

УДК 004.932.2

DOI: [10.20535/2312-1807.2017.22.1.80745](https://doi.org/10.20535/2312-1807.2017.22.1.80745)Дрозд В. П., аспірант, OrcID [0000-0002-8701-678X](https://orcid.org/0000-0002-8701-678X)e-mail: drozdvadym@gmail.com

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»,

ПОРІВНЯЛЬНИЙ АНАЛІЗ ЗОБРАЖЕНЬ ЗА ДОПОМОГОЮ ГЛИБИННИХ НЕЙРОННИХ МЕРЕЖ

Побудова стабільної системи для виявлення коефіцієнту подібності зображень є досить складною задачею. Насамперед необхідно враховувати відмінності між різними класами об'єктів, що можуть бути на зображенні, а також відмінності між внутрішньокласовими представниками. В даній роботі розглянуто підхід до навчання системи порівняння зображень, що використовує техніку глибинного навчання, а саме об'єднання усіх етапів порівняння в одну глибинну нейронну мережу. Даний підхід дозволяє побудувати систему, що має набагато більшу навчальну ємність в порівнянні з іншими системами. Експериментальна перевірка системи проведена на тестових зображеннях, що представляють собою реальні фото одягу. Результати підтверджують, що підхід об'єднання системи в одну глибинну нейронну мережу дозволяє підвищити якість в порівнянні з іншими підходами, а також дозволяє зменшити розмір отриманого вектору ознак.

Бібл. 16, рис. 1, табл. 2.

Ключові слова: цифрові зображення; глибинні нейронні мережі; порівняння зображень; дескриптори; ознаки.

Вступ. Порівняння зображень і їх частин є однією з фундаментальних задач в сфері комп'ютерного зору і аналізу цифрових зображень. Величезна кількість алгоритмів побудовано саме на порівнянні зображень, серед них можна виділити такі як: структура з руху [1], загальна реєстрація зображень, побудова панорам, пошук схожих зображень, класифікація і категоризація зображень. Звичайно, проблема порівняння зображень є досить складною, оскільки існує багато факторів, які впливають на фінальне представлення зображення. Вони можуть включати в себе зміну точки спостереження, варіації освітлення сцен, перекриття об'єктів на зображенні, тіні, різні налаштування камери і т.д. Більшість існуючих методів порівняння зображень використовує порівняння за категорією. Наприклад, в роботах [2] і [3] зображення вважаються схожими, якщо вони належать до одного класу. Такий підхід може бути використаний в разі пошуку зображень схожих за змістом, але не надає можливості визначити міру схожості між зображеннями однієї категорії.

Одним із підходів до побудови системи порівняння зображень є виділення певних особливостей з зображень, таких як SIFT [4], HOG [5] та інші, і потім навчання певної моделі для порівняння, основаної на отриманих ознаках, як наприклад описано в роботі [6]. Зазвичай ефективність даних підходів обмежується репрезентативною потужністю вибраних ознак. Іншим

недоліком даних підходів є розділення на два незв'язані етапи, а саме виділення ознак і навчання класифікатора. Таким чином ми припускаємо, що об'єднання цих двох етапів в один дозволить збільшити ефективність і якість системи. Для підтвердження цього припущення вирішено використати техніку глибинного навчання.

Глибинні нейронні мережі (ГНМ) досягли надзвичайних показників якості в задачах класифікації зображень [7]. Глибокі згорткові нейронні мережі представляють собою набір зв'язаних між собою згорткових і повнозв'язаних шарів. Також можуть використовуватися зв'язані вагові коефіцієнти та агрегувальні шари. Ця архітектура дозволяє ГНМ отримувати переваги від двовимірної структури вхідних даних, якими і є зображення. Та задача класифікації дещо відрізняється від задачі порівняння зображень. Оскільки для системи класифікації «червоний автомобіль», «білий автомобіль» і «жовтий автомобіль» належать до класу «автомобіль», а в задачі порівняння зображень потрібно для двох автомобілів схожого кольору поставити у відповідність більший коефіцієнт подібності, ніж для автомобілів різних кольорів. Таким чином, існуючі моделі повинні бути розширені для можливості коректного порівняння зображень.

Отже порівняння зображень є важливою прикладною задачею для додатків систем технічного зору, що диктує актуальність роботи



в напрямку побудови нових та покращення існуючих підходів до порівняння зображень. Метою даної роботи є побудова та експериментальна перевірка системи порівняння зображень на основі глибинних нейронних мереж.

