Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 2 (32)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

УДК 004
DOI: https://doi.org/10.30837/2522-9818.2025.2.079

О. Прокопенко, С. Смеляков

РОЗРОБЛЕННЯ ОБ'ЄКТНО-ОРІЄНТОВАНОГО АЛГОРИТМУ ПОРІВНЯННЯ ЗОБРАЖЕНЬ ДЛЯ ЇХ ЕФЕКТИВНОГО ПОШУКУ

Об'єктом у статті є пошук зображень на основі змісту. **Предмет дослідження** – моделі та методи пошуку зображень на основі змісту (CBIR) і управління значними обсягами медіаконтенту у великих системах збереження зображень. **Мета статті** – розроблення алгоритму порівняння об'єктно-орієнтованих дескрипторів зображень, що передбачає використання передових моделей комп'ютерного зору для виявлення об'єктів і побудови ефективних методів порівняння й пошуку цих дескрипторів. Запропонований дескриптор і алгоритм порівняння мають на меті підвищити ефективність і точність процесів пошуку зображень і управління ними. **Завдання:** аналіз сучасних підходів і рішень для створення та порівняння дескрипторів зображень та їх використання в пошуку зображень на основі змісту (CBIR); розроблення метрик і алгоритмів порівняння дескрипторів зображень, що ефективно використовують інформацію про виявлені об'єкти, такі як їх типи, розміри та місце розташування для пошуку зображень у великих сховищах даних; проведення експериментів для оцінювання запропонованого алгоритму пошуку за зображенням і порівняння ефективності з наявними рішеннями. **Методологія** передбачає всебічний огляд передових методів створення дескрипторів зображень, зокрема: геш-дескрипторів, створених вручну дескрипторів, дескрипторів на основі глибокого навчання; аналіз використання наявних дескрипторів у системах CBIR, зважаючи на їх переваги й обмеження; аналіз найкращих алгоритмів пошуку за зображенням, зокрема з підходами, що використовують глибоке навчання; розроблення алгоритму порівняння об'єктних дескрипторів для завдань пошуку за тегами, зображенням тощо. **Досягнуті результати**: розроблено дескриптор зображення, оснований на об'єктах, виявлених за допомогою сучасних моделей машинного навчання; розроблено метрики й алгоритми порівняння запропонованих дескрипторів, що дають змогу використовувати їх для пошуку зображень на основі змісту у великих сховищах даних; проведено серію експериментів для оцінювання ефективності та якості пошуку у великих системах збереження зображень за допомогою запропонованого дескриптора та алгоритмів. Експерименти дали змогу порівняти їх ефективність з наявними методами, виявивши їх переваги й обмеження, а саме: більш швидке створення дескриптора, більш швидке порівняння дескрипторів, ніж гешовані, створені вручну, та дескриптори на основі глибокого навчання, ефективне фільтрування зображень у сховищі, вища якість та швидкість пошуку зображень, але ефективність дескриптора залежить від якості моделі та даних, що використовуються для виявлення об'єктів, оскільки зображення без виявлених об'єктів не з'являються внаслідок пошуку, що може обмежувати повноту пошуку. **Висновки**. Розроблений алгоритм порівняння об'єктно-орієнтованих дескрипторів зображень є ефективним інструментом для розв'язання низки завдань пошуку зображень на основі змісту. Досягнуті результати є задовільними, оскільки розроблений алгоритм пошуку зображень перевершує більшість аналогів за швидкістю та якістю пошуку. Перспективним напрямом цього дослідження є побудова системи пошуку зображень на основі змісту з використанням розробленого дескриптора та алгоритмів, посилене застосування паралельних і розподілених обчислень, доопрацювання під конкретні потреби, що дасть змогу використовувати його не тільки в контексті зображень загального призначення, а й для більш точних наукових напрямів.

Ключові слова: оброблення зображень; детекція об'єктів; глибоке навчання; дескриптор зображень; пошук зображень; великі дані; зберігання зображень; оптимізація пошуку; інформаційні технології.

1. Аналіз сучасного стану розвитку

1.1. Мотивація

Люди отримують величезну кількість інформації з навколишнього середовища за допомогою зору. Як наслідок, з появою комп'ютерних технологій значні ресурси були спрямовані на вилучення корисної інформації із зображень, її формалізацію та використання для аналізу ситуацій і прийняття рішень у системах комп'ютерного зору, моніторингу, підтримки рішень і штучного інтелекту [1].

Спочатку через обмеження ранніх комп'ютерних ресурсів, зображення могли бути подані лише в низькій якості, займаючи всього кілька сотень байтів або кілобайтів.

Зі швидким розвитком технологій інформаційні ресурси також зазнали вибухового зростання [2]. Тепер зображення можуть займати до 32 гігабайтів за секунду в разі використання високошвидкісного відео для спеціальних фільмів. Навіть не беручи до уваги високоякісні зображення, повсякденні зображення зазвичай займають від кількох до десятків мегабайтів. Хоча це дає змогу отримати

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 2 (32)

майже реалістичне візуальне подання, це також створює значні виклики для розроблення ефективних алгоритмів, здатних обробляти такі великі обсяги інформації в режимі реального часу. Цей виклик ще більше ускладнюється швидким зростанням розмірів і кількості сучасних систем збереження інформації. Крім того, зображення часто зазнають дублювання й трансформації протягом свого життєвого циклу, чи то внаслідок покращень, ініційованих користувачами, чи через автоматичні сервіси оброблення, що стискають зображення за допомогою втратних форматів, таких як JPEG.

Багатство візуальної інформації передбачає неабиякі вимоги до пам'яті та обчислювальних ресурсів для збереження та оброблення сцен. У завданнях навігації в реальному часі така кількість інформації може стати непосильною. Тому необхідно подавати зображення за допомогою дескрипторів, що зводять інформацію до вектора ознак, але зберігають здатність розпізнавати зображення з-поміж інших у базі даних [3]. Оскільки час і увага людини є принципово обмеженими ресурсами, лише незначна частина завантажених зображень буде реально переглянута згодом [4].

Пошук зображень на основі контенту (CBIR) є дуже важливою та складною проблемою, що висвітлюється в багатьох наукових і бізнесових сферах діяльності людини, зокрема медичній промисловості [5], віртуальній реальності [6], іграх, роздрібній торгівлі, безпеці [7], соціальних медіаплатформах [8], візуальних пошукових системах [9], автомобілебудівництві тощо. Кожна із згаданих сфер потребує різного підходу до аналізу вмісту зображень. Залежно від цілей оброблення можемо детально аналізувати об'єкти на зображенні, а також оточення, текстури фону, елементи переднього плану сцени тощо. Тому CBIR є дуже широким напрямом досліджень у комп'ютерних науках. Проте, незалежно від цілей застосування та обраного методу, ефективність завжди відіграє найважливішу роль [10]. Ефективність системи вилучення визначається тим, наскільки добре вона може відновлювати ознаки за допомогою дескриптора ознак [11].

Основною ідеєю є використання зображення, предмета, теми або інтересу як вхідних даних, а відповідні зображення можуть бути отримані з бази даних за допомогою CBIR. Швидке зростання баз даних зображень для системи CBIR ускладнило отримання схожих зображень із вищою точністю [12].

Незважаючи на значні дослідження, CBIR залишається складним завданням, особливо зі зростанням розмірів наборів даних. У CBIR статистичні характеристики є надзвичайно важливими, особливо коли вони поєднуються з класифікаторами машинного навчання. Точність пошуку зображень залежить від атрибутів, що містять кольори, текстури та форми, які вилучаються із зображень [13].

Технологія пошуку контенту є обчислювальним методом, основаним на поведінці розпізнавання, що імітує взаємодію та співпрацю між індивідуумами для досягнення мети глобальної оптимізації [14].

Сучасні тенденції в пошуку зображень на основі контенту (CBIR) спрямовані на потребу в передових моделях і алгоритмах для вилучення, порівняння та пошуку інформації на основі вмісту зображень. Щоб забезпечити спеціалізованим сервісам можливість пошуку, аналізу та прийняття рішень у реальному часі, зазвичай створюється база даних дескрипторів. У цьому підході дескриптори генеруються та зберігаються в допоміжному сховищі. Коли зображення потрапляє в пошукову систему, його дескриптор створюється під час операції пошуку. Фактичний пошук і порівняння зображень у сховищі виконуються за допомогою цих дескрипторів, що значно зменшує час пошуку та зберігає ресурси системи. Використання дескрипторів є набагато ефективнішим рішенням порівняно з виконанням операцій безпосередньо на повних файлах зображень. Однак вимоги до дескрипторів постійно зростають, оскільки якість результатів пошуку й адекватність подання зображень залежать від них. Це зі свого боку впливає на складність операцій пошуку.

Постановка проблеми

Наявні дескриптори або пріоритизують незначний розмір і високу швидкість оброблення (але здебільшого спираються на кольорові параметри, ігноруючи об'єкти на зображеннях), або використовують високовимірні вектори ознак, вилучені за допомогою глибокого навчання, які часто гарантують високу точність пошуку й глибоке розуміння контексту (але вони відстають у швидкості обчислень, що робить це рішення дорогим для великих сховищ зображень і високонавантажених потоків даних).

1.2. Аналіз наявних аналогів

Будь-які параметри, статистичні дані або трансформації зображень, що дають змогу ефективно

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 2 (32) ISSN 2524-2296 (online)

ідентифікувати й розрізняти зображення у сховищі, можуть бути використані як дескриптори зображень. Однак одним з найпоширеніших типів дескрипторів для пошуку схожих зображень є геші, згенеровані алгоритмами перцептивного гешування. Гешування є широко вживаним методом, основаним на пошуку найближчих сусідів, і використовується у завданнях пошуку зображень у великих масштабах [15]. Ці алгоритми є функціями, розробленими для генерації порівняних гешів. Приклади містять Average Hash, DCT-Based Hash, Radial Variance-Based Hash та Marr-Hildreth Operator-Based Hash. Узагальнений процес роботи цих алгоритмів передбачає:

– застосування особливих трансформацій до зображення для вилучення шуму;

– нормалізацію зображення до фіксованого розміру;

– виконання математичних трансформацій, визначених алгоритмом;

– обчислення гешу.

Для порівняння геш-дескрипторів використовують такі метрики, як відстань Хеммінга, нормалізована відстань Хеммінга та пікова функція взаємної кореляції. Деякі алгоритми гешування стійкі до певних трансформацій зображень, зокрема розмиття, масштабування, стиснення, затемнення та незначні повороти. Важливою перевагою геш-дескрипторів є швидкість їх порівняння, оскільки геші зазвичай займають до 128 байтів. Однак суттєвим недоліком є те, що вони не беруть до уваги об’єкти на зображенні. Це може призводити до ситуацій, коли зовсім різні об’єкти реального світу з подібними кольоровими схемами неправильно ідентифікуються як схожі, а колізії є досить поширеними.

Іншим широко використовуваним типом дескрипторів зображень є ті, що основані на ключових точках. Традиційно підхід до проєктування локальних дескрипторів зображень передбачав застосування створених вручну ознак, серед яких SIFT [16] та його варіанти (SURF [17], GLOH [18], PCA-SIFT [19]), є, мабуть, найпопулярнішими. Деякі створені вручну дескриптори генерують кодування в просторі Хеммінга – їх найчастіше називають бінарними дескрипторами. Прикладами таких є BRIEF [20], ORB [21], BRISK [22], FREAK [23] та LDAHash [24].

Ці алгоритми зазвичай передбачають два етапи:

– виявлення ключових точок – ділянок, що значно відрізняються від свого оточення, наприклад контури або кути об’єктів;

– генерація дескрипторів для цих точок, що дає змогу порівнювати точки на різних зображеннях, щоб визначити, чи належать вони до одного об’єкта.

Сукупність дескрипторів ключових точок утворює дескриптор зображення. Порівняння цих дескрипторів дає змогу ідентифікувати однакові об’єкти на різних зображеннях, навіть за різних кутів, масштабів, умов освітлення та інших трансформацій.

Порівнюючи дескриптори на основі ключових точок із геш-дескрипторами, перші мають значні переваги. На відміну від геш-дескрипторів, вони допомагають виявляти об’єкти, реконструювати 3D-сцени, зшивати зображення та виконувати інші складні завдання комп’ютерного зору. Однак створення та використання дескрипторів ключових точок є обчислювально витратним. Це робить їх менш придатними для порівняння великих обсягів зображень, оскільки час порівняння значно залежить від кількості ключових точок, виявлених на кожному зображенні.

