

**Галаган Наталія Вікторівна***Державний університет інформаційно-комунікаційних технологій, Київ*  
ORCID 0000-0001-8582-3126**Борисенко Ірина Ігорівна***Київський національний університет імені Тараса Шевченка, Київ*  
ORCID 0009-0001-8645-3881**Яковець Всеволод Петрович***Державний університет інформаційно-комунікаційних технологій, Київ*  
ORCID 0009-0002-3866-8017**Бойко Олег Володимирович***Державний університет інформаційно-комунікаційних технологій, Київ*  
ORCID 0009-0001-3761-7528

### КОНЦЕПТУАЛЬНА МОДЕЛЬ СИСТЕМИ УПРАВЛІННЯ РОЗПІЗНАВАННЯ ОБРАЗІВ ІЗ ЗАСТОСУВАННЯМ ШІ

**Анотація:** Останні інновації в галузі комп'ютерного зору дозволили впровадження складних моделей розпізнавання, здатних обробляти складні образи і робити надійні прогнози на основі величезних обсягів відеоданих. Головними проблемами тренування нейронних мереж в галузі комп'ютерного зору є упередженість даних, що виникає при навчанні систем комп'ютерного зору на упереджених наборах інформації, інтерпретованість даних, що ускладнює інтерпретацію процесів прийняття рішень та потреба в значних обчислювальних ресурсах, особливо в сценаріях застосування в режимі реального часу на пристроях з обмеженими ресурсами. Робота присвячена дослідженню методів підвищення ефективності комп'ютерного зору, зокрема в режимі реального часу та з використанням мінімального обчислювального ресурсу. В статті розглядаються принципи роботи згорткових нейронних мереж та порівнюються характеристики актуальних алгоритмів розпізнавання, таких як R-CNN, R-FCN та YOLO. Головними характеристиками для порівняння були обрані швидкість, точність та використання обчислювальних ресурсів.

На підставі проведеного аналізу для запропонованої моделі системи керування розпізнаванням образів було обрано алгоритм You Only Look Once (YOLO), який вирізняється поєднанням високої швидкості та точності. Це робить його особливо ефективним у завданнях, що потребують миттєвої обробки даних. YOLO виконує обробку зображення за один прохід крізь нейронну мережу, що суттєво пришвидшує процес виявлення об'єктів порівняно з класичними двоетапними методами. У межах реалізації моделі визначено ключові етапи функціонування алгоритму, зокрема використання Non-Maximum Suppression для усунення надлишкових обмежувальних рамок і вибору найточнішої, а також застосування функції втрат, що враховує координати, розміри, рівень впевненості та клас, з метою покращення загальної точності.

**Ключові слова:** штучний інтелект, розпізнавання образів, згорткові нейронні мережі, You Only Look Once

**Halahan Nataliia***State University of Information and Communication Technologies, Kyiv*  
ORCID 0000-0001-8582-3126**Borysenko Iryna***Taras Shevchenko National University of Kyiv, Kyiv*  
ORCID 0009-0001-8645-3881**Yakovets Vsevolod***State University of Information and Communication Technologies, Kyiv*  
ORCID 0009-0002-3866-8017**Boiko Oleh***State University of Information and Communication Technologies, Kyiv*  
ORCID 0009-0001-3761-7528

## CONCEPTUAL MODEL OF AN AI-BASED PATTERN RECOGNITION MANAGEMENT SYSTEM

**Abstract:** Recent innovations in computer vision have enabled the introduction of sophisticated recognition models capable of processing complex images and making reliable predictions based on huge amounts of video data. The main problems of training neural networks in the field of computer vision are data bias, which occurs when training computer vision systems on biased sets of information, data interpretability, which complicates the interpretation of decision-making processes, and the need for significant computing resources, especially in real-time application scenarios on devices with limited resources. This paper is devoted to the study of methods for improving the efficiency of computer vision, in particular in real time and using minimal computing resources. The paper discusses the principles of convolutional neural networks and compares the characteristics of current recognition algorithms such as R-CNN, R-FCN, and YOLO. The main characteristics chosen for comparison are speed, accuracy, and computing resource utilization.

