

О. І. Лактіонов, Н. М. Педченко, А. С. Янко

Національний університет «Полтавська політехніка імені Юрія Кондратюка», Полтава, Україна

ПРАКТИЧНІ КЕЙСИ СТВОРЕННЯ ЗГОРТКОВИХ МОДЕЛЕЙ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ЗАДАЧ РОЗПІЗНАВАННЯ ОБРАЗІВ

Анотація. Мета статті підвищення ефективності розпізнавання образів шляхом створення бінарного класифікатора об'єктів цивільної інфраструктури засобами глибокого навчання. **Результати дослідження.** Створено згорткові моделі штучного інтелекту для розпізнавання об'єктів цивільної інфраструктури. Побудовано згорткову нейронну мережу Sequential, яка нараховує три згорткових шари, шари пулінгу, шар перетворення, пов'язаний шар та вихідний шар. Оптиміальне значення помилки при навчанні/тестуванні становить 0,0650/0,4424, де точність 0,98/0,92 відповідно, у той час як результати третьої епохи демонструють помилку 0,2442/0,2595 відповідно, а точність – 0,93. Використано попередньо навчену модель мережі VGG16, котру перенавчено на досліджуваному датасеті. Вказана модель демонструє мінімальне значення помилки для навчання/тестування 0,0278/0,1538 при точності 0,99/0,96 відповідно. **Наукова новизна.** Набуло подальшого розвитку використання згорткових нейронних мереж для завдань розпізнавання рівня руйнувань цивільної інфраструктури. **Практичне значення.** Побудовано дві моделі згорткової нейронної мережі, зокрема модель Sequential й VGG16 для вирішення завдань розпізнавання зруйнованих і цілих будинків. Для виявлення вказаних об'єктів необхідною передумовою є використання камери й відповідного апаратуного забезпечення, на кшталт raspberry пі або персонального комп'ютера чи ноутбука.

Ключові слова: комп'ютерний зір, завдання класифікації, глибоке навчання, нейронні мережі.

Постановка проблеми

Використання супутникових технологій потребує створення нових інструментів штучного інтелекту, котрий дозволяє автоматизувати процеси виконувани вручну [1]. Інструменти штучного інтелекту відомі принципом постійного навчання на помилках, що дозволяє будувати моделі, точність яких сягає 90% і більше [2].

Розвідка територій й формування датасету з фото потребує швидкої обробки даних й побудови моделей штучного інтелекту. Довга за часом процедура створення моделі впливає на виконання завдань, оскільки за час навчання моделі змінюються розміщення досліджуваних об'єктів.

Існуючі тактики створення нових рішень щодо побудови моделей штучного інтелекту різноманітні й потребують низку відповідних витрат часу. Наприклад повний час створення моделі штучного інтелекту, що розглянута у попередньому дослідженні [3], сягає орієнтовно чотири-п'ять годин. При цьому рішення працює лише з числовими даними.

Крім витрат часу існуючі підходи зорієнтовані на вивчення конкретних завдань, зокрема для автоматичної оцінки типу та ступеня пошкодження дерев'яних конструкцій [4], де потрібно здійснювати перенавчання моделі на інших датасетах. З іншої сторони, більшість нейронних мереж будуються за принципом гібридизації, що покращує точність і збільшує складність моделі [5]. Втім, існуюча низка програмного забезпечення теж потребує ретельного вивчення функціоналу й можливостей.

Таким чином актуалізується питання створення класифікатора об'єктів на основі датасету цивільної інфраструктури.

Аналіз останніх досліджень і публікацій. Останні досягнення у сфері штучного інтелекту та робототехніки породили нові підходи щодо використання нейронних мереж для вирішення різних

практичних задач. Так напрацювання [6] пропонує підходи щодо використання штучного інтелекту у сфері будівництва. За висновками авторів [6], інтелектуальні технології використовуватимуться на етапах проектування, виробництва та експлуатації будівельних конструкцій.

