

Кравець А.М.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Сімоненко В.П.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

МЕТОД ЗНАХОДЖЕННЯ ІСТОРИЧНИХ ПАМ'ЯТОК НА ЗОБРАЖЕННІ ЗА ДОПОМОГОЮ ТЕХНОЛОГІЙ КОМП'ЮТЕРНОГО ЗОРУ

У статті описано метод знаходження визначних місць та історичних пам'яток на зображенні в системі для проведення екскурсій у віртуальній та доповненій реальності. Зважаючи на стрімкий розвиток та впровадження нейромереж і машинного навчання, людство отримало змогу залучити певні системи штучного інтелекту для виконання практичних завдань, проте далеко не всі завдання було охоплено. Система надає користувачеві вебресурс, який може бути використаний для проведення екскурсій у віртуальній та доповненій реальності за попередньо складеним сценарієм. Екскурсії поділені на точки, кожна з яких демонструє користувачу доповнену або віртуальну реальність. Задля покращення наявної системи було ухвалено рішення додати функцію розпізнавання визначних місць природи та архітектури для точки, де використовується доповнена реальність. Для вирішення цього завдання розглянуто наявні рішення у сфері технологій комп'ютерного зору. Використання технологій комп'ютерного зору для вирішення цього завдання є виправданим із точки зору надійності та швидкодії. Однак завдання розпізнавання пам'яток не є тривіальним та потребує більшої уваги, ніж просто розпізнавання об'єкта якогось класу (людина, предмет, природа тощо). Вона потребує не лише виокремлення об'єкта в деякий клас, а і визначення того, чи є цей об'єкт пам'яткою архітектури, чи це просто будівля, чи взагалі на зображенні наявна пам'ятка. Цей напрям у технології комп'ютерного зору є порівняно новим та, безперечно, цікавим і важливим для навчальних, наукових та розважальних цілей. Метою розроблення є покращення наявних рішень у форматі вебресурсу. Для досягнення кращих результатів було оцінено рішення та методи з точки зору можливості їх використання у вебдодатку. Виявлено недоліки та шляхи вдосконалення. Запропоновано підхід, який дає високу точність визначення пам'яток. Цей підхід може бути використаний у системі для проведення екскурсій у віртуальній та доповненій реальності. Актуальність проблеми підкріплюється необхідністю вдосконалення наявних програмних рішень та підходів до розпізнавання об'єктів за допомогою комп'ютерного зору.

Ключові слова: комп'ютерний зір, машинне навчання, згорткова нейронна мережа, доповнена реальність, пам'ятки архітектури.

Постановка проблеми. Розпізнавання визначних місць за допомогою комп'ютерного зору відрізняється від інших проблем розпізнавання зображень щонайменше за трьома характеристиками:

1) проблема розпізнавання на рівні екземпляру, а не категорії: замість визначення загальних об'єктів, таких як будівлі чи природа, метою є визначення конкретних історичних пам'яток або природних заповідників, таких як Біг-Бен чи Великий каньйон;

2) кількість різних сутностей набагато більше, ніж у класичних завданнях розпізнавання зображень [1]; деякі пам'ятки архітектури не мають достатньо фотографій;

3) історичні пам'ятки – це статичні об'єкти, які рідко змінюються. Варіації в їх зображеннях виникають унаслідок умов зйомки, таких як експозиція або оглядова площа. Це є відмінною рисою від загального розпізнавання зображень, коли предмети (люди, тварини, будівлі тощо) мають багато варіантів розташування та зовнішності.

Аналіз останніх досліджень і публікацій. Виявлення зображень із відомими спорудами та розпізнавання історичних надбань людства – основні теми досліджень, які проводились десятиліттями у сфері технологій комп'ютерного зору. Загальне завдання пошуку зображень полягає в ранжуванні зображень у деякий словник (так званий індекс) та встановлення відповідності щодо запиту.

Завдання розпізнавання визначних місць та споруд полягає у визначенні конкретного екземпляра класу об'єктів (наприклад екземпляр «Мона Ліза» з класу об'єкта «живопис») [2]. Оскільки методи для виконання обох завдань розвиваються і дотепер, у результаті підходи стали більш надійними та масштабованими, що дозволяє використовувати навіть маловідомі набори даних. Більше того, хоча масштаби класифікаційних наборів даних, таких як ImageNet, COCO і OpenImages (які вже зарекомендували себе як еталон), зростають із кожним днем, отримання достовірних результатів можливе і на малих датасетах. Наприклад, оригінальний Oxford5k та датасет Paris6k, випущені у 2007 р. та 2008 р. відповідно, мають лише 55 зображень у кожному класі, але сьогодні все ще широко використовуються. Це пов'язано з тим, що обидва набори даних містять лише зображення з одного міста, тому результати можна не узагальнювати для масштабованих налаштувань.

