УДК 004.896.8:004.451.426

СИСТЕМА РОЗПІЗНАВАННЯ ОБРАЗІВ НА БАЗІ НЕЙРОННОЇ МЕРЕЖІ З ГЛИБОКИМ НАВЧАННЯМ

Сагун Андрій Вікторович
к.т.н., доцент, Національний університет біоресурсів і природокористування України
ORCID https://orcid.org/0000-0002-5151-9203
E-mail: a.sagun@nubip.edu.ua
Панаско Олена Миколаївна
к.т.н., доцент, Черкаський державний технологічний університет
ORCID https://orcid.org/0000-0002-0510-7742
E-mail: lena.pa@ukr.net

Анотація. Розроблена система розпізнавання об'єктів, яка ґрунтується на нейронних мережах з глибоким навчанням. Така система розпізнавання образів здатна забезпечити точне та швидке розпізнавання відомих їй та близьких до таких зображень у відеоконетні, отриманому з IP-камер відеонагляду. Залежно від умов зйомки відео та кута огляду IP-камери, алгоритм роботи даної системи досягає точності розпізнавання, що складає 96,38%. Такий відсоток розпізнавання образів є практично сталим відносно 11 класів об’єктів ідентифікації та розпізнавання. Такі високі результати досягаються завдяки використанню навчальної бази відеоданих CamVid в якості навчальної вибірки нейронної мережі. Така база сформована на основі 421 навчального та 280 тестових відеозображень. В моделі системи розпізнавання передбачені можливості оптимізації параметрів функції навчання та ідентифікації, а також зміни методу вимірювання відстані між векторами ознак (метрики вимірювання відстані точок).

Ключові слова: нейронна мережа, глибоке навчання мережі, класифікатор ознак, опорний інформаційний вектор, графічний фрейм, машинний зір.

Мета дослідження полягає в розробці ефективного алгоритму розпізнавання графічних об’єктів з відеоконтенту, отриманого з камер IP-камер системи відеонагляду. Точність розпізнавання образів в разі практичної реалізації алгоритма має складати не менше 95% для відеоданих, максимально наближених до реальних умов.

Аналіз матеріалів досліджень за напрямком роботи. На підставі проведеного аналізу наукових досліджень, присвячених алгоритмам та методам розпізнавання об'єктів типу «графічний образ», які наведені в численних джерелах, зокрема в [1-9], можна зробити певні висновки стосовно моделей та математичного апарату прикладних систем розпізнавання. В такій ролі найбільш доцільно застосовувати нейронні мережі (НМ). Моделі на базі НМ є найбільш ефективними для розробки систем розпізнавання образів, які оброблюють графічний або відеоконент. Для створення системи розпізнавання образів важливо сформулювати адекватну математичну модель, на основі якої можливо буде реалізувати підсистему ідентифікації та класифікації графічних об'єктів. Цілком логічно використати модель НМ, що ґрунтується на алгоритмах нейронних мереж з навчанням або без такого. Перевагою подібних моделей є можливість навчання. Навчання моделі НМ дає змогу актуалізувати і значно підвищувати якість процесу розпізнавання і наступної класифікації, що вказано в джерелах [6,7]. В комп'ютерних інтелектуальних системах для розпізнавання об’єктів у відеоконтенті машинне навчання часто є фактором, який суттєво підвищує адекватність та точність алгоритму визнання та ідентифікації, як зазначено у джерелах [6-8]. Тому вибір моделі системи на базі НМ є цілком виправданий.

Основна частина. Система розпізнавання образів з відеотрафіку IP-камер є актуальною. Подібні системи умовно складаються з декількох підсистем: 1) технічна частина (підсистема відеонагляду); 2) підсистема аналізу та класифікації елементів зображення; 3) пісистема реєстрації та узагальнення даних розпізнавання. Структурна схема системи розпізнавання образів в загальному вигляді показана на рис.1.