Крім будівництва штучний інтелект активно використовують у наземній та повітряній робототехніці, зокрема для виявлення об'єктів різного типу [7]. Цікаво, що виявлення об'єктів відбувається у режимі реального часу й без затримок за допомогою моделі глибокого навчання YOLOv7. Вказана модель виявляє одночасно кілька об'єктів з точністю 0,97 й 92 кадрами в секунду при 0,02 мс на одне зображення.

Та можливість нейронних мереж на цьому не зупиняються, де у напрацюванні [8] розглянуто підходи щодо інтелектуальної діагностики несправностей, інструменти котрої характеризуються високими обчислювальними можливостями. Адекватність підходу доведено за допомогою датасету про підшипники кочення та даних несправності ротора. Пошук шляхів удосконалення обчислювальних можливостей комп'ютерних систем здійснювався також за допомогою ортогональних базисів й процесу паралельної обробки проєкцій вихідного числа [9]. Такий спосіб підвищує процес діагностування даних у 1,2 рази.

Нарощування обчислювальних можливостей комп'ютерної техніки корелюється з інструментарієм глибокого навчання. Як зазначено у напрацюванні [10], виявлення об'єктів на фото під різними кутами є складною задачею навіть й для глибоких нейронних мереж. У зв'язку з цим автори [10] пропонують метод візуальної кластеризації, що визначає різницю форми об'єкта дослідження з точністю 86,3%.

Паралельно з вивченням поведінки згорткових мереж щодо класифікації зображень досліджуються

й питання прогнозування різних подій. Наприклад у роботі [11] вивчається гібридна ймовірнісна модель на основі згорткових та рекурентних мереж для прогнозування навантаження домогосподарств. Принциповим у дослідженні є поєднання двох типів нейронних мереж й створення комбінованого варіанту мережі. Для вивчення вказаних тактик досліджень передбачається існування відповідного програмного забезпечення. Так, відома освітня платформа [12] поєднує Moodle та Jupyter для навчання у галузі STEM.

Створенням гібридних згорткових моделей також присвячено роботу [13], де обиралася оптимальна стратегія. Перевага запропонованого підходу досягається за рахунок нової ідеї оптимізації ознак. На відміну від результатів [13] у роботі [14] вивчалися питання паралельно діючих класифікаторів побудованих на основі методу опорних векторів та згорткової мережі.

Залежно від вихідних зображень, що являють собою досліджувану вибірку, одна й та сама технологія здатна вирішувати різні завдання, наприклад картографування зсувів [15]. За результатами досліджень встановлено території, схильні до зсувів ґрунту та визначено рівні ризику будівель.

Як видно з досліджень [6–15], використання нейронних мереж розглядається у різних сферах, де будуються як звичайні так і гібридні мережі. Побудова мережі зорієнтована на використання конкретного програмного забезпечення та вміння його об'єкти.

Для побудови власного класифікатора об'єктів цивільної інфраструктури, відповідно до рекомендації, розглянемо використання згорткової нейронної мережі.

Мета – підвищення ефективності розпізнавання образів шляхом створення бінарного класифікатора об'єктів цивільної інфраструктури засобами глибокого навчання.

Методологія дослідження

Процес створення нейронної мережі передбачає існування формальної постановки завдання дослідження. Нехай A та B є множинами світлин, що являють класи 0 й 1 відповідно. Завдання полягає у створенні класифікатора $y=f(x)$, де x є фотографією з множини A чи B . Створена модель класифікатора використовується у задачах розпізнавання образів й передбачає існування вихідного датасету.

У якості вихідного датасету дослідження використано зображення будинків з ресурсу [16]. Датасет включає 2735 зображень будівель, з них 1237 пошкоджених і 1498 непошкоджених. Формування навчального і тестового наборів фото здійснювалося випадковим чином у співвідношенні, де розмір навчальної вибірки складає $split_ratio = 0,8$.

