

<https://doi.org/10.31891/2219-9365-2024-79-16>

УДК 004.9

АХМЕТШИНА Людмила

Дніпровський Національний Університет ім. Олесея Гончара

ORCID: <https://orcid.org/0000-0002-5802-0907>

e-mail: [akhmlu1@gmail.com](mailto:akhmlu1@gmail.com)

НЕСТЕРЕНКО Антон

Дніпровський Національний Університет ім. Олесея Гончара

e-mail: [an.nestere@gmail.com](mailto:an.nestere@gmail.com)

## АНАЛІЗ АЕРОФОТОЗНІМКІВ З ВИКОРИСТАННЯМ ЗГОРТКОВОЇ НЕЙРОННОЇ МЕРЕЖІ-АВТОКОДУВАЛЬНИКА

Метою цієї роботи було створення автономної системи, здатної автоматично виявляти аномалії на аерофотознімках, отриманих за допомогою безпілотних літальних апаратів. Аномальними у даному контексті вважаються об'єкти та сегменти зображення, які мають нетипові для даних зображень характеристики: яскравість, колір, текстуру або форму (наприклад рукотворні об'єкти та транспортні засоби у загородній місцевості, люди у обмежених зонах тощо). У дослідженні розглядаються інформативні можливості згорткових нейронних мереж (ЗНМ) для вирішення задачі виявлення аномальних об'єктів на аерофотознімках. Запропоновано структуру згорткового автокодувальника з трьома кодуєчими та трьома декодуєчими згортковими нейронними шарами, що забезпечує можливість навчання на немаркованих зображеннях і розпізнавання раніше невідомих типів аномалій. Крім того, ми розглядаємо вплив застосування алгоритмів попередньої обробки на швидкість і ефективність системи. Запропоновано метод локалізації потенційно аномальних сегментів за допомогою алгоритму знаходження контурів. Розроблена система складається з моделі згорткового автокодувальника, двох алгоритмів попередньої обробки, заснованих на виявленні порогів кольору та яскравості, алгоритму виявлення контурів і класифікатора, що зважає на вихідні дані НМ. Розроблена модель НМ була навчена на декількох наборах даних, що містять аерофотознімки без аномалій. Отримана система була протестована на реальних даних, результати цього тестування наведені у даній статті.

Ключові слова: машинне навчання, комп'ютерний зір, згорткова нейронна мережа, автокодувальник, аерофотозйомка.

AKHMETSHINA Ludmila, NESTERENKO Anton

Oles Honchar Dnipro National University

## ANALYSING AERIAL PHOTOGRAPHS USING A CONVOLUTIONAL AUTOENCODER NETWORK

The purpose of this work was to create a functioning autonomous system capable of automatically detecting anomalies in aerial images obtained using unmanned aerial vehicles. Anomalous, in this context, are objects and segments of the image that do not fit with the general picture of the investigated area, such as man-made objects and vehicles in the wilderness, people in restricted areas, etc. The research considers the informative capabilities of convolutional neural networks (CNN) for solving the task of detecting abnormal objects in photographs. The structure of convolutional autoencoder with three encoding and three decoding convolutional neural layers is proposed, providing the capability of learning on unlabeled images and recognizing previously unknown types of anomalies. In addition, we study the impact of applying preprocessing algorithms on the speed and effectiveness of the system. Such algorithms include detection of color and brightness threshold, as well as finding the contours of objects that stand out in the image. We also propose a method of localizing potentially abnormal segments using the combination of these algorithms. We show that this approach allows for a great increase in performance at the cost of a very marginal increase in the number of anomalies missed.

The resulting system consists of a convolutional autoencoder model, two preprocessing algorithms based on detecting thresholds of color and brightness, an algorithm for contour detection and a final classifier, weighting the outputs. The developed NN model was trained on several data sets containing aerial photography images with no anomalies. The resulting system was tested out on real-world data, and the results of this testing are provided here.

Keywords: machine learning, computer vision, convolutional neural network, autoencoder, aerial imagery.

## ПОСТАНОВКА ПРОБЛЕМИ У ЗАГАЛЬНОМУ ВИГЛЯДІ ТА ЇЇ ЗВ'ЯЗОК ІЗ ВАЖЛИВИМИ НАУКОВИМИ ЧИ ПРАКТИЧНИМИ ЗАВДАННЯМИ

Зображення, зняті з літальних апаратів, зазвичай містять складну картину різних типів земної поверхні, рослинності та рельєфу, а також можуть містити стандартні для місцевості рукотворні об'єкти. У багатьох практичних завданнях важливим етапом їх аналізу є виявлення об'єктів інтересу (аномалій), які мають нетипові для даних зображень характеристики, наприклад, яскравість, колір, текстуру або форму (плавальний засіб на воді, літак у небі, автомобіль на місцевості). Для автоматизації вирішення таких завдань дедалі більшого поширення набувають методи штучного інтелекту. Зокрема для аналізу складних зображень використовуються згорткові нейронні мережі (НМ), які здатні синтезувати багаторівневі карти ознак, що дозволяє ефективно вирішувати завдання класифікації візуальних об'єктів [1].

