
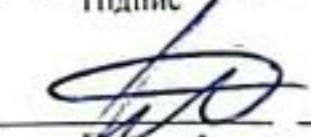



КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА

на тему Метод визначення кількості людей у натовпі у відеопотоці
нейромережевими засобами

Галузь знань 12 – Інформаційні технології
Шифр і назва галузі знань
Спеціальність 122 – Комп'ютерні науки
Шифр і назва спеціальності
Освітня програма Комп'ютерні науки
Назва освітньої програми

Виконав: студент 2 курсу, група КНм-23-2  Юрій Голуб
Курс, група виконавця Підпис ім'я, прізвище
Керівник: к.т.н., доцент кафедри КН  Руслан Багрій
Науковий ступінь, посада Підпис ім'я, прізвище
Нормоконтроль: к.т.н., доцент кафедри КН  Руслан Багрій
Науковий ступінь, посада Підпис ім'я, прізвище

До захисту допускаю:
Зав. кафедри КН, д.т.н., професор  Олександр Бармак
Підпис ім'я, прізвище

19 12 2024 р.

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет інформаційних технологій
Кафедра комп'ютерних наук
Освітній ступінь магістр
Галузь знань 12 – Інформаційні технології
Спеціальність 122 – Комп'ютерні науки

ЗАТВЕРДЖУЮ

Завідувач кафедри комп'ютерних наук

(підпис)

д.т.н. професор Олександр Бармак

« 02 » вересня 2024 року

ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ МАГІСТРА

1.Тема кваліфікаційної роботи магістра: Метод визначення кількості людей у натовпі у відеопотоці нейромережевими засобами

2.Завдання видано студенту Юрію Голубу
(ім'я, прізвище)

3.Керівник роботи доцент кафедри КН Руслан Багрій
(ім'я, прізвище)

4.Затверджені наказом університету від « 26 » серпня 2024 р. № 60

5.Зміст пояснювальної записки (перелік задач) та вихідні дані: Метою роботи є підвищення точності та ефективності визначення кількості людей у натовпі в реальному часі у відеопотоці нейромережевими засобами. Для досягнення мети визначено такі задачі: провести аналіз нейромережових моделей та існуючих підходів для визначення кількості людей у натовпі у відеопотоці; розробити метод визначення кількості людей у натовпі у відеопотоці з використанням попередньо натренованої нейромережової моделі; підготувати набір даних, що містить різноманітні зображення натовпу для донавчання моделі; провести валідацію запропонованого методу визначення кількості людей у натовпі за відомими статистичними показниками та порівняти результати з відомими підходами.

Реферат

У сучасному світі визначення кількості людей у натовпі набуває важливого значення в різних сферах життя. Цей процес є ключовим для забезпечення громадської безпеки, організації масових заходів, моніторингу транспортних вузлів, а також для оцінки потоків людей у торговельних центрах, стадіонах та інших місцях великого скупчення. Точний і швидкий аналіз натовпу дозволяє запобігати можливим небезпекам, таким як тисняви або неконтрольовані ситуації, а також сприяє оптимізації управлінських рішень у реальному часі.

Зі зростанням обсягів відеоінформації та збільшенням кількості камер спостереження виникає необхідність автоматизації процесу обробки даних.

Традиційні методи, засновані на ручному аналізі або класичних алгоритмах комп'ютерного зору, стають менш ефективними. У цьому контексті використання нейромережових засобів, зокрема сучасних моделей глибокого навчання, відкриває нові можливості для розв'язання такого типу завдань.

Ця кваліфікаційна робота магістра спрямована на розробку та вдосконалення методу автоматичного визначення кількості людей у натовпі на основі сучасних підходів до обробки відео та нейромережових технологій. Результатом роботи є метод визначення кількості людей у натовпі за допомогою спеціально навченої нейромережевої моделі.

Актуальність теми. Сучасні тенденції урбанізації та зростання кількості населення в містах призводять до збільшення концентрації людей у громадських місцях, таких як транспортні вузли, площі, спортивні та культурні заходи. Ефективний моніторинг цих локацій є критично важливим для забезпечення безпеки, управління потоками людей і зменшення ризику інцидентів, таких як тисняви чи паніка.

Існуючі підходи до визначення кількості людей у натовпі, що базуються на традиційних методах аналізу відеопотоків, часто виявляються недостатньо точними або вимагають значних ресурсів для їх реалізації. Зі збільшенням обсягів відеоінформації від камер спостереження необхідно впроваджувати автоматизовані

в рішення, які можуть працювати в реальному часі та адаптуватися до змінних умов, таких як освітлення, щільність натовпу або зміна ракурсів зйомки.

У цьому контексті методи на основі нейронних мереж відкривають нові можливості для аналізу складних даних. Зокрема, моделі глибокого навчання, які здатні виконувати завдання розпізнавання об'єктів із достатньо високою точністю та швидкістю. Їх застосування для підрахунку людей у натовпі може значно підвищити ефективність систем відеоспостереження, знизити людський фактор і забезпечити адаптивність до різних сценаріїв.

Мета і задачі роботи. Метою роботи є підвищення точності та ефективності визначення кількості людей у натовпі в реальному часі у відеопотоці нейромережевими засобами. Для досягнення поставленої мети потрібно виконати наступні завдання:

- провести аналіз нейромережових моделей та існуючих підходів для визначення кількості людей у натовпі у відеопотоці;
- розробити метод визначення кількості людей у натовпі у відеопотоці з використанням нейромережової моделі;
- підготувати набір даних, що містить різноманітні зображення натовпу для донавчання моделі;
- провести валідацію запропонованого методу за відомими статистичними показниками та порівняти результати з відомими підходами.

Об'єкт дослідження □ процес визначення кількості людей у натовпі на основі відеопотоку з використанням нейромережових моделей для обробки зображень.

Предмет дослідження – моделі нейронної мережі для підрахунку людей у натовпі.

Методи дослідження. Для виконання завдання було застосовано згорткову нейронну мережу для розпізнавання людей у відеокадрі, спеціалізований датасет для підвищення точності та алгоритми трекінгу об'єктів, для збереження стабільності та підвищення ефективності у динамічних сценах.

Наукова новизна одержаних результатів. В результаті проведеної роботи удосконалено метод визначення кількості людей у натовпі за рахунок використання сучасних архітектур глибоких нейронних мереж, оптимізованих для обробки відеопотоків, який відрізняється від існуючих донавчанням моделі на спеціалізованих наборах даних та використанням алгоритмів трекінгу об'єктів у кадрі для покращення результатів у динамічних сценах, що дало можливість підвищити точність та ефективність визначення кількості людей у натовпі в реальному часі у відеопотоці.

Практичне значення отриманих результатів полягає в створенні ефективного інструменту для автоматичного підрахунку людей у натовпі, що може бути інтегрований у системи відеоспостереження, що використовуються для забезпечення безпеки на масових заходах, в транспортних вузлах, та в інших громадських місцях.

Розроблений метод дозволяє значно знизити людський фактор, забезпечуючи більш точний та оперативний аналіз в реальному часі, що особливо важливо для запобігання небезпечних ситуацій.

Апробація результатів кваліфікаційної роботи магістра та публікації. Основні положення роботи опубліковані в збірнику наукових праць – Визначення кількості людей в натовпі у відеопотоці нейромережевими засобами / Голуб Ю.В., Багрій Р.О., Вознюк Л.О. // Збірник наукових праць за матеріалами XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН□2024». Хмельницький. 2024.

Структура і обсяг роботи. Кваліфікаційна робота магістра складається з завдання, реферату, змісту, переліку скорочень, вступу, 4 розділів, висновків, переліку посилань із 50 найменувань та 3 додатки. Загальний обсяг роботи кваліфікаційної роботи магістра становить 104 сторінки з них 75 сторінок основного тексту та 14 сторінок додатків. У роботі наведено 28 рисунків, 7 формул, 2 таблиці.

Ключові слова: натовп, нейромережі, моделі, відеопотік, YOLO, детекція, згортка.

Зміст

Перелік скорочень	4
Вступ	5
РОЗДІЛ 1	8
Огляд існуючих методів та моделей для підрахунку кількості людей у натовпі на основі відеопотоків	8
1.1 Натовп та його класифікація	8
1.2 Глибокі та згорткові нейронні мережі	9
1.3 Аналіз існуючих методів машинного навчання для підрахунку людей у відеопотоці	10
1.4 Розвиток моделей YOLO як інструменту комп'ютерного зору	16
1.5 Алгоритм трекінгу для об'єктів DeepSORT	20
1.6 Постановка задачі	21
РОЗДІЛ 2	23
Розробка методу визначення людей у натовпі у відеопотоці нейромережевими засобами	23
2.1 Метод визначення людей у натовпі за допомогою нейромереж	23
2.2 Архітектура нейронної мережі YOLOv11 для підрахунку об'єктів у натовпі	26
2.3 Варіації моделей YOLOv11 за розмірами	31
2.4 Тренування моделі	34
2.4.1 Функції втрат	37
2.4.2 Регуляризація та моніторинг	42
2.5 Критерії оцінювання точності нейронної мережі	44
Висновки до розділу 2	51
РОЗДІЛ 3	53
Програмна реалізація методу визначення людей у натовпі у відеопотоці нейромережевими засобами	53
3.1 Засоби розробки та тестування	53

3.2 Середовище WandB.....	55
Висновки до розділу 3	59
РОЗДІЛ 4	60
Дослідження методу визначення людей у натовпі та порівняння результатів.....	60
4.1 Налаштування параметрів навчання нейронної мережі	60
4.2 Характеристика датасету.....	62
4.3 Налаштування алгоритму трекінгу DeepSORT.....	66
4.4 Отримання результатів загальної точності методу визначення людей у натовпі	68
4.4.1 Середнє значення точності моделі	68
4.4.2 Точність і повнота моделі	71
4.4.3 Втрати на валідаційних та тренувальних даних	73
4.5 Порівняння результатів покращенох моделі з стандартною попередньо тренованою версією	75
Висновки до розділу 4	79
Загальні висновки.....	81
Перелік посилань.....	82
Додатки	

Перелік скорочень

Скорочення, термін, позначення	Пояснення
ШНМ	Штучні нейронні мережі
CNN	Convolutional Neural Network (Конволюційна нейронна мережа)
DNN	Deep Neural Network (Глибока нейронна мережа)
RNN	Recurrent Neural Network (Рекурентна нейронна мережа)
GPU	Graphics Processing Unit (Графічний процесор)
YOLO	You Only Look Once (Ви дивитесь лише один раз)
CSRNet	Convolutional Spatial-Relation Network (Конволюційна мережа з просторовими взаєминами)
R ² CNN	Region-based Convolutional Neural Network (Областєво-орієнтована конволюційна нейронна мережа)
SSD	Single Shot MultiBox Detector (Детектор об'єктів одночасного розпізнавання декількох об'єктів)
SPPF	Spatial Pyramid Pooling Fusion (Просторова пірамідальна об'єднана агрегація)
CPU	Central Processing Unit (Центральний процесор)
NMS	Non-Maximum Suppression (Немаксимальне подавлення)
AMP	Automatic Mixed Precision (Автоматичне змішане точне обчислення)

Вступ

У сучасному світі визначення кількості людей у натовпі набуває важливого значення в різних сферах життя. Цей процес є ключовим для забезпечення громадської безпеки, організації масових заходів, моніторингу транспортних вузлів, а також для оцінки потоків людей у торговельних центрах, стадіонах та інших місцях великого скупчення. Точний і швидкий аналіз натовпу дозволяє запобігати можливим небезпекам, таким як тисняви або неконтрольовані ситуації, а також сприяє оптимізації управлінських рішень у реальному часі.

Зі зростанням обсягів відеоінформації та збільшенням кількості камер спостереження виникає необхідність автоматизації процесу обробки даних.

Традиційні методи, засновані на ручному аналізі або класичних алгоритмах комп'ютерного зору, стають менш ефективними. У цьому контексті використання нейромережових засобів, зокрема сучасних моделей глибокого навчання, відкриває нові можливості для розв'язання такого типу завдань.

Ця кваліфікаційна робота магістра спрямована на розробку та вдосконалення методів автоматичного визначення кількості людей у натовпі на основі сучасних підходів до обробки відео та нейромережових технологій. Результатом роботи є метод визначення кількості людей у натовпі за допомогою спеціально навченої нейромережевої моделі.

Актуальність теми. Сучасні тенденції урбанізації та зростання кількості населення в містах призводять до збільшення концентрації людей у громадських місцях, таких як транспортні вузли, площі, спортивні та культурні заходи. Ефективний моніторинг цих локацій є критично важливим для забезпечення безпеки, управління потоками людей і зменшення ризику інцидентів, таких як тисняви чи паніка.

Існуючі підходи до визначення кількості людей у натовпі, що базуються на традиційних методах аналізу відеопотоків, часто виявляються недостатньо точними або вимагають значних ресурсів для їх реалізації. Зі збільшенням обсягів відеоінформації від камер спостереження необхідно впроваджувати автоматизовані

рішення, які можуть працювати в реальному часі та адаптуватися до змінних умов, таких як освітлення, щільність натовпу або зміна ракурсів зйомки.

У цьому контексті методи на основі нейронних мереж відкривають нові можливості для аналізу складних даних. Зокрема, моделі глибокого навчання, які здатні виконувати завдання розпізнавання об'єктів із достатньо високою точністю та швидкістю. Їх застосування для підрахунку людей у натовпі може значно підвищити ефективність систем відеоспостереження, знизити людський фактор і забезпечити адаптивність до різних сценаріїв.

Мета і задачі роботи. Метою роботи є підвищення точності та ефективності визначення кількості людей у натовпі в реальному часі у відеопотоці нейромережевими засобами. Для досягнення поставленої мети потрібно виконати наступні завдання:

- провести аналіз нейромережових моделей та існуючих підходів для визначення кількості людей у натовпі у відеопотоці;
- розробити метод визначення кількості людей у натовпі у відеопотоці з використанням нейромережової моделі;
- підготувати набір даних, що містить різноманітні зображення натовпу для донавчання моделі;
- провести валідацію запропонованого методу за відомими статистичними показниками та порівняти результати з відомими підходами.

Об'єкт дослідження □ процес визначення кількості людей у натовпі на основі відеопотоку з використанням нейромережових моделей для обробки зображень.

Предмет дослідження – моделі нейронної мережі для підрахунку людей у натовпі.

Методи дослідження. Для виконання завдання було застосовано згорткову нейронну мережу для розпізнавання людей у відеокадрі, спеціалізований датасет для підвищення точності та алгоритми трекінгу об'єктів, для збереження стабільності та підвищення ефективності у динамічних сценах.

Наукова новизна одержаних результатів. В результаті проведеної роботи удосконалено метод визначення кількості людей у натовпі за рахунок використання сучасних архітектур глибоких нейронних мереж, оптимізованих для обробки відеопотоків, який відрізняється від існуючих донавчанням моделі на спеціалізованих наборах даних та використанням алгоритмів трекінгу об'єктів у кадрі для покращення результатів у динамічних сценах, що дало можливість підвищити точність та ефективність визначення кількості людей у натовпі в реальному часі у відеопотоці.

Практичне значення отриманих результатів полягає в створенні ефективного інструменту для автоматичного підрахунку людей у натовпі, що може бути інтегрований у системи відеоспостереження, що використовуються для забезпечення безпеки на масових заходах, в транспортних вузлах, та в інших громадських місцях.

Розроблений метод дозволяє значно знизити людський фактор, забезпечуючи більш точний та оперативний аналіз в реальному часі, що особливо важливо для запобігання небезпечних ситуацій.

Апробація результатів кваліфікаційної роботи магістра та публікації. Основні положення роботи опубліковані в збірнику наукових праць – Визначення кількості людей в натовпі у відеопотоці нейромережевими засобами / Голуб Ю.В., Багрій Р.О., Вознюк Л.О. // Збірник наукових праць за матеріалами XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН□2024». Хмельницький. 2024.

Структура і обсяг роботи. Кваліфікаційна робота магістра складається з завдання, реферату, змісту, переліку скорочень, вступу, 4 розділів, висновків, переліку посилань із 50 найменувань та 3 додатки. Загальний обсяг роботи кваліфікаційної роботи магістра становить 104 сторінки, з них 75 сторінок основного тексту та 14 сторінок додатків. У роботі наведено 28 рисунків, 7 формул, 2 таблиці.

РОЗДІЛ 1

Огляд існуючих методів та моделей для підрахунку кількості людей у натовпі на основі відеопотоків

1.1 Натовп та його класифікація

Натовп – це велике скупчення людей на будь-якій території (Рис 1.1), що мають безпосередній контакт і спонтанно реагують подібним чином на одні і ті ж стимули [1]82.

Натовпи можна класифікувати на різні типи залежно від мотивації, емоційного стану та соціальних умов, які їх формують [2].



Рисунок 1.1 □ Натовп [3]

Найпоширенішими є такі підходи:

- агресивні натовпи □ виникають у ситуаціях конфлікту, протестів або паніки та демонструють схильність до руйнівних дій;
- панічні натовпи □ утворюються в ситуаціях, що загрожують безпеці учасників, часто уникаючи логіки та спонтанно діють;

- емоційні натовпи □ люди об'єднуються на основі позитивних або негативних емоцій, наприклад, під час спортивних змагань, концертів або релігійних заходів;
- випадкові натовпи □ формуються без попереднього планування, зазвичай внаслідок спонтанного інтересу до певної події

1.2 Глибокі та згорткові нейронні мережі

Штучні нейронні мережі □ це такі мережі, в основу яких покладена аналогія роботи нейронів людського мозку [4].

У загальному випадку штучні нейронні мережі (ШНМ) представляють собою систему, що складається з штучних нейронів, які взаємодіють між собою за допомогою синаптичних зв'язків. Нейрони та їхні взаємозв'язки формують структуру нейронної мережі [5].

Вона може складатися з тисяч або навіть мільйонів простих обробних вузлів, які взаємодіють між собою. Більшість сучасних нейронних мереж організовані у вигляді шарів вузлів, що мають «прямий» тип зв'язків, тобто дані рухаються через них лише в одному напрямку. Один вузол може з'єднуватися з кількома вузлами на нижньому рівні, з яких він отримує інформацію, а також із кількома вузлами на наступному рівні, до яких він передає дані [6].

Базова структура штучної нейронної мережі (Рис.1.2) складається з вхідного шару, одного або кількох прихованих шарів і вихідного шару, де кожен нейрон обробляє інформацію відповідно до встановлених ваг і функцій активації [4].

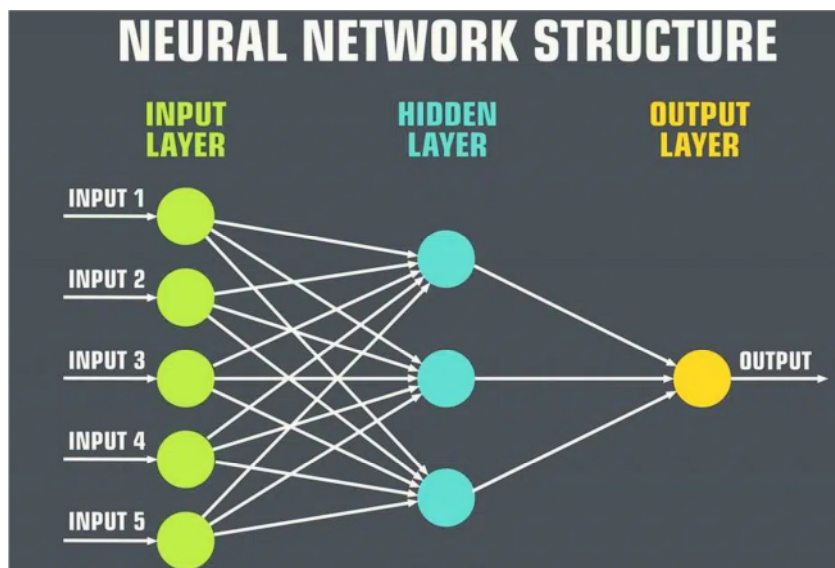


Рисунок 1.2 □ Структура ШНМ [7]

Апаратна реалізація нейронної мережі складається з обчислювальних компонентів (процесорів), які функціонують як штучні нейрони. Кожен з них має кілька зв'язків для отримання та передачі інформації, а кожен зв'язок має вагу, яка множить значення, що до нього надходить. Ці ваги визначають, як нейрони реагують на вхідну інформацію, що дозволяє мережі навчатися та адаптуватися до різних завдань [7].

Для завдань, пов'язаних із виявленням об'єктів у відеопотоці, найбільш ефективні нейронні мережі типу CNN (Convolutional Neural Networks) та DNN (Deep Neural Networks). Обидва типи мереж демонструють високу ефективність завдяки здатності справлятися зі складністю та варіативністю зображень.

Глибокі нейронні мережі (Deep Neural Networks, DNN) складаються з численних прихованих шарів, що дає їм змогу моделювати складні залежності та розпізнавати складні патерни у зображеннях. Проте для їх тренування необхідні значні обчислювальні ресурси та великі обсяги даних, що може призводити до тривалого часу навчання. Хоча такі мережі демонструють високу точність у задачах розпізнавання об'єктів, їх застосування в реальному часі обмежується через велику обчислювальну складність [8].

Конволюційні нейронні мережі (Convolutional Neural Networks, CNN) є спеціалізованими для обробки та розпізнавання зображень. Вони застосовують конволюційні шари для виявлення локальних ознак зображень, таких як краї, форми та текстури, що дозволяє автоматично навчатися репрезентативним характеристикам. CNN ефективно справляються із завданнями класифікації та розпізнавання об'єктів.

Завдяки використанню оптимізацій, зокрема паралельної обробки на графічних процесорах, ці мережі можуть працювати дуже швидко. Сучасні аналоги CNN дозволяють обробляти зображення у реальному часі, що є особливо корисним для таких застосувань, як розпізнавання об'єктів у відеопотоці чи автономна навігація роботів [9].

Структура CNN складається з двох основних етапів (Рис 1.3), що визначають її функціональність:

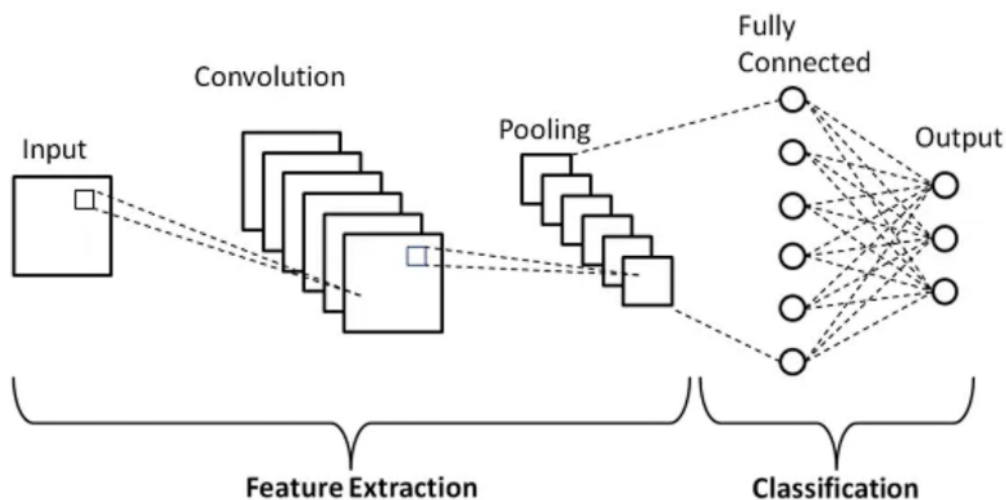


Рисунок 1.3 □ Структура CNN [10]

- етап екстракції ознак, де мережа використовує згорткові та пулінгові шари для вилучення важливих ознак із вхідних даних (наприклад, зображень);
- етап класифікації, де на основі екстрагованих ознак мережа приймає рішення, класифікуючи об'єкти або передбачаючи їхні координати у випадку виявлення об'єктів [11].

CNN складається з трьох основних типів шарів: згорткових шарів, шарів об'єднання та повністю зв'язаних (FC) шарів. Поєднання цих шарів утворює структуру згорткової нейронної мережі [12].

CNN служать основою для багатьох сучасних моделей та інструментів у сфері штучного інтелекту.

CNN також інтегруються у складні гібридні архітектури для досягнення високої точності у вирішенні складних завдань. Одним із прикладів є поєднання CNN із рекурентними нейронними мережами (RNN). Це дозволяє враховувати часові контексти у завданнях, пов'язаних з відеопотоками, послідовностями даних та аналізом часових рядів [13].

Іншою перспективною технологією є використання механізмів уваги (Attention Mechanisms). Ці механізми дозволяють нейронній мережі фокусуватися на найбільш важливих аспектах вхідних даних, оптимізуючи обробку інформації, що є критичним для задач із високими обчислювальними вимогами [14].

CNN та їх різновиди, як частина глибоких нейронних мереж, застосовуються в багатьох галузях, зокрема:

- медицина □ для виявлення патологій на медичних зображеннях;
- автономний транспорт □ для розпізнавання перешкод та аналізу навколишнього середовища;
- системи безпеки □ для моніторингу та аналізу зображень у режимі реального часу;
- розпізнавання облич □ для ідентифікації осіб у різноманітних ситуаціях.

Оптимізація архітектур CNN та їх застосування у різноманітних галузях відкриває нові можливості в галузі штучного інтелекту. Подальші дослідження в напрямку архітектур, навчальних алгоритмів та обчислювальної ефективності сприятимуть ще більш швидкому розвитку цієї технології.

1.3 Аналіз існуючих методів машинного навчання для підрахунку людей у відеопотоці

Методи детекції. Методи детекції об'єктів на основі машинного навчання спрямовані на виявлення кожної окремої людини у кадрі та підрахунок їх кількості на основі координат детекції. Найчастіше ці методи базуються на нейронних мережах, які навчені на великих наборах даних для ідентифікації людей. Сучасні методи детекції об'єктів, такі як YOLO (You Only Look Once), SSD (Single Shot MultiBox Detector) або Faster R-CNN, показали високу ефективність у задачах виявлення людей у реальному часі [15;16].

Для таких моделей характерна здатність до швидкого обчислення координат кожної людини, що дозволяє підраховувати їх у відеопотоці з мінімальною затримкою. Ці методи широко застосовуються в системах моніторингу, розумних транспортних системах, робототехніці та безпекових системах.

Демонстрацію роботи методів детекції об'єктів можна переглянути на (Рис. 1.4).

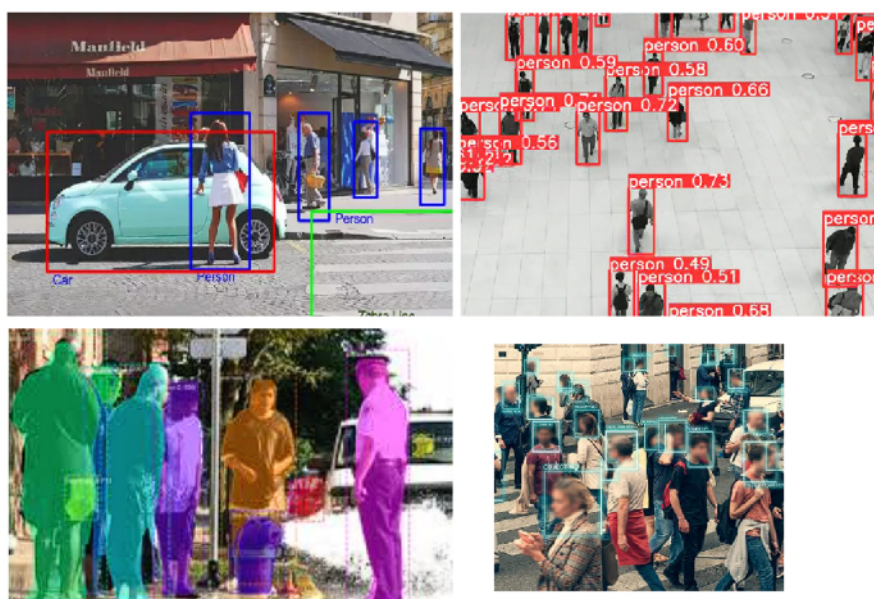


Рисунок 1.4 □ Робота НМ детекторів

Комп'ютерний зір □ це стрімко зростаюча галузь, що дозволяє машинам аналізувати та розпізнавати візуальну інформацію. Однією з ключових складових цієї сфери є виявлення об'єктів, що передбачає точне визначення та локалізацію об'єктів на зображеннях або у відеопотоках [18;17].

