Хмельницький національний університет
Факультет інформаційних технологій
Кафедра комп'ютерних наук

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

на тему *Захоплення та трасування об'єктів на зображеннях в системах наведення цілі за допомогою нейронних мереж*

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 4 курсу, група КН-19-1* _____ *С.В. Горохольський*
Курс, група виконавця — Підпис — Ініціали, прізвище

Керівник: *д.т.н., доцент кафедри КН* _____ *Е.А. Манзюк*
Науковий ступінь, посада — Підпис — Ініціали, прізвище

Нормоконтроль: *к.т.н., доцент кафедри КН* _____ *Р.О. Багрій*
Науковий ступінь, посада — Підпис — Ініціали, прізвище

До захисту допускаю:
Зав. кафедри КН, д.т.н., професор _____ *О.В. Бармак*
Підпис — Ініціали, прізвище

01 06 2023 р.

Хмельницький 2023

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет *інформаційних технологій*
Кафедра *комп'ютерних наук*
Освітній ступінь *бакалавр*
Галузь знань *12 – Інформаційні технології*
Спеціальність *122 – Комп'ютерні науки*
Освітня програма *освітньо-професійна програма підготовки бакалавра*

ЗАТВЕРДЖУЮ
Завідувач кафедри *комп'ютерних наук*

_____
(підпис)
д.т.н., професор О.В. Бармак
«*06*» *03* *2023* року

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ БАКАЛАВРА

1. Тема кваліфікаційної роботи бакалавра: *«Захоплення та трасування об'єктів на зображеннях в системах наведення цілі за допомогою нейронних мереж»*

2. Завдання видано студентці *Горохольському Станіславу В'ячеславовичу*
(прізвище, ім'я, по батькові)

3. Керівник роботи *доцент кафедри КН Манзюк Едуард Андрійович*
(посада, прізвище, ім'я, по батькові)

4. Затверджено наказом університету від «*01*» *03* *2023* р. № *5*

5. Дата видачі завдання студенту: «*03*» *03* *2023* р.

6. Зміст пояснювальної записки (перелік задач) та вихідні дані:

Провести аналіз предметної області, здійснити огляд методів захоплення об'єктів та слідкування за їх переміщенням на послідовних зображеннях. Розробити спосіб захоплення та трасування об'єктів на зображеннях в системах наведення цілі за допомогою нейронних мереж. Визначити послідовність застосування способу захоплення та трасування об'єктів на зображеннях в системах наведення цілі. Розробити програмну систему з практичної реалізації розробленого способу з використанням сучасних засобів проєктування та розробки програмного забезпечення.

7. Календарний план виконання кваліфікаційної роботи бакалавра:

№	Назва етапів (розділів) кваліфікаційної роботи бакалавра	Термін виконання	Примітка
1	*Вибір напряму дослідження та узгодження тематики кваліфікаційної роботи бакалавра з керівником*	*грудень 2022*	*виконано*
2	*Ознайомлення з предметною областю, формулювання мети та задач дослідження, визначення об'єкта та предмета дослідження*	*січень 2023*	*виконано*
3	*Робота над розділом 1 – Характеристика предметної області та постановка задачі*	*січень 2023*	*виконано*
4	*Робота над розділом 2 – Захоплення та трасування об'єктів на зображеннях в системах наведення цілі за допомогою нейронних мереж*	*березень 2023*	*виконано*
5	*Робота над розділом 3 – Програмна реалізація способу захоплення та трасування об'єктів на зображеннях в системах наведення цілі за допомогою нейронних мереж*	*квітень 2023*	*виконано*
6	*Оформлення пояснювальної записки згідно вимог*	*травень 2023*	*виконано*
7	*Попередній захист кваліфікаційної роботи бакалавра*	*травень 2023*	*виконано*
8	*Захист кваліфікаційної роботи бакалавра на засіданні Екзаменаційної комісії*	*червень 2023*	*виконано*

Виконавець: *студент 4 курсу, група КН-19-1* (Курс, група виконавця) _____ (Підпис) *С.В.Горохольський* (Ініціали, прізвище)

Керівник: *д.т.н., доцент кафедри КН* (Науковий ступінь, посада) _____ (Підпис) *Е.А. Манзюк* (Ініціали, прізвище)

Анотація

Тема кваліфікаційної роботи бакалавра: *Захоплення та трасування об'єктів на зображеннях в системах наведення цілі за допомогою нейронних мереж.*

Виконавець кваліфікаційної роботи бакалавра: *студент групи КН-19-1 Горохольський Станіслав В'ячеславович.*

Керівник кваліфікаційної роботи бакалавра: *д.т.н., доцент кафедри КН Манзюк Едуард Андрійович*

Кваліфікаційна робота бакалавра містить:

Пояснювальна записка				Кількість додатків
Сторінок	Рисунків	Таблиць	Джерел інформації	
68	12	8	56	1

Мета кваліфікаційної роботи бакалавра полягає в розробці способу захоплення та трасування об'єктів на зображеннях в системах наведення цілі за допомогою нейронних мереж.

Для досягнення поставленої мети визначені наступні задачі дослідження: визначити послідовність застосування способу захоплення об'єкту на зображенні з наступним трасуванням його переміщення; розробити спосіб захоплення та слідкування за переміщення об'єкта на зображеннях за допомогою нейронної мережі; спроєктувати та реалізувати програмну систему.

Результатом виконання кваліфікаційної роботи бакалавра є створення способу захоплення та трасування об'єктів на зображеннях в системах наведення цілі за допомогою нейронних мереж.

Ключові слова: аналіз зображення, ідентифікації об'єктів на зображенні, трасування об'єктів на зображенні.

