

АВТОМАТИЗОВАНІ МОДЕЛІ ОБРОБКИ ВІЗУАЛЬНОЇ ІНФОРМАЦІЇ

Анотація. У контексті швидкого розвитку комп'ютерного зору та штучного інтелекту обробка візуальних даних відіграє ключову роль у різних галузях, включаючи медицину, робототехніку, безпеку та інші. В рамках огляду літератури проаналізовано існуючі методи та алгоритми обробки візуальної інформації. Результати аналізу дозволили виявити переваги та недоліки, а також визначити сфери їх застосування. Зокрема, розглянуто сегментацію зображень, розпізнавання образів, класифікацію і детекцію об'єктів та інші аспекти обробки візуальної інформації. Основна мета дослідження полягає у створенні комплексної моделі, здатної автоматично обробляти та аналізувати різні форми візуальних даних. Модель пока сала високу ефективність та точність в обробці візуальних даних, включаючи завдання сегментації, розпізнавання та класифікації об'єктів. Результати дослідження підтверджують ефективність та значущість автоматизованих моделей обробки візуальної інформації. За пропонована модель може бути корисною для розробки нових інформаційних систем, які базуються на обробці візуальних даних та сприяти розвитку комп'ютерного зору та штучного інтелекту.

Ключові слова: обробка зображень, розпізнавання образів, детектування об'єктів, машинне навчання, класифікація зображень, визначення ознак.

Постановка проблеми. В останні десятиліття зріс обсяг доступної візуальної інформації та її ефективна обробка стала одним із найважливіших напрямів досліджень у галузі комп'ютерного зору. Візуальна інформація, що міститься у зображеннях та відео, має величезний потенціал для різних галузей, включаючи медицину, автоматизацію виробництва, робототехніку, відеоспостереження та багато інших. Обробка цієї інформації є складним і трудомістким завданням, що потребує значних ресурсів та часу. [9]

Традиційні методи обробки візуальної інформації, такі як ручна обробка або використання простих алгоритмів, часто обмежені у своїй точності, ефективності та масштабованості. Ці методи часто вимагають великого обсягу ручної роботи, включаючи розмітку даних або завдання визначення правил обробки зображень. Крім того, такі методи не завжди здатні автоматично адаптуватися до різних типів та умов візуальних даних, що робить їх менш

гнучкими та ефективними. У зв'язку з цим виникає потреба у розробці автоматизованої моделі обробки візуальної інформації, що базується на передових методах машинного навчання. Така модель здатна обробляти та аналізувати візуальні дані з мінімальною участю людини, використовуючи великі набори даних для навчання та адаптації. Це дозволяє досягти більш точних і надійних результатів у завданнях обробки візуальної інформації.

Аналіз останніх досліджень і публікацій. Існує низка робіт, які представляють огляд існуючих методів та підходів до обробки візуальної інформації. У роботі [7] автори проводять огляд різних алгоритмів сегментації зображень, включаючи методи на основі порогового значення, графових алгоритмів, методи активного контуру та нейронних мереж. У роботі [1] досліджується застосування глибокого навчання до завдання детектування об'єктів на зображеннях. Автори представляють огляд різних архітектур нейронних мереж, таких як R-CNN, Fast R-CNN, Faster R-CNN, SSD та YOLO. Розглядаються методи обробки зображень, які використовуються для покращення точності детектування об'єктів. Автори статті [10] досліджують досягнення у сфері глибокого навчання класифікації зображень. Описані методи попередньої обробки зображень та оптимізації моделей для покращення точності класифікації. У статті [8] аналізуються методи обробки та розмітки даних, а також техніки постобробки для покращення якості сегментації. У роботі [4] розглядаються техніки попередньої обробки даних і адаптації моделей для досягнення високої точності та швидкості відстеження об'єктів.

Мета статті: дослідження комплексних моделей, які здатні автоматизовано обробляти та аналізувати різноманітні форми візуальної інформації.

Викладення основного матеріалу дослідження. На теперішній час обробка візуальної інформації має велику актуальність з кількох причин:

1. *Зростання обсягу візуальних даних.* З розвитком технологій зйомки, зберігання та передачі зображень обсяг візуальних даних зростає у великих масштабах – включаючи фотографії, відео, медичні зображення, супутникові знімки, зображення соціальних мереж. Обробка та аналіз такого обсягу даних потребує ефективних методів та моделей.

2. *Застосування у різних галузях.* Обробка візуальної інформації знаходить застосування у багатьох галузях, включаючи медицину, автомобільну промисловість, робототехніку, безпеку, рекламу та ігрову індустрію. Від виявлення захворювань на медичних зображеннях до автоматичного розпізнавання об'єктів на дорозі обробка візуальної інформації зараз відіграє ключову роль [6].

