

DOI: <https://doi.org/10.15276/ict.02.2025.22>

УДК 004'25

Аналіз методів та технологій інтелектуального відеоспостереження

Арсирій Олена Олександрівна¹⁾

Д-р техніч. наук., професор, завідувачка каф. Інформаційних систем
ORCID: <https://orcid.org/0000-0001-8130-9613>; e.arsiriy@gmail.com. Scopus Author ID: 54419480900

Голованчук Микола Юрійович¹⁾

Магістр каф. Інформаційних систем
ORCID: <https://orcid.org/0009-0005-2542-1930>; k.holovanchuk@gmail.com.

¹⁾ Національний університет «Одеська політехніка», пр. Шевченка, 1. Одеса, 65044, Україна

АНОТАЦІЯ

Інтелектуальні системи відеоспостереження є одним із ключових напрямів розвитку сучасних інформаційних технологій, оскільки вони забезпечують автоматичний аналіз відеопотоків, виявлення об'єктів і подій у реальному часі. Завдяки розвитку методів глибинного навчання та комп'ютерного зору з'явилася можливість створення високоточних систем, здатних працювати в умовах змінного освітлення, шумів і складного фону. Водночас залишається актуальним завдання підвищення ефективності таких систем за рахунок вибору оптимальних архітектур нейронних мереж і технологій апаратного прискорення, що дозволяють зменшити затримку обробки кадрів і споживання енергії на вбудованих пристроях (Edge AI). В роботі проведено аналіз сучасних методів і технологій створення інтелектуальних систем відеоспостереження, визначено найбільш ефективні архітектури глибинних нейронних мереж для задач детекції й трекінгу об'єктів, а також проведено оцінку можливостей їх оптимізації для роботи у реальному часі. Розглянуто архітектури згорткових нейронних мереж (CNN) та трансформерні моделі (ViT, DETR), а також гібридні підходи, що поєднують просторовий і часовий аналіз відео. Проведено порівняння фреймворків YOLOv5/YOLOv8, OpenVINO, TensorRT і RKNN Toolkit, які забезпечують апаратне прискорення на платформах GPU та NPU. Додатково проаналізовано ефективність алгоритмів трекінгу DeepSORT і ByteTrack, що забезпечують стійке відстеження об'єктів у потоковому відео. Результати дослідження показали, що поєднання компактних CNN-моделей з апаратно оптимізованими бібліотеками дозволяє зменшити затримку обробки кадру до 30 мс при збереженні точності детекції понад 90 %. Отримані висновки підтверджують доцільність використання гібридних архітектур і технологій апаратного прискорення для створення ефективних систем інтелектуального відеоспостереження нового покоління.

Ключові слова: інтелектуальне відеоспостереження; комп'ютерний зір; глибинне навчання; YOLOv8; Edge AI; детекція об'єктів; трекінг; OpenVINO; TensorRT; RKNN

Актуальність. Сучасний етап розвитку інформаційних технологій характеризується стрімким зростанням обсягів відеоданих, що потребують автоматизованого аналізу для виявлення об'єктів, подій та аномальної поведінки. Традиційні системи відеоспостереження, засновані лише на детекції руху або порогових алгоритмах, не здатні забезпечити належний рівень точності та швидкодії в умовах змінного освітлення, шумів або динамічного середовища.

Завдяки розвитку глибинного навчання (Deep Learning) і алгоритмів комп'ютерного зору стало можливим створення інтелектуальних систем відеоаналізу, що автоматично ідентифікують об'єкти, класифікують сцени та оцінюють поведінкові патерни у режимі реального часу. Однак, попри значний прогрес, актуальною залишається проблема забезпечення високої продуктивності при роботі на обмежених апаратних ресурсах.

Особливої актуальності набуває дослідження методів оптимізації моделей штучного інтелекту для реалізації на вбудованих пристроях (Edge AI), де критично важливими є мінімальні затримки, низьке енергоспоживання та стабільність роботи в автономному режимі.