Однією з численних варіацій НМ є архітектура «автокодувальник», яка базується на реконструюванні вхідного зображення на підставі складеного в процесі навчання набору ознак. Класичні НМ-автокодувальники широко використовуються в дефектоскопії та в аналізі часових рядів [2, 3]. Вони показують високу ефективність у завданнях аналізу простих, однотипних даних, що характеризуються обмеженою кількістю складових структур з незначною мінливістю та варіативністю, у тому числі простих зображень (що не містять складні сцени). Однак, зі збільшенням складності зображень та за наявності на них безлічі різноманітних об'єктів, точність та швидкість їх роботи недостатні.

Ця робота присвячена розробці НМ-автокодувальника, архітектура якого включає згорткові нейронні шари, що забезпечують виявлення аномальних сегментів зображень з високою точністю. Окрім цього використовуються декілька алгоритмів первинного аналізу з метою підвищення швидкості роботи НМ.

### АНАЛІЗ ВІДОМИХ РІШЕНЬ І ПОСТАНОВКА ЗАДАЧІ ДОСЛІДЖЕННЯ

Класичні методи аналізу зображень засновані на виконанні комплексу операцій з метою поліпшення їх візуального сприйняття, або перетворення у форму, зручну для візуального або машинного аналізу. Наприклад, у роботі [4] наводиться порівняння ефективності модифікацій популярного алгоритму Reed-Xiaoli (RX), в основі якого лежить порівняльний аналіз вихідних характеристик яскравості пікселів зображення з усередненим значенням фону у вікні обробки. У статті [5] описується аналогічний підхід, у якому використовуються матриці Кірхгофа для порівняння характеристик сусідніх пікселів.

Нині у напрямках "Комп'ютерний зір" і "Обчислювальний інтелект" з'явилась велика кількість наукових праць, присвячених процесу комп'ютерного отримання інформації (синтезу описів, зразків) із зображень. Наприклад, у галузі машинного навчання застосування НМ-автокодувальників є поширеним підходом для вирішення задач пошуку аномалій у галузях виявлення дефектів та аналізу медичних знімків завдяки можливості їх навчання методом «без вчителя» [6, 7]. З метою виключення надмірності та шумових складових навчальних вихідних даних, використовуються змагальні моделі автокодувальників або генеративні нейронні мережі [8, 9, 10]. У свою чергу, для вирішення завдань класифікації фото-даних активно використовуються «глибокі» згорткові НМ з великою кількістю шарів.

Основною для сучасних архітектур згорткових НМ стала архітектура VGG16, яка була представлена в 2015 році і включала 16 згорткових шарів [11].

На відміну від завдань класифікації об'єктів, задача виявлення аномалій ускладнюється тим, що на етапі створення та навчання моделі, як правило, доступні лише "базові зображення" – тобто такі, що не містять аномалій. Крім цього, «аномаліями», залежно від сфери застосування, можуть вважатися найрізноманітніші об'єкти та явища, які неможливо врахувати під час навчання повною мірою. Виходячи з цього система, яка здійснює пошук аномальних ділянок на зображеннях, повинна мати наступні характеристики:

1. можливість навчання з використанням лише "базових зображень";
2. здатність кваліфікувати об'єкти інтересу, які не зустрічаються в навчальному наборі даних як аномальні;
3. відсутність чи мінімізація кількості хибних спрацьовувань під час роботи з даними, схожими, але не ідентичними навчальним;
4. висока швидкодія, що забезпечує роботу системи у реальному часі.

### МЕТА СТАТТІ

Метою роботи є створення аналітичної системи для виявлення аномалій на фотознімках, знятих з літальних апаратів на основі НМ-автокодувальника, яка, за рахунок використання згорткових нейронних шарів та алгоритмів первинного виявлення потенційно аномальних сегментів дозволяє швидко виявляти складні багаторівневі ознаки об'єктів інтересу.

### ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

У рамках розв'язання задачі аналізу зображень з метою пошуку аномальних об'єктів інтересу була розроблена модель НМ, що являє собою НМ-автокодувальник із шарами згортки. Автокодувальники, як правило, складаються з вхідного та вихідного шарів однакової розмірності, а також набору прихованих шарів, що знаходяться між ними. Для вирішення більшості завдань вони проектуються з використанням повністю пов'язаних шарів, однак у запропонованій архітектурі кожен шар автокодувальника складається з шару згортки та об'єднання.

Основними механізмами роботи згорткових нейронних мереж є процедури згортки та об'єднання. Операція згортки (рис. 1) полягає в покроковому застосуванні операції скалярного добутку до сегментів зображення з покроковим зсувом. У якості множника використовується так званий «фільтр» - невелика матриця чисел фіксованого розміру. Результатом операції згортки є набір матриць, які називають «картами ознак».