Виконавець: *студент 4 курсу, група КН-19-1* (Курс, група виконавця) _____ (Підпис) *С.В.Горохольський* (Ініціали, прізвище)

Зміст

Перелік скорочень

Скорочення, термін, позначення	Пояснення
КРБ	Кваліфікаційна робота бакалавра
КН	Комп’ютерні науки
CNN	Згорткова нейронна мережа
RPN	Мережа регіональних пропозицій
SSD	Single Shot Multi-Box Detector
YOLO	Одноходова нейронна мережа
FCN	Повна згорткова мережа
FCIS	Повна згорткова сегментація екземплярів
CDA	Корекційна доменна адаптація
KCF	Кореляційний фільтр ядра

Вступ

Кваліфікаційна робота бакалавра присвячена розробці способу захоплення та трасування об’єктів на зображеннях в системах наведення цілі за допомогою нейронних мереж, що дозволяє за зображеннями ідентифікувати об’єкти та здійснювати стеження за ними при зміні зображення як за окремими фрагментами, так за послідовними змінами зображень.

Актуальність – виявлення та відслідковування об’єктів є важливими задачами комп'ютерного зору, які стали ключовими завданнями для багатьох практичних застосувань, таких як відеоспостереження, інтелектуальні транспортні системи, охоронні системи. Завдяки технологіям глибокого навчання, таким як згорткові нейронні мережі, сучасні системи виявлення та відслідковування об'єктів досягають значно кращої точності у практичних застосуваннях. В роботі розроблена система виявлення та відслідковування об’єктів на основі глибокого навчання.

Для покращення виявлення об’єктів використано семантичну контекстну інформація для виявлення об’єктів, та семантичні ознаки, які отримані із застосуванням семантичних масок сегментації. Ці маски сегментації діють як механізм отримання областей і дозволяють детекторам зосереджуватися на тих ділянках зображення, де найімовірніше з'являться потенційні кандидати в об'єкти.

Об’єкт дослідження – процес обробки зображень в задачах ідентифікації та слідування переміщення об’єктів.

Предмет дослідження – моделі, методи, алгоритми та засоби для створення способу захоплення та трасування об’єктів на зображеннях в системах наведення цілі за допомогою нейронних мереж.

Мета кваліфікаційної роботи бакалавра полягає в розробці способу захоплення та трасування об’єктів на зображеннях в системах наведення цілі за допомогою нейронних мереж.

Завдання кваліфікаційної роботи бакалавра.

Для досягнення поставленої мети визначені наступні задачі дослідження:

– визначити послідовність застосування способу захоплення об’єктів на зображеннях в системах наведення цілі за допомогою нейронних мереж;

– визначити послідовність застосування способу трасування об’єктів на зображеннях в системах наведення цілі;

– реалізувати програмну систему захоплення та трасування об’єктів на зображеннях в системах наведення цілі за допомогою нейронних мереж;

– провести експериментальне тестування розробленого способу.

Розділ 1 Характеристика предметної області та постановка задачі

1.1 Аналіз предметної області виявлення та відслідковування об’єктів в задачах комп'ютерного зору

Виявлення та відслідковування об’єктів є важливими задачами комп'ютерного зору, які стали ключовими завданнями для багатьох практичних застосувань, таких як відеоспостереження, інтелектуальні транспортні системи, охоронні системи. Завдяки технологіям глибокого навчання, таким як згорткові нейронні мережі, сучасні системи виявлення та відслідковування об'єктів досягають значно кращої точності у практичних застосуваннях. В останній час, обчислювальні машини стають дедалі потужнішими, машинний інтелект досяг значних успіхів у багатьох реальних сферах застосування, таких як система розпізнавання облич, машинний перекладач, безпілотний автомобіль, моніторинг безпеки. Усі ці програми роблять штучний інтелект незамінним у повсякденному житті суспільства. Завдяки розвитку мобільних пристроїв, соціальних мереж і високошвидкісного зв'язку у світі постійно зростає кількість зображень, що робить все менш можливим керування всіма цими даними вручну. Тому розробка комп'ютерних систем для автоматичної обробки та розуміння великої кількості даних стає необхідним завданням. Однак, як відомо, комп'ютери добре справляються із завданнями, які можна визначити за допомогою формальних і математичних правил, таких як обчислення, зберігання, пошук та інші. Але вирішувати інтуїтивно зрозумілі та абстрактні завдання, такі як розпізнавання зображень, для комп'ютерів є складним завданням. Це спричинено семантичним розривом між людиною і машиною, тобто файли зображень зберігаються у вигляді низькорівневих піксельних даних на машинах, але для аналізу зображень необхідна високорівнева семантична інформація. Комп'ютерний зір намагається скоротити цей розрив і навчити машини розуміти зображення.

В роботі розроблена система виявлення та відслідковування об’єктів на основі глибокого навчання. Для покращення виявлення об’єктів використаємо

семантичну контекстну інформацію для виявлення об’єктів, та семантичні ознаки, які отримані із застосуванням семантичних масок сегментації. Ці маски сегментації діють як механізм отримання областей і дозволяють детекторам зосереджуватися на тих ділянках зображення, де найімовірніше з'являться потенційні кандидати в об'єкти. Крім того, проаналізувавши деякі широко застосовні набори даних, визначено, що якість анотацій для малих об'єктів і об'єктів скупчення об’єктів не задовільняє необхідним вимогам для їх визначення. Водночас було запропоновано базовий метод виявлення об’єктів, який використовує особливості напряму руху для підвищення ефективності виявлення. Результати експерименту показують, що підхід значно покращує точність виявлення для базових об’єктів.

Також запропоновано перенести глибинну функцію, яка спочатку навчалася для класифікації зображень, до області візуального відслідковування. Адаптація до домену досягається за допомогою об’єднання допоміжних мереж, які навчаються шляхом регресії розташування об'єктів у кадрах відслідковування. Крім того, адаптацію також використано для введення концепції об'єктності у візуальне відслідковування. Це усуває невизначеність напрямку переміщення об’єкта в задачах візуального відслідковування, і дозволяє проілюструвати емпіричну перевагу більш чітко визначеної задачі. Експериментально продемонстровано ефективність метода відслідковування на корпусі даних.