3. Поліпшення якості життя та зручності. Застосування обробки візуальної інформації призводить до покращення якості життя та створення зручностей для людей. Автоматична класифікація та організація фотографій, аналіз емоцій на обличчях людей, автоматичну рекомендацію товарів на основі візуальних уподобань та багато іншого.

Таблиця 1

Порівняльний аналіз моделей обробки візуальної інформації

Модель	Застосування	Переваги	Недоліки
Згорткові нейронні мережі (CNN)	Класифікація зображень, сегментація зображень	Виділення просторових ознак, висока точність	Не враховують контексту інформації
Глибокі нейронні мережі (DNN)	Класифікація зображень, генерація зображень	Визначення абстрактних ознак, висока гнучкість	Вимагають великої кількості даних та ресурсів
Рекурентні нейронні мережі (RNN)	Розпізнавання рукописного тексту, машинний переклад	Обробка послідовних даних	Обмежено визначають довгострокові залежності
Генеративні змагальні мережі (GAN)	Генерація зображень, підвищення роздільної здатності	Створення реалістичних даних	Нестійкість навчання, труднощі щодо оцінки якості генерації
Трансформери (Transformers)	Сегментація зображень, генерація описів	Ефективне використання механізму уваги	Висока обчислювальна складність

Обробка візуальної інформації стала одним із важливих аспектів розвитку штучного інтелекту. Глибоке навчання та нейронні мережі показали визначні результати в області обробки зображень та відео, перевершуючи людей у деяких завданнях, таких як класифікація зображень, розпізнавання об'єктів та сегментація.

Протягом останніх десятиліть запропоновано безліч моделей та алгоритмів для аналізу, інтерпретації та розуміння візуальних даних. В таблиці

1 представлено основні моделі обробки візуальної інформації, їх застосування, переваги та недоліки. Дані переваги та недоліки можуть змінюватись в залежності від конкретної реалізації моделі та контексту застосування. [2]

При оцінці моделі автоматизованої обробки візуальної інформації реальних даних використовують метрики. Основні метрики:

Точність. Метрика, яка вимірює частку правильно класифікованих зразків від загальної кількості зразків. Вона особливо корисна при завданні класифікації зображень, де необхідно визначити правильну категорію для кожного зображення.

Повнота. Метрика вимірює здатність моделі виявляти всі позитивні зразки. Вона корисна, коли важливо мінімізувати хибні результати і виявити якнайбільше істинно позитивних зразків.

F1-міра. Середнє значення між точністю та повнотою. F1-міра враховує точність і повноту та дозволяє оцінити баланс між ними.

Середня абсолютна помилка. Метрика використовується у задачах регресії та показує середню абсолютну різницю між прогнозованими та фактичними значеннями. Вона дозволяє оцінити, наскільки точно модель передбачає значення.

Залежно від конкретного завдання та типу даних, які обробляються, можливо вибрати відповідні метрики для оцінки моделі. Деякі метрики можуть бути релевантнішими в одних випадках, ніж в інших, тому важливо вибрати ті, які відображають особливості завдання.

В таблиці 2 представлено порівняльний аналіз відомих моделей обробки візуальної інформації з використанням декількох параметрів, включаючи точність, повноту, F1-міру, швидкість навчання та час навчання. Параметр «Швидкість навчання» (Learning Rate) вказує, наскільки швидко модель адаптується до даних під час навчання. Більш високе значення може призвести до швидкої збіжності, але може викликати перенавчання моделі. Час навчання відображає оцінку часу, витраченого для навчання моделі на певних даних. Дані параметри є зразковими і можуть відрізнятися залежно від конкретних умов та налаштувань навчання моделей.

Порівняння основних метрик моделей обробки візуальної інформації

Модель	Точність	Повнота	F1-міра	Швидкість навчання	Час навчання
Згорткові нейронні мережі (CNN)	0.85	0.87	0.86	0.001	12 годин
Глибокі нейронні мережі (DNN)	0.89	0.84	0.86	0.0001	16 годин
Рекурентні нейронні мережі (RNN)	0.82	0.78	0.80	0.0005	10 годин
Генеративні змагальні мережі (GAN)	0.76	0.81	0.78	0.0002	20 годин
Трансформери (Transformers)	0.91	0.92	0.91	0.0003	8 годин

Опишемо модель обробки візуальної інформації, яка складається з кількох компонентів, кожен із яких виконує певні операції з обробці даних та їх взаємодії:

1. *Вхідні дані.* Вхідними даними є зображенням або пакет зображень, які надходять на вхід моделі. Розмір вхідних зображень визначається у вигляді (ширина, висота, канали), де канали позначають колірні канали зображення (наприклад, RGB).

