



УДК 004.92

METHODS FOR DETERMINING VISUAL SIMILARITY OF IMAGES**МЕТОДИ ВИЗНАЧЕННЯ ВІЗУАЛЬНОЇ ПОДІБНОСТІ ЗОБРАЖЕНЬ****Porublov K.Y. / Порубльов К.Ю.***student of magistracy / студент магістратури***Bushyn I.M. / Бушин І.М.***Cand. Sc. (Phys.-Math.), Associate Professor / к.ф.-м.н., доцент*

ORCID: 0009-0003-1385-6924

*Черкаський національний університет імені Богдана Хмельницького,**Черкаси, Бульвар Шевченка, 81, 18031*

Анотація. У статті розглядаються наявні підходи та методи визначення рівня візуальної подібності зображень. Визначено різні види візуальної подібності зображень. Проаналізовано особливості кожного з підходів, визначено їх переваги та недоліки, оцінено їх ефективність при роботі з визначеними категоріями подібності зображень. Також у статті наведено результати експерименту, проведеного для перевірки зроблених припущень.

Ключові слова: обробка зображень, нейронна мережа, комп'ютерний зір, ознаки зображень

Вступ.

У сучасному інформаційному середовищі візуальний контент становить значну частку всієї інформації, що обробляється в Інтернеті. Це зумовлює створення розумних систем пошуку, які дозволяють користувачеві здійснювати пошук, використовуючи зображення, повертаючи результати на основі візуальної подібності вмісту до зображення-запиту.

Метою цієї статті є дослідження наявних методів визначення візуальної подібності зображень та порівняння їх ефективності та швидкодії при роботі з вибірками зображень значного розміру (10 тис. зображень).

Критерії подібності зображень. Термін «візуальна подібність зображень» не має єдиного визначення, отже, критерії, за якими пара зображень може вважатися подібною, залежатимуть від конкретної задачі. Критерії подібності можна класифікувати на три загальні категорії:

- майже ідентичні – дублікати з мінімальними перетвореннями;
- містять ідентичні образи – зображення, що містять однакові символи, логотипи тощо;
- містять подібні за сутністю об'єкти – зображення з візуально різними, але семантично подібними об'єктами.



Приклади пар зображень, що відповідають описаним критеріям, наведено на рис. 1.



Майже ідентичні зображення



Зображення, що містять ідентичні образи



Зображення, що містять подібні за сутністю образи

Рисунок 1 - Категорії подібності зображень

Підходи до визначення схожості зображень. Наявні алгоритми визначення візуальної подібності зображень беруть за основу один з трьох підходів: перцептивне ґешування, співставлення ключових точок та порівняння векторів ознак, обчислених згортковими нейронними мережами (англ. *Convolutional Neural Networks*, далі – *CNN*).

Алгоритми на основі перцептивного ґешування обчислюють для вхідного зображення ґеш-значення, що складається з малої кількості байтів. Подібність зображень можна визначити як відстань Гемінга між їх ґеш-значеннями. Обчислення ґеш-значення виконується за допомогою дискретного косинусного перетворення (англ. *Discrete Cosine Transform*, *DCT*) вхідного зображення [1]. Для зображення розміром $N \times N$, DCT-матрицю M можна обчислити за формулою:

$$M[n, m] = \sqrt{\frac{2}{N}} * \cos\left(\frac{(2m + 1) * n\pi}{2N}\right), \quad (1)$$

де $m, n \in [0, N)$.

Отриману матрицю M можна обчислити зарані і застосувати до вхідного зображення за формулою:

$$DCT(I) = M * I * M', \quad (2)$$

де I – вхідне зображення розміру $N \times N$.

В результаті отримуємо матрицю коефіцієнтів розміром $N \times N$, де значення, ближчі до позиції $(0,0)$, відповідають більш значущим компонентам. Коефіцієнти, достатньо далекі від верхнього лівого кута, відкидаються, таким



чином кінцеве геш-значення формується лише на основі значимих коефіцієнтів. Відкидання малозначимих коефіцієнтів забезпечує стійкість геш-значень до малих змін у зображенні, таких як шуми чи зміна яскравості. Для підвищення швидкодії перед застосуванням DCT зображення масштабується до малого розміру, наприклад 32x32 пікселів.

Цей підхід забезпечує стійкість до незначних модифікацій зображення, таких як масштабування, шуми або водяні знаки (англ. *watermark*). Водночас цей підхід не здатний розпізнавати подібність у більш складному візуальному контексті, тому його застосування є доречним лише при пошуку майже подібних зображень.