Одним із новаторських методів у цій галузі є алгоритм You Only Look Once (YOLO), який був розроблений Redmon та співавторами у 2015 році. Назва алгоритму відображає його унікальний підхід: обробка всього зображення відбувається лише один раз для визначення об'єктів та їх координат. На відміну від традиційних підходів, що застосовують двоетапні процеси детектування, YOLO формулює задачу виявлення об'єктів як регресійну проблему [15].

Модель SSD, вона менш точна, ніж YOLO в деяких випадках, але може працювати з меншими ресурсами. SSD є однопрохідним детектором, який не використовує мережу регіональних пропозицій. Замість цього він напряму прогнозує межі об'єктів та їхні класи з карт ознак за один прохід [19].

Модель Faster R-CNN, вона була створена командою дослідників із Microsoft.

Це глибока згорткова нейронна мережа для виявлення об'єктів, яка функціонує як єдина наскрізна архітектура. Faster R-CNN здатна швидко та точно визначати місцезнаходження об'єктів на зображеннях. Щоб повністю зрозуміти Faster R-CNN, варто ознайомитися з її попередниками – моделями R-CNN та Fast R-CNN, оскільки Faster R-CNN є розширенням Fast R-CNN. Як вказує сама назва, Faster R-CNN забезпечує вищу швидкість у порівнянні з Fast R-CNN завдяки використанню мережі регіональних пропозицій (RPN) [19].

Ці пропозиції регіонів потім уточнюються R-CNN, який класифікує об'єкти та коригує координати обмежувальної коробки. Використовуючи карти ознак зі згорткових шарів, Faster R-CNN ефективно виявляє об'єкти з високою точністю, але його двоетапний процес призводить до затримки в порівнянні з одноступеневим детектором YOLO [20].

Методи оцінки щільності натовпу є ефективними підходами для підрахунку кількості людей у відеопотоці або зображенні без необхідності точного виявлення кожної окремої людини. Застосування цих методів дозволяє швидко та точно аналізувати ситуацію в кадрі, особливо у випадках, коли люди мають високу ступінь скупчення, що ускладнює застосування методів детекції об'єктів.

Ці методи засновані на концепції карт щільності (Рис 1.5), де кожен піксель у карті містить значення, що відображає щільність людей у відповідній ділянці кадру.

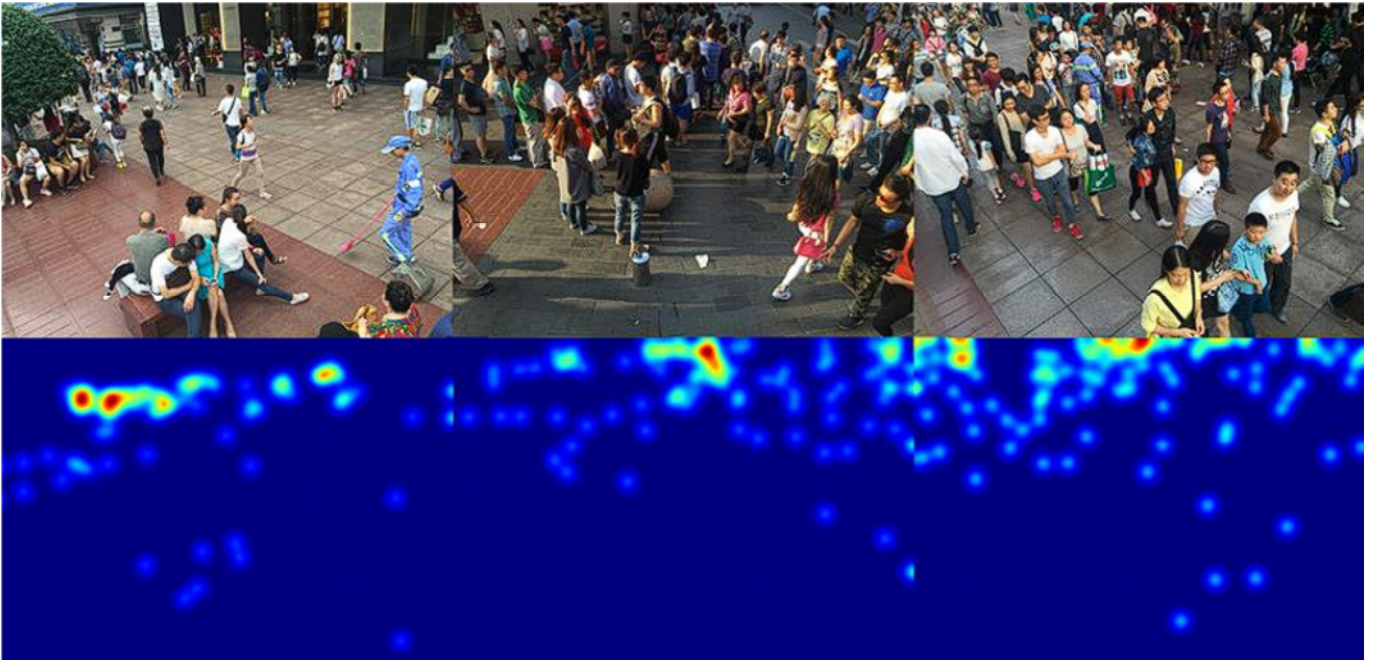


Рисунок 1.5 □ Приклад карти щільності для 3 зображень із набору даних Shanghai data set [21]

Це значення показує ймовірність наявності людини в певній області зображення, що дозволяє оцінювати кількість людей без необхідності ідентифікації кожного з них індивідуально.

Ключовим аспектом є те, що модель навчається на основі великих наборів даних, де кожна точка на карті щільності відповідає розташуванню людей у кадрі. Моделі використовують нейронні мережі, такі як згорткові мережі (CNN), для створення та прогнозування таких карт щільності [22].

Однією із найбільш ефективних моделей для оцінки щільності натовпу □ це MCNN (Multi-column Convolutional Neural Network), який використовує кілька колонок згорткових нейронних мереж для роботи з даними на різних масштабах. Цей підхід особливо ефективний у випадках, коли люди дуже близько стоять одне до одного або сховані одна за одною, що робить завдання підрахунку людей у кадрі

складнішим. Метод MCNN дозволяє генерувати більш точні карти щільності завдяки адаптації до різноманітності масштабів у зображенні [22].

MCNN працює на основі кількох згорткових мереж, які функціонують паралельно на різних рівнях масштабування. Кожна з колонок обробляє зображення з різним рівнем деталізації, що дозволяє моделі фокусуватися на ознаках у різних масштабах і забезпечувати високу точність у складних сценаріях. Завдяки цьому метод здатний ефективно працювати з високою щільністю, коли окремі люди на зображенні схожі між собою або перекриваються [23].

Інші модель, що використовує цей метод □ це CSRNet (Crowd Subnet), вона використовує розширені згортальні мережі з розширеними шарами. Розширені згортки дозволяють мережі зберігати більше просторової інформації, що важливо для генерування точних карт щільності, навіть у дуже щільних натовпах. CSRNet показав чудові результати в складних умовах, де інші методи можуть зазнавати труднощів через втрату просторової інформації або недостатню точність [24].

Але методи на основі генерації карт щільності мають недоліки. CSRNet та MCNN моделі працюють у три етапи: попередня обробка зображень, створення карти щільності за допомогою згорток та підрахунок об'єктів на основі цієї карти. Ця структура забезпечує високу ефективність підрахунку особливо для щільних натовпів, але потребує значних обчислювальних ресурсів через складності в обробці.

Загалом, усі методи мають свої переваги та недоліки в залежності від задачі та умов застосування. Метод детекції об'єктів є найпоширенішим для підрахунку людей у відеопотоці, оскільки він дозволяє отримати достатньо точні результати при високій швидкості обробки та має простішу реалізацію. Проте для ситуацій з високою щільністю натовпу або у випадках, коли розрізнення між людьми складне, методи оцінки щільності та трекінгу можуть показати кращі результати в точності на щільних натовпах, але так як це двоетапні моделі, їхня генерація карт щільності займає додатковий час.

1.4 Розвиток моделей YOLO як інструменту комп'ютерного зору

Моделі YOLO (You Only Look Once) залишаються одними з найпопулярніших підходів у детекції об'єктів, демонструючи високу швидкодію, точність та здатність працювати в реальному часі. Їх успіх обумовлений спрощеною архітектурою, яка дозволяє об'єднувати всі етапи обробки зображення в одну операцію. Завдяки постійним вдосконаленням, нові версії, як от YOLOv8, пропонують підвищену точність, кращу адаптацію до різних завдань і зниження обчислювальних витрат, що робить їх особливо ефективними для роботи на пристроях із обмеженими ресурсами (наприклад, вбудованих системах) [25].

Останні оновлення моделей YOLO демонструють суттєві покращення точності та продуктивності. Зокрема, YOLOv11 має нову архітектуру, що включає вдосконалені блоки для обробки характеристик (C3k2 та CBS), оптимізовану обчислювальну ефективність і зменшення параметрів моделі на 22% порівняно з YOLOv8, при цьому досягаючи вищої точності. Додатково, YOLOv11 покращує підтримку задач, таких як сегментація, оцінка позицій, відстеження об'єктів та класифікація зображень, що розширює її застосування в медицині, безпілотних системах та промисловості **Ошибка! Источник ссылки не найден.**]

Архітектура моделей YOLO (You Only Look Once) базується на кількох основних компонентах (Рис. 1.6), що забезпечують їх високу швидкодію та точність у задачах детекції об'єктів. Архітектура включає три основні блоки, через які проходять дані: backbone, neck та head (Dense Prediction).

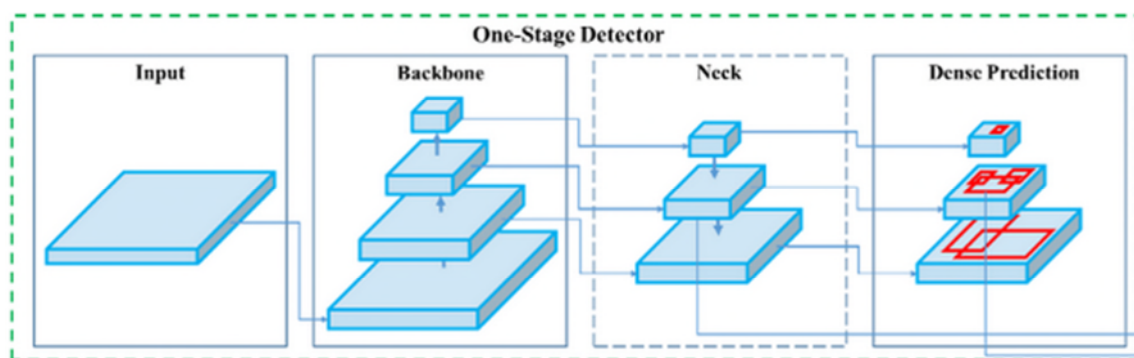


Рисунок 1.6 □ Архітектура моделей YOLO [**Ошибка! Источник ссылки не найден.**]

Backbone є основним модулем, відповідальним за вилучення ознак із вхідного зображення. Цей блок зазвичай включає конволюційні нейронні мережі, такі як ResNet або MobileNet, які виконують обробку зображень на різних рівнях абстракції. Саме тут відбувається виділення базових ознак, необхідних для подальшого аналізу [28].

Neck служить посередницьким модулем, що об'єднує ознаки з різних рівнів, отриманих із блоку backbone, для покращення виявлення об'єктів різного масштабу. Одним із популярних механізмів у цьому блоці є Feature Pyramid Network (FPN), що дозволяє ефективно комбінувати ознаки з різних масштабів [29].

Head є кінцевим етапом архітектури, який відповідає за генерацію предсказань, таких як координати обмежувальних прямокутників (bounding boxes) та класифікація об'єктів. Цей блок використовує вилучені ознаки, щоб здійснити виявлення об'єкта у заданому зображенні або відеопотоці [27].

У сучасних версіях YOLO значна увага приділяється оптимізації кожного з цих блоків, зокрема покращенню швидкості обробки, зменшенню кількості обчислювальних витрат та підвищенню точності через удосконалення [30].

У порівнянні з YOLOv5, одинадцята версія забезпечує не лише вищу точність, але й швидшу обробку, що є критичним для задач реального часу, таких як відеоаналітика та автономне керування. Наприклад, YOLOv11 може обробляти до 5×6 мс на зображення, підтримуючи гнучкість на різних платформах, включаючи хмарні рішення та пристрої на межі (edge devices). Графічне порівняння точності моделей від YOLOv5 до YOLOv11 можна переглянути на рисунку 1.7 [26;31].

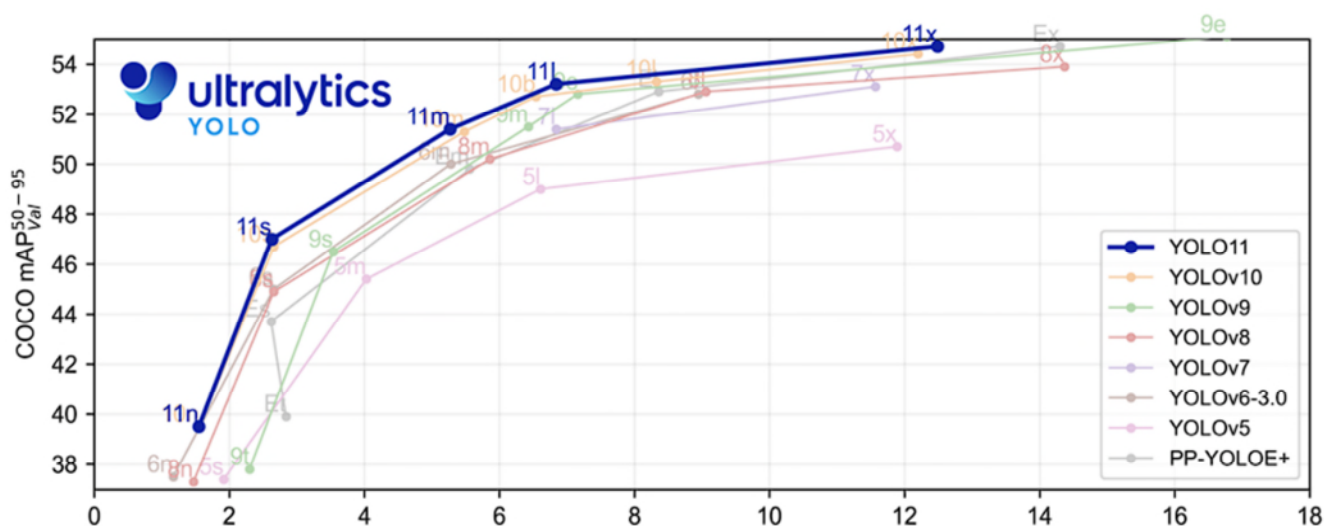


Рисунок 1.7 □ Прогрес в точності різних версій моделей YOLO [33]

Основний прогрес моделей YOLOv5(2020p), YOLOv8(2023p) та YOLOv11(2024p):

Модель YOLOv5 є однією з найпопулярніших ітерацій сімейства YOLO (You Only Look Once), яка поєднує високу продуктивність із зручністю використання. Її успіх обумовлений оптимізованою архітектурою, яка забезпечує швидку обробку зображень у реальному часі та високоточну детекцію об'єктів. Основною перевагою YOLOv5 є впровадження підтримки PyTorch, що спрощує інтеграцію моделі у сучасні платформи машинного навчання [31].

Модель YOLOv8 є однією з найсучасніших ітерацій у серії YOLO, яка продовжує удосконалювати баланс між точністю детекції та швидкістю обробки. Однією з основних інновацій YOLOv8 є впровадження нових архітектурних елементів, зокрема оновленого детекційного блоку та вдосконалених механізмів обробки ознак, що дозволяє моделі забезпечувати високу продуктивність у задачах реального часу. Крім того, YOLOv8 підтримує функцію автоматичної аугментації, що дозволяє підвищити точність навчання без збільшення кількості вихідних даних [26].

У YOLOv8 реалізовано покращені блоки neck (наприклад, PANet) та head, які оптимізують обробку багатомасштабних характеристик для детекції об'єктів різного

розміру. Крім того, модель є енергоефективною, що робить її придатною для використання на пристроях із обмеженими ресурсами [32].

Ще однією важливою перевагою починаючи з YOLOv8 є її сумісність із хмарними платформами та edge-пристроями, що забезпечує широкий спектр застосувань. Наприклад, YOLOv8 демонструє високу ефективність у задачах аналізу відеопотоків, медичній діагностиці, промисловій автоматизації та робототехніці. Завдяки цим покращенням YOLOv8 стала одним із найкращих виборів для задач комп'ютерного зору [26].

Особливою перевагою моделі є її здатність адаптуватися до різноманітних апаратних платформ, від edge-пристроїв до хмарних обчислювальних центрів. Завдяки оптимізованій архітектурі та покращеним блокам обробки, YOLOv11 працює швидше за попередні версії та показує високу ефективність у задачах реального часу, таких як відеоспостереження, безпілотні транспортні засоби, а також медична діагностика [33].

1.5 Алгоритм трекінгу для об'єктів DeepSORT

DeepSORT – це покращення оригінального алгоритму SORT (Simple Online and Realtime Tracking), він є одним із найпопулярніших методів для здійснення відстеження об'єктів у режимі реального часу в системах комп'ютерного зору. Ця модель забезпечує ефективне спостереження за об'єктами, використовуючи методи детекції об'єктів у поєднанні з траєкторіями на основі фільтрації Калмана та класифікації через асоціативний аналіз. Основною особливістю DeepSORT є його здатність до поєднання інформації з візуальних даних та просторових координат, що дозволяє відстежувати об'єкти навіть у складних умовах із частковим перекриттям, відсутністю візуальної інформації або короткими перервами в даних [35].

Метод DeepSORT працює у поєднанні з детекторами об'єктів, такими як YOLO (Рис.1.8) або SSD, де основна роль детектора полягає у виявленні об'єктів на кадрах відеопотоку, а DeepSORT забезпечує їх відстеження на основі асоціативної інформації. Завдяки цьому, DeepSORT ефективно використовує методи

евристичного аналізу та машинного навчання для визначення найбільш ймовірних асоціацій між детекціями у послідовності кадрів, що дозволяє зменшити помилки відстеження, які можуть виникати при складному середовищі.

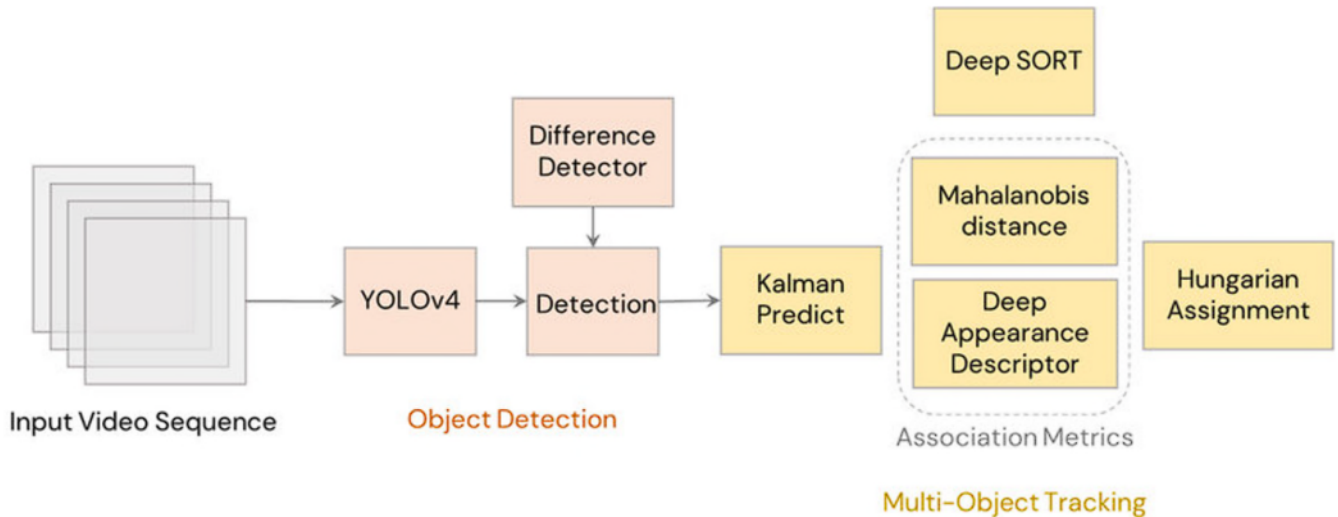


Рисунок 1.8 □ Схема роботи алгоритму трекінгу DeepSORT в поєднанні з YOLO [36]

DeepSORT часто застосовується у різноманітних галузях, включаючи відеоспостереження, автономні транспортні засоби, медичні системи моніторингу, робототехніку та інші сфери, де потрібне ефективне відстеження великої кількості об'єктів. Головною перевагою DeepSORT є його висока точність, швидкість обробки та адаптивність до різних варіантів відеопотоків, що робить його універсальним інструментом для задач реального часу [36].

1.6 Постановка задачі

Метою роботи є підвищення точності та ефективності визначення кількості людей у натовпі в реальному часі у відеопотоці нейромережевими засобами. Для досягнення поставленої мети потрібно виконати наступні завдання:

- провести аналіз нейромережових моделей та існуючих підходів для визначення кількості людей у натовпі у відеопотоці;

- розробити метод визначення кількості людей у натовпі у відеопотоці з використанням нейромережевої моделі;
- підготувати набір даних, що містить різноманітні зображення натовпу для донавчання моделі;
- провести валідацію запропонованого методу за відомими статистичними показниками та порівняти результати з відомими підходами.

Успішним виконанням завдання можна вважати підвищення точності методу у процесі обчислень.

РОЗДІЛ 2

Розробка методу визначення людей у натовпі у відеопотоці нейромережевими засобами

2.1 Метод визначення людей у натовпі за допомогою нейромереж

Метод визначення кількості людей за допомогою згорткової мережі YOLOv11 полягає у детекції об'єктів у кадрі, схема методу зображена на рисунку 2.1, сам метод виконується у наступні кроки:

– *Підготовка до обробки відеоматеріалу.* Кадри з відеоматеріалу завантажуються в модель, масштабуються та нормалізуються відповідно до форматів вхідних даних YOLOv11, які включають адаптивну обробку з урахуванням високої роздільної здатності.

– *Завантаження початкових згорткових шарів.* Виконується виділення базових ознак, таких як лінії, контури та текстури, з урахуванням покращеної ефективності завдяки оптимізованим шаровим блокам у YOLOv11.

– *Застосування багаторівневої підвибірki та агрегації ознак.* Підвибірка виконується через адаптивний \max -pooling та нові алгоритми агрегації ознак, що зберігають більше контекстної інформації та забезпечують стійкість до масштабів.

– *Обробка глибокими згортковими шарами.* Модель аналізує високорівневі узагальнені ознаки, такі як форма, розташування та взаємодія об'єктів. YOLOv11 використовує вдосконалені механізми багатомасштабного навчання для поліпшення якості детекції.

– *Прогнозування обмежувальних рамок та класів об'єктів.* Модель за допомогою розширених блоків прогнозування визначає обмежувальні рамки та ймовірності належності до класу “людина”, забезпечуючи більш точну локалізацію завдяки гібридним втратам.

– *Оптимізація обмежувальних рамок.* Використовується оновлений алгоритм NMS (Non-Maximum Suppression) або його вдосконалена варіація, яка враховує співвідношення розмірів і перекриття для ефективного усунення зайвих рамок.

– *Трекінг та підрахунок.* Підрахунок кількості людей базується на об'єднанні детекції з механізмами трекінгу. Використовується алгоритм трекінгу об'єктів DeepSORT у реальному часі з урахуванням попередніх кадрів та ідентифікації об'єктів для покращення стійкості результатів.

– *Виведення результатів обробки.* Отримані дані, включаючи ідентифікацію людей, їхню фактичну, та додану кількість унікальних об'єктів (людей), що були помічені протягом певного періоду, яка виводиться на монітор у реальному часі.

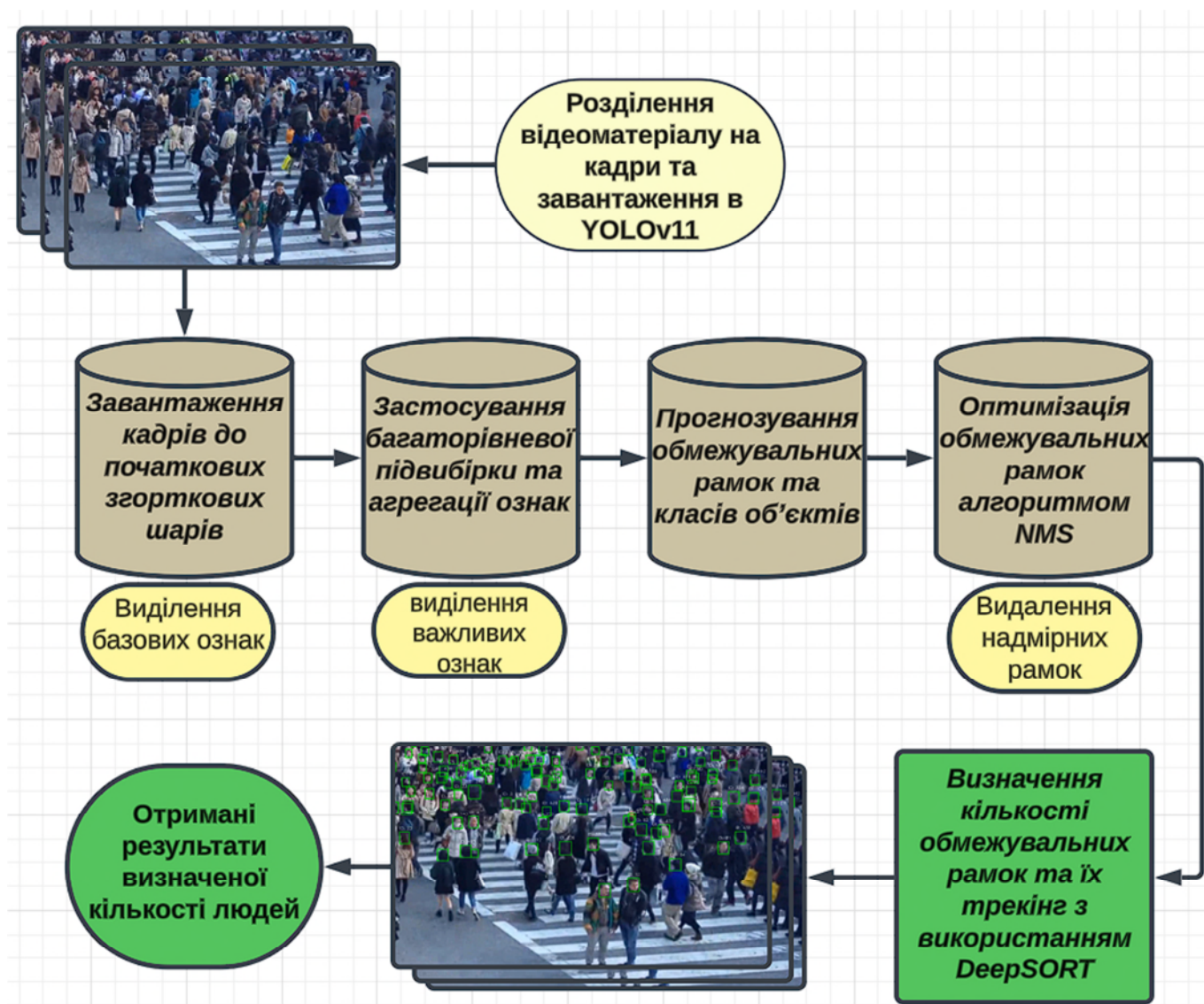


Рисунок 2.1 □ Метод визначення кількості людей в натовпі у відеопотоці

Покращення методу полягає у додаванні алгоритму DeepSORT, який забезпечує стабільність і точність детекції об'єктів (людей) у відеопотоці та

тренуванні з індивідуальними налаштуваннями моделі на спеціалізованому датасеті, що має розбіжності з попереднім тренуванням у тому, що детекція буде відбуватися не на все тіло людини а лише на обличчя та голову, що покращить результати у щільності. Схема методу визначення кількості людей у натовпі у відеопотоці за допомогою YOLOv11 та DeepSORT показана на рисунку 2.2, основні покращення полягають у підвищенні точності, стабільності відстеження та адаптивності системи.