Отже, дослідження й удосконалення методів інтелектуального відеоспостереження є актуальним напрямом, що сприяє підвищенню ефективності систем безпеки, аналітики відеоданих і автоматизованого моніторингу об'єктів.

Метою дослідження є проведення системного аналізу сучасних методів і технологій інтелектуального відеоспостереження, визначення ефективних архітектур глибинних

нейронних мереж для детекції та трекінгу об'єктів у режимі реального часу, а також оцінка можливостей їх оптимізації для роботи на вбудованих пристроях із обмеженими обчислювальними ресурсами. Дослідження спрямоване на узагальнення підходів до побудови автономних систем спостереження, що поєднують методи комп'ютерного зору, машинного навчання та апаратного прискорення на основі GPU і NPU, із метою підвищення точності, стабільності та енергоефективності таких систем.

Сучасні інтелектуальні системи відеоспостереження базуються на використанні алгоритмів комп'ютерного зору, що реалізують багаторівневу обробку зображення — від попередньої фільтрації до семантичної сегментації сцени. Основу таких систем становлять згорткові нейронні мережі (Convolutional Neural Networks, CNN), які забезпечують автоматичне виділення ознак об'єктів [1, 4]. Найпоширенішими архітектурами є серії моделей YOLO (You Only Look Once) [1–3], EfficientDet [4], SSD та Faster R-CNN [5].

Архітектури сімейства YOLO (версії v5, v7, v8) поєднують високу швидкість та точність завдяки оптимізованим механізмам розподілу обчислень і використанню фокусних втрат (Focal Loss) [2, 3]. У той час як Faster R-CNN демонструє вищу точність на статичних даних [5], моделі YOLOv8 здатні працювати в реальному часі навіть на вбудованих обчислювальних платформах, що робить їх більш придатними для систем безпilotного моніторингу [3, 10].

Новим етапом розвитку стали моделі на основі візуальних трансформерів (Vision Transformers, ViT) [6] та DETR (Detection Transformer) [7], що застосовують механізм самоуваги (Self-Attention) для побудови просторових залежностей між об'єктами сцени. Такі підходи забезпечують більш глибоке розуміння контексту, однак потребують значних обчислювальних ресурсів, тому їх практичне застосування в системах Edge AI потребує додаткової оптимізації [13, 15].

З метою скорочення затримки обробки кадрів активно використовуються технології апаратного прискорення, зокрема OpenVINO™ (Intel) [9], TensorRT (NVIDIA) [8] та RKNN Toolkit (Rockchip) [10]. Ці фреймворки дозволяють виконувати оптимізацію моделі за рахунок квантування, прунінгу та компіляції під специфічну архітектуру процесора або нейронного співпроцесора. На практиці це зменшує час обробки відеокадру в кілька разів без суттєвого зниження точності.

Важливою складовою інтелектуального відеоспостереження є трекінг об'єктів — відстеження їхніх координат у послідовності кадрів. Серед найефективніших алгоритмів можна виокремити DeepSORT [11], який поєднує детектор YOLO з рекурентною мережею для ідентифікації об'єкта, та ByteTrack [12], що оптимізує асоціацію детекцій з урахуванням пропущених кадрів. Обидва підходи забезпечують стійкість відстеження навіть у разі короточасних перекриттів або зміни масштабу.

Одним із ключових викликів є баланс між точністю та швидкістю. Глибокі моделі, як-от Faster R-CNN або Swin Transformer, досягають високої точності на статичних зображеннях, проте їхня складність не дозволяє застосовувати їх у реальному часі на вбудованих пристроях. У свою чергу, легкі моделі (YOLOv8n, MobileNetV3, EfficientDet-D0) орієнтовані на швидкість, що робить їх оптимальними для використання у портативних або автономних системах.

Важливою тенденцією є перехід від серверної обробки до периферійних обчислень (Edge AI). Завдяки цьому знижується затримка передачі даних, забезпечується приватність інформації, а системи стають незалежними від стабільності мережевого з'єднання. Для реалізації таких рішень застосовуються платформи Rockchip RK3588, NVIDIA Jetson, Google Coral, які мають інтегровані прискорювачі нейронних обчислень (NPU, TPU).