1.2 Огляд нейронних мереж для задач обробки зображення та виявлення об’єктів

Виявлення об'єктів є фундаментальною проблемою комп'ютерного зору і є важливим завданням для багатьох реальних застосувань. Сучасна система виявлення об'єктів в основному складається з двох етапів: локалізація набору об'єктів-кандидатів і класифікація цих об'єктів за певною категорією. За останнє десятиліття, з розвитком глибокого навчання, згорткові нейронні мережі

(Convolutional Neural Networks – CNN) [21] стали фактично стандартом для вирішення цієї задачі, і було запропоновано велику кількість методів на основі CNN [10, 19, 36, 42, 43, 46, 54]. Крім того, сучасні об'єктні методи можна розділити на два типи: одно етапні (однокрокові) підходи та двокрокові підходи.

Відомі двокрокові методи, такі як RCNN [3], Faster RCNN [2, 39] та Mask RCNN [44, 47], розділюють завдання виявлення об'єктів на два кроки: виділення області інтересу (RoI) та класифікація RoI на передній/задній план.

У роботі [29] автори запропонували метод вибіркового пошуку для генерації набору пропозицій-кандидатів, які містять об'єкти всіх категорій, відфільтровуючи більшість негативних місць на першому етапі, а потім використовують дескриптори SIFT [20] як представлення ознак для навчання класифікаторів. Далі класифікують пропозиції за різними категоріями на другому етапі. RCNN [16] замінює дескриптори SIFT на згорнуті ознаки, що дозволило досягти значного покращення точності розпізнавання. Були запропоновані покращені версії RCNN, такі як Fast RCNN [28], Faster RCNN [35] та Mask RCNN [48]. Fast RCNN та Faster RCNN переглянули процес видалення ознак в RCNN та запропонували більш ефективні стратегії видалення ознак, які дозволяють мережі використовувати ту ж саму магістральну мережу з регресорами обмежувальних рамок. Такі підходи значно покращили двокрокові методи як в точності, так і в швидкості.

Однак слід зазначити, що методи на основі Region Proposal Network – RPN вводять надмірну кількість гіперпараметрів, таких як розміри анкерів, крок анкера і співвідношення сторін анкера, які необхідно ретельно налаштовувати для різних наборів даних (особливо для малих цілей), щоб досягти прийнятної точності.

Однокрокові методи, такі як YOLO [9, 13, 18], Retina Net [17, 56], відмовляються від процесу генерації RoI у двоетапних системах виявлення і безпосередньо регресують і класифікують набір попередньо визначених якірних кандидатів.

Система YOLO [18] розбиває вхідне зображення на сітку $S \times S$ і одночасно прогнозує обмежувальні рамки та категорії в цих рамках, що дозволило досягти дуже високої швидкості виводу. Аналогічно, SSD [4] заздалегідь визначає набір стандартних областей і використовує глибинні ознаки з різних рівнів згорткових шарів для регресії та класифікації цих стандартних областей.

В системі Retina-Net розроблено функцію втрат для усунення дисбалансу між позитивними і негативними зразками, але все ще значною мірою покладається на якірні об’єкти. Такі підходи продемонстровано в роботах [34, 49], які були налаштовані на високу швидкість виведення, але їхня ефективність виявлення не відрізняється від більшості двоступеневих методів. Система SSD регресує малі цілі на більш дрібних згорткових шарах, що призволить до гіршої точності на малих об'єктах. Крім того, оскільки кожна комірка сітки в YOLO прогнозує лише два поля і може мати лише один клас, вона не може добре працювати на малих або скупчених об'єктах.

Останнім часом було запропоновано багато безякірних одноступеневих методів. CornerNet [24] виявляє рамку об'єкта як пару ключових точок (верхній лівий і нижній правий кути), однак для групування пар кутів, що належать до одного екземпляра, потрібен складний етап постобробки. Fcos [41] формулює задачу виявлення об'єктів у вигляді прогнозування для кожного пікселя зображення і досягає хороших результатів на відомих наборах даних. ExtremeNet [32] визначає чотири крайні точки (крайню верхню, крайню ліву, крайню нижню, крайню праву) і центральну точку об'єктів за допомогою мережі оцінки ключових точок, потім п'ять ключових точок групуються в обмежувальну рамку за допомогою геометричних правил.

1.3 Огляд нейронних мереж для задач відслідковування об’єктів

Візуальне відслідковування має за мету слідувати за переміщенням конкретного об'єкта, позначеного на першому кадрі відеопослідовності. До появи глибокого навчання, традиційні алгоритми відслідковування приділяли

найбільшу увагу розробці надійної моделі з точки зору стратегії оновлення моделі, ансамблевого постпроцесора, моделі спостереження та інші. Деякі з них досягли значних успіхів як у точності, так і у швидкості.

За останні роки згорткові нейронні мережі [25, 52] досягли успіху завдяки своїй здатності до автоматичного видалення ознак. Експертам більше не потрібно витрачати час на розробку різних ознак, створених вручну. Згідно з сучасними дослідженнями, модель створена на основі нейронний мереж відіграє важливу роль у надійній системі візуального відслідковування [50]. Однак замінити вручну створені ознаки на глибокі згорткові ознаки було неефективно в деяких ранніх алгоритмах відслідковування, заснованих на глибокому навчанні. Відома робота [51], яка вивчає глибинні ознаки для задачі відслідковування візуальних об'єктів. Розробляють глибоку модель, яка працює в автономному режимі з великою кількістю зображень, оновлюючись в режимі онлайн для поточної відеопослідовності [26]. Витягують ієрархічні згорткові ознаки з різних рівнів глибокої нейронної мережі, а потім поміщують ознаки в кореляційні фільтри для регресії карти відгуку [31, 53]. Ці методи можна розглядати як комбінацію між глибоким навчанням і швидким поверхневим трекером на основі кореляційних фільтрів. Останнім часом все більше і більше сучасних глибоких трекерів використовують наскрізне навчання і тестування [6, 12, 55]. В роботі [11] запропоновано попередньо навчати згорткові мережі на декількох доменах, де кожен домен відповідає одній навчальній відеопослідовності. Стверджується, що існують деякі спільні властивості, які є бажаними для представлення цілей у таких областях, як зміни освітлення та переміщення об’єкта. Щоб виділити ці спільні риси, відокремлюють незалежну від домену інформацію від специфічних для домену шарів. Отриманий трекер досягає хороших показників відслідковування, хоча швидкість відслідковування становить лише 1 кадр/с. В роботі [1] досліджено глибокий регресор, який може передбачити місцезнаходження поточного об'єкта на основі його вигляду в останньому кадрі. Трекер отримує набагато вищу швидкість відслідковування (понад 100 к/с)

порівняно зі звичайними глибокими трекерами. Однак, все ще існує явний розрив у продуктивності цих систем.