Вихідні набори зображень використовувалися для побудови двох згорткових нейронних мереж. Побудова першої моделі нейронної мережі (Sequential) здійснювалася за відомою технологією [17] мовою програмування Python та з використанням бібліотек Tensor Flow й Keras. Для прискорення

пошуку гіперпараметрів та оптимальної архітектури мережі використано бібліотеку Keras Tuner [18], де перевага надавалася методу пошуку гіперпараметрів Random Search при $max_trials=10$.

Обмеженнями для пошуку архітектури нейронної мережі були значення кількості фільтрів у згорткових шарах, що досліджувалися у діапазоні [1, 4] з кроком 1.

Параметри пов'язаного шару теж вивчалися у діапазоні [1, 4] з кроком 1. При компіляції моделі швидкість навчання обмежувалася діапазоном [0,05, 1] з кроком 0,05, де функція втрати $loss='binary_crossentropy'$. З метою зменшення часу навчання моделі використовувався підхід ранньої зупинки, де $patience=8$.

Якість побудованих результатів дослідження фіксувалася метриками accuracy, precision, recall та f1 score, де здійснювався порівняльний аналіз навчального й тестового наборів. Додатково здійснювалася перевірка помилки моделі залежно від епохи.

Графічна інтерпретація отриманих результатів дослідження реалізовувалася засобами бібліотеки matplotlib, де будувалися візуалізації результатів процесу навчання моделі.

Побудовані моделі зберігалися за допомогою функціоналу бібліотеки Keras у файловий формат .h5, котрий використовувався у зовнішніх сервісах (Netron) для графічної інтерпретації моделі мережі [19].

Для порівняльного аналізу будувалася ще одна згорткова нейронна мережа VGG16 [17], котра є попередньо навченою. Це друга мережа. З метою зменшення обсягу часу навчання мережі здійснювалося перенавчання мережі на датасеті дослідження за обмежень використання верхнього шару шляхом додавання нових шарів. Моделювання мережі VGG16 проводилося протягом 3 епох.

Виклад основного матеріалу

Процес створення моделі згорткової нейронної мережі щодо класифікації об'єктів на зображеннях передбачає вибір оптимальної архітектури на основі вихідного датасету, табл. 1.

Як видно з табл. 1, згорткова нейронна мережа Sequential нараховує три згорткових шари, шари пулінгу, шар перетворення, пов'язаний шар та вихідний шар. Визначена архітектура нейронної мережі класифікації об'єктів на зображеннях використовується для побудови моделі.

Відповідно до методології дослідження процес побудови моделі нейронної мережі нараховував 10 епох, де фіксувалася точність на навчальній й тестовій вибірках, рис. 1.

Як видно з рис. 1 оптимальна точність навчальної й тестової вибірок досягається під час першої епохи та становить 0,93. Наступні епохи вказують на перенавчання моделі, оскільки збільшується різниця точності між навчальною й тестовою вибірками. Це пояснюється обмеженою кількістю вхідних зображень.

Результати визначення точності та помилка на навчальній й тестовій вибірках подано у табл. 2.

Таблиця 1 – Досліджувана архітектура згорткової нейронної мережі Sequential для класифікації об'єктів на зображеннях

Назва шару	Вихідна форма	Досліджувані параметри
Перший згортковий шар conv2d	(None, 254, 254, 2)	56
Шар пулінгу max_pooling2d	(None, 127, 127, 2)	0
Другий згортковий шар conv2d_1	(None, 125, 125, 2)	38
Шар пулінгу max_pooling2d_1	(None, 62, 62, 2)	0
Третій згортковий шар conv2d_2	(None, 60, 60, 2)	38
Шар пулінгу max_pooling2d_2	(None, 30, 30, 2)	0
Шар перетворення flatten	(None, 1800)	0
Пов'язаний шар з двома виходами dense	(None, 2)	3602
Вихідний шар dense_1	(None, 1)	3