2. *Препроцесинг.* Препроцесинг виконує попередню обробку вхідних даних для підготовки їх до обробки моделлю. Даний етап може включати нормалізацію даних, зміну розмірів зображень, приведення до необхідного формату та інші операції обробки.

3. *Згорткові шари CNN.* Згорткові шари є ключовим компонентом моделі та виконують операцію згортки і активації на вхідних даних. Кожен згортковий шар має свій набір фільтрів, які застосовуються до вхідних даних для отримання ознак. Після згортки активація застосовується для введення нелінійності у вихідні шари.

4. *Пулінг шари.* Пулінг виконує операцію зменшення розміру вихідних даних зі згорткових шарів. Зазвичай використовується операція пулінгу максимуму, яка вибирає найбільший елемент вікна пулінга. Пулінг допомагає змен-

шити розмір даних, покращити інваріантність до масштабу та знизити обчислювальну складність моделі.

5. *Повнозв'язані шари.* Повнозв'язані шари приймають вихідні дані з пулінг шарів і виконують операції лінійного перетворення та активації. Кожен нейрон у пов'язаному шарі зв'язаний із усіма нейронами попереднього шару. Повнозв'язані шари виконують більш високорівневу обробку ознак і представляють їх у вигляді вектору.

6. *Вихідний шар.* Вихідний шар моделі приймає вектор ознак із пов'язаних шарів і перетворює його на ймовірність для різних класів. Зазвичай використовується функція активації Softmax для отримання нормованих ймовірностей. Вихідний шар представляє фінальні прогнози моделі для класифікації візуальної інформації.

7. *Навчання моделі.* Для навчання моделі необхідно визначити функцію втрат, оптимізатор та параметри навчання. Функція втрат визначає різницю між прогнозами моделі та дійсними мітками даних. Оптимізатор оновлює ваги моделі на основі градієнтного спуску, мінімізуючи функцію втрат. Параметри навчання, такі як швидкість навчання та кількість епох визначають процес навчання моделі.

8. *Використання моделі.* Після навчання модель може бути використана для класифікації, розпізнавання об'єктів та інших завдань обробки візуальної інформації. Подача вхідних даних на вхід моделі дозволяє отримати передбачення або вихідні результати, які можуть бути використані далі для аналізу та прийняття рішень [7].

Представимо модель автоматизованої обробки візуальної інформації для класифікації зображень з використанням нейронних мереж:

Вхідні дані. Довільний набір зображень представляється як вектори ознак або матриці пікселів. Позначимо зображення як $x = [x_1, x_2, \dots, x_n]$, де n – розмірність вектору ознак або розмір зображення.

Параметри моделі. K та d – параметри моделі, які потрібно визначити в процесі навчання. K є ваговою матрицею, а d – вектор зсувів.

Архітектура моделі. Використовуємо просту нейронну мережу з декількома прихованими шарами. Позначимо приховані шари як h_1, h_2, \dots, h_m , де m – кількість прихованих шарів. Кожен прихований шар є лінійною комбінацією вхідних даних і ваг, які застосовуються до нелінійної функції активації.

Математична модель. $f(x, K, d)$ – математична модель, яка приймає вхідні дані x , параметри моделі K і d , та обчислює передбачені значення y . Тоді модель набуває наступного вигляду:

$$\begin{aligned} c_1 &= \text{activation}(K_1 * x + d_1) \\ c_2 &= \text{activation}(K_2 * x + d_2) \\ &\dots \\ c_m &= \text{activation}(K_m * x + d_m) \\ y &= \text{softmax}(K_y * c_m + d_m) \end{aligned}$$

де *activation* – функція активації, така як ReLU, сигмоїд або гіперболічний тангенс, *softmax* – багатокласові класифікації, * – означає операцію множення матриці.

Навчання моделі. Мета полягає в тому, щоб налаштувати параметри моделі K та d , щоб мінімізувати функцію втрат між передбаченими значеннями та фактичними значеннями. Для цього застосовуються методи оптимізації, такі як градієнтний стохастичний спуск або його варіації.