Зіставлення ознак зображень є невід’ємним завданням для багатьох застосунків комп’ютерного зору, таких як відстеження об’єктів, пошук зображень тощо. Зображення можна зіставляти незалежно від того, як вони змінюються внаслідок геометричних трансформацій (наприклад, повороти та зсуви), освітлення тощо. Також завдяки успішному застосуванню глибокого навчання в обробленні зображень метод глибокого навчання має перевагу у вилученні ознак із зображень [25].

Дескриптори зображень на основі глибокого навчання стали потужною альтернативою традиційним методам водночас із використанням можливості згорткових нейронних мереж (CNN) та інших нейронних архітектур для вилучення високодискримінантних ознак із зображень. На відміну від створених вручну дескрипторів, основаних на попередньо визначених алгоритмах, моделі глибокого навчання навчаються подавати ознаки безпосередньо з даних, захоплюючи як низькорівневі деталі (наприклад, краї, текстури), так і високорівневі семантичні дані (зокрема об’єкти, сцени). Ці дескриптори генеруються способом пропускання зображення крізь навчену мережу та вилучення карт ознак або вбудовувань із проміжних шарів. Зокрема для пошуку зображень низка підходів безпосередньо використовує активації мережі як ознаки зображень і успішно виконує пошук зображень [26].

Популярні підходи передбачають застосування попередньо навчених моделей, зокрема VGG, ResNet

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 2 (32)

та EfficientNet, як екстракторів ознак, або навчання спеціалізованих мереж для конкретних завдань. Дескриптори на основі глибокого навчання мають переваги в таких завданнях, як пошук зображень, розпізнавання об'єктів і семантична сегментація, оскільки вони можуть узагальнювати інформацію з різних наборів даних і адаптуватися до складних візуальних шаблонів. Однак вони часто вимагають значних обчислювальних ресурсів для навчання та виведення, а їх ефективність значно залежить від якості та різноманітності навчальної інформації. Незважаючи на окреслені виклики, здатність захоплювати багаті, ієрархічні ознаки робить їх основним елементом сучасних систем комп'ютерного зору. Вони застосовуються в багатьох сферах, особливо в медицині [27].

Останні дослідження показують зростання інтересу до вилучення деталізованих ознак і крос-модальних підходів. Ці методи зосереджуються на навчанні деталізованої семантичної інформації для збереження високорівневої семантичної релевантності за умови усунення модальних ознак. Це приводить до більш точної та адекватної репрезентації оригінального зображення [28].

Узагальнене подання моделі CBIR зображено на рис. 1.

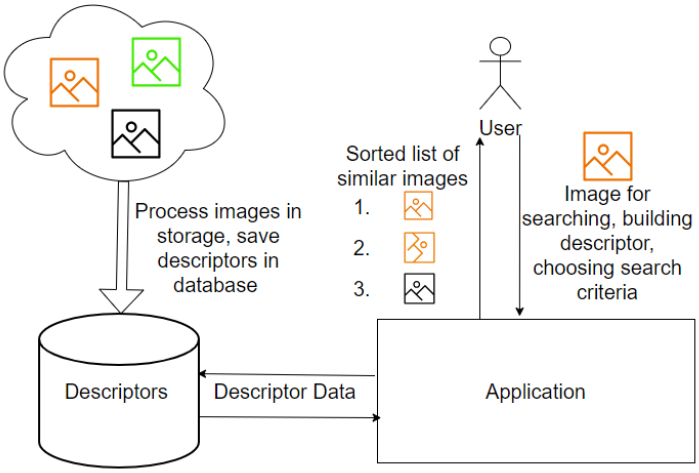

Рис. 1. Подання моделі CBIR

Пошук зображень на основі контенту (CBIR) набув значного розвитку за останні кілька десятиліть, перейшовши від традиційних методів вилучення створених у ручний спосіб ознак до сучасних підходів, основаних на глибокому навчанні. Власне, CBIR має на меті отримати зображення з бази даних, які візуально схожі на запитуване зображення, на основі їх контенту, а не текстових метаданих. Ранні системи CBIR покладалися на низькорівневі ознаки, такі як кольорові гістограми, текстури та дескриптори країв. Хоча ці методи були обчислювально ефективними, вони часто не могли охопити високорівневі семантичні дані, що призводило до обмеженої точності в реальних застосунках. Наприклад, два зображення з подібними кольоровими розподілами, але зовсім різними об'єктами могли бути неправильно зіставлені. Незважаючи на зазначені обмеження, базові техніки заклали основу для більш складних систем пошуку.

Поява глибокого навчання революціонізувала CBIR, даючи змогу системам навчатися багатого, ієрархічного подання ознак безпосередньо з даних. Сучасні системи CBIR використовують згорткові нейронні мережі (CNN) та інші архітектури глибокого навчання для вилучення високовимірних векторів ознак або вбудовувань, що охоплюють як низькорівневі, так і високорівневі візуальні дані. Попередньо навчені моделі, зокрема VGG, ResNet та EfficientNet, часто застосовуються як екстрактори ознак, тоді як більш спеціалізовані підходи, а саме сіамські мережі та мережі з трійками, упроваджуються для навчання подання ознак, оптимізованих для порівняння схожості. Ці досягнення значно покращили точність пошуку, даючи змогу системам CBIR обробляти складні запити, що містять об'єкти, сцени й навіть деталізовані деталі. Наприклад, запитуване зображення певного виду птаха тепер може отримати зображення того самого виду з різних ракурсів та умов освітлення

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 2 (32)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

завдяки семантичній багатогранності дескрипторів на основі глибокого навчання.

Наявні дескриптори для порівняння та пошуку зображень часто пріоритизують незначний розмір та високу швидкість оброблення, але вони здебільшого спираються на кольорові параметри, ігноруючи об'єкти на зображеннях. Хоча цей підхід добре працює в застосунках з унікальними зображеннями, він страждає від колізій, коли зображення мають подібні інтегральні властивості. Це призводить до неадекватних результатів порівняння та пошуку, оскільки дескриптори базуються винятково на інтегральних властивостях яскравості та кольору.

Іншою помітною тенденцією в сучасному CBIR є інтеграція крос-модальних технік пошуку, що уможливлюють пошук серед різних типів даних, наприклад, отримувати зображення за текстовими запитами або навпаки. Це досягається за допомогою моделей, зокрема CLIP (*Contrastive Language-Image Pretraining*), які навчають спільні вбудовування для зображень і тексту в єдиному просторі ознак. Такі системи особливо ефективні в таких сферах, як електронна комерція, де користувачі можуть шукати товари за описовим текстом, або в медичній візуалізації, де діагностичні звіти використовуються для пошуку відповідних медичних зображень. Крім того, механізми уваги й трансформери все частіше впроваджуються в CBIR для зосередження на найбільш релевантних частинах зображення, що значно підвищує точність пошуку.

Незважаючи на ці досягнення, у сфері CBIR залишаються виклики. Однією з основних проблем є масштабованість систем пошуку, оскільки обчислювальна вартість порівняння високовимірних векторів ознак зростає зі збільшенням розміру бази даних. Техніки, такі як гешування та векторне квантування, часто використовуються для стиснення векторів ознак і прискорення пошуку, але ці методи іноді можуть погіршувати точність. Іншим викликом є інтерпретованість систем CBIR на основі глибокого навчання, оскільки їх процеси прийняття рішень часто не прозорі. Дослідники активно вивчають способи зробити ці системи більш прозорими та зручними для користувачів. Крім того, продуктивність систем CBIR значно залежить від якості та різноманітності навчальних даних, що може призводити до упереджень та обмежувати узагальнюваність.

1.3. Мета й підходи

Метою дослідження є розроблення алгоритму порівняння об'єктно-орієнтованих дескрипторів зображень, що передбачає використання передових моделей комп'ютерного зору для виявлення об'єктів і побудови ефективних методів порівняння й пошуку цих дескрипторів. Запропонований дескриптор і алгоритм порівняння мають на меті підвищити ефективність і точність пошуку й управління зображеннями. Цей підхід забезпечує швидкість обчислень завдяки обмеженій кількості інформації, що використовується, а також надає глибокі контекстуальні дані для порівняння. Об'єкти, як складні сутності, несуть значну частину інформаційного вмісту зображення, що забезпечує більш змістовні й точні результати пошуку. Цей підхід пропонує збалансоване рішення для пошуку зображень на основі контенту, значно покращуючи точність традиційних дескрипторів, і водночас є значно швидшим, ніж високовимірні вектори ознак, що застосовуються в дескрипторах на основі глибокого навчання. Топові результати, відібрані запропонованим методом, можуть бути додатково проаналізовані за допомогою більш обчислювально витратних технік, що забезпечує як ефективність, так і точність у процесі пошуку.

Основна ідея розробленого дескриптора полягає в поєднанні низьких обчислювальних витрат дескрипторів, основаних на низькорівневих ознаках зображень, із високорівневими даними про об'єкти, які надаються згортковими нейронними мережами (CNN), що зазвичай вимагають значних обчислювальних ресурсів. Цей підхід має на меті збалансувати ефективність і точність, використовуючи переваги обох методів для створення надійного та масштабованого рішення для завдань пошуку зображень.

Розроблений дескриптор має відповідати кільком критеріям / вимогам.

- Ефективність: використовувати мінімальну кількість даних про об'єкти для забезпечення обчислювальної можливості та ефективності зберігання.
- Адаптивність: бути сумісним з різними моделями детекції об'єктів, що забезпечує гнучкість у застосуванні.
- Налаштовуваність: уможливлювати налаштування для задоволення особливих потреб різних наборів даних і сценаріїв використання.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 2 (32)

– Баланс: забезпечувати оптимальний баланс між швидкістю пошуку та якістю результатів, щоб гарантувати як швидкий пошук, так і точність.

– Простота побудови: бути простим у створенні, що дає змогу ефективно будувати базу дескрипторів і легко інтегрувати їх у пошукові системи.

Для досягнення окресленої мети необхідно виконати такі завдання:

– визначити поля даних для збереження об'єктів, щоб забезпечити легкість дескриптора й достатню інформацію для ефективного пошуку;

– розробити дизайн збереження дескриптора, що буде адаптивним до різних моделей детекції об'єктів і налаштовуваним для задоволення особливих вимог застосунків;

– розробити метрики для порівняння запропонованих дескрипторів, алгоритми для порівняння дескрипторів, які використовуються для розв'язання завдань пошуку зображень на основі змісту;

– обрати відповідну модель детекції об'єктів для експериментів, що дасть змогу справедливо порівнювати з іншими дескрипторами;

– обрати набір даних для проведення серії експериментів, які окреслять переваги й недоліки розробленого дескриптора, метрик і алгоритмів;

– створити програмне забезпечення для реалізації побудови бази дескрипторів для зберігання зображень і забезпечення ефективного порівняння дескрипторів.

З розробленим алгоритмом і реалізованим програмним забезпеченням можемо застосовувати обраний набір даних і модель детекції об'єктів для проведення точного й швидкого пошуку у великих сховищах зображень.

2. Матеріали та методи дослідження

2.1. Вибір моделі для детекції об'єктів

Детекція об'єктів є важливим складником систем комп'ютерного зору. Вона може застосовуватися в багатьох сферах, таких як відеоспостереження [29], медична візуалізація [30] та навігація роботів [31]. Для розв'язання цього завдання можна використовувати чимало алгоритмів, зокрема вилучення фону, тимчасове диференціювання, оптичний потік, фільтрація Калмана, метод опорних векторів і зіставлення контурів [32].

Ми обрали нейронну мережу, спеціалізовану на детекції об'єктів, для вилучення інформації із зображень. Обрана модель має відповідати таким критеріям:

– бути широко використовуваною та часто застосовуватися для завдань детекції об'єктів, що гарантує її ефективність у різних умовах;

– бути відносно простою у використанні, здатною виявляти значну кількість об'єктів і уможливлювати легке перетренування для адаптації до нових класів об'єктів або спеціалізації на певних типах об'єктів;

– мати кілька версій, які розрізняються кількістю параметрів, швидкістю й точністю, що дає змогу використовувати одну й ту саму архітектуру як в умовах обмежених ресурсів (наприклад, на мобільних пристроях), так і в сценаріях із значною обчислювальною потужністю.

Важливо наголосити, що різні модифікації нейронної мережі можуть бути оптимізовані для зображень певних розмірів. Це є критично важливим, оскільки швидкість оброблення залежить від цього. Нейронні мережі зазвичай добре обробляють масштабування великих зображень, але не завжди можуть адекватно масштабувати малі зображення.

Моделі детекції об'єктів можуть підвищити зрозумілість, розпізнаючи ділянки інтересу на запитуваних і знайдених зображеннях. Детекція об'єктів у глибокому навчанні охоплює дві основні парадигми: двохетапні алгоритми детекції, основані на якорних прямокутниках, та одноетапні алгоритми детекції, основані на без'якорних прямокутниках. До першої категорії належать відомі моделі, такі як R-CNN, Fast R-CNN, Faster R-CNN та R-FCN, що зазвичай досягають високої точності, але потребують більшого часу на детекцію. На противагу цьому, до другої категорії належать алгоритми SSD та YOLO, які використовують пряму мережу виведення для ефективного визначення місця розташування цілей та генерації результатів класифікації [33].