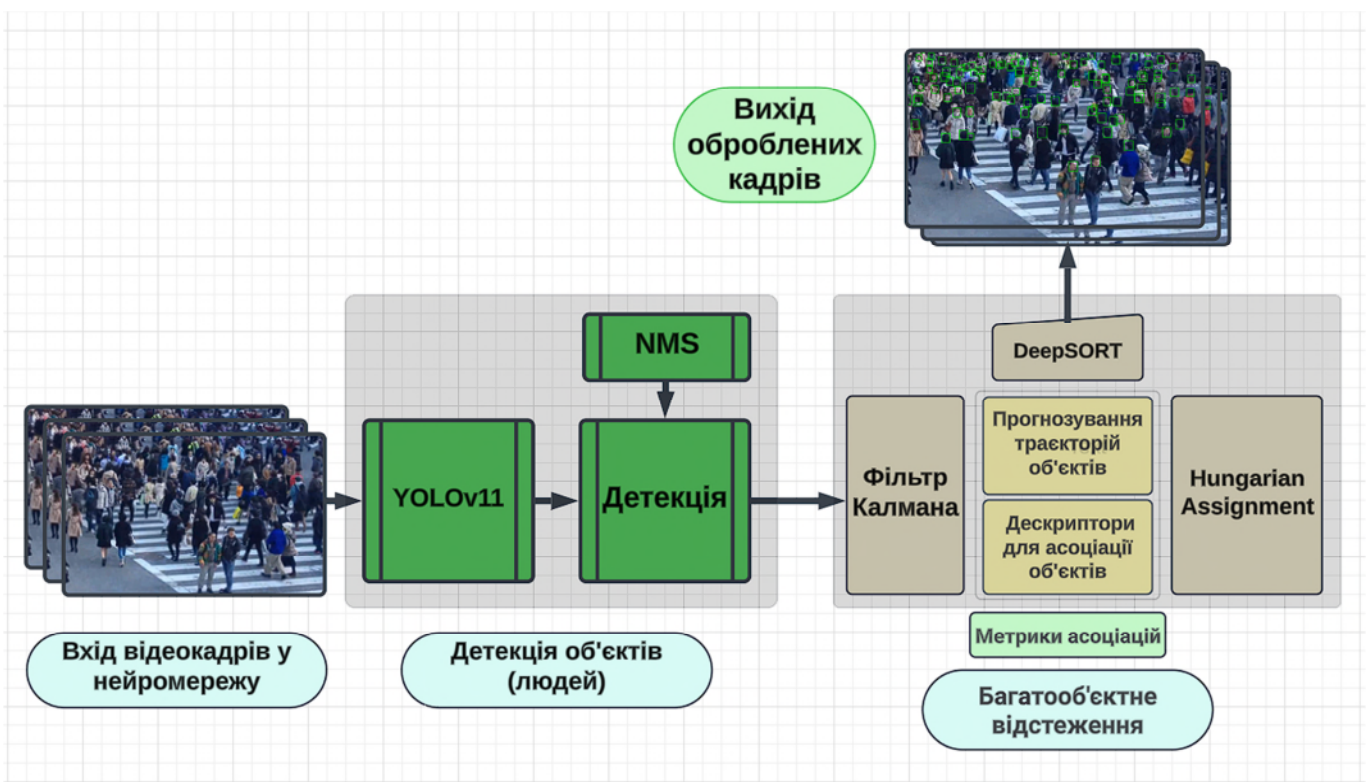


Рисунок 2.2 □ Візуалізація проходу кадрів через YOLOv11 та DeepSORT

DeepSORT, як трекер об'єктів, додає можливість точного відстеження руху осіб між кадрами відео. DeepSORT покращує підрахунок людей у кадрі завдяки використанню комбінованого підходу до ідентифікації та відстеження. Основою алгоритму є асоціація виявлених об'єктів із їхніми траєкторіями через застосування косинусної схожості ознак об'єкта, отриманих за допомогою нейронної мережі, і калманівського фільтра для прогнозування місця розташування об'єкта у наступних кадрах. Такий підхід дозволяє точно ідентифікувати кожну людину, зменшуючи ризик дублювання при повторному входженні об'єкта в кадр чи його тимчасовому

зникненні. Крім того, врахування просторово-часового контексту забезпечує точне розпізнавання та облік навіть у динамічних чи складних сценах, що значно підвищує ефективність підрахунку людей. Оптимізація обчислювальних витрат у цих алгоритмах дозволяє працювати зі значною кількістю кадрів за секунду, що є важливим для практичного застосування на відеопотоках.

В основі DeepSORT лежать два компоненти: методи оцінки руху на основі фільтра Калмана і механізм порівняння ознак між об'єктами, що дозволяє ефективніше ідентифікувати об'єкти на основі їхніх характеристик, навіть у складних ситуаціях, таких як перекриття об'єктів та зміна руху.

Hungarian assignment (угорський метод) використовується для вирішення задачі призначення, яка полягає у зіставленні виявлених об'єктів між кадрами відео. Це дозволяє відстежувати об'єкти в реальному часі, навіть якщо вони рухаються або змінюють свою позицію.

Основні етапи включають:

- обчислення матриці відстаней \square визначаються відстані між об'єктами в поточному кадрі та об'єктами в попередньому кадрі;
- призначення об'єктів \square використовуючи угорський метод, об'єкти з поточного кадру призначаються об'єктам з попереднього кадру на основі мінімізації загальної відстані;
- оновлення траєкторій \square після призначення об'єктів оновлюються їхні траєкторії, що дозволяє точно відстежувати їх у наступних кадрах.

2.2 Архітектура нейронної мережі YOLOv11 для підрахунку об'єктів у натовпі

Модель YOLOv11 є одним із найшвидших і найефективніших інструментів для завдань, пов'язаних із детекцією об'єктів у реальному часі, зокрема підрахунком людей у натовпах через відеопотоки. Завдяки своїй високій швидкодійності, низькій затримці обробки та сучасним методам нейронних мереж, вона добре підходить для аналізу ситуацій у великих скупченнях людей.

Переваги YOLOv11 для завдання підрахунку людей заключаються у наступному:

- у YOLOv11 використовується трансформерна архітектура для більш ефективного вилучення ознак із зображень, що забезпечує вищу точність детекції, ніж традиційні CNN. YOLOv11 використовує новітні методи обробки зображень, що дозволяє забезпечити високу точність навіть при великій кількості людей у кадрі та в умовах поганого освітлення чи складного фону;

- завдяки низькій затримці та ефективному використанню обчислювальних ресурсів, YOLOv11 дозволяє вести моніторинг у реальному часі, що є критично важливим для ситуацій у великих натовпах;

- ця функціональність дозволяє моніторити значні території з високою щільністю людей, що важливо для планування безпеки, організації заходів і підтримки громадського порядку.

YOLOv11 представляє собою нові, покращення в архітектурі порівнянно з попередніми версіями (Рис 2.3), яка впроваджує значні інновації в сфері швидкого та точного розпізнавання об'єктів у режимі реального часу. Ця версія оптимізована для різних завдань комп'ютерного зору, таких як детекція об'єктів, сегментація, оцінка пози, а також виявлення об'єктів у складних умовах а також трекінг, завдяки покращеним старим та додаванню нових модулів, які збільшують її можливості.

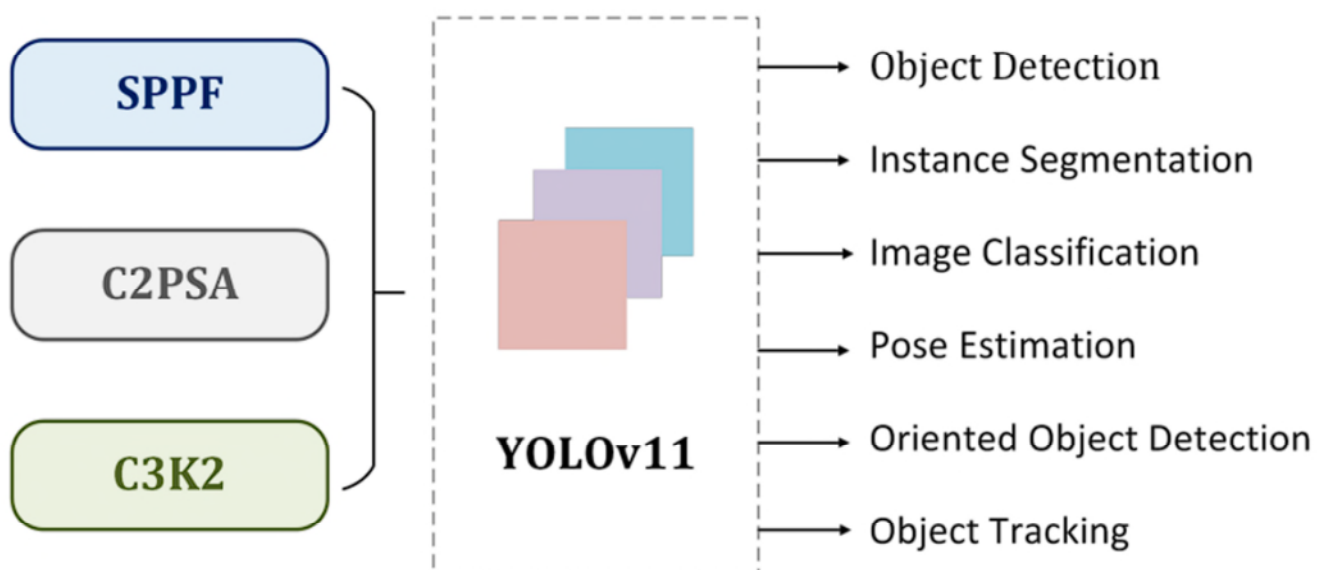


Рисунок 2.3 □ Покращення архітектури YOLOv11 [33]

Зокрема, архітектура нейронної мережі YOLOv11, яка використовується у методі визначення кількості людей у натовпі (Рис.2.4) в цілому містить наступні модулі:

- *Conv*. Звичайний 2D-згортковий шар для виділення ознак.
- *C3k2*. Блок згортки з кількома пропусками (residual blocks) для кращого навчання.
- *SPPF*. Блок просторового пірамідального пулінгу для зменшення масштабної варіативності об'єктів.
- *C2PSA*. Просторово-середовищний механізм уваги (PSA) для фокусування на значущих ознаках.
- *Upsample*. Збільшує розмір ознак для об'єднання.
- *Concat*. Об'єднання ознак по каналах (конкатенація).
- *Detect*. Основний блок для детекції об'єктів на основі ознак.

Архітектура нової моделі значно покращена в порівнянні з попередниками завдяки впровадженню блоків уваги, більш ефективним методам попередньої обробки, а також оптимізованим блокам зменшення розмірів параметрів. Зокрема, механізм C2PSA (Cross Stage Partial with Spatial Attention) допомагає моделі фокусуватися на важливих ділянках зображення, що є важливим фактором для точної детекції навіть у випадках, коли об'єкти частково закриті або дрібні.

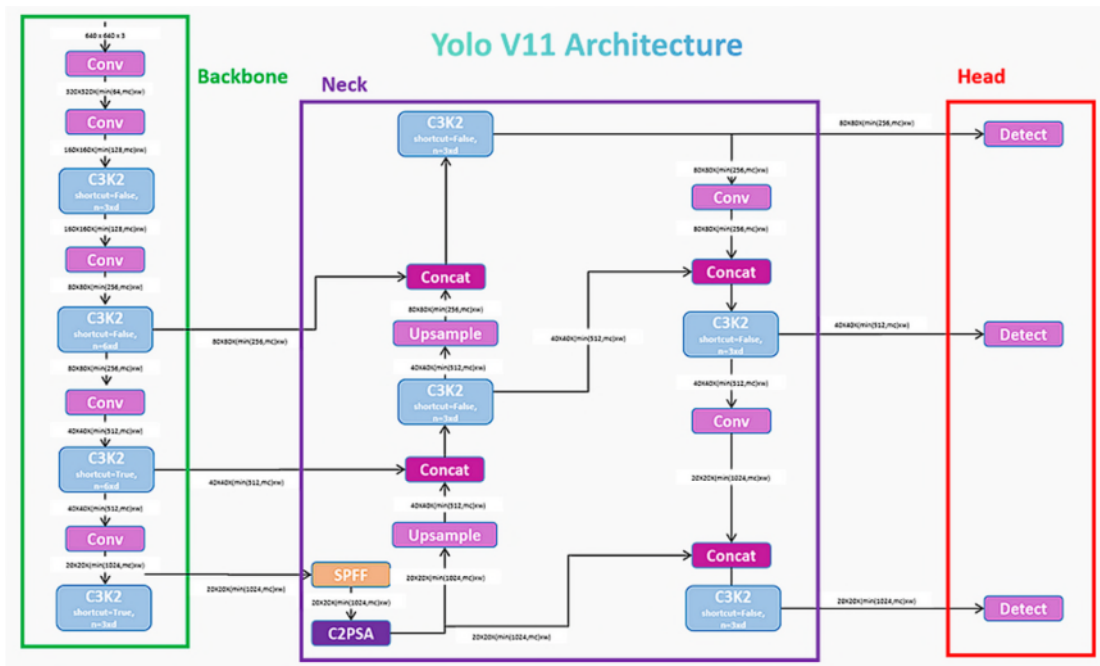


Рисунок □ 2.4 Архітектура YOLOv11 [26]

Окрім цього, YOLOv11 використовує блоки C3K2 для ефективного витягання ознак із зображень, що дозволяє зменшити обчислювальну складність у порівнянні з попередніми версіями. Блок C3K2 працює на основі невеликих 3×3 ядер замість більших, що дозволяє досягати високої продуктивності з меншою кількістю обчислювальних ресурсів.

Spatial Pyramid Pooling Fast (SPFF) □ механізм дозволяє об'єднувати інформацію з різних масштабів зображень, що підвищує точність розпізнавання об'єктів різних розмірів, особливо дрібніших. Це забезпечує баланс між швидкістю роботи та якістю виявлення, що є важливим у завданнях реального часу.

Таким чином, архітектура YOLOv11 поєднує в собі швидкість, точність та енергоефективність завдяки інтеграції новітніх методів оптимізації, механізмів уваги, швидших блоків і ефективного витягання ознак. Завдяки цьому YOLOv11 стає однією з найпотужніших моделей для задач візуального розпізнавання в реальному часі.

Регресійні модулі в архітектурі YOLOv11 додають функціонал для створення теплових карт щільності натовпів. Ці модулі значно розширюють можливості стандартної детекції об'єктів, дозволяючи моделі не лише виявляти окремі об'єкти, а

й аналізувати концентрацію об'єктів (людей) у різних ділянках кадру. Регресійні модулі функціонують на основі додаткових регресійних шарів, які обробляють дані функціональних карт для підрахунку кількості об'єктів у певній ділянці зображення. Ці функції дозволяють моделі створювати теплові карти, які візуалізують місця з найбільшою концентрацією об'єктів у кадрі. Теплові карти надають інформацію про щільність натовпу, що важливо для аналізу в завданнях безпеки, організації масових заходів або оцінювання евакуаційних сценаріїв.

Регресійні модулі обробляють великі масиви даних з допомогою ефективних механізмів, SPFF (Spatial Pyramid Pooling Fast) та C2PSA (Cross Stage Partial with Spatial Attention), які забезпечують фокусування на ключових ділянках зображення, що поліпшує точність аналізу в складних ситуаціях, коли об'єкти можуть перекриватися або бути в різних масштабах.

Хоча додавання регресійних модулів підвищує точність моделі завдяки візуалізації щільності натовпу, це також створює певний вплив на швидкодію. Додаткові обчислювальні ресурси, необхідні для обробки цих шарів, можуть знижувати швидкість моделі порівняно зі стандартною версією детекції.

Координати обмежувальної коробки як зміщення щодо заздалегідь визначеного набору розмірів якорної коробки. Ці розміри якоря мають важливе значення для ініціалізації процесу прогнозування та можуть суттєво вплинути на продуктивність моделі (Рис.2.5).

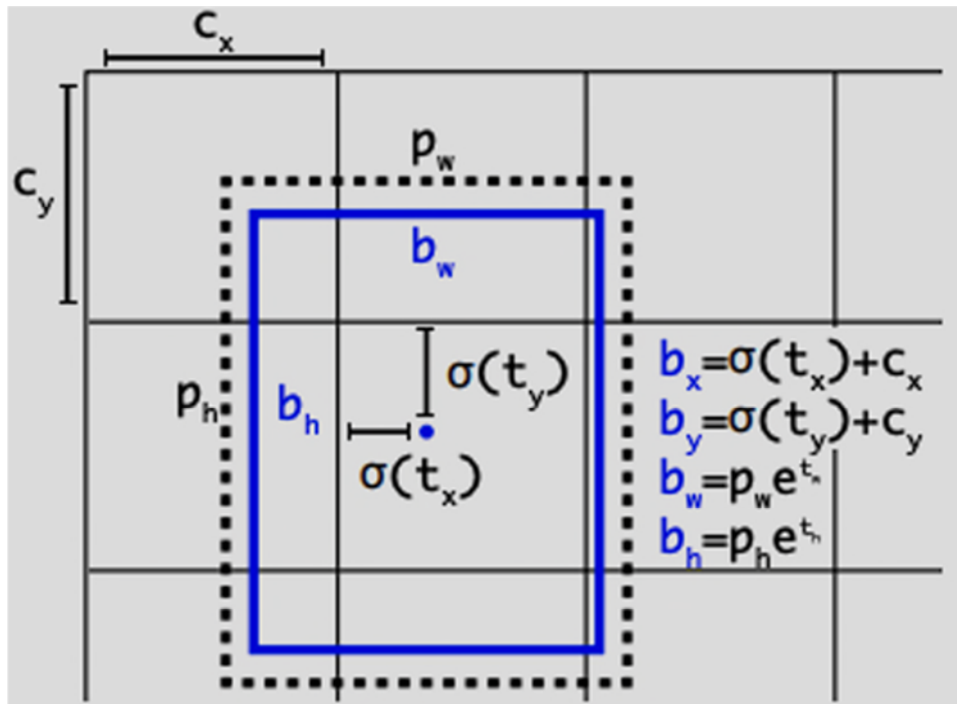


Рисунок 2.5 □ Прогнозування обмежувальної коробки [37]

Отже, для завдання підрахунку людей у відеопотоці модель YOLOv11 підходить по багатьом параметрам. Її вибір обґрунтований високою ефективністю та адаптивністю до систем із різними обчислювальними можливостями, швидкістю обробки та здатністю працювати в режимі реального часу. Ці фактори є важливими, коли необхідна висока точність, низька затримка та оперативний аналіз відеопотоку, наприклад, у відеоспостереженні.

2.3 Варіації моделей YOLOv11 за розмірами

Модель YOLOv11 є найсучаснішою з серії моделей, які спеціалізуються на виявленні об'єктів на зображеннях та відеопотоках. Вона пропонує ефективний підхід до задач детекції об'єктів завдяки своїй одноетапній архітектурі, що забезпечує баланс між швидкістю та точністю. Модель відзначається можливістю ефективного навчання на різних наборах даних і використовує найсучасніші методи для оптимізації часу обчислень без втрати якості результатів.

Вона підтримує різноманітні конфігурації, що дозволяє адаптувати її до задач із різними рівнями складності та технічними можливостями обладнання. Архітектура підвищує ефективність на етапах виявлення об'єктів у реальному часі завдяки оптимізованим обчислювальним шарам і методам навчання.

Для попереднього навчання моделей використовується стандартний набір даних COCO [39], який широко застосовують у задачах детекції об'єктів, але за бажанням, модель можна покращити, перенавчивши на власному наборі даних.

Моделі YOLOv11 доступні у п'яти основних розмірах (Табл. 2.1): nano (n), small (s), medium (m), large (l) та extra-large (x). Ці розміри оптимізовані для різних завдань, ресурсів і швидкості обчислень.

Таблиця 2.1 Різниця між розмірами моделей YOLO

Model	size (pixels)	mAPval 50□95	Speed CPU ONNX (ms)	Speed T4 TensorRT10 (ms)	params (M)	FLOPs (B)
YOLOv11n	640	39.5	56.1 ± 0.8	1.5 ± 0.0	2.6	6.5
YOLOv11s	640	47.0	90.0 ± 1.2	2.5 ± 0.0	9.4	21.5
YOLOv11m	640	51.5	183.2 ± 2.0	4.7 ± 0.1	20.1	68.0
YOLOv11l	640	53.4	238.6 ± 1.4	6.2 ± 0.1	25.3	86.9
YOLOv11x	640	54.7	462.8 ± 6.7	11.3 ± 0.2	56.9	194.9

YOLOv11n (Nano) □ Ця модель є найменшою за кількістю параметрів. Вона швидка, оптимізована для використання на ресурсах із обмеженою обчислювальною

потужністю, проте має нижчу точність порівняно з іншими розмірностями. Наприклад, вимірюючи параметри mAP 50□95, вона досягає 39,5 на наборі COCO [39].

YOLOv11s (Small) □ модель "small" є більш точною за рахунок більшої кількості обчислювальних ресурсів. mAP 50□95 дорівнює 47,0. Вона балансує між швидкістю роботи та точністю в завданнях детекції

YOLOv11m (Medium) □ ця версія вже більш оптимізована для завдань, де важлива точність, та досягає міри 51,5 у mAP 50□95 при детекції на COCO. Це підходить для середнього рівня завдань, що потребують балансу обчислювальної складності та швидкості

YOLOv11l (Large) □ вона забезпечує високий рівень точності завдяки вищій кількості обчислювальних ресурсів. Значення mAP 50□95 тут вже 53,4, що робить цю модель сильною в обробці більш складних завдань детекції

YOLOv11x (Extra□large) □ найпотужніша версія у пакеті YOLOv11. Модель забезпечує максимальну точність завдяки більшій кількості параметрів та складнішій архітектурі, досягаючи міри 54,7 у mAP 50□95. Цей варіант підходить для складніших завдань і рішень на великих наборах даних.

Загалом на Рисунку 2.6 можна побачити порівняння всіх розмірів моделі для завдання виявлення.

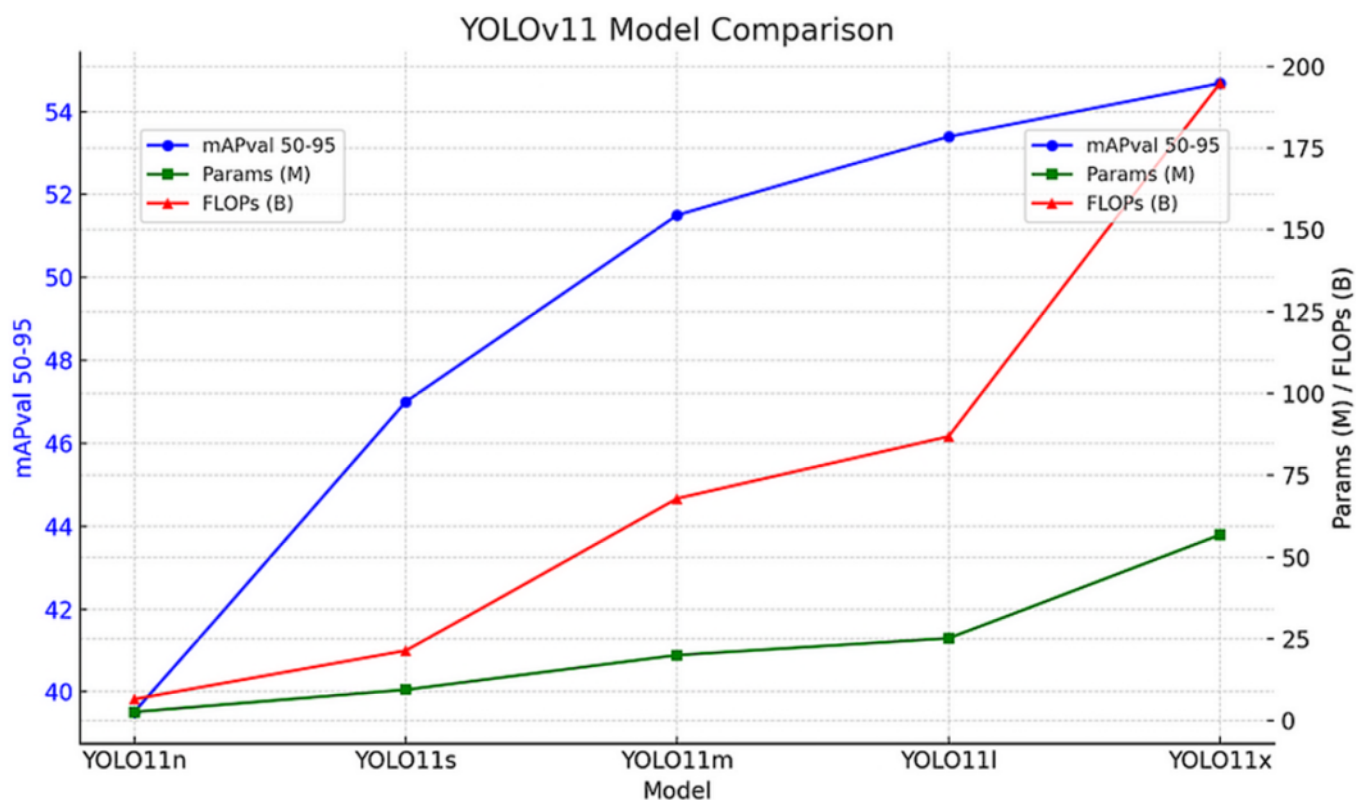


Рисунок 2.6 Порівняння розмірів моделі YOLOv11 по точності та параметрам [40]

Ключові особливості порівняння моделей:

- моделі більшого розміру мають кращі результати у вимірі точності, але потребують більше ресурсів;
- версія nano оптимізована для середовищ з обмеженими обчислювальними можливостями;
- в залежності від задачі (реальний час, обробка складних зображень, обмежені ресурси), вибирають відповідну модель.

Ці різні розміри дозволяють адаптувати YOLOv11 до різних завдань, включаючи багато сфер застосування, де потрібні швидка обробка та висока точність.

2.4 Тренування моделі

Донавчання (fine-tuning) — це процес додаткового навчання попередньо навчених моделей на специфічному наборі даних для досягнення кращих

результатів у конкретних задачах. Цей метод дозволяє використовувати знання, отримані під час навчання на великих загальних наборах даних, і адаптувати модель до конкретної предметної області, зменшуючи при цьому час і обчислювальні витрати на тренування.

Донавчання моделі YOLOv11 на специфічному наборі даних дозволяє адаптувати архітектуру до задач виявлення об'єктів у конкретному відеопотоці, зображеннях або конкретному середовищі. Завдяки цьому можна досягти вищої точності, враховуючи характеристики об'єктів і фонові особливості, що притаманні даному завданню.

Перед донавчанням необхідно підготувати відповідний набір даних та вибрати оптимальні попередньо треновані ваги YOLOv11(n,s,m,l,x). Підготовка включає наступні кроки:

- збір зображень або відеопотоків, які будуть використовуватися для донавчання. Дані повинні містити об'єкти, які потрібно виявляти;
- кожне зображення повинно бути анотоване, тобто об'єкти повинні бути позначені прямокутниками (bounding boxes) з відповідними мітками класів. Для цього можуть бути використані інструменти, такі як LabelImg або CVAT.

В YOLOv11YOLOv також є можливість нможлалаштування пових значень Ov11. IoU та Conf. Порогові значення можуть бути налаштовані користувачем залежно від завдань, метрик оцінки та пріоритетів:

- для задач з низькою ціною помилки низький поріг довірчої ймовірності та IoU дозволяє виявити максимальну кількість об'єктів, навіть якщо деякі з них є хибними;
- для задач з високою ціною помилки встановлення високих порогів гарантує, що модель виводить лише високоякісні передбачення, зменшуючи кількість хибних позитивних спрацьовувань.