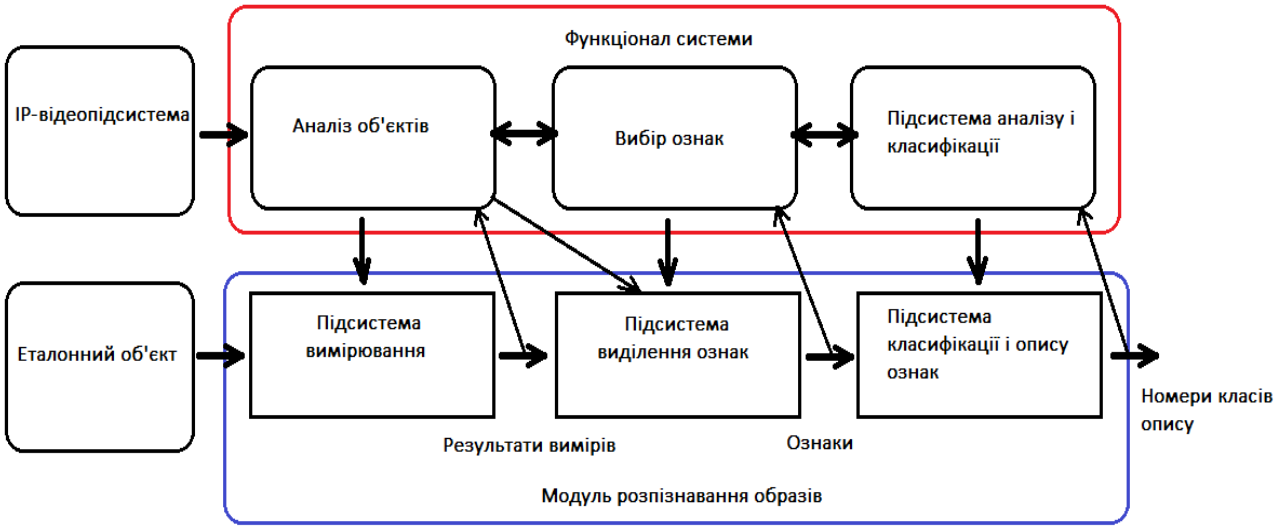

Рисунок - 1 Структурна схема системи розпізнавання образів

Важливим етапом у формуванні алгоритму є вибір моделі і типу навчання НМ. На сьогодні існують декілька типів моделей систем розпізнавання та ідентифікації за видами навчання [10]. З огляду на використання в алгоритмі математичного апарату НМ можна обрати наглядне навчання (Supervised Learning). Цей тип навчання використовується для навчання моделі на основі пар вхідних даних та відповідних міток (позначок), що узгоджується з використанням існучих у вільному доступі тренувальних виборок. Модель НМ з глибоким навчанням вчиться передбачати відповідні мітки для нових вхідних даних на основі попередньо навчених прикладів.

У побудованій НМ, яка обрана в якості основи для підсистеми ідентифікації графічних об'єктів, передбачено кілька шарів нейронів (рис.2).

З огляду на наведену на рис.2 схему нейронів НМ математичної моделі системи розпізнавання образів з ознакою Хаара, найбільш ефективної для випадку розроблюваної системи буде НМ з глибоким навчанням. Обрана модель НМ дозволяє проводити навчання НМ для передбачення результатів на основі вхідних даних.

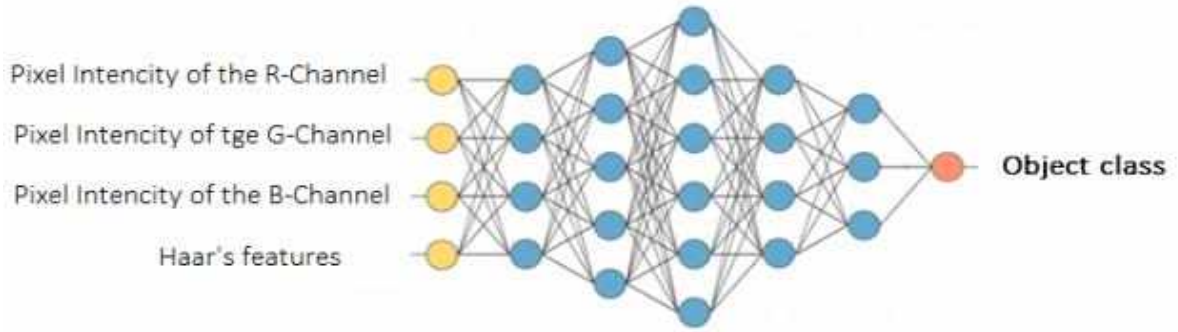

Рисунок – 2 Схема нейронних шарів прикладу нейромережі системи розпізнавання з ознакою Хаара