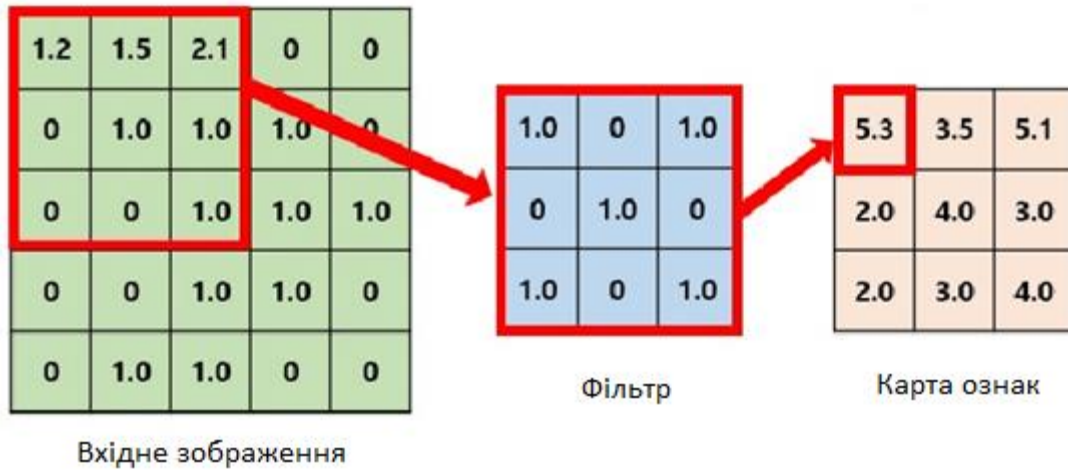


Рис. 1. Операція згортки

Таким чином, для вхідного зображення  $f$  та фільтра  $h$  значення карти ознак можна обчислити за формулою:

$$G[m, n] = \sum_j^s \sum_k^s h[j, k] f[m - j, n - k], \quad (1)$$

де  $m$  і  $n$  – індекси рядів та стовпців результуючої матриці, а  $s$  – розмір фільтра. Після формування карт ознак кожне значення модифікується за допомогою активаційної функції. У згорткових НМ широко використовується функція ReLU (лінійний випрямляч):

$$G_1(x) = \max(0, G), \quad (2)$$

Операція об'єднання проводиться для зниження розмірності карт ознак і полягає у зведенні кількох значень у вхідній матриці одного значення. Найбільш поширеним є використання максимального значення у вікні обробки

$$G_2[m, n] = \max_{s \times s}(G_2), \quad (3)$$

однак є інші варіації, наприклад, знаходження середнього значення.

У класичній архітектурі згорткової НМ за кожним шаром згортки слід шар об'єднання, а потім додаються кілька пов'язаних повністю шарів, які визначають належність вхідного зображення до одного з відомих класів. Цей підхід забезпечує отримання багаторівневих ознак, які представлені у вигляді ієрархічних матриць. На відміну від цього, НМ-автокодувальник включає дві частини, в першій приховані шари (кодувальник) мають розмірність, що послідовно зменшується, а в другій - збільшується (декодувальник). У процесі його навчання на кожному шарі формуються карти ознак, властиві вхідним даним. На рис. 2 показаний приклад карт ознак, які витягуються кожному з шарів під час аналізу зображення.

Шари декодувальника пов'язані з шарами кодувальника і дублюють карти їх ознак для використання в процесі декодування. Коли модель навчається реконструкції вхідного зображення, у якості цільової функції використовується мінімізація середньоквадратичного відхилення (MSE).

$$L = \frac{1}{nm} \sum_j^n \sum_j^m (I'_{ij} - I_{ij})^2 \quad (4)$$

де  $I$  – вхідне зображення, а  $I'$  – результуюче.

У розробленій архітектурі використовується три згорткові шари, для кодування зображення і ще три – для декодування. Кожен шар меншої розмірності має пропорційно більше згорткових фільтрів. Використання такої архітектури дозволяє автокодувальнику отримувати кілька рівнів ознак, які дозволяють більш детально кодувати реальні зображення. Схематично згорткова НМ-автокодувальник представлена на рис. 3.