З огляду на ці вимоги до моделі детекції об'єктів, яка буде інтегрована в систему управління великими сховищами зображень, ми обрали модель YOLOv8 для цієї роботи. На момент написання YOLOv8 є одним із найбільш передових рішень у сфері комп'ютерного зору для завдань детекції об'єктів, пропонуючи баланс швидкості, точності та гнучкості.

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 2 (32) ISSN 2524-2296 (online)

YOLOv8 є найновішою моделлю, побудованою на успіху попередніх версій YOLO, і містить нові функції та покращення для подальшого підвищення продуктивності та гнучкості. YOLOv8 розроблена як швидкий, точний і зручний інструмент, що робить її ефективним вибором для широкого спектра завдань, зокрема детекції та відстеження об'єктів, сегментації, класифікації зображень і оцінювання постав.

Основною перевагою моделі YOLO над конкурентами в розв'язанні поставленого завдання є швидкість, що відіграє вирішальну роль. Вона у 2–2,5 раза швидша за інші популярні моделі, такі як Faster R-CNN, SSD, RetinaNet, маючи водночас порівняну (а іноді навіть вищу) точність, на відміну від своїх конкурентів. YOLOv8 має одноетапну архітектуру детекції, якщо порівнювати із сімейством нейронних мереж R-CNN, які виявляють об'єкти за два проходи.

Ще одним важливим аргументом на користь вибору YOLOv8 є можливість легко використовувати графічний процесор для прискорення детекції об'єктів способом виклику функції детекції об'єктів з одним додатковим параметром, що пришвидшує оброблення зображень у 5–10 разів.

Мережа масштабує оригінальне зображення до кількох карт ознак за допомогою пропускних зв'язків та інших архітектурних технік. Отримані карти ознак зменшуються до однієї роздільної здатності за допомогою апсемплінгу й конкатенації. Потім прогнозуються класи та обмежувальні прямокутники для об'єктів, і за допомогою придушення неповних максимумів обирається найбільш імовірний обмежувальний прямокутник для кожного об'єкта. Інформація про кожен обмежувальний прямокутник подана шістьма значеннями: номер класу, координати центру X та Y, ширина, висота й рівень довіри.

2.2. Загальний алгоритм пошуку зображення

На початковому етапі, щоб забезпечити функціональність пошуку, ми виявляємо об'єкти на кожному зображенні в сховищі за допомогою нейронної мережі для детекції об'єктів. На основі цих виявлень створюємо та заповнюємо базу даних дескрипторів зображень, основаних на об'єктах. Ця база даних дескрипторів слугує основою для ефективних операцій пошуку зображень у майбутньому. Детальний опис моделей нейронних мереж, дескрипторів і методів їх побудови наведені нижче.

Далі використовуються розроблені метрики й алгоритми порівняння для оцінювання схожості між зображеннями (точніше, їх дескрипторами), а також алгоритм пошуку, який використовує ці дескриптори та метрики для оцінювання схожості зображень.

Розглянемо загальний процес пошуку.

1. Побудова дескриптора: генерація дескриптора для запитуваного зображення.

2. Налаштування пошуку: визначення критеріїв пошуку та встановлення параметрів (наприклад, пороги схожості, типи об'єктів).

3. Виконання пошуку:

– ітерація по дескрипторах у сховищі;

– порівняння кожного дескриптора з дескриптором запитуваного зображення за допомогою метрики схожості;

– сортування зображень за оцінкою їх схожості;

– відтворення результатів за схожістю.

В ідеальному сценарії відсортовані результати пошуку відтворюватимуть безпосередньо запитуване зображення (якщо воно є в сховищі) на першому місці, потім трансформації або варіації запитуваного зображення (наприклад, змінені розміри, обрізані або відредаговані версії), а потім інші зображення зі сховища, ранжовані за їх схожістю до запиту.

Для перевірки алгоритмів і програмного забезпечення розроблено допоміжні моделі, методи, критерії, метрики та інструменти програмного забезпечення, зокрема метрики для оцінювання якості пошуку. Експериментальна установка, результати й детальний аналіз подані у відповідних розділах нижче.

2.3. Дані дескрипторів

Дескриптор зображення є компактним поданням, що зберігає основну інформацію про зображення в сховищі. Це передбачає: розташування зображення та його формальні параметри (наприклад, роздільна здатність, формат), а також інформацію про об'єкти, виявлені на зображеннях, зокрема їх розташування (координати обмежувального прямокутника), розміри та площі, рівень довіри моделі комп'ютерного зору, що обмежувальний прямокутник містить об'єкт певного типу (наприклад, тип i). Завдяки цьому мінімальному, але всеосяжному набору інформації можна ефективно будувати методи пошуку, фільтрації та порівняння зображень на основі їх вмісту, зокрема об'єктів, які вони містять.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 2 (32)

2.3.1. Дескриптор об'єкта

Кожен об'єкт подано вектором із семи чисел, як показано в табл. 1. Дескриптор об'єкта не містить номера класу, до якого належить об'єкт. Для неявного зберігання номера класу кожного об'єкта використовується структура дескриптора класу.

Таблиця 1. *Дані дескриптора об'єкта*

Назва	Опис
x	Відносна координата за віссю *x* центра обмежувальної рамки об'єкта (діапазон 0–1)
y	Відносна координата за віссю *y* центра обмежувальної рамки об'єкта (діапазон 0–1)
w	Відносна ширина об'єкта (діапазон 0–1)
h	Відносна висота об'єкта (діапазон 0-1)
area	Відносна площа об'єкта
ratio	Відношення меншої до більшої сторони обмежувальної рамки об'єкта
conf	Впевненість моделі в тому, що рамка містить об'єкт класу *i* (діапазон 0–1)

2.3.2. Дескриптор класу

Дескриптор класу описує набір об'єктів, що належать до певного класу в межах зображення. Він містить колекцію дескрипторів для кожного об'єкта цього класу, виявленого на зображенні, разом із додатковою інформацією про групу об'єктів.

Ця додаткова інформація передбачає кількість об'єктів класу, загальну площу, яку вони займають, та арифметичний центр (центроїда) групи на зображенні (див. табл. 2). Хоча ці дані можна обчислити безпосередньо з дескрипторів об'єктів, їх попереднє обчислення усуває зайві розрахунки під час порівнянь, значно скорочуючи час порівняння та підвищуючи загальну ефективність.

Таблиця 2. *Дані дескриптора класу*

Назва	Опис
class	Ідентифікатор класу
number	Кількість об'єктів цього класу на зображенні
area	Сума відносних площ, які займають об'єкти цього класу на зображенні
center	Арифметичний центр групи об'єктів цього класу на зображенні
objects	Колекція дескрипторів об'єктів цього класу на зображенні

2.3.3. Дескриптор класів

Дані дескриптора класів агрегують дескриптори класів для всіх можливих типів об'єктів. Завдання полягає в забезпеченні зручного стандартизованого способу зберігання дескрипторів класів, що дає змогу проводити майбутні обчислення на графічному процесорі (табл. 3). Кількість класів відповідає кількості класів, використаних з метою навчання нейронної мережі для детекції об'єктів.

Таблиця 3. *Дані дескриптора класів*

#	Клас	Кількість	Площа	
1	*person*	2	0.012	…
2	*car*	0	0	…
3	*dog*	0	0	…
4	*cat*	0	0	…
5	*tie*	1	0.074	…

2.3.4. Дескриптор зображення

Дескриптор зображення інкапсулює основні дані, необхідні для локалізації та опису оригінального зображення. Він містить:

– розташування зображення: шлях до зображення, наприклад, локальний шлях до файлу, URI або інший ідентифікатор, щоб отримати повний файл зображення за потреби;

– розміри зображення: абсолютна ширина й висота зображення, необхідні для обчислення абсолютних координат і розмірів обмежувальних прямокутників, оскільки вся інформація про об'єкти зберігається у відносних величинах;

– дескриптори класів: колекція дескрипторів для кожного класу об'єктів, виявлених на зображенні.

Структура дескриптора зображення є гнучкою та може бути адаптована для збереження додаткової інформації про зображення та його об'єкти за потреби. Хоча дескриптор може бути налаштований для конкретних завдань, основною вимогою є збереження його компактності для швидкого оброблення під час роботи з великими сховищами зображень. Дані, пов'язані з моделлю дескриптора зображення, узагальнені в табл. 4.

Таблиця 4. Дані дескриптора зображення

Назва	Опис
path	Шлях до файлу із зображенням
width	Ширина зображення
height	Висота зображення
classes	Дескриптор класів, знайдених на зображенні

2.4 Моделі пошуку, алгоритми та метрики

Дані, отримані з нейронної мережі та використані для побудови дескрипторів, можуть застосовуватися в різних моделях пошуку.

Фільтрація за класами об'єктів

Дескриптори дають змогу фільтрувати зображення в сховищі, ідентифікуючи ті, що містять

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 2 (32) ISSN 2524-2296 (online)

певну кількість об'єктів із зазначених класів. Залежно від налаштувань, модель може шукати зображення, що містять або точний набір об'єктів, або принаймні зазначений набір об'єктів. Результати такого пошуку можуть бути корисними самі по собі або як попередній фільтр для звуження ділянки пошуку. Це значно зменшує обсяг для більш складних і часозатратних процедур пошуку, підвищуючи загальну ефективність.

Пошук на основі зображень

Дескриптори також полегшують пошук на основі зображень, даючи змогу користувачам знаходити конкретні зображення або ідентифікувати схожі на основі об'єктів, зображених на них, та їх розташування в межах зображення. Ця модель пошуку може знаходити точні відповідники в сховищі та ранжувати схожі зображення за їх візуальним вмістом. За замовчуванням це вважається основною моделлю пошуку.

Пошук зображень за вказаним складом об'єктів

Вимога полягає в тому, щоб повернути всі зображення в наборі даних, які містять об'єкти, зазначені користувачем. Запит містить класи й кількість об'єктів цих класів, які мають бути присутні на зображенні. Запит перетворюється на вектор, де індекси, що відповідають класам, містять кількість таких об'єктів, утворюючи вектор, аналогічний вектору *Numbers*, що зберігається в кожному дескрипторі. Відповідь на запитання – чи необхідно повертати зображення? – зрештою зводиться до порівняння двох векторів однакової довжини.

Це завдання може бути виконано в строгій та нестрогій формах, де строга форма означає, що зображення, яке шукається, містить лише об'єкти із запиту у вказаній кількості, тобто вектор запиту та вектор дескриптора ідентичні. У разі нестрогого пошуку зображення має містити принаймні типи й кількості об'єктів із запиту, тобто значення кожного елемента вектора зображення має бути більшим або рівним відповідному елементу вектора запиту. Рівняння нижче пояснюють різницю між строгою та нестрогою фільтрацією за кількістю об'єктів.

a_{ij} – кількість об'єктів

j-го класу на i-му фото

$\begin{cases} Vec_1 = (a_{11}, a_{12}, \ldots, a_{1n}) - \text{вектор запиту} \\ Vec_2 = (a_{21}, a_{22}, \ldots, a_{2n}) - \text{вектор дескриптора} \end{cases}$

$Vec_2 = Vec_1,\ if\ all\ a_{1i} = a_{2i}$ (строгий)

$Vec_2 = Vec_1,\ if\ all\ a_{1i} \le a_{2i}$ (нестрогий)

Аналогічно можемо фільтрувати за параметрами об'єктів, записаними в дескриптор. За потреби можна використовувати двосторонні обмеження та логічні схеми для формулювання критеріїв фільтрації.

Розв'язання цього завдання корисне не лише для надання відповідей на такі запити, але й для використання подібного підходу для виконання складніших завдань попередньої фільтрації всього набору даних. Складні методи порівняння зображень, які передбачають порівняння ділянок локалізації та площ, займаних об'єктами певних класів, зазвичай є більш обчислювально витратними. Має сенс зменшити кількість порівнянь на порядок, оскільки немає сенсу порівнювати зображення, що не містять жодної пари об'єктів одного класу для такого запиту.

2.5. Метрика ступеня схожості дескрипторів

Як метрику пропонуємо використовувати відстань між зображеннями. Очікується, що ідентичні зображення матимуть відстань, рівну нулю, і що більша відстань, то менш схожим є зображення на бажане. Нагадаємо, що метод, який розробляється, має бути стійким до типових трансформацій, що часто трапляються в сховищах зображень, таких як помірні графічні зміни (наприклад, зміни яскравості, насиченості, кольорового охоплення, стиснення, масштабування тощо). Тому необхідно додатково брати до уваги те, що зображення може бути збільшене або зменшене, і працювати здебільшого з відносними, а не абсолютними розмірами зображення та об'єктів на ньому.

