

SECTION: TECHNICAL SCIENCES.

SEKCJA: NAUKI TECHNICZNE.

How to cite: Sahun, A., Khaidurov, V., & Ivanova, N. (2024). Pattern Identification System Based on a Neural Network with Deep Learning. *International Conference on Science, Innovations and Global Solutions*. (pp. 205-212). Futurity Research Publishing. <https://futuraity-publishing.com/international-conference-on-science-innovations-and-global-solutions-archive/>

Pattern Identification System Based on a Neural Network with Deep Learning

Сагун Андрій Вікторович¹, Хайдров Владислав Володимирович², Іванова Наталія Валеріївна³

¹*к.т.н., доцент, Національний університет біоресурсів і природокористування України, м. Київ, Україна, a.sagun@nubip.edu.ua*

²*к.т.н, старший дослідник, НТУУ "КПІ ім. Ігоря Сікорського", м. Київ, Україна, allif0111@gmail.com*

³*аспірант, НТУУ "КПІ ім. Ігоря Сікорського", м. Київ, Україна, nataustyanovska@ukr.net*

Accepted: July 3, 2024 | **Published:** July 17, 2024 | **Language:** Ukrainian

Abstract: The developed object identification system is based on neural networks (NN) with deep learning. This identification system provides accurate and fast recognition of known and close to known objects in the video data obtained from IP video surveillance cameras. Depending on the video shooting conditions and the viewing angle of the IP camera, the system's algorithm achieves a maximum recognition accuracy of 96.38%. This percentage of pattern recognition is consistent across 11 classes of identification objects. The obtained results are achieved by using the CamVid training database as a training sample of the neural network. This database consists of 421 training and 280 test video images.

The system model provides for the possibility of optimizing the parameters of the learning function, identifying and changing the method of measuring the distance between feature vectors (metrics).

Keywords: deep learning, classifier, machine vision, neural network, reference information vector.

Вступ

Системи розпізнавання та ідентифікації графічних об'єктів на основі НМ є найсучаснішим і ефективним напрямком в комп'ютерному зорі та штучного інтелекту. Здатність НМ навчатися на великих вибірках даних та їх адаптація до різноманітних умов забезпечує високу точність та надійність розпізнавання зображень.

Сучасні НМ здатні автоматично витягувати ознаки зображень і класифікувати їх на основі шаблонів, що є корисним для розпізнавання графічних об'єктів. Моделі на НМ типу VGG16 і VGG19, використовуються в задачах комп'ютерного зору, аналізу відеоконтенту. Для таких застосувань розвиток НМ розширяє можливості систем розпізнавання збільшенням точності, швидкості та адаптації.

Огляд літератури

На основі аналізу наукових досліджень, присвячених алгоритмам і методам розпізнавання та ідентифікації об'єктів, представлених в (Moskalenko, 2019; Lakhno, 2021; Herasina, 2010; Lakhno, 2020; Schuster, 1992; Ljung, 2020; Nelles, 2001; Goodfellow, 2016; Lee, 2019), видно, що для подібних задач переважно використовуються нейронні мережі (НМ). Наприклад, в (Ioannis, 2022) показано, що моделі на основі НМ у ряді специфічних застосувань виявляються найефективнішими для розробки систем розпізнавання образів та ідентифікації графічних об'єктів. Для створення системи розпізнавання образів важливо сформулювати математичну модель, яка дозволить реалізувати підсистему ідентифікації графічних об'єктів. Доцільно обрати модель на базі НМ, що дозволяє актуалізувати та покращувати якість ідентифікації та подальшої класифікації об'єктів (Ljung, 2020; Nelles, 2001). У комп'ютерних інтелектуальних системах для розпізнавання об'єктів у відеоконтенті НМ є ключовим фактором підвищення точності та адекватності алгоритмів ідентифікації та розпізнавання (Ljung, 2020; Nelles, 2001; Goodfellow, 2016).

