

Прокопенко О. С., Смеляков С. В.

РОЗРОБЛЕННЯ ДЕСКРИПТОРА ЗОБРАЖЕНЬ ДЛЯ ЕФЕКТИВНОГО ПОШУКУ НА ОСНОВІ ВМІСТУ У ВЕЛИКИХ СХОВИЩАХ ДАНИХ

Об'єктом дослідження є пошук зображень на основі вмісту. **Предметом** є моделі та методи пошуку зображень на основі вмісту (CBIR). **Мета дослідження** – розроблення дескриптора зображень, що використовує передові моделі комп'ютерного зору для виявлення об'єктів. Запропонований дескриптор підвищує ефективність і точність процесів пошуку зображень і управління ними. **Завдання передбачає:** аналіз сучасних підходів і рішень для створення дескрипторів зображень та їх використання в пошуку зображень на основі вмісту; виявлення обмежень наявних дескрипторів зображень і визначення вимог до дескриптора; розроблення нового дескриптора зображень, що ефективно застосовує інформацію про виявлені об'єкти; проведення експериментів для оцінювання запропонованого дескриптора й порівняння його ефективності з наявними рішеннями. **Методологія дослідження:** усебічний огляд передових методів створення дескрипторів зображень; аналіз використання наявних дескрипторів у системах CBIR; розроблення дескриптора, що застосовує сучасні моделі виявлення об'єктів; валідація ефективності запропонованого дескриптора порівняно з наявними дескрипторами на задачах пошуку. **Досягнуто таких результатів:** розроблено дескриптор зображення, оснований на об'єктах, виявлених за допомогою сучасних моделей машинного навчання; проведено серію експериментів для оцінювання ефективності та якості пошуку в великих системах зберігання зображень за допомогою запропонованого дескриптора. Експерименти виявили певні переваги, зокрема: швидше створення й порівняння дескрипторів, якщо порівнювати з гешованими, створеними вручну та дескрипторами, що ґрунтуються на глибокому навчанні; ефективне фільтрування зображень у сховищі на основі вмісту об'єктів, але ефективність дескриптора залежить від якості моделі й даних, оскільки зображення без виявлених об'єктів не з'являються в результатах пошуку, що може обмежувати його повноту. **Висновки:** розроблений дескриптор зображень є ефективним інструментом для розв'язання низки завдань пошуку зображень на основі вмісту завдяки швидкості побудови й пошуку. Перспективним напрямом дослідження є побудова системи пошуку зображень на основі вмісту з використанням розробленого дескриптора, посилене використання паралельних і розподілених обчислень.

Ключові слова: машинне навчання; дескриптори зображень; оброблення зображень; оптимізація пошуку зображень.

Вступ

Люди отримують більшість інформації зі свого оточення за допомогою органів зору. Тому з появою комп'ютерних технологій значні ресурси були спрямовані на витягування 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 обчислювальних ресурсів зображення могли мати лише низьку якість і займати лише кількості байтів або кілобайт. Однак нині зображення можуть сягати 32 гігабайти за секунду, якщо йдеться про відео з надвисокою роздільною здатністю, що використовується в спеціалізованих кінопродукціях. Навіть якщо не брати до уваги такі високоякісні знімки, повсякденні зображення зазвичай займають від кількох до десятків мегабайтів. Це дає змогу наблизитися до реалістичного візуального

відтворення, але водночас створює суттєві труднощі для розроблення ефективних алгоритмів, здатних обробляти такі великі масиви даних у реальному часі. Окреслена проблема ще більше ускладнюється стрімким зростанням обсягів і кількості сучасних систем зберігання інформації. Крім того, зображення часто зазнають дублювання й перетворень протягом їх життєвого циклу незалежно від того, чи це покращення, виконане користувачем, чи автоматичні сервіси оброблення, що стискають зображення за допомогою форматів із втратами (наприклад, JPEG).

Сучасні тенденції в пошуку зображень на основі вмісту (CBIR) [2] демонструють необхідність розроблення передових моделей і алгоритмів для витягування, порівняння та пошуку інформації, що ґрунтується на вмісті зображень [3–5]. Щоб спеціалізовані сервіси могли шукати, аналізувати зображення та приймати рішення в реальному часі, зазвичай створюється база даних дескрипторів. У цьому підході дескриптори генеруються й зберігаються в окремому сховищі.

Коли зображення надходить у систему пошуку, створюється його дескриптор. Безпосередній пошук і порівняння зображень у сховищі здійснюються за допомогою цих дескрипторів, що суттєво скорочує час пошуку й економить системні ресурси. Використання дескрипторів є значно ефективнішим рішенням, ніж виконання операцій безпосередньо над повнорозмірними файлами зображень. Однак вимоги до дескрипторів постійно підвищуються, оскільки від них залежать якість результатів пошуку й адекватність подання зображення. Це зі свого боку впливає на складність операцій пошуку [6].

Визначення проблеми

Наявні дескриптори мають знаходити компроміс між швидкістю та якістю і надавати пріоритет високій швидкості оброблення, спираючись на примітивні кольорові властивості, ігноруючи об'єкти на зображенні; або використовують високорозмірні векторні ознаки, отримані за допомогою глибокого навчання, що надають глибоке розуміння контексту й високу якість пошукових результатів, але мають відносно низьку швидкість обчислень, що робить таке рішення неефективним для великих сховищ зображень та інтенсивно завантажених потоків даних.

Аналіз останніх досліджень і публікацій

Будь-які параметри, статистичні показники або перетворення зображень, які дають змогу ефективно ідентифікувати й розрізнити зображення в сховищі, можуть використовуватися як їх дескриптори. Однак одним із найбільш поширених типів дескрипторів для пошуку подібних зображень є геші, що генеруються за допомогою алгоритмів перцептивного гешування. Ці алгоритми є функціями, що створюють порівняні геші [7, 8]. Прикладами є *Average Hash*, *DCT-Based Hash*, *Radial Variance-Based Hash* та *Marr-Hildreth Operator-Based Hash*. Узагальнений робочий процес для цих алгоритмів передбачає [9–11]:

- застосування специфічних трансформацій до зображення з метою усунення шуму;
- нормалізація зображення до фіксованого розміру;
- математичні трансформації, визначені алгоритмом;
- обчислення гешу.

Для порівняння дескрипторів гешів використовуються метрики, як-от відстань Хеммінга, нормалізована відстань Хеммінга та функція пік-кроскореляції. Деякі алгоритми гешування є стійкими до певних трансформацій зображень, наприклад розмиття, масштабування, стиснення, затемнення й незначні обертання [12, 13]. Основною перевагою геш-дескрипторів є швидкість їх порівняння, оскільки геші зазвичай займають до 128 байтів. Однак значним недоліком є те, що вони не беруть до уваги об'єкти в зображенні. Це може призвести до ситуацій, коли абсолютно різні реальні об'єкти з подібними кольорними схемами помилково ідентифікуються як подібні, і зіткнення не є рідкісним явищем [10, 14].

Іншим широко використовуваним типом дескриптора зображень є алгоритми, що базуються на ключових точках [15]: *SIFT (Scale-Invariant Feature Transform)*, *SURF (Speeded-Up Robust Features)*, *GLOH (Gradient Location and Orientation Histogram)*, *ORB (Oriented FAST and Rotated BRIEF)*, *BRISK* та *AKAZE (Accelerated KAZE)*. Ці алгоритми зазвичай передбачають двоетапний процес:

- виявлення ключових точок – ділянок, що суттєво відрізняються від навколишнього середовища, як-от контури чи кути об'єктів [16, 17];
- генерація дескрипторів для цих точок, що дає змогу порівнювати точки на різних зображеннях для визначення, чи належать вони до одного й того самого об'єкта [18–20].

Колекція дескрипторів ключових точок утворює дескриптор зображення. Порівняння цих дескрипторів дає змогу ідентифікувати однакові об'єкти на різних зображеннях, навіть під різними кутами, масштабами, умовами освітлення та іншими трансформаціями [21, 22]. Порівнюючи основані на ключових точках дескриптори з геш-дескрипторами, можна наголосити на значних перевагах останніх. Адже вони дають змогу виявляти об'єкти, здійснювати 3D-реконструкцію сцени, зшивати зображення й виконувати інші складні завдання комп'ютерного зору [23, 24]. Однак побудова й використання дескрипторів ключових точок є обчислювально складним процесом. Це робить їх менш придатними для порівняння великих обсягів зображень, оскільки час порівняння значною мірою залежить від кількості основних точок, виявлених на кожному зображенні [25, 26].