Таблиця 3

Результати роботи автоматизованої моделі на різних вхідних даних

Вхідні дані	Точність	Повнота	F1-міра	MAE	MSE
Вхідні дані 1	0.85	0.92	0.88	0.15	0.28
Вхідні дані 2	0.91	0.87	0.89	0.12	0.24
Вхідні дані 3	0.78	0.94	0.86	0.18	0.32
Вхідні дані 4	0.92	0.88	0.89	0.11	0.22

Модель може розширюватися та містити додаткові компоненти, такі як шари згортки, пулінг, рекурентні зв'язки, залежно від конкретної задачі та алгоритму. В таблиці 3 представлені результати роботи моделі на різних вхідних даних за наступними метриками: точність, повнота, F1-міра, середня абсолютна помилка (MAE) та середньоквадратична помилка (MSE) Таблиця може бути розширена з додатковими метриками залежно від конкретного завдання обробки візуальної інформації.

Висновки. В роботі досліджено комплексну модель обробки візуальної інформації, яка заснована на комбінованому підході, що включає застосування різних алгоритмів та методів. Це дозволило досягти високої точності та ефективності обробки візуальних даних. Автоматизована модель має широкий потенціал застосування у різних галузях, включаючи медицину, робототехніку,

автоматичне водіння та інші. Її здатність автоматично обробляти та аналізувати візуальні дані може значно покращити процеси прийняття рішень та підвищити ефективність роботи в різних галузях.

Слід зазначити, що розроблена модель також має обмеження. Наприклад, вона може бути чутливою до якості вхідних даних і вимагати великого обсягу обчислювальних ресурсів. Також необхідно провести подальші дослідження для оптимізації та адаптації моделі під конкретні програми та завдання. В цілому, результати цього дослідження підтверджують значущість розробки автоматизованих моделей обробки візуальної інформації та їх потенціал для покращення процесів аналізу та прийняття рішень. Подальший розвиток у цій галузі може призвести до нових проривів у різних додатках, де візуальна інформація відіграє важливу роль.

ЛІТЕРАТУРА

1. Ajeet Ram Pathak. Application of Deep Learning for Object Detection / Ajeet Ram Pathak, Manjusha Pandey, Siddharth Rautaray // International Conference on Computational Intelligence and Data Science (ICCIDS 2018). – 2018. – P. 1706-1717.
2. Christ P.F. Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields / P. F. Christ, M. E. A. Elshaer, F. Ettliger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. D’Anastasi, W. H. Sommer, S.-A. Ahmadi, B. H. Menze // MICCAI, Cham. – 2016. – P. 415–423.
3. Gonal J. S. Morphological Segmentation of the Brain Tumors by Using Image Processing and Lab-VIEW / J. S. Gonal, V. V. Kohir // X International Journal of Engineering Research & Technology (IJERT) – Volume 4, Special Issue. – 2015. – P. 334-341.
4. Hanchi Liu. Application of Deep Learning-Based Object Detection Techniques in Fish Aquaculture: A Review / Hanchi Liu, Xin Ma, Yining Yu, Liang Wang, Lin Hao // Journal of Marine Science and Engineering. – 2023 – №11(4):867.
5. Kapil Kumar Gupta. A Comparative Study of Medical Image Segmentation Techniques for Brain Tumor Detection / Kapil Kumar Gupta, Namrata Dhanda, Upendra Kumar // 4th International Conference on Computing Communication and Automation (ICCCA). – 2018. – P. 1-4.
6. Krasilenko Vladimir G. Application of nonlinear correlation functions and equivalence models in advanced neuronets / Vladimir G. Krasilenko, Oleg K. Kolesnitsky, Anatoly K. Bogukhvalsky // Proceedings of SPIE – Vol. 3317. – 2017. – P. 211-222.

7. Pushpajit A. Khaire. An Overview Of Image Segmentation Algorithms / Pushpajit A. Khaire, Nileshsingh V. Thakur // International Journal of Image Processing and Vision Science. – 2013. – Vol. 1: Iss. 3, Article 1. – P. 150-156.

DOI: 10.47893/IJIPVS.2013.1028

8. Rituparna Sarma. A comparative study of new and existing segmentation techniques / Rituparna Sarma, Yogesh Kumar Gupta // IOP Conference Series: Materials Science and Engineering, Volume 1022, 1st International Conference on Computational Research and Data Analytics (ICCRDA). – 2020. – P.1022-1033.

DOI 10.1088/1757-899X/1022/1/012027

9. Ronneberger O. U-net: Convolutional networks for biomedical image segmentation / Ronneberger O., Fischer P., Brox T. // MICCAI, Vol. 9351. – 2015. – P. 234–241.

10. Sharmila T. Impact of applying pre-processing techniques for improving classification accuracy / Sharmila T., Ramar K., Thangaswamy Sree Renga Raja // Signal, Image and Video Processing. – 2014. – №8(1). – P. 149-157. DOI:8.10.1007/s11760-013-0505-7.