Google вже дозволяє розпізнавання пам'яток у мобільних телефонах за допомогою API Firebase ML-kit, який пропонує спеціальну функцію ідентифікації пам'ятки. Цей спосіб розпізнавання вимагає від користувача підключення до мережі інтернет та встановлення спеціального додатка, що є недоліком та обмеженням певних груп користувачів.

Інші компанії, такі як Blippar, також розробляють реальні програми розпізнавання об'єктів, які включають розпізнавання відомих об'єктів. Ці програми доступні для користувачів Android та iOS (рис. 1).

Перші спроби Google розпізнати відомі споруди датуються 2009 р., коли була розроблена система розпізнавання пам'яток, яка вважається

точною на 80%. Сьогодні системи здатні розпізнавати зображення з вищою точністю.

Постановка завдання. Метою роботи є отримання більш сучасного та покращеного методу розпізнавання об'єктів на зображенні та його впровадження. Для досягнення цієї мети було поставлено такі основні завдання: дослідження наявних алгоритмів та методів комп'ютерного зору і розпізнавання образів пам'яток, аналіз їх сильних та слабких сторін, розроблення способу інтеграції компонентів для розпізнавання образів, пошук та постачання контекстно-релевантної інформації, розроблення складного програмного продукту, який поєднує цей функціонал та розширює його.

Завдання цієї роботи є розроблення більш продуктивного та оптимізованого методу розпізнавання зображень, проектування та розроблення прототипу з використанням розробленого підходу з технологіями машинного навчання та віртуальної й доповненої реальності з подальшою інтеграцією цього функціоналу з навчальними цілями.

Предметом дослідження є покращення наявних характеристик нейронних мереж для роботи у сфері розпізнавання об'єктів, що спрощує сприйняття нової інформації під час екскурсії, аналогів якого в широкому доступі знайдено не було.

Виклад основного матеріалу дослідження. Для досягнення мети розроблення було спроектовано вдосконалений метод розпізнавання об'єктів на зображенні. Як і більшість методів пошуку зображень, цей метод заснований на глибокій згортковій нейронній мережі (далі – ЗНМ) [3]. Для отримання результату використовується прототипний підхід, де всі класи характеризуються відповідними центроїдами. Цей тип метричного навчання поєднує швидкість і точність для широ-



Рис. 1. Додаток Blippar для iOS

комасштабного завдання розпізнавання. Далі описується архітектура мережі, функція втрат, методологія навчання та аспекти ухвалення рішень.

Нещодавні успіхи в розвитку технології комп'ютерного зору та пошуку зображень тісно пов'язані зі згортковими нейронними мережами [4]. Еталонна ЗМН складається з трьох частин: згорткової мережі, агрегувального шару та повноз'язного шару (рис. 2).

Насамперед слід було обрати оптимальну архітектуру мережі. Для того, щоб отримати ЗМН, придатну для навчання моделі розпізнавання пам'яток, було використано тонке налаштування, як дуже ефективна техніка в різних сферах застосування комп'ютерного зору. Ця архітектура мережі продемонструвала компроміс між якістю та часом виведення [5].

По-друге, був доданий шар вбудовування в магістраль ЗМН. Потім були вирізані останні повноз'єднані шари після середнього агрегувального шару мережі, а замість них додано один повноз'єднаний шар та нормалізацію.

Розмір шару вбудовування є важливим параметром. Після серії експериментів із даними, 512 був обраний оптимальним. Зазвичай цей параметр визначається емпіричним шляхом [6]. Третім компонентом ЗМН є класифікаційний повноз'єднаний шар. Для перевірки вивченої ЗМН використано карти активації класів, отримані за найбільшим значенням активацій із вбудованого шару. карти активації класів є одним із найпоширеніший методів визначення класу образу на зображенні. Таким чином, ЗМН виділяє частини зображень із правильним зображенням.

Конвеєр розпізнавання образів під час логічної оцінки використовує центроїди, що описують

кожен клас образу. Щоб максимізувати ефективність цього підходу, довелося навчити мережу так, щоб члени кожного класу були якомога ближче до якогось елемента, який називається «центр» (або «центроїд») [7].

Були проведені експерименти з кількома придатними функціями втрат: контрастними втратами, втратами від Arcface та Center loss. Найкращі результати тестів були досягнуті емпіричним шляхом за допомогою Center loss, який одночасно навчає центр для вбудовувань кожного класу й оцінює відстань між вбудованими зображеннями та відповідними центрами класів (рис. 3). Важливою перевагою функції втрати Center loss є те, що це просте додавання до втрати Softmax, і, на відміну від контрастної втрати або втрати триплетом, немає необхідності відбирати негативні пари, що може бути досить складним завданням [8].