Для досягнення цього використано пропорції всього зображення, оскільки такі трансформації, як збільшення або зменшення, не впливають на пропорції зображення, що беруться як відношення меншої сторони до більшої сторони зображення. Отже, перший компонент метрики – це різниця в пропорціях фотографії Δp. У формулі меншу сторону позначено як h, а більшу – як w, з яких береться модуль для інваріантності:

$$\Delta p = \left| \frac{h_1}{w_1} - \frac{h_2}{w_2} \right|, \qquad (1)$$

де Δp – різниця у пропорціях двох зображень; h_1 – менша сторона першого зображення; w_1 – більша

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 2 (32)

сторона першого зображення; h_2 – менша сторона другого зображення; w_2 – більша сторона другого зображення.

Різниця в пропорціях, менша за порогове значення, вважається нульовою, щоб усунути вплив похибок масштабування, які зазвичай виникають за умови округлення довжини та ширини масштабованого зображення.

Наступна частина метрики – це порівняння параметрів зображених об'єктів, а саме співвідношення їх площ і центроїдів. Необхідно зауважити, що в цій роботі центроїд не є геометричним центром точок, які описують центри об'єктів, і не бере до уваги різні розміри об'єктів. Для обчислення центроїда о б'єктів розраховано його координати як середнє арифметичне координат усіх об'єктів i-го класу. Об'єкти обробляються за типами, кожне зображення містить нуль або більше об'єктів i-го типу, і для порівняння групи об'єктів на першому зображенні з тими, що на другому, використано такий метод.

Обчислюється відстань між центроїдами групи об'єктів i-го класу d_i на одному зображенні та на іншому за допомогою евклідової відстані. Обчислюється другий компонент метрики, який позначений Δo_i і показує відстань між зображеннями за об'єктами i-го класу. Для його обчислення множимо відстань між центроїдами на різницю площ, займаних усіма об'єктами i-го типу на двох зображеннях, і додаємо штраф, якщо зображення містять різну кількість об'єктів i-го класу. Варто зазначити, що існують два особливих випадки: по-перше, коли обидва зображення не мають об'єктів i-го типу, тоді вважаємо внесок цього типу нульовим, і, по-друге, якщо одне зображення містить об'єкти, а інше – ні, тоді аналогічно множимо різницю займаних площ на максимально можливу відстань між центроїдами, додаючи штраф за різну кількість об'єктів на зображеннях.

$$d_i = \sqrt{\left(x_{i1} - x_{i2}\right)^2 + \left(y_{i1} - y_{i2}\right)^2}, \qquad (2)$$

де d_i – відносна відстань між центроїдами групи об'єктів i-го класу на двох зображеннях; x_{i1} – координата x центроїда i-го класу на першому зображенні; x_{i2} – координата x центроїда i-го класу на другому зображенні; y_{i1} – координата y центроїда i-го класу на першому зображенні; y_{i2} – координата y центроїда i-го класу на другому зображенні.

$$\Delta o_i = d_i * \left|area_{i1} - area_{i2}\right| * \left(\left|num_{i1} - num_{i2}\right| + 1\right), \qquad (3)$$

де Δo_i – відстань між зображеннями за об'єктами i-го класу; d_i – відносна відстань між центроїдами групи об'єктів i-го класу на двох зображеннях; $area_{i1}$ – сума віносних площ об'єктів i-го класу на першому зображенні; $area_{i2}$ – сума віносних площ об'єктів i-го класу на другому зображенні; num_{i1} – кількість об'єктів i-го класу на першому зображенні; num_{i2} – кількість об'єктів i-го класу на другому зображенні.

Формула (2) використовується для обчислення відстані між центрами групи об'єктів i-го класу на двох зображеннях. Формула (3) застосовується для обчислення відстані між двома зображеннями за об'єктами i-го класу. Для досягнення кращих результатів можна запропонувати кілька модифікацій.

По-перше, як згадувалося раніше, немає сенсу порівнювати пари зображень і обчислювати відстань для зображень, які не містять жодної пари об'єктів спільного класу, тому ми попередньо фільтруємо всі такі зображення з порівняння.

По-друге, порівняння загальних площ не бере до уваги кількість об'єктів i-го типу на зображеннях, тому пропонуємо поняття штрафу для зображень, які мають різну кількість об'єктів i-го типу, що збільшує різницю в площах пропорційно до різниці в кількості об'єктів на зображеннях. Поєднуючи всі компоненти, отримуємо метрику, що дає змогу знаходити зображення в наборі даних, отримані деякими трансформаціями з оригіналу, а також ранжувати інші зображення в наборі даних за їх схожістю з оригінальним зображенням для пошуку схожих зображень у наборі даних, якщо таке завдання поставлено перед користувачем.

Кожен компонент формули (4) змінюється від 0 до 1, тому вони мають однаковий відносний вплив на результат. Якщо досягнуті результати вказують на надмірний або, навпаки, незначний вплив певного компонента на кінцеве значення, то є можливість компенсувати це, додавши або помноживши на константу. Остаточна формула має такий вигляд:

$$diff\left(image_1, image_2\right) = \Delta p + \frac{\sum_{i=0}^{n} \Delta o_i}{n}, \qquad (4)$$

де Δp – різниця в пропорціях двох зображень; Δo_i – відстань між зображеннями за об'єктами i-го класу; n – кількість унікальних класів об'єктів на двох зображеннях.

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 2 (32) ISSN 2524-2296 (online)

Отже, розроблена формула, що дає змогу швидко обчислювати міру схожості двох зображень, маючи їх дескриптори, отримані за допомогою моделі комп'ютерного зору.

Алгоритм пошуку зображення

1. На зображенні шукаються об'єкти за допомогою моделі YOLOv8.
2. За даними про знайдені об'єкти будується описаний дескриптор зображення.
3. Зі сховища даних відфільтровуються всі зображення, дескриптори яких не містять жодного об'єкта класів, знайдених на шуканому зображенні.
4. Для всіх дескрипторів із відфільтрованої вибірки розраховується відстань до шуканого зображення за формулою (4).
5. Результати сортуються та повертаються зображення з найменшими відстанями до шуканого.

Розроблений алгоритм використовує деякі параметри зображень і властивості, притаманні всій групі об'єктів, зокрема пропорції зображення, центроїди груп об'єктів, загальна площа, зайнята групою об'єктів на зображенні, та кількість виявлених об'єктів певного типу. Теоретично розроблені методи дають змогу повертати результати пошуку в сховищах, що містять сотні тисяч зображень, за час до 3 с.

Розроблений алгоритм використовує лише загальні параметри зображень і властивості, притаманні всій групі об'єктів, замість властивостей окремих об'єктів, оскільки порівняння двох зображень об'єкт за об'єктом не може бути дозволено заради мінімізації часу виконання пошуку у великих сховищах зображень. Розроблений метод має час порівняння двох дескрипторів, який лінійно залежить від кількості класів об'єктів, на яких навчена нейронна мережа, і не залежить від фактичної кількості виявлених об'єктів на зображеннях.

Для більш точного пошуку можна обрати знайдені зображення за певним порогом відстані або, наприклад, топ-20, топ-100 зображень, і для них провести порівняння, що бере до уваги розміри та положення окремих об'єктів замість груп. Використання таких обмежень є дуже ефективним з практичного боку. Однак це питання не розглядається детально в цій роботі. Ця тема є актуальною для подальших досліджень.

Розроблений алгоритм призначений для управління великими сховищами зображень за умови, що зображення здатні зазнавати різних трансформацій: можуть мати різні формати файлів, скориговану яскравість, контрастність, підвищену різкість, виконане квантування, зображення може бути стиснуте або масштабоване. Водночас, це дослідження спрямоване на найпоширеніші трансформації у великих сховищах зображень; повороти, дзеркальне відтворення, додавання, модифікація та вилучення окремих об'єктів, а також вбудовування іншого зображення в оригінальне на цьому етапі не розглядаються.

2.6. Метрика якості пошуку

Якість пошуку (q) визначається насамперед порядком, у якому зображення, відсортовані за зростанням відстані від оригіналу, повертаються зі сховища.

Припустимо, що сховище містить $n1$ схожих зображень (зокрема оригінальне зображення та його трансформації). Нехай *max* – порядковий номер останнього зображення цієї групи у відсортованому результаті пошукового запиту.

Далі знайдемо кількість $n2$ зображень у сховищі, що не належать до розглянутої групи зображень, але передують зображенню з номером *max* у порядку. Наявність таких зображень вказує на те, що трансформовані зображення зрештою не утворюють суцільного списку. Серед них також є нетипові, що свідчить про зниження якості пошуку. Тому оцінюємо якість пошуку таким чином:

$$q = 1 - \frac{n2}{n1 + n2}, \quad (5)$$

де q – якість пошуку; $n1$ – кількість трансформацій шуканого зображення; $n2$ – кількість зображень, що не належать до трансформацій шуканого зображення, проте мають меншу відстань під час пошуку.

Коли $n2 = 0$, перші $n1$ зображень, як наслідок, належать до бажаної групи зображень та йдуть одне за одним без проміжків. Ця ситуація вважається ідеальною $(q = 1)$.

Після опису розроблених моделей дескрипторів зображень, метрик, моделей пошуку та нейронної мережі, яка використовується для побудови дескрипторів із вихідних зображень, тепер опишемо програмне забезпечення, використане для проведення експериментів.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 2 (32)

2.7. Опис програмного забезпечення

Для реалізації запропонованих методів пошуку зображень за допомогою дескрипторів розроблено застосунки з простим вебінтерфейсом. Ці застосунки полегшують проведення експериментів і наочно демонструють результати запитів. Розроблене програмне забезпечення виконує кілька ключових функцій.

1. Побудова дескрипторів. Створює дескриптори для окремих зображень або цілих колекцій. Будує та керує сховищем зображень.

2. Інтерфейс сховища. Надає зручний інтерфейс для перегляду всіх зображень у сховищі. Дає змогу фільтрувати зображення за тегами й шукати схожі зображення в межах сховища.

3. Детекція об’єктів. Виявляє об’єкти на заданому зображенні та візуально подає виявлені об’єкти (наприклад, обмежувальні рамки).

4. Перегляд зображень. Дає змогу користувачам переглядати будь-яке зображення в сховищі. Відтворює колекцію виявлених об’єктів та інтерактивно виділяє їх на зображенні.

5. Трансформація зображень. Генерує дев’ять заданих трансформацій оригінального зображення, таких як редагування кольору, стиснення, масштабування тощо. Ці трансформації використовуються в експериментах щодо стійкості й точності пошуку зображень для подальших досліджень.

На рис. 2 показано інтерфейс розробленого застосунку, який використовується для проведення експериментів.

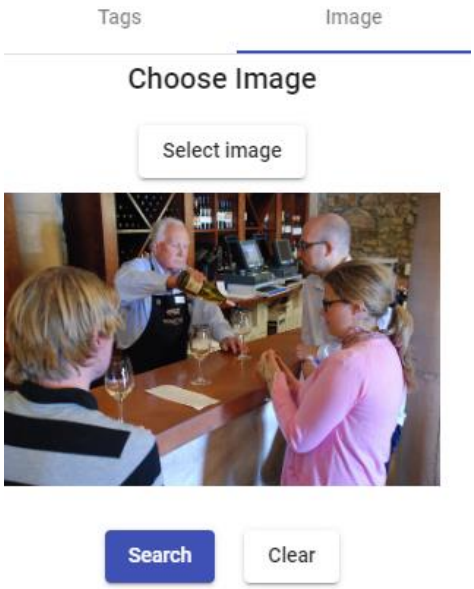

Рис. 2. Інтерфейс застосунку для експериментів

Для тестування запропонованого алгоритму пошуку за зображенням та програмного забезпечення було розроблено деякі компоненти. Опишемо їх.

Сервіс YOLO. Реалізований на Python, цей сервіс абстрагує взаємодію з моделлю YOLO. Приймає шлях до зображення або набору зображень і генерує дескриптори, повертаючи їх безпосередньо або зберігаючи в базі даних. Використовує прискорення за допомогою GPU (якщо доступне) для прискорення обчислень.

База даних. В експериментальній установці використовується MongoDB, обрана за свою гнучкість як NoSQL база даних. Під час досліджень і розроблення моделі дескрипторів часто розширюються та модифікуються, і MongoDB дає змогу вносити зміни без постійної міграції даних. MongoDB також підтримує масштабованість за допомогою таких технік, як шардування, що робить її придатною для роботи з великими сховищами даних зображень. Побудова гнучкого рішення для бази даних є критично важливою для передбачення та розв’язання проблем, що виникають у роботі зі сховищами зображень великого масштабу.