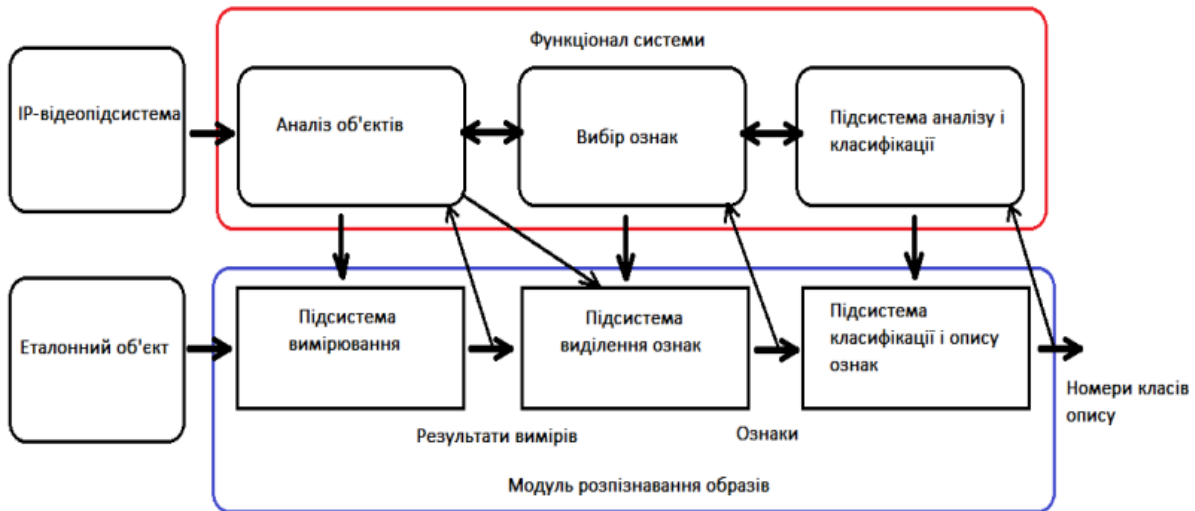
Мета дослідження полягає в розробці ефективного алгоритму розпізнавання графічних об'єктів з відеоконтенту, отриманого з IP-камер системи відеонагляду. Точність розпізнавання образів при практичній реалізації алгоритму повинна становити не менше 95%, що є показником для подібних прикладних реальних систем.

Результати дослідження

Системи ідентифікації/розпізнавання образів складаються з підсистем: технічна частина (підсистема відеонагляду); підсистема аналізу та класифікації елементів зображення; підсистема реєстрації та узагальнення даних розпізнавання. Структурна схема системи розпізнавання образів в загальному вигляді показана на рис. 1.

Рисунок 1

Структурна схема системи розпізнавання образів



Джерело: власна розробка авторів

Важливим етапом у формуванні алгоритму є вибір моделі і типу навчання НМ. На сьогодні існують такі НМ за видами навчання (Lee, 2019): наглядне навчання (Supervised Learning); ненаглядне навчання (Unsupervised Learning); підглядне навчання (Semi-Supervised Learning); підкріплене навчання (Reinforcement Learning).

У побудованій НМ передбачено кілька шарів нейронів. Вхідний шар отримує початкові дані, включаючи інтенсивність кожного пікселя та ознаки Хаара для об'єктів (дерева, автомобілі, дороги, елементи тротуару, пішоходи тощо). Ці дані передаються до першого прихованого шару НМ, де відбуваються математичні обчислення з використанням вхідних даних, а вихідний шар формує фінальний результат - ідентифікацію типу об'єкта, присутнього на зображенні. Отже, ефективною для випадку розроблюваної системи є НМ з глибоким навчанням (Lee, 2019), а практична реалізація розробленої моделі системи розпізнавання образів проведена в середовищі розробки MATLAB. В ньому доступна функція `vgg16()`, яка реалізує архітектуру глибокої НМ з 16 згорткових та повністю зв'язаних шарів нейронів, що включає 13 згорткових шарів і 3 повністю зв'язані шари. Дана функція повертає об'єкт НМ, але не містить кріткого методу обчислення відстані між векторами ознак для оброблюваних зображень. Тому, для розв'язку таких задач використовуються такі методи обчислень (Sahun, Khaidurov & Bobkov, 2024): евклідова відстань; квадрат евклідової відстані; відстань Чебишева; хемінгова відстань; мангеттенська відстань. Для моделі системи ідентифікації є сенс використовувати відстань Евкліда для визначення відстані між точками тестових та еталонних зображень. Для розрахунку відстані (метрики) скористаємося наступним виразом:

$$\rho(x, x') = \sqrt{\sum_{i=1}^n (x_i - x'_i)^2}.$$

У наведеному вище виразі: x_i - перший n-вимірний вектор точок зображення; x'_i -другий n-вимірний вектор точок зображення. Інформаційні n-вимірні вектори в даній моделі формуються за принципом таких виразів: $x = (x_1; x_2; \dots x_n)$, $x' = (x'_1; x'_2; \dots x'_n)$.

Тренування та тестування системи образів

Ефективне розпізнавання образів можливе завдяки застосуванню алгоритмів моделей, розроблених на основі глибокого навчання. Оптимальні результати досягаються після тренування моделі на відеовибірці високої якості. Для навчання використовувались еталонні дані відеобазы CamVid, що містять спеціальні мітки. Ця база даних для тренування складається з колекції відеофрагментів, які мають семантичні мітки для відповідних класів об'єктів та доповнені метаданими. Еталонна база CamVid асоціює кожен піксель зображення з одним із 32 семантичних класів об'єктів, які ідентифікує система. Серед класів об'єктів, які модель ідентифікує, можна виокремити: «building» (будівля), «pole» (стовп), «sky» (небо), «tree» (дерево), «signSymbol» (знак), «road» (дорога), «pavement» (тротуар), «fence» (паркан), «car» (автомобіль), «pedestrian» (пішохід), «bicyclist» (велосипедист). Роздільна здатність графічних кадрів під час тренування обмежена параметрами 360x480 пікселів.

Для того, щоб отримати середні значення для впорядкованого набору тестових даних можна використати функції `median()`, яка присутня в середовищі моделювання та тестування створеного алгоритму НМ Matlab. Для функції `trainingOptions()` в Matlab змінимо значення трьох з дев'яти аргументів, які впливають на якість та можливості навчання обраної НМ, а саме: `momentum` (значення імпульсу для прискорення сходження градієнтного спуску в навчанні НМ); `initialLearnRate` (ініціалізаційне значення для навчання нейронної мережі); `L2Regularization` (визначає силу регуляризації L2, що застосовується до ваг мережі під час навчання), на наступні значення параметрів ініціалізації та функції втрат для тренування НМ (таблиця 1).

Таблиця 1

Параметри ініціалізації функції тренування нейронної мережі

Назва аргументу функції <code>trainingOptions()</code>	Значення параметра
<code>Momentum</code>	0.9
<code>InitialLearnRate</code>	$1 \cdot 10^{-3}$
<code>L2Regularisation</code>	0,0005

Джерело: власна розробка авторів

З метою виконання кластерного аналізу та наступної фінальної класифікації розпізнаних та ідентифікованих графічних об'єктів в підсистемі ідентифікації слід визначити шкалу оцінок для класифікатора отриманих з відеоконтенту IP-камер графічних зображень. Для основи формування шкали оцінок в роботі використано інформаційну складову існуючих в комп'ютерній моделі RGB

каналів, що визначають графічний фрейм у відповідному графічному форматі: $\{R \in (0; 255), G \in (0; 255), B \in (0; 255)\}$.

Еталонні вектори, з якими порівнюються отримані графічні кадри з IP-камер, сформовані відповідно до RGB-моделі кольору. Еталонні вектори в спеціалізованих джерелах часто згадаються, як опорні (назва-аналог). Крім того, такі вектори для створюваної системи розділені на 11 класів розпізнавання: 1) будівлі, інфраструктурні об'єкти; 2) небо, горизонт; 3) дорожня інфраструктури та її елементи; 4) дорожнє покриття; 5) пішохідні доріжки, пандуси, пішоходи тощо; 6) дерева, кущі, рослинність; 7) дорожні, інформаційні знаки, світлофори; 8) паркани та огорожі; 9) автівки, великогабаритні засоби пересування та потяги; 10) дорослі пішоходи та діти, тварини, візки; 11) мотоцикли та мопеди, скутери.