Постановка задачі. Задачею дослідження є порівняння зображень, отже в якості вхідних даних маємо два зображення: I_1 та I_2 . В результаті порівняння необхідно отримати певну міру подібності двох зображень. Виберемо шкалу від 0 до 1, де 1 – це максимально схожі зображення, 0 – це зовсім різні зображення. При такій шкалі прийнято вважати схожими усі зображення, для яких міра подібності більша від 0.5. Таким чином потрібно побудувати систему S , яка для двох вхідних зображень I_1 та I_2 буде визначати міру їх подібності M в межах $[0; 1]$, тобто: $M = S(I_1, I_2)$.

Пов'язані роботи та методи. Найпростішим підходом є використання прямого порівняння пікселів між двома зображеннями, як то сумарна різниця між інтенсивністю пікселів:

$$D(I_1, I_2) = \sum_d |I_{1(d)} - I_{2(d)}|$$

де I_1 – перше зображення для порівняння, I_2 – друге зображення для порівняння, d – індекс пікселя.

Такий підхід використовується не дуже часто, оскільки він є надзвичайно чутливим до змін у зображенні, як геометричних так і умов отримання зображення.

Іншим, більш популярним підходом, є використання нормалізованої кореляції:

$$NC(I_1, I_2) = \frac{\sum_d (I_{1(d)} - \bar{I}_1)(I_{2(d)} - \bar{I}_2)}{\sigma_1 \sigma_2}$$

де I_1, I_2 – зображення для порівняння, d – індекс пікселя, \bar{I}_1, \bar{I}_2 – середні арифметичні значення інтенсивності пікселів кожного із зображень, σ_1, σ_2 – девіація інтенсивності пікселів.

Даний метод є більш ефективним, оскільки враховуються середні значення інтенсивності і девіації. Зазвичай використовується в розпізнаванні по фрагментам [10], але даний метод також є дуже чутливим до геометричних спотворень.

Ще одним традиційним підходом до порівняння зображень є використання дескрипторів і евклідової відстані між ними. В якості дескрип-

торів найчастіше використовують SIFT, SURF, MSER [8] або DAISY [9]. Та дані методи мають обмеження закладене в репрезентативну можливість вибраного дескриптору.

В роботі [12] аналізується ефективність згорткових дескрипторів отриманих з глибинної нейронної мережі, а також проводиться їх порівняння з ознаками SIFT. Показано, що в більшості випадків саме глибокі нейронні мережі показують кращу якість.

Таким чином вирішено зупинитися на використанні техніки глибинного навчання [7].

Архітектура обраної глибинної нейронної мережі. Вхідними даними нейронної мережі є пара зображень. Модель не передбачає певних обмежень на кількість каналів вхідного зображення, що дозволяє працювати як з кольоровими так і сірими зображеннями. Для оцінки якості моделей вирішено використовувати реальні фото одягу, що дозволить перевірити чутливість моделей до варіативності форми та кольору об'єктів на зображеннях. Часто інформація про колір дозволяє отримати набагато точнішу оцінку схожості зображень, тому вирішено обрати модель з каналами RGB. В багатьох сучасних архітектурах ГНМ в якості вхідних даних використовують зображення 256 на 256 пікселів. В своїй моделі ми використовуємо зображення яке складається з пари зображень розмірами 128 на 256 пікселів, оскільки фото одягу зазвичай має співвідношення 1:2.

Існує велика кількість підходів до побудови мережі, що опрацьовуватиме пару зображень. В роботі вирішено розглянути два підходи, які на нашу думку якнайкраще пристосовані до нашої задачі, а саме: «об'єднану» і «siamese» архітектури.

«Об'єднана» архітектура. Цей тип нейронної мережі передбачає попередню обробку вхідних зображень, а саме об'єднання пари в одне зображення. Таким чином за основу можна використовувати будь-яку з доступних архітектур ГНМ. Вся робота на визначення міри схожості двох частин зображення покладається на алгоритм навчання. Явним недоліком даної архітектури є неможливість роздільного обчислення дескриптору для кожного із зображень, що в реальних системах призводить до збільшення часу роботи.