Сервіс Backend. Постійно оновлюється для підтримки актуальності моделей дескрипторів. Відповідає за виконання операцій пошуку та порівняння, обробляє запити користувачів на детекцію об’єктів у зображеннях та додавання фотоархівів до бази даних дескрипторів. Реалізований на .NET, він зберігає дескриптори в пам’яті для підвищення продуктивності та використовує ImageMagick для трансформації зображень.

Сервіс UI. Легкий фронтенд, розроблений на Angular для демонстраційних цілей. Дає змогу користувачам візуалізувати можливості системи, додавати фотографії до сховища, будувати дескриптори для зображень у певних директоріях і виконувати пошукові запити (наприклад, за тегом або схожістю до запитуваного зображення).

Крім того, був створений простий набір інструментів для генерації трансформованих версій вхідних зображень (наприклад, зміни масштабу, стиснення, гами тощо). Ці трансформації використовуються для тестування стійкості й точності методу пошуку зображень у великих сховищах.

2.8. Опис даних

Для експериментів використовуємо зображення з набору даних COCO 2017. Він містить понад

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 2 (32) ISSN 2524-2296 (online)

163 000 зображень із анотаціями для 80 класів об'єктів. Така значна кількість зображень дає змогу оцінити швидкість пошуку в сховищі із сотнями тисяч екземплярів.

Повна назва – *Common Objects in Context* (Поширені об'єкти в контексті) – вказує на характер зображених об'єктів, переважно повсякденних фотографій, зокрема людей, природи, міських пейзажей, спортивних подій тощо. Серед розпізнаваних класів об'єктів – люди, автомобілі, тварини, пристрої, предмети й аксесуари. Окрім оригінальних зображень з набору даних COCO 2017, додаємо трансформовані копії тестових зображень, які будуть використовуватися для пошуку.

Повний список трансформацій оригінальних зображень передбачає: зменшення яскравості на 20%, стиснення до 70%, підвищення контрастності, збільшення розміру фотографії на 20%, зменшення розміру фотографії на 20%, перетворення в градації сірого, квантування до 128 кольорів, збільшення насиченості на 30% і підвищення різкості зображення. Разом з оригінальним зображенням очікуємо, що оригінал і трансформації з'являться в топових результатах пошуку для тестового зображення.

2.9. Опис обладнання

Експерименти проводилися на робочій станції з процесором Intel(R) Core(TM) i5-9300H, що працює на частоті 2,4 ГГц, 24 ГБ оперативної пам'яті, 128 ГБ SSD та 1 ТБ HDD. Робоча станція також має графічну карту NVIDIA GeForce GTX 1650 з 4 ГБ відеопам'яті GDDR5, побудовану на архітектурі Turing. Ця графічна карта має 896 CUDA ядер, які використовуються для машинного навчання. Примітиви CUDA підтримують роботу з даними на GPU і використовуються для прискорення детекції об'єктів моделлю YOLOv8.

2.10. Побудова дескрипторів та їх репозиторію

Сховище дескрипторів було побудовано способом зберігання набору даних на робочій станції та оброблення їх створеним програмним забезпеченням. Програмне забезпечення отримувало назви директорій, де зберігалися зображення, потім обробляло всі зображення в цих директоріях, виконувало детекцію об'єктів на кожному зображенні, перетворювало дані про об'єкти та зображення в описану модель дескриптора та зберігало отримані дескриптори в базу даних. Набір даних містить приблизно 164 000 зображень, і побудова дескрипторів для всіх зображень зайняла близько п'яти з половиною годин. Це передбачало детекцію об'єктів, побудову дескрипторів та їх збереження в базу даних.

2.11. Підготовка експериментів

Експерименти зосереджені на швидкості виконання, використанні ресурсів та якості пошуку зображень. Зокрема нас цікавить наявність трансформованих зображень у топових результатах, повернених нашою системою.

Тест пошуку зображень на основі дескрипторів

Виконуємо пошук для 10 тестових зображень, вимірюємо час виконання та якість пошуку результатів, а також відзначаємо кількість оригінальних і трансформованих зображень із цільової групи, які безперервно з'являються у верхній частині вибірки. Це надає практичні результати для розроблених дескрипторів, моделей пошуку та метрик.

Порівняння з перцептивними гешами

Проводиться пошук зображень за допомогою відомого алгоритму для порівняння з тим, що подано в цій роботі. Перцептивні геші використовуються для цього порівняння, оскільки методи, що застосовують ключові точки, є значно більш часозатратними. Геші попередньо обчислюються для всіх зображень у сховищі, і ми обчислюємо відстань як різницю суми квадратів між гешами зображень, сортуємо результати за зростанням і аналогічно вимірюємо час виконання та якість пошуку.

Нарешті, порівнюються результати, досягнуті за допомогою вилучення ознак із зображень набору даних, з тими, що згенеровані сучасною моделлю Vision Transformer, яку запропонували наприкінці 2020 р. А. Досовіцький, Л. Бейєр, О. Колесніков та ін. [34]. Для цього обчислюємо вектори ознак для всіх зображень у наборі даних і зберігаємо їх у базі даних. Під час пошуку зображень обчислюватимемо вектори ознак для цільового зображення та порівнюватимемо їх із тими, що зберігаються в базі даних, використовуючи пошук за косинусною схожістю. Це порівняння дасть змогу оцінити запропонований метод порівняно із сучасними моделями в аналогічних умовах. Наш основний фокус буде спрямований на якість пошуку порівняно із запропонованим методом, а також на час виконання.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 2 (32)

Обирається 10 тестових зображень, що містять різні конфігурації об'єктів (один об'єкт, група об'єктів одного класу, група об'єктів різних класів, малі та великі об'єкти тощо), для яких додаємо дев'ять трансформацій до сховища. Ці трансформації передбачають: зменшення яскравості на 20%, стиснення до 70%, підвищення контрастності, збільшення розміру фотографії на 20%, зменшення розміру фотографії на 20%, перетворення в градації сірого, квантування до 128 кольорів, збільшення насиченості на 30% та підвищення різкості зображення. Обираються зображення, що містять різні класи та кількості об'єктів, а також різні конфігурації та розташування об'єктів у межах зображень. Вимірюється час виконання для пошукових запитів із тестовими зображеннями та якість пошуку. Для тестів із гешами також фіксуємо кількість зображень, які випадають з топ-100 результатів пошуку. Мета цього експерименту – визначити ефективність розробленого методу, порівняти його результативність з наявними аналогами та виявити крайні ситуації, коли цей метод може бути неефективним для пошуку трансформованих зображень.

3. Результати та обговорення

3.1. Пошук за зображенням

У наступному експерименті тестові зображення разом з їх трансформаціями шукаються, в ідеалі повертаючи топ-10 результатів пошуку. Пошук передбачає виявлення об'єктів на зображенні та побудову дескрипторів; потім зображення без жодного перетину в складі об'єктів фільтруються, а решта зображень порівнюються за допомогою розробленого алгоритму. Кілька скриншотів результатів пошуку наведено на рис. 3–5. Усі скриншоти досягнутих результатів пошуку за допомогою дескрипторів та перцептивних гешів, а також скриншоти розробленого програмного забезпечення можна знайти в нашому репозиторії: https://github.com/alex-prokopenko-nure/image-search-results

Рис. 3. Результати пошуку для тестового зображення № 3

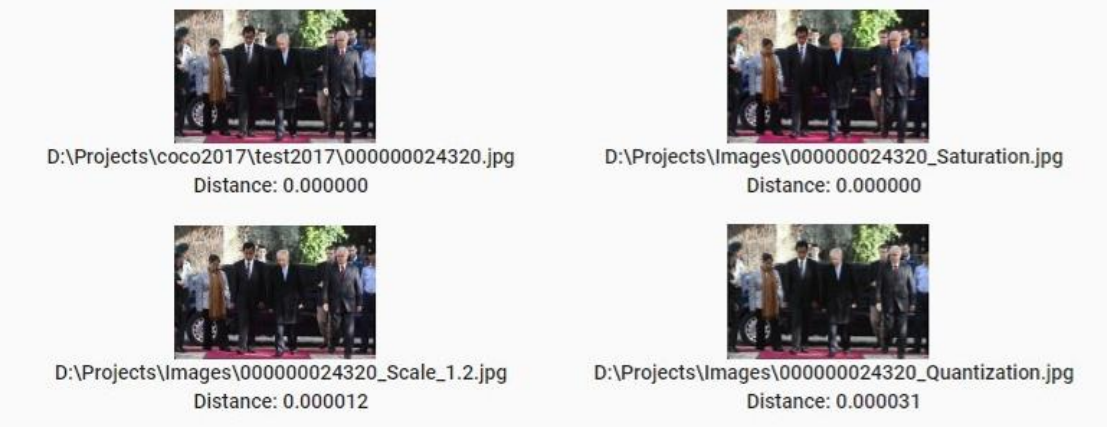

Рис. 4. Результати пошуку для тестового зображення № 6

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 2 (32)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Рис. 5. Результати пошуку для тестового зображення № 10

У табл. 4 узагальнено результати експерименту. Як і очікувалося, для всіх 10 тестових зображень оригінальне зображення отримано як перший результат, і всі трансформації містяться в межах топ-15 результатів для всіх протестованих зображень. Для 7 із 10 тестових зображень досягнуто ідеальну якість пошуку, крім цього, оригінальне зображення та всі трансформації з'являються в топ-10 результатах пошуку. Найгірший результат спостерігається для зображення № 6, з якістю пошуку 76,92%. Три сторонніх зображення знайдено перед деякими трансформаціями, до того ж сторонні зображення знайдено на індексах 9, 11 та 12. Середня якість пошуку серед 10 тестових зображень становить 95,12 %.

Таблиця 4. *Результати пошуку за зображенням з розробленими дескрипторами*

	Час пошуку	**Максимальний індекс трансформованого зображення, *max***	**Якість пошуку, *q***	**Кількість трансформованих зображень підряд у топі вибірки**
1	144 мс	10	1	10
2	122 мс	12	0.8333	9
3	123 мс	10	1	10
4	130 мс	10	1	10
5	154 мс	10	1	10
6	132 мс	13	0.7692	8
7	132 мс	10	1	10
8	131 мс	11	0.9091	9
9	155 мс	10	1	10
10	142 мс	10	1	10

Під час експерименту з пошуку зображень спостерігається, що обчислення метрики відстані для всіх зображень у сховищі за заданих умов є швидшим, ніж попередній метод фільтрації за складом об'єктів з подальшим обчисленням відстані для меншого набору зображень, оскільки копіювання даних у цьому разі займає більше часу, ніж обчислення відстаней для решти зображень. Нетипові зображення іноді заважають результатам пошуку, випереджаючи деякі трансформації, через те, що модель виявляє інший склад об'єктів порівняно з оригінальним зображенням. Найбільша різниця спостерігається, коли в трансформованому зображенні не виявлено жодного об'єкта класу, який був виявлений в оригінальному зображенні. Іноді відстань до нетипового зображення з подібним складом об'єктів є меншою, ніж відстань до трансформованого зображення з неповним набором об'єктів у таких випадках.

3.2. Пошук за зображенням із використанням перцептивних гешів

У цьому експерименті ті самі тестові зображення з їх трансформаціями, що й у другому експерименті, шукаються, але з використанням перцептивних гешів зображень, і для порівняння обчислюється сума квадратів різниць між гешами. Досягнуті результати для порівняння із застосуванням гешів подано в табл. 5.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 2 (32)

Таблиця 5. *Результати пошуку за зображенням із використанням перцептивних гешів*

	Час пошуку	Максимальний індекс трансформованого зображення, *max*	Якість пошуку, *q*	Кількість трансформованих зображень підряд у топі вибірки
1	214 мс	>1000	<0.01	8
2	221 мс	>1000	<0.01	7
3	230 мс	>1000	<0.01	7
4	242 мс	>1000	<0.01	7
5	215 мс	>1000	<0.01	6
6	226 мс	>1000	<0.01	6
7	223 мс	>1000	<0.01	7
8	252 мс	>1000	<0.01	7
9	235 мс	>1000	<0.01	8
10	228 мс	>1000	<0.01	7

Пошук на основі гешів відбувається за порівняний час. Однак, попри те, що 6–8 трансформацій містяться в топових результатах пошуку, трансформація в градації сірого для всіх тестових зображень опиняється за межами топ-1000 результатів. Крім того, для 8 із 10 тестових зображень ще 1–2 зображення опиняються за межами топ-1000 вибірки через маніпуляції з кольоровим охопленням. Це робить метод на основі гешів непрактичним для пошуку змін кольорового охоплення в зображеннях, навіть за умови перетворення зображень у градації сірого перед отриманням гешу. Це контрастує з методом на основі дескрипторів, де ті самі трансформації успішно з'являлися в топових результатах.

