasu obrobki prostorovoyi Informatsiyi” [Decomposition of a clay learning machine based on sets of specialized data networks to reduce the processing time of spatial information]. *Viskovo-tehnichniy zbirnik*, (28), pp.60–68. doi: 10.33577/2312-4458.28.2023.60-68. [in Ukrainian].
19. Google Colab. URL: <https://colab.research.google.com/>. [Accessed 29 February 2024].
20. Mahrovskaya, N.A. and Pogromskaya, G.S. (2019), “Algoritmi i strukturi danih: navchalno-metodichniy poslbnik” [Algorithms and data structures: a basic methodological textbook]. *MikolaYiv: MikolaYivskiy natsionalniy univrsitet Imeni V.O. Suhomlinskogo*. 279 p. [in Russian].

ASSESSMENT OF THE COMPLEXITY OF THE SET OF INPUT DATA FOR A COMPUTER VISION SYSTEM

Tsybulia S., Tymchuk V., Volokyta A.

The article considers the actual problems of using neural networks in the development of computer vision systems for military purposes, with an emphasis on the analysis of input datasets. The purpose of the article is to develop a method for assessing the complexity of data sets based on the size and quality of input images, and the level and structure of their annotation. The application of a preliminary assessment of the complexity of datasets will allow analyzing the influence of the specified properties of the input data on the duration of training, the consumption of hardware resources, and the overall effectiveness of computer vision models designed for the detection and identification of masked objects.

To achieve the goal, the research used methods of analyzing the internal structure and properties of datasets, including the analysis of image quality, sizes, variety of objects, and complexity of annotations. The obtained results demonstrate a direct relationship between the complexity of datasets and the efficiency of model training, revealing the key parameters that affect the accuracy and speed of training.

Special attention was paid in the study to the development of a universal method of assessing the complexity of datasets, which allows combining various parameters into a single quantitative indicator. This approach not only contributes to a deeper analysis of the influence of the characteristics of datasets on the learning process, but also allows to predict the costs of computing resources. The theoretical significance of the work lies in the formation of a holistic view of the factors affecting the effectiveness of machine learning in the field of computer vision for military applications. Practical significance is revealed through the possibility of using the results to optimize the development and training of neural networks aimed at increasing the effectiveness of military recognition systems.

The research results contribute to the development of the theory and practice of creating computer vision systems, offering a toolkit for more effective selection or creation of datasets taking into account the requirements of military and technical tasks. In further research, it is planned to expand the use of the proposed indicator on various types of datasets and neural network architectures, which should contribute to the universality and scalability of the proposed approach for evaluating the complexity of datasets.

Keywords: machine learning, computer vision, masker, hiding, terrain image, complexity index, dataset, dataset, image quality, input, object detection, object classification, meta-information, annotation, annotation.