Дескриптори зображень, побудовані на глибокому навчанні, демонструють високу результативність у завданнях пошуку зображень, розпізнавання об'єктів чи семантичної сегментації, оскільки дають змогу узагальнювати різноманітні дані та ефективно аналізувати складні візуальні структури. Проте їх застосування пов'язано зі значними витратами обчислювальних потужностей як на етапі навчання, так і під час практичного використання. Якість роботи цих моделей, наприклад VGG, ResNet і EfficientNet, критично залежить від ретельного підбору, обсягу й різноманіття навчальної інформації. Незважаючи на це, саме здатність таких дескрипторів відтворювати глибокі, багаторівневі ознаки зображень викликають значну потребу в них у сучасних технологіях комп'ютерного зору. Останні дослідження продемонстрували зростання інтересу до екстракції дрібних ознак і крос-модальних підходів. Ці методи зосереджуються на навчанні дрібних семантичних ознак, щоб зберегти високий рівень семантичної релевантності й одночасно усунути особливі для модальності властивості.

Це приводить до більш точної та адекватної репрезентації оригінального зображення [11, 27].

Пошук зображень за вмістом (CBIR) зазнав суттєвого розвитку протягом останніх десятиліть, перейшовши від традиційних методів ручного виділення ознак до сучасних підходів, основаних на глибокому навчанні. Сутність CBIR полягає в тому, щоб знаходити в базі даних зображення, візуально схожі на задане, аналізувати безпосередньо їх вміст, а не текстові метадані. Ранні системи CBIR ґрунтувалися на низькорівневих ознаках, як-от кольорові гістограми, текстури й краї. Хоча ці методи були обчислювально ефективними, вони часто виявляли нездатність адекватно відтворювати високорівневу семантичну інформацію, що призводило до низької точності в реальних застосуваннях. Наприклад, зображення зі схожим розподілом кольорів, але абсолютно різними об'єктами могли бути помилково ідентифіковані як схожі. Такі системи містили значну кількість недоліків, але ці базові техніки заклали початок для створення більш складних пошукових систем. На рис. 1 зображено загальне подання такої системи.

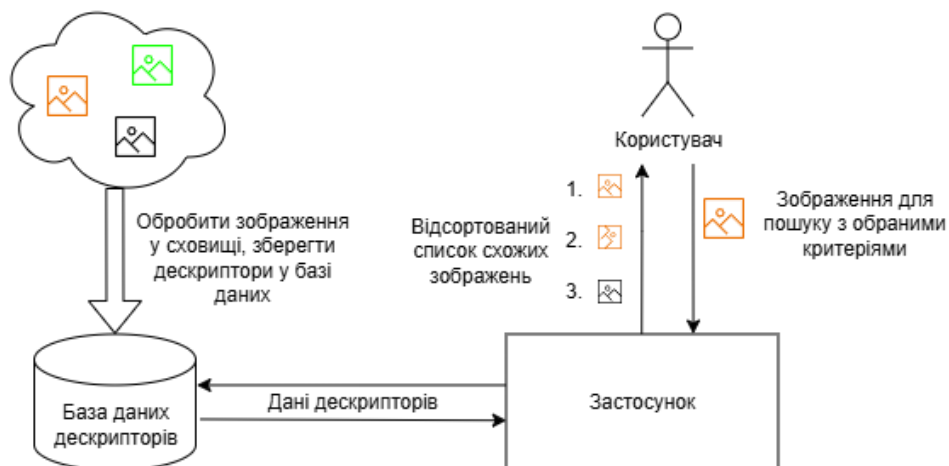


Рис. 1. Подання системи пошуку зображень за вмістом

Поява глибокого навчання сприяла революційним змінам у CBIR, надавши системам можливість самостійно навчатися складному, ієрархічному поданню ознак безпосередньо з даних. Сучасні системи CBIR використовують згорткові нейронні мережі (CNN) та інші архітектури глибокого навчання для отримання високовимірних векторів ознак або ембедингів, що інкорпують як низькорівневу, так і високорівневу візуальну інформацію. Попередньо навчені моделі, зокрема VGG, ResNet або EfficientNet, часто відіграють роль

екстракторів ознак, тоді як спеціалізовані архітектури, зокрема сіамські та триплетні мережі, застосовуються для навчання векторних ознак, оптимізованих саме для оцінювання схожості. Ці досягнення суттєво підвищили точність пошуку, даючи змогу системам CBIR ефективно обробляти складні запити, пов'язані з конкретними об'єктами, сценами чи навіть дрібними деталями. Наприклад, запит із зображенням певного виду птаха тепер може успішно знаходити зображення цього самого виду, зняті з багатьох ракурсів і за різних умов освітлення,

саме завдяки семантичній багатогранності дескрипторів, побудованих на основі глибокого навчання.

Наявні дескриптори для порівняння й пошуку зображень часто надають пріоритет маленькому розміру й високій швидкості оброблення, але здебільшого вони орієнтовані на колірні параметри, ігноруючи об'єкти на зображеннях. Хоча цей підхід добре працює в застосунках з унікальними зображеннями, він страждає від зіткнень, коли зображення мають подібні загальні властивості. Це призводить до недостатньо точних результатів порівняння й пошуку, оскільки дескриптори основані лише на загальних властивостях яскравості й кольору [28, 29].

Окремою значною тенденцією в сучасних CBIR є впровадження методів крос-модального пошуку, що дають змогу шукати співвідношення між різними типами даних, як-от пошук зображень за текстовим описом чи текстом за зображенням. Такий функціонал реалізується завдяки використанню моделей на кшталт CLIP (*Contrastive Language-Image Pretraining*), які навчаються подавати зображення й текст у єдиному спільному векторному просторі. Ці системи виявляються особливо цінними в таких сферах, як електронна комерція, де користувачі можуть шукати товари, описуючи їх словами, або в медичній діагностиці, де текстові висновки застосовуються для пошуку відповідних знімків. Крім того, у системах CBIR дедалі ширше використовуються механізми уваги (*attention*) й архітектури-трансформери, що допомагають моделі зосереджуватися на найбільш важливих ділянках зображення, що суттєво підвищує точність і релевантність результатів пошуку.

Попри ці вдосконалення, у сфері CBIR залишаються нерозв'язані завдання. Одним з основних є масштабованість систем пошуку, оскільки обчислювальна вартість порівняння високорозмірних векторів ознак значно зростає з розміром сховищ. Пошук прискорюють способом гешування й квантування векторів, але ці методи призводять до погіршення точності результуючої вибірки зображень, знайденої системою. Іншою проблемою є інтерпретованість систем CBIR на основі глибокого навчання, оскільки їх внутрішні механізми прийняття рішень часто залишаються "чорною скринькою". Крім цього, продуктивність таких систем критично залежить від даних, на яких вони навчаються. Недостатня репрезентативність або упередженість цих даних може спричинити систематичні помилки

й значно обмежувати здатність моделей коректно працювати з новими, незнайомими зображеннями.

Сучасні застосунки, де обсяги візуальної інформації (наприклад, у соціальних мережах, медичних базах даних чи системах спостереження) зростають експоненційно, наявні методи пошуку на основі вмісту часто є повільними або ресурсомісткими. Необхідний метод, що використовує обмежену кількість високорівневих даних і дає змогу швидко обробляти мільйони зображень, що критично для хмарних сервісів чи мобільних застосунків.

Визначення нерозв'язаних частин загальної проблеми. Мета й завдання роботи

Метою цього дослідження є розроблення дескриптора зображень, який використовує об'єкти на зображенні, що визначаються сучасними моделями виявлення об'єктів (наприклад, YOLO, Faster R-CNN), і застосовує типи об'єктів, їх розміри й місце розташування як ознаки зображення. Такий підхід забезпечує швидкість обчислень завдяки обмеженій кількості використовуваних даних, водночас надаючи глибоку контекстуальну інформацію для порівняння. Об'єкти, як складні сутності, несуть значну частину інформаційного вмісту зображення, що дає змогу забезпечити більш значущі й точні результати пошуку. Цей підхід пропонує збалансоване рішення для пошуку зображень на основі вмісту, значно покращуючи точність традиційних дескрипторів, і є набагато швидшим за високорозмірні вектори ознак, які використовуються в дескрипторах, що ґрунтуються на глибокому навчанні. Топові результати, обрані запропонованим методом, можна потім додатково аналізувати за допомогою обчислювально більш інтенсивних технік, забезпечуючи як ефективність, так і точність пошуку.

Ключова ідея розробленого дескриптора полягає в поєднанні низької обчислювальної вартості дескрипторів на основі низькорівневих ознак зображень із високорівневими даними об'єктів, що надаються згортковими нейронними мережами (CNN), які зазвичай потребують значних обчислювальних ресурсів. Цей підхід має на меті забезпечити баланс між ефективністю й точністю, використовуючи переваги обох методів для створення надійного й масштабованого рішення для завдань пошуку зображень.

На відміну від "чорних скриньок" глибокого навчання з високорозмірними векторами, цей дескриптор базується на зрозумілих властивостях об'єктів (тип, розмір, розташування), що полегшує пояснення, чому певні зображення вважаються подібними. Це важливо для галузей, де необхідна прозорість для прийняття рішень.