В якості архітектури глибинної нейронної мережі вибрано AlexNet, про яку детально описано в роботі [7]. Архітектура представляє собою нейронну мережу, що складається з 8 шарів: 5 згорткових і 3 повнозв'язних. Для застосування даної мережі до нашої задачі слід заміни-

ти 3 повнозв'язні шари 2 шарами, кількість виходів останнього шару встановити в 1 і використати сигмоїдальну активаційну функцію. Надалі отриману архітектуру іменуватимемо AlexNetUnion. Детальний опис конфігурації мережі, кількості нейронів та активаційних функцій наведено в табл. 1.

«Siamese» архітектура. Цей тип архітектури повністю повторює підхід з дескрипторами. В

архітектурі присутні дві окремі нейронні мережі (НМ), що мають абсолютно однакові конфігурації і набір вагових коефіцієнтів. Кожна з НМ приймає на вхід одне зображення з пари і обраховує свій вихід. Виходи двох нейронних мереж об'єднуються декількома повнозв'язними шарами, які і обраховують міру схожості зображень.

Таблиця 1.

Деталізована конфігурація мережі AlexNetUnion

Шар НМ	Розмір вхідних даних	Параметри шару	Активаційна функція
Згортковий шар-1	256x256x3	96 ядр розміром 11x11x3 з кроком 4 пікселі	ReLU
Шар субдискретизації-1 по максимуму	62x62x96	розмір ядра 3x3, крок - 2	-
Згортковий шар-2	31x31x96	256 ядр розміром 5x5x48 з кроком 1 піксель	ReLU
Шар субдискретизації-2 по максимуму	31x31x256	розмір ядра 3x3, крок - 2	-
Згортковий шар-3	15x15x256	384 ядр розміром 3x3x256 з кроком 1 піксель	ReLU
Згортковий шар-4	15x15x384	384 ядр розміром 3x3x192 з кроком 1 піксель	ReLU
Згортковий шар-5	15x15x384	256 ядр розміром 3x3x192 з кроком 2 пікселі	ReLU
Шар субдискретизації-5 по максимуму	15x15x256	розмір ядра 3x3, крок - 2	-
Повнозв'язний шар-1	7x7x256		ReLU
Повнозв'язний шар-2	1x4096		сигмоїдальна

Кожна з цих двох нейронних мереж може розглядатися як певний дескриптор (вектор ознак) що обраховується для кожного зображень, а повнозв'язні шари можуть розглядатися як функція обрахування відстані між дескрипторами. Основним недоліком даного підходу є збільшення складності обрахунку. Основною перевагою є незалежність обрахування ознак кожною з двох нейронних мереж, що дає змогу для попереднього обрахунку дескрипторів для певної групи зображень і в подальшому обрахунку тільки результату з повнозв'язних шарів.

В якості згорткової нейронної мережі для «siamese» архітектури теж обрано AlexNet. Для

кожної з двох мереж слід відкинути повнозв'язні шари. Об'єднати виходи двох мереж потрібно двома повнозв'язними шарами, кількість виходів останнього шару встановити в 1 і використати сигмоїдальну активаційну функцію. Надалі дану архітектуру іменуватимемо AlexNetSiamese. Архітектура кожної з двох нейронних мереж аналогічна мережі AlexNetUnion з шару «Згортковий-1» до шару «Субдискретизації-5». Відмінність полягає у розмірах результуючих матриць та вхідного зображення.

Навчання мережі. Навчання ГНМ здійснюється за допомогою стандартного алгоритму зворот-

ного поширення помилки. Уточнення вагових коефіцієнтів може здійснюватися стохастичним найшвидшим спуском. [14]

Навчання досліджуваних моделей проводилося строго в режимі навчання з учителем зі значеннями на виході нейронної мережі в межах $[0, 1]$, відповідно 0 для зображень, що не є схожими і 1 – для схожих зображень. В якості методу оптимізації використано метод стохастичного найшвидшого спуску без модифікацій. Параметр регулювання величини кроку навчання рівний 0.01. Зменшення регулювання величини навчання на 1 порядок відбувається кожні 30 епох. В якості критерію зупинки використано навчання протягом 90 епох. Початкова ініціалізація вагових коефіцієнтів довільна з нормальним розподілом, середньоквадратичне відхилення рівне 0.01. Розмір пачки зображень рівний 128. Частину параметрів для навчання обрано відповідно до рекомендацій наданих в роботі [7], а кількість епох, розмір пачки для обчислення градієнту та зміну параметру регулювання навчання вибрано експериментальним шляхом.