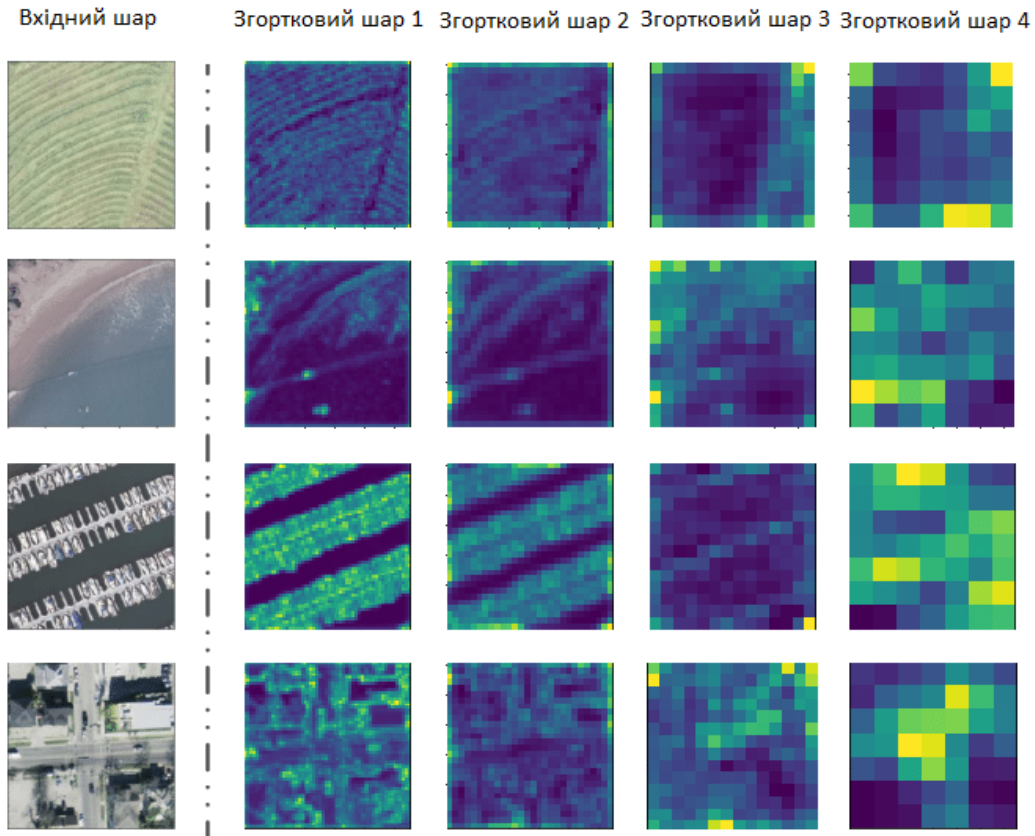


Рис. 2. Візуалізація прихованих шарів згорткової НМ

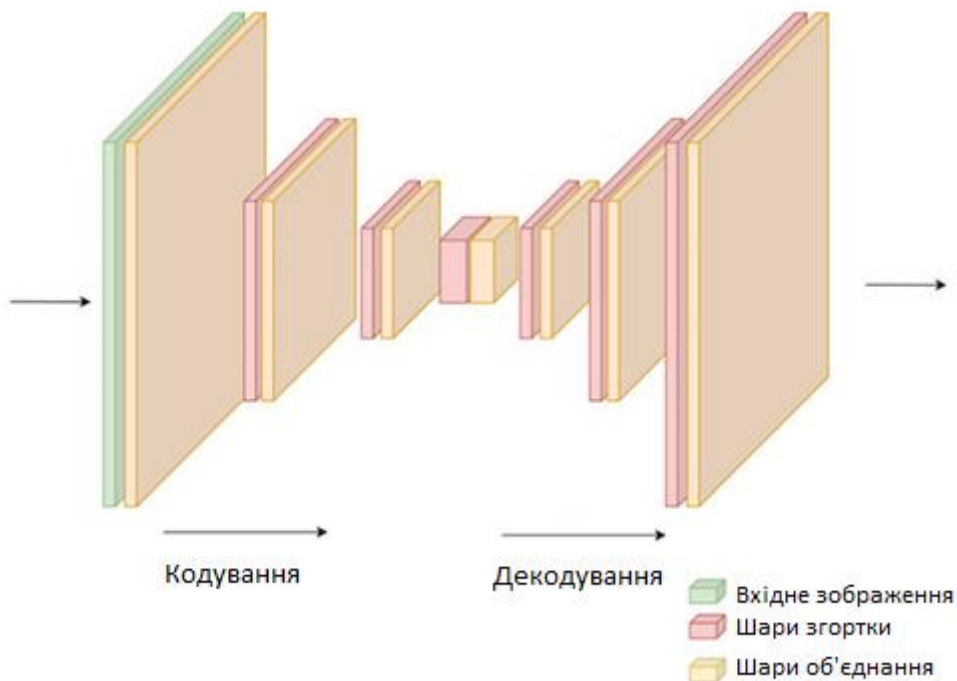


Рис. 3. Схематичне представлення архітектури згорткового НМ-автокодувальника

Для тренування розробленого НМ-автокодувальника використовується набір базових зображень без аномалій, які розбиваються на сегменти. Як метод оптимізації використовувалася варіація методу

стохастичного градієнтного спуску, що описаний у роботі [12]. Далі, до навченої моделі передаються нові зображення. Ті з них, які моделі вдалося реконструювати найменш точно – вважаються аномальними. Для оцінки точності реконструкції використовується функція середньоквадратичного відхилення (MSE), її поріг визначається експериментально.

На додачу до НМ, з метою підвищення швидкодії системи, розроблено алгоритми первинної обробки зображень та визначення на ньому потенційно аномальних ділянок. Цими алгоритмами є аналіз порогів кольору та яскравості, а також пошук закритих контурів. Аналіз перепадів кольору та яскравості виконується за допомогою швидкого та ефективного алгоритму знаходження оператора Лапласа для матриці сусідніх пікселів, що описано у роботі [13].