В цілому, тренуваавчання YOLOv11 є важливим і складним етапом процесу вдосконалення в системах комп'ютерного зору, які будутьбуду призначені для виявлення об'єктів на в потоках, чиокдрах та статичних зображеннях. YOLOv11 □ це одна з новітніх версій моделі сімейства You Only Look Once (YOLO), що

використовуються для швидкого та точного виявлення об'єктів у реальному часі. Навчання такої моделі передбачає серію методичних кроків, щоб підвищити її продуктивність, точність та надійність у процесі виконання завдання виявлення об'єктів.

Процес самого навчання моделі YOLOv11 можна розділити на кілька ключових етапів.

Прямий прохід є першим етапом процесу навчання нейронної мережі YOLOv11. На цьому етапі зображення подається як вхідний сигнал до мережі, де воно проходить через усі шари нейронної мережі, обробляючись відповідно до архітектури моделі. Мережа на основі параметрів, що містяться у вагових матрицях, генерує передбачення щодо об'єктів на зображенні. Результати передбачень включають ймовірність наявності об'єкта, його клас та координати обмежувальної рамки (bounding box). Ці прогнози визначаються шляхом обчислення функціональних перетворень між вхідними даними, вагами та функціями активації кожного шару нейронної мережі.

Після завершення прямого проходу модель отримує передбачення, які порівнюються зі справжніми мітками (істинними значеннями) об'єктів на зображенні. На основі цього порівняння обчислюється функція втрат, яка вимірює ступінь помилки між передбаченнями мережі та справжніми значеннями. Функція втрат є критичним компонентом під час навчання, оскільки вона дозволяє оцінити, наскільки добре модель виконала завдання. Чим менше значення функції втрат, тим кращими є прогнози моделі. Основні функції втрат, що використовуються в задачах об'єктного детектування, враховують помилки в координатах обмежувальних рамок, ймовірності класифікації та інші аспекти.

Після обчислення функції втрат мережа переходить до етапу зворотного поширення помилки (backpropagation). Цей процес є основою для навчання нейронної мережі, оскільки він дозволяє мережі адаптувати свої параметри (ваги) для зменшення помилки на наступних ітераціях. Зворотний прохід включає обчислення градієнтів функції втрат відносно всіх ваг мережі за допомогою методу

диференціювання. Ці градієнти показують, як саме потрібно змінити ваги для мінімізації помилки, що виникла під час навчання.

Оновлення ваг є заключним етапом у процесі навчання. Після обчислення градієнтів, оптимізатор здійснює коригування ваг нейронної мережі. Найпопулярніші методи оптимізації включають AdamW та SGD (Stochastic Gradient Descent). Оптимізатор аналізує обчислені градієнти та на їх основі оновлює ваги з метою мінімізації значення функції втрат. Це дозволяє мережі адаптуватися до навчальних даних, покращуючи якість передбачень на кожному наступному етапі. Оптимізатор регулює кроки оновлення ваг, щоб забезпечити ефективний та стабільний процес навчання, уникнувши при цьому перенавчання або нестабільності під час процесу.

Таким чином, навчання та оновлення ваг YOLOv11 є циклічним процесом, що включає прямий прохід, обчислення втрат, зворотний прохід та оновлення ваг. Цей процес дозволяє мережі навчатися на великій кількості зображень, покращуючи здатність до виявлення об'єктів у реальних умовах.

2.4.1 Функції втрат

У моделі YOLOv11 використовується складна комбінована функція втрат, яка поєднує кілька ключових елементів, необхідних для оптимального навчання. Ці елементи дозволяють моделі враховувати як класифікацію об'єктів, так і точність їх локалізації (виявлення обмежувальних рамок). Загалом, функція втрат включає три основні складові: Focal Loss, Loss for bounding box regression (SIOU Loss) та DFL (Distribution Focal Loss).

Однак особливо важливим у контексті задач класифікації є Focal Loss, яка спрямована на коригування впливу об'єктів, які важко класифікуються, та зменшення впливу легких прикладів.

Focal Loss □ це функція втрат, яка модифікує стандартний Cross□Entropy Loss, щоб зменшити домінуючий вплив легких прикладів. Завдання Focal Loss □

сфокусувати навчання на складніших прикладах, які модель класифікує з помилками.

Застосування Focal Loss особливо важливо в ситуаціях, коли класифікація включає багато прикладів, які легко розрізнити, але деякі випадки є більш складними для навчання.

Focal Loss обчислюється за такою формулою:

$$L_{Focal} = -\alpha(1-p_t)^\gamma \times \log(p_t) \quad (1)$$

Де:

p_t □ ймовірність правильної класифікації (для відповідного класу).

α □ коефіцієнт ваги, що надає додаткову вагу важливим об'єктам та балансує вплив між класами.

γ □ параметр фокусування, що регулює, наскільки сильно модель сфокусується на складних прикладах.

Focal Loss використовується для боротьби з дисбалансом класів, коли в даних багато прикладів, які моделі легко класифікувати, а також складних прикладів, які вимагають більшої уваги.

Уявімо задачу виявлення об'єктів у відеокадрах, де більшість об'єктів □ це легко класифіковані автомобілі, але є деякі складні випадки, такі як мало видимі пішоходи або об'єкти, які частково перекриті.

– Звичайний Cross□Entropy Loss може надмірно фокусуватися на легко розпізнаваних автомобілях.

– Focal Loss дозволить моделі краще фокусуватися на складніших випадках, наприклад, на пішоходах, які з'являються в кадрі лише частково.

Focal Loss □ це потужний інструмент для оптимізації класифікації в задачах з великою кількістю легких прикладів або дисбалансом класів. Завдяки цьому, модель YOLOv11 краще навчається на складних, мало представлених прикладах, що робить її більш ефективною у реальних умовах застосування.

У моделі YOLOv11 функція втрат для регресії обмежувальних рамок (Box_loss) реалізована за допомогою методу SIoU Loss (Scaled Intersection over Union). Цей метод є ключовим інструментом у навчанні мережі для точної локалізації об'єктів у зображенні, визначаючи координати обмежувальних рамок, що описують місце розташування об'єкта.

Основні принципи роботи SIoU Loss. SIoU Loss орієнтується на обчислення IoU (Intersection over Union) між передбачуваною та істинною обмежувальною рамкою, що дозволяє моделі вимірювати, наскільки добре передбачена рамка збігається з реальною. IoU є ключовим показником у задачах регресії обмежувальних рамок, оскільки він кількісно вимірює перетин між об'єктами у двовимірному просторі. Формула IoU обчислюється як відношення площі перекриття між двома рамками до площі їхнього об'єднання, та виглядає наступним чином:

$$IoU = \frac{Interaction\ Area}{Union\ Area} \quad (2)$$

Крім обчислення значення IoU, SIoU Loss включає й інші геометричні аспекти, що роблять цей метод більш ефективним у порівнянні з базовими методами, такими як CIoU Loss (Complete Intersection over Union). До них можна віднести наступні параметри:

- врахування відстані між геометричними центрами обмежувальних рамок дозволяє моделі коректніше адаптуватися до різних позицій об'єктів на зображенні;
- узгодження орієнтації рамок (кутів між передбачуваною та істинною рамкою) дозволяє враховувати не тільки розмір, а й орієнтацію об'єкта. Це важливо для складніших ситуацій, коли об'єкти розташовані під різними кутами.

SIoU Loss має кілька важливих переваг, які забезпечують високу ефективність та швидке навчання моделі YOLOv11:

- завдяки врахуванню не лише значень IoU, а й геометричних характеристик обмежувальних рамок (відстані та орієнтації), SIoU забезпечує більш

швидке досягнення оптимального значення функції втрат під час навчання. Це дозволяє моделі швидше навчатися на даних, що зменшує час, необхідний для досягнення високої ефективності;

- у порівнянні з CIoU Loss, SIoU демонструє вищу стабільність, що дозволяє моделі мінімізувати помилки більш надійно та уникати коливань під час оптимізації. Це особливо важливо для задач, де дані можуть мати значну варіативність;

- завдяки комбінованому врахуванню IoU, відстані між центрами рамок та їхніх орієнтацій, SIoU дозволяє точно визначати місце розташування об'єкта на зображенні. Це дає можливість моделі ефективніше локалізувати об'єкти у різних масштабах та умовах.

Під час навчання YOLOv11, SIoU Loss обчислюється між передбаченою обмежувальною рамкою та істинною рамкою. Функція втрат враховує такі аспекти:

- визначає перетин об'єкта між передбачуваною та цільовою рамкою, виражений як відношення площі перетину до площі об'єднання двох рамок;

- міряє відстань між геометричними центрами двох рамок. Чим менша ця відстань, тим точніша локалізація;

- оцінює, наскільки орієнтації обмежувальних рамок збігаються, що дозволяє моделі краще враховувати об'єкти, які можуть бути нахилені або частково перекриті.

Таким чином, функція втрат SIoU дозволяє YOLOv11 швидко та стабільно навчатися, покращуючи локалізацію об'єктів на зображеннях і підвищуючи загальну точність моделі.

Distribution Focal Loss (DFL) є однією з найбільш значущих інновацій, представлених у моделі YOLOv11, та спрямована на підвищення точності регресії координат об'єктів на зображенні. Основною особливістю DFL є те, що вона прогнозує розподіли ймовірностей, замість того щоб передбачати лише окремі значення координат. Цей підхід дозволяє моделі більш точно визначати місце розташування об'єкта, а також підвищує стійкість до шуму в даних та складні

ситуації, коли об'єкти можуть мати невелике переміщення або нестандартний розмір.

DFL дозволяє більш гнучко працювати з варіативністю координат об'єкта на зображенні, оскільки кожне значення координати не є фіксованим числом, а розподіляється у певному діапазоні на основі ймовірностей. Це дає змогу моделі враховувати невпевненість у локалізації об'єкта, що значно покращує якість прогнозів, особливо в ситуаціях з високим рівнем шуму або складними умовами освітлення.

Формула для обчислення DFL виглядає наступним чином:

$$L_{DFL} = -\sum p_k \times \log(p_k) \quad (3)$$

де:

p_k □ ймовірність для кожного дискретного значення координати.

\log □ натуральний логарифм.

Ця формула є крос-ентропією між розподілом ймовірностей, який передбачає модель, та істинним значенням координати об'єкта. Ідея полягає в тому, що модель не просто передбачає одне конкретне значення, а створює розподіл ймовірностей на основі діапазону можливих координат. Це дає їй змогу враховувати варіативність у значеннях координат об'єкта та адаптуватися до складніших сценаріїв, коли об'єкти можуть бути неповністю або частково перекриті іншими об'єктами на зображенні.

Використання DFL у моделі YOLOv11 має кілька важливих переваг, які значно покращують ефективність у задачах виявлення об'єктів:

- завдяки розподілу ймовірностей замість однозначного значення, DFL дозволяє моделі точніше визначати місцезнаходження об'єкта. Модель може більш ефективно обробляти невпевненість і мінімізувати помилки у випадках, коли об'єкти займають неоднозначні позиції на зображенні;

– у ситуаціях, коли об'єкти можуть перекриватися, мати нестандартний розмір або коли зображення містять шум, DFL забезпечує більш стійке навчання. Завдяки ймовірнісному прогнозу модель краще впорається з такими умовами;

– модель, що використовує DFL, може краще обробляти випадкові зміни в даних (наприклад, збої у виділенні об'єкта через зміну умов освітлення або інших факторів), оскільки ймовірнісний підхід дає можливість зробити менш жорсткі прогнози;

– завдяки тому, що DFL прогнозує ймовірності, замість фіксованих координат, модель має змогу враховувати невпевненість у передбаченнях. Це дозволяє моделі більш точно оцінювати складні ситуації та робити прогнози з урахуванням потенційних помилок.

DFL працює шляхом створення ймовірнісного розподілу для кожного значення координат. У процесі навчання модель аналізує розподіли ймовірностей і порівнює їх з істинними значеннями координат об'єкта. Важливо підкреслити, що це відрізняється від класичного підходу, де регресія використовує лише одне найкраще значення для передбачення координат.

Прикладом може бути ситуація, коли об'єкт частково перекритий іншими об'єктами, що робить його точне локалізування складним. У такому випадку модель через DFL генерує ймовірнісний розподіл, що дає змогу моделі враховувати варіативність у значеннях координат і робити більш ефективні прогнози.

Distribution Focal Loss (DFL) є важливим інструментом у моделі YOLOv11, що дозволяє підвищити ефективність регресії координат об'єкта на зображенні. Застосування DFL забезпечує підвищену точність локалізації, покращену стійкість до шуму та адаптацію до складних сценаріїв завдяки використанню ймовірнісного підходу. Ця функція втрат демонструє високу ефективність при навчанні моделі, дозволяючи їй краще орієнтуватися у складних умовах, а також підвищує здатність до узагальнення на нові дані.

2.4.2 Регуляризація та моніторинг

Для уникнення перенавчання в методах машинного навчання та нейронних мережах активно застосовуються техніки регуляризації. Ці методи дозволяють обмежити складність моделі, забезпечуючи її ефективну здатність до узагальнення на нові дані, а не лише до запам'ятовування тренувальних зразків. Однією з поширених технік регуляризації є $L1$ та $L2$ регуляризація, які додають штрафи до функції втрат на основі значень вагів у моделі. Також широко використовуються методи dropout, що випадковим чином вимикають окремі нейрони під час навчання, що зменшує ймовірність надмірної залежності моделі від окремих ознак.

Іншим важливим аспектом, пов'язаним із забезпеченням високої якості навчання та уникненням перенавчання, є моніторинг прогресу моделі. Під час навчального процесу важливо регулярно відстежувати значення таких метрик, як точність, функція втрат та інші показники, наприклад, mAP (mean Average Precision), на валідаційному наборі даних. Це дозволяє оцінити, як добре модель навчається на даних, які не використовуються безпосередньо для тренування, і визначити, чи не відбувається перенавчання. Моніторинг прогресу також дозволяє вчасно зупинити навчання, якщо спостерігається значний розрив між точністю на навчальних та валідаційних даних.

Після завершення навчання модель повинна бути оцінена на тестовому наборі даних, який вона не бачила під час навчання та валідації. Це дає змогу визначити, наскільки добре модель здатна узагальнювати отримані знання на невідомі дані, що є основною метою застосування методів машинного навчання. Під час тестування зазвичай обчислюються метрики точності, такі як mAP та IoU (Intersection over Union). Ці метрики дозволяють аналізувати, наскільки точно модель виявляє об'єкти, а також визначають оптимальний поріг для прийняття рішень про те, чи вважається об'єкт знайденим.

На основі результатів тестування та за необхідності модель можна адаптувати до конкретних задач або специфічних умов. Адаптація може включати додаткове навчання на вузькоспеціалізованих наборах даних, що містять конкретні типи об'єктів, а також налаштування гіперпараметрів (наприклад, швидкості навчання або розміру пакету даних), щоб досягти кращих результатів у відповідних

умовах. Такий підхід дозволяє підвищити ефективність моделі, зменшуючи ризик помилок та підвищуючи її узагальнювальну здатність.

2.5 Критерії оцінювання точності нейронної мережі

Оцінка точності та швидкості роботи донавченої моделі є головним етапом в процесі розробки і впровадження моделі для реального застосування. Точність визначає, наскільки ефективно модель виконує своє завдання – виявляє та класифікує об'єкти у відеопотоці, а швидкість роботи є критично важливою для застосувань у реальному часі, де затримки можуть бути неприємні до користувача або системи.

Для оцінки ефективності моделі зазвичай використовуються кілька ключових метрик, таких як:

Точність (Precision) – показує, який відсоток позитивних прогнозів моделі є правильними.

Формула розрахунку:

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

де:

TP – кількість вірно передбачених позитивних класів (True Positives),

FP – кількість помилково класифікованих негативних класів (False Positives).

Повнота (Recall) – вказує, скільки відсотків усіх дійсних позитивних класів модель змогла правильно передбачити.

Формула:

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

де:

FN □ кількість пропущених позитивних класів (False Negatives).

$F1$ □ метрика (Рис. 3.2) □ це статистичний показник, що представляє собою середнє гармонійне між точністю (Precision) та повнотою (Recall). Ця метрика широко використовується у задачах класифікації, обробки зображень, а також у задачах детекції об'єктів, щоб дати зважену оцінку роботи моделей.

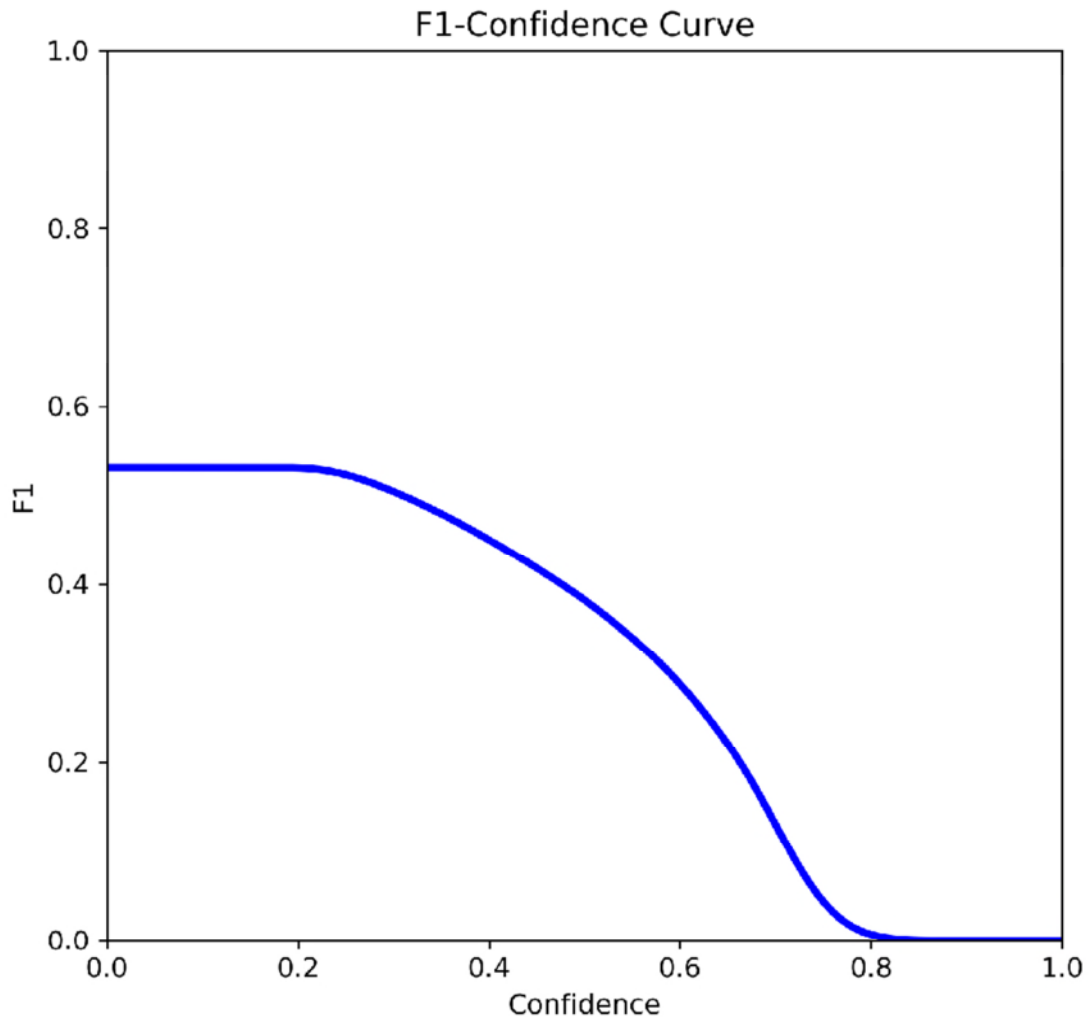


Рисунок 3.2 □ Приклад $F1$ □ метрика

Основна особливість $F1$ □ метрики полягає у її здатності обчислювати баланс між точністю та повнотою, особливо коли важливо враховувати обидва ці аспекти.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Крива Precision-Recall (Рис 3.3) є важливим інструментом для оцінювання продуктивності моделей у задачах класифікації та виявлення об'єктів, особливо коли класи є незбалансованими. Вона показує взаємозалежність між точністю (precision) та повнотою (recall) моделі, допомагаючи оцінити її здатність виявляти об'єкти при різних порогових значеннях.

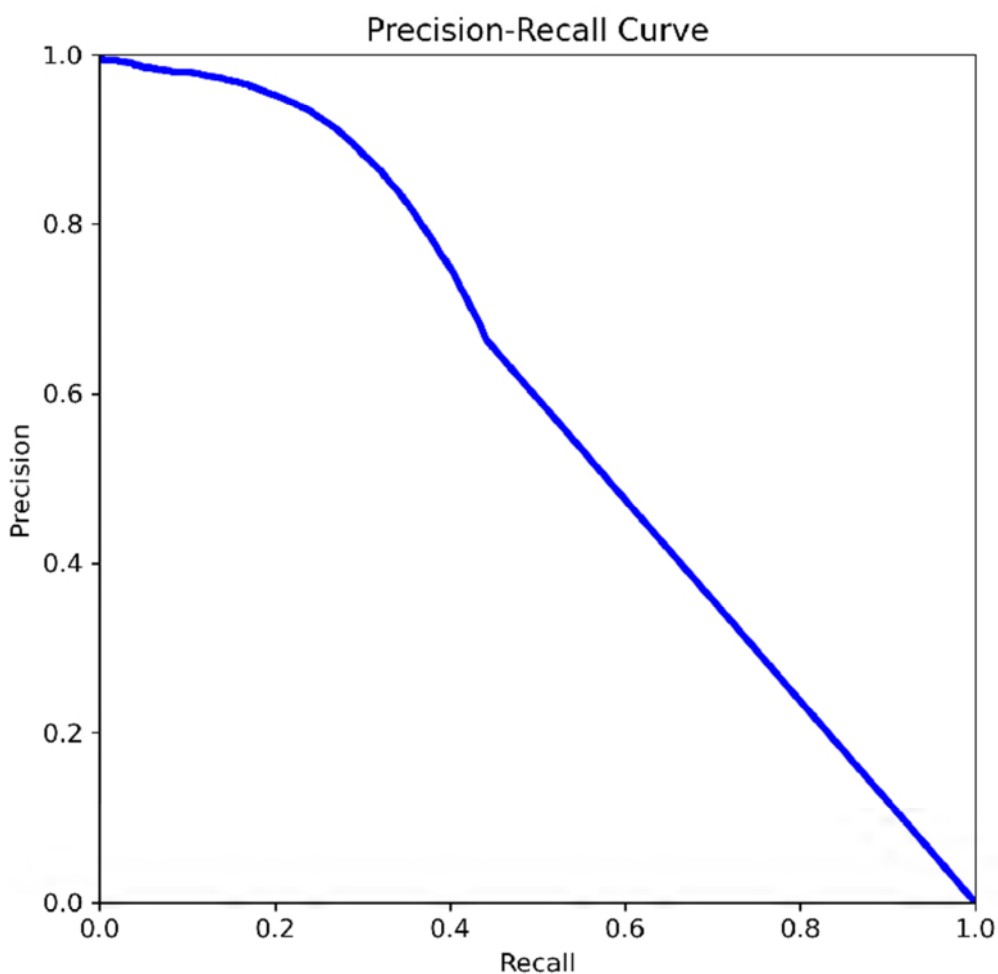


Рисунок 3.3 □ Приклад кривої Precision-Recall

Для побудови кривої Precision-Recall обчислюють точність та повноту для різних значень порогу, який може бути пов'язаний з такими показниками, як довірча ймовірність передбачення або показник IoU (Intersection over Union).

Крива демонструє, як змінюється компроміс між точністю та повнотою залежно від обраного порогу.

Висока точність за низької повноти. У такому випадку модель робить мало передбачень, але вони є високоточними. Це може свідчити про високу обережність моделі в класифікації, коли вона видає результат лише за високого рівня впевненості.

Висока повнота за низької точності. Модель намагається виявити всі можливі позитивні приклади, але при цьому допускає велику кількість помилкових позитивних передбачень.

Ідеальна модель має одночасно високу точність і повноту, і її P-R крива близька до верхнього правого кута графіка.

Вплив порогів на P-R Криву. Зміна порогового значення безпосередньо впливає на положення точки на P-R кривій.

Низькі порогові значення: Модель класифікує більшу кількість прикладів як позитивні, що підвищує повноту, але знижує точність.

Високі порогові значення: Модель класифікує лише ті приклади, для яких вона впевнена, що вони позитивні. Це підвищує точність, але знижує повноту.

У задачах виявлення об'єктів поріг часто залежить від довірчої ймовірності передбачення моделі або мінімального значення IoU, необхідного для того, щоб передбачений прямокутник вважався збігом із істинним об'єктом.

P-R крива є особливо корисною в задачах із незбалансованими класами, коли кількість негативних прикладів значно перевищує кількість позитивних. У таких умовах традиційна ROC-крива (Receiver Operating Characteristic) може показувати завищені оцінки через велику кількість хибнонегативних прикладів, тоді як P-R крива надає більш чутливу оцінку якості класифікації.

mAP (mean Average Precision) — це загальноприйнята метрика в задачах детекції об'єктів, яка характеризує якість моделі, враховуючи баланс між точністю (precision) та повнотою (recall). Вона обчислюється як середнє значення середніх точностей (AP, Average Precision) для всіх класів об'єктів, представлених у наборі

даних. Метрика mAP дозволяє оцінити не лише, наскільки добре модель виявляє об'єкти, але й наскільки точно вона визначає їхні межі.

Формула для обчислення mAP виглядає так:

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (7)$$

де:

N □ загальна кількість класів у задачі детекції,

AP_i □ середня точність для i -го класу.

Таким чином, mAP є усередненим показником, який об'єднує результати моделі по всіх класах. Він дозволяє зробити висновки про її узагальнюючу здатність, а не лише про продуктивність на окремих об'єктах.

Одним із ключових аспектів обчислення mAP є метрика IoU (Intersection over Union), яка використовується для оцінки відповідності між передбаченими межами об'єкта (bounding box) та їхніми істинними значеннями (ground truth). IoU визначається як відношення площі перетину (intersection) передбаченого та істинного обмежуючих прямокутників до площі їх об'єднання (union):

Для обчислення AP (Average Precision) і mAP, спочатку визначається, чи є передбачена межа об'єкта правильною (True Positive) на основі порогового значення IoU. Наприклад, для завдань об'єктної детекції часто використовують пороги IoU = 0.5 або 0.75, де:

$\text{IoU} \geq \text{порог}$ □ передбачення вважається коректним.

$\text{IoU} < \text{порог}$ □ передбачення вважається некоректним.

Після цього для кожного класу будуються криві PR (Precision-Recall), а AP розраховується як площа під цією кривою. Результати усереднюються по всіх класах, що й дає значення mAP.

У багатьох завданнях об'єктної детекції використовують різні порогові значення IoU для аналізу продуктивності моделі. Наприклад:

- $mAP@0.5 \square mAP$, розрахований для одного порогового значення IoU = 0.5.
- $mAP@[0.5:0.95] \square$ середнє значення mAP для IoU від 0.5 до 0.95 з кроком 0.05. Це дає більш точну оцінку продуктивності моделі на різних рівнях строгості.

Таким чином, mAP є комплексною метрикою, яка враховує як точність локалізації об'єктів, так і їхню класифікацію, забезпечуючи об'єктивний аналіз роботи моделей детекції.

Високий mAP є індикатором того, що модель демонструє відмінну продуктивність у задачах детекції об'єктів для всіх класів. Це означає, що модель здатна ефективно визначати об'єкти різних розмірів, розташованих у різних положеннях сцени, забезпечуючи як високу точність, так і правильну локалізацію об'єктів. Високий mAP свідчить про добре збалансовану роботу моделі з точки зору виявлення, класифікації та обчислення меж.

Низький mAP є тривожним сигналом, який може вказувати на кілька проблем у роботі моделі. Серед них:

- Низька точність (Precision). Висока кількість хибнопозитивних спрацьовувань (False Positives) є однією з основних причин низького mAP. Це свідчить про те, що модель часто ідентифікує об'єкти, яких насправді немає, або неправильно класифікує знайдені об'єкти.

- Низька повнота (Recall). Велика кількість хибнонегативних спрацьовувань (False Negatives) також може призводити до низького mAP. У таких випадках модель пропускає багато істинних об'єктів, які вона повинна була б виявити.