Поєднання низькорівневих ознак із високорівневими даними від нейронних мереж пропонує нові гібридні архітектури систем пошуку зображень на основі вмісту. Дослідження в цьому напрямі потрібні для того, щоб наблизитися до точки рівноваги, де ефективність і точність максимізуються, сприяючи створенню більш стійких до шумів чи варіацій зображень систем пошуку.

Важливою перевагою такого дескриптора є робота з об'єктами як абстрактними сутностями. Це дає змогу не прив'язуватися до конкретних моделей згорткових нейронних мереж для пошуку об'єктів, а використовувати найбільш сучасні та ефективні моделі. Залежно від розв'язуваного завдання й даних зображень, з якими необхідно працювати, можна обрати найбільш доцільний метод пошуку об'єктів щодо наявних обчислювальних ресурсів і часу, що не обов'язково має ґрунтуватися на машинному навчанні. Хоча в статті застосовується згорткова нейронна мережа для пошуку звичайних об'єктів на зображеннях, гнучкість дає змогу впроваджувати цей підхід для особливих зображень і наборів об'єктів (наприклад, медичних або військових) з використанням будь-яких методів для ефективного пошуку на основі вмісту у великих сховищах даних.

Розроблений дескриптор має відповідати таким критеріям:

- **ефективність** – застосування якомога менше даних про об'єкти для забезпечення обчислювальної та сховищної ефективності;
- **адаптивність** – сумісність із різними моделями виявлення об'єктів і водночас забезпечення гнучкості у використанні;
- **налаштуваність** – допущення налаштування під конкретні потреби різних наборів даних і випадків застосування;
- **баланс** – забезпечення оптимального балансу між швидкістю пошуку та якістю пошуку для забезпечення як швидкого, так і точного результату;
- **легкість побудови** – простота в створенні, що дає змогу ефективно будувати базу дескрипторів і безперешкодно інтегрувати її в системи пошуку.

Для досягнення окресленої мети необхідно виконати такі завдання:

- визначити поля даних для збережених об'єктів, щоб дескриптор був легким, але містив достатньо інформації для ефективного пошуку;
- розробити дизайн зберігання для дескриптора, який буде адаптивним до різних моделей виявлення об'єктів і налаштовуваним відповідно до конкретних вимог застосування;
- обрати відповідну модель виявлення об'єктів для експериментів, що дасть змогу провести справедливе порівняння з іншими дескрипторами;
- обрати набір даних для проведення серії експериментів, що визначать переваги й недоліки розробленого дескриптора;
- створити програмне забезпечення для реалізації побудови бази дескрипторів для зберігання зображень і забезпечення ефективного порівняння дескрипторів.

З розробленим дескриптором і реалізованим підтримувальним програмним забезпеченням можемо використовувати обраний набір даних і модель виявлення об'єктів та виконувати точний, але швидкий пошук у великих сховищах зображень.

Матеріали й методи

Обрання моделі для пошуку об'єктів

Для обрання моделі, що задовольнить потреби дескрипторів у поставленому завданні, було визначено певні критерії. Зокрема необхідно, щоб модель:

- широко й часто застосовувалася для пошуку об'єктів, що має гарантувати її ефективність за різних умов;
- була простою у використанні, здатною виявляти значну кількість об'єктів і уможлилювала просте перенавчання для адаптації до нових класів об'єктів або спеціалізації на певних типах об'єктів;
- мала кілька версій, які відрізняються кількістю параметрів, швидкістю і точністю, що уможлилює застосування одної й тої самої архітектури як у ресурсозабезпечених, так і в обмежених середовищах (наприклад, мобільні пристрої).

Важливо зауважити, що різні модифікації нейронної мережі можуть бути оптимізовані для зображень конкретних розмірів. Це має вирішальне значення, оскільки швидкість оброблення залежить від цього; нейронні мережі зазвичай добре

обробляють зменшення великих зображень, але вони не завжди можуть адекватно збільшувати маленькі зображення [30–32].

З огляду на перелічені вимоги до моделі для виявлення об'єктів, яку планується інтегрувати в систему керування масштабними сховищами зображень, обрано модель YOLOv8. На момент написання цього дослідження YOLOv8 є одним із найсучасніших рішень у сфері комп'ютерного зору для завдань детекції, що пропонує оптимальний баланс між швидкістю, точністю та адаптивністю. Ця модель є найновішою ітерацією успішної серії YOLO, яка інкорпорує низку нових функцій і вдосконалень для ще більшого підвищення продуктивності й гнучкості. YOLOv8 розроблена як швидкий, точний і зручний у використанні інструмент, що робить її ідеальним вибором для широкого спектра завдань, зокрема детекції та відстеження об'єктів, сегментації, класифікації зображень і оцінювання поз.

Основною перевагою моделей сімейства YOLO для розв'язання поставленого завдання є швидкість, яка відіграє вирішальну роль. Вона працює у 2–2,5 рази швидше за інші популярні моделі, такі як Faster R-CNN, SSD чи RetinaNet, демонструючи порівняну, а іноді навіть вищу точність. Це досягається завдяки її одноетапній архітектурі детекції, на відміну від дворівневих підходів, властивих для сімейства мереж R-CNN.

Ще одним вагомим аргументом на користь YOLOv8 є її зручна підтримка апаратного прискорення. Модель дає змогу легко долучити графічний процесор (GPU) для пришвидшення детекції, часто за допомогою лише одного додаткового параметра в разі виклику, що може збільшити швидкість оброблення зображень у 5–10 разів.

Архітектурно модель трансформує вхідне зображення в кілька карт ознак, використовуючи пропускання з'єднання та інші сучасні техніки. Ці карти ознак потім об'єднуються та спрощуються до єдиного вихідного шару внаслідок комбінації операцій апсемплінгу та об'єднання. На цьому етапі прогноуються класи об'єктів та їх обмежувальні прямокутники, після чого для кожного об'єкта за допомогою алгоритму *Non-Maximum Suppression* (NMS) обирається найбільш вірогідний прямокутник. Кожен такий прямокутник описується шістьма значеннями: ідентифікатор класу, координати центру (X , Y), ширина, висота й рівень впевненості прогнозу.

Методологія пошуку

На першому етапі виявляються об'єкти на кожному зображенні в репозиторії за допомогою обраної моделі. Дескриптори створюються з використанням об'єктів, виявлених на зображенні, що після перетворення всієї необхідної інформації зберігається як база даних дескрипторів зображень.

Ця база даних стає основою для подальшого пошуку зображень. Вона ефективно порівнює класи, положення та інші ознаки об'єктів. Детальні описи нейронних мереж, дескрипторів і методів їх побудови надано нижче в цьому розділі.

Далі використовуємо просту метрику для оцінювання схожості між зображеннями (точніше, їх дескрипторами) й алгоритм пошуку, що застосовує ці дескриптори й метрики для визначення подібності зображень.

Загальний процес пошуку передбачає такі дії:

- 1) побудова дескриптора – його генерування для запитуваного зображення;
- 2) налаштування пошуку – визначення його критеріїв і налаштування параметрів (наприклад, пороги схожості, типи об'єктів);
- 3) безпосереднє виконання пошуку:
 - перебір дескрипторів у репозиторії;
 - порівняння кожного дескриптора з дескриптором запитуваного зображення за допомогою метрики схожості;
 - сортування зображень за балом подібності;
 - виведення результатів за схожістю.

За ідеальних умов відсортовані результати пошуку мають показати саме запитуване зображення (якщо воно є в репозиторії) як перший результат, далі трансформації або варіації запитуваного зображення (наприклад, змінене за розміром, обрізане або відредаговане зображення), а потім інші зображення з репозиторію, відсортовані за їх схожістю із запитуваним.

Для перевірки алгоритмів і програмного забезпечення розробляються допоміжні моделі, методи, критерії, метрики й програмні інструменти, зокрема метрики для оцінювання якості пошуку. Експериментальне налаштування, результати й детальний аналіз подано у відповідних розділах нижче.

Моделі дескриптора

Дескриптор зображення – це компактне подання, що інкапсулює основну інформацію про зображення в репозиторії. Він містить як технічні метадані (місце

зберігання, роздільну здатність, формат файлу), так і семантичні показники про виявлені на зображенні об'єкти. Останні охоплюють координати обмежувальних прямокутників, геометричні розміри (площу), а також оцінку впевненості моделі комп'ютерного зору в належності об'єкта до певного класу (наприклад, класу i). Такий мінімалістичний, але вичерпний набір атрибутів забезпечує основу для ефективного розроблення алгоритмів пошуку, фільтрації та порівняння зображень на базі їх візуального вмісту, зокрема за присутністю й властивостями об'єктів.

Кожен об'єкт подано вектором із семи чисел, як показано в табл. 1.