Алгоритми на основі співставлення ключових точок використовують техніки машинного зору для виявлення у зображенні характерних точок, що описуються координатами та дескриптором – об'єктом, сформованим на основі пікселів навколо точки. Структура і спосіб формування дескриптора залежать від конкретного алгоритму. Серед поширених алгоритмів виявлення ключових точок можна виділити SIFT [2], ORB [3], BRIEF [4].

Перевагою цього підходу є здатність алгоритмів виділяти ознаки на основі невеликих фрагментів зображення, що дозволяє виявляти подібні символи або візерунки на зображеннях, навіть якщо вони знаходяться у різному візуальному контексті. Недоліком є значно вища обчислювальна складність у порівнянні з алгоритмами перцептивного гешування.

На рис. 2 зображено приклад співставлення ключових точок, виявлених алгоритмом ORB. Варто зазначити, що хоча алгоритм коректно зіставив ключові точки навколо логотипу на обох зображеннях, чіткі збіги склали малу частину всіх ключових точок, виявлених у парі зображень. Більшість ключових точок у парі зображень не є подібними через відмінності в інших елементах зображення.

Алгоритми на основі CNN покладаються на заздалегідь натреновані згорткові нейронні мережі для виявлення ознак зображення. На рис. 3 зображено узагальнену структуру CNN, що виконує задачу мультикласової класифікації.



Частка чітких відповідностей: 0.25 (126 / 500)



Рисунок 2 - Результат співставлення ключових точок

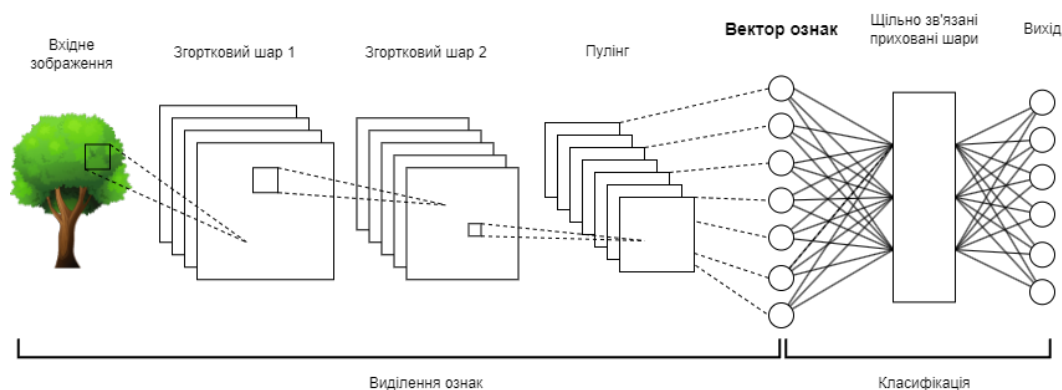


Рисунок 3 - Структура згорткової нейронної мережі

На етапі виділення ознак нейромережа формує вектор ознак – числовий вектор, що представляє зображення у багатовимірному просторі. Під час тренування нейромережі отриманий вектор використовується як вхідні дані до багатoshарового перцептрон, який виконує задачу класифікації зображення. При порівнянні зображень за допомогою CNN етап класифікації відкидається, натомість подібність зображень визначається як відстань або косинусна подібність (англ. *cosine similarity*) між їх векторами ознак.

Цей підхід дозволяє виявляти подібність між зображеннями, що містять



подібні за сутністю образи, навіть якщо ознаки низького рівня, такі як контури, є повністю відмінними. Варто зазначити, що точність порівнянь при використанні CNN залежить від її архітектури, а також об'єму та особливостей навчальної вибірки. У цій роботі використано моделі, навчені на вибірках загального призначення: Google ViT, OpenAI CLIP, Timm.

Порівняння точності та швидкодії підходів. Для порівняння можливостей описаних підходів при роботі з різними видами подібності зображень було проведено експеримент на основі трьох вибірок даних. Вибірka LogoDet-3k [5] складається з груп зображень, що містять ідентичні образи. Вибірka GPR1200 [6] складається з груп зображень з подібними за сутністю об'єктами. Вибірku з майже ідентичних зображень (далі – дублікатів) було сформовано автоматично на основі зображень з GPR1200 та LogoDet-3k. Кожна група в цій вибірці містить оригінальне зображення та його копії, до яких було застосовано різні трансформації: обертання, віддзеркалення, додавання JPEG-артефактів, зміна яскравості, зміна масштабу зображення. Розмір кожної вибірки було обмежено до 10000 зображень.

Експеримент складається з двох частин. Під час першої частини алгоритм обчислює та кешує ознаки всіх зображень у вибірці, не виконуючи при цьому операцій порівняння. В результаті отримано швидкість обробки зображень у вибірках. У другій частині експерименту було визначено попарну точність (англ. *pairwise accuracy*) порівняння ознак. Для цього з кожної групи у вибірці було обрано «базове» зображення та 5 пар, що складаються з подібного та відмінного зображень.