Для генерування більшої кількості вхідних даних застосовувалися наступні техніки перетворення зображень: дзеркальний розворот пари зображень по горизонталі, поворот зображень на 15, 30, 45 градусів. Усі яскравості пікселів зображення масштабовано в інтервал $[0; 1]$. Додаткових попередніх операцій обробки вхідних зображень не проводилося.

В якості інструменту обрано фреймворк Caffe [15]. Навчання проводилося на тестовій машині наступної конфігурації: Intel Core i7, NVIDIA GTX 660 Ti, 16 Gb RAM. Час навчання однієї моделі близько 72 годин.

Експериментальна перевірка. Всі тести проведено на власній вибірці реальних фото одягу. Приклади показано на рис. 1.



Рис. 1. Приклади вхідних зображень

Розмір вибірки сягає 50000 зображень. Для генерування завідомо негативних пар, тобто пар не схожих зображень, були обрані ті пари, для яких розмір нормалізованої кореляції виявився мінімальним. Таким чином в навчальній вибірці отримали 80000 зображень, в тестовій – 20000. В таблиці 2 представлено результати тестування отриманих моделей. Для порівняння було навчено SVM класифікатор поверх ознак SIFT. При тестуванні зображення вважаються схожими, якщо коефіцієнт подібності на виході НМ більший або рівний 0.5.

В ході експериментів було проведено тестування моделей з різними параметрами та з різними конфігураціями деяких шарів нейронної мережі, зміна даних параметрів не призводила до суттєвої зміни точності, тому порівняння проводилося для стандартної конфігурації.

Таблиця 2.

Точність отриманих моделей

Метод	Точність
AlexNetUnion	64.2%
AlexNetSiamese	77.7%
SIFT + SVM	53.8%

З таблиці 2 помітно, що архітектура AlexNetSiamese значно випередила інші в якості, що в першу чергу пов'язано з розділенням етапу виділення ознак зображень на дві окремі нейронні мережі. Для підтвердження правильності навчання дескриптору було досліджено активації різних шарів отриманих нейронних мереж. Для цього методом описаним в роботі

[16] було визначено яким чином кожен з пікселів зображення впливає на активації різних нейронів отриманої мережі. Визначено, що завдяки мережі виділяється інформація про колір, текстуру і контур одягу, що свідчить про високу якість дескриптору.

Висновки. Порівняння зображень і об'єктів на зображеннях є досить складною задачею, яка не має точного рішення на даний час. В даній роботі досліджено новий спосіб побудови системи порівняння зображень з використанням глибоких нейронних мереж. Для підтвердження гіпотези проведено огляд двох різних конфігурацій нейронних мереж. Кращою виявилася архітектура AlexNetSiamese, що насамперед пов'язано з використанням двох незалежних нейронних мереж для кожного із зображень.

Проведено порівняння якості побудованих систем з існуючими підходами до визначення

схожості зображень, а саме до системи, що використовує дескриптор SIFT і класифікатор SVM. Встановлено, що на тестовій вибірці розглянута система показала значно кращий результат. Слід зазначити, що збільшення розміру навчальної вибірки повинно збільшити якість побудованої системи.

Подальші дослідження будуть проводитись у напрямку семантичного порівняння зображень і збільшення швидкодії існуючої системи.

Надійшла до редакції 21 жовтня 2016 р.