Згорткова нейронна мережа на основі регіонів (RCNN) досягла хороших результатів для виявлення типових об'єктів, тому останнім часом було запропоновано багато методів виявлення об’єктів на основі RCNN. Однак, на відміну від виявлення загальних об'єктів, ділянки зображення об’єктів гірше відрізняються від фону, що спричинено внутрішньокласовими відмінностями при освітленні та оклюзії об’єктів, які, як було показано, негативно впливають на ефективність виявлення [45]. Іншими словами, дискримінатор об’єктів повинен більше покладатися на семантичну контекстну інформацію, щоб досягти хороших результатів.

У дослідженнях комбінування додаткових ознак розглядається як ефективний підхід для покращення характеристик RGB-зображень, оскільки можна ввести зовнішню семантичну інформацію. Реалізують інтегровану систему для використання можливостей RGBD зображень для виявлення та сегментації об'єктів [7]. Розробляють гістограму орієнтованих глибин (Histogram of Oriented Depths - HOD) для вдосконалення детектора [14]. Використовують маску сегментації для видалення дискримінаційних ознак для задач пошуку об’єктів [37]. Пропонують модель, яка керується маскою, для повторної ідентифікації об’єктів. Застосовують латентну модель з мінімальною ентропією, яка навчається за допомогою карти ідентифікації об'єкта, щоб мінімізувати випадковість локалізації [27]. Порівняно з оригінальними RGB-зображеннями, такі характеристики, як краї, градієнти, теплові карти, карти глибини, оптичний потік, карта ідентифікації об'єкта і маска сегментації, можуть забезпечити додаткове джерелом інформації, а також орієнтувати CNN на отримання більш кращих результатів, що є ключем до покращення ефективності виявлення.

Досить часто методи на основі CNN, будувалися на низькорівневих ознаках зовнішнього вигляду (краях, градієнтах та інше) і сформованих вручну ознаках.

Ці характеристики часто недостатньо сильні для досягнення задовільної точності, особливо для зображень з низькою роздільною здатністю. Крім того, у багатьох випадках методи можуть отримати користь від розширення інформації. Для отримання інформації про глибину зображення зазвичай потрібен датчик глибини, наприклад, лазерний, які не завжди є доступними. Було проведено кілька досліджень, які виявили можливості масок сегментації [10, 37, 44]. Сегментація зображення має за мету виведення маски сегментації, яка призначає семантичні мітки кожному пікселю зображення. Такі маски сегментації значно збільшують семантичну інформацію, і це може бути потужним інструментом для покращення роботи методів. Відповідно до цього пропонується простий, але ефективний підхід з використанням масок сегментації як зовнішнього каналу для забезпечення додаткового семантичного контексту в методах виявлення об’єктів. Результати експериментів показують, що метод перевершує базові методи, які використовують лише RGB-канали. Тобто використаємо високорівневий семантичний контекст, наданий масками сегментації. Такий контекст використаємо для того, щоб допомогти детекторам об’єктів отримати більше дискримінаційних ознак для виявлення об’єктів на фоні.

1.4 Аналіз рішень в задачах виявлення та відслідковування переміщення об’єктів на зображеннях

Для ефективного захоплення та відслідковування об’єктів використовують інтеграцію декількох ознак для отримання додаткової семантики. Додаткові ознаки, такі як градієнт, створені вручну функції, маски глибини та семантичної сегментації, були використані як джерело додаткової семантики для підвищення продуктивності згорткових нейронних мереж у широкому спектрі задач, таких як візуальне відслідковування об'єктів, пошук об’єктів, ідентифікація об’єктів.

Існує два підходи до об'єднання додаткових ознак з оригінальними RGB-об'єктами. Перший і найпоширеніший метод полягає у використанні зовнішніх шарів згортки для вивчення додаткових ознак, а потім конкатенації цих двох карт ознак на пізнішому етапі (рисунок 1.1).