Наступним етапом аналізу зображення є застосування алгоритму пошуку контурів для попереднього виявлення кластерів пікселів, які можуть являти собою об'єкти, що виділяються. Як вхідні дані цьому алгоритму передаються три варіанти зображень: вихідне, а також два результуючих зображення після фільтрації двома попередніми методами. Алгоритм пошуку контурів реалізує метод, описаний у публікації [14]. В результаті роботи алгоритм і визначає координати виявлених кластерів пікселів, що виділяються, розмір яких перевищує задане порогове значення. Такі кластери потенційно можуть бути об'єктами інтересу. Ці кластери вписуються у квадрати фіксованого розміру і підлягають подальшому аналізу за допомогою нейронної мережі. На рис. 4 показаний результат роботи алгоритму пошуку контурів, в результаті якої були виявлені дві ділянки, що підлягають подальшому аналізу, з яких одна дійсно містить аномалію (надувний човен).

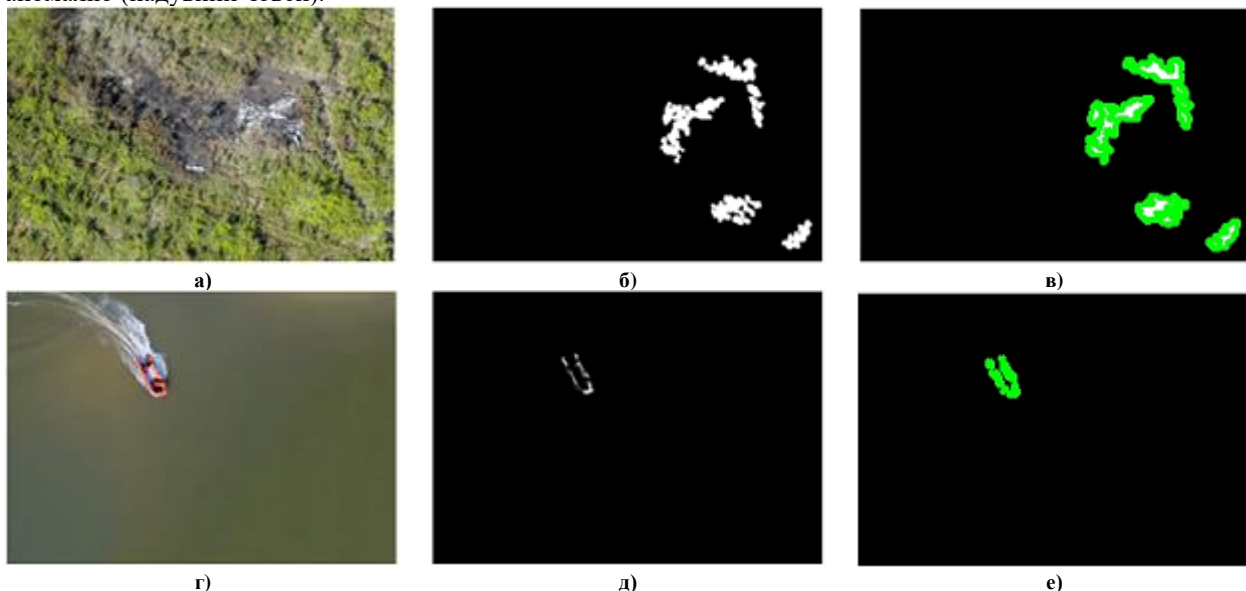


Рис. 4. Приклад роботи алгоритмів первинної обробки. а), г) – оригінальні зображення. б) д) – порогові алгоритми. в) е) – знайдені контури

На рис. 4 представлена загальна схема розробленого програмного забезпечення для аналізу зображень з використанням НМ-автокодувальника із згортковими шарами. Аналітична система реалізована мовою програмування Python із використанням бібліотек OpenCV, Scikit-image, Tensorflow та Keras.

У якості вихідних даних для тренування та тестування були використані кольорові RGB-зображення аерофотозйомки, що містять типову для помірної кліматичної зони місцевість, схожі за типом, але взяті з різних джерел та різних географічних локацій.

Оскільки в більшості сфер потенційного застосування такої системи аномальні ділянки зустрічаються не часто, передбачено використання алгоритмів попередньої локалізації потенційно аномальних ділянок зображення на основі виявлення кольорних і яскравих перепадів, а також пошуку закритих контурів. Лише ті ділянки зображень, які є потенційно аномальними, передаються в нейронну мережу для аналізу. У процесі навчання вихідні дані поділяються на набір навчання, набір валідації та набір тестування, але лише набір тестування містив аномалії. Результатом навчання є вагова модель нейронної мережі, яка надалі використовується для виявлення аномалій у нових даних.

У розробці та тестуванні системи був використаний набір даних із 192 зображення високої роздільної здатності, отриманих у різних географічних локаціях. З усього набору 28 зображень містили аномальні об'єкти, 164 - не містили. Оскільки деякі зображення містили більше одного аномального об'єкта, загальна кількість аномальних об'єктів у тестовому наборі даних становила 36. При тестуванні аномальними вважалися рукотворні об'єкти нестандартної форми, автомобілі, плавзасоби на водній поверхні тощо. Будинки стандартної форми аномальними об'єктами не вважалися.