Крім того, для підвищення ефективності інтелектуальних систем відеоспостереження активно застосовуються методи компресії та оптимізації нейронних мереж, серед яких найпоширенішими є квантування (quantization), прунінг (pruning) та знання-дистиляція (knowledge distillation) [13]. Застосування цих підходів дозволяє зменшити розмір моделі, скоротити кількість обчислень і, відповідно, підвищити швидкість без істотної втрати

точності. У поєднанні з апаратним прискоренням на GPU, NPU чи TPU такі методи забезпечують реальну можливість виконання глибоких моделей безпосередньо на периферійних пристроях.

Окрему увагу дослідники приділяють енергоефективності нейронних моделей, оскільки тривала робота систем спостереження в автономному режимі потребує мінімального енергоспоживання. Для цього розробляються спеціальні архітектури, зокрема MobileNetV4, ShuffleNetV2 та EfficientNet-Lite, орієнтовані на високу продуктивність при обмежених ресурсах [14].

Для узагальнення сучасних підходів у сфері інтелектуального відеоспостереження доцільно провести порівняльний аналіз найпоширеніших моделей і технологій, які використовуються для автоматизованої детекції та відстеження об'єктів у відеопотоці. Такий аналіз дозволяє оцінити ефективність різних архітектур з позицій точності виявлення об'єктів, швидкодії, стійкості до змін умов зйомки, ресурсомісткості алгоритмів та можливості їх адаптації до обмежених апаратних ресурсів вбудованих систем. Оскільки більшість практичних застосувань передбачає роботу в режимі реального часу, особливо важливим критерієм є не лише якість розпізнавання, а й здатність моделі стабільно працювати при мінімальній затримці обробки кадрів. Порівняння також враховує підтримку технологій оптимізації, таких як квантування, прунінг і компіляція під архітектуру GPU або NPU, що є ключовими для впровадження інтелектуальних систем відеоспостереження у форматі Edge AI. На рис. 1 наведена загальна архітектура моделі YOLOv8, яка складається з трьох модулів Backbone, Neck та Head.