Зверніть увагу, що дескриптор об'єкта не зберігає номер класу, до якого належить об'єкт. Для зберігання номеру класу кожного об'єкта використовується структура дескриптора класу, що містить перелік знайдених об'єктів.

Таблиця 1. Модель дескриптора об'єкта

Назва	Опис
x	Відносна координата x -осі центра обмежувальної рамки об'єкта (у діапазоні 0–1)
y	Відносна координата y -осі центра обмежувальної рамки об'єкта (у діапазоні 0–1)
w	Відносна ширина об'єкта (у діапазоні 0–1)
h	Відносна висота об'єкта (у діапазоні 0–1)
$area$	Відносна площа об'єкта
$ratio$	Відношення меншої до більшої сторони обмежувальної рамки об'єкта
$conf$	Впевненість нейронної мережі в тому, що рамка містить об'єкт певного класу (у діапазоні 0–1)

Дескриптор класу описує сукупність об'єктів, що належать до певного класу на зображенні. Він містить набір дескрипторів для кожного виявленого об'єкта цього класу, а також додаткову інформацію, що характеризує групу об'єктів загалом.

Ця додаткова інформація містить кількість об'єктів певного класу, сумарну площу, яку вони займають, а також центр мас (центроїд) всієї групи на зображенні (див. табл. 2).

Таблиця 2. Модель дескриптора класу

Назва	Опис
$class$	Ідентифкатор класу
$number$	Кількість об'єктів цього класу на зображенні
$area$	Сума відносних площ зайнятими об'єктами цього класу на зображенні
$center$	Арифметичний центр групи об'єктів цього класу на зображенні
$objects$	Колекція дескрипторів об'єкта цього класу на зображенні

Хоча ці показники можна обчислити безпосередньо з дескрипторів окремих об'єктів, їх попереднє обчислення й зберігання дає змогу уникнути повторних розрахунків під час операцій порівняння. Це суттєво скорочує час виконання таких операцій і підвищує загальну ефективність системи.

Дескриптор зображення охоплює основну інформацію, необхідну для локалізації та опису оригінального зображення, а саме:

- місце розташування зображення, зокрема шлях до файлу (локальний або мережевий), URI або інший унікальний ідентифікатор, що дає змогу в разі необхідності отримати повнорозмірне зображення;
- габарити зображення, тобто абсолютну ширину й висоту в пікселях. Ці параметри є необхідними для перерахунку відносних координат і розмірів обмежувальних прямокутників об'єктів у абсолютні значення, оскільки всі дані про об'єкти зберігаються у нормованому, відносному вигляді;
- дескриптори класів – колекцію дескрипторів для кожного класу об'єктів, виявлених на зображенні.

Важливо зауважити, що структура дескриптора зображення є гнучкою та може бути розширена для долучення додаткової інформації про зображення або його об'єкти відповідно до конкретних вимог завдання. Незважаючи на можливість адаптації, основним критерієм залишається збереження компактності дескриптора, що є критично важливим для забезпечення високої швидкості оброблення під час роботи з масштабними сховищами зображень. Дані, що формують модель дескриптора зображення, узагальнені в табл. 3.

Таблиця 3. Модель дескриптора зображення

Назва	Опис
$path$	Шлях до файлу зображення
$width$	Ширина зображення
$height$	Висота зображення
$classes$	Дескриптори класів, знайдених на зображенні

Пошукові моделі, алгоритми й метрики

Дані, отримані від нейронної мережі та використані для формування дескрипторів, можуть застосовуватися в різних пошукових моделях.

Фільтрація за класами об'єктів

Дескриптори дають змогу відбирати зображення в репозиторії, визначаючи ті, що містять задану кількість об'єктів певних класів. Залежно від налаштувань, система може шукати зображення,

що містять точний набір об'єктів або ж хоча б зазначений набір. Результати такого пошуку можуть бути корисними самі по собі або слугувати попереднім фільтром для звуження простору пошуку. Це значно зменшує обсяг інформації для наступних, більш складних та ресурсомістких процедур, і підвищує загальну ефективність системи.

Пошук за зображеннями

Дескриптори також забезпечують можливість пошуку за зображеннями, допомагаючи користувачам знаходити конкретні зображення або визначати схожі на основі присутніх об'єктів та їх розташування. Ця модель пошуку може знаходити точні відповідності в репозиторії та ранжувати схожі зображення за їх візуальним вмістом. Вона вважається основною (за замовчуванням).

На цьому етапі пошук за зображенням реалізується за допомогою найпростіших порівняльних операцій: зіставляємо кількість об'єктів одного класу на двох різних зображеннях. Для кожного класу обчислюється різниця в кількості об'єктів, яка потім ділиться на більше з двох значень (між порівнюваними дескрипторами). Отримані значення для всіх класів, присутніх хоча б в одному з дескрипторів, підсумовуються, а сума ділиться на загальну кількість унікальних класів у обох дескрипторах. Такий підхід забезпечує комутативність операції порівняння (результат не залежить від порядку дескрипторів), дає змогу виявляти повні копії зображень у сховищі й гарантує, що зображення з ідентичним набором об'єктів повертатимуться як найбільш релевантні результати пошуку.

$$S = \sum_{s=1}^n w^* \frac{|O_1^i - O_2^i|}{\max(O_1^i, O_2^i)}, \quad (1)$$

де S – коефіцієнт подібності; n – кількість різних типів об'єктів на зображенні; w – ваговий коефіцієнт типу (розраховується як $1/n$); O_1^i – кількість об'єктів i -го типу на першому зображенні; O_2^i – кількість об'єктів i -го типу на другому зображенні.

Ось приклад алгоритму, який можна використати для порівняння двох дескрипторів, що дає змогу зіставити час виконання пошуку з іншими дескрипторами.

Опис програмного забезпечення

Для втілення запропонованих методів пошуку зображень на основі дескрипторів розроблено

застосунки з мінімалістичним вебінтерфейсом. Ці інструменти полегшують проведення експериментів і наочно відтворюють результати пошукових запитів. Для комплексного тестування запропонованої моделі пошуку, метрик, алгоритмів і програмного забезпечення створено чотири компоненти.

1. **Служба YOLO.** Реалізована на *Python*, абстрагує взаємодії 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 збільшення швидкості обчислень.

2. **База даних.** Для експериментальної установки застосовано *MongoDB*, обраний за свою гнучкість як база даних NoSQL. Під час досліджень і розроблення моделі дескрипторів часто розширюються й модифікуються, і *MongoDB* дає змогу безперешкодно коригувати їх без постійної міграції даних. *MongoDB* також підтримує масштабованість завдяки таким технікам, як шардинг, що робить її придатною для роботи з великими репозиторіями інформації зображень. Створення гнучкого рішення для бази даних є критичним для передбачення й розв'язання проблем, що виникають під час роботи з масштабованими системами зберігання зображень [33–35].

3. **Сервіс на сервері.** Постійно оновлюється, щоб підтримувати актуальність моделей дескрипторів. Відповідає за виконання операцій пошуку й порівняння. Обробляє запити користувачів на детектування об'єктів на зображеннях і додавання фотоархівів до бази даних дескрипторів. Реалізовано в .NET, зберігає дескриптори в пам'яті з метою підвищення швидкості виконання та використовує *ImageMagick* для трансформацій зображень.

4. **UI-сервіс.** Легкий фронтенд, побудований на *Angular* для демонстраційних цілей. Дає змогу користувачам візуалізувати можливості системи, додавати фото до репозиторію, створювати дескриптори для зображень у визначених директоріях і здійснювати пошукові запити (наприклад, за тегами або схожістю до запитуваного зображення). Крім того, створено простий набір інструментів для генерування трансформованих версій вхідних зображень (наприклад, зміна масштабу, стиснення, гамма-корекція тощо). Ці трансформації використовуються для тестування стійкості й точності методу пошуку зображень у великих репозиторіях.

Результати досліджень та їх обговорення

Опис даних

Для експериментів використовуємо зображення з набору даних COCO 2017. Цей набір містить понад 163 тис. зображень, анотованих для 80 категорій об'єктів. Такий великий обсяг даних дає змогу оцінити швидкодію пошуку в репозиторії, що налічує сотні тисяч елементів.

Повна назва набору – *Common Objects in Context* ("Загальні об'єкти в контексті") – відтворює властивості зображень, які переважно є повсякденними фотографіями: на них зображені люди, природа, міські пейзажі, спортивні події тощо. Серед розпізнаваних класів об'єктів – люди, транспортні засоби, тварини, пристрої, предмети побуту й аксесуари.

Окрім оригінальних зображень із COCO 2017, додаємо до сховища модифіковані копії тестових зображень, які використовуватимуться як цільові запити. Список трансформацій, застосованих до оригіналів, передбачає: зниження яскравості на 20%, стиснення якості до 70%, підвищення контрастності, збільшення розміру зображення на 20%, зменшення розміру на 20%, перетворення в чорно-білу палітру, квантування кольорів до 128 відтінків, підвищення насиченості на 30% і використання ефекту різкості. Очікується, що в результатах пошуку для кожного тестового запиту серед перших позицій буде наявне як оригінальне зображення, так і всі його трансформовані версії.