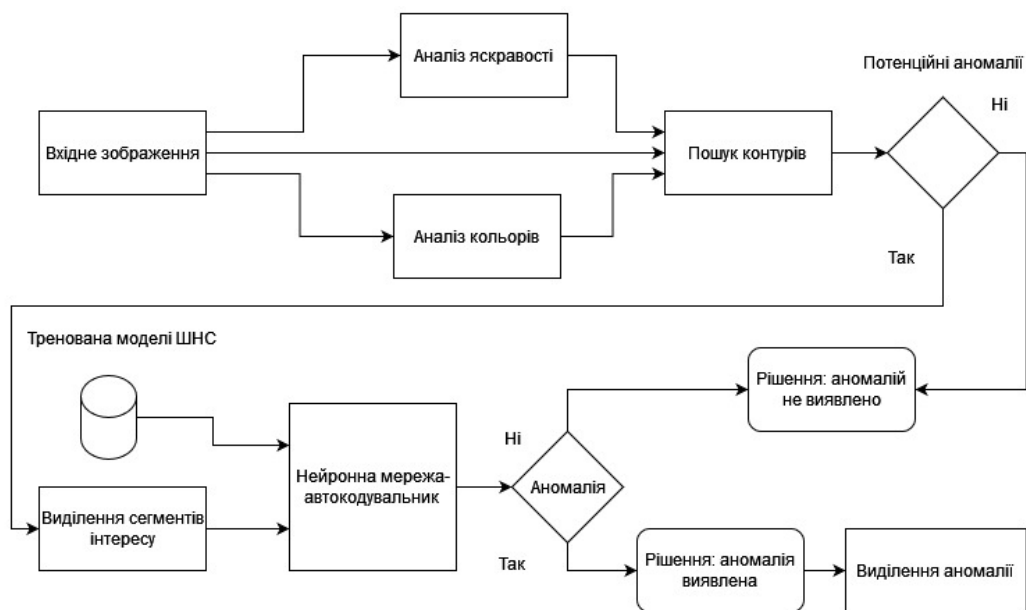


Рис. 5. Схема роботи аналітичної системи

Експериментально було встановлено, що з урахуванням дозволу та деталізації зображень при аналізі доцільно використовувати сегменти даних розміром 128x128 пікселів. Тоді, приховані шари, мають розмірність 64x64, 32x32 і 16x16 відповідно. На першому та другому рівнях шарів використовувалися фільтри, розмірністю 5x5, на третьому та четвертому – 3x3. Операція об'єднання здійснюється за допомогою вікна зсуву 2x2 на кожному із шарів.

Для навчання було використано 100 зображень різних розмірів, розбиті на фрагменти по 128x128 пікселів. Загальна кількість фрагментів склала 46875. Навчання автокодувальника проводилося протягом 40 epoch. На рис. 6 наведено графік сходження функції втрати під час навчання.

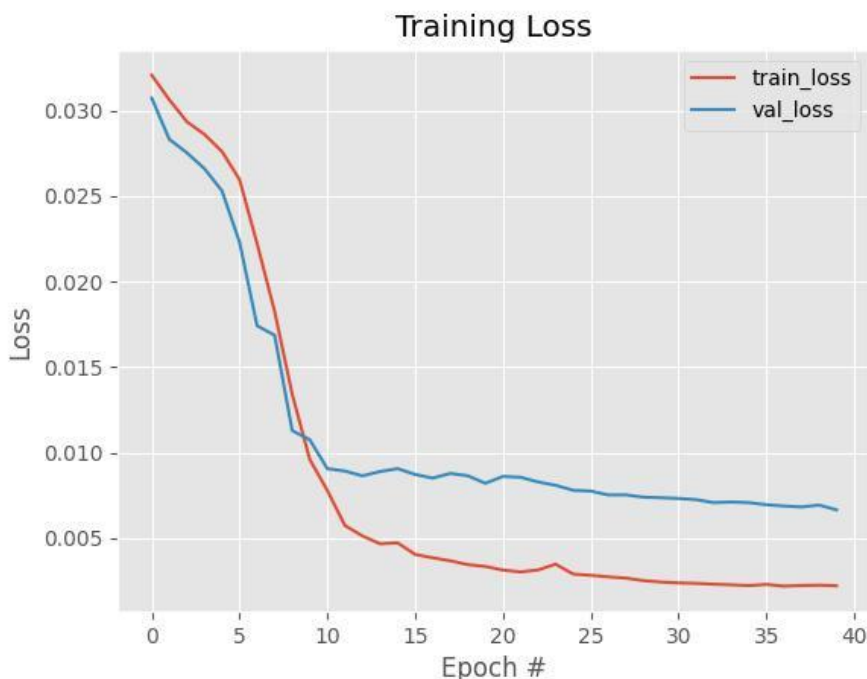


Рис. 6. Графік сходження функції втрати (тренувальний та валідаційний набори даних)

Після навчання моделі, 28 зображень, що містять аномалії, а також 64 зображення, що не містять їх (загалом 92) було використано для тестування системи. За результатами тестування було досягнуто наступних результатів. З 36 аномальних об'єктів було успішно розпізнано 31, що становить 86% від загальної кількості аномалій. Хибно-позитивні спрацювання (кількість сегментів зображень, некоректно

виділених як аномальні) склали 15 сегментів. Середній час обробки зображень становив 3.2 с, максимальне – ~8 с. На рис. 7 наведено приклад успішно виявленої аномалії (надувний човен) на зображенні, що містить берег річки. Лише сегмент, що містить човен, перевищив поріг функції середньоквадратичного відхилення, що дорівнює 0.05.