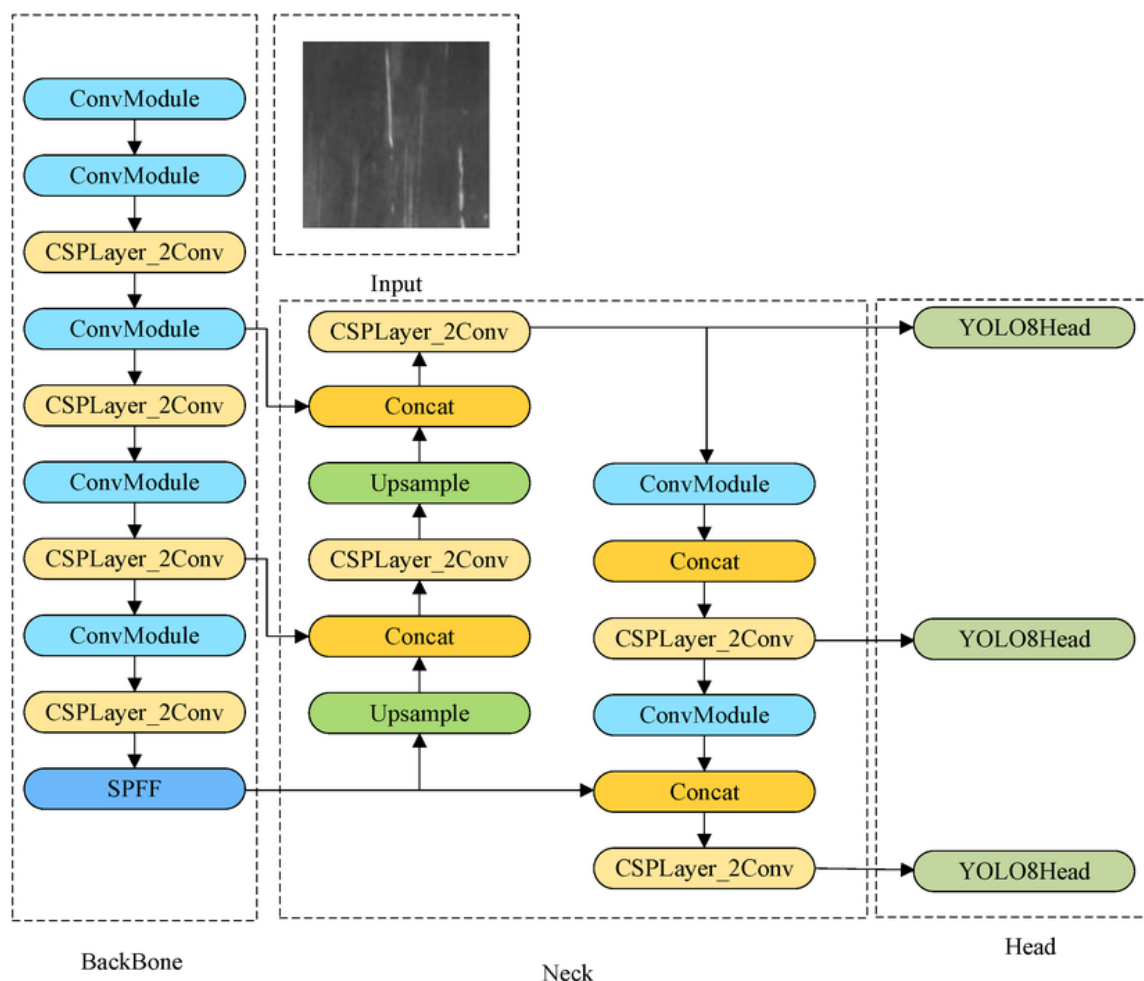


Рис. 1. Взаємозв'язок трьох модулів Backbone, Neck та Head в загальній архітектурі моделі YOLOv8

Результати порівняльного аналізу методів та технологій інтелектуального відеоспостереження показано в таблиці 1 та зазначно, що YOLOv8n є найбільш збалансованою моделлю для організації відеоспостереження в реальному часі, оскільки забезпечує високу точність при низькому енергоспоживанні й легко адаптується під вбудовані платформи. Моделі Faster R-CNN і ViT/DETR демонструють кращі результати при аналізі статичних зображень, проте потребують значно більших обчислювальних ресурсів і не є оптимальними для Edge AI.

Таблиця 1. Порівняльний аналіз технологій

Технологія	Архітектура	Точність(mAP)	Продуктивність (FPS)	Обчислювальна складність	Edge-оптимізація
YOLOv5	CNN	~87–89	40–60	Середня	Так
YOLOv8n	CNN	~90–92	70–120	Низька	Так
EfficientDet-D0	CNN	~88	45–55	Середня	Частково
Faster R-CNN	Two-stage CNN	~92–94	10–20	Висока	Ні
Vision Transformer (ViT)	Transformer	~93–95	8–15	Дуже висока	Ні
DETR	CNN + Transformer	~91–93	10–25	Висока	Частково

У сучасних моделях комп'ютерного зору для відеоспостереження активно використовується механізм уваги (Attention Mechanism), який дає змогу моделі фокусуватися на найбільш інформативних ділянках зображення або відеокадру. Такий підхід значно підвищує точність детекції об'єктів, особливо у складних сценах із великою кількістю об'єктів, частковими перекриттями або тінями. Механізм уваги дає можливість мережі динамічно розподіляти ресурси обчислень, концентруючи увагу на областях, які містять ознаки цільових об'єктів, ігноруючи несуттєві деталі фону.

Зокрема, у моделях на основі Vision Transformer (ViT) увага реалізується через багатоголовий механізм самоуваги (Multi-Head Self-Attention), який дозволяє одночасно аналізувати взаємозв'язки між різними частинами зображення. Це дає змогу враховувати як локальні, так і глобальні контексти, що особливо важливо для виявлення об'єктів у динамічних сценах або при зміні перспективи камери. Модифікації таких моделей, як Swin Transformer або DETR, демонструють покращену узагальнювальну здатність завдяки ієрархічній структурі уваги та комбінуванню CNN-шарів для попереднього вилучення ознак.