Based on the analysis, the You Only Look Once recognition algorithm was chosen for the presented model of the pattern recognition control system because it combines speed and accuracy, making it suitable for applications that require immediate processing. YOLO performs all calculations simultaneously and allows detecting objects in one pass through the neural network, which significantly increases the speed and significantly differs it from traditional two-stage algorithms. The main stages of the algorithm in the context of the proposed model are defined, including the use of Non-Maximum Suppression to discard unnecessary frames and finally determine the optimal frame, and loss functions to optimize accuracy by taking into account coordinates, size, trust, and classification

**Keywords:** artificial intelligence, pattern recognition, convolutional neural networks, You Only Look Once

### 1. Постановка проблеми

Комп'ютерний зір – це технологія, що дозволяє машинам сприймати, обробляти та інтерпретувати візуальні дані. Останні інновації в галузі комп'ютерного зору призвели до розробки складних моделей, здатних вивчати складні образи і робити надійні прогнози на основі величезних обсягів відеоданих. Однією з найбільш значимих стала поява глибокого навчання та згорткових нейронних мереж (Convolutional Neural Network – CNN), що дозволили впровадження автоматичного навчання та наскрізних системних архітектур.

Робота комп'ютерного зору забезпечується завдяки класифікації зображень та розпізнаванню образів, що є частинами концепції розпізнавання об'єктів, що ідентифікує об'єкти на зображеннях незалежно від місця та умов. Виявлення об'єктів на основі глибокого навчання використовує більш надійні та репрезентативні характеристики та поділяється на два види: одноступеневе та двоступеневе. Одноступеневі методи, такі як YOLO та SSD підходять для сценаріїв застосування в реальному часі, тоді як двоступеневі методи поділяють процес на пропонування регіону, виведення характеристик та класифікацію. Цей процес вимагає більше часу та обчислювальної потужності, тому двоступеневі методи використовуються в сценаріях, де необхідна максимальна точність розпізнавання.

Для досягнення ефективності систем комп'ютерного зору важливо чітко окреслити задачі, які вони мають вирішувати, а також визначити основні вимоги, дотримання яких гарантує надійну роботу системи та високу точність розпізнавання образів. Однією з проблем галузі є упередженість даних, що виникає при навчанні систем комп'ютерного зору на упереджених наборах інформації, оскільки система може успадкувати та часто навіть посилити упередження. Іншою проблемою є інтерпретованість даних, що виникає із-за непрозорого характеру моделей глибокого навчання та ускладнює інтерпретацію процесів прийняття рішень, що обмежує їх застосування в критично важливих сферах [1]. Також залишається серйозною перешкодою потреба в значних обчислювальних ресурсах, особливо в сценаріях застосування на пристроях з обмеженими ресурсами та в режимі реального часу.

### 2. Аналіз останніх досліджень

Розпізнавання об'єктів за допомогою різних алгоритмів досліджується для багатьох сценаріїв застосування. Ця тема набула розповсюдженості в першу чергу завдяки прогресу в згорткових нейронних мережах, методах глибокого навчання та збільшенню потужності паралельних обчислень, пропонованих графічними процесорами. В [2] досліджується застосування алгоритму YOLO для виявлення та розпізнавання зіниць ока в біометричних системах, з зосередженням на точності та високій конфіденційності. Збільшення ефективності розпізнавання досягається за рахунок тренування моделі на основі декількох налаштованих наборів даних. Автори в [3] розглядають можливість покращення ефективності виявлення диму та вогню за допомогою алгоритму You Only Look Once версії 11. Для вирішення проблем з масштабованістю та практичністю моделі було запропоновано інтегрувати YOLO11 з механізмом MSCA (Multi-Scale Convolutional Attention), завдяки чому загальна точність розпізнавання збільшилась на 2.6%, а ефективність розпізнавання диму збільшилась на 2.8%.

В [4] досліджується розпізнавання об'єктів в умовах низької освітленості з застосуванням алгоритму You Only Look Once версії 8, що досяг найбільшого середнього значення точності 0.5513 при порівнянні моделей YOLO версій 3, 5, 6 та 8. В [5] представлено інтегроване рішення для покращення ефективності розпізнавання об'єктів алгоритмом R-CNN (Region-based Convolutional Neural Network – Обласна згорткова нейронна мережа) за допомогою методу розповсюдження довіри в обчисленні потоку LFRFCF між локальними матрицями ознак та двома несуміжними кадрами у послідовності. В [6] представлено метод динамічної генеративної R-CNN (DGR-CNN), що зменшує кількість пропонуваніх регіонів з використанням динамічної програмованої моделі на основі методу подібності графів замість сегментації зображень на основі графів, зменшуючи загальні витрати часу на виявлення та класифікацію, збільшуючи при цьому точність розпізнавання. В [7] представлено систему виявлення для AMR (Autonomous Mobile Robots) на основі CNN та наборів даних з реальними характеристиками транспорту. В [8] автори розглядають використання методів комп'ютерного зору, таких як CNN, для виявлення об'єктів за допомогою тактильних датчиків з високою роздільною здатністю, використовуючи завчасно натреновану CNN на основі набору даних з RGB-зображеннями та спеціально створену CNN (TactNet), натреновану на основі тактильної інформації. В [9] автори представляють метод розумного розпізнавання сторонніх предметів в реальному часі на основі CNN для сценаріїв контролю дорожнього руху, з ефективністю до 92% та 88% в складних середовищах.