Важливою особливістю розпізнавання історичних пам'яток є те, що фотографії запитів можуть взагалі не містити їх, тому функція втрат повинна

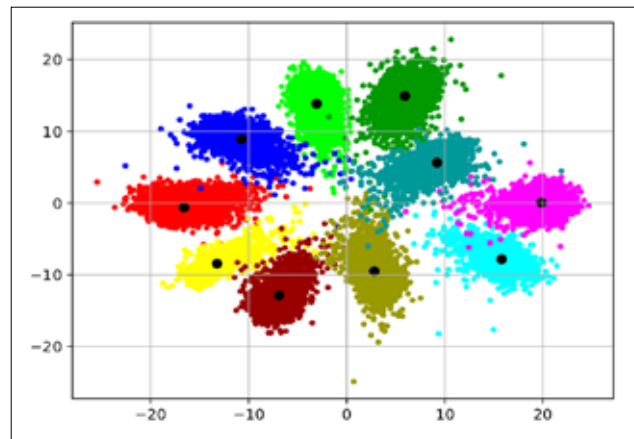


Рис. 3. Поведінка Center loss

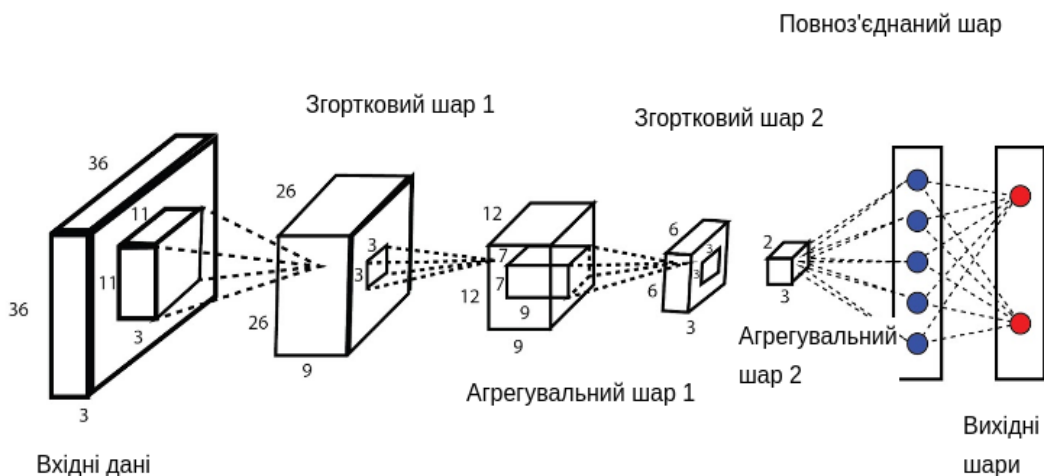


Рис. 2. Структура конволюційної нейронної мережі

це обов'язково враховувати. Таким чином, розглядається додатковий клас не-пам'ятки, який включає буквально все, що не є пам'яткою (люди, тварини, будівлі, природні об'єкти, машини тощо). Отже, кількість елементів не-пам'яткового класу набагато більше загальної кількості елементів у всіх класах пам'яток. Крім того, на відміну від пам'яток, не-пам'ятки класу не мають структури (оскільки вона включає неструктуровані об'єкти), тому немає сенсу застосовувати функцію втрати для цього класу. Щоб вирішити цю проблему, було модифіковано Center loss, щоб обчислити центри лише для класів пам'яток, але не для класу, що не є пам'яткою.

Для того, щоб вирішити, чи містить зображення запити пам'ятку, необхідно порівняти вектор вбудовування, отриманий із ЗНМ із деякими іншими векторами, що описують класи пам'яток [9]. Назвемо ці вектори центроїдами, тому що вони обчислюються шляхом усереднення вкладених векторів.

Основне питання використання центроїд полягає в даних, які використовувати для їх розрахунку. Спочатку було використано повні навчальні дані для кожної пам'ятки. Цей підхід дав задовільні результати, але цього було не зовсім достатньо для врахування всіх особливостей архітектури будівель. Незважаючи на очищення даних у датасеті, під час якого було видалено більшість невідповідних даних, деякі з них все-таки могли залишитись, що могло вплинути на центроїд. Наприклад, якщо пам'яткою архітектури, яка нас цікавить, є собор, який розташований на міській площі, то в ансамбль собору може бути помилково включена сусідня будівля.