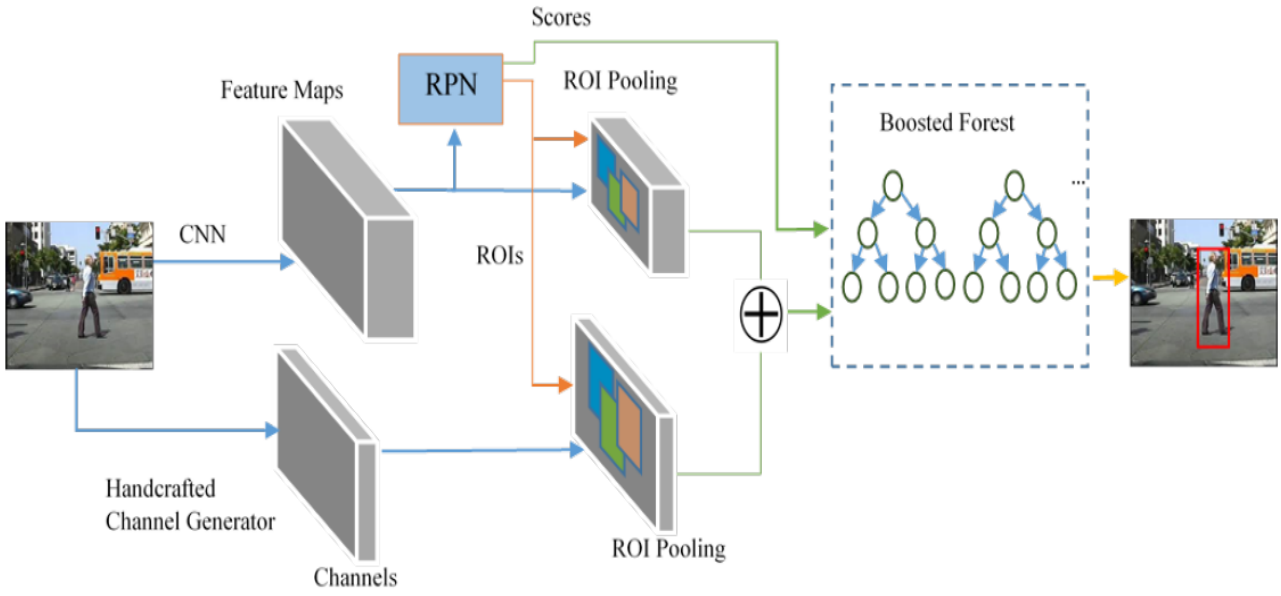


Рисунок 1.1 – Архітектура гібрідного канала детектування [40]

Наприклад, у роботі [40] запропоновано Hybrid Channel Detector, який вивчає представлення ознак каналу з додаткового контексту і об'єднує карти додаткових ознак з ознаками, отриманими опорною мережею.

Інший підхід полягає в тому, що додатковий семантичний контекст безпосередньо додається до вихідних RGB-каналів. Зауважимо, що перший підхід можна розглядати як окремий випадок другого підходу. Для другого підходу, коли перші кілька згорткових шарів використовують групові згортки, він стає першим випадком.

Автори [54] використовують маску сегментації для розділення оригінальних RGB-зображень на частину переднього плану та частину фону.

Модель навчається наскрізно з втратою багатозадачності. Використано RPN для генерації кандидатів і варіанти, пов'язані з позначеною людиною, як пропозиції з маскою, а пропозиції, пов'язані з немаркованою особою, як пропозиції без маски, оскільки лише частково позначено маски сегментації для позначених осіб на необробленому зображенні. Особливості векторів всіх кандидатів ідуть на втрату регресії, втрату класифікації та втрату ідентифікації. Лише карти ознак пропозицій з маскою подаються до гілок масок (рисунок 2.2).

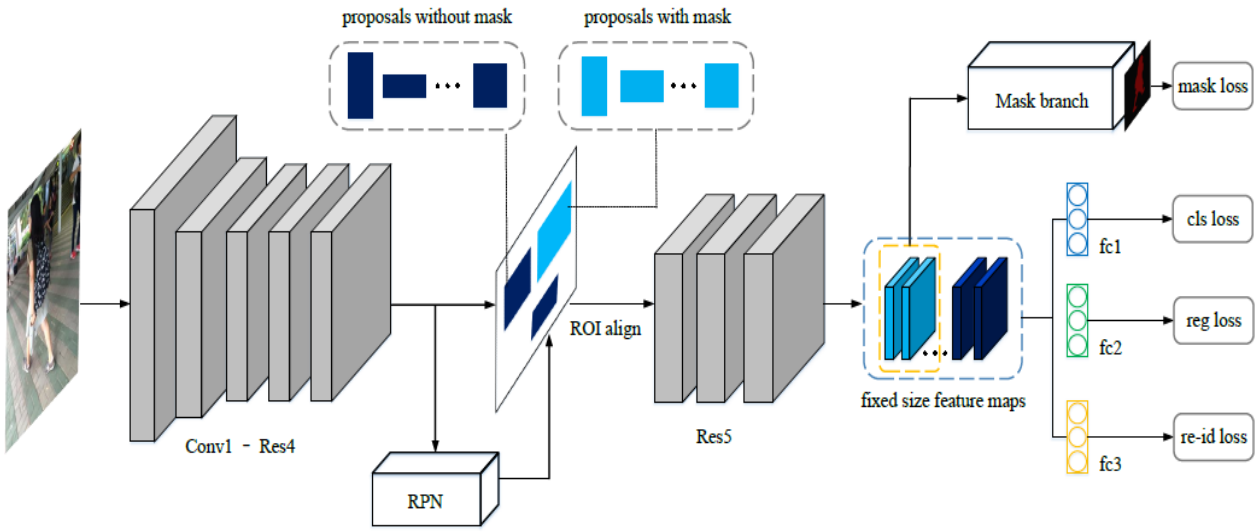


Рисунок 2.2 – Схема системи для пошуку людей за допомогою масок сегментації [54].

В роботі [15] пропонується семантичний сегментний інфузійний шар для кодування семантичних масок у спільні карти ознак. У роботі [23] реалізовано підмережі для навчання дискримінативних представлень ознак з декількох сигналів, які містять оригінальні RGB-зображення та зовнішній градієнтний контекст, і всі ознаки, витягнуті цими підмережами, об'єднуються у векторну класифікацію ознак. Недоліком є те, що доводиться вивчати вдвічі або навіть втричі більше додаткових параметрів, що призводить до збільшення обчислювальних витрат. Крім того, більшість з вищезгаданих робіт розробляють нову структуру нейронної мережі, яку важче адаптувати до різних систем виявлення, і залишається незрозумілим, чи є це оптимальним підходом для оригінальних і додаткових ознак, які вивчаються окремо у різних підмережах. Можна припустити, що пряма конкатенація додаткового каналу інформації з оригінальними RGB-каналами і подача їх разом у CNN може бути простішим, але ефективнішим підходом з точки зору використання додаткової інформації. Таким чином, для використання нової семантичної інформації вивчається лише кілька нових параметрів.

Згорткова нейронна мережа на основі регіонів (RCNN) продемонструвала ефективність використання пропозицій регіонів з глибокими нейронними мережами, і досягає належної продуктивності для виявлення типових об'єктів. В останні роки було запропоновано багато методів на основі RCNN.