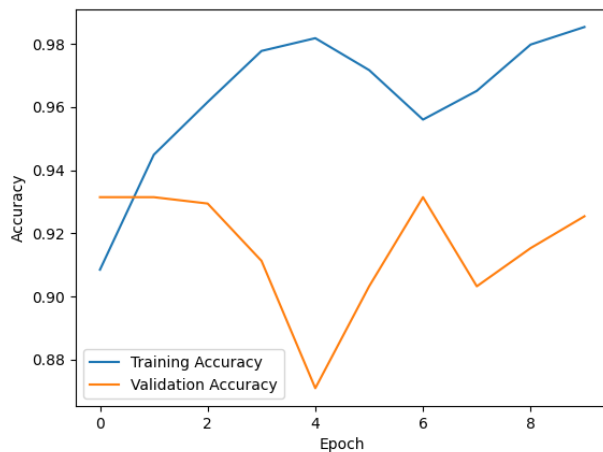


Рис. 1. Порівняльний аналіз залежності точності згорткової нейронної мережі (навчальна й тестова вибірки) Sequential від кількості епох

З точки зору балансу точності між навчальною й тестовою вибірками зображень перевагу має модель побудована під час другої епохи, де точність становить по 0,93 при помилці 0,2442/0,2595 відповідно. Але помилка під час другої епохи для навчального й тестового наборів велика. Оптимальне значення помилки при навчанні/тестуванні досяга-

Таблиця 2 – Результати навчання згорткової нейронної мережі Sequential для класифікації об'єктів на зображеннях

№ епохи	Точність на навчальній/ тренувальній вибірках	Помилка на навчальній/ тренувальній вибірках
1	0,88/0,93	0,3857/0,2843
2	0,93/0,93	0,2442/0,2595
3	0,95/0,92	0,1751/0,2628
4	0,97/0,91	0,1234/0,3922
5	0,98/0,87	0,0985/0,7616
6	0,97/0,9	0,1003/0,566
7	0,94/0,93	0,1587/0,328
8	0,96/0,9	0,1128/0,5265
9	0,97/0,91	0,0763/0,6152
10	0,98/0,92	0,0650/0,4424

ється аж на десятій епісі й становить 0,0650/0,4424 відповідно, де точність 0,98/0,92 відповідно. З одного боку, модель має ознаки перенавчання, з іншого – мінімальне значення помилки при Precision 0,93; Recall 0,91 й F1 Score 0,92.

На рис. 2 подано результати дослідження матриці плутанини найліпшої моделі мережі Sequential.

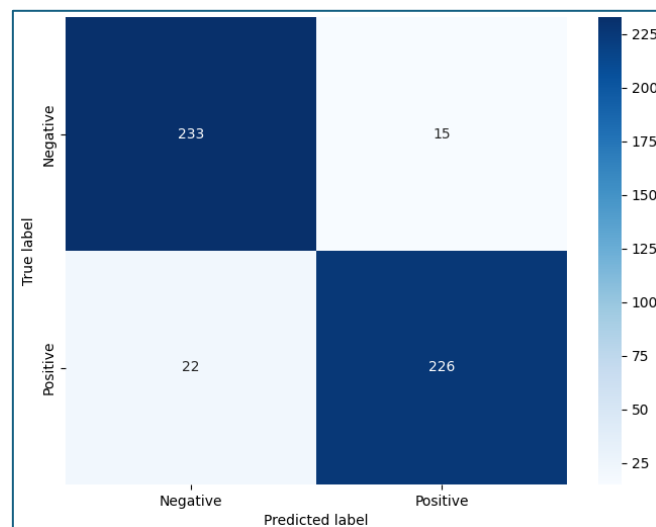


Рис. 2. Результати дослідження матриці плутанини оптимальної класифікаційної моделі мережі Sequential

З рис. 2 видно, що 233 й 226 зображень відповідних класів були вірно класифіковані, у той час як 22 й 15 зображень класифіковано хибно. Отримані хибно класифіковані зображення пояснюються обранням оптимальної моделі, котра має незначну різницю показників точності на навчальній й тестовій вибірках, а з іншої сторони мінімальне значення помилки.