### 3. Мета і задачі дослідження

Метою роботи є визначення кращого алгоритму розпізнавання на основі аналізу існуючих методів розпізнавання на основі глибокого навчання для застосування в сценаріях, що потребують ідентифікації об'єктів в режимі реального часу. В роботі аналізується ефективність різних алгоритмів розпізнавання в контексті системи управління. Серед відомих алгоритмів було обрано методи на основі згорткових нейронних мереж, такі як R-CNN, Fast R-CNN з оптимізацією на основі RoI Pooling (Об'єднання регіонів інтересу), Faster R-CNN з інтеграцією RPN (Region Proposal Network – Мережа регіональних припущень), R-FCN (Region-based Fully Convolutional Network – Обласна повністю згорткова мережа) та YOLO (You Only Look Once). Досліджується відповідність кожного алгоритму для застосування в сценаріях реального часу за рахунок швидкості, точності та використання обчислювальних ресурсів.

### 4. Результати дослідження

Згорткова нейронна мережа (CNN) це алгоритм глибокого навчання, основними елементами якого є локальне підключення, розподіл ваг та зменшення вибірки, що дозволяє використання CNN для обробки зображень. Архітектура CNN складається з згорткового шару (Convolution Layer), що містить вхідні дані для фільтрування, об'єднувального шару (Pooling Layer), який зменшує вибірку шляхом об'єднання даних та відкидання непотрібних параметрів, та шару повного об'єднання (Fully connected Layer), що містить кінцевий результат об'єднання.

You Only Look Once (YOLO) – це алгоритм виявлення об'єктів на основі CNN. Модель YOLO стала справжньою інновацією у виявленні об'єктів у реальному часі, сформулювавши її як єдину регресійну задачу. Вона поєднує в собі швидкість і точність, що робить її придатною для застосувань, які потребують негайної обробки [10]. Архітектура YOLO складає 24 згорткових шари, які функціонують для отримання ознак із зображення, за якими слідує 2 об'єднувальні шари, що функціонують для прогнозування ймовірності та координат. В таблиці 1 представлені характеристики алгоритмів розпізнавання об'єктів та їх відповідність для використання в сценаріях реального часу.

На основі проведеного порівняльного аналізу алгоритмів розпізнавання можна дійти висновку, що для задач ідентифікації об'єктів у режимі реального часу оптимально використовувати методи, здатні забезпечити високу точність і швидкодію при мінімальних затримках. Одним із найбільш ефективних рішень у цій сфері є алгоритм YOLO, який здійснює детекцію в один прохід крізь нейронну мережу, що вигідно відрізняє його від традиційних двоетапних підходів.

YOLO належить до одноетапних методів виявлення об'єктів і ґрунтується на використанні нейромережі для одночасного передбачення меж об'єктів та їхніх класів. Принцип його роботи полягає в наступному:

вхідне зображення ділиться на сітку за певним розміром; кожна клітинка сітки формує прогноз для обмежувальних рамок; алгоритм паралельно визначає координати рамок, ймовірність належності до певного класу та коефіцієнт довіри.

Таблиця 1

## Порівняння алгоритмів розпізнавання

| Алгоритм     | Принцип роботи                                                        | Швидкість (FPS) | Точність (mAP) | Використання ресурсів                      | Підходить для реального часу? |
|--------------|-----------------------------------------------------------------------|-----------------|----------------|--------------------------------------------|-------------------------------|
| R-CNN        | Обрізає регіони, обробляє їх CNN та класифікує через SVM              | ~0.5-1 FPS      | Висока         | Дуже високе споживання пам'яті та ресурсів | Ні                            |
| Fast R-CNN   | CNN обробляє все зображення, а RoI Pooling оптимізує регіони          | ~2 FPS          | Висока         | Високе, але ефективніше, ніж у R-CNN       | Ні                            |
| Faster R-CNN | Додає RPN (Region Proposal Network) для швидкого визначення регіонів  | ~5-7 FPS        | Дуже висока    | Високе                                     | Обмежено                      |
| R-FCN        | Використовує повністю згорткову мережу для спрощення обробки регіонів | ~10 FPS         | Висока         | Середнє                                    | Обмежено                      |
| YOLO         | Ділить зображення на сітку та одночасно визначає об'єкти              | ~45-60 FPS      | Висока         | Оптимізоване для GPU                       | Так                           |

Для усунення зайвих рамок застосовується метод Non-Maximum Suppression, що дозволяє отримати остаточний набір виявлених об'єктів.