Faster RCNN є набагато швидшою і гнучкішою альтернативою оригінальному RCNN, і стає одним з найбільш широко розповсюджених системуів виявлення об’єктів. Faster RCNN складається з двох етапів. На першому за допомогою RPN генерується ряд об'єктів-кандидатів, які обмежують об'єкти для знаходження. Далі ознаки кожного RoI витягуються мережею за допомогою RoIPool, який був в [5]. Потім всі ці ознаки подаються в регресори та класифікатори для остаточного прогнозування. Для прискорення виведення ознаки, що використовуються на цих двох етапах, можуть бути спільними.

Одноходові методи, такі як Single Shot Multi-Box Detector (SSD) [22] та YOLO [34], відмовляються від модуля пропозиції регіону для спрощення проєктування і дозволяють виявляти об’єкти за один хід, який безпосередньо прогнозує обмежувальні рамки та мітки категорій. У SSD вихідний простір обмежувальних рамок дискретизується на набір "рамок за замовчуванням" з різними співвідношеннями сторін і масштабами для декількох шарів згортки, і кожен шар примусово фокусується на прогнозуванні об'єктів певного масштабу. Таким чином, для прогнозування малих і середніх об'єктів система повинна використовувати ознаки з неглибоких шарів з невеликими рецептивними полями. Це може призвести до зниження продуктивності на малих і середніх об'єктах через нестачу семантичної інформації. Тому додавання додаткового семантичного контексту до цих методів на основі одиночних пробних детектувань може бути корисним для покращення їхньої роботи.

Метою семантичної сегментації зображень є передбачення мітки категорії для кожного пікселя зображення. Останнім часом згорткові мережі стали головним чинником прогресу в семантичній сегментації, і в цьому напрямку було досягнуто значних успіхів.

Серед методів, що базуються на CNN, популярною стала повна згорткова мережа (Fully Convolutional Network, FCN) [33]. FCN бере вхідне зображення довільного розміру і застосовує серію згорткових шарів. Потім мережа прогнозує попіксельні карти оцінки ймовірності для всіх семантичних категорій.

Використовуючи технології глибокого навчання, FCN забезпечує наскрізне рішення для точної семантичної сегментації. DeepLab [8], метод семантичної сегментації на основі FCN, досягає найсучасніших показників за останні роки. DeepLabv3+, остання версія системи DeepLab, використовує атомарну згортку для керування роздільною здатністю, а структура кодер-декодер розгорнута для подальшого уточнення результатів сегментації, особливо пікселів між границями об'єктів. Такий систему значно покращує продуктивність семантичної сегментації.

На відміну від семантичної сегментації, сегментація екземплярів має за мету ідентифікувати окремі екземпляри різних семантичних класів на зображенні. Оскільки зовнішній вигляд об'єктів з однієї категорії може бути дуже схожим, сегментація екземплярів часто вважається набагато складнішою задачею, ніж традиційна семантична сегментація.

Методи на основі FCN також широко використовуються і добре працюють у задачі сегментації екземплярів. Наприклад, пропонують повну згорткову сегментацію екземплярів (FCIS). FCIS виявляє та сегментує екземпляри об'єктів спільно та одночасно.

Розроблено систему Mask RCNN, який поєднує систему RCNN для виявлення обмежувальних рамок та систему FCN для задач з великою кількістю вихідних даних, що дозволило досягнути хорошої продуктивності для багатьох завдань, включаючи виявлення об'єктів, сегментацію екземплярів та оцінювання положення об’єктів.

1.5 Мета, задачі до реалізації програмної системи

Відповідно до проведеного аналізу, метою кваліфікаційної роботи бакалавра є розробка способу захоплення та трасування об’єктів на зображеннях в системах наведення цілі за допомогою нейронних мереж.

Для досягнення поставленої мети визначені наступні задачі дослідження:

– визначити послідовність застосування способу захоплення об’єкту на зображенні з наступним трасуванням його переміщення;

– розробити спосіб захоплення та слідкування за переміщення об’єкта на зображеннях за допомогою нейронної мережі;

– спроєктувати та реалізувати програмну систему запропонованого способу;

– провести експериментальне тестування розробленого способу.

Результатом виконання кваліфікаційної роботи бакалавра є створення способу захоплення та трасування об’єктів на зображеннях в системах наведення цілі за допомогою нейронних мереж.

Розділ 2 Спосіб захоплення та трасування об’єктів на зображеннях в системах наведення цілі за допомогою нейронних мереж

2.1 Спосіб захоплення об’єктів на зображеннях в системах наведення цілі за допомогою нейронних мереж

На основі проведеного аналізу було визначено, що результати сегментації зображень можуть бути корисними з точки зору продуктивності глибоких згорткових нейронних мереж. Однак процедура використання маски сегментації може бути дуже різною. Так для спільного навчання задач виявлення та сегментації використаємо багатогілкову нейронну мережу, таким чином: спільна магістральна мережа може навчатися дискримінативним ознакам з обох задач одночасно. Результати аналізу показують, що метод виявлення об’єктів, навчений спільно із задачею сегментації, може отримати незначне покращення порівняно з детектором, навченим для однієї задачі. Задача виявлення об'єктів формулюється як задача сегментації, тоді початкова локалізація об'єкта може бути уточнена за допомогою масок сегментації.

Однак, вищезазначені методи можуть мати ряд обмежень:

1. Обмеженість даних: Для навчання декількох задач, наприклад, виявлення обмежувальних рамок та сегментації екземплярів, використаємо процедура спільного навчання. Однак для більшості наборів даних виявлення об'єктів попіксельні анотації недоступні. Таким чином, така процедура навчання навряд чи може бути адаптована до тих наборів даних, які не мають міток пікселів.

2. Стійкість: Для видалення глибоких ознак декілька задач використовують одну і ту ж саму магістральну мережу. Ця стратегія значно підвищує ефективність мережі. Однак ознаки, отримані різними задачами, можуть не вплинути на продуктивність інших задач.