3.3. Пошук за зображенням з використанням дескрипторів отриманих з Vision Transformer

У фінальному експерименті порівнюємо тестові зображення разом з їх попередньо створеними трансформаціями за допомогою вилучених ознак із моделі Vision Transformer. Ці ознаки порівнюються за допомогою косинусної схожості між векторами. У табл. 6 подано результати пошуку зображень за допомогою Vision Transformer.

Таблиця 6. *Результати пошуку за зображенням із використанням дескрипторів, отриманих з Vision Transformer*

	Час пошуку	Максимальний індекс трансформованого зображення, *max*	Якість пошуку, *q*	Кількість трансформованих зображень підряд у топі вибірки
1	69.7 с	10	1	10
2	62.6 с	>20	<0.5	9
3	63.2 с	18	0.5556	9
4	66.4 с	>20	<0.5	9
5	58.1 с	15	0.6667	9
6	68 с	12	0.8333	8
7	61.7 с	10	1	10
8	64.9 с	11	0.9091	9
9	62.7 с	10	1	10
10	62.8 с	>20	<0.5	9

Пошук за допомогою косинусної схожості векторів ознак, отриманих за допомогою моделі Vision Transformer, є значно довшим порівняно з іншими використаними методами. Подібно до експерименту з перцептивними гешами, трансформація у градації сірого часто випадає з топових результатів пошуку, опиняючись за межами топ-20 результатів у трьох випадках. Однак інші трансформації переважно містяться в межах першої десятки. Якщо вимкнути трансформацію в градації сірого, якість пошуку для деяких зображень є вищою, ніж у запропонованого методу, але вимагає більш тривалого часу.

Пошук способом порівняння векторів ознак зображень з подальшим уточненням або поєднанням з іншими методами може бути використаний для більш точного порівняння зображень у попередньо відфільтрованій вибірці із сотень або тисяч зображень, спочатку відфільтрованих за допомогою нашого запропонованого методу.

3.4. Обговорення

Досягнуті результати свідчать про те, що розроблений алгоритм успішно досягає поставлених

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 2 (32) ISSN 2524-2296 (online)

цілей, а саме пошуку схожих зображень у великих сховищах даних. Несподівано розроблені моделі та алгоритми дають змогу порівнювати зображення набагато швидше, ніж очікувалося, використовуючи базу даних дескрипторів. Простота обчислень, закладених у алгоритм, призводить до швидкого розрахунку відстаней між зображеннями, коли накладні витрати на копіювання даних займають більше часу, ніж фактичні обчислення відстаней. Якість пошуку для всіх тестових зображень перевищує 75% у разі пошуку зображень за допомогою дескрипторів, і ми обговоримо причини та можливі способи подолання цього далі.

По-перше, обговоримо час виконання пошуку, який у середньому становить близько 150 мс для понад 164 тисяч зображень. Це свідчить про те, що розроблений алгоритм може комфортно застосовуватися для великих сховищ із мільйонами зображень і забезпечувати результати пошуку в межах прийнятного часу для користувача. Паралельні та розподілені обчислення є перспективним напрямом для подальшого скорочення часу виконання.

По-друге, обговоримо якість пошуку зображень для всіх алгоритмів. Розроблений алгоритм повертає всі 10 трансформованих зображень як найкращі 15 результатів для всіх 10 тестових зображень. Сім із 10 тестових зображень повертають всі 10 трансформацій як найкращі результати без будь-якої домішки зображень, не пов'язаних із оригіналом. Навіть у разі, коли не всі трансформовані зображення потрапляють у топ-10 результатів, вони всі наявні серед 15 найближчих зображень, що дає змогу порівнювати з іншими методами з-поміж відфільтрованих топових зображень для досягнення більш точних кінцевих результатів.

Нижча якість пошуку для деяких зображень пов'язана з чутливістю алгоритму до розбіжностей у складі об'єктів зображень. Якщо зображення містить кілька об'єктів певного класу, і один з них втрачається, це зазвичай не впливає на пошук. Ця ситуація спостерігається, зокрема, у зображенні № 9, що має високу якість пошуку, незважаючи на це. Однак проблема виникає, коли втрачаються кілька об'єктів, що належать до різних класів, порівняно з оригіналом. Тоді зображення отримує високий штраф за відстань за кількома класами об'єктів, і іноді воно отримує вищий бал відстані, ніж ті зображення, що не пов'язані з оригіналом, але мають подібний склад об'єктів. Зокрема на деяких трансформаціях зображення № 8 і № 10 втрачають класи під час детекції, а зображення № 2 містить клас об'єкта, що не був знайдений в оригінальному зображенні.

Найпоширенішою причиною того, що трансформоване зображення не потрапляє до топового списку під час експерименту, є проблема, пов'язана з роботою моделі детекції об'єктів. Зміни, виконані для трансформованих зображень, переважно незначні, але навіть вони можуть призвести до того, що малі об'єкти в оригінальному зображенні пропускаються моделлю в трансформованому зображенні, або, навпаки, об'єкти, не знайдені в оригіналі, з'являються в трансформованому.

Порівнюючи результати розробленого алгоритму пошуку зображень за допомогою дескрипторів з пошуком зображень за допомогою перцептивних гешів, а також із порівнянням векторів ознак, отриманих за допомогою моделі Vision Transformer, спостерігаємо, що метод, поданий у цій роботі, є порівняним за швидкістю (на 30–40% швидшим за перцептивні геші в проведених тестах і на порядок швидшим за порівняння векторів ознак). Крім того, він більш ефективний для пошуку зображень, що можуть містити поширені трансформації, зокрема зміни кольорового охоплення. Якість пошуку на основі запропонованої метрики в середньому перевищує 95%, тоді як якість пошуку на основі порівняння векторів ознак зазвичай нижча за 75%, а за умови використання відстані перцептивних гешів без додаткових модифікацій – менше ніж 1%. Це пов'язано насамперед з тим, що за певних трансформацій кольорового охоплення нейронна мережа може не виявити конкретні об'єкти з оригінального зображення. Однак ці об'єкти зазвичай малі й мають незначний вплив на кінцеву відстань, тому навіть у разі втрати об'єктів трансформації все ще розташовані близько до топових результатів пошуку з найгіршою продуктивністю під час тестів понад 75%.

Порівнюючи із сучасною моделлю, часто спостерігаємо, що внаслідок обчислення косинусної схожості між отриманими векторами ознак трансформації в градації сірого часто випадають з бажаних топ-10 результатів, іноді навіть за межі топ-20, вимагаючи в сотні разів більше часу на обчислення під час експериментів. Це свідчить про те, що запропонований метод може мати значні переваги в практичному використанні. Водночас, якщо зміни кольорового охоплення є суттєвими,

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 2 (32)

перцептивний геш дуже змінюється, і трансформоване зображення навіть не потрапляє до топ-1000 результатів пошуку. Це еквівалентно тому, що трансформація взагалі не виявлена.

Третій пункт – обговорення потенційних напрямів досліджень. Одним з них є поєднання з іншими методами та алгоритмами пошуку. Оскільки швидкість алгоритму висока, інші методи порівняння можна застосувати до топ-100 групи зображень, де їх можна порівнювати, наприклад, на основі спеціальних точок або гешів, згаданих на початку статті, для уточнення об'єктно-орієнтованого пошуку. Також у відфільтрованій групі можна більш точно порівняти розташування об'єктів або навіть інші дескриптори цих об'єктів, щоб отримати кінцевий результат.

Серед модифікацій алгоритму можна розглянути зміну розрахункового центру групи об'єктів із середнього арифметичного центрів на певний "центр мас", де центр буде прагнути до більших об'єктів, що робить втрату менших об'єктів менш помітною за умови трансформацій. Також серед параметрів, що присутні зараз у дескрипторі зображення, але фактично не використовуються для обчислення відстані між зображеннями, є пропорції окремих об'єктів і впевненість моделі в тому, що прямокутник належить об'єкту певного класу.

На нашу думку, обчислення деяких усереднених показників пропорцій об'єктів у групі та середньої впевненості в об'єктах цього класу не є сильним показником і може бути приблизно однаковим для абсолютно різних наборів об'єктів на основі інших параметрів. Більш того, спостережувана ситуація, коли об'єкти "випадають" під час трансформацій, може суттєво вплинути на показники та збільшити відстань між трансформованими зображеннями, що є небажаним для нас. Вважаємо згадані невикористані показники перспективними, особливо для ситуацій порівняння зображень на основі об'єктів з певного невеликого топу, коли вони є сильним показником того, чи є два зображення близькими чи далекими одне від одного.

Хотіли б наголосити, що якість розробленого дескриптора зображення та методу порівняння повністю залежить від якості даних, отриманих із нейронної мережі. У разі, коли об'єкти на зображенні не знайдені, це зображення ніколи не обирається. Тому дуже важливо ретельно обирати модель для пошуку зображень, можливо, налаштовуючи її за допомогою трансферного навчання, знаючи особливість сховища зображень, з яким потрібно працювати в межах практичного завдання.

Логічним напрямом подальшого розвитку побудованої системи є використання гібридних моделей з одночасним використанням різних дескрипторів, що сприяє покращенню якості пошуку та усуненню недоліків конкретних моделей.

Нарешті, на основі результатів експериментів варто зазначити, що розроблений дескриптор зображення та метод порівняння успішно виконують завдання пошуку схожих зображень, маючи високу швидкість порівняння та спираючись не на формальні параметри зображень, а на зображені об'єкти. Крім упровадження розробленого методу пошуку як автономного алгоритму, його також можна використовувати для попередньої фільтрації, звуження ділянки пошуку та формування вибірки зображень, до яких можна застосувати більш складні та обчислювально витратні методи порівняння.

4. Висновки

У цій роботі досліджено проблему компромісу між швидкістю та точністю в дескрипторах зображень, які використовуються для пошуку зображень на основі контенту у великих сховищах даних.

Для розв'язання окресленої проблеми запропоновано об'єктно-орієнтований алгоритм порівняння дескрипторів зображень, що використовує дані, отримані моделями детекції об'єктів, для створення ефективного та збалансованого рішення для пошуку зображень на основі контенту, значно покращуючи точність порівнянь, отриманих за допомогою традиційних дескрипторів, і залишаючись у цьому разі більш швидким за високовимірні вектори ознак, що використовуються в дескрипторах на основі глибокого навчання. Для розроблення об'єктно-орієнтованого алгоритму було виконано кілька завдань, зокрема: визначення полів даних для зберігання об'єктів, розроблення дизайну сховища для дескриптора, вибір відповідної моделі детекції об'єктів, вибір набору даних, розроблення метрик і алгоритмів для порівняння запропонованих дескрипторів та створення програмного забезпечення з метою реалізації порівняння дескрипторів у сховищі зображень.

Для перевірки розробленого алгоритму порівняння зображень і застосунків, призначених

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 2 (32)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

для управління великими сховищами зображень, було заплановано та проведено серію експериментів. Вони тестували різні аспекти розробленої системи за допомогою набору даних, призначеного для детекції об'єктів, що містить понад 160 000 зображень. Було побудовано базу даних із дескрипторами для всіх зображень у наборі даних. Результати розробленого алгоритму пошуку потім було порівняно з результатами, досягнутими за допомогою сучасної моделі Vision Transformer та відомого алгоритму пошуку зображень за допомогою перцептивних гешів.

Результати експериментів показали, що розроблений дескриптор і алгоритм порівняння чудово впоралися з поставленими завданнями. Основні критерії, що тестувалися: витрачений час, застосування пам'яті й точність, що необхідно для висновків про доцільність використання дескриптора й алгоритмів для виконання завдань пошуку зображень на основі змісту.

Розроблений алгоритм виявився достатньо точним для пошуку схожих трансформованих зображень, оскільки майже всі копії розміщені серед топових результатів пошуку для тестового зображення. Навіть коли сторонні зображення були ближчими до оригіналу, всі 10 трансформованих зображень усе ще були в межах топ-15 результатів. Експерименти продемонстрували переваги розробленого алгоритму пошуку порівняно з аналогічним пошуком за допомогою моделі Vision Transformer, особливо щодо якості пошуку й значно вищої швидкості виконання. Це наголошує на перевагах розробленого алгоритму, на відміну від сучасної моделі, оскільки він пропонує достатню швидкість пошуку у великих сховищах із збереженням високої якості пошуку. Порівняно з пошуком на основі перцептивних гешів, розроблений алгоритм також пропонує значно кращу якість пошуку з порівняною швидкістю виконання.