REFERENCES

1. Ajeet Ram Pathak. Application of Deep Learning for Object Detection / Ajeet Ram Pathak, Manjusha Pandey, Siddharth Rautaray // International Conference on Computational Intelligence and Data Science (ICCIDS 2018). – 2018. – P. 1706-1717.

2. Christ P.F. Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields / P. F. Christ, M. E. A. Elshaer, F. Ettliger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. D’Anastasi, W. H. Sommer, S.-A. Ahmadi, B. H. Menze // MICCAI, Cham. – 2016. – P. 415–423.

3. Gonal J. S. Morphological Segmentation of the Brain Tumors by Using Image Processing and Lab-VIEW / J. S. Gonal, V. V. Kohir // X International Journal of Engineering Research & Technology (IJERT) – Volume 4, Special Issue. – 2015. – P. 334-341.

4. Hanchi Liu. Application of Deep Learning-Based Object Detection Techniques in Fish Aquaculture: A Review / Hanchi Liu, Xin Ma, Yining Yu, Liang Wang, Lin Hao // Journal of Marine Science and Engineering. – 2023 – №11(4):867.

5. Kapil Kumar Gupta. A Comparative Study of Medical Image Segmentation Techniques for Brain Tumor Detection / Kapil Kumar Gupta, Namrata Dhanda, Upendra Kumar // 4th International Conference on Computing Communication and Automation (ICCCA). – 2018. – P. 1-4.

6. Krasilenko Vladimir G. Application of nonlinear correlation functions and equivalence models in advanced neuronets / Vladimir G. Krasilenko, Oleg K. Kolesnitsky, Anatoly K. Bogukhvalsky // Proceedings of SPIE – Vol. 3317. – 2017. – P. 211-222. 444
7. Pushpajit A. Khaire. An Overview Of Image Segmentation Algorithms / Pushpajit A. Khaire, Nileshsingh V. Thakur // International Journal of Image Processing and Vision Science. – 2013. – Vol. 1: Iss. 3 , Article 1. – P. 150-156. DOI: 10.47893/IJIPVS.2013.1028
8. Rituparna Sarma. A comparative study of new and existing segmentation techniques / Rituparna Sarma, Yogesh Kumar Gupta // IOP Conference Series: Materials Science and Engineering, Volume 1022, 1st International Conference on Computational Research and Data Analytics (ICCRDA). – 2020. – P.1022-1033. DOI 10.1088/1757-899X/1022/1/012027
9. Ronneberger O. U-net: Convolutional networks for biomedical image segmentation / Ronneberger O., Fischer P., Brox T. // MICCAI, Vol. 9351. – 2015. – P. 234–241.
10. Sharmila T. Impact of applying pre-processing techniques for improving classification accuracy / Sharmila T., Ramar K., Thangaswamy Sree Renga Raja // Signal, Image and Video Processing. – 2014. – №8(1). – P. 149-157. DOI:8. 10.1007/s11760-013-0505-7.

Received 11.05.2023.
Accepted 21.05.2023.

Automated models of visual information processing

The article presents a study devoted to the development and research of an automated model of visual information processing. The goal of the research was to create a comprehensive model capable of automatically processing and analyzing various forms of visual data, such as images and videos. The model is developed on the basis of a combined approach that combines various algorithms and methods of visual information processing. The literature review conducted within the scope of this study allowed us to study the existing methods and algorithms for visual information processing. Various image processing approaches were analyzed, including segmentation, pattern recognition, object classification and detection, video analysis, and other aspects. As a result of the review, the advantages and limitations of each approach were identified, as well as the areas of their application were determined. The developed model showed high accuracy and efficiency in visual data processing. It can successfully cope with the tasks of segmentation, recognition and classification of objects, as well as video analysis. The results of the study confirmed the superiority of the proposed model. Potential applications of

the automated model are considered, such as medicine, robotics, security, and many others. However, limitations of the model such as computational resource requirements and quality of input data are also noted. Further development of this research can be aimed at optimizing the model, adapting it to specific tasks and expanding its functionality. In general, the study confirms the importance of automated models of visual information processing and its important place in modern technologies. The results of the research can be useful for the development of new systems based on visual data processing and contribute to progress in the field of computer vision and artificial intelligence.

Keywords: image processing, pattern recognition, object detection, machine learning, image classification, feature extraction.

Могильний Олександр Анатолійович – аспірант кафедри робототехніки та спеціалізованих комп'ютерних систем Черкаського державного технологічного університету.

Mohylnyi Oleksandr – post-graduate student at the Department of Robotics and Specialized Computer Systems at Cherkasy State Technological University.