Для моделей, що працюють у реальному часі, зокрема серії YOLOv8, важливим є поєднання згорткових шарів і легкого механізму фокусування (Spatial Attention Block), який допомагає моделі виділяти найбільш значущі області у просторі кадру. Це дозволяє підвищити точність без збільшення кількості параметрів, що є критичним для виконання на вбудованих системах.

На етапі попередньої обробки даних важливу роль відіграють методи нормалізації освітлення, фільтрації шумів та масштабування. Їхнє правильне застосування забезпечує стабільність роботи детекторів у різних умовах навколишнього середовища – наприклад, при переході з денного на нічний режим спостереження.

Для забезпечення високої швидкодії застосовуються методи квантування (quantization) та прунінгу (pruning). Квантування знижує точність представлення ваг (наприклад, з 32-бітних до 8-бітних значень), що суттєво зменшує обсяг пам'яті та прискорює виконання без значних втрат точності. Прунінг, своєю чергою, видаляє малозначущі з'єднання між нейронами, оптимізуючи структуру мережі. У сукупності ці підходи дають змогу розгортати складні нейронні архітектури навіть на пристроях із низькою обчислювальною потужністю.

Особливу увагу дослідники приділяють поєднанню CNN і Transformer-архітектур. Такі гібридні моделі, як ConVNext або YOLOv10, використовують CNN-шари для вилучення просторових ознак і механізми самоуваги для моделювання довготривалих залежностей у послідовності кадрів. Це забезпечує кращу узгодженість між об'єктами у часі та дозволяє ефективніше реалізовувати трекінг.

У задачах відстеження об'єктів (object tracking) сучасні системи комбінують детекцію + асоціацію. Наприклад, алгоритм DeepSORT використовує CNN для вилучення ознак об'єкта та метод косинусної схожості для їх порівняння між кадрами. Алгоритм ByteTrack, своєю чергою, оптимізує асоціацію об'єктів, зберігаючи навіть ті детекції, які мають низьку впевненість, що підвищує стійкість системи при короткочасній втраті об'єкта.

Порівняльний аналіз демонструє, що методи на основі CNN із механізмом уваги досягають середньої точності mAP понад 90 %, тоді як класичні підходи на базі HOG або SIFT показують лише 65–70 %. Крім того, завдяки апаратному прискоренню (OpenVINO, TensorRT, RKNN Toolkit) середня затримка обробки кадру знижується з 60–80 мс до 25–30 мс, що дозволяє забезпечити роботу системи в реальному часі навіть на енергоефективних платформах.

Таким чином, поєднання легких CNN-архітектур, механізмів уваги та апаратної оптимізації є найбільш ефективним підходом до побудови сучасних інтелектуальних систем відеоспостереження. Це забезпечує високу точність, швидкодію й стабільність навіть за умов обмежених обчислювальних ресурсів.