Кількість хибно класифікованих зображень є різною для епох 1, 3-9, проте на вказаних епохах спостерігається процес перенавчання моделі, що доведено й на рис. 1.

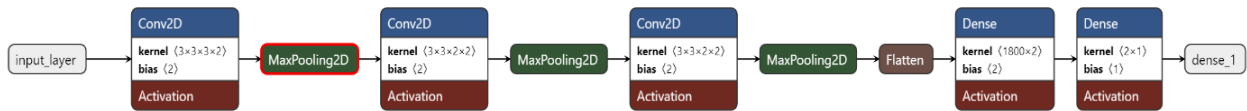


Рис. 3. Архітектура побудованої згорткової нейронної мережі для класифікації об'єктів на зображеннях створена засобами Netron

Вказана візуалізація використовувалася для демонстрації особливостей побудованої мережі й планування аспектів її майбутнього удосконалення. З метою порівняльного аналізу точності отриманих результатів першої згорткової нейронної мережі побудовано нейронну мережу VGG16. Оскільки відома мережа VGG16 попередньо навчена, але не на досліджуваному датасеті, до мережі додано досліджувану вибірку зображень – перенавчання, табл. 3.

Таблиця 3 – Результати навчання згорткової нейронної мережі VGG16 для класифікації об'єктів цивільної інфраструктури

№ епохи	Точність на навчальній/ тренувальній вибірках	Помилка на навчальній/ тренувальній вибірках
1	0,81/0,96	0,735/0,0905
2	0,97/0,96	0,0531/0,1056
3	0,99/0,96	0,0278/0,1538

З точки зору компромісу точності між навчальною й тестовою вибірками зображень цивільної інфраструктури він спостерігається за результатами другого етапу та становить 0,97/0,96 й вказує на незначні ознаки перенавчання. За критерієм помилки, мінімальне значення котрої спостерігається за результатами третьої епохи й становить для навчання/тестування 0,0278/0,1538 відповідно при точності 0,99/0,96.

Практичні рекомендації результатів дослідження

Проведене дослідження процесу побудови нейронної мережі на основі датасету будівель, за наведених у методології дослідження обмежень, дозво-

Варто відмітити, час навчання моделі такого типу складає орієнтовно до 30 хвилин, за вказаних у методології дослідження обмеженнях. У реаліях сьогодення, мінімізація витрат часу для навчання моделі є важливою з точки зору заощадження як енергоресурсів, так і швидкості впровадження моделі до використання. Остаточне рішення щодо показників якості згорткової моделі нейронної мережі схвалюється людиною, котра приймає рішення.

Додатково, для поглибленого аналізу побудованої мережі Sequential, відповідно до методології дослідження, здійснювалася її візуалізація, рис. 3.

ляє сформулювати практичні рекомендації щодо створення аналогічних моделей.

Процес побудови згорткової мережі прискорюється за рахунок використання автоматичного підбору архітектури й гіперпараметрів мережі, що реалізовано у Keras Tuner. При цьому важливим є використання Random Search, у той час як послідовний пошук збільшує час побудови моделі до 5 годин лише на одну епоху. Це важливо в умовах обмеження у часі.

Використання попередньо навченої моделі VGG16 й інтеграція до моделі датасету дослідження дозволяє перенавчити модель за 4 години, відповідно до умов вказаного дослідження.

Створені моделі можуть бути використані для бінарної класифікації зруйнованих і цілих будівель. Для автоматизації цього процесу рекомендовано здійснити інтеграцію моделі інструментами OpenCV [20] й реалізувати відповідні рішення апаратним забезпеченням, зокрема raspberry pi.