На відміну від двоетапних методик, YOLO виконує всі обчислення одночасно, що істотно підвищує його швидкість. Далі розглянемо математичні особливості алгоритму.

**Поділ зображення на сітку та визначення регіонів.**

На початковому етапі алгоритм YOLO отримує вхідне зображення й розбиває його на сітку розміром  $S \times S$ . Кожна клітинка цієї сітки відповідає за перевірку, чи розташований центр об'єкта саме в її межах. Формально це можна описати за допомогою індикаторної функції:

$$I_{i,j}^{obj} = \begin{cases} 1, & \text{якщо центр об'єкта знаходиться у клітинці}(i,j) \\ 0, & \text{в іншому випадку} \end{cases} \quad (1)$$

де  $i,j$  – положення клітинки сітки у координатному просторі  $S \times S$ ;  $I_{i,j}^{obj}$  – двійковий параметр, що відображає факт розташування центру об'єкта в межах клітинки сітки.

**Визначення меж об'єкта у кожній комірці.** Кожна клітинка прогнозує  $B$  меж об'єкта, кожна з яких представлена п'ятьма параметрами:  $(x,y,w,h,C)$

де:

- $B$  – означає кількість передбачених обмежувальних рамок для кожної комірки  $S \times S$ .
- $(x,y)$  – координати центру об'єкта у межах цієї клітинки.
- $w,h$  – ширина та висота передбаченої рамки у відносних координатах.
- $C$  – рівень довіри (confidence score), який обчислюється як добуток ймовірності об'єкта  $P(Object)$  та  $IoU$  (Intersection over Union) між передбаченим і реальним полями:

$$C = P(Object) \times IoU \text{ (передбачена рамка, реальна рамка)} \quad (2)$$

де  $P(Object)$  – ймовірність наявності об'єкта у клітинці;  $IoU$  – метрика, яка оцінює, наскільки добре прогнозована рамка збігається з реальною.

**Визначення класу об'єкта.** Кожна клітинка також передбачає ймовірності для кожного класу  $C_i: P(Class_i|Object)$ . Загальний вектор ймовірностей класів для кожної клітинки представлений як:

$$\vec{P} = (P(Class_1|Object), P(Class_2|Object), \dots, P(Class_C|Object)) \quad (3)$$

Остаточна ймовірність для кожного класу обчислюється як:

$$P(Class_i) = P(Class_i|Object) \times P(Object) \quad (4)$$

**Остаточне визначення найкращої рамки.** Кожна клітинка сітки відповідає за  $B$  кандидатів обмежувальних рамок. Для відбору найоптимальнішої рамки застосовується метод Non-Maximum Suppression (NMS): спочатку обирається рамка з максимальним значенням коефіцієнта довіри, після чого всі інші рамки з високим рівнем перекриття видаляються, які мають високий перекриття ( $IoU$  вище певного порогу, наприклад, 0.5); повторюємо, поки не залишаться лише оптимальні bounding boxes.

**Остаточний вихід YOLO.** Алгоритм повертає тензор  $S \times S \times (B \times C)$ , що містить: Обмежувальні рамки –  $(x, y, w, h, C)$ ; Ймовірність класів –  $P(Class_i)$ .

Загальна функція виходу YOLO виглядає як:

$$y = \{(x, y, w, h, C, P(Class_1), \dots, P(Class_c))\} S \times S \quad (5)$$

**Повна функція втрат YOLO.** YOLO мінімізує функцію втрат, що складається з чотирьох компонентів:

1. Втрати координат рамки (Bounding Box Loss):

$$L_{coord} = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \quad (6)$$

2. Втрати розміру рамки (Width & Height Loss):

$$L_{size} = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} [(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2] \quad (7)$$

3. Втрати довіри (Confidence Loss):

$$L_{conf} = \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} (C_i - \hat{C}_i)^2 + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{noobj} (C_i - \hat{C}_i)^2 \quad (8)$$

4. Втрати класифікації (Classification Loss):

$$L_{class} = \sum_{i=0}^{S^2} I_i^{obj} \sum_{c \in classes} (p_i(c) - \hat{p}_i(c))^2 \quad (9)$$

Фінальна функція втрат:

$$L_{YOLO} = L_{coord} + L_{size} + L_{conf} + L_{class} \quad (10)$$

де:  $I_{ij}^{obj}$  – індикатор того, що об'єкт присутній у клітинці  $(i, j)$ ,  $\lambda_{coord}$  – коефіцієнт ваги для координат,  $\lambda_{noobj}$  – штраф за передбачення фонового об'єкта.