Для модуля сегментації використаємо два готові методи генерації масок сегментації: DeepLabv3+ для семантичної маски сегментації та Mask RCNN для маски сегментації екземплярів. Потім і семантична маска сегментації, і маска сегментації екземплярів, згенеровані модулем сегментації, переносяться у бінарну маску сегментації (рисунок 2.1).

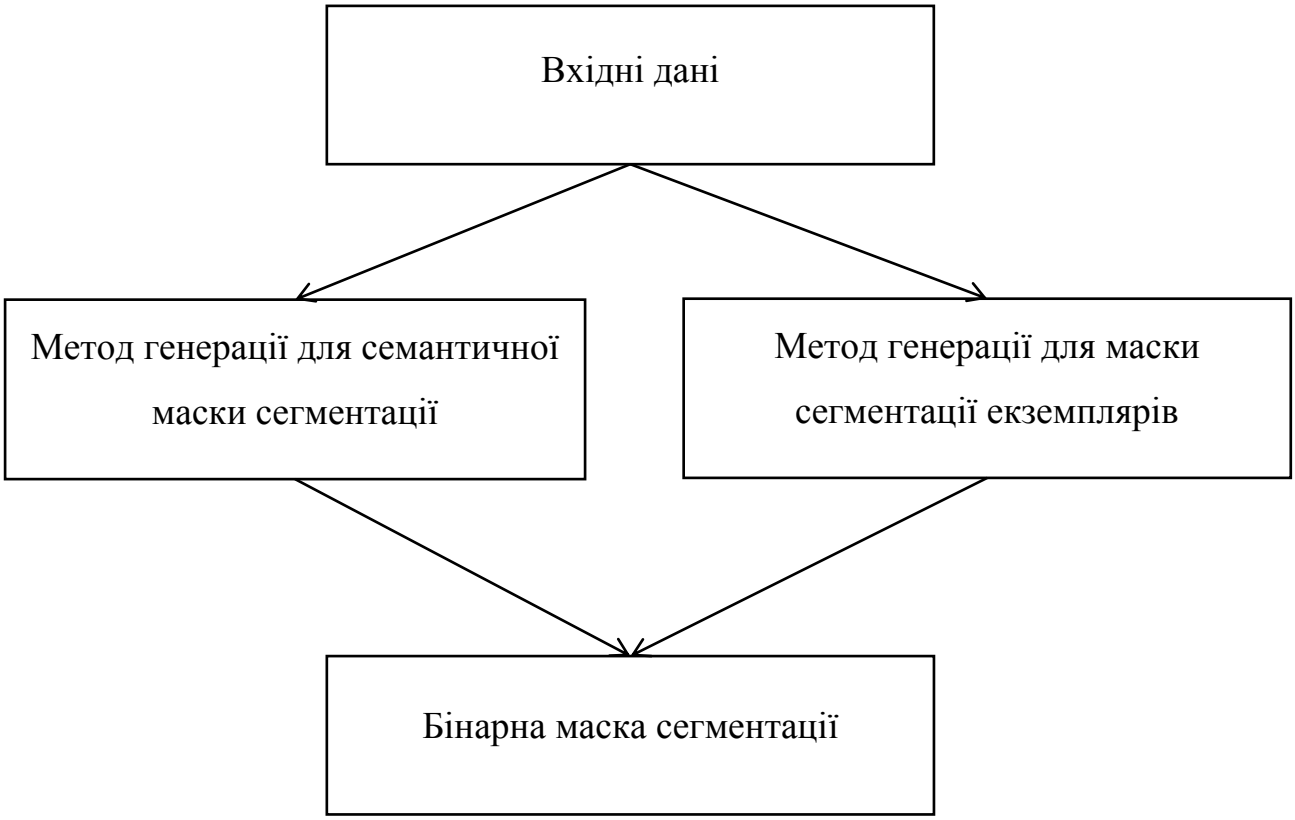


Рисунок 2.1 – Модуль сегментації

2.2 Виявлення об’єктів за допомогою маски

Важливо зазначити, що Mask RCNN може генерувати маски екземплярів. Використаємо лише семантичні маски для перевірки ефективності підходу.

Після того, як маски семантичної сегментації згенеровано, їх інтегруємо з оригінальними RGB-зображеннями та подамо до модуля виявлення (рисунок 2.2). Реалізуємо метод на двох популярних загальних системах виявлення об'єктів – Faster RCNN та SSD.