Якщо алгоритм стверджує, що в черговій парі базове зображення є більш подібним на відмінне зображення, це вважається негативним результатом, інакше – позитивним. Тоді попарна точність обчислюється за формулою:

$$A = \frac{N_{positive}}{N_{positive} + N_{negative}} \quad (3)$$

Результати експерименту. У таблиці наведено результати експерименту для ряду алгоритмів для кожної з вибірок.



Таблиця 1 - Швидкість та точність алгоритмів порівняння за візуальною подібністю зображень.

Назва методу	Вибірка	Швидкість виявлення ознак (зображень/с)	Швидкість порівняння ознак (пара ознак/с)	Попарна точність
SIFT	Дублікати	20,57	215,40	0,930
	LogoDet-3k	12,30	364,36	0,651
	GPR1200	21,73	444,86	0,685
ORB	Дублікати	65,55	449,19	0,924
	LogoDet-3k	63,42	520,92	0,636
	GPR1200	73,11	525,82	0,652
pHash	Дублікати	77,93	1348,90	0,653
	LogoDet-3k	68,86	1369,49	0,526
	GPR1200	79,13	1423,02	0,485
ViT	Дублікати	20,65	1060,63	0,996
	LogoDet-3k	20,18	893,20	0,981
	GPR1200	20,35	1342,30	0,821
CLIP	Дублікати	18,71	1102,81	0,999
	LogoDet-3k	18,08	1101,03	0,992
	GPR1200	18,84	1372,625	0,964
Timm	Дублікати	41,84	1169,17	0,996
	LogoDet-3k	35,99	801,96	0,974
	GPR1200	40,92	1530,26	0,773

Бачимо, що перцептивне ґешування демонструє найвищу швидкодiю серед усіх алгоритмiв, але низькi показники точностi навiть при роботi з вибiркою дублiкатiв. Водночас швидкiсть порiвняння ознак пiсля кешування є дуже високою. Алгоритми, що базуються на спiвставленнi ключових точок, продемонстрували найнижчi результати. Їх швидкодiя близька до нейромережевих моделей, але точнiсть є значно нижчою.

Висновки.

Було проаналiзовано наявнi методи порiвняння зображень за вiзуальною подiбнiстю, описано їх принцип роботи, переваги та недолiки. Також було



описано три категорії критеріїв подібності зображень та проведено експеримент з метою визначення швидкодії та точності розглянутих методів при роботі з вибірками даних, що містять групи зображень, подібних відповідно до описаних критеріїв. В результаті експерименту було визначено, що нейромережевий підхід є найбільш точним та забезпечує високу швидкодію за умови кешування ознак зображень. Алгоритми на основі співставлення ключових точок мають нижчу точність при близькій швидкодії. Перцептивне гешування показало високу швидкодію при виявленні та порівнянні ознак при низькій точності. Подальші дослідження плануються спрямувати на детальніше оцінювання точності розглянутих алгоритмів.

Література:

1. C. Zauner, «Implementation and Benchmarking of Perceptual Image Hash Functions,» Hagenberg, 2010.
2. D. G. Lowe, «Object Recognition from Local Scale-Invariant Features,» в Proc. of the International Conference on Computer Vision, Corfu, 1999.
3. E. Rublee, V. Rabaud, K. Konolige та G. Bradski, «ORB: an efficient alternative to SIFT or SURF,» 2011.
4. M. Calonder, V. Lepetit, C. Strecha and P. Fua, “BRIEF: Binary Robust Independent Elementary Features,” CVLab, EPFL, Lausanne, Switzerland, 2010.
5. lyly99, «Kaggle,» 22 Вересень 2020. [Онлайновий]. Available: <https://www.kaggle.com/datasets/lyly99/logodet3k>. [Дата звернення: 11 09 2025].
6. K. Schall, K. U. Barthel, N. Hezel та K. Jung, «GPR1200: A Benchmark for General-Purpose Content-Based Image Retrieval,» в MultiMedia Modeling: 28th International Conference, MMM 2022, Phu Quoc, Vietnam, June 6–10, 2022, Proceedings, Part I, Phu Quoc, Vietnam, 2022.

Abstract. This article examines existing approaches and methods for determining the level of visual similarity between images. Different types of visual image similarity are defined. The characteristics of each approach are analyzed, their advantages and disadvantages are identified, and their effectiveness is evaluated when applied to specific categories of image similarity. The article also presents the results of an experiment conducted to verify the assumptions made.

Keywords: image processing, neural network, computer vision, image features