Література

1. Structure from motion [Електронний ресурс] – Режим доступу до ресурсу: https://en.wikipedia.org/wiki/Structure_from_motion.
2. R. Hadsell. Dimensionality reduction by learning an invariant mapping / R. Hadsell, S. Chopra, Y. LeCun // Computer vision and pattern recognition. – 2006. - pp. 1735-1742. DOI: [10.1109/CVPR.2006.100](https://doi.org/10.1109/CVPR.2006.100).
3. G. W. Taylor. Learning invariance through imitation / G. W. Taylor, I. Spiro, C. Bregler, and R. Fergus // Computer vision and pattern recognition. - 2011. - pp. 2729–2736. DOI: [10.1109/CVPR.2011.5995538](https://doi.org/10.1109/CVPR.2011.5995538).
4. D. G. Lowe. Object recognition from local scale-invariant features / D. G. Lowe // ICCV. - 1999. - Vol. 2. – pp. 1150–1157. DOI: [10.1109/ICCV.1999.790410](https://doi.org/10.1109/ICCV.1999.790410).
5. N. Dalal. Histograms of Oriented Gradients for Human Detection / N. Dalal. B. Triggs // Computer vision and pattern recognition. - 2005. - pp. 886-893. DOI: [10.1109/CVPR.2005.177](https://doi.org/10.1109/CVPR.2005.177).
6. Y. L. Boureau. Learning mid-level features for recognition / Y. L. Boureau, F. Bach, Y. LeCun, J. Ponce // Computer vision and pattern recognition. - 2010. - pp. 2559–2566. DOI: [10.1109/CVPR.2010.5539963](https://doi.org/10.1109/CVPR.2010.5539963).
7. A. Krizhevsky. Imagenet classification with deep convolutional neural networks / A. Krizhevsky, I. Sutskever, G. Hinton // NIPS. - 2012. - pp. 1106–1114.
8. T. Tuytelaars. Local invariant feature detectors: A survey / T. Tuytelaars, A. Mikolajczyk // Foundations and Trends in Computer Graphics and Vision. - 2007. - Vol.3, No. 3. - pp. 177-280.
9. E. Tola. A Fast Local Descriptor for Dense Matching. / E. Tola, V. Lepetit, P. Fua. // Computer Vision and Pattern Recognition. - 2008. – p. 501. DOI: [10.1109/CVPR.2008.4587673](https://doi.org/10.1109/CVPR.2008.4587673).
10. V. Athitsos. Boostmap: A method for efficient approximate similarity rankings / V. Athitsos, J. Alon, S. Sclaroff, G. Kollios // " IEEE Conference on Computer Vision and Pattern Recognition. – 2004. DOI: [10.1109/CVPR.2004.1315173](https://doi.org/10.1109/CVPR.2004.1315173).
11. A. S. Razavian. CNN features off-the-shelf: An astounding baseline for recognition / A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson // Computer Vision and Pattern Recognition, CVPR. - 2014. - pp. 512–519. DOI: [10.1109/CVPRW.2014.131](https://doi.org/10.1109/CVPRW.2014.131).
12. Fisher P. Descriptor matching with convolutional neural networks: a comparison to SIFT [Електронний ресурс] / P. Fisher, A. Dosovitskiy // arXiv preprint arXiv:1405.5769. – 2014. – Режим доступу до ресурсу: <https://arxiv.org/abs/1405.5769>.
13. C. Szegedy. Going deeper with convolutions [Електронний ресурс] / C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich // arXiv preprint arXiv:1409.4842. – 2014. – Режим доступу до ресурсу: <https://arxiv.org/abs/1409.4842>
14. Глибине навчання [Електронний ресурс] – Режим доступу до ресурсу: https://uk.wikipedia.org/wiki/Глибине_навчання.
15. Y. Jia. Caffe: Convolutional architecture for fast feature embedding [Електронний ресурс] / Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell // arXiv preprint arXiv:1408.5093. – 2014. – Режим доступу до ресурсу: <https://arxiv.org/abs/1408.5093>



16. Zeiler M.D. Visualizing and Understanding Convolutional Networks / Zeiler M.D., Fergus R. // Computer Vision ECCV. – 2014. - vol. 8689. - pp. 818–833. DOI: [10.1007/978-3-319-10590-1_53](https://doi.org/10.1007/978-3-319-10590-1_53)

УДК 004.932.2

Дрозд В. П., аспирант, OrcID [0000-0002-8701-678X](https://orcid.org/0000-0002-8701-678X)

e-mail: drozdvdym@gmail.com

Национальный технический университет Украины
«Киевский политехнический институт имени Игоря Сикорского»,

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ИЗОБРАЖЕНИЙ С ПОМОЩЬЮ ГЛУБИННЫХ НЕЙРОННЫХ СЕТЕЙ

Построение стабильной системы для оценки коэффициента схожести изображений очень сложная задача. В этой работе рассмотрен подход к обучению системы сравнения изображений, использующий техники глубокого обучения, а именно объединение всех этапов сравнения в одну глубинную нейронную сеть. Данный подход позволяет построить систему, которая имеет гораздо большую учебную емкость по сравнению с другими. Экспериментальная проверка системы проведена на тестовых изображениях, представляющих собой реальные фото одежды. Результаты подтверждают, что подход объединения системы в одну глубинную нейронную сеть позволяет повысить качество по сравнению с другими подходами, а также позволяет уменьшить размер полученного вектора признаков.

Библ. 16, рис. 1, табл. 2.

Ключевые слова: цифровые изображения; глубокие нейронные сети; сравнение изображений; дескрипторы; признаки.