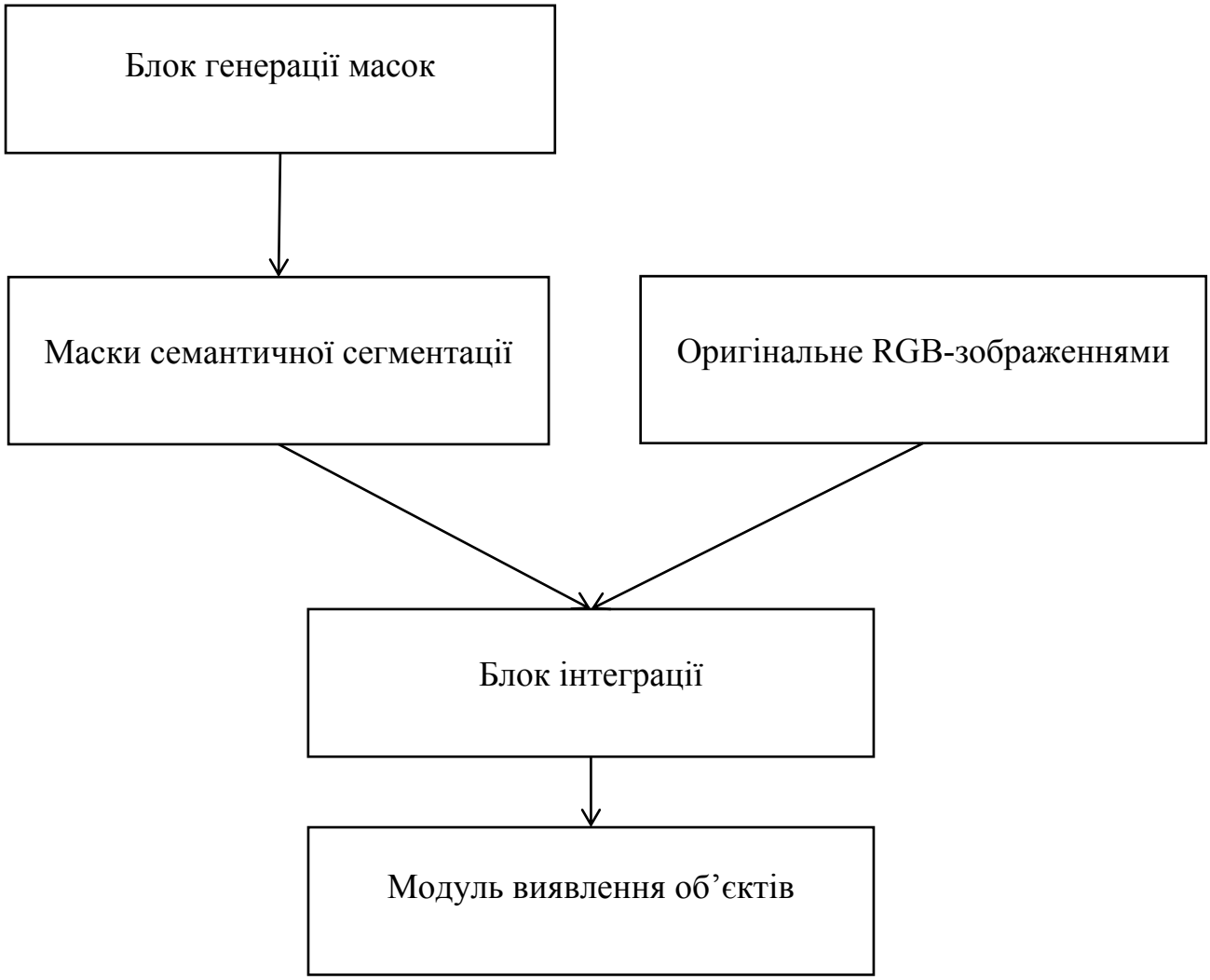


Рисунок 2.2 – Структурна схема блоку виявлення об’єктів

Щоб дослідити ефективність вхідних масок сегментації, використаємо декілька налаштувань для DeepLabv3+ та Mask RCNN, щоб згенерувати маски сегментації різної якості.

Для DeepLabv3+ розробимо два типи масок сегментації:

– бінарна маска семантичної сегментації;

– оціночна маска семантичної сегментації з балами.

Вони позначаються як $Mask_{binary}$ та $Mask_{score}$ відповідно. Оціночна семантична маска сегментації використовують одну і ту саму опорну мережу для отримання ознак (рисунок 2.3).

Однак, $Mask_{binary}$ не містить інформації про оцінки, яку зберігає $Mask_{score}$.

Бінарна семантична маска сегментації визначається як:

$$Mask_{binary} = f\left(\frac{e^{X_i}}{\sum_{j=1}^{T} e^{X_i}}\right), \quad \text{i=1,...,T} \tag{2.1}$$

Оціночна маска семантичної сегментації $Mask_{score}$ визначається як:

$$Mask_{score} = f\left(\frac{e^{X_i}}{\sum_{j=1}^{T} e^{X_i}}\right), \quad \text{j=1,...,T} \tag{2.2}$$

де X – матриця згенерована моделлю DeepLabv3+.

$Mask$ – елемент у матриці сегментації;

T – кількість елементів матриці сегментації.

Рисунок 2.3 – Модуль сегментації

Використовуємо $f(x)$ для перетворення маски сегментації в єдину двійкову матрицю.

Для маски RCNN використаємо декілька різних магістральних мереж для генерації масок сегментації екземплярів різного рівня якості. Наприклад, використовуємо магістральні мережі RestNet

2.3 Розпізнавання об’єктів за допомогою глибоко вивчених семантичних масок

Маски сегментації екземплярів можна легко перенести в обмежувальні рамки. Якщо збережемо інформацію про екземпляри, методи (люди) зможуть легко підбирати маски сегментації екземплярів під час навчання, що може призвести до поганих показників виявлення під час тестування без масок сегментації екземплярів. Тому маски сегментації екземплярів перетворюються в одну бінарну маску сегментації *Mask*. Іншими словами, під час навчання і тестування детектора використовуються лише семантичні маски.

Маска сегментації екземплярів *Mask*

$$Mask = \sum_{i=1}^{T} f\left(Mask_{type,i}\right), \quad \text{i=1,...,T} \tag{2.3}$$

де $Mask_i$ – маска сегментації для *і-го* екземпляра на одному зображенні, згенерованому за допомогою маски RCNN, з відповідним типом маски сегментації *type*;

T – кількість екземплярів на зображенні.

Для порівняння з ефективністю семантичних масок сегментації та зменшення обчислювального комплексу, більшість масок сегментації переводяться в єдину бінарну маску сегментації для кожного зображення за $f(x)$, окрім семантичної маски з оцінками, яка зберігає інформацію про оцінку. Тобто, кожен піксель вказує на ймовірність певної категорії. Бінарна маска сегментації, яка подібна до механізму уваги, змушує методи приділяти більше уваги тим областям, які виділені семантичними масками сегментації. У той же час, такі маски бінарної сегментації можуть природним чином розділити одне зображення на частину переднього плану і частину фону, що може допомогти детекторам вивчити високодискримінаційні ознаки для класифікації цільових об'єктів і фону.

Початкові маски сегментації, згенеровані модулем сегментації, перетворюються на бінарну маску за допомогою $f(x)$:

$$f(x)=\begin{cases}0, & x_{i,j} \le Th; \\ 1, & x_{i,j} > Th, \end{