Побудова бази даних дескрипторів

Репозиторій дескрипторів створено способом розміщення набору даних на робочій станції та його оброблення створеним програмним забезпеченням. Це програмне забезпечення приймає шляхи до каталогів зі зображеннями, послідовно обробляє всі зображення в цих каталогах, виявляє об'єкти на кожному зображенні, трансформує дані про об'єкти й зображення в описану модель дескриптора й зберігає отримані дескриптори в базі даних. Набір даних містить приблизно 164 тис. зображень, а процес генерації дескрипторів для всього масиву зайняв майже п'ять з половиною годин. Цей час передбачає виявлення об'єктів, побудову дескрипторів та їх індексацію в базі даних.

Підготовка експериментів

Експерименти зосереджені на швидкості побудови дескрипторів, швидкості фільтрації датасету й швидкості виконання порівняно з аналогами.

По-перше, порівняємо швидкість побудови дескрипторів із дескрипторами на основі глибокого навчання й перцептуальними гешами та побудуємо всі три дескриптори на зображеннях датасету.

По-друге, виконаємо тест пошуку за тегами. Цей тест фільтрує загальну вибірку, а саме лише екземпляри, що містять зазначений набір класів об'єктів. Проведемо пошукові запити, вказуючи по черзі від одного до п'яти класів об'єктів, які мають бути присутні в зображенні, виконаємо п'ять запитів із кожною конфігурацією та виміряємо час виконання.

Далі проведемо пошук для десяти тестових зображень, використовуючи розроблені дескриптори, перцептуальні геші та дескриптори на основі глибокого навчання, отримані за допомогою видобутку ознак із зображень датасету, а також згенеровані за допомогою сучасної моделі *Vision Transformer*, яку наприкінці 2020 р. запропонували О. Досовіцький, Л. Баєр, О. Колесніков та ін. [36, 37]. Виміряємо час виконання, а також визначимо кількість оригінальних і трансформованих зображень із цільової групи, які постійно з'являються в топових результатах вибору. Це дасть практичні результати для розроблених дескрипторів і моделей пошуку.

Побудова дескрипторів

Дескриптори побудовано для понад 163 тис. зображень у датасеті. Визначено середню швидкість побудови дескрипторів для кожного типу дескриптора. Варто зауважити, що дескриптори, зокрема на основі глибокого навчання, створені як на CPU, так і на GPU, тоді як перцептуальні геші обчислювалися лише на CPU. Результати подано в табл. 4.

Таблиця 4. Порівняння часу побудови дескрипторів

Тип дескриптора	Середня швидкість (зображень за секунду)
Перцептивний геш (CPU)	4.39
Об'єктний дескриптор (CPU)	2.56
Дескриптор ознак зображення (CPU)	2.14
Об'єктний дескриптор (GPU)	12.89
Дескриптор ознак зображення (GPU)	12.68

Результати експериментів демонструють, що побудова дескрипторів на основі об'єктів виявилася трохи швидшою, ніж генерування дескрипторів зображень, оснований на ознаках (наприклад, SIFT або інші класичні дескриптори), але, як і очікувалося, більш повільною за обчислення перцептивного гешу за однакових умов. Однак у процесі застосування апаратного прискорення на GPU час створення як об'єктних дескрипторів, так і дескрипторів на основі глибокого навчання скоротився у 5–6 разів у нашому тестовому середовищі. Ці результати свідчать про те, що час побудови об'єктних дескрипторів є порівняним, а іноді навіть меншим за час створення аналогічних дескрипторів.

Пошук за тегами

Цей експеримент передбачає запити на пошук присутніх на них зображень із кількістю класів, що коливається від одного до п'яти, із загальною кількістю понад 163 тис. зображень у репозиторії.

Як можна побачити з табл. 5, час пошуку не збільшується – навпаки, він зменшується із зростанням кількості класів на зображенні. Це здебільшого зумовлено порівнянням векторів, що забезпечує стабільне оброблення незалежно від кількості класів. Однак із збільшенням кількості класів кількість зображень, що задовольняють критерії пошуку, зменшується, що знижує витрати на виділення пам'яті за умови повернення результатів.

Середній час виконання запиту (50–100 мс для 164 тис. зображень) вказує на те, що фільтрація за складом об'єктів може обробляти понад мільйон

зображень за секунду, залишаючи вибірку з кількох тисяч зображень для більш детальних порівнянь.

Таблиця 5. Пошук зображень за тегами, час виконання

Кількість класів у запиті	Тест № 1	Тест № 2	Тест № 3	Тест № 4	Тест № 5	Середній час
1	121 мс	78 мс	79 мс	99 мс	106 мс	96.6 мс
2	92 мс	84 мс	80 мс	99 мс	89 мс	88.8 мс
3	71 мс	76 мс	57 мс	71 мс	80 мс	71 мс
4	67 мс	92 мс	64 мс	64 мс	61 мс	69.6 мс
5	54 мс	52 мс	50 мс	46 мс	59 мс	52.2 мс

Такі результати доводять, що розроблений об'єктно-орієнтований дескриптор може використовуватися у великих сховищах зображень як швидкий інструмент фільтрації на основі вмісту.

Пошук за зображенням

У межах цього експерименту здійснено пошук для 10 тестових зображень у великому репозиторії із застосуванням різних типів дескрипторів для порівняльного аналізу. Основним критерієм оцінювання був час виконання пошукових запитів, що дало змогу встановити, наскільки продуктивність розробленого дескриптора відповідає показникам відомих альтернатив. Отримані результати доводять ефективність запропонованого підходу, а розвиток і оптимізація алгоритмів пошуку залишатимуться пріоритетом для подальших наукових досліджень.

Кілька скриншотів результатів пошуку за зображенням подано на рис. 2–4. Усі скриншоти можна знайти в репозиторії <https://github.com/alex-prokopenko-nure/image-search-results>.