Створену модель мережі Sequential рекомендовано до використання у майбутньому, зокрема для перенавчання на нових датасетах, оскільки досвід використання VGG16 показав тенденцію зменшення помилки й підвищення точності.

Висновки

Задача побудови бінарного класифікатора об'єктів вирішується шляхом створення згорткової мережі Sequential та мережі VGG16.

Запропоновані рішення рекомендовано до використання для класифікації об'єктів цивільної інфраструктури, побудови майбутніх моделей глибокого навчання.

Крім того, запропоновані моделі у поєднанні з бібліотекою комп'ютерного зору дозволяють реалізувати принципи комп'ютерного зору.

СПИСОК ЛІТЕРАТУРИ

1. S V. Satellite Image Classification using CNN with Particle Swarm Optimization Classifier [Electronic resource] / Vidhya S, Balaji M, Kamaraj V // Procedia Computer Science. – 2024. – Vol. 233. – P. 979–987. – Mode of access: <https://doi.org/10.1016/j.procs.2024.03.287>

2. O'Shaughnessy D. Understanding Automatic Speech Recognition [Electronic resource] / Douglas O'Shaughnessy // Computer Speech & Language. – 2023. – P. 101538. – Mode of access: <https://doi.org/10.1016/j.csl.2023.101538>
3. Practical principles of integrating artificial intelligence into the technology of regional security predicting [Electronic resource] / Oleksandr Shefer [et al.] // Advanced Information Systems. – 2024. – Vol. 8, no. 1. – P. 86–93. – Mode of access: <https://doi.org/10.20998/2522-9052.2024.1.11>
4. Computing the characteristics of defects in wooden structures using image processing and CNN [Electronic resource] / Rana Ehtisham [et al.] // Automation in Construction. – 2024. – Vol. 158. – P. 105211. – Mode of access: <https://doi.org/10.1016/j.autcon.2023.105211>
5. Hybrid framework for respiratory lung diseases detection based on classical CNN and quantum classifiers from chest X-rays [Electronic resource] / G. V. Eswara Rao [et al.] // Biomedical Signal Processing and Control. – 2024. – Vol. 88. – P. 105567. – Mode of access: <https://doi.org/10.1016/j.bspc.2023.105567>
6. Artificial-intelligence-led revolution of construction materials: From molecules to Industry 4.0 [Electronic resource] / Xing Quan Wang [et al.] // Matter. – 2023. – Vol. 6, no. 6. – P. 1831–1859. – Mode of access: <https://doi.org/10.1016/j.matt.2023.04.016>
7. Yasmine G. Anti-drone systems: An Attention Based Improved YOLOv7 model for a real-time detection and identification of multi-airborne target [Electronic resource] / Ghazlane Yasmine, Gmira Maha, Medromi Hicham // Intelligent Systems with Applications. – 2023. – P. 200296. – Mode of access: <https://doi.org/10.1016/j.iswa.2023.200296>
8. Deep hypergraph autoencoder embedding: An efficient intelligent approach for rotating machinery fault diagnosis [Electronic resource] / Mingkuan Shi [et al.] // Knowledge-Based Systems. – 2022. – P. 110172. – Mode of access: <https://doi.org/10.1016/j.knsys.2022.110172>
9. Onyshchenko S. Improving the efficiency of diagnosing errors in computer devices for processing economic data functioning in the class of residuals [Electronic resource] / Svitlana Onyshchenko, Alina Yanko, Alina Hlushko // Eastern-European Journal of Enterprise Technologies. – 2023. – Vol. 5, no. 4 (125). – P. 63–73. – Mode of access: <https://doi.org/10.15587/1729-4061.2023.289185>
10. Lv X.-L. Visual clustering network-based intelligent power lines inspection system [Electronic resource] / Xian-Long Lv, Hsiao-Dong Chiang // Engineering Applications of Artificial Intelligence. – 2024. – Vol. 129. – P. 107572. – Mode of access: <https://doi.org/10.1016/j.engappai.2023.107572>
11. A hybrid CNN-GRU based probabilistic model for load forecasting from individual household to commercial building [Electronic resource] / Ming-Chuan Chiu [et al.] // Energy Reports. – 2023. – Vol. 9. – P. 94–105. – Mode of access: <https://doi.org/10.1016/j.egy.2023.05.090>
12. Development of Shared Modeling and Simulation Environment for Sustainable e-Learning in the STEM Field [Electronic resource] / Anatolijs Zabasta [et al.] // Sustainability. – 2024. – Vol. 16, no. 5. – P. 2197. – Mode of access: <https://doi.org/10.3390/su16052197>
13. Sikakollu P. Ensemble of multiple CNN classifiers for HSI classification with Superpixel Smoothing [Electronic resource] / Prasanth Sikakollu, Ratnakar Dash // Computers & Geosciences. – 2021. – Vol. 154. – P. 104806. – Mode of access: <https://doi.org/10.1016/j.cageo.2021.104806>
14. Cross-evaluation of a parallel operating SVM – CNN classifier for reliable internal decision-making processes in composite inspection [Electronic resource] / Sebastian Meister [et al.] // Journal of Manufacturing Systems. – 2021. – Vol. 60. – P. 620–639. – Mode of access: <https://doi.org/10.1016/j.jmsy.2021.07.022>
15. Hybridizing genetic random forest and self-attention based CNN-LSTM algorithms for landslide susceptibility mapping in Darjiling and Kurseong, India [Electronic resource] / Armin Moghimi [et al.] // Quaternary Science Advances. – 2024. – P. 100187. – Mode of access: <https://doi.org/10.1016/j.qsa.2024.100187>
16. Turkiye_Earthquake_2023 [Electronic resource] // Kaggle: Your Machine Learning and Data Science Community. – Mode of access: <https://www.kaggle.com/datasets/buraktaci/turkiye-earthquake-2023?resource=download>
17. François Chollet. Deep Learning with Python / François Chollet. – Shelter Island, NY : Manning Publications Co., 2018. – 361 p.
18. Introduction to the Keras Tuner | TensorFlow Core [Electronic resource] // TensorFlow. – Mode of access: https://www.tensorflow.org/tutorials/keras/keras_tuner
19. Netron [Electronic resource] // Netron. – Mode of access: <https://netron.app/>
20. Home [Electronic resource] // OpenCV. – Mode of access: <https://opencv.org/>