- Погана локалізація об'єктів (низький IoU). Навіть якщо модель правильно визначає об'єкти, але межі (bounding boxes) неточно відповідають справжнім положенням об'єктів, це призводить до зниження IoU. Внаслідок цього модель може не досягти порогових значень IoU, що також впливає на mAP.

Для перевірки точності моделі на тестових даних використовують Confusion Matrix (Рис 3.4). Це таблиця, яка використовується для оцінки продуктивності

моделей класифікації. Вона показує, як модель передбачає класи у порівнянні зі справжніми значеннями.

Модель може бути перевірена на конкретних реальних відео, що містять різні сцени з різними варіаціями наговпу та освітлення. Це дає змогу оцінити, наскільки добре модель працює у практичних умовах.

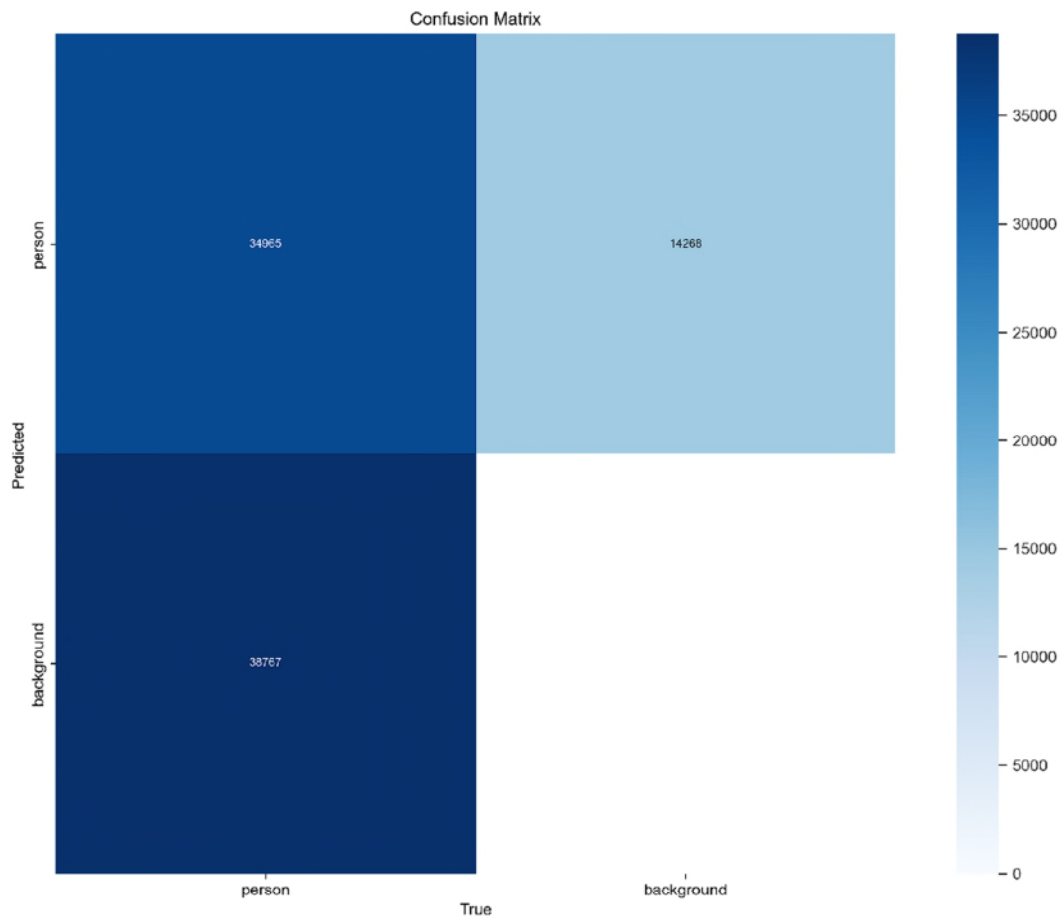


Рисунок 3.4 □ Приклад метрики Confusion matrix

Оскільки задача полягає у виявленні кількості об'єктів у відеопотоці, дуже важливо, щоб модель могла працювати в режимі реального часу. Швидкість роботи оцінюється за такими показниками:

Час обробки одного кадру □ визначає, скільки часу модель витрачає на обробку одного відеокадру. Це важливо для відео з високою частотою кадрів.

Час обробки на систему \square може вимірюватися час виконання на конкретному апаратному забезпеченні (CPU, GPU). Моделі, оптимізовані для використання на GPU (наприклад, з використанням бібліотеки TensorRT), зазвичай демонструють вищі показники швидкості.

Для оцінки реальної швидкості роботи модель можна тестувати на реальних відео або потоках з різними параметрами \square роздільною здатністю відео, кількістю об'єктів у кадрі тощо. У реальному застосуванні важливо забезпечити час відповіді на кожен кадр, який повинен бути мінімальним, щоб модель могла працювати без затримок.

Якщо модель показує низьку швидкість обробки, можна застосувати такі стратегії оптимізації:

Применшення роздільної здатності вхідних зображень або кадрів відео, що дозволяє знизити обсяг обчислень.

Приведення моделі до меншої кількості параметрів без значних втрат точності (наприклад, через методи pruning або quantization).

Оцінка точності та швидкості моделі є важливим етапом, який дозволяє зрозуміти, наскільки ефективно модель виконує завдання виявлення об'єктів у відеопотоці. Висока точність забезпечує правильну детекцію об'єктів, а оптимізована швидкість роботи гарантує, що модель може працювати в режимі реального часу. Оцінка цих показників дозволяє здійснити необхідні коригування та підвищити ефективність моделі для конкретних застосувань.

Висновки до розділу 2

У цьому розділі був розроблений метод для визначенні кількості людей у натовпі у відеопотоці, що покращується інтеграцією алгоритму DeepSORT. Це дозволяє значно покращити можливість обробки відеоматеріалу та відстеження об'єктів у режимі реального часу. Особлива увага була приділена архітектурі моделі YOLOv11, що поєднує високу точність та швидкість обробки. Основні інновації,

такі як блоки C2PSA та C3k2, зробили модель більш ефективною в роботі з різними типами зображень та сценаріями завдань.

Аналіз різних розмірів моделі YOLOv11 підкреслив їхню універсальність і можливість застосування до широкого спектру завдань комп'ютерного зору. Окрім цього, було розглянуто методи навчання, що включають цілі, функцію втрат, оновлення ваг критерії оцінювання точності, техніки регуляризації, а також моніторинг прогресу моделі під час навчання. Ці аспекти забезпечують стабільність, високу точність та ефективність роботи моделі.

Таким чином, результати дослідження підтверджують, що інтеграція алгоритму DeepSORT з моделлю YOLOv11, а також впровадження сучасних технік навчання, дозволять досягти високих показників точності в задачі визначення кількості людей у натовпі у відеопотоці.

РОЗДІЛ 3

Програмна реалізація методу визначення людей у натовпі у відеопотоці нейромережевими засобами

3.1 Засоби розробки та тестування

Для реалізації запропонованого методу виявлення об'єктів у відеопотоці було обрано мову програмування Python. Вона є одним із найпопулярніших інструментів у сфері розробки нейронних мереж завдяки своїй простоті, універсальності та підтримці сучасних бібліотек. Python дозволяє швидко створювати ефективні рішення для задач комп'ютерного зору, обробки відео та інших напрямів штучного інтелекту. Її гнучкість забезпечує можливість інтеграції різних модулів і налаштування системи відповідно до конкретних потреб проєкту.

Однією з ключових переваг Python є його простий та зрозумілий синтаксис. Це значно скорочує час розробки, знижує ймовірність помилок у коді та полегшує роботу навіть для менш досвідчених програмістів. Python підтримує об'єктно-орієнтоване програмування, що дозволяє створювати структурований код, легко керувати складністю системи та використовувати концепції, такі як спадкування, інкапсуляція та поліморфізм.

Додатковою перевагою є величезний набір бібліотек, які надає Python. Для задач нейромережевого навчання широко застосовується бібліотека PyTorch, що забезпечує повний функціонал для роботи з глибоким навчанням [41]. Для роботи із зображеннями та відеопотоками використовується OpenCV [42], а для числових обчислень і аналізу даних — WandB [43]. Ці інструменти дозволяють легко організувати процес розробки, від попередньої обробки даних до візуалізації результатів.

Python також відомий своєю мультиплатформеністю. Програми, створені цією мовою, можуть запускатися на різних операційних системах, таких як Windows, macOS та Linux. Це забезпечує зручність і універсальність у виборі середовища для розробки [44].

Важливою перевагою Python є його активна та велика спільнота. Завдяки цьому доступні численні навчальні матеріали, документація та готові приклади, які спрощують процес роботи з мовою. Розробники можуть легко знайти відповіді на питання, отримати поради чи поділитися своїми ідеями.

Загалом, Python є вдалим вибором для реалізації методу, додатково можна використати такі бібліотеки:

Ultralytics □ це популярний набір інструментів для навчання та застосування моделей, які використовують архітектуру YOLO для задач детекції об'єктів у зображеннях і відео. Цей інструмент надає зручний функціонал для створення, тренування та валідації моделей, які можна застосовувати в реальному часі в задачах комп'ютерного зору [45].

Ultralytics забезпечує підтримку наступних функціональних можливостей:

- навчання моделей YOLO з різними параметрами;
- можливість завантаження попередньо навчених моделей або створення моделей з нуля;
- гнучке налаштування параметрів тренування через YAML конфіги;
- інтеграція з платформою WandB для моніторингу метрик навчання;
- оптимізація процесу виявлення за допомогою різних технік, таких як data augmentation, multi-scale training, early stopping, тощо;
- підтримка навчання для різних архітектур, таких як YOLOv5, YOLOv8, YOLOv11.

Matplotlib □ бібліотека для побудови графіків. Використовується для візуалізації метрик навчання та відображення результатів детекції на окремих зображеннях [46].

TensorBoard використовується для моніторингу процесу навчання. Вона дозволяє відстежувати метрики, такі як втрати, точність і зміни ваг моделі. Це допомагає аналізувати продуктивність моделі на різних етапах навчання, щоб ефективніше налаштувати параметри тренування [47].

DeepSORT Realtime є високоефективною бібліотекою, призначеною для трекінгу об'єктів у відеопотоках у реальному часі. Вона базується на методах асоціативного трекінгу, які застосовують вектори-ембедінги для ідентифікації об'єктів між окремими кадрами відеопотоку. Ця бібліотека широко застосовується в задачах, пов'язаних зі спостереженням за рухом об'єктів, у таких галузях, як відеоспостереження, робототехніка та системи автоматичного моніторингу.

DeepSORT Realtime оптимізована для роботи у швидкому режимі, що дозволяє інтегрувати її в реальні застосування без значних затримок. Основується на методах, розроблених в алгоритмі DeepSORT, які поєднують систему трекінгу з використанням моделей, що забезпечують інтерпретацію руху об'єктів у відеопотоках [48].

Scipy □ це універсальна бібліотека математичних функціональних інструментів, що широко застосовується в наукових розробках та інженерних розрахунках. Вона включає функції для інтеграції, оптимізації, статистичного аналізу, обробки сигналів та вирішення завдань лінійної алгебри. Завдяки своїй універсальності, Scipy активно використовується у фізичних науках, обробці даних, оптимізації алгоритмів та інших складних обчислювальних завданнях [49].

Бібліотека надає набір високоефективних функціональних методів, що дозволяють інтегрувати математичні функції, розв'язувати оптимізаційні задачі та аналізувати статистичні дані з великою швидкістю та точністю. Scipy є потужним інструментом, що забезпечує ефективний аналіз даних у багатьох галузях.

3.2 Середовище WandB

Weights & Biases (WandB) □ це сучасна платформа для відстеження, аналізу та оптимізації експериментів у машинному навчанні. Вона дозволяє розробникам, дослідникам та інженерам ефективно моніторити процеси навчання моделей (Рис 3.1), аналізувати результати, порівнювати різні конфігурації та забезпечує зручний інтерфейс для візуалізації [43].

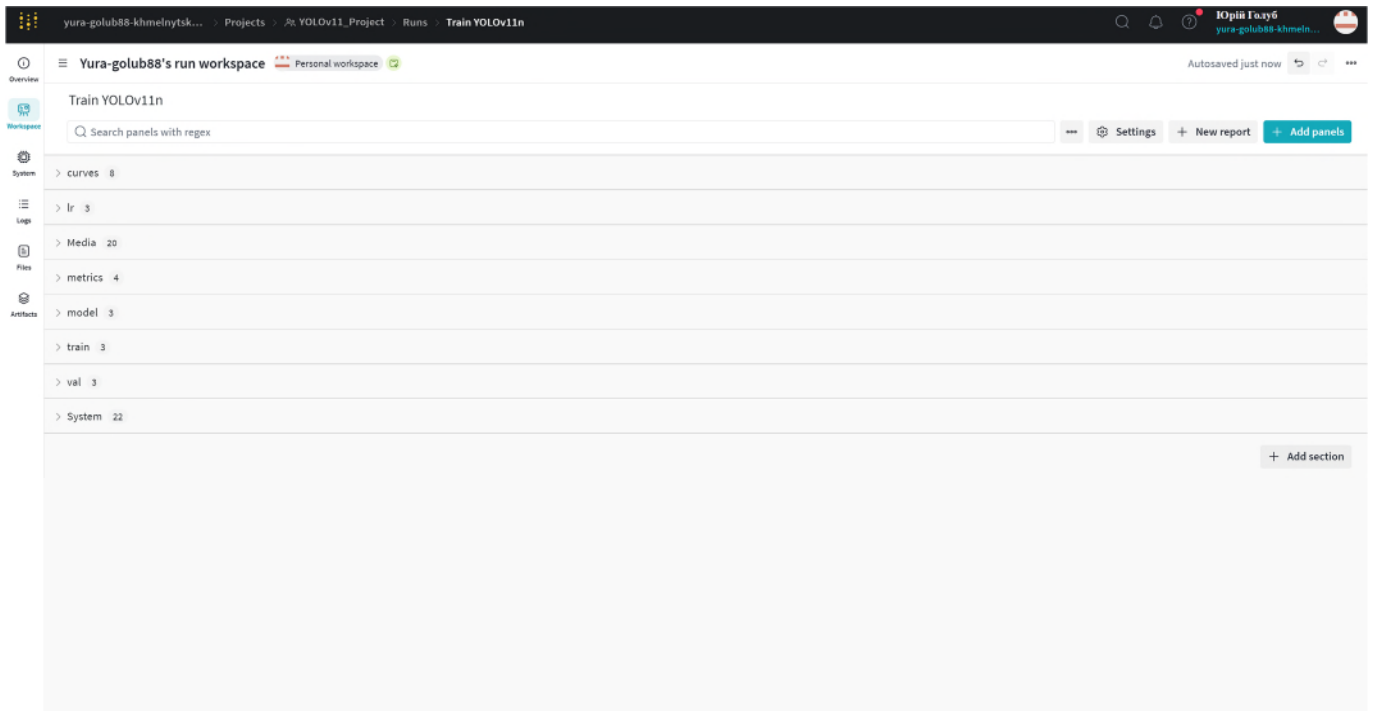


Рисунок 3.1 □ Як виглядає середовище WandB

Основні функції WandB складаються з:

- відстеження експериментів (Experiment Tracking) □ WandB дозволяє автоматично фіксувати параметри моделей, функції втрат, метрики та інші важливі дані під час навчання. Це дає змогу відстежувати, як змінюються ці показники в часі, що допомагає швидко діагностувати помилки та ефективніше налаштувати модель.
- візуалізація результатів □ платформа надає інтуїтивно зрозумілий інтерфейс для візуалізації метрик, таких як точність (accuracy), функція втрат (loss) та інші важливі показники. Це дозволяє краще розуміти процес навчання та ефективність обраної архітектури моделі.
- порівняння експериментів (Comparisons) □ WandB дає можливість порівнювати результати кількох експериментів, що допомагає визначити найкращі параметри, моделі та налаштування. Це особливо корисно, коли проводяться численні тести для вибору оптимального підходу до розв'язання задачі.
- логування конфігурацій та параметрів □ завдяки інтеграції з кодом, WandB автоматично реєструє всі параметри, що використовуються під час навчання,

включно з розміром навчального набору, швидкістю навчання, типом функції втрат тощо. Це дозволяє легко відновити умови навчання для аналізу результатів або повторного запуску моделі.

Платформа WandB забезпечує інструменти для спільної роботи над проєктами. Це особливо важливо у великих командах, де кілька учасників працюють над спільною моделлю чи завданням.

Інтеграція з популярними бібліотеками машинного навчання: WandB легко інтегрується з такими бібліотеками, як TensorFlow, PyTorch, Keras, Scikit-learn та іншими популярними інструментами для машинного навчання.

Використання середовища Weights & Biases (WandB) в системах машинного навчання забезпечує значну кількість переваг, які полегшують розробку моделей, моніторинг навчання, аналіз результатів та оптимізацію робочих процесів. Основні переваги включають:

1. Ефективний моніторинг, однією з основних переваг якого є можливість переглядати метрики, функції втрат та інші важливі параметри моделі в реальному часі під час навчання. Це дозволяє розробникам миттєво оцінювати продуктивність моделі та аналізувати зміни у функції втрат протягом кожного етапу навчального процесу. Завдяки цьому можна:

- виявляти моменти, коли модель перестає навчатися або демонструє перенавчання;
- аналізувати, як швидко модель досягає бажаних показників ефективності;
- оптимізувати параметри моделі (наприклад, швидкість навчання) відповідно до поточного стану метрик. Моніторинг у реальному часі дозволяє мінімізувати час на пошук помилок та забезпечує ефективніше управління навчальним процесом.

2. Легке налаштування та інтеграція WandB, середовище підтримує інтеграцію з найпопулярнішими бібліотеками машинного навчання, такими як PyTorch, TensorFlow, Keras, Scikit-learn та інші. Завдяки цій підтримці платформа дозволяє швидко налаштуватися для роботи з наявним кодом, не вимагаючи

складної конфігурації. Переваги моніторингу результатів тренування у середовищі WandB:

- для використання WandB достатньо кількох рядків коду, що значно скорочує час, необхідний для підключення до навчального середовища;
- можна підключати до будь-якої моделі або експерименту, незалежно від складності коду;
- завдяки простому інтерфейсу інтеграції, нові користувачі можуть швидко почати роботу з платформою, без необхідності в глибокому навчанні або додаткових налаштуваннях;

3. У WandB є функціонал збереження всіх конфігурацій, метрик, візуалізацій та результатів, що були отримані під час навчальних експериментів. Ця функціональність розкривається в наступному:

- кожен експеримент зберігати в історії, що дає можливість аналізувати, як зміни у налаштуваннях впливають на кінцеві результати;
- порівнювати результати кількох експериментів, завдяки збереженій історії можна легко порівнювати різні моделі, архітектури та значення параметрів, щоб визначити, які з них є найбільш ефективними;
- якщо модель працює некоректно або потрібно повернутися до попереднього стану, можна швидко відновити будь-які параметри чи конфігурації;
- історія надає повний огляд тенденцій та закономірностей, що допомагає визначати нові напрями для оптимізації та аналізу.

4. Завдяки розширеним інструментам логування та аналізу метрик, WandB дозволяє швидко виявляти помилки, які можуть виникати в процесі навчання. Це значно скорочує час на пошук причин помилок і оптимізує навчання моделей. Конкретно, це дозволяє:

- виявляти невідповідності в даних або моделі. Помилки, пов'язані зі зміною параметрів, некоректними даними чи неправильним налаштуванням, можна швидко діагностувати завдяки логуванню метрик;

- якщо модель показує високу ефективність на навчальних даних, але погано працює на валідаційних наборах, можна своєчасно виявити перенавчання та вжити відповідних заходів;
- WandB дозволяє моніторити, як функція втрат змінюється під час навчання, що допомагає виявити проблеми, такі як занадто висока швидкість навчання або нестабільний процес;
- за допомогою логування можна швидко перевіряти, як зміни у параметрах (наприклад, швидкість навчання, розмір пакету даних) впливають на кінцеві результати.

Таким чином, Weights & Biases (WandB) є потужним та простим інструментом, що забезпечує глибокий аналіз, зручний моніторинг та швидку діагностику в процесах машинного навчання. Завдяки функціоналу, що включає моніторинг в реальному часі, легку інтеграцію, збереження історії експериментів та діагностику помилок, платформа допомагає розробникам швидше знаходити оптимальні рішення, покращувати моделі та проводити детальний аналіз на всіх етапах розробки.

Висновки до розділу 3

У цьому розділі розглянуто основні засоби, що використовуються для розробки та тестування моделі YOLOv11. Основною мовою програмування обрано Python завдяки його простоті, універсальності та широкій підтримці потужних бібліотек, таких як PyTorch, OpenCV, WandB та Ultralytics. Ці інструменти забезпечують швидку інтеграцію, ефективний аналіз даних, візуалізацію та підтримку архітектур, таких як YOLOv11.

Використання сучасних бібліотек і методів оптимізації сприяє досягненню високої ефективності на всіх етапах розробки та донавчання.

РОЗДІЛ 4

Дослідження методу визначення людей у натовпі та порівняння результатів

4.1 Налаштування параметрів навчання нейронної мережі

Для тренування було обрано оптимальний розмір моделі, а саме YOLOv11s, що забезпечує баланс між швидкістю обробки даних та ефективністю виявлення об'єктів. Модель YOLOv11s характеризується високою швидкодією та достатньою потужністю для розв'язання завдання з виявлення людей у натовпі. Для забезпечення збалансованого та ефективного тренування було використано crowd counting dataset □ набір даних, що містить зображення натовпів із різною кількістю людей у різних умовах [50].

Оптимізація параметрів моделі є ключовим етапом тренування для досягнення високої продуктивності. Параметри впливають на різні аспекти роботи моделі, зокрема на швидкість навчання, здатність до узагальнення, ефективність обробки даних та якість виявлення об'єктів. Нижче наведено огляд найбільш важливих параметрів, що впливають на процес тренування моделі YOLOv11s.

1. Базові параметри навчання:

- `epochs` (кількість епох) □ визначає кількість проходів моделі через весь навчальний набір даних. У цьому прикладі встановлено 150 епох, що є достатнім для глибоких моделей, але потребує моніторингу, щоб уникнути перенавчання;
- `batch` (розмір батчу) □ кількість зображень, що обробляються одночасно. Установлення малого розміру (2) знижує вимоги до оперативної пам'яті, але може уповільнити навчання;
- `optimizer` (оптимізатор) □ використання стохастичного градієнтного спуску (SGD) забезпечує стабільне оновлення ваг моделі, хоча може бути менш швидким порівняно з сучасними варіантами, такими як AdamW;

– `device` □ визначає, на якому пристрої проводиться тренування. Установлення 0 означає використання першого GPU, що забезпечує високу продуктивність у порівнянні з CPU.

2. Параметри навчання і регуляризації:

– `lr0` (початкове значення швидкості навчання) □ початкова швидкість навчання встановлена на рівні 0.01. Це стандартне значення, яке підходить для більшості завдань, однак може потребувати адаптації залежно від специфіки даних.

– `warmup_epochs` (розігрів) □ визначає кількість початкових епох, протягом яких швидкість навчання поступово збільшується. Це допомагає стабілізувати навчання на ранніх етапах;

– `weight_decay` (регуляризація ваг) □ значення 0.0005 використовується для зменшення переваг великих ваг, що сприяє зменшенню перенавчання;

– `patience` (терпимість) □ цей параметр (20) визначає кількість епох без покращення метрики до зупинки навчання. Це сприяє збереженню ресурсів і запобігає зайвому тренуванню.

3. Обробка та аугментація даних:

– `imgsz` (розмір зображення) □ встановлено 640 пікселів, що є стандартним компромісом між якістю детекції та швидкістю обчислень;

– `augment` (аугментація) □ додавання аугментації (True) дозволяє моделі навчатися на різноманітних варіаціях даних, підвищуючи її узагальнюючу здатність;

– `flipr`, `flipud` (дзеркальні відображення) □ значення 0.5 для горизонтального та 0.0 для вертикального відображення вказує, що горизонтальні зміни застосовуються з високою ймовірністю, що є корисним для більшості наборів даних;

– `translate` (трансляція) □ встановлене значення 0.1 дозволяє невеликі зміщення об'єктів у межах зображення, зберігаючи контекст об'єкта;

– `mosaic`, значення 1.0 вказує на використання мозаїчної аугментації, яка поєднує кілька зображень в одному, покращуючи здатність моделі працювати з об'єктами на різному масштабі.

4. Метрики та обмеження:

- `iou` (Intersection over Union) встановлене значення 0.2 визначає мінімальний поріг для відповідності передбачених і реальних рамок об'єктів. Це впливає на точність обчислення метрик, таких як mAP;
- `conf` (поріг впевненості), значення 0.2 використовується для фільтрації низьковпевнених передбачень, забезпечуючи баланс між точністю та повнотою;
- `max_det` (максимальна кількість об'єктів) параметр обмежує кількість об'єктів, які можуть бути виявлені на одному зображенні. Значення 1000 дозволяє враховувати складні сцени.

5. Додаткові налаштування

- `single_cls` (єдиний клас) включення цього параметра (True) дозволяє моделі навчатися на даних з одним класом, що зменшує складність задачі;
- `agnostic_nms` (агностичний NMS) використання агностичного NMS (True) дозволяє моделі видаляти дублікати об'єктів незалежно від їх класу;

Правильний вибір параметрів моделі має вирішальний вплив на її продуктивність. Поєднання основних налаштувань, таких як швидкість навчання, аугментація даних та регуляризація, разом із спеціалізованими метриками та обмеженнями, забезпечує оптимальне навчання і точність детекції об'єктів.

4.2 Характеристика датасету

Характеристики датасету є одним із найважливіших факторів, що впливають на точність моделей машинного навчання, особливо в задачах детекції, класифікації та інших областях застосування штучного інтелекту. Залежно від якості, кількості, різноманітності та збалансованості даних, які використовуються для тренування, модель може мати різний рівень ефективності. Для донавчання моделі було використано датасет `crowd counting dataset` [50].

Цей датасет був створений для досліджень у сфері підрахунку натовпів із використанням моделей YOLO. Мета його використання полягає у тому, щоб забезпечити різноманітні дані для тренування та тестування моделей у завданнях виявлення та підрахунку людей у складних середовищах.

Датасет складається з трьох частин (Рис. 4.): навчальна вибірка (79%, 2285 зображень), валідаційна вибірка (13%, 382 зображення) та тестова вибірка (8%, 231 зображення).

Підготовлений набір даних розділений на тренувальні, валідаційні та тестові дані (Рис.4.1).

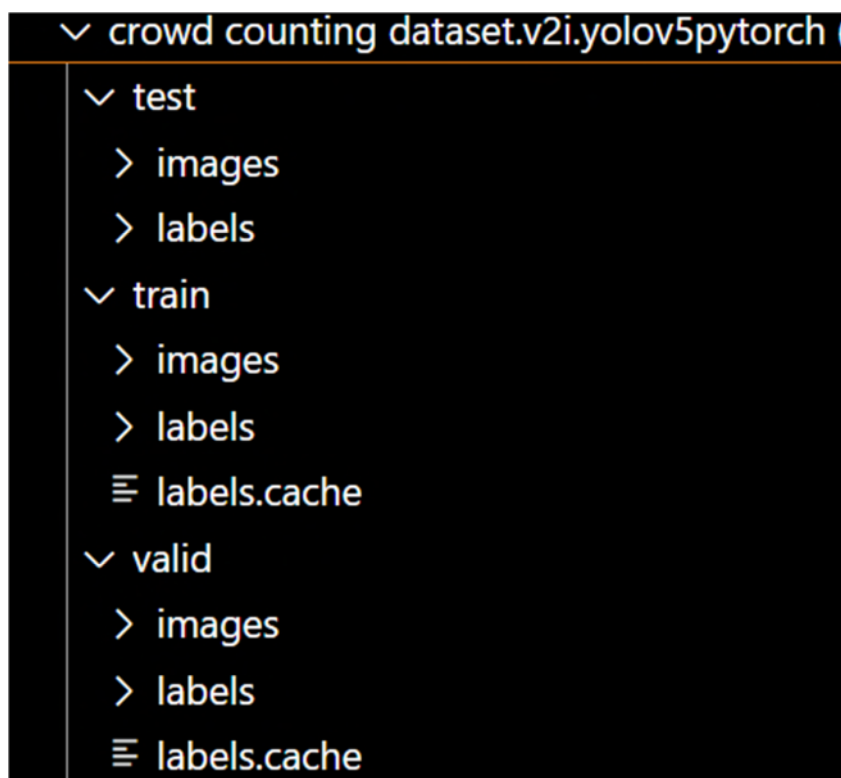


Рисунок 4.1 □ Структура папки з датасетом

Усі зображення у датасеті мають стандартний розмір 640x640 пікселів, де акцент на виявлення кількості людей зроблено на головах, для отримання кращих результатів у щільних натовпах.