## 5. Висновки

На основі проведеного в статті аналізу актуальних методів розпізнавання для моделі системи управління розпізнавання образів було обрано алгоритм You Only Look Once. YOLO є високошвидкісним алгоритмом, що виконує обробку зображення за один прохід, поділяючи його на сітку, де кожна комірка відповідає за визначення координат, ймовірності наявності об'єкта та його класу. Для усунення зайвих рамок використовується метод Non-Maximum Suppression, а функція втрат дозволяє підвищити точність за рахунок врахування просторових координат, розмірів об'єкта, рівня довіри та результатів класифікації. Завдяки цьому YOLO є оптимальним рішенням для задач автоматичного розпізнавання образів у режимі реального часу з мінімальними обчислювальними витратами.

## Список використаної літератури

1. Druzhynin V., Gladka M., Hladkyi Y., Borysenko I. A model for allocating labor resources to project work based on task prioritization. *Information Technology and Implementation (IT&I-2024)*. Kyiv, 2024. P. 42–54.
2. Parthasarathy C, Rachel G P., Sherin J R. Real-Time Iris Detection and Recognition System Using You Only Look Once Version 8. *International Journal of Inventive Engineering and Sciences*. 2025. Vol. 12, no. 2. P. 13–17. URL: <https://doi.org/10.35940/ijies.11096.12020225>.
3. Improving Fire and Smoke Detection with You Only Look Once 11 and Multi-Scale Convolutional Attention / Y. Li et al. *Fire*. 2025. Vol. 8, no. 5. P. 165. URL: <https://doi.org/10.3390/fire8050165>.
4. Advancing low-light object detection with you only look once models: An empirical study and performance evaluation / S. U. A. Shovo et al. *Cognitive Computation and Systems*. 2024. URL: <https://doi.org/10.1049/ccs2.12114>.
5. Local Fast R-CNN Flow for Object-Centric Event Recognition in Complex Traffic Scenes / Q. Gu et al. *Image and Video Technology*. Cham, 2018. P. 439–452. URL: [https://doi.org/10.1007/978-3-319-92753-4\\_34](https://doi.org/10.1007/978-3-319-92753-4_34).
6. Dynamic generative R-CNN / R. Saffarini et al. *Neural Computing and Applications*. 2025. URL: <https://doi.org/10.1007/s00521-024-10739-5>.
7. A new CNN-BASED object detection system for autonomous mobile robots based on real-world vehicle datasets / U. Aulia et al. *Heliyon*. 2024. Vol. 10, no. 15. P. e35247. URL: <https://doi.org/10.1016/j.heliyon.2024.e35247>.
8. Gandarias J. M., Garcia-Cerezo A. J., Gomez-de-Gabriel J. M. CNN-Based Methods for Object Recognition With High-Resolution Tactile Sensors. *IEEE Sensors Journal*. 2019. Vol. 19, no. 16. P. 6872–6882. URL: <https://doi.org/10.1109/jsen.2019.2912968>.
9. Research on Intelligent Recognition System of Traffic Image Based on CNN and Intelligent Recognition of Foreign Object Intrusion / J. Tan et al. *Applied Mathematics and Nonlinear Sciences*. 2025. Vol. 10, no. 1. URL: <https://doi.org/10.2478/amns-2025-0400>.
10. Bhalani V. Advancements in Computer Vision: Techniques, Applications, and Future Trends. *Journal of Information and Computational Science*. 2024. URL: <https://doi.org/10.12733/JICS.2024.V14I12.535569.110001>.
11. Скіцько, О., Складанний, П., Ширшов, Р., Гуменюк, М., & Ворохоб, М. (2023). Загрози та ризики використання штучного інтелекту. *Кібербезпека: освіта, наука, техніка*, 2(22), 6–18. <https://doi.org/10.28925/2663-4023.2023.22.618>
12. Zhebka, V., Skladannyi, P., Zhebka, S., Shlianchak, S., Bondarchuk, A. Methodology for Predicting Failures in a Smart Home based on Machine Learning Methods / *Ceur Workshop Proceedings* Open source preview, 2024, 3654, p. 322–332
13. Zhebka, V., Gertsyuk, M., Sokolov, V., Malinov, V., Sablina, M. Optimization of Machine Learning Method to Improve the Management Efficiency of Heterogeneous Telecommunication Network / *Ceur Workshop Proceedings* Open source preview, 2022, 3288, p. 149–155