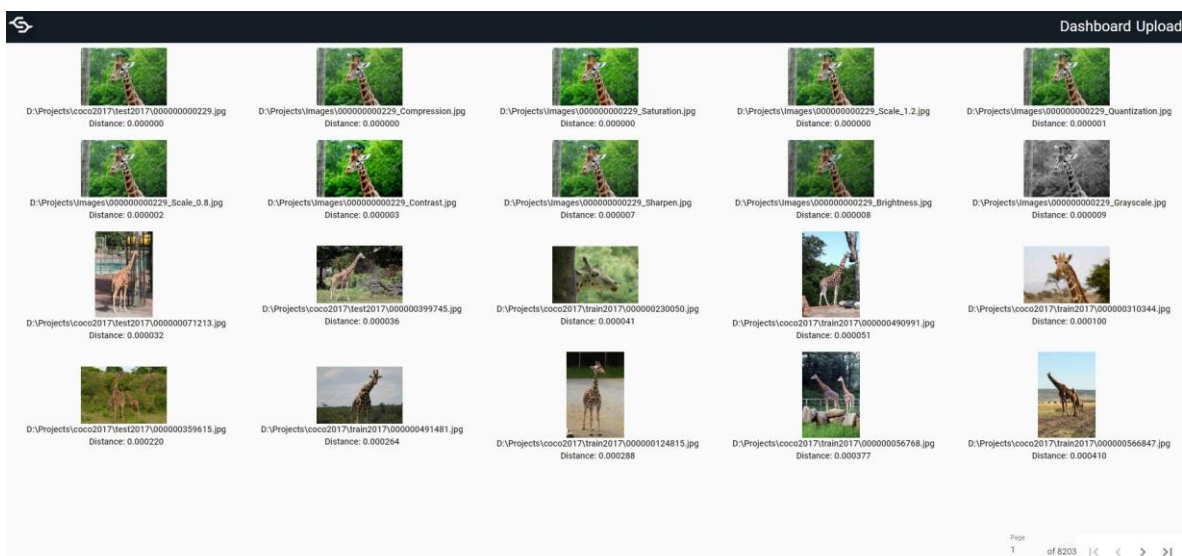


Рис. 2. Результат пошуку для тестового зображення № 3

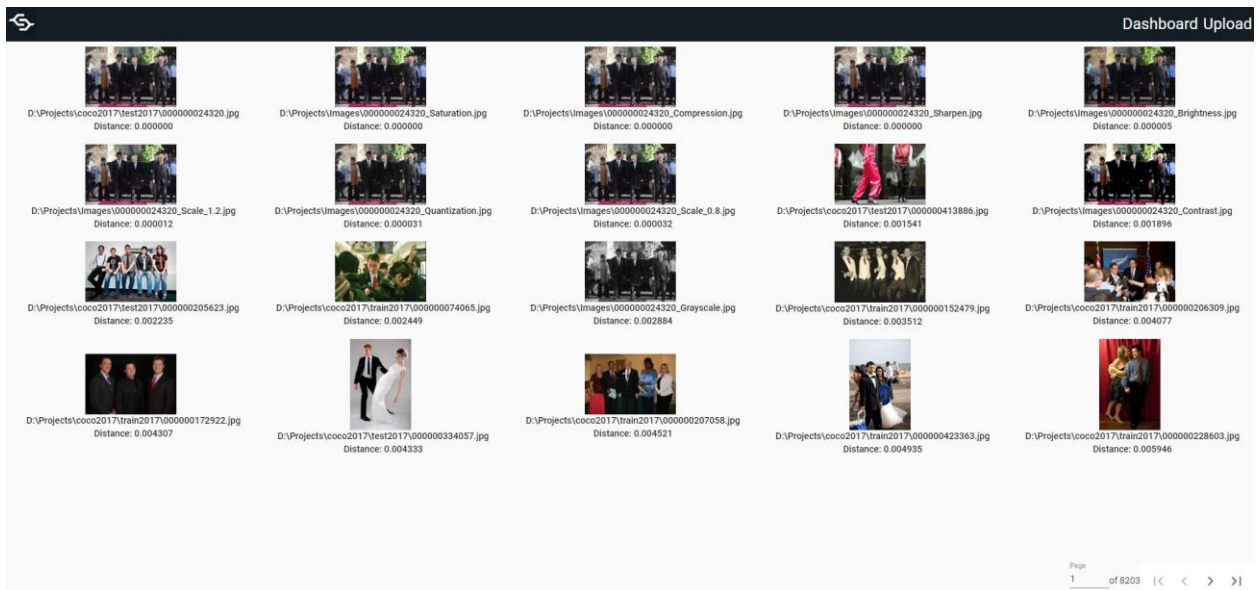


Рис. 3. Результат пошуку для тестового зображення № 6

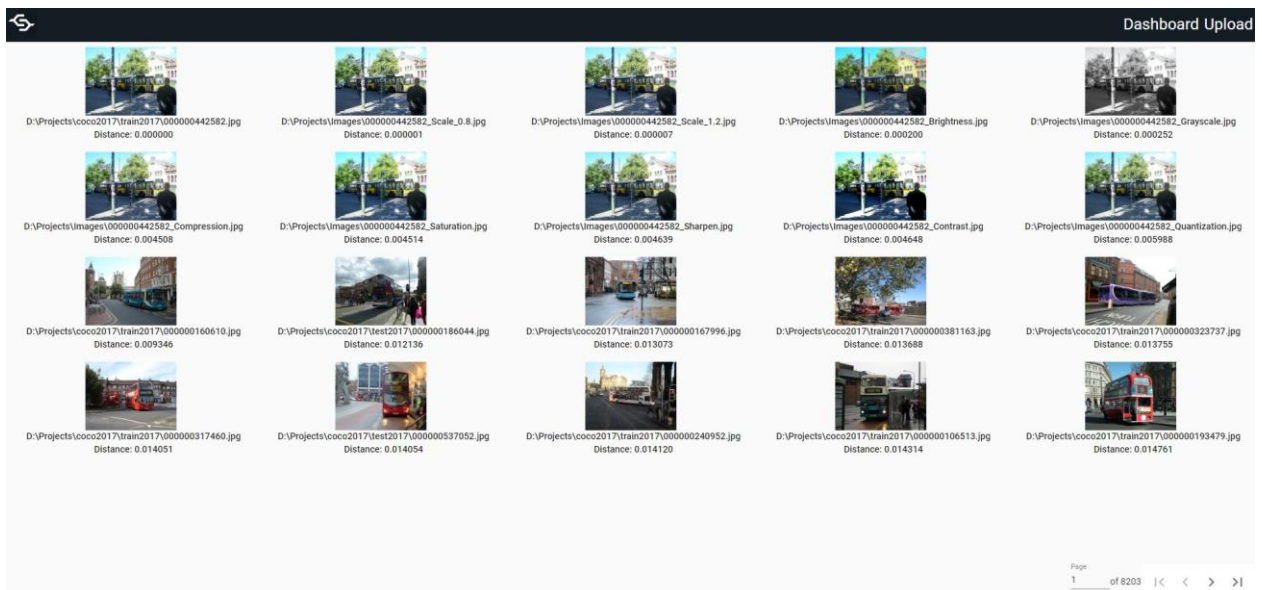


Рис. 4. Результат пошуку для тестового зображення № 10

Пошук зображень у цій роботі реалізовано за допомогою трьох різних методів: перцептивних гешів, дескрипторів на основі глибокого навчання й розроблених об'єктно-орієнтованих дескрипторів. Для перцептивних гешів порівняння відбувається способом обчислення суми квадратів різниць між бінарним поданням гешів. Дескриптори на основі глибокого навчання порівнюються за допомогою обчислення косинусної подібності між їх високовимірним векторним поданням (ембедингом). Об'єктно-орієнтовані дескриптори порівнюються за допомогою примітивного алгоритму, детально описаного в попередніх розділах, який бере

до уваги кількість і розподіл об'єктів різних класів на зображеннях.

У табл. 6 подано результати експерименту. Час пошуку наведено для перцептивних гешів (PH), дескрипторів ознак зображень (FD) та об'єктних дескрипторів (OD).

Деякі нетипові зображення можуть знижувати точність пошуку, зрештою опиняючись вище за трансформовані версії цільового зображення. Це відбувається, коли модель детекції об'єктів виявляє на них інший набір об'єктів, якщо порівнювати з оригіналом. Найбільш суттєве відхилення виникає, коли на трансформованому зображенні (наприклад,

затемненому або зі зміненими кольорами) не виявлено жодного об'єкта з певного класу, наявного на оригіналі. Іноді відстань до такого нетипового зображення зі схожим, але неідентичним набором об'єктів виявляється меншою, ніж до трансформованої версії з неповним або спотвореним набором об'єктів.

Таблиця 6. Час виконання пошуку за зображенням

	Час пошуку (PH)	Час пошуку (FD)	Час пошуку (OD)
1	214 мс	69.7 с	144 мс
2	221 мс	62.6 с	122 мс
3	230 мс	63.2 с	123 мс
4	242 мс	66.4 с	130 мс
5	215 мс	58.1 с	154 мс
6	226 мс	68 с	132 мс
7	223 мс	61.7 с	132 мс
8	252 мс	64.9 с	131 мс
9	235 мс	62.7 с	155 мс
10	228 мс	62.8 с	142 мс

Пошук на основі перцептивних гешів виконується за аналогічний проміжок часу. Однак, хоча 6–8 трансформацій з кожного тестового запиту потрапляють до верхніх позицій результатів, перетворення в чорно-білу палітру для всіх тестових зображень опиняється поза першою тисячею знайдених елементів. Крім того, для 8 з 10 тестових запитів ще 1–2 інші трансформації (пов'язані з маніпуляціями кольоровим діапазоном, як-от зміна насиченості або яскравості) також випадають за межі першої тисячі результатів.

Пошук за допомогою косинусної подібності між векторами ознак, отриманими за допомогою моделі *Vision Transformer* (ViT), значно довший за часом, якщо порівнювати з іншими методами. Аналогічно до експерименту з перцептивними гешами, перетворення зображень у чорно-білу палітру часто не потрапляє до найрелевантніших результатів, опиняючись за межами першої двадцятки в трьох випадках. Однак інші види трансформацій (як-от зміна яскравості, контрасту, розміру) переважно посідають місця в першій десятці. Якщо вилучити з оцінки перетворення в чорно-біле, то якість пошуку для частини зображень за цим методом вища, ніж у запропонованого об'єктно-орієнтованого підходу, але досягається це ціною значно більших витрат обчислювального часу.

Пошук за допомогою порівняння векторів ознак зображень з подальшим доопрацюванням або комбінацією з іншими методами може використовуватися для більш точного порівняння зображень у попередньо відфільтрованій вибірці із

сотень чи тисяч зображень, спочатку відфільтрованих за допомогою запропонованого в цій роботі методу.

Обговорення

Досягнуті результати підтверджують, що створений дескриптор успішно виконує поставлені завдання, зокрема ефективний пошук схожих зображень у великих сховищах даних. Він поєднує низькі обчислювальні витрати, властиві для дескрипторів на основі низькорівневих ознак, із високорівневою семантичною інформацією про об'єкти, яку зазвичай надають згорткові нейронні мережі (CNN), що вимагають значно більших обчислювальних ресурсів.

Дескриптори на основі об'єктів будуються швидше, ніж генерується векторне подання (ембединг) на базі глибокого навчання, але, як і передбачалося, поступається за швидкістю обчисленню перцептивних гешів. Така продуктивність у створенні дескрипторів робить їх практично застосовними і надає певну перевагу перед більш повільними методами на основі глибокого навчання.