Батчі тренування (Рис.4.2) □ це невеликі групи даних, які використовуються для навчання моделі під час кожного кроку оптимізації. У контексті тренування моделей, таких як YOLO, батчі розділяють весь датасет на менші частини, які модель обробляє поетапно замість навчання на всьому датасеті одночасно.

Батч містить певну кількість зображень, які обробляються разом на одному кроці навчання. Наприклад, розмір батчу $batch=2$ означає, що модель за один раз обробляє дві картки.

Після обробки всіх зображень у батчі, обчислюється середнє значення втрат (loss), і модель оптимізує свої ваги на основі цих даних. Це дозволяє ефективніше навчати модель, ніж обробляти кожне зображення окремо.

Кількість кроків (steps) за епоху залежить від розміру датасету та розміру батчу. Наприклад:

Якщо є 2285 зображень у навчальній вибірці, і розмір батчу складає 2, то кількість кроків на епоху буде 1143 кроки за епоху. Батчі часто обробляються паралельно через декілька ядер процесора чи GPU, щоб пришвидшити навчання.



Рисунок 4.2 □ Тренувальний батч

Батч валідації □ це невеликі групи даних, які використовуються під час тренування моделей для перевірки їх ефективності (Рис.4.3). Ці дані відокремлені від навчального набору, тобто модель не бачить їх під час навчання. Основна функція батчів валідації □ перевірити, як добре модель працює на невідомих даних, та виміряти метрики, такі як точність (accuracy), повнота (recall) і точність детекції. Це дозволяє зрозуміти, чи модель перенавчається на навчальних даних.

Батчі валідації працюють наступним чином: валідаційний набір розділяється на менші групи (батчі), і кожен із них обробляється окремо. Під час обробки кожного батчу обчислюються функції втрат і метрики, які показують, як модель працює з цими даними. Потім після обробки всіх батчів метрики обчислюються в середньому, щоб отримати загальну оцінку ефективності моделі.

Батчі валідації дуже важливі, щоб уникати перенавчання. Перенавчання □ це ситуація, коли модель добре вивчає лише навчальні дані, але погано працює з новими, невідомими даними. Батчі валідації дозволяють визначати, чи модель працює стабільно та чи добре узагальнює знання на нові приклади.

Розмір батчу валідації може відрізнятись від розміру навчального батчу. Іноді його збільшують, щоб отримати більш точну оцінку метрик. Під час кожного циклу тренування метрики з батчів валідації використовуються для прийняття рішень, наприклад, про те, чи зберігати поточну модель, чи зменшити значення функції втрат.



Рисунок 4.3 □ Батч валідації

4.3 Налаштування алгоритму трекінгу DeepSORT

Алгоритм DeepSORT (Deep Simple Online and Realtime Tracking) є потужним інструментом для трекінгу об'єктів у відеопотоках. У поєднанні з детектором об'єктів YOLOv11, DeepSORT дозволяє ефективно вирішувати завдання підрахунку кількості людей у кадрі завдяки інтеграції даних детекції та трекінгу. Ключовими параметрами, що визначають роботу DeepSORT, є:

- `max_age=10`. Цей параметр визначає максимальну кількість кадрів, протягом яких трек може залишатися активним без нових асоційованих детекцій. Значення 10 забезпечує стійкість до короточасних втрат об'єкта з поля зору, наприклад, при часткових перекриттях або швидкому русі. Водночас, це значення запобігає надмірному продовженню "мертвих" треків, що знижує ризик помилкових підрахунків.

– `n_init=2`. Цей параметр задає кількість детекцій, необхідних для ініціалізації нового треку. Значення 2 забезпечує баланс між швидкістю виявлення нових об'єктів і надійністю, зменшуючи ризик хибних треків через шум у даних. Особливо це корисно для динамічних сцен із великою кількістю людей, де поодинокі хибні детекції можуть спотворювати результати.

– `nn_budget=30`. Параметр визначає обмеження на кількість історичних ознак (features) для кожного треку. Значення 30 забезпечує достатньо інформації для відстеження об'єкта, водночас знижуючи обчислювальну складність. Це налаштування оптимальне для сцен із помірною кількістю об'єктів, оскільки дозволяє ефективно використовувати ресурси.

Для інтеграції DeepSORT із YOLOv11 використовуються вихідні дані YOLO, включно з координатами обмежувальних рамок, класами об'єктів та впевненістю детекції. Цей процес складається з таких етапів:

Детекція об'єктів. YOLOv11 виконує детекцію людей у кожному кадрі відео. Результати детекції включають координати рамок, клас "людина" та значення впевненості.

Передача результатів до DeepSORT. Вихідні дані YOLO передаються до трекера DeepSORT, який використовує їх для ініціалізації або оновлення треків.

Трекінг об'єктів. DeepSORT зберігає унікальні ідентифікатори для кожного треку, враховуючи позицію, розмір рамки та дескриптори, створені за допомогою нейронної мережі.

Інтеграція DeepSORT із YOLOv11 помірно впливає на швидкість системи, без суттєвої шкоди для швидкодії. YOLOv11 вже забезпечує високошвидкісну детекцію об'єктів, а DeepSORT додає до цього обчислювальні витрати, пов'язані з трекінгом і обробкою дескрипторів об'єктів. У реальному часі продуктивність може залежати від кількості об'єктів у кадрі, але за налаштувань, таких як `max_age=10`, `n_init=2` та `nn_budget=30`, обчислювальні витрати залишаються помірними і прийнятними для сучасного апаратного забезпечення.

Таку систему цілком можна інтегрувати навіть у відеоспостереження, особливо якщо використовуються GPU або інші апаратні прискорювачі. Вона буде

ефективною для підрахунку людей у кадрі, аналізу натовпів або відстеження поведінки в режимі реального часу. Однак для сценаріїв із надвеликою кількістю об'єктів може знадобитися додаткова оптимізація або кластеризація обробки даних.

4.4 Отримання результатів загальної точності методу визначення людей у натовпі

4.4.1 Середнє значення точності моделі

Метрика mAP@50 (mean Average Precision at IoU=0.5) використовується для вимірювання ефективності моделей у задачах детекції об'єктів. Вона вимірює, як добре модель може локалізувати об'єкти на зображеннях та правильно їх класифікувати, враховуючи лише ті випадки, коли IoU (Intersection over Union) між передбачуваними рамками та істинними обмежувальними рамками перевищує 0,5. Цей показник дозволяє визначати точність моделі на основі її здатності знаходити об'єкти та локалізувати їх у межах встановленого порогу.

Коли метрика mAP@50 досягає значення 59,2% після 150 епох тренування (Рис.4.4), це вказує на те, що модель навчилася добре класифікувати об'єкти та визначати їх координати у відповідності до заданих критеріїв. Досягнення такого результату за 150 епох означає ефективний процес оптимізації, добре підібрані параметри навчання, а також що оброблений датасет містить достатньо даних для навчання.

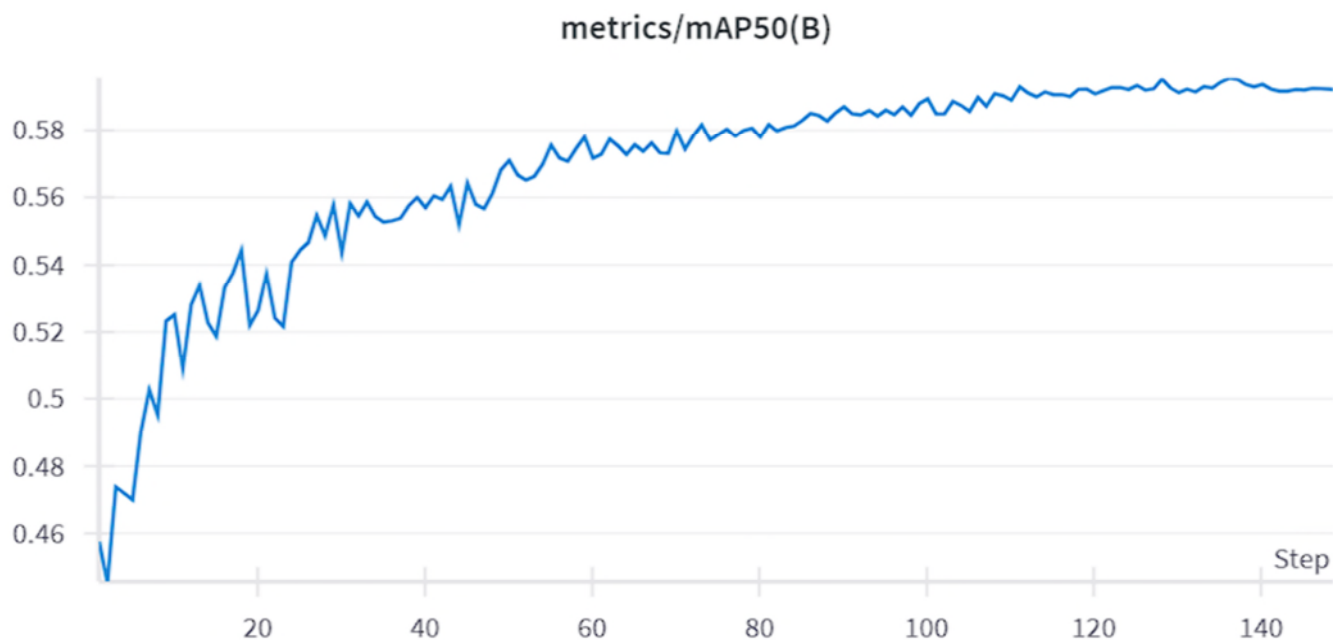


Рисунок 4.4 □ Метрика точності mAP50 під час навчання

Така метрика може досягатися завдяки декільком ключовим факторам:

- оптимізатор SGD, Adam або інші оптимізатори разом із правильною функцією втрат допомагають мінімізувати помилки у прогнозах;
- якщо дані для навчання якісні, різноманітні та мають коректно позначені мітки, модель тренується швидше та досягає вищих показників;
- батчі розміром 2×4 разом із використанням AMP прискорюють навчання та допомагають досягати більш стабільного оптимізаційного процесу;
- архітектура YOLOv11 оптимізована для швидкого навчання, що дозволяє досягати високої точності в розпізнаванні об'єктів навіть за обмежений період тренування;

В підсумку, досягнення метрики $mAP@50 = 59,2\%$ вказує на те, що модель на етапі навчання ефективно оптимізувалася і добре сприймає закономірності у навчальних даних, що дозволяє їй успішно детектувати об'єкти з достатньою точністю.

Метрика $mAP@50 \square 95$ (mean Average Precision at IoU thresholds from 0.50 to 0.95) є ключовою метрикою для оцінки ефективності моделей детекції об'єктів, що вимірює точність моделі на основі координат обмежувальних рамок та класової

предикції. Ця метрика обчислюється через середнє значення середніх точностей при різних значеннях IoU (Intersection over Union) від 0.50 до 0.95. Чим вищі ці значення, тим точніша модель у визначенні об'єктів та їхніх обмежувальних рамок.

Досягнення значення 26,9% для $mAP@50\text{-}95$ у завданні підрахунку натовпу за 150 епох (Рис.4.5) тренування свідчить про те, що модель навчилася ефективно визначати об'єкти (людей у натовпі), навіть у складних умовах, коли відстані, перекриття та варіативність у зображеннях можуть впливати на точність. До такого результату можуть призводити коректно підібраний навчальний датасет, правильне налаштування функцій втрат, оптимізація навчання та ефективна архітектура нейронної мережі.

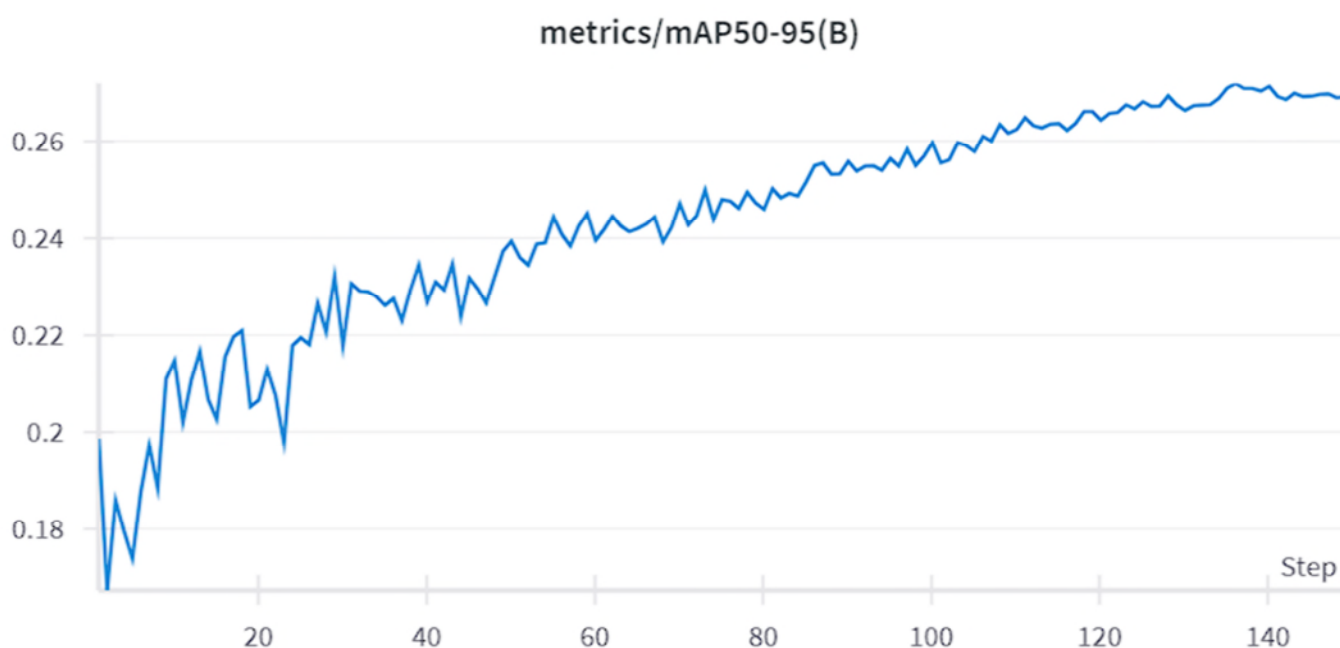


Рисунок 4.5 □ Метрика точності $mAP50\text{-}95$

Оцінка на основі map-метрики дозволяє з'ясувати, наскільки модель може узагальнювати навички на невідомі зображення, а також мінімізувати помилки, пов'язані з неправильною локалізацією об'єктів чи класовою класифікацією. Результат 26,9% є значущим для завдання підрахунку людей у натовпі, оскільки такий тип завдання зазвичай характеризується великою кількістю об'єктів, перекриттями, зміною перспективи та відстані між об'єктами.

Цей показник метрики дозволяє визначити ступінь ефективності навчання, а також спланувати подальші кроки, такі як розширення датасету, корекція функцій втрат, збільшення розмірів навчального батчу або оптимізація архітектури моделі.

4.4.2 Точність і повнота моделі

Метрика precision (точність) є однією з основних метрик для оцінки якості моделей детекції об'єктів, яка вимірює співвідношення між кількістю правильно класифікованих об'єктів та кількістю всіх об'єктів, які модель класифікувала як позитивні. За результатами тренування моделі, метрика precision досягла значення 72,3% після 150 епох навчання (Рис. 4.6). Це свідчить про те, що модель ефективно виявляла об'єкти та класувала їх із високою точністю, досягаючи відносно високих показників правильної класифікації. Однак протягом всього процесу навчання показник precision був нестабільним. Це означає, що його значення змінювалося протягом окремих епох, не маючи чіткого стабільного рівня.

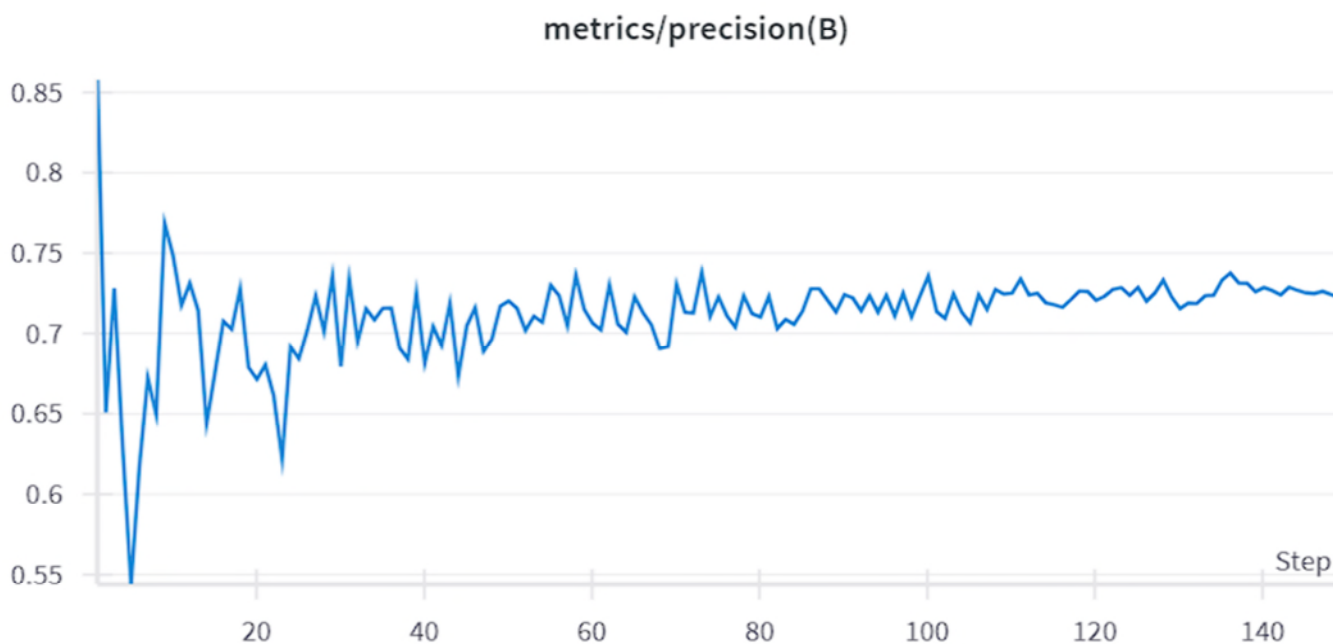


Рисунок 4.6 □ Метрика точності precision

Нестабільність метрики може бути спричинена різними факторами, такими як варіативність у даних, зміна розподілу даних у навчальних батчах, недостатня кількість тренувальних прикладів, помилки в процесі оптимізації чи гіперпараметри, що потребують додаткового налаштування. Проте після досягнення приблизно 72,3%, метрика трималася близько цього значення, що вказує на те, що модель змогла досягти певного рівня стабільності у своїх передбаченнях.

Результат 72,3% precision дозволяє стверджувати про те, що модель є ефективною у виявленні об'єктів у більшості випадків.

Метрика recall (повнота) є ключовою метрикою для оцінки якості моделей в задачах детекції об'єктів. Вона вимірює здатність моделі виявляти всі позитивні об'єкти, які повинні бути виявлені, незалежно від помилок класифікації.

Після 150 епох тренування модель досягла значення 41,1% (Рис. 4.7) для метрики recall. Це означає, що модель змогла виявити близько 41,1% всіх об'єктів, які були присутні у вхідних даних. Показник recall є важливим, оскільки він демонструє, наскільки повно модель здатна ідентифікувати об'єкти, навіть якщо це відбувається не завжди точно.

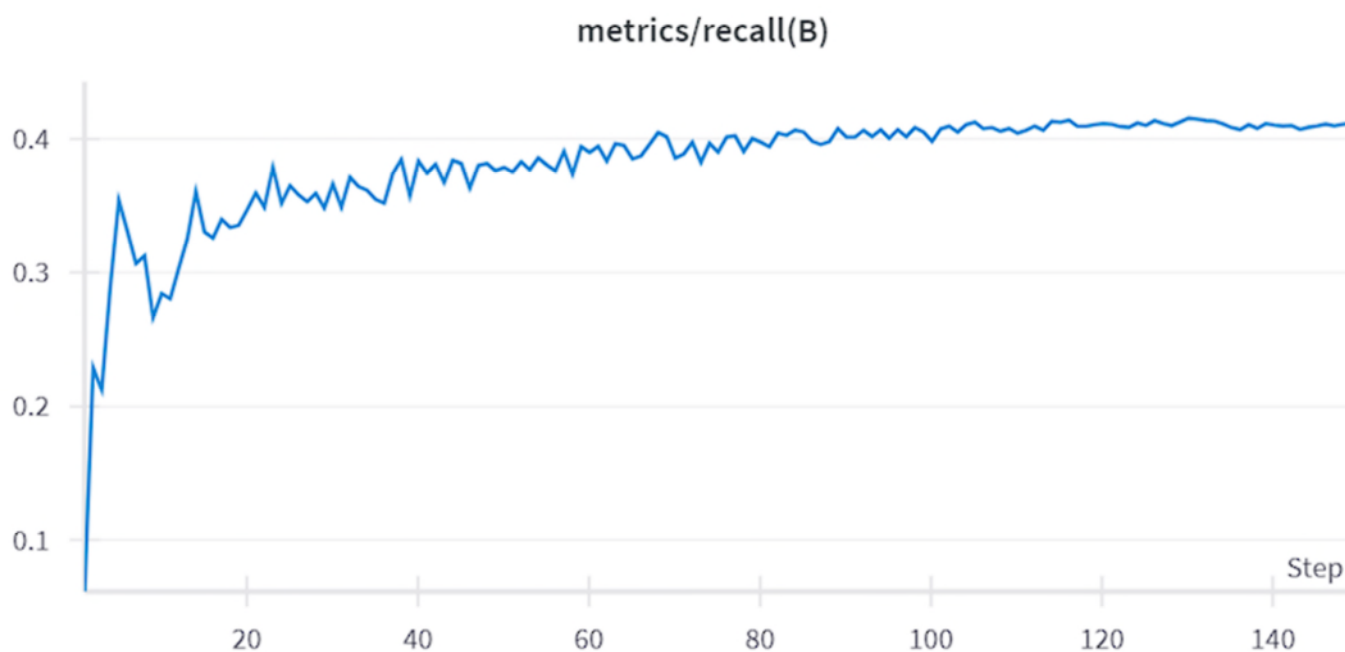


Рисунок 4.7 □ Метрика точності recall

Хоча досягнутий рівень recall відображає певний прогрес у процесі навчання, він вказує на те, що значна частина об'єктів у даних залишається невиявленою. Можливі причини такої ситуації можуть включати недостатню кількість тренувальних прикладів, складність архітектури, недостатнє налаштування параметрів моделі або високу варіативність у зображеннях, що впливає на здатність моделі правильно виявляти об'єкти.

Загалом отримане значення 41,1% recall добрий показник для цього завдання але в подальшому можлива необхідність оптимізації моделі, збільшення кількості та різноманітності тренувальних даних або налаштування гіперпараметрів з метою підвищення цього показника.

4.4.3 Втрати на валідаційних та тренувальних даних

Під час тестування моделі на валідаційному та тестовому наборах даних (Рис 4.8) було отримані значення втрат, які виявилися близькими до значень на навчальному наборі.

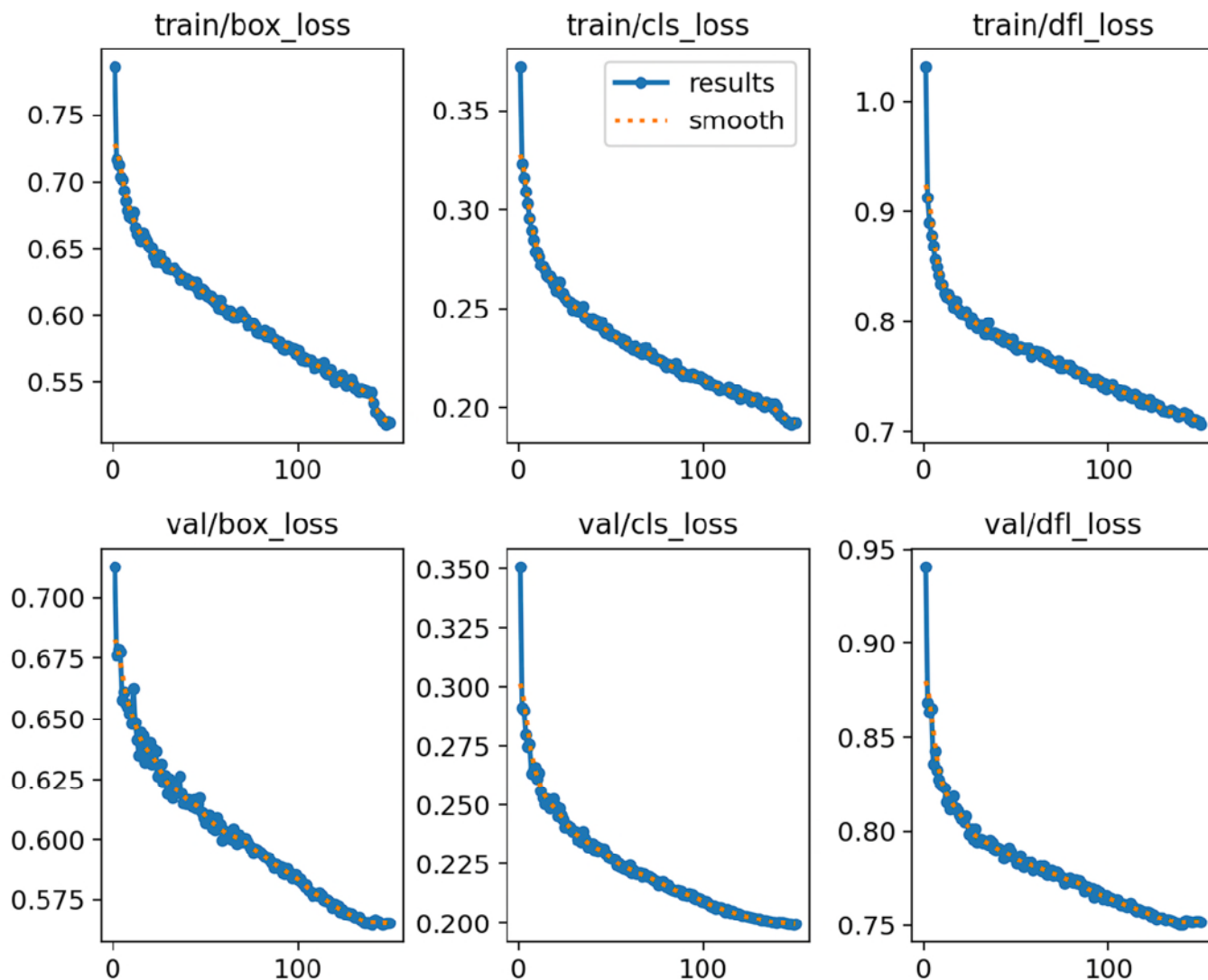


Рисунок 4.8 □ Функції втрат при навчанні моделі

Ці показники вказують на те, що модель не демонструє значного перенавчання (overfitting). Перенавчання характеризується суттєвим зростанням втрат на валідаційних даних у порівнянні з навчальними, що свідчить про те, що модель занадто сильно адаптувалася до специфіки навчальних даних і погано узагальнює знання на нові приклади. Однак отримані результати показують, що модель ефективно навчається, підтримуючи низькі значення втрат як на навчальному, так і на валідаційному наборах. Це підтверджує стійкість моделі та її здатність узагальнювати на невідомі дані без ознак перенавчання.