Received (Надійшла) 29.05.2024

Accepted for publication (Прийнята до друку) 14.07.2024

Practical use cases for creating convolutional neural network models for image recognition tasks

O. Laktionov, N. Pedchenko, A. Yanko

Abstract. The aim of the article is to improve the efficiency of image recognition by creating a binary classifier for objects of civil infrastructure using deep learning techniques. **Research results.** Convolutional neural network models have been created for the recognition of civil infrastructure objects. A Sequential convolutional neural network has been constructed, consisting of three convolutional layers, pooling layers, a transformation layer, a fully connected layer, and an output layer. The optimal error values during training/testing are 0.0650/0.4424, with accuracies of 0.98/0.92 respectively. Results from the third epoch show errors of 0.2442/0.2595 and an accuracy of 0.93. A pre-trained VGG16 model was also utilized, fine-tuned on the dataset, demonstrating minimal error values of 0.0278/0.1538 during training/testing, with accuracies of 0.99/0.96 respectively. **Scientific novelty.** Further development of using convolutional neural networks for recognizing the level of civil infrastructure damage has been achieved. **Practical significance.** Two convolutional neural network models, Sequential and VGG16, have been built to address the recognition of damaged and intact buildings. The prerequisite for detecting these objects is the use of a camera and appropriate hardware such as a Raspberry Pi or a personal computer/laptop.

Keywords: computer vision, classification tasks, deep learning, neural networks.