Швидкість фільтрації, досягнута завдяки розробленому дескриптору, дає змогу формувати підмножини зображень, ґрунтуючись на їх об'єктному складі, з продуктивністю майже в один мільйон зображень за секунду. Ця можливість відкриває шлях до ефективної комбінації запропонованого дескриптора з більш складними методами пошуку, використовуючи попередню фільтрацію для різкого звуження ділянки пошуку і, як наслідок, отримання точних і релевантних результатів високої якості.

Швидкість порівняння зображень за допомогою об'єктних дескрипторів у сотні разів перевищує швидкість порівняння з використанням векторів ознак, отриманих за допомогою глибокого навчання, зберігаючи в цьому разі ключовий семантичний контекст зображення в пошуковій операції. Якість пошуку в цьому дослідженні не вимірювалася, але можна припустити, що вона може бути дещо нижчою за дескриптори на основі глибокого навчання. Саме це робить поданий дескриптор ідеальним інструментом для попереднього фільтрування та різкого зменшення обсягу вхідної інформації для більш складних і точних моделей пошуку на базі глибокого навчання, які зможуть провести фінальне уточнення порядку результатів для заданого запиту.

Подальші дослідження необхідно зосередити на створенні швидкого й точного алгоритму пошуку для розробленого дескриптора. Такий алгоритм

використовуватиме інформацію про розмір і розташування об'єктів для пошуку подібних зображень, а також трансформацій зображень, як-от зміни яскравості, кольору, масштабу, ступеня стиснення над оригіналом у сховищі.

Ще одним логічним удосконаленням є використання розподілених обчислювальних систем, оскільки поставлене завдання має високу автономність. Результати обчислень, виконаних для одного зображення, не потрібні для обчислень для іншого зображення. Тому навіть у великих репозиторіях, що містять мільярди зображень, швидкий пошук за зображенням відповідно до розробленого алгоритму можна виконати, оскільки складне завдання припустімо поділити на підзавдання й одночасно їх розв'язувати на різних частинах набору даних десятками або сотнями різних машин.

6. Висновки

й перспективи подальшого дослідження

У статті порушено проблему компромісу між швидкістю й точністю дескрипторів зображень, що застосовуються для пошуку за вмістом у великих сховищах даних, і способи підвищення ефективності таких систем за допомогою використання нових видів дескрипторів зображень.

Для розв'язання окресленої проблеми запропоновано об'єктно-орієнтований дескриптор зображень, який застосовує дані, отримані за допомогою моделей детекції об'єктів, з метою створення ефективного й збалансованого рішення для пошуку за вмістом (CBIR). Цей підхід значно покращує точність традиційних дескрипторів, залишаючись набагато швидшим за високовимірне векторне подання, що використовується в дескрипторах на основі глибокого навчання. Для розроблення дескриптора на базі об'єктів було виконано такі завдання: визначено поля даних для збережених об'єктів, створено дизайн зберігання для дескриптора, обрано відповідну модель виявлення об'єктів, обрано набір даних і розроблено програмне забезпечення для порівняння дескрипторів у сховищах зображень.

Результати експериментів, що порівнюють об'єктний дескриптор з методами на основі перцептивних гешів і глибокого навчання, демонструють переваги запропонованого підходу. Зокрема побудова об'єктних дескрипторів відбувається швидше, ніж створення векторного подання за допомогою глибокого

навчання. Крім того, запропонований дескриптор забезпечує найвищу швидкість пошуку у великих сховищах, перевершуючи навіть швидкі обчислення на основі перцептивних гешів. Такі порівняння в 100 разів швидші за порівняння зображень за допомогою дескрипторів на основі глибокого навчання, тоді як вони також зберігають контекстну інформацію про зображення. Як і очікувалося, якість пошуку об'єктного дескриптора є дещо нижчою, ніж у передових методів глибокого навчання, але значно вищою, ніж у методів на основі перцептивних гешів. Отже, запропонований дескриптор пропонує аналогічний час побудови й значно швидший час виконання пошуку порівняно з дескрипторами на основі глибокого навчання, які останнім часом вважаються передовими у сфері пошуку зображень за вмістом. Це робить його ідеальним рішенням для ефективною попередньої фільтрації та швидкого пошуку в масштабних архівах.

Крім того, визначено такі напрями для подальших досліджень:

- розроблення метрик і алгоритмів пошуку для створення системи CBIR з використанням розробленого дескриптора на основі об'єктів;
- поєднання розробленого дескриптора з більш складними алгоритмами пошуку для уточнення топових результатів пошуку;
- виконання завдання пошуку зображень з порівнянням кожного об'єкта найбільш схожих зображень для більш точного ранжування близьких результатів;
- покращення використання паралельних і розподілених обчислень у побудові та порівнянні дескрипторів.

Дескриптор на основі об'єктів балансує ефективність і точність, застосовуючи переваги наявних методів з метою створення надійного й масштабованого рішення для завдань пошуку зображень.

Конфлікт інтересів

Автори заявляють, що не мають конфлікту інтересів, зокрема фінансового, особистого, авторського чи будь-якого іншого характеру, який міг би вплинути на дослідження, а також на результати, опубліковані в цій статті.

Фінансування

Це дослідження було проведено без фінансової підтримки.

Доступність даних

Робота має супутні матеріали в репозиторії даних. Вихідний код програмного забезпечення буде доступний за обгрутованим запитом.

Використання засобів штучного інтелекту

Автори підтверджують, що не застосовували технології штучного інтелекту для написання цієї роботи.

References

- Gonzalez, R. C., Woods, R. E. (2018), "Digital Image Processing", 4th. ed. Pearson/Prentice Hall. 1168 p. DOI/ISBN: 9780133356724
- Liu, X. et al. (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>
- Zheng, Y. et al. (2018), "Size-Scalable Content-Based Histopathological Image Retrieval From Database That Consists of WSIs", *IEEE Journal of Biomedical and Health Informatics*, Vol. 22, No. 4, P. 1278–1287. DOI: <https://doi.org/10.1109/JBHI.2017.2723014>
- Padma, Y. (2024), "Advancements in Non-Linear Content-Based Image Retrieval (CBIR) Systems for Image Analysis", *Communications on Applied Nonlinear Analysis*, Vol. 31, No. 2s, P. 253–265. DOI: <https://doi.org/10.52783/cana.v31.639>
- Wang, X. et al. (2020), "A Storage Method for Remote Sensing Images Based on Google S2", *IEEE Access*. Vol. 8, P. 74943–74956. DOI: <https://doi.org/10.1109/ACCESS.2020.2988631>
- Yang, W. et al. (2020), "Design of Intelligent Search Engine Service Performance Evaluation System", *5th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS)*, P. 86–91. DOI: <https://doi.org/10.1109/ACIRS49895.2020.9162611>
- Žižakić, N., 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>
- Amorós, F. et al. (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>
- Staszewski, P. et al. (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>
- Suárez, I., Buenaposada, J. M., Baumela, L. (2021) "Revisiting Binary Local Image Description for Resource Limited Devices", *IEEE Robotics and Automation Letters*, Vol. 6, No. 4, P. 8317–8324. DOI: <https://doi.org/10.1109/LRA.2021.3107024>
- Wang, B. et al. (2023), "Multi-level adversarial attention cross-modal hashing", *Signal Processing: Image Communication*, Vol. 117, 117017, ISSN 0923-5965. DOI: <https://doi.org/10.1016/j.image.2023.117017>
- Reato, T., Demir, B., Bruzzone L. (2019), "An Unsupervised Multicode Hashing Method for Accurate and Scalable Remote Sensing Image Retrieval", *IEEE Geoscience and Remote Sensing Letters*, Vol. 16, No. 2, P. 276–280. DOI: <https://doi.org/10.1109/LGRS.2018.2870686>
- Zhu, Y., Li, Y., Wang, S. (2019), "Unsupervised Deep Hashing With Adaptive Feature Learning for Image Retrieval", *IEEE Signal Processing Letters*, Vol. 26, No. 3, P. 395–399. DOI: <https://doi.org/10.1109/LSP.2019.2892233>
- Bai, J. et al. (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>
- Zhou, Z. et al. (2020), "Integrating SIFT and CNN Feature Matching for Partial-Duplicate Image Detection", *IEEE Transactions on Emerging Topics in Computational Intelligence*, Vol. 4, No. 5, P. 593–604. DOI: <https://doi.org/10.1109/TETCI.2019.2909936>
- Akita, K., Ukita, N. (2023) "Context-Aware Region-Dependent Scale Proposals for Scale-Optimized Object Detection Using Super-Resolution", *IEEE Access*, Vol. 11, P. 122141–122153. DOI: <https://doi.org/10.1109/ACCESS.2023.3329302>
- Rao, Y. et al. (2023), "Learning Enriched Feature Descriptor for Image Matching and Visual Measurement", *IEEE Transactions on Instrumentation and Measurement*, Vol. 72, P. 1–12. DOI: <https://doi.org/10.1109/TIM.2023.3249237>
- Migliorati, A. et al. (2021), "Learnable Descriptors for Visual Search", *IEEE Transactions on Image Processing*, Vol. 30, P. 80–91. DOI: <https://doi.org/10.1109/TIP.2020.3031216>
- Zhang, Y. (2024), "Graph Convolution Based Efficient Re-Ranking for Visual Retrieval", *IEEE Transactions on Multimedia*, Vol. 26, P. 1089–1101. DOI: <https://doi.org/10.1109/TMM.2023.3276167>
- Iida, K., Kiya, H. (2020), "Privacy-Preserving Content-Based Image Retrieval Using Compressible Encrypted Images", *IEEE Access*, Vol. 8, P. 200038–200050. DOI: <https://doi.org/10.1109/ACCESS.2020.3035563>
- Brogan, J. et al. (2021), "Fast Local Spatial Verification for Feature-Agnostic Large-Scale Image Retrieval", *IEEE Transactions on Image Processing*, Vol. 30, P. 6892–6905. DOI: <https://doi.org/10.1109/TIP.2021.3097175>
- Imbriaco, R. et al. (2022), "Toward Multilabel Image Retrieval for Remote Sensing", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 60, P. 1–14. DOI: <https://doi.org/10.1109/TGRS.2021.3095957>