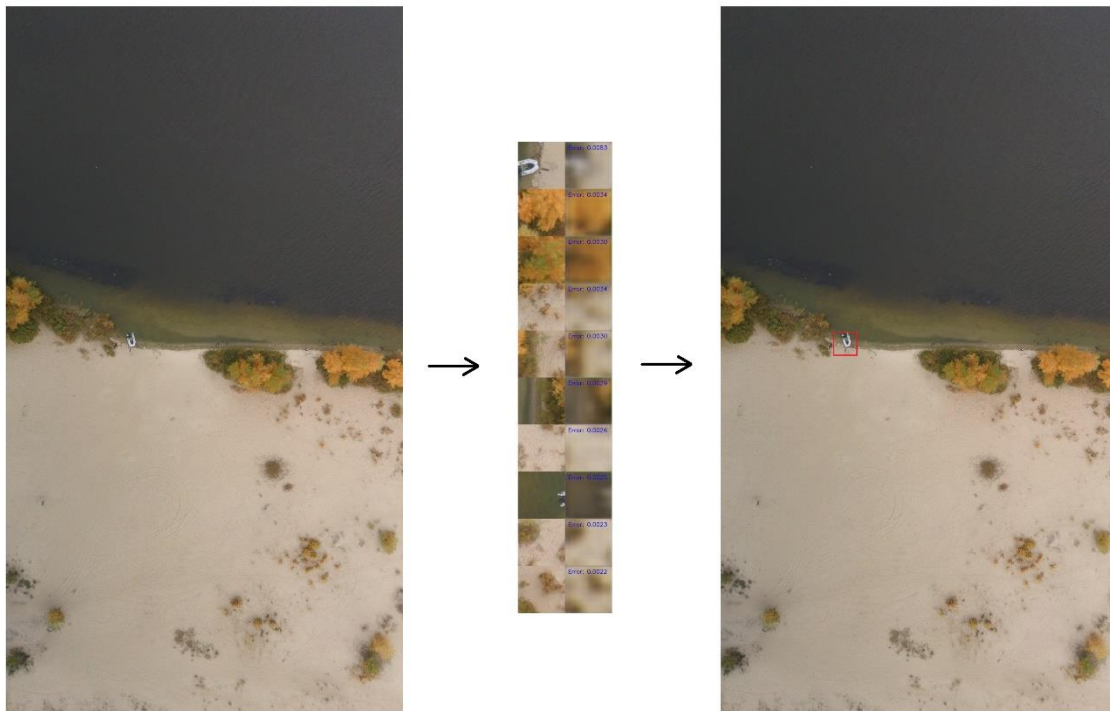


Рис. 7. Приклад успішно розпізнаного аномального сегмента

З метою порівняння ефективності розробленої системи було зроблено порівняння швидкості її роботи з методом Single-Class SVM (Support Vector Machine), а також із використанням автокодувальника без попереднього відсікання зображень. Метод Single-Class SVM продемонстрував надзвичайно високий ступінь хибно-позитивних спрацьовувань зважаючи на його нездатність адаптуватися до комплексних елементів ландшафту, представлених на зображеннях. На 92 тестових зображеннях 326 сегментів було розпізнано як аномальні, при цьому майже 30% справжніх аномалій були пропущені.

Автокодувальник без попереднього аналізу зображень більш простими алгоритмами показав незначний приріст успішно розпізнаних аномалій за рахунок значного зниження швидкості та підвищення хибно-позитивних спрацьовувань. Це порівняння продемонструвало, що запропонований підхід дозволяє досягти найменшого значення хибно-позитивних спрацьовувань за практично повної відсутності зростання пропущених аномалій. Швидкість роботи системи знаходиться на прийнятному рівні. Візуалізація результатів показана рис. 8.

