



АНАЛІЗ МЕТОДІВ ДЕТЕКЦІЇ ТА МОНІТОРИНГУ ЛЮДИНИ У ВІДЕОПОТОКАХ

Четвериков Г.Г., професор, кафедра ПІ, ХНУРЕ
Миронюк С.А., студент, кафедра ПІ, ХНУРЕ

В останні роки відеоаналітика, що базується на комп'ютерному зорі, стала універсальним інструментом у багатьох сегментах діяльності людини. Ринок цієї технології очікує значний зріст із річним темпом в 22,3% до 2028 року, з використанням інтелектуальних камер та передових технологій, таких як глибоке навчання [1]. Системи відеоаналітики забезпечують автоматичний аналіз відеоматеріалів для виявлення та розпізнавання подій за часом та місцем їх відбуття, що значно покращує ефективність відеоспостереження. Це сприяє оперативній реакції на потенційні загрози та оптимізації бізнес-процесів. Важливим завданням є детекція та моніторинг людей у відеопотоці, що застосовується в автоматичному керуванні з метою виявлення пішоходів на вулицях, аналізі поведінки споживачів у місцях торгівлі, клієнтської активності, визначенні навантаження на персонал, моніторингу пасажиропотоку, а також в системах «розумного будинку» та офісних комплексах. Метою роботи є дослідити існуючі алгоритми детектування людини, визначити слабкі та сильні сторони.

Методи виявлення об'єктів розвивалися через два історичні етапи: перший (до 2014 року) – використання класичних алгоритмів, що використовували ручно створені ознаки; другий (після 2014 року) – поява методів, що ґрунтуються на глибокому навчанні, зокрема на згорткових нейронних мережах (CNN). Традиційний метод Віоли-Джонса відзначився високими результатами у виявленні об'єктів, але його навчання повільне через використання великої кількості функцій [2]. Гістограма напрямлених градієнтів працює з малими локальними зонами і залишається стійкою до геометричних та фотометричних змін [3]. Проте класичні методи більш придатні для обмежених тренувальних наборів, а CNN виявляються більш ефективними для аналізу відеопослідовностей через свою унікальну архітектуру та здатності автоматично вчитися розпізнавати характеристики зображень.

Нейронні мережі вражають своєю універсальністю та ефективністю в різних завданнях, завдяки паралельній обробці інформації у всіх їхніх шарах. Здатність до навчання дозволяє узагальнювати знання та розуміти взаємозв'язки між різними частинами зображення. Що робить згорткові нейронні мережі успішними для розпізнавання об'єктів у реальному часі, забезпечуючи точні та надійні результати навіть в умовах зміни параметрів.

Сучасні методи детекції об'єктів можна визначити як одно- та двоетапові методи виявлення об'єктів. Двоетапові детектори спершу формують області-кандидати, які потенційно містять об'єкти, після цього проводиться класифікація та виконується регресія обмежувальної рамки з використанням характеристик цих областей. Ці моделі відзначаються високою точністю у



виявленні об'єктів, але зазвичай працюють повільніше. Мережа R-CNN (Region-based Convolutional Neural Networks) та її варіації (Fast R-CNN, Faster R-CNN, Mask R-CNN, Mesh R-CNN) є прикладами двоступеневих детекторів. З метою прискорення процесу виявлення об'єктів та зменшення кількості непотрібних обчислень у Faster R-CNN використовується додатковий модуль RPN («Region Proposal Network»), який відповідає за генерацію пропозицій регіонів для подальшого вивчення. На відміну від регіональних підходів однокаскадні детектори застосовують лише один прохід через мережу, яка напряму прогнозує обмежувальні прямокутники та класи об'єктів. Це може призвести до меншої точності у порівнянні з регіональними методами, але вони працюють набагато швидше, що робить їх дуже ефективними для завдань, які потребують миттєвого виявлення та відслідковування об'єктів. Методами однопрохідного детектування є YOLO, SSD і RetinaNet. Вони відрізняються базовою згортковою мережею, що використовується для екстрагування карт ознак; способом знаходження якірних обмежувальних рамок; та функцією втрат.

Хоча в сфері виявлення об'єктів досягнуто значний прогрес у використанні різних моделей, проблеми виникають у реальних сценаріях відстеження. Ці проблеми включають взаємодії між об'єктами, обмеження видимості (оклюзії) та перешкоди у фоні, різні масштаби об'єктів. Навчання мереж та сам процес детектування можуть забирати багато часу. Все це може ускладнювати завдання відслідковування.

Для забезпечення реалістичності та підвищення рівня складності завдань у контексті порівняльного аналізу за точністю і швидкодією алгоритмів детекції: Faster R-CNN, YOLO, SSD і RetinaNet було використано набір даних Human-Aligned Bounding Boxes from Overhead Fisheye cameras (HABBOFcameras) від лабораторії обробки візуальної інформації (VIP) університету Бостона з різноманітністю сцен, дій та людей. Для оцінки ефективності алгоритмів використовувались такі параметри: середня точність (mAP), швидкість обробки кадрів у секунду (FPS) та метрики точності (precision) і повноти (recall).

Отже, двоетаповий детектор Faster R-CNN продемонстрував високу точність, але повільно обробляє відео. RetinaNet і SSD підходять для точного виявлення, але уступають в швидкості YOLO. SSD працює з об'єктами різних масштабів. Високий показник повноти у RetinaNet підтверджує точність класифікації. Серед усіх розглянутих мереж, YOLO – найшвидша і може бути використана в реальному часі.

Список літератури

1. Markets and Markets. (n.d.). Video Analytics Market worth \$22.6 billion by 2028. <http://www.marketsandmarkets.com/PressReleases/iva.asp>.
2. Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. Computer Vision and Pattern Recognition Conference (CVPR).
3. Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. Computer Vision and Pattern Recognition Conference (CVPR).
4. Shengye Li, M. Ozan Tezcan, Prakash Ishwar, Janusz Konrad. (2019). People counting using overhead fisheye cameras. <https://vip.bu.edu/projects/vsns/lossy/fisheye/li-et-al/>.