Вхідний шар НМ отримує початкові дані, включаючи інтенсивність кожного пікселя та ознаки Хаара для різних об'єктів (дерева, автомобілі, дороги, небо, елементи тротуару або

огорожі, пішоходи тощо). Даний метод базується на використанні фільтрів Хаара для виявлення різниць у значеннях пікселів у певних областях зображення. В якості основної ідеї використання ознак Хаара для ідентифікації графічних об’єктів є те, що на зображеннях шукають певні характеристики об'єктів, такі як границі, лінії або текстури, які можуть бути присутні у зображенні. Наприклад, для задачі розпізнавання автомобілів можуть використовуватися ознаки Хаара для виявлення коліс, фар, або інших характерних деталей автомобіля. Використовуючи такі ознаки алгоритм на основі глибокої нейронної мережі відрізняємо на графічному фоні інформаційний вміст об’єкту ідентифікації (на рис.3 цей об'єкт - автомобіль) [10].

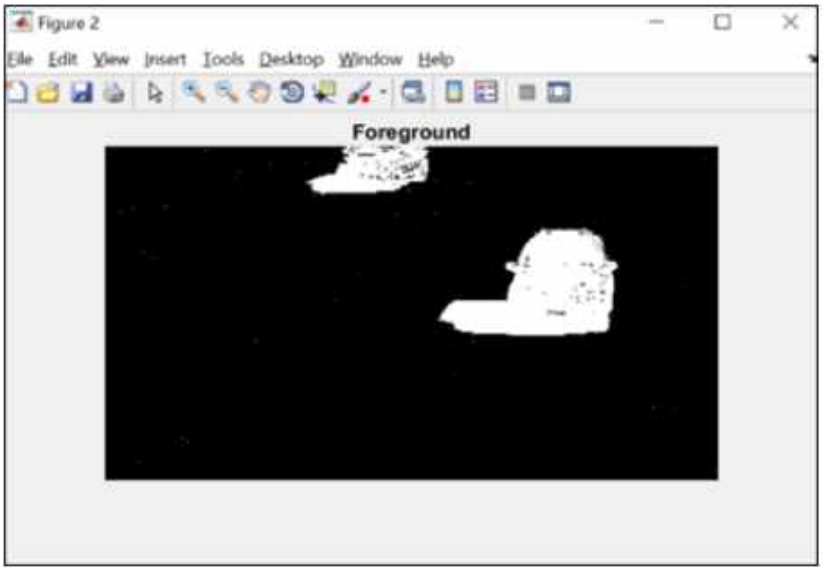

Рисунок - 3 Приклад виокремлення на графічному фоні інформаційний вміст об’єкту ідентифікації розробленим алгоритмом

Далі, ці дані передаються до першого прихованого шару НМ, де відбуваються математичні обчислення з використанням вхідних даних. Нарешті, вихідний шар формує фінальний результат - ідентифікацію типу об'єкта, присутнього на конкретному зображенні.

Для практичної реалізації розробленої математичної моделі системи розпізнавання обазів було вибрано середовище розробки MATLAB. У цьому середовищі доступна вбудована функція vgg16(), що реалізує в собі архітектуру глибокої нейронної мережі з 16 згорткових і повністю зв'язаних шарів нейронів, включаючи 13 згорткових шарів та 3 повністю зв'язаних шарів. Дану функцію можна використати для класифікації зображень в процесі розпізнавання. Згадана функція vgg16, що присутня MATLAB, повертає об'єкт нейронної мережі, але вона не містить конкретного методу для обчислення відстані (метрики) між векторами ознак оброблюваних зображень. Через це для розв’язку подібних задач використовуються спеціальні методики обчислень [9]:

- евклідова відстань;
- квадрат евклідової відстані;
- відстань Чебишева;
- хемінгова відстань.
- мангеттенська відстань;

Для моделі системи ідентифікації ми будемо використовувати класичну відстань Евкліда в якості методу визначення відстані між точками тестових та еталонних зображень. Обчислення відстані за допомогою метрики Евкліда є найпростішим та ефективним способом отримання значення прямої відстані між двома точками у просторі. При ідентифікації графічних зображень відстань Евкліда використовується, як міра порівняння виводу моделі з

очікуваними значеннями, які були отримані під час навчання нейронної мережі, а також для подальшої класифікації зображень на основі їхніх властивостей.

Для розрахунку метрик відстаней точок застосовуємо вираз для визначення Евклідової відстані, наведений нижче:

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

Відповідно, у наведеному вище виразі: x_i - перший n-вимірний вектор точок зображення; x_i' -другий n-вимірний вектор точок зображення. Дані інформаційні n-вимірні вектори в створеній моделі формуються за принципом, які можуть бути описані наступними виразами: $x = (x_1; x_2; \dots x_n)$. $x' = (x_1'; x_2'; \dots x_n')$.

Тренування та тестування системи образів.

Високі показники системи розпізнавання образів можливо отримати лише тоді, коли її алгоритм моделі, побудований на базі глибокого навчання, пройде тренування на високоякісній відеовибірці. Для тренування нейронної мережі використовувалися еталонні дані відеобази CamVid, що включає спеціальні мітки. Дана тренувальна база даних складається з колекції відеофрагментів з семантичними мітками відповідних класів об'єктів, доповнених метаданими. Основний набір міток еталонної бази CamVid асоціює кожен піксель з одним з 32 семантичних класів об'єктів, що ідентифікує система.

Однозначно визначені класи об'єктів ідентифікації моделі системи розпізнавання образів за типом:

- "Небо" ("Sky");
- "Будівля" ("Building");
- "Стовп" ("Pole");
- "Дорога" ("Road");
- "Тротуар" ("Pavement");
- "Дерево" ("Tree");
- "ЗнакСимволу" ("SignSymbol");
- "Паркан" ("Fence");
- "Автомобіль" ("Car");
- "Пішохід" ("Pedestrian");
- "Велосипедист" ("Bicyclist").

Для режиму тренування роздільна здатність графічних фремів була обмежена параметром 360x480 точок у відеовиборці.

Для того, щоб отримати середні значення для впорядкованого набору тестових даних можна використати функції median(), яка присутня в середовищі моделювання та тестування створеного алгоритсу НМ Matlab. Значення для параметрів ініціалізації, функції втрат та тренування нейронної мережі наведені в таблі 1.

Таблиця 1 - Параметри ініціалізації функції тренування нейронної мережі

Назва аргумента функції trainingOptions()	Значення параметра
Momentum	0.9
InitialLearnRate	$1*10^{-3}$
L2Regularisation	0,0005

З метою виконання кластерного аналізу та наступної фінальної класифікації розпізнаних та ідентифікованих графічних об'єктів в підсистемі ідентифікації слід визначати 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 ∈ (0; 255), G ∈ (0; 255), B ∈ (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:

- будівлі, інфраструктурні об’єкти;
- небо, горизонт;
- дорожня інфраструктури та її елементи;
- дорожнє покриття;
- пішоходні доріжки, пандуси, пішоходи тощо;
- дерева, кущі, рослинність;
- дорожні, інформаційні знаки, світлофори;
- паркани та огорожі;
- автівки, великогабаритні засоби пересування та потяги;
- дорослі пішоходи та діти, тварини, візки;
- мотоцикли та мопеди, скутери.

Графічна карта RGB-каналів кольорів для відокремлення класів ідентифікованих графічним об'єктів містить певний набір значень каналу кольору для всіх існуючих а системі розпізнавання 11 основних класів об’єктів. В таблиці 2 наведені значення параметрів для розпізнавання базових об’єктів у основних класах ідентифікації на базі RGB-моделі.