По-друге, кожне визначне місце може мати кілька типових ракурсів зйомки. Змішування

їх в одному центроїді призводить до його недостатньої чутливості. Тому ефективніше розраховувати окремий центроїд для кожного типового ракурсу окремо. Для досягнення цієї цілі застосовано ієрархічний агломеративний алгоритм кластеризації та розділення даних тренування на кілька кластерів для кожного орієнтира. Кластером вважається множина, що містить більше 50 елементів. Якщо всі кластери для пам'ятки не є дійсними, то обчислюється лише один центроїд, використовуючи найбільший кластер.

Висновки. Був запропонований та обґрунтований спосіб покращення класичного використання алгоритму комп'ютерного зору, який дає можливість гарантувати підвищення ефективності роботи та швидкості навчання мережі.

Основою методу є використання вбудовувань у глибокій згортковій нейронній мережі. Ця нейронна мережа пройшла навчання з використанням функції втрат Center loss. Ця функція втрат забезпечує кращі результати швидкодії та точності (порівняно з відомим алгоритмом). Для визначення того, чи є пам'ятка на цьому зображенні, використовуються відстані між вектором вбудовування зображення та центроїдом класу. Для кожної пам'ятки розраховуються центроїди над кластерами, отриманими в результаті ієрархічної процедури кластеризації. На думку авторів, це є однією з характеристик, які покращують класичний метод розпізнавання образів на зображенні.

Практичне значення одержаних результатів роботи визначається тим, що розроблений підхід дає можливість ефективно використовувати ресурси системи та підвищити ефективність розпізнавання конкретних об'єктів на зображенні.

Список літератури:

1. Höppner, F., Klawonn, F., Kruse, R., & Runkler, T. Fuzzy cluster analysis: methods for classification, data analysis and image recognition. John Wiley & Sons. 1999.
2. Созыкин, А. В. Обзор методов обучения глубоких нейронных сетей. *Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика*, 2017, № 6.3. 32 с.
3. Бринк Х., Ричардс Д., Феверолф М. Машинное обучение. Издательский дом «Питер», 2017. 54 с.
4. David A., Ponce J. Computer vision: a modern approach. Prentice Hall Professional Technical Reference, 2002. p. 78.
5. Hoff, W. A., Nguyen, K., & Lyon, T. Computer-vision-based registration techniques for augmented reality. In *Intelligent Robots and Computer Vision XV: Algorithms, Techniques, Active Vision, and Materials Handling* (Vol. 2904). International Society for Optics and Photonics. 1996, pp. 538–548.
6. Wang, C., Komodakis, N., & Paragios, N. Markov random field modeling, inference & learning in computer vision & image understanding: A survey. *Computer Vision and Image Understanding*, 117(11), 2013, 1610–1627.
7. Liu, Y., Song, G., Shao, J., Jin, X., & Wang, X. Transductive centroid projection for semi-supervised large-scale recognition. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018. (pp. 70–86).
8. Kalchbrenner, N., Grefenstette, E., & Blunsom, P. A convolutional neural network for modelling sentences. 2014.
9. O'Shea, K., & Nash, R. An introduction to convolutional neural networks. 2015.

**Kravets A.M., Simonenko V.P. LANDMARKS RECOGNITION METHOD
USING COMPUTER VISION TECHNOLOGIES**

The article describes the method of detecting landmarks and historical monuments in the image in the system for excursions in virtual and augmented reality. Due to the rapid development and implementation of neural networks and machine learning, humanity has been able to attract certain artificial intelligence systems to solve practical problems, but not all tasks were covered. The system provides the user with a web resource that can be used to conduct excursions in virtual and augmented reality according to a pre-compiled script. Excursions are divided into points, each of which shows the user augmented reality or virtual. To improve the existing system, it was decided to add the function of recognizing the sightseeing places and architecture for the point where augmented reality is used. To solve this problem, the existing solutions in the field of computer vision technology were considered. The use of computer vision technology to solve this problem is justified in terms of reliability and speed. However, the task of recognizing monuments is not trivial and requires more attention than just recognizing an object of a class – person, object, nature, etc. It requires not only the extraction of an object into a class, but also the determination of whether the object is an architectural monument, whether it is just a building, or whether there is a landmark in the image at all. This direction in computer vision technology is relatively new and, of course, interesting and important for educational, scientific and entertainment purposes. The purpose of the development is to improve existing solutions in the format of a web resource. For best results, solutions and methods were evaluated in terms of their use in the web application. Shortcomings and ways to improve have been identified. An approach that provides high accuracy in the identification of monuments is proposed. This approach can be used in the system to conduct excursions in virtual and augmented reality. The urgency of the problem is supported by the need to improve existing software solutions and approaches to object recognition using computer vision.

Key words: computer vision, machine learning, convolutional neural network, augmented reality, landmarks.