СПИСОК ЛІТЕРАТУРИ

1. Redmon J., Farhadi A. “YOLOv3: An Incremental Improvement.” *arXiv*. 2018. DOI: <https://doi.org/10.48550/arXiv.1804.02767>.
2. Jocher G., Stoken A., Borovec J., Chaurasia A., Qiu J., Ultralytics Team. “YOLOv5 – Object Detection on the Edge”. *GitHub Repository*. 2022.
3. Ultralytics. “YOLOv8 Documentation”. *Ultralytics AI Ltd*. 2023.
4. Tan M., Le Q. V. “EfficientDet: Scalable and Efficient Object Detection”. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle. 2020. p. 10778–10787. DOI: <https://doi.org/10.1109/CVPR42600.2020.01079>.
5. Ren S., He K., Girshick R., Sun J. “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017; 39 (6): 1137–1149.
6. Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., et al. “An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale.” *arXiv*. 2020. DOI: <https://doi.org/10.48550/arXiv.2010.11929>.
7. Carion N., Massa F., Synnaeve G., Usunier N., Kirillov A., Zagoruyko S. “End-to-End Object Detection with Transformers (DETR).” *European Conference on Computer Vision (ECCV)*. Glasgow. 2020. p. 213–229.
8. “TensorRT Developer Guide. Version 8.6”. *NVIDIA. Corporation Santa Clara, CA*. 2023.
9. “OpenVINO™ Toolkit Documentation. Release 2023.0”. *Intel Corporation. Santa Clara, CA*. 2023.
10. “RKNN Toolkit User Guide”. *Rockchip Electronics Co., Ltd*. 2023.
11. Wojke N., Bewley A. “Simple Online and Realtime Tracking with a Deep Association Metric”. *arXiv*. 2017. DOI: <https://doi.org/10.48550/arXiv.1703.07402>.
12. Zhang Y., Sun P., Jiang Y., Yu F., Wang X. “ByteTrack: Multi-Object Tracking by Associating Every Detection Box”. *arXiv*. 2021. DOI: <https://doi.org/10.48550/arXiv.2110.06864>.
13. Tao Y., Chen M., Wang Y., Duan Q. “Edge AI: A Survey on Edge Computing Technologies for Artificial Intelligence”. *IEEE Access*. 2020; 8: 175204–175219. DOI: <https://doi.org/10.1109/ACCESS.2020.3027920>.

14. Huang Z., Wang J. “Survey on Deep Learning for Video Surveillance”. *IEEE Access*. 2022; 10: 21547–21569.

15. Chen L., Li H., Yang Y., Zhou Z. “Hybrid CNN-Transformer Architectures for Real-Time Video Analysis”. *Pattern Recognition Letters*. 2023; 169: 94–104.

DOI: <https://doi.org/10.15276/ict.02.2025.22>

UDC 004’25

Analysis of methods and technologies of intelligent video surveillance

Olena O. Arsirii¹⁾

Doctor of Engineering Sciences, Professor, Head of the Department of Information Systems
ORCID: <https://orcid.org/0000-0001-8130-9613>; e.arsirii@gmail.com. Scopus Author ID: 54419480900

Mykola Y. Holovanchuk¹⁾

Master of the Department of Information Systems
ORCID: <https://orcid.org/0009-0005-2542-1930>; k.holovanchuk@gmail.com
¹⁾ Odesa Polytechnic National University, 1, Shevchenko Ave. Odesa, 65044, Ukraine

ABSTRACT

Intelligent video surveillance systems are one of the key areas of development of modern information technologies, as they provide automatic analysis of video streams, detection of objects and events in real time. Thanks to the development of deep learning methods and computer vision, it has become possible to create high-precision systems capable of operating in conditions of variable lighting, noise and complex background. At the same time, the task of increasing the efficiency of such systems by selecting optimal neural network architectures and hardware acceleration technologies that reduce frame processing latency and energy consumption on embedded devices (Edge AI) remains relevant. The purpose of the work is to analyze modern methods and technologies for creating intelligent video surveillance systems, determine the most effective architectures of deep neural networks for object detection and tracking tasks, and evaluate the possibilities of their optimization for real-time operation. The study considers convolutional neural network architectures (CNN) and transformer models (ViT, DETR), as well as hybrid approaches that combine spatial and temporal video analysis. A comparison of the YOLOv5/YOLOv8, OpenVINO, TensorRT, and RKNN Toolkit frameworks, which provide hardware acceleration on GPU and NPU platforms, is carried out. Additionally, the effectiveness of the DeepSORT and ByteTrack tracking algorithms, which provide stable tracking of objects in streaming video, was analyzed. The results of the study showed that the combination of compact CNN models with hardware-optimized libraries allows reducing the frame processing delay to 30 ms while maintaining detection accuracy of over 90%. The obtained conclusions confirm the feasibility of using hybrid architectures and hardware acceleration technologies to create effective intelligent video surveillance systems of the new generation

Keywords: intelligent video surveillance; computer vision; deep learning; YOLOv8; Edge AI; object detection; tracking; OpenVINO; TensorRT; RKNN