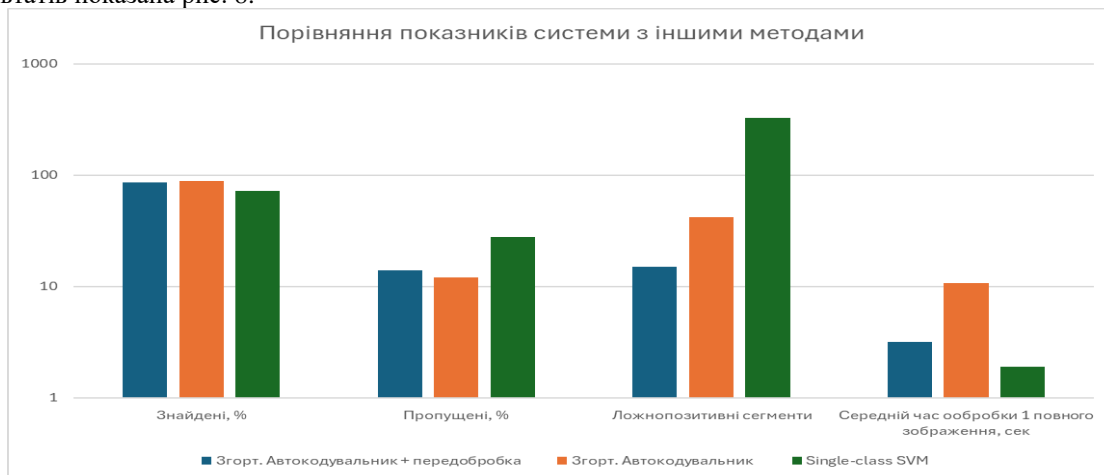


Рис. 8. Порівняння результатів різних підходів (логарифмічна шкала)

## ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШОГО РОЗВИТКУ У ДАНОМУ НАПРЯМІ

Використання в архітектурі НМ-автокодувальника із згортковими шарами забезпечує принципову можливість виявлення аномалій на комплексних, насичених різноманітними об'єктами зображеннях з достатньою для прикладних завдань ефективністю. У порівнянні з іншими методами, не заснованими на використанні згорткових НМ, такий підхід демонструє значно меншу кількість хибних спрацьовувань і високу точність виявлення, що говорить про більш глибоке уявлення системою характеристик досліджуваної місцевості.

Одночасно з цим комбінування НМ-автокодувальника з алгоритмами попереднього аналізу дозволяє значно підвищити швидкодію системи.

Тим не менш, зберігається певна кількість помилкових спрацьовувань, що обумовлено складністю зображень та різноманітністю таких рис, як крони дерев або кам'янистої місцевості. Для вирішення практичних завдань з високою достовірністю доцільним є створення кількох попередньо вивчених моделей та декількох шаблонів налаштувань для алгоритмів попереднього аналізу під кожен тип задачі.

## References

- [1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, *Commun. ACM* 60, 6 (June 2017), p 84–90. doi:10.1145/3065386
- [2] A. Mujeeb, et al, One Class Based Feature Learning Approach for Defect Detection Using Deep Autoencoders, *Advanced Engineering Informatics*, vol. 42, Oct. 2019. doi:10.1016/j.aei.2019.100933
- [3] J. Pereira and M. Silveira, Unsupervised Anomaly Detection in Energy Time Series Data Using Variational Recurrent Autoencoders with Attention, 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 2018, pp. 1275-1282. doi:10.1109/ICMLA.2018.00207
- [4] S. Küçük and S. E. Yüksel, Comparison of RX-based anomaly detectors on synthetic and real hyperspectral data, 2015 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2015, pp. 1-4. doi:10.1109/WHISPERS.2015.8075504
- [5] A. Legrand, H. Trannois and A. Courmier, Use of Uncertainty with Autoencoder Neural Networks for Anomaly Detection, 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), 2019, pp. 32-35. doi:10.1109/AIKE.2019.00014
- [6] C. Baur, et al. Autoencoders for Unsupervised Anomaly Segmentation in Brain MR Images: A Comparative Study, *Medical Image Analysis*, vol. 69, Apr. 2021, doi:10.1016/j.media.2020.101952
- [7] J. Chow, et al. Anomaly detection of defects on concrete structures with the convolutional autoencoder, *Advanced Engineering Informatics*, vol 45, 2020. doi:10.1016/j.aei.2020.101105.
- [8] C. Zhou, R. C. Paffenroth, Anomaly Detection with Robust Deep Autoencoders, *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining August 2017* pp. 665–674. doi:10.1145/3097983.3098052
- [9] L. Beggel, M. Pfeiffer, B. Bischl, Robust Anomaly Detection in Images Using Adversarial Autoencoders, *Machine Learning and Knowledge Discovery in Databases*, pp. 206–222. doi:10.1007/978-3-030-46150-8\_13
- [10] A. Marco, et al, Generative Adversarial Networks for anomaly detection in aerial images, *Computers and Electrical Engineering*, Volume 106, March 2023. doi:10.1016/j.compeleceng.2022.108470.
- [11] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, *Arxiv pre-print*, 2014. doi:10.48550/arXiv.1409.1556
- [12] L. Bottou, Large-Scale Machine Learning with Stochastic Gradient Descent, *Proceedings of COMPSTAT'2010*, 2010. doi:10.1007/978-3-7908-2604-3\_16
- [13] F. Verdoja, M. Grangetto, Graph Laplacian for Image Anomaly Detection, *Machine Vision and Applications*, vol. 31, no. 1, Feb. 2020. doi:10.1007/s00138-020-01059-4
- [14] Suzuki, S. Abe, K., Topological Structural Analysis of Digitized Binary Images by Border Following. *CVGIP* 30 1, pp 32-46 (1985)