Графічна карта RGB-каналів кольорів для відокремлення класів ідентифікованих графічних об'єктів містить певний набір значень каналу кольору для всіх існуючих у системі розпізнавання 11 основних класів об'єктів. У таблиці 2 наведено значення параметрів для розпізнавання базових об'єктів у основних класах ідентифікації на базі RGB-моделі. Математичну модель навчають на вибірці, що складається з 421 еталонних зображень, з яких 280 використовуються для подальшого тестування. Наприклад, вага об'єктів класу «Sky» у графічних фреймах відеопотоку з IP-камери становить 0.318184709354742 при частоті 0,9266.

Таблиця 2

Базові об'єкти основних класів ідентифікації та їх характеристики

Назва класу розпізнавання	Характеристики класів розпізнавання (RGB-ознаки)
Sky	128 128 128
Building	128 0 0
Pole	192 192 192
Road	128 64 128
Pavement	60 40 222
Tree	128 128 0
SignSymbol	192 128 128
Fence	64 64 128
Car	64 0 128
Pedestrian	64 64 0
Bicyclist	0 128 192

Джерело: власна розробка авторів

Тренування одержаного алгоритму на навчальному наборі відеоданих базується на глибокому навчанні нейронної мережі. Такий підхід дає змогу ефективно відрізнити фон зображення від інформаційного вмісту об'єктів ідентифікації та розпізнавання. Практичне застосування розробленої системи ідентифікації передбачає, що результати ідентифікації не в візуальному відображенні, а у вигляді числового масиву даних для того, щоб надалі їх можна було б використати в більш складних системах. Гістограма частот зустрічі ідентифікованих класів і підкласів об'єктів в зоні ідентифікації камер відеонагляду показано на рис.2.

Рисунок 1

Гістограма частоти зустрічей ідентифікованих класів і підкласів графічних об'єктів в зоні ідентифікації

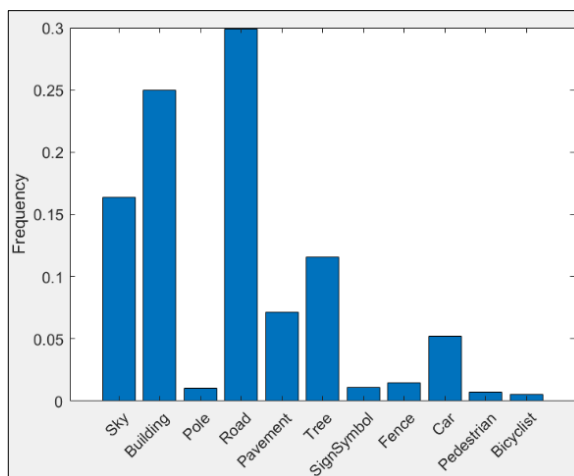


Рис. 2.

Джерело: власна розробка авторів

Порівнюючи відхилення в на отриманих числових масивах порівняно з референсною виборкою, можна побачити що розроблена система ідентифікації графічних об'єктів може приймати рішення щодо виявлення специфічних станів на графічних об'єктах або визначати деякі специфічні реакції всієї системи.

Після серій експериментів з ідентифікації, проведених в Matlab, алгоритм продемонстрував точність ідентифікації з максимальним показником 96,38%, а середній показник точності розпізнавання, отримани в 5 експериментах складає 75,42% для відеопотоку IP-камери. Для моделі НМ були взяті тестові дані з 12 фрагментів, кожен з яких тривалістю 68,35 секунди. Порівняння результатів точності розпізнавання об'єктів з наведеними в (Lee, 2019) даними та близькою за використанням доменною зоною показано в таблиці 3.