23. Zhao, Z. -Q. et al. (2019), "Object Detection With Deep Learning: A Review", *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 30, No. 11, P. 3212–3232. DOI: <https://doi.org/10.1109/TNNLS.2018.2876865>
24. Xu, Y., Zhao, X., Gong, J. (2019), "A Large-Scale Secure Image Retrieval Method in Cloud Environment", *IEEE Access*, Vol. 7, P. 160082–160090. DOI: <https://doi.org/10.1109/ACCESS.2019.2951175>
25. Cheng, M. -L., Matsuoka, M. (2020), "An Enhanced Image Matching Strategy Using Binary-Stream Feature Descriptors", *IEEE Geoscience and Remote Sensing Letters*, Vol. 17, No. 7, P. 1253–1257. DOI: <https://doi.org/10.1109/LGRS.2019.2943237>
26. Clark, C. N. et al. (2023), "Investigating the Resolution-Performance Trade-off of Object Detection Models in Support of the Sustainable Development Goals", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 16, P. 5695–5713. DOI: <https://doi.org/10.1109/JSTARS.2023.3284309>
27. Sun, C. et al. (2023), "Multi-label adversarial fine-grained cross-modal retrieval", *Signal Processing: Image Communication*, Vol. 117, 117018, ISSN 0923-5965, DOI: <https://doi.org/10.1016/j.image.2023.117018>
28. Radenović, F., Tolias, G., 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>
29. Vynokurova O., Peleshko D., Peleshko M. (2020), "Hybrid Deep Convolutional Neural Network with Multimodal Fusion", *Data Stream Mining & Processing. DSMP 2020. Communications in Computer and Information Science*, Vol. 1158, P. 62–78. DOI: https://doi.org/10.1007/978-3-030-61656-4_4
30. Galvez, R. L. et al. (2018), "Object Detection Using Convolutional Neural Networks", *TENCON 2018 – 2018 IEEE Region 10 Conference*, P. 2023–2027. DOI: <https://doi.org/10.1109/TENCON.2018.8650517>
31. Fang, F. et al. (2020), "Combining Faster R-CNN and Model-Driven Clustering for Elongated Object Detection", *IEEE Transactions on Image Processing*, Vol. 29, P. 2052–2065. DOI: <https://doi.org/10.1109/TIP.2019.2947792>
32. Sun, C.-Y. et al. (2021), "Cascade Faster R-CNN Detection for Vulnerable Plaques in OCT Images", *IEEE Access*, Vol. 9, P. 24697–24704. DOI: <https://doi.org/10.1109/ACCESS.2021.3056448>
33. Dhayne, H. et al. (2021), "In Search of Big Medical Data Integration Solutions – A Comprehensive Survey", *IEEE Access*, Vol. 7, P. 91265–91290. DOI: <https://doi.org/10.1109/ACCESS.2019.2927491>
34. Kathiravelu, P., Sharma, A., Sharma, P. (2021), "Understanding Scanner Utilization With Real-Time DICOM Metadata Extraction", *IEEE Access*, Vol. 9, P. 10621–10633. DOI: <https://doi.org/10.1109/ACCESS.2021.3050467>
35. Ebrahimian, A. et al. (2020), "Class-Aware Image Search for Interpretable Cancer Identification", *IEEE Access*, Vol. 8, P. 197352–197362. DOI: <https://doi.org/10.1109/ACCESS.2020.3033492>
36. Wolf, T. et al. (2020), "Transformers: State-of-the-Art Natural Language Processing", *Natural Language Processing: System Demonstrations*, P. 38–45. DOI: <https://doi.org/10.18653/v1/2020.emnlp-demos.6>
37. Dosovitskiy, A. et al. (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>

Received (Надійшла) 22.01.2026

Accepted for publication (Прийнята до друку) 14.02.2026

Publication date (Дата публікації) 12.03.2026

Відомості про авторів / About the Authors

Прокопенко Олександр Сергійович – Харківський національний університет радіоелектроніки, аспірант кафедри програмної інженерії; Харків, Україна;

Oleksandr Prokopenko – Kharkiv National University of Radio Electronics, Postgraduate Student at the Department of Software Engineering; Kharkiv, Ukraine;

e-mail: oleksandr.prokopenko1@nure.ua

ORCID ID: <http://orcid.org/0000-0003-0489-6820>

Смеляков Сергій Вячеславович – доктор фізико-математичних наук, професор, Харківський національний університет радіоелектроніки, професор кафедри програмної інженерії; Харківський національний університет Повітряних Сил імені Івана Кожедуба, професор кафедри математичного та програмного забезпечення АСУ; Харків, Україна;

Sergii Smelyakov – Doctor of Physical and Mathematical Sciences, Professor, Kharkiv National University of Radio Electronics, Professor at the Software Engineering Department; Ivan Kozhedub Kharkiv National Air Force University, Professor at the Department of Mathematics and Software for Automated Control Systems; Kharkiv, Ukraine;

e-mail: serhii.smeliakov@nure.ua

ORCID ID: <http://orcid.org/0000-0002-5791-2479>

DEVELOPMENT OF AN IMAGE DESCRIPTOR FOR EFFECTIVE CONTENT-BASED RETRIEVAL IN LARGE DATA REPOSITORIES

The object of the research is content-based image retrieval. *The subject of this research* is models and methods of content-based image retrieval (CBIR). *The purpose of the research* is the development of an image descriptor that uses advanced computer vision models for object detection. The proposed descriptor increase the efficiency and accuracy of image search and management processes. *The objectives of the study* include: analysis of modern approaches and solutions for creating image descriptors and their use in content-based image retrieval; identification of the limitations of existing image descriptors and determination of requirements for a new, improved descriptor; development of a new image descriptor that effectively uses information about detected objects; conducting experiments to evaluate the proposed descriptor and compare its effectiveness with existing solutions. *The methodology* includes conducting a comprehensive review of state-of-the-art methods for creating image descriptors; analysis of the use of existing descriptors in CBIR systems; development of a descriptor that utilizes modern object detection models; validation of the effectiveness of the proposed descriptor in comparison with existing descriptors on retrieval tasks. The following *results* were obtained: an image descriptor based on objects detected using modern machine learning models was developed; a series of experiments was conducted to evaluate the efficiency and quality of retrieval in large image storage systems using the proposed descriptor. The experiments revealed the following advantages: faster creation and comparison of descriptors than hashed, handcrafted, and deep learning-based descriptors; effective filtering of images in the repository based on object content. However, the effectiveness of the descriptor depends on the quality of the model and the data, since images without detected objects do not appear in the search results, which may limit recall. *Conclusions*: the developed image descriptor is an effective tool for solving a range of content-based image retrieval tasks due to the speed of construction and search. A promising direction for further research is the development of a content-based image retrieval system using the proposed descriptor, with enhanced use of parallel and distributed computing.

Keywords: machine learning; image descriptors; image processing; image search optimization.

Бібліографічні описи / Bibliographic descriptions

Прокопенко О. С., Смеляков С. В. Розроблення дескриптора зображень для ефективного пошуку на основі вмісту у великих сховищах даних. *Автоматизовані системи управління та прилади автоматики*. 2026. № 1 (188). С. 129–144. DOI: <https://doi.org/10.30837/0135-1710.2026.188.129>

Prokopenko, O., Smelyakov, S. (2026), "Development of an image descriptor for effective content-based retrieval in large data repositories", *Management Information System and Devices*, No. 1 (188), P. 129–144. DOI: <https://doi.org/10.30837/0135-1710.2026.188.129>