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

Тренування одержаного алгоритму на навчальному наборі відеоданих базується на глибокому навчанні нейронної мережі. Такий підхід дає змогу ефективно відрізняти фон зораюення від інформаційного вмісту об'єктів ідентифікації та розпізнавання. Математичну модель навчають на вибірці, що складається з 421 еталонних зображень, з яких 280 використовуються для подальшого тестування. Після тренування визначаються ваги об’єктів для кожного з відомих класів (таблиця 2). Наприклад, вага об’єктів класу «Sky» на графічних фреймах відеопотоку з IP-камери становить 0.318184709354742 при частоті 0,9266.

Висновки. Під час серії експериментів з розпізнаванням графічних образів, проведених в середовищі Matlab, алгоритм системи продемонстрував високу точність у роботі з графічною інформацією. В якості відеоданних для отримання результатів тестування був використаний відеоряд, що містив базові об’єкти основних класів ідентифікації, перераховані в таблиці 2**.** Точність розпізнавання ідентифікованих об’єктів склала 96,38% для відеопотоку з IP-камери. Для моделі НМ використовувалися тестові дані, що включали 12 фрагментів, кожен з яких мав тривалість 68,35 секунди. Порівнюючи результати точності розпізнавання графічних об’єктів

з максимально наближеними за математичним апаратом та доменною зоною використання можна зробити певні висновки за результатати, показаними на табл.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

Для розпізнавання графічних виборок на наборах тестових даних, позначених знаком «*» при навчанні використовувалися спеціалізовані для кожного набору навчальні вибірки [12]. В якості мамематичної основи для алгоритмів розпізнавання наборів тестових даних таблиці 3 використаного нейронну мережу з глибоким навчанням VGG19. Така модель має на 3 шари нейронів більше, порівняно з тією НМ, яка використана в даній роботі. Відомим фактом є те, що зі збільшенням кількості шарів та нейронів у нейронній мережі, як правило, збільшуються обчислювальні витрати на розпізнавання образів. Отже, збішаться і вимоги до апаратного забезпечення цільової платформи ідентифікації і розпізнавання графічних об’єктів.

Можна стверджувати, що комбінація НМ типу VGG16 з методом для обчислення відстані (метрики) між векторами ознак оброблюваних зображень на базі евклідової відстані дають прийнятну точність розпізнювання, достатню для застосування в ряді систем машинного зору. Отримані показники співставні з більш продуктивними, але складними алгоритмами на базі НМ VGG-16 при більшій обчислювальній складності останніх.

Перспективи вдосконалення. Для подальшого вдосконалення точності розпізнавання та ідентифікації графічних об’єктів можна використати інший метод обчислення метрики між векторами ознак зображень, наприклад, на основі чотирьох альтернатив: квадрату евклідової відстані; мангеттенська відстань; відстань Чебишева; хемінгова відстань. Хоча ці методи є обчислювально більш складними для потенційної апаратної бази застосування, але їх використання може підняти точність розпізнавання графічних образів в прикладному застосуванні.

References

1. 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.
2. Lakhno V., Akhmetov B., Chubaievskyi V., Desiatko A., Palaguta K., Blozva A., Chasnovskyi Y. (2021). Information Security Audit Method Based on the Use of a Neuro-Fuzzy System. Lecture Notes in Networks and Systems, 232 LNNS, pp. 171–184.
3. Herasina, O.V., & Korniienko, V.I. (2010). Alhorytmy hlobalnoi i lokalnoi optymizatsii v zadachi identyfikatsii skladnykh dynamichnykh system. Systemy obrobky informatsii, (6), 73–77.
4. Lakhno V., Kryvoruchko O., Desiatko A., Blozva A., Semidotska V. Development strategy model of the informational management logistic system of a commercial enterprise by neural network apparatus (2020). CEUR Workshop Proceedings, 2746, pp. 87–98.

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

6. 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

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

8. Goodfellow I., Bengio Y., & Courville A. (2016). Deep Learning. The MIT Press, London, England, 777 pages.

9. 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.

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

11. Sturgess, Paul. (2016). Oxford Brookes University. CamVid Database.

12. Ioannis D. Apostolopoulos, Mpesiana Tzani (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