4.5 Порівняння результатів покращенох моделі з стандартною попередньо тренованою версією

Після донавчання моделі YOLOv11s на датасеті Crowd Counting Dataset та її тестування, результати якого наведені у таблиці 4.1, було проведено аналіз за основними метриками: mAP (mean Average Precision), Precision та Recall. Ці метрики дозволяють оцінити ефективність моделі з різних аспектів: mAP показує середню точність моделі на основі класу об'єктів люди, Precision вимірює частку коректно виявлених об'єктів серед усіх передбачень, а Recall показує здатність моделі ідентифікувати всі наявні об'єкти в натовпі. Результати аналізу вказують на те, наскільки модель в цілому для такого завдання ефективно навчилася узагальнювати інформацію та добре справляється з виявленням людей у різноманітних умовах.

Таблиця 4.1 Результати дотренованої моделі

Модель	Precision	Recall	mAP50	mAP50 \square 95
Моя модель	72,3%	41,1%	59,2%	26,9%

Також важливо відзначити, що швидкість моделі складає 8,87 мс на один кадр, це є важливим показником, що характеризує час, необхідний моделі для обробки одного окремого зображення або кадру з відеопотоку. Цей показник вимірює, за який час модель здійснює аналіз вхідного кадру, виявляє об'єкти (людей у натовпі) та проводить їх підрахунок. Швидкість на рівні 8,87 мс демонструє високу ефективність обчислювального процесу, що дозволяє моделі працювати майже миттєво після отримання кадру.

Така швидкість є достатньою для застосування у режимі реального часу, оскільки відеопотоки зазвичай містять велику кількість кадрів, що надходять зі швидкістю 25 \square 60 кадрів за секунду. Для забезпечення плавного функціонування та своєчасного прийняття рішень у реальних умовах швидкість моделі повинна бути меншою за час, відведений на обробку кожного кадру. У цьому випадку 8,87 мс

дозволяють моделі обробляти до 113 кадрів за секунду, що робить її цілком придатною для завдань, де потрібна швидка реакція.

Швидкість на рівні 8,87 мс підтверджує, що модель ефективно навчена та оптимізована для мінімізації обчислювальних витрат, забезпечуючи баланс у параметрах. Це дозволяє застосовувати її в реальних сценаріях, таких як моніторинг масових заходів, спостереження за натовпами, відеоаналітика для безпеки, системи автоматичного підрахунку людей та інших прикладних завданнях, де потрібна миттєва обробка даних.

На рисунку 4.9 продемонстровано, як модель виконує завдання з визначення кількості людей у відеопотоці з використанням алгоритму трекінгу DeepSORT, що забезпечує відстеження руху об'єктів та дозволяє моделі не лише покращувати виявлення кількості людей у кадрах, а й підтримувати їхні ідентифікації між окремими кадрами. Алгоритм DeepSORT працює на основі поєднання системи ідентифікації векторів-ембедінгів та методів асоціативного трекінгу, що дозволяє ефективно відстежувати кожного індивідуума протягом усієї тривалості відео. Результати, представлені на рисунку, демонструють точність і надійність алгоритму в умовах різноманітного фону, різних ракурсів та неповного перекриття об'єктів.



Рисунок 4.9 □ Визначення кількості людей у кадрі за допомогою донавченої моделі YOLOv11s

Базова модель YOLOv11s, яка тренувана на датасеті COCO, показує обмежену ефективність у завданнях виявлення людей у натовпах (Рис. 4.10) через кілька факторів. По-перше, COCO містить різноманітні об'єкти та сцени, однак він не спеціалізується на умовах натовпів, де об'єкти мають високі рівні перекриття, різні масштаби та близькість між ними. Це призводить до того, що модель може неправильно класифікувати або упустити об'єкти через складність розрізнення індивідуальних людей у натовпі.

По-друге, вхідні дані в COCO не завжди достатньо репрезентативні для ситуацій з високою щільністю людей, що також погіршує ефективність. Більшість зображень у COCO фокусуються на окремих об'єктах або добре визначених сценах, де особливі випадки, такі як зміни у масштабах об'єктів, динамічні пози, тіні та перекриття між людьми, не зустрічаються в достатній кількості. Ця обмеженість означає, що модель вчиться на менш складних випадках, а тому погано адаптується до складніших сценаріїв натовпів.

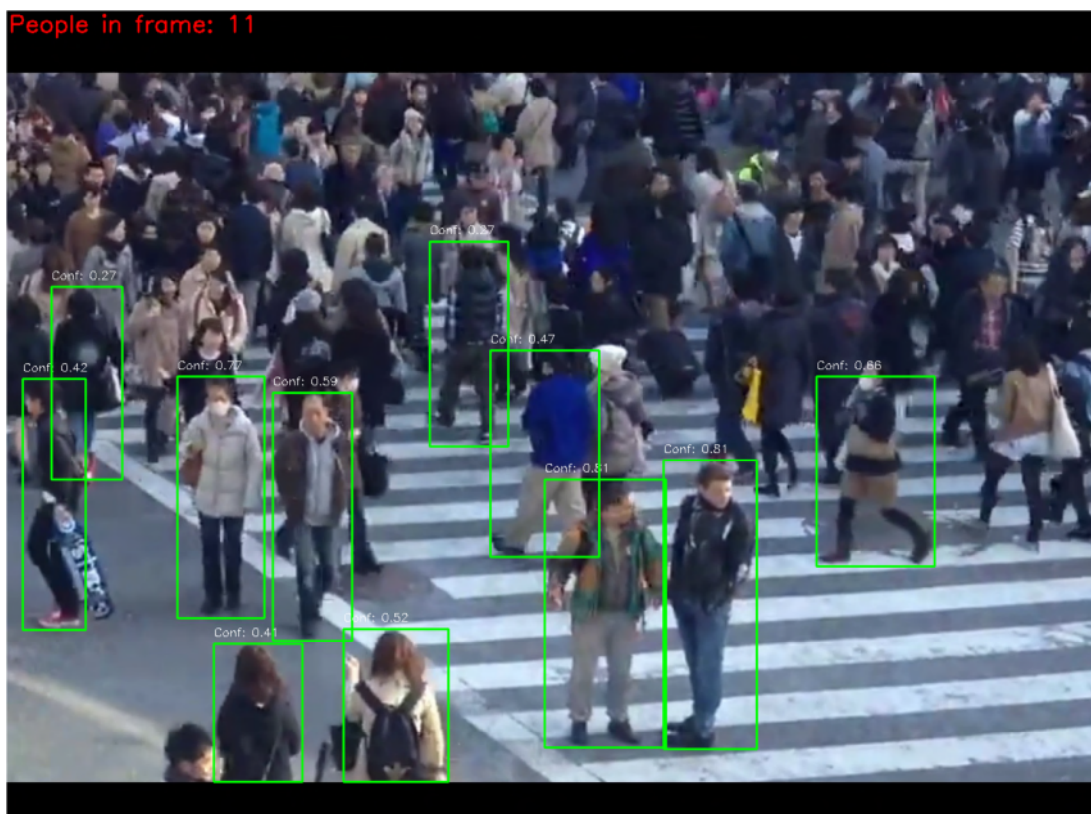


Рисунок 4.10 □ Визначення кількості людей у кадрі за допомогою стандартної моделі YOLOv11s

Крім того, методи, що застосовуються в базовій моделі YOLOv11s, не орієнтовані на детекцію багатьох схожих об'єктів із високим рівнем перекриття. Модель не має механізмів для аналізу складних взаємодій між об'єктами у натовпі, що обмежує її здатність до розпізнавання нащадкових та багат шарових варіантів, характерних для таких ситуацій.

Як можна побачити по рисунку, стандартна модель YOLOv11s, тренувана на COCO датасеті, демонструє суттєво нижчу ефективність у завданні з виявлення людей у натовпах у порівнянні з покращеною версією. Це зумовлено недостатньою адаптацією базової моделі до специфіки задачі, зокрема до високої щільності об'єктів, варіативності положень людей та складних умов зйомки. У той час як покращена модель, донавчена на спеціалізованому датасеті crowd counting dataset, базова модель демонструє обмежену здатність до точного розпізнавання об'єктів та знижену продуктивність у таких умовах. Це підтверджує важливість використання спеціалізованих підходів та адаптації моделей до конкретних задач для досягнення високої точності.

Висновки до розділу 4

У цьому розділі було розглянуто налаштування моделі YOLOv11s з алгоритмом трекінгу DeepSORT, тренування й тестування у задачі виявлення кількості людей у кадрі. Основними критеріями оцінювання ефективності моделі стали mAP (Mean Average Precision), Precision, та Recall, що дозволяють об'єктивно оцінити здатність моделі виявляти об'єкти з високою точністю, враховуючи як правильні спрацювання, так і кількість помилкових визначень.

Для досягнення оптимальних результатів, модель YOLOv11s було налаштовано з використанням специфічних гіперпараметрів, таких як значення порогової впевненості (confidence threshold) та порогу збігу (IoU threshold). Ці параметри дозволили адаптувати модель до виявлення об'єктів високої щільності та варіативності у натовпах. Поряд із цим, алгоритм трекінгу DeepSORT було

налаштовано для точного відстеження окремих людей у кадрі. Це включало оптимізацію таких параметрів, як максимальний вік треків, кількість підтверджених кадрів для активації треку та використання векторів ознак для зіставлення об'єктів.

Модель була натренована на спеціалізованому датасеті crowd counting dataset, що враховує особливості задачі виявлення людей у натовпах. Тренування на такому датасеті дозволило суттєво підвищити точність моделі у складних умовах зйомки, коли присутня велика кількість людей, які частково перекривають одне одного. Під час тестування було отримано загальну оцінку точності моделі, що підтверджує її високу здатність до виконання поставленої задачі.

Впровадження алгоритму трекінгу DeepSORT дало змогу покращити результати, забезпечивши стабільне відстеження об'єктів між кадрами. Це дозволило уникнути дублювання ідентифікації та збільшило загальну кількість правильно розпізнаних людей. Порівняння донавченої моделі з базовою версією YOLOv11s, попередньо тренуваної на датасеті COCO, продемонструвало значну перевагу донавченої моделі. Стандартна модель виявилася недостатньо точною через відсутність адаптації до задачі виявлення людей у натовпах, що підтверджує важливість налаштування і донавчання моделей для специфічних завдань.

Таким чином, виконані дослідження й оптимізації дозволили створити рішення, здатне з високою точністю виконувати задачу виявлення й підрахунку людей у кадрі, що підтверджує ефективність обраного підходу та необхідність спеціалізації моделей для складних задач комп'ютерного зору.

Загальні висновки

У цій роботі було запропоновано вдосконалений метод виявлення кількості людей у натовпі у відеопотоці за допомогою детектора YOLOv11. Метою дослідження було підвищення точності та ефективності визначення кількості людей у реальному часі, використовуючи нейромережеві технології. Для досягнення поставленої мети було проведено комплекс досліджень, що включав аналіз сучасних методів, розробку власного підходу, підготовку спеціалізованого набору даних, донавчання моделі YOLOv11, інтеграцію трекінгового алгоритму DeepSORT та валідацію отриманих результатів.

Першочерговим етапом роботи став аналіз сучасних нейромережевих моделей, які використовуються для задач визначення кількості людей у відеопотоці. Було встановлено, що базові моделі, як-от стандартний YOLO, не забезпечують достатньої точності у випадках високої щільності об'єктів у натовпі, що зумовило необхідність доопрацювання моделі через донавчання на специфічному наборі даних.

Метод, розроблений у рамках роботи, заснований на адаптації попередньо натренованої моделі YOLOv11s, донавчаної на спеціалізованих даних з датасету. Використання алгоритму DeepSORT дозволило забезпечити ефективне трекінгове відстеження об'єктів у відеопотоці, що значно підвищило точність підрахунку, мінімізуючи помилки через повторний підрахунок тих самих індивідуумів.

Отримані результати підтвердили ефективність запропонованого методу завдяки значному покращенню метрик, таких як mAP, Precision та Recall, у порівнянні з базовими методами.

Перелік посилань

1. Характеристика і приклади натовпу. Моя освіта.
URL: <https://moyaosvita.com.ua/psihologija/xarakteristika-i-prikladi-natovpu/>.
2. Психологія натовпу. Вікіпедія. URL:
https://uk.wikipedia.org/wiki/Психологія_натовпу.
3. Foro.T. BriefCam Bytes: Crowd Counting.BriefCam.
URL: <https://www.briefcam.com/resources/blog/briefcam-bytes-crowd-counting/>.
4. Матвієнко С. М. Штучні нейронні мережі – IT Master □ електроніка та програмування.Головна – IT Master □ електроніка та програмування.
URL:<https://www.itmaster.biz.ua/programming/vision/neural-networks.html>.
5. КПІ ім. Ігоря Сікорського. ШТУЧНІ НЕЙРОННІ МЕРЕЖІ: БАЗОВІ ПОЛОЖЕННЯ.
URL:<https://ela.kpi.ua/server/api/core/bitstreams/9fee52b6-83fc-4e99-8541-c2767f634c7c/content>.
6. Explained: Neural networks.MIT News | Massachusetts Institute of Technology.
URL:<https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414>.
7. MC.today. Що таке нейронна мережа? [Електронний ресурс]. 2021. URL:
<https://mc.today/uk/shho-take-nejronna-merezha/>.
8. Кушнір Н.О. Використання згорткових нейронних мереж у задачах розпізнавання та класифікації об'єктів зображень / Н.О. Кушнір, Т.М. Локтікова, А.В. Морозов, В.О. Юрченко // Технічна інженерія №1 (89), 2022. – С. 93-100.
9. Синеглазов В. Глибокі нейронні мережі для вирішення завдань розпізнавання і класифікації зображення [Електронний ресурс] / В. Синеглазов, О. Чумаченко // «Інформаційні технології та комп'ютерне моделювання» : матеріали ст. Міжнар. наук. □практ. конф., м. Івано-Франківськ, 15-20 трав. 2017 р. – Івано-Франківськ, 2017. – С. 274-277.
10. Balaji.S. Binary Image classifier CNN using TensorFlow. Medium.
URL: <https://medium.com/techiepedia/binary-image-classifier-cnn-using-tensorflow-a3f5d6746697>.

11. Krizhevsky, A., Sutskever, I., & Hinton, G.E. ImageNet Classification with CNNs. *Nature*, 521(7553), 436–444. DOI: 10.1038/nature14539.
12. Basic CNN Architecture: Explaining 5 Layers of Convolutional Neural Network.upgrad. URL:<https://www.upgrad.com/blog/basic-cnn-architecture/>.
13. Bińkowski M.B., Donnat .P., Marti G. Autoregressive Convolutional Neural Networks for Asynchronous Time Series.arxiv. URL: <https://doi.org/10.48550/arXiv.1703.04122>.
14. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. Attention is all you need. arXiv preprint arXiv:1706.03762.
15. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. arXiv preprint arXiv:1506.02640. DOI: 10.48550/arXiv.1506.02640.
16. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., & Fu, C.Y. SSD: Single Shot MultiBox Detector. *European Conference on Computer Vision (ECCV)*, 21–37. DOI: 10.1007/978-3-319-46448-0_2.
17. M. A. Ansari, A. Crampton, and S. Parkinson. A layer-wise surface deformation defect detection by convolutional neural networks in laser powder-bed fusion images. *Materials*, 15(20):7166, Oct. 2022.
18. Muhammad Hussain. When, where, and which?: Navigating the intersection of computer vision and generative ai for strategic business integration. *IEEE Access*, 11:127202–127215, 2023.
19. Sojasingarayar A. Faster R-CNN vs YOLO vs SSD–Object Detection Algorithms. *Medium*. URL: <https://medium.com/ibm-data-ai/faster-r-cnn-vs-yolo-vs-ssd-object-detection-algorithms-18badb0e02dc>.
20. Mastering Object Detection: A Comprehensive Guide to YOLO Series, Faster R-CNN, SSD Multibox, and Mask R-CNN [Электронный ресурс] // Awareeye. – 2023. – Режим доступа: <https://www.awareeye.com/post/mastering-object-detection-a-comprehensive-guide-to-yolo-series-faster-r-cnn-ssd-multibox-and-mask-r-cnn>.

21. Yuhong Li , Xiaofan Zhang , Deming Chen. University of Illinois at Urbana-Champaign. Beijing University of Posts and Telecommunications “CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes” <https://github.com/vivekbombatkar/CSRNet-pytorch>
22. KIZRAK M. A., BOLAT B. Crowd Density Estimation by Using Attention Based Capsule Network and Multi-Column CNN. IEEE Xplore. URL:<https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9433545>.
23. Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, “Single-image crowd counting via multi-column convolutional neural network,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 589–597.
24. Y. Li, X. Zhang, and D. Chen, “CSRNet: Dilated convolutional neural networks for understanding the highly congested scenes,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 1091–1100
25. Muhammad Hussain. YOLOv5, YOLOv8 and YOLOv10: The Go-To Detectors for Real-time Vision. arXiv preprint arXiv:2407.02988. 2024. C. 1–15. DOI: 10.48550/arXiv.2407.02988.
26. YOLOv11 Architecture Explained: Next-Level Object Detection with Enhanced Speed and Accuracy [Електронний ресурс]. // YOLOv11 Blog. 2024. 13 листопада. URL: <https://yolov11.wordpress.com/2024/11/13/10/>
27. Alexey Bochkovskiy, Chien-Yao Wang and Hong-Yuan Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection", arXiv preprint, 2020.
28. Redmon et al. YOLO9000: Better, Faster, Stronger. 2016. DOI: 10.48550/arXiv.1612.08242.
29. Lin et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. 2017. IEEE. DOI: 10.1109/TPAMI.2017.2699184.
30. Wang C.-Y., Bochkovskiy A., Liao H.-Y. M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. IEEE Xplore. URL: <https://ieeexplore.ieee.org/document/10204762>.

31. Khanam R., Hussain M. What is YOLOv5: A deep look into the internal features of the popular object detector. arXiv.org. URL: <https://arxiv.org/html/2407.20892v1>.
32. He, Z., Wang, K., Fang, T., Su, L., Chen, R., & Fei, X. (2024). Comprehensive Performance Evaluation of YOLOv11, YOLOv10, YOLOv9, YOLOv8 and YOLOv5 on Object Detection of Power Equipment. arXiv preprint arXiv:2411.18871.
33. Rahima Khanam, Muhammad Hussain .YOLOv11: an overview of the key architectural enhancements. October 24, 2024. [Електронний ресурс]. Режим доступу: <https://arxiv.org/pdf/2410.17725>.
34. Wojke, N., Bewley, A., & Paulus, D. (2017). Simple Online and Realtime Tracking with a Deep Association Metric. arXiv preprint arXiv:1703.07402.
35. Kouidri, A. (2023). Deep SORT: Realtime Object Tracking Guide. Ikomia. URL: <https://www.ikomia.ai/blog/deep-sort-object-tracking-guide>. [Nanonets. (2023). DeepSORT: Deep Learning to Track Custom Objects in a Video. Nanonets. URL: <https://nanonets.com/blog/object-tracking-deepsort/>
36. **Ошибка! Источник ссылки не найден.**
37. Roboflow Blog. What is an anchor box? URL: <https://blog.roboflow.com/what-is-an-anchor-box/>. Accessed: 2024-07-25.
38. Ultralytics YOLO. Supported tasks and modes. 2023. [Електронний ресурс]. URL: <https://docs.ultralytics.com/models/yolo11/#supported-tasks-and-modes>
39. COCO Dataset. URL: <https://cocodataset.org/#home>
40. Ultralytics: Next-Level Object Detection [Електронний ресурс]. // YOLOv11 Blog. 2024. 13 листопада. URL: <https://yolov11.wordpress.com/2024/11/13/10/>
41. PyTorch documentation. URL: <https://pytorch.org/>
42. OpenCV documentation. URL: <https://opencv.org/>
43. WandB documentation. (2023). Weight & Biases платформа для моніторингу машинного навчання. URL: <https://wandb.ai/>
44. Python.org. (2021). The Python Language Reference. URL: <https://www.python.org/>
45. Ultralytics documentation. (2023). YOLOv5, YOLOv8, YOLOv11. URL: <https://ultralytics.com/>

46. Matplotlib documentation. (2021). URL: <https://matplotlib.org/>
47. TensorBoard documentation. (2021). URL: <https://www.tensorflow.org/tensorboard>
48. PyPI, 2023. [deep-sort-realtime]. URL: <https://pypi.org/project/deep-sort-realtime/>
49. Scipy.org, 2023. [Scipy]. URL: <https://www.scipy.org/>
50. Crowd Counting Dataset. [Электронный ресурс]. 2023. URL: <https://universe.roboflow.com/crowd-dataset/crowd-counting-dataset-w3o7w>

Додатки

Додаток А

Світлини наукових публікацій, виконаних при роботі над кваліфікаційною роботою магістра

(ксерокопії титульної сторінки, сторінки змісту та всіх сторінок із публікацією)

Наукова публікація:

Збірник наукових праць – Визначення кількості людей в натовпі у відеопотоці нейромережевими засобами / Голуб Ю.В., Багрій Р.О., Вознюк Л.О. // Збірник наукових праць за матеріалами XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2024». Хмельницький. 2024.

Міністерство освіти і науки України
Хмельницький національний університет



ЗБІРНИК НАУКОВИХ ПРАЦЬ
за матеріалами XVI Всеукраїнської науково-практичної конференції
«Актуальні проблеми комп'ютерних наук АПКН-2024»

15-16 листопада 2024

Хмельницький 2024

УДК 004.8

Голуб Ю.В., Багрій Р.О., Вознюк Л.О.

Хмельницький національний університет

ВИЗНАЧЕННЯ КІЛЬКОСТІ ЛЮДЕЙ В НАТОВПІ У ВІДЕОПОТОЦІ НЕЙРОМЕРЕЖЕВИМИ ЗАСОБАМИ

Розглянуто сучасні моделі машинного навчання для визначення кількості людей у натовпі на основі відеопотоків. Визначено ефективну модель, яка справляється з завданням виявлення та підрахунку натовпу, зберігаючи оптимальну швидкість і точність, забезпечуючи меншу витрату ресурсів.

Modern machine learning models for determining the number of people in a crowd based on video streams have been examined. An effective model has been identified that addresses the task of detecting and counting crowds while maintaining optimal speed and accuracy, ensuring lower resource consumption.

Сучасні міста стикаються з численними викликами у сфері управління великими скупченнями людей, що виникають під час масових заходів, спортивних подій, фестивалів або у людних місцях, таких як вокзали, аеропорти, торгові центри тощо. Ці ситуації потребують не лише уважного моніторингу, а й швидкого реагування з боку відповідних служб. Традиційні методи підрахунку людей часто виявляються недостатньо ефективними через їхню трудомісткість, низьку точність або неможливість роботи в реальному часі. З огляду на цю ситуацію, необхідно розробити автоматизовані системи, що використовують алгоритми машинного навчання для точного та швидкого підрахунку людей у натовпі на основі відеопотоків.

Важливість цієї проблеми підкреслюється не лише в контексті громадської безпеки, а й у сфері досліджень поведінки людей, управління натовпами, а також у комерційних цілях, таких як аналіз поведінки споживачів. Наявність ефективного рішення для підрахунку людей у натовпі має потенціал для істотного покращення якості надання послуг, оптимізації ресурсів та зменшення ризиків у ситуаціях з великою кількістю людей. Це створює підґрунтя для застосування новітніх технологій у цій галузі, де точність, швидкість і надійність алгоритмів відіграють критичну роль у досягненні успіху [1].

Сучасні дослідження пропонують низку моделей для підрахунку об'єктів у натовпі, таких як R-CNN, MCNN, CSRNet та YOLO. Кожна модель має певні переваги та недоліки залежно від специфічних умов застосування.

R-CNN (Region-based Convolutional Neural Network) є однією з перших моделей для виявлення об'єктів, що використовує глибокі нейронні мережі. Вона розроблена для покращення точності виявлення об'єктів на зображеннях. R-CNN

спочатку генерує регіони інтересу, а потім класифікує їх, що забезпечує високу точність. Проте ця модель має суттєві недоліки, такі як висока обчислювальна вартість і затримки в обробці, що робить її малоефективною для динамічних сцен, зокрема у відеопотоках [2].

MCNN (Multi-Column Convolutional Neural Network) спеціально розроблена для адаптації до різної щільності натовпу, що дозволяє їй працювати з високою точністю на зображеннях із щільним розташуванням людей [3]. Однак її здатність до обробки відеопотоків є обмеженою.

CSRNet (Convolutional Neural Network with Dilated Convolutions) використовує розширені згортки, які збільшують поле огляду моделі без втрати просторової інформації, що робить її ефективною для високощільних натовпів на статичних зображеннях. Але CSRNet не призначена для швидкої обробки динамічних сцен [4].

YOLO (You Only Look Once) вирізняється серед згаданих моделей завдяки швидкій обробці відеопотоків та одноетапній архітектурі, яка дозволяє ефективно обробляти зображення в реальному часі. Вона підходить для завдань з високими вимогами до продуктивності завдяки здатності одночасно ідентифікувати та підраховувати об'єкти за один прохід по кадру, що суттєво відрізняє її від R-CNN, MCNN та CSRNet [5].

Метою дослідження є знаходження найбільш придатної до підрахунку людей у натовпі моделі, яка поєднує в собі оптимальну точність та швидкість в обробці даних, що необхідні для роботи з відеопотоками.

Під час огляду було визначено, що модель YOLOv5 є більш придатною для завдань з обробкою відеопотоку завдяки своїй оптимізованій структурі та алгоритмам обробки даних, які забезпечують високу продуктивність із мінімальними втратами точності. Це робить YOLOv5 особливо ефективною для застосувань, що потребують швидкого підрахунку об'єктів у реальному часі, як-от відеоспостереження або автономне водіння, де важлива обробка динамічних сцен із мінімальними затримками.

Архітектура моделі складається з трьох компонентів, таких як: магістраль (Backbone), шия (Neck) та голова (Head) для здійснення щільних прогнозів (Dense predictions), як показано на рисунку нижче (рисунок 1).

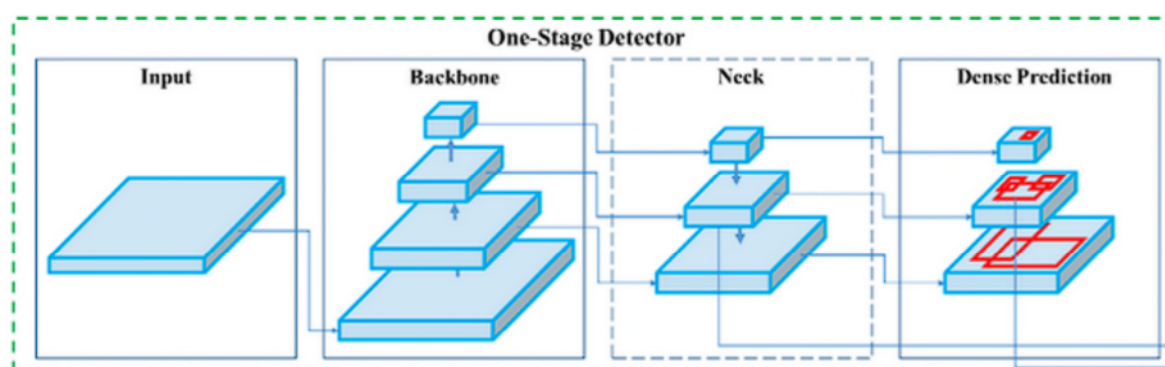


Рисунок 1 – Архітектура YOLOv5 [6]

Backbone — це попередньо навчена мережа, яка використовується для отримання широкого представлення функцій для зображень. Це сприяє зменшенню просторової роздільної здатності зображення та збільшенню його функціональної (каналної) роздільної здатності.

Neck використовується для вилучення пірамід особливостей. Це допомагає моделі добре узагальнювати на об'єкти різних розмірів і масштабів.

Head використовується для виконання операцій завершального етапу. Він застосовує якірні блоки на картах об'єктів і обробляє кінцеві виводи: класи, оцінки об'єктів та обмежувальні рамки.