Окрім позитивних результатів, досягнутих унаслідок експериментів, було зібрано достатньо велику кількість інформації для роздумів і виявлено значну кількість перспективних напрямів досліджень, які плануємо розвивати в найближчому майбутньому.

Деякі з цих напрямів передбачають застосування розробленого алгоритму для пошуку зображень із проведенням порівняння об'єктів для більш точного ранжування близьких результатів, посилене використання паралельних і розподілених обчислень для ще кращого масштабування запропонованої системи до значно більших обсягів даних, а також ідеї щодо модифікацій та покращень, якими необхідно доповнити розроблені методи й модулі системи для тонкого налаштування під конкретні потреби окремих сховищ і користувачів, що дасть змогу застосовувати її не лише в контексті зображень загального призначення, але й для більш точних наукових сфер.

Об'єктно-орієнтований дескриптор балансує ефективність і точність, використовуючи переваги наявних методів для створення надійного й масштабованого рішення для завдань пошуку зображень.

Список літератури

1. Gonzalez R. C., Woods R. E. Digital Image Processing: monograph. 4th ed., New York, 2018. 1168 p. ISBN: 9780133356724
2. Yang W., Zhao H., Wang M. Design of Intelligent Search Engine Service Performance Evaluation System. 2020 5th Asia-Pacific *Conference on Intelligent Robot Systems (ACIRS)*. Singapore, 2020. P. 86–91. DOI: https://doi.org/10.1109/ACIRS49895.2020.9162611
3. Amorós F., Payá L., Mayol-Cuevas W. Holistic Descriptors of Omnidirectional Color Images and Their Performance in Estimation of Position and Orientation. *IEEE Access*. 2020. Vol. 8. P. 81822–81848. DOI: https://doi.org/10.1109/ACCESS.2020.2990996
4. Liu X., Cheung G., Lin C.-W. Prior-Based Quantization Bin Matching for Cloud Storage of JPEG Images. *IEEE Transactions on Image Processing*. 2018. Vol. 27. № 7. H. 3222–3235. DOI: https://doi.org/10.1109/TIP.2018.2799704
5. Carvalho E. D., Filho A. O. C., Silva R. R. V. Breast Cancer Diagnosis from Histopathological Images Using Textural Features and CBIR. *Artificial Intelligence in Medicine*. 2020. Vol. 105. P. 101845. DOI: https://doi.org/10.1016/j.artmed.2020.101845
6. Nakazato M., Huang T. S. 3D MARS: Immersive Virtual Reality for Content-Based Image Retrieval. *IEEE International Conference on Multimedia and Expo (ICME)*. Tokyo, Japan, 2001. P. 44–47. DOI: https://doi.org/10.1109/ICME.2001.1237651
7. Iqbal K., Odetayo M. O., James A. Content-Based Image Retrieval Approach for Biometric Security Using Colour, Texture and Shape Features Controlled by Fuzzy Heuristics. *Journal of Computer and System Sciences*. 2012. Vol. 78. № 4. P. 1258–1277. DOI: https://doi.org/10.1016/j.jcss.2011.10.013

8. Popescu A., Grefenstette G. Social Media Driven Image Retrieval. *Proceedings of the 1st ACM International Conference on Multimedia Retrieval (ICMR '11)*. New York, NY, USA, 2011. P. 1–8. DOI: https://doi.org/10.1145/1991996.1992029
9. Liu Y., Mei T., Hua X.-S. CrowdReranking: Exploring Multiple Search Engines for Visual Search Reranking. *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '09)*. New York, NY, USA, 2009. P. 500–507. DOI: https://doi.org/10.1145/1571941.1572027
10. Staszewski P., Jaworski M., Cao J. A New Approach to Descriptors Generation for Image Retrieval by Analyzing Activations of Deep Neural Network Layers. *IEEE Transactions on Neural Networks and Learning Systems*. 2022. Vol. 33. № 12. P. 7913–7920. DOI: https://doi.org/10.1109/TNNLS.2021.3084633
11. Sahmoudi Y., El-Ogri O., El-Mekkaoui J. An Efficient Biomedical Color Image Retrieval System Based on Continuous Orthogonal Legendre Fourier Quaternion. 2024 Sixth *International Conference on Intelligent Computing in Data Sciences (ICDS)*. Marrakech, Morocco, 2024. P. 1–6. DOI: https://doi.org/10.1109/ICDS62089.2024.10756406
12. Bano M., Matta P., Chandel S. Content Based Image Retrieval: A Study of Approaches and Techniques. 2024 4th *International Conference on Technological Advancements in Computational Sciences (ICTACS)*. Tashkent, Uzbekistan, 2024. P. 16–22. DOI: https://doi.org/10.1109/ICTACS62700.2024.10840489
13. Anand A., Saxena A., Singh K. Statistical Features Based Content Based Image Retrieval Using Machine Learning Classifiers. 2024 *IEEE 3rd World Conference on Applied Intelligence and Computing (AIC)*. Gwalior, India, 2024. P. 1102–1109. DOI: https://doi.org/10.1109/AIC61668.2024.10731120
14. Debin H., Yue Z., Shuai J. Application of Content-Based Retrieval Technology in Image Archive Management. 2024 *Global Conference on Communications and Information Technologies (GCCIT)*. Bangalore, India, 2024. P. 1–6. DOI: https://doi.org/10.1109/GCCIT63234.2024.10862277
15. Bai J., Ni B., Wang M. Deep Progressive Hashing for Image Retrieval. *IEEE Transactions on Multimedia*. 2019. Vol. 21. № 12. P. 3178–3193. DOI: https://doi.org/10.1109/TMM.2019.2920601
16. Lowe D. G. Object Recognition from Local Scale-Invariant Features. *Proceedings of the Seventh IEEE International Conference on Computer Vision*. Kerkyra, Greece, 1999. Vol. 2. P. 1150–1157. DOI: https://doi.org/10.1109/ICCV.1999.790410
17. Bay H., Ess A., Tuytelaars T. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*. 2008. Vol. 110. № 3. P. 346–359. DOI: https://doi.org/10.1016/j.cviu.2007.09.014
18. Mikolajczyk K., Schmid C. A Performance Evaluation of Local Descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2005. Vol. 27. № 10. P. 1615–1630. DOI: https://doi.org/10.1109/TPAMI.2005.188
19. Ke Y., Sukthankar R. PCA-SIFT: A More Distinctive Representation for Local Image Descriptors. Proceedings of the 2004 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. Washington, DC, USA, 2004. Vol. 2. P. II–II. DOI: https://doi.org/10.1109/CVPR.2004.1315206
20. Calonder M., Lepetit V., Strecha C. BRIEF: Binary Robust Independent Elementary Features. *Computer Vision – ECCV 2010. Lecture Notes in Computer Science*. 2010. Vol. 6314. P. 778–792. DOI: https://doi.org/10.1007/978-3-642-15561-1_56
21. Rublee E., Rabaud V., Konolige K. ORB: An Efficient Alternative to SIFT or SURF. 2011 *International Conference on Computer Vision*. Barcelona, Spain, 2011. P. 2564–2571. DOI: https://doi.org/10.1109/ICCV.2011.6126544
22. Leutenegger S., Chli M., Siegwart R. Y. BRISK: Binary Robust Invariant Scalable Keypoints. 2011 *International Conference on Computer Vision*. Barcelona, Spain, 2011. P. 2548–2555. DOI: https://doi.org/10.1109/ICCV.2011.6126542
23. Alahi A., Ortiz R., Vandergheynst P. FREAK: Fast Retina Keypoint. 2012 *IEEE Conference on Computer Vision and Pattern Recognition*. *Providence*, RI, USA, 2012. P. 510–517. DOI: https://doi.org/10.1109/CVPR.2012.6247715
24. Žižakić N., Pižurica A. Efficient Local Image Descriptors Learned with Autoencoders. *IEEE Access*. 2022. Vol.10. P. 221–235. DOI: https://doi.org/10.1109/ACCESS.2021.3138168
25. Liu Y., Xu X., Li F. Image Feature Matching Based on Deep Learning. 2018 *IEEE 4th International Conference on Computer and Communications (ICCC)*. Chengdu, China, 2018. P. 1752–1756. DOI: https://doi.org/10.1109/CompComm.2018.8780936
26. Radenović F., Tolias G., Chum O. Fine-Tuning CNN Image Retrieval with No Human Annotation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2019. Vol. 41. № 7. P. 1655–1668. DOI: https://doi.org/10.1109/TPAMI.2018.2846566
27. Song L., Lin J., Wang Z. J. An End-to-End Multi-Task Deep Learning Framework for Skin Lesion Analysis. *IEEE Journal of Biomedical and Health Informatics*. 2020. Vol. 24. № 10. P. 2912–2921. DOI: https://doi.org/10.1109/JBHI.2020.2973614
28. Wang B., Zhang H., Zhu L. Multi-Level Adversarial Attention Cross-Modal Hashing. *Signal Processing: Image Communication*. 2023. Vol. 117. 117017 p. DOI: https://doi.org/10.1016/j.image.2023.117017
29. Gajjar V., Khandhediya Y., Gurnani A. Human Detection and Tracking for Video Surveillance: A Cognitive Science Approach. 2017 *IEEE International Conference on Computer Vision Workshops (ICCVW)*. Venice, Italy, 2017. P. 2805–2809. DOI: https://doi.org/10.1109/ICCVW.2017.330
30. Adel M., Moussaoui A., Rasigni M. Statistical-Based Tracking Technique for Linear Structures Detection: Application to Vessel Segmentation in Medical Images. *IEEE Signal Processing Letters*. 2010. Vol. 17. № 6. P. 555–558. DOI: https://doi.org/10.1109/LSP.2010.2046697

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 2 (32)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

31. Truong X.-T., Yoong V. N., Ngo T.-D. RGB-D and Laser Data Fusion-Based Human Detection and Tracking for Socially Aware Robot Navigation Framework. 2015 *IEEE International Conference on Robotics and Biomimetics (ROBIO).* Zhuhai, China, 2015. P. 608–613. DOI: https://doi.org/10.1109/ROBIO.2015.7418835
32. Galvez R. L., Bandala A. A., Dadios E. P. Object Detection Using Convolutional Neural Networks. *TENCON 2018 – 2018 IEEE Region 10 Conference*. Jeju, Korea (South), 2018. P. 2023–2027. DOI: https://doi.org/10.1109/TENCON.2018.8650517
33. Wehbe A., Hotiet H., Minetti I. Integrating YOLO for Advanced Content-Based Image Retrieval in Lung Cancer Imaging. 2024 31st *IEEE International Conference on Electronics, Circuits and Systems (ICECS)*. Nancy, France, 2024. P. 1–4. DOI: https://doi.org/10.1109/ICECS61496.2024.10848862
34. Dosovitskiy A., Beyer L., Kolesnikov A. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale / A. Dosovitskiy та ін. *International Conference on Learning Representations.* 2021. DOI: https://doi.org/10.48550/arXiv.2010.11929