UDC 004.932.2

Droz V. P., postgraduate, OrcID [0000-0002-8701-678X](https://orcid.org/0000-0002-8701-678X)

e-mail: drozdvdym@gmail.com

National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute»

COMPARATIVE ANALYSIS OF IMAGES USING DEEP NEURAL NETWORKS

Building a stable system for assessing the similarity of the images is very difficult task. In this paper proposed approach to learning image comparison system that uses techniques of deep learning, namely the combining of all stages of the comparison into one deep neural network. This approach allows one to construct a system that has a much greater training capacity than others. Experimental verification of the system was done on digital images, which represent photos of clothes. Results confirm that the proposed system can improve quality of comparison, as well as reduce the size of the feature vector.

References 16, figures 1, tables 2.

Keywords: digital images; deep neural networks; image comparison; descriptors; features.

References:

- [1]. Structure from motion. [Online]. Available: https://en.wikipedia.org/wiki/Structure_from_motion.
- [2]. R. Hadsell, S. Chopra, and Y. LeCun. Dimensionality reduction by learning an invariant mapping. In CVPR, volume 2, pages 1735–1742. IEEE, 2006. DOI: [10.1109/CVPR.2006.100](https://doi.org/10.1109/CVPR.2006.100).
- [3]. G. W. Taylor, I. Spiro, C. Bregler, and R. Fergus. Learning invariance through imitation. In CVPR, pages 2729–2736. IEEE, 2011. DOI: [10.1109/CVPR.2011.5995538](https://doi.org/10.1109/CVPR.2011.5995538)
- [4]. D. G. Lowe. Object recognition from local scale-invariant features. In ICCV, volume 2, pages 1150–1157. IEEE, 1999. DOI: [10.1109/ICCV.1999.790410](https://doi.org/10.1109/ICCV.1999.790410)
- [5]. N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In CVPR, pages 886–893. IEEE, 2005. DOI: [10.1109/CVPR.2005.177](https://doi.org/10.1109/CVPR.2005.177).



- [6]. Y.-L. Boureau, F. Bach, Y. LeCun, and J. Ponce. Learning mid-level features for recognition. In CVPR, pages 2559–2566. IEEE, 2010. DOI: [10.1109/CVPR.2010.5539963](https://doi.org/10.1109/CVPR.2010.5539963).
- [7]. A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, pages 1106–1114, 2012.
- [8]. Tuytelaars, T and Mikolajczyk, K (2008) Local invariant feature detectors: A survey Foundations and Trends in Computer Graphics and Vision, 3 (3). pp. 177-280.
- [9]. E. Tola, V. Lepetit, and P. Fua. A Fast Local Descriptor for Dense Matching. In Proceedings of Computer Vision and Pattern Recognition, Alaska, USA, 2008. DOI: [10.1109/CVPR.2008.4587673](https://doi.org/10.1109/CVPR.2008.4587673).
- [10]. V. Athitsos, J. Alon, S. Sclaroff, and G. Kollios. Boostmap: A method for efficient approximate similarity rankings. In IEEE Conf. on Computer Vision and Pattern Recognition, Madison, WI, June 2004. DOI: [10.1109/CVPR.2004.1315173](https://doi.org/10.1109/CVPR.2004.1315173).
- [11]. A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. CNN features off-the-shelf: An astounding baseline for recognition. In IEEE Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2014, Columbus, OH, USA, June 23-28, 2014, pages 512–519, 2014. DOI: [10.1109/CVPRW.2014.131](https://doi.org/10.1109/CVPRW.2014.131).
- [12]. P. Fischer, A. Dosovitskiy, and T. Brox. Descriptor matching with convolutional neural networks: a comparison to SIFT, arXiv preprint arXiv:1405.5769, 2014. Available: <https://arxiv.org/abs/1405.5769>.
- [13]. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, Going deeper with convolutions, arXiv preprint arXiv:1409.4842, 2014. Available: <https://arxiv.org/abs/1409.4842>.
- [14]. Глибоке навчання [Deep Learning]. Available: https://uk.wikipedia.org/wiki/Глибоке_навчання.
- [15]. Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, Caffe: Convolutional architecture for fast feature embedding, arXiv preprint arXiv:1408.5093, 2014. Available: <https://arxiv.org/abs/1408.5093>.
- [16]. Zeiler M.D., Fergus R. (2014) Visualizing and Understanding Convolutional Networks. Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8689. Springer, Cham. DOI: [10.1007/978-3-319-10590-1_53](https://doi.org/10.1007/978-3-319-10590-1_53)