На рисунку 2 показано схему методу визначення кількості людей у натовпі за допомогою моделі YOLOv5, що включає такі кроки:

1. Підготовка до обробки відеоматеріалу – кадри з відеоматеріалу завантажуються в модель, масштабуються та нормалізуються до розмірів, відповідних для обробки YOLOv5.
2. Початкові згорткові шари – відбувається виділення базових ознак (лінії, краю, текстури) з кадрів.
3. Шари підвибірки та виділення ознак – шари підвибірки (max-pooling) зменшують розміри карт ознак, залишаючи лише найбільш значущі характеристики.
4. Глибокі згорткові шари – модель обробляє більше узагальнених ознак, таких як форма та розташування об'єктів.
5. Шари прогнозування обмежувальних рамок – модель визначає ймовірність належності кожного об'єкта до класу “людина”.

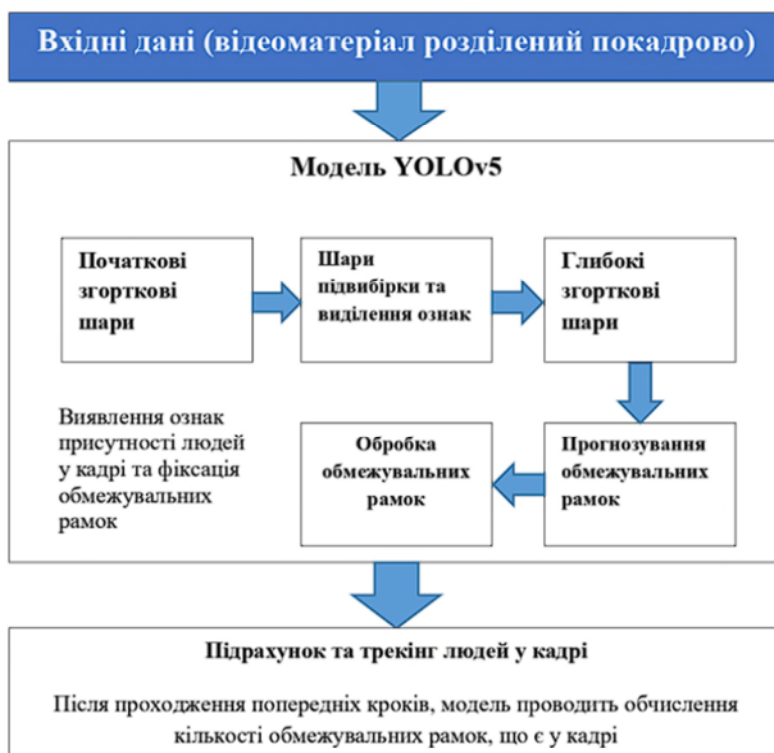


Рисунок 2 – Метод визначення кількості людей в натовпі у відеопотоці

6. Обробка обмежувальних рамок – застосовується NMS алгоритм для видалення повторюваних рамок та вибору найвірогідніших для кожного об'єкта.

7. Підрахунок та трекінг – модель підраховує кількість людей у кадрі та зберігає ідентифікацію кожного об'єкта в динаміці.

8. Результат обробки – отримані дані (кількість людей та їх позиції) відображаються на моніторі в реальному часі.

У порівнянні з іншими моделями, такими як MCNN та CSRNet, YOLOv5 виявляє свою більшу перевагу у динаміці. MCNN і CSRNet здебільшого спеціалізуються на статичних сценах з високою щільністю натовпів, що може обмежувати їх застосування в умовах швидкої зміни середовища. Водночас YOLOv5 є одноетапною архітектурою, яка дозволяє зберігати високу швидкість обробки, досягаючи при цьому конкурентоспроможних результатів за точністю.

Отже, під час дослідження було виявлено, що модель YOLOv5 є оптимальним вибором для задачі з підрахунком людей у роботі з відеопотоками, де натовп має постійну динаміку, вона забезпечує достатню швидкість та точність в обробці даних з відео. Порівняльний аналіз моделей показав, що YOLOv5 у основних потрібних аспектах переважає інші моделі та має перевагу у менших витратах ресурсів за рахунок простішої реалізації, що робить її доцільною у використанні в системах моніторингу безпеки та управління натовпами. Також продемонстровано покрокову реалізацію методу визначення кількості людей у відеопотоці з застосуванням YOLOv5 та візуалізацію цього методу.

Перелік посилань

1. Zhang, Y., Chen, S., Wang, D., & et al. (2021). Crowd Counting via Density Map Estimation and Optimization. *IEEE Transactions on Image Processing*, 30, 3486-3498.
2. Huang, Z., & Wang, J. (2020). A comprehensive review of deep learning-based object detection. *IEEE Transactions on Neural Networks and Learning Systems*, 31(4), 1234-1248.
3. Zhang, Y., Zhou, D., Chen, S., Gao, S., & Ma, Y. (2016). Single-Image Crowd Counting via Multi-Column Convolutional Neural Network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 589–597.
4. Li, Y., Zhang, X., & Chen, D. (2018). CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1091–1100.
5. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779–788.
6. Bochkowski, Alexey, WANG, Chien-Yao, et LIAO, Hong-Yuan Mark. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*, 2020.

Додаток Б

Сертифікат учасника XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2024»

Сертифікат № 2024-069-1



Міністерство освіти і науки України
Хмельницький національний університет

СЕРТИФІКАТ



Голуб Юрій Віталійович

учасник XVI Всеукраїнської науково-практичної конференції
«Актуальні проблеми комп'ютерних наук АПКН-2024»
24 години участі (0,8 ECTS credits)

Голова оргкомітету АПКН-2024

Олег СИНЮК

проректор Хмельницького національного
університету з наукової роботи,
доктор технічних наук, професор

м. Хмельницький
15-16 листопада 2024

E-mail: apkt.khnu@gmail.com

Додаток В

Презентація

КВАЛІФІКАЦІЙНА
РОБОТА
МАГІСТРА

Метод визначення кількості людей у натовпі у відеопотоці нейромережевими засобами

Виконав: студент II курсу ОР «Магістр» група, КНм-23-2, Ю.В. Голуб
Керівник: к.т.н., доцент кафедри КН, Р.О. Багрій

Актуальність

Сучасні тенденції урбанізації та зростання кількості населення в містах призводять до збільшення концентрації людей у громадських місцях, таких як транспортні вузли, площі, спортивні та культурні заходи. Ефективний моніторинг цих локацій є критично важливим для забезпечення безпеки, управління потоками людей і зменшення ризику інцидентів, таких як тисняви чи паніка.

Існуючі підходи до визначення кількості людей у натовпі, що базуються на традиційних методах аналізу відеопотоків, часто виявляються недостатньо точними або вимагають значних ресурсів для їх реалізації. Зі збільшенням обсягів відеоінформації від камер спостереження необхідно впроваджувати автоматизовані рішення, які можуть працювати в реальному часі та адаптуватися до змінних умов, таких як освітлення, щільність натовпу або зміна ракурсів зйомки.

У цьому контексті методи на основі нейронних мереж відкривають нові можливості для аналізу складних даних. Зокрема, моделі глибокого навчання, які здатні виконувати завдання розпізнавання об'єктів із достатньо високою точністю та швидкістю. Їх застосування для підрахунку людей у натовпі може значно підвищити ефективність систем відеоспостереження, знизити людський фактор і забезпечити адаптивність до різних сценаріїв.

Мета роботи

Метою роботи є підвищення точності та ефективності визначення кількості людей у натовпі в реальному часі у відеопотоці нейромережевими засобами.

Завдання

Для досягнення поставленої мети потрібно виконати наступні завдання:

- провести аналіз нейромережових моделей та існуючих підходів для визначення кількості людей у натовпі у відеопотоці;
- розробити метод визначення кількості людей у натовпі у відеопотоці з використанням нейромережової моделі;
- підготувати набір даних, що містить різноманітні зображення натовпу для донавчання моделі;
- провести валідацію запропонованого методу за відомими статистичними показниками та порівняти результати з відомими підходами.

Об'єкт та предмет дослідження

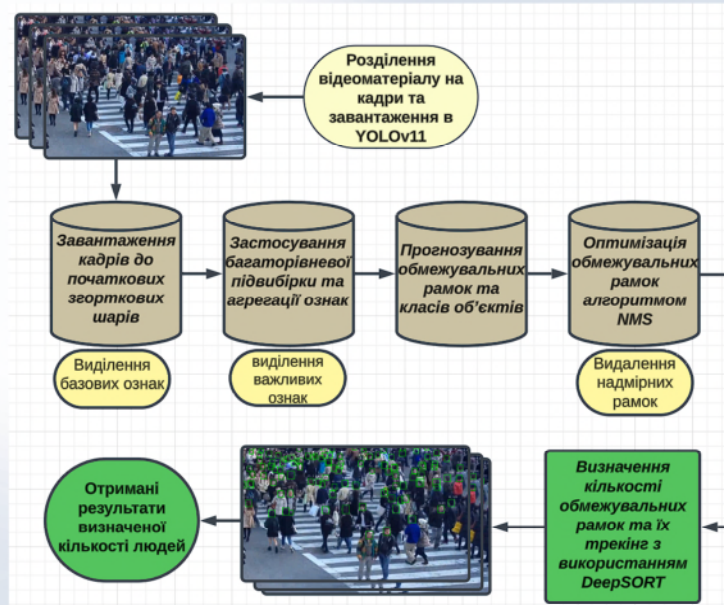
Об'єкт дослідження - процес визначення кількості людей у натовпі на основі відеопотоку з використанням нейромережових моделей для обробки зображень.

Предмет дослідження – моделі нейронної мережі для підрахунку людей у натовпі.

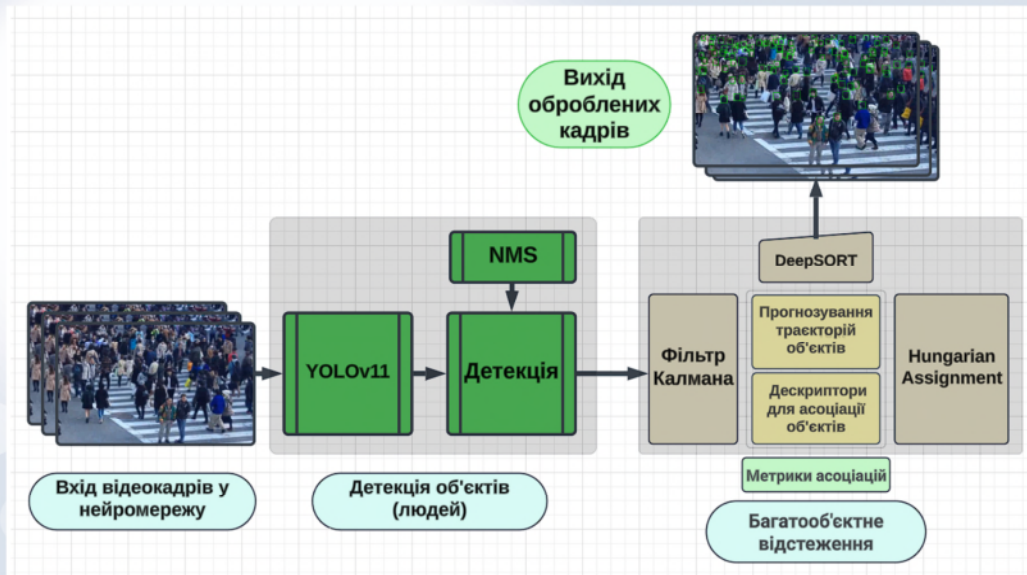
Наукова новизна

В результаті проведеної роботи удосконалено метод визначення кількості людей у натовпі за рахунок використання сучасних архітектур глибоких нейронних мереж, оптимізованих для обробки відеопотоків, який відрізняється від існуючих донавчанням моделі на спеціалізованих наборах даних та використанням алгоритмів трекінгу об'єктів у кадрі для покращення результатів у динамічних сценах, що дало можливість підвищити точність та ефективність визначення кількості людей у натовпі в реальному часі у відеопотоці.

Схема методу виявлення кількості людей у натовпі у відеопотоці



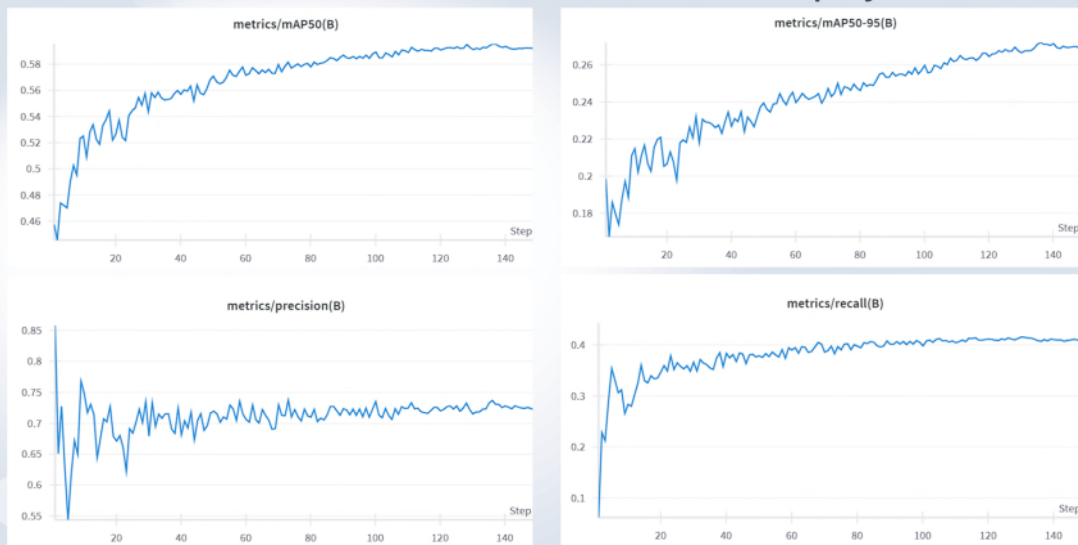
Візуалізація проходження кадрів відео через модель YOLOv11 та алгоритми трекінгу DeepSORT



Навчання нейромережі

Модель навчалася на унікальних зображеннях, адаптованих для задач високої щільності натовпу, де мітки рамок розташовані на головах людей, для уникнення помилок через перекриття. Це значно покращило точність розпізнавання кількості натовпу в умовах щільності. Для обробки відеопотоку застосовано алгоритм трекінгу DeepSORT, який запобігає повторному підрахунку тих самих осіб і підвищує точність шляхом відстеження їх переміщень між кадрами.

Основні показники точності YOLOv11s після тренування



Модель	Precision	Recall	mAP50	mAP50-95
Покращена модель	72,3%	41,1%	59,2%	26,9%

Порівняння результатів виявлення людей стандартної моделі з донавченою на прикладі відео



Робота стандартної, попередньо тренованої моделі



Робота донавченої моделі

Висновки

В результаті виконання кваліфікаційної роботи магістра було розроблено метод виявлення кількості людей у натовпі у відеопотоці неймережевими засобами.

В процесі виконання дослідження було виконано наступні задачі:

- проведено аналіз неймережових моделей та існуючих підходів для визначення кількості людей у натовпі у відеопотоці;
- розроблено метод визначення кількості людей у натовпі у відеопотоці з використанням неймережової моделі, запропонований метод складається з згорткової мережі та удосконалюється донавчанням моделі та використанням алгоритму трекінгу об'єктів DeepSORT, для стабільності результатів обробки.
- проведено валідацію запропонованого методу за відомими статистичними показниками та порівняти результати з відомими підходами.

Дякую за увагу!

Anti-Plagiarism v-15.257

Максимальне співпадіння з одним документом 1.0%

Словники перевірки: en_US, ru_RU, ua_UA. Помилки в документах: 12%

ID: 160829 Назва: КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА на тему Метод визначення кількості людей у натовпі у відеопотоці нейромережевими засобами Додано в БД: 2024-12-18 Автора: Юрій Голуб Керівники: Руслан Багрій Консультанти: Опоненти:	Документ		Сумарний збіг по Базі Даних	
	Символи	Лексеми	Символи	Лексеми
	98943	1459	1621 (2%)	26 (2%)

Джерело плагіату

ID	Опис	Наявність плагіату в документі	
		Символи	Лексеми

Протокол аналізу звіту подібності науковим керівником

Заявляю, що я ознайомився (-лась) з Повним звітом подібності, який був згенерований Системою виявлення і запобігання плагіату щодо роботи:

Автор: Юрій Голуб

Співавтор:

Назва: Метод визначення кількості людей у натовпі у відеопотоці нейромережевими засобами

Науковий керівник: Руслан Багрій, к.т.н., доцент

Підрозділ: Кафедра комп'ютерних наук

Коефіцієнт подібності 1: 2.8%

Коефіцієнт подібності 2: 1.1%

Мікропробіли: 0

Заміна букв: 0

Інтервали: 0

Білі знаки: 1

Дата створення звіту: 2024-12-18 08:31:28.0

Після аналізу Звіту подібності констатую наступне:

Запозичення, виявлені в роботі є законними і не є плагіатом. Рівень подібності не перевищує допустимої межі. Таким чином робота незалежна і приймається.

Запозичення не є плагіатом, але перевищено граничне значення рівня подібностей. Таким чином робота повертається на доопрацювання.

Виявлено запозичення і плагіат або навмисні текстові спотворення (маніпуляції), як передбачувані спроби укриття плагіату, які роблять роботу невідповідною вимогам законодавства (Ст. 32. ЗУ Про вищу освіту, пункт 3.1, Ст. 42. ЗУ Про освіту) та вимог НАЗЯВО (Критерій 5), а також кодексу етики і процедурам. Таким чином робота не приймається.

Обґрунтування:

Дата

18.12.2024

експерт

Левобільський С.Р.

РІШЕННЯ ЕКСПЕРТНОЇ КОМІСІЇ

КАФЕДРИ КОМП'ЮТЕРНИХ НАУК

ПРО ДОПУСК КВАЛІФІКАЦІЙНОЇ РОБОТИ МАГІСТРА ДО ЗАХИСТУ ЗА РЕЗУЛЬТАТАМИ АНАЛІЗУ ЗВІТУ ПОДІБНОСТІ

Підтверджуємо ознайомлення з результатом звіту подібності щодо роботи, генерованого системою виявлення текстових збігів/ідентичності/схожості:

Назва: Метод визначення кількості людей у натовпі у відеопотоці нейромережевими засобами

Автор: Юрій ГОЛУБ

Спеціальність: 122 – Комп'ютерні науки

Освітня програма: освітньо-професійна

Науковий керівник: к.т.н., доц. Руслан БАГРІЙ

Після аналізу звіту подібності зроблено такий висновок:

№	Висновок	Позначка про відповідність
1	Запозичення, виявлені в роботі, є законними і не є плагіатом. Робота приймається до захисту.	відповідає
2	Виявлені запозичення не є плагіатом, розміщені в розділах, які не описують безпосередньо авторське дослідження, але кількість цитат перевищує обсяг, виправданий поставленою метою роботи. Робота приймається до захисту, але має бути відкоригована. Відкоригований варіант має бути поданий на кафедру за 2 дні до захисту, разом із заявою щодо самостійності виконання письмової роботи та ідентичності друкованої та електронної версії роботи	
3	Виявлені запозичення не є плагіатом, але частково розміщені в розділах, які описують безпосередньо авторське дослідження, а кількість цитат перевищує обсяг, виправданий поставленою метою роботи. В зв'язку з цим мета роботи та поставлені завдання не були досягнені. Робота може бути допущена до захисту (наступного року) після того як буде відкоригована та допрацьована і успішно пройде повторну перевірку на академічний плагіат.	
4	Робота містить навмисні текстові спотворення, передбачувані спроби укриття запозичень або інші прояви академічного плагіату. Робота містить фабрикацію або фальсифікацію даних. Робота не допускається до захисту.	

Підтвердження:

Запозичення, виявлені в роботі, є законними і не є плагіатом, оскільки:

- 1) за програмою Anti-Plagiarism виявлені 1% є фрагментарними – містять поширені конструкції, загальновідомі терміни, скорочення та визначення.
- 2) За програмою StrikePlagiarism виявлені за коефіцієнтом подібності КП1 – 2.8%, за коефіцієнтом подібності КП2 – 1.1%, що є запозиченнями, які розміщені в розділах аналізу існуючих технологій та прототипів, які не описують безпосередньо авторське дослідження і не стосуються результатів роботи.

Сумарний обсяг всіх запозичень, визначений системою виявлення збігів/ідентичності/схожості, складає 1% та 2.8% і 1.1% відповідно, що, з урахуванням наведених обґрунтувань, відповідає характеру наукового дослідження і свідчить на користь кваліфікаційної роботи.

Керівник роботи

Гарант ОП

Завідувач кафедри КН



Руслан БАГРІЙ

Руслан БАГРІЙ

Олександр БАРМАК



ВІДГУК ОПОНЕНТА

на кваліфікаційну роботу магістра

гр. КНм-23-2 Юрія ГОЛУБА за темою: Метод визначення кількості людей у натовпі у відеопотоці нейромережевими засобами

1. Актуальність обраної теми

Розробка методу визначення кількості людей у натовпі у відеопотоці є достатньо актуальною на даний час, особливо з урахуванням тенденцій щодо розвитку штучного інтелекту та зростаючої потреби в автоматизації процесів моніторингу. Подібні методи знаходять застосування у різних сферах, таких як забезпечення безпеки на масових заходах, управління натовпами, аналізу поведінки у громадських місцях та оптимізації роботи транспортних вузлів. Актуальність обраної теми детально обґрунтована в роботі, що підтверджується її практичною значущістю та потенціалом для подальших досліджень.

2. Відповідність роботи предметній області спеціальності 122 Комп'ютерні науки та загальним вимогам до наукових робіт

Тема кваліфікаційної роботи студента повністю відповідає предметній області спеціальності 122 Комп'ютерні науки та загальним вимогам до кваліфікаційної роботи магістра. Робота вирішує науково-прикладну задачу у сфері комп'ютерних наук (а саме застосування інтелектуальних методів обробки інформації (моделей глибокого навчання) до визначення кількості людей у натовпі). У роботі присутнє дослідження, яке розвиває існуючі підходи, та отримані нові результати, які за певними показниками – вищі, ніж в аналогічних підходах.

3. Повнота розкриття мети та завдань дослідження

Часткові завдання дослідження сформульовані коректно та повністю відповідають меті роботи.

4. Наявність наукової новизни

Магістерська робота характеризується наявністю наукової новизни. Положення наукової новизни сформульовані коректно та відображають їх сутність.

Матеріали магістерської роботи апробовані на науковій конференції «Актуальні проблеми комп'ютерних наук».

5. Зміст кожного розділу роботи

Робота складається з чотирьох розділів,

р., м. Хмельницький, Україна, доповідь на тему «Визначення кількості людей в натовпі у відеопотоці нейромережевими засобами».

6. Ступінь оволодіння методами дослідження

Студент впевнено застосовував сучасні методи дослідження, включаючи глибоке навчання, згорткові нейронні мережі, алгоритми трекінгу об'єктів та статистичні методи оцінки ефективності.

7. Повнота та якість розкриття теми роботи

Поставлені задачі повністю виконані. Результати підтверджують наукову обґрунтованість та практичну значимість дослідження.

8. Логічність, послідовність, аргументованість, літературна грамотність викладу матеріалу

Робота відзначається структурованим викладом, логічною послідовністю та науковим стилем. Використаний стиль відповідає сучасним стандартам наукового письма, що забезпечує легкість сприйняття матеріалу.

9. Можливість практичного застосування кваліфікаційної роботи, окремих її частин

Розроблений метод має значний потенціал для інтеграції у системи відеоспостереження для моніторингу великих скупчень людей у реальному часі.

10. Висновок про можливість допуску кваліфікаційної роботи до захисту, на яку оцінку заслуговує робота

Кваліфікаційну роботу магістра Юрія ГОЛУБА рекомендовано до захисту. Вважаю, що робота заслуговує на оцінку «добре».

Науковий керівник _____



к.т.н., доц. Руслан Багрій



**ВІДГУК НАУКОВОГО КЕРІВНИКА
на кваліфікаційну роботу магістра**

гр. КНм-23-2 Юрія ГОЛУБА за темою: «Метод визначення кількості людей у натовпі у відеопотоці нейромережевими засобами»

1. Актуальність теми

Актуальність теми кваліфікаційної роботи обумовлена необхідністю вдосконалення систем відеоспостереження для моніторингу громадських місць, що є критично важливим для забезпечення безпеки на масових заходах. Запропонований метод на основі нейромережових засобів значно підвищує точність та ефективність визначення кількості людей у натовпі в реальному часі.

2. Відповідність роботи предметній області спеціальності 122 Комп'ютерні науки та загальним вимогам до наукових робіт

Тема повністю відповідає предметній області спеціальності 122 – Комп'ютерні науки. Об'єктом дослідження є процес автоматичного визначення кількості людей у відеопотоці, а предметом – нейромережові моделі для обробки зображень та алгоритми трекінгу об'єктів.

3. Професійні та особистісні якості магістранта

Юрій ГОЛУБ у процесі роботи продемонстрував високий рівень знань у галузі нейромережових технологій та комп'ютерного зору. Студент виявив цілеспрямованість, самостійність у прийнятті рішень та вміння працювати над складними дослідницькими завданнями..

4. Ступінь самостійності під час виконання кваліфікаційної роботи

Робота виконана самостійно, академічного плагіату не виявлено. Всі наукові запозичення мають відповідні посилання на джерела.

5. Наукова новизна та оригінальність запропонованих підходів

В роботі запропоновано метод удосконалення підрахунку людей у натовпі за допомогою моделі YOLOv11 у поєднанні з алгоритмом DeepSORT. Доновчання моделі на спеціалізованих наборах даних забезпечило високу точність та адаптивність у динамічних сценах. Отримані результати оприлюднені на XVI всеукраїнській науково-практичній конференції «Актуальні проблеми комп'ютерних наук АПКН-2024», 15-16 листопада 2024

У першому розділі – «Огляд існуючих методів та моделей для підрахунку кількості людей у натовпі на основі відеопотоків» автором проведено аналіз предметної області, існуючих публікацій, щодо методів визначення кількості людей у натовпі та моделей глибокого навчання, що використовуються для цього.

У другому розділі – «Розробка методу визначення людей у натовпі у відеопотоці нейромережевими засобами» магістрантом розроблено та удосконалено метод визначення кількості людей у натовпі за моделлю глибокого навчання та покращено даний метод застосуванням алгоритму трекінгу.

У третьому розділі – «Програмна реалізація методу визначення людей у натовпі у відеопотоці нейромережевими засобами» автором описана засоби розробки та засоби для відстеження результатів тренування моделі.

У четвертому розділі – «Дослідження методу визначення людей у натовпі та порівняння результатів» магістрантом проведено оцінку ефективності вдосконаленого авторського методу. В якості результатів, що порівнювались із запропонованим методом, було обрано порівняння з базовою попередньо тренованою моделлю без покращень.

6. Ступінь розкриття теми роботи

Тема роботи розкрита достатньою мірою та мета досягнута.

7. Якість оформлення кваліфікаційної роботи

Робота написана літературною мовою та оформлена згідно відповідних вимог достатньо якісно.

8. Недоліки кваліфікаційної роботи

Разом з тим, у роботі є і певні недоліки. Так, у роботі відсутні відповіді на питання щодо ефективності авторського удосконалення на інших наборах даних, крім тих, що досліджені в магістерській роботі.

Проте зазначений недолік загалом не впливає на комплексну оцінку роботи.

9. Загальний висновок (допускається чи не допускається до захисту), якої оцінки заслуговує кваліфікаційна робота.

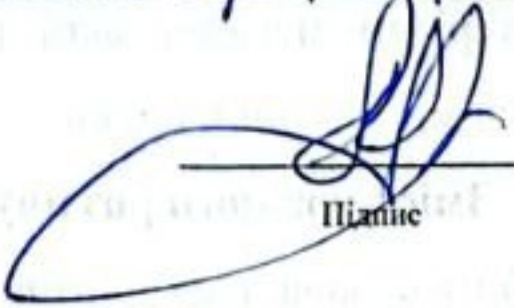
Рекомендую допустити роботу до захисту.

Кваліфікаційна робота заслуговує на оцінку «Добре».

Опонент (прізвище, імя, по батькові, посада, місце роботи)

Лисенко Сергій Михайлович д.т.н. професор КІІС

«19» 12 2024 р.


Підпис