References

1. Gonzalez, R.C. and Woods, R.E. (2018), *Digital Image Processing*, 4th ed., Pearson/Prentice Hall, 1168 p. DOI/ISBN: 9780133356724
2. Yang, W., Zhao, H., Wang, M. and Ji, J. (2020), "Design of Intelligent Search Engine Service Performance Evaluation System", *2020 5th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS)*, Singapore, P. 86–91. DOI: https://doi.org/10.1109/ACIRS49895.2020.9162611
3. Amorós, F., Payá, L., Mayol-Cuevas, W., Jiménez, L.M. and Reinoso, O. (2020), "Holistic Descriptors of Omnidirectional Color Images and Their Performance in Estimation of Position and Orientation", *IEEE Access*, vol. 8, P. 81822–81848. DOI: https://doi.org/10.1109/ACCESS.2020.2990996
4. Liu, X., Cheung, G., Lin, C.-W., Zhao, D. and Gao, W. (2018), "Prior-Based Quantization Bin Matching for Cloud Storage of JPEG Images", *IEEE Transactions on Image Processing*, Vol. 27, No. 7, P. 3222–3235. DOI: https://doi.org/10.1109/TIP.2018.2799704
5. Carvalho, E.D., Filho, A.O.C., Silva, R.R.V., Araújo, F.H.D., Diniz, J.O.B., Silva, A.C., Paiva, A.C. and Gattass, M. (2020), "Breast Cancer Diagnosis from Histopathological Images Using Textural Features and CBIR", *Artificial Intelligence in Medicine*, Vol. 105, 101845 p. DOI: https://doi.org/10.1016/j.artmed.2020.101845
6. Nakazato, M. and Huang, T.S. (2001), "3D MARS: Immersive Virtual Reality for Content-Based Image Retrieval", *IEEE International Conference on Multimedia and Expo (ICME)*, Tokyo, Japan, P. 44–47. DOI: https://doi.org/10.1109/ICME.2001.1237651
7. Iqbal, K., Odetayo, M.O. and James, A. (2012), "Content-Based Image Retrieval Approach for Biometric Security Using Colour, Texture and Shape Features Controlled by Fuzzy Heuristics", *Journal of Computer and System Sciences*, Vol. 78, No. 4, P. 1258–1277. DOI: https://doi.org/10.1016/j.jcss.2011.10.013
8. Popescu, A. and Grefenstette, G. (2011), "Social Media Driven Image Retrieval", *Proceedings of the 1st ACM International Conference on Multimedia Retrieval (ICMR '11)*, New York, NY, USA, Article 33, P. 1–8. DOI: https://doi.org/10.1145/1991996.1992029
9. Liu, Y., Mei, T. and Hua, X.-S. (2009), "CrowdReranking: Exploring Multiple Search Engines for Visual Search Reranking", *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '09)*, New York, NY, USA, P. 500–507. DOI: https://doi.org/10.1145/1571941.1572027
10. Staszewski, P., Jaworski, M., Cao, J. and Rutkowski, L. (2022), "A New Approach to Descriptors Generation for Image Retrieval by Analyzing Activations of Deep Neural Network Layers", *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, No. 12, P. 7913–7920. DOI: https://doi.org/10.1109/TNNLS.2021.3084633
11. Sahmoudi, Y., El-Ogri, O., El-Mekkaoui, J. and Hjouji, A. (2024), "An Efficient Biomedical Color Image Retrieval System Based on Continuous Orthogonal Legendre Fourier Quaternion", *2024 Sixth International Conference on Intelligent Computing in Data Sciences (ICDS)*, Marrakech, Morocco, P. 1–6. DOI: https://doi.org/10.1109/ICDS62089.2024.10756406
12. Bano, M., Matta, P. and Chandel, S. (2024), "Content Based Image Retrieval: A Study of Approaches and Techniques", *2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS)*, Tashkent, Uzbekistan, P. 16–22. DOI: https://doi.org/10.1109/ICTACS62700.2024.10840489
13. Anand, A., Saxena, A. and Singh, K. (2024), "Statistical Features Based Content Based Image Retrieval Using Machine Learning Classifiers", *2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC)*, Gwalior, India, P. 1102–1109. DOI: https://doi.org/10.1109/AIC61668.2024.10731120
14. Debin, H., Yue, Z. and Shuai, J. (2024), "Application of Content-Based Retrieval Technology in Image Archive Management", *2024 Global Conference on Communications and Information Technologies (GCCIT)*, Bangalore, India, P. 1–6. DOI: https://doi.org/10.1109/GCCIT63234.2024.10862277

15. Bai, J., Ni, B., Wang, M., Li, Z., Cheng, S. and Yang, X. (2019), "Deep Progressive Hashing for Image Retrieval", *IEEE Transactions on Multimedia*, Vol. 21, No. 12, P. 3178–3193. DOI: https://doi.org/10.1109/TMM.2019.2920601
16. Lowe, D.G. (1999), "Object Recognition from Local Scale-Invariant Features", *Proceedings of the Seventh IEEE International Conference on Computer Vision*, Kerkyra, Greece, vol. 2, P. 1150–1157. DOI: https://doi.org/10.1109/ICCV.1999.790410
17. Bay, H., Ess, A., Tuytelaars, T. and Van Gool, L. (2008), "Speeded-Up Robust Features (SURF)", *Computer Vision and Image Understanding*, vol. 110, No. 3, P. 346–359. DOI: https://doi.org/10.1016/j.cviu.2007.09.014
18. Mikolajczyk, K. and Schmid, C. (2005), "A Performance Evaluation of Local Descriptors", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 10, P. 1615–1630. DOI: https://doi.org/10.1109/TPAMI.2005.188
19. Ke, Y. and Sukthankar, R. (2004), "PCA-SIFT: A More Distinctive Representation for Local Image Descriptors", *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Washington, DC, USA, Vol. 2, P. II–II. DOI: https://doi.org/10.1109/CVPR.2004.1315206
20. Calonder, M., Lepetit, V., Strecha, C. and Fua, P. (2010), "BRIEF: Binary Robust Independent Elementary Features", *Computer Vision – ECCV 2010. Lecture Notes in Computer Science*, Vol. 6314, P. 778–792. DOI: https://doi.org/10.1007/978-3-642-15561-1_56.
21. Rublee, E., Rabaud, V., Konolige, K. and Bradski, G. (2011), "ORB: An Efficient Alternative to SIFT or SURF", *2011 International Conference on Computer Vision*, Barcelona, Spain, P. 2564–2571. DOI: https://doi.org/10.1109/ICCV.2011.6126544
22. Leutenegger, S., Chli, M. and Siegwart, R.Y. (2011), "BRISK: Binary Robust Invariant Scalable Keypoints", *2011 International Conference on Computer Vision*, Barcelona, Spain, P. 2548–2555. DOI: https://doi.org/10.1109/ICCV.2011.6126542
23. Alahi, A., Ortiz, R. and Vandergheynst, P. (2012), "FREAK: Fast Retina Keypoint", *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, USA, P. 510–517. DOI: https://doi.org/10.1109/CVPR.2012.6247715
24. Žižakić, N. and Pižurica, A. (2022), "Efficient Local Image Descriptors Learned with Autoencoders", *IEEE Access*, Vol. 10, P. 221–235. DOI: https://doi.org/10.1109/ACCESS.2021.3138168
25. Liu, Y., Xu, X. and Li, F. (2018), "Image Feature Matching Based on Deep Learning", *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, Chengdu, China, P. 1752–1756. DOI: https://doi.org/10.1109/CompComm.2018.8780936
26. Radenović, F., Tolias, G. and Chum, O. (2019), "Fine-Tuning CNN Image Retrieval with No Human Annotation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, No. 7, P. 1655–1668. DOI: https://doi.org/10.1109/TPAMI.2018.2846566
27. Song, L., Lin, J., Wang, Z.J. and Wang, H. (2020), "An End-to-End Multi-Task Deep Learning Framework for Skin Lesion Analysis", *IEEE Journal of Biomedical and Health Informatics*, vol. 24, No. 10, P. 2912–2921. DOI: https://doi.org/10.1109/JBHI.2020.2973614
28. Wang, B., Zhang, H., Zhu, L., Nie, L. and Liu, L. (2023), "Multi-Level Adversarial Attention Cross-Modal Hashing", *Signal Processing: Image Communication*, Vol. 117, 117017 p. DOI: https://doi.org/10.1016/j.image.2023.117017
29. Gajjar, V., Khandhediya, Y. and Gurnani, A. (2017), "Human Detection and Tracking for Video Surveillance: A Cognitive Science Approach", *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, Venice, Italy, P. 2805–2809. DOI: https://doi.org/10.1109/ICCVW.2017.330
30. Adel, M., Moussaoui, A., Rasigni, M., Bourennane, S. and Hamami, L. (2010), "Statistical-Based Tracking Technique for Linear Structures Detection: Application to Vessel Segmentation in Medical Images", *IEEE Signal Processing Letters*, Vol. 17, No. 6, P. 555–558. DOI: https://doi.org/10.1109/LSP.2010.2046697
31. Truong, X.-T., Yoong, V.N. and Ngo, T.-D. (2015), "RGB-D and Laser Data Fusion-Based Human Detection and Tracking for Socially Aware Robot Navigation Framework", *2015 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Zhuhai, China, P. 608–613. DOI: https://doi.org/10.1109/ROBIO.2015.7418835
32. Galvez, R.L., Bandala, A.A., Dadios, E.P., Vicerra, R.R.P. and Maningo, J.M.Z. (2018), "Object Detection Using Convolutional Neural Networks", *TENCON 2018 – 2018 IEEE Region 10 Conference*, Jeju, Korea (South), P. 2023–2027. DOI: https://doi.org/10.1109/TENCON.2018.8650517
33. Wehbe, A., Hotiet, H., Minetti, I. and Dellapiane, S. (2024), "Integrating YOLO for Advanced Content-Based Image Retrieval in Lung Cancer Imaging", *2024 31st IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, Nancy, France, P. 1–4. DOI: https://doi.org/10.1109/ICECS61496.2024.10848862
34. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. and Houlsby, N. (2021), "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale", *International Conference on Learning Representations*. DOI: https://doi.org/10.48550/arXiv.2010.11929

Надійшла (Received) 15.04.2025

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 2 (32)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Відомості про авторів / About the Authors

Прокопенко Олександр Сергійович – Харківський національний університет радіоелектроніки, аспірант кафедри Програмної інженерії, Харків, Україна. e-mail: oleksandr.prokopenko1@nure.ua, ORCID ID: https://orcid.org/0000-0003-0489-6820

Смеляков Сергій Вячеславович – доктор фізико-математичних наук, професор, Харківський національний університет радіоелектроніки, професор кафедри Програмної інженерії, Харків, Україна. e-mail: serhii.smeliakov@nure.ua, ORCID ID: https://orcid.org/0000-0002-5791-2479, Scopus Author ID: 24527617600

Prokopenko Oleksandr – Kharkiv National University of Radio Electronics, PhD Student, Software Engineering Department, Kharkiv, Ukraine.

Smelyakov Serhii – Kharkiv National University of Radio Electronics, Doctor of Science, Professor at the Software Engineering Department, Kharkiv, Ukraine.

DEVELOPMENT OF AN OBJECT-ORIENTED IMAGE COMPARISON ALGORITHM FOR EFFICIENT SEARCH

The **object** of research is content-based image retrieval (CBIR). The **subject** of this study is models and methods for content-based image retrieval (CBIR) and managing large volumes of media content in extensive image storage systems. The **goal** of the research is to develop an algorithm for comparing object-oriented image descriptors, which involves using advanced computer vision models for object detection and constructing efficient methods for comparing and searching these descriptors. The proposed descriptor and comparison algorithm aim to enhance the efficiency and accuracy of image search and management processes. The **tasks** include: analyzing modern approaches and solutions for creating and comparing image descriptors and their use in CBIR; developing metrics and algorithms for comparing image descriptors that effectively utilize information about detected objects – such as their types, sizes, and locations – for image search in large data repositories; conducting experiments to evaluate the proposed image search algorithm and comparing its efficiency with existing solutions. The **methodology** includes: conducting a comprehensive review of advanced image descriptor generation methods, including hash-based descriptors, handcrafted descriptors, and deep learning-based descriptors; analyzing the use of existing descriptors in CBIR systems, focusing on their advantages and limitations; evaluating the best image search algorithms, including deep learning-based approaches; developing an object descriptor comparison algorithm for tag-based search, image-based search, and other tasks. The **results** obtained are as follows: an object-based image descriptor was developed using state-of-the-art machine learning models for object detection; metrics and comparison algorithms for the proposed descriptors were developed, enabling their use for CBIR in large data repositories; a series of experiments were conducted to assess the efficiency and search quality of the proposed descriptor and algorithms in large-scale image storage systems. These experiments compared their performance with existing methods, revealing their advantages and limitations, namely: faster descriptor generation; faster descriptor comparison than hashed, handcrafted, and deep learning-based descriptors; efficient image filtering in storage; higher search quality and speed for image-based queries. However, the descriptor's effectiveness depends on the quality of the model and data used for object detection, as images without detected objects do not appear in search results, which may limit search completeness. **Conclusions:** The developed algorithm for comparing object-oriented image descriptors is an effective tool for solving various CBIR tasks. The obtained results are satisfactory, as the proposed image search algorithm outperforms most alternatives in terms of speed and search quality. A promising direction for future research is the development of a CBIR system using the proposed descriptor and algorithms, enhanced by parallel and distributed computing, and further refinement for specific applications. This would allow its use not only for general-purpose images but also for more precise scientific domains.

Keywords: image processing; object detection; deep learning; image descriptor; image retrieval; big data; image storage; search optimization; information technology.

Бібліографічні описи / Bibliographic descriptions

Прокопенко О. С., Смеляков С. В. Розроблення об'єктно-орієнтованого алгоритму порівняння зображень для їх ефективного пошуку. *Сучасний стан наукових досліджень та технологій в промисловості*. 2025. № 2 (32). С. 79–101. DOI: https://doi.org/10.30837/2522-9818.2025.2.079

Prokopenko, O., Smelyakov, S. (2025), "Development of an object-oriented image comparison algorithm for efficient search", *Innovative Technologies and Scientific Solutions for Industries*, No. 2 (32), P. 79–101. DOI: https://doi.org/10.30837/2522-9818.2025.2.079