Таблиця 3

Порівнювана точність розпізнавання графічних образів

Тип наборів тестових даних	Точність, %
Casting*	77,62
Defect*	97,88
Magnetic*	92,67
Tech*	94,23
Bridge*	99,02
Solar*	76,78
Базові об'єкти основних класів ідентифікації створеного алгоритму (таблиця 2)	96,38

Джерело: власна розробка авторів

Для розпізнавання графічних вибірок на наборах тестових даних, позначених знаком «*» в таблиці 3 при навчанні використовувалися спеціалізовані навчальні вибірки для кожного набору (Sturgess, 2016). Як математична основа алгоритмів розпізнавання наборів тестових даних (таблиця 3), використовувалася НМ з глибоким навчанням - VGG19. Ця модель має на три шари нейронів більше, ніж та, що використана в даній роботі. Але, зі збільшенням кількості шарів і нейронів у НМ зростають обчислювальні витрати на розпізнавання образів. Відповідно, зростають вимоги до апаратного забезпечення платформи ідентифікації та розпізнавання графічних об'єктів.

Висновки

Отже, можна стверджувати, що комбінація НМ типу VGG16 з методом обчислення відстані між векторами ознак зображень на основі евклідової відстані забезпечує прийнятну точність розпізнавання, достатню для використання в системах машинного зору. Отримані показники є співставними продуктивнішими, але складнішими алгоритмами на базі VGG19, які мають вищу обчислювальну складність.

Література

Apostolopoulos, I. D., & Tzani, M. (2022). Industrial object, machine part and defect recognition towards fully automated industrial monitoring employing deep learning. The case of multilevel VGG19. *Journal of Ambient Intelligence and Humanized Computing*. <https://doi.org/10.1007/s12652-021-03688-7>

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. The MIT Press, London, England.

Herasina, O. V., & Korniienko, V. I. (2010). Alhorytmy hlobalnoi i lokalnoi optymizatsii v zadachi identyfikatsii skladnykh dynamichnykh system. *Systemy obrobky informatsii*, (6), 73–77.

Lakhno, V., Akhmetov, B., Chubaievskiy, V., Desiatko, A., Palaguta, K., Blozva, A., & Chasnovskiy, Y. (2021). Information Security Audit Method Based on the Use of a Neuro-Fuzzy System. In *Lecture Notes in Networks and Systems* (Vol. 232, pp. 171–184).

Lakhno, V., Kryvoruchko, O., Desiatko, A., Blozva, A., & Semidotska, V. (2020). Development strategy model of the informational management logistic system of a commercial enterprise by neural network apparatus. *CEUR Workshop Proceedings*, 2746, 87–98.

Lee, J.-H. (2019). Minimum Euclidean distance evaluation using deep neural networks. *AEU - International Journal of Electronics and Communications*, 112, 152964. <https://doi.org/10.1016/j.aeue.2019.152964>

Ljung, L., Andersson, C., Tiels, K., & Schön, T. B. (2020). Deep Learning and System Identification. *IFAC-PapersOnLine*, 53(2), 1175–1181. <https://doi.org/10.1016/j.ifacol.2020.12.1329>

Nelles, O. (2001). *Nonlinear System Identification: From Classical Approaches to Neural and Fuzzy Models*. Springer.

Sahun, A., Khaidurov, V., & Bobkov, V. (2024). Model of Graphic Object Identification in a Video Surveillance System based on a Neural Network. In *CPITS-2024: Cybersecurity Providing in Information and Telecommunication Systems*. <https://ceur-ws.org/Vol-3654/short2.pdf>

Schuster, H. G. (1992). Deterministic Chaos: Introduction and Recent Results. In H. Thomas (Ed.), *Nonlinear Dynamics in Solids* (pp. 1–25). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-95650-8_2

Sturgess, P. (2016). Oxford Brookes University. CamVid Database.

Moskalenko, V. V., & Korobov, A. H. (2019). Optimization parameters of intellectual identification system of objects on the terrain. *Radioelectronic and Computer Systems*, (2), 32–39. <https://doi.org/10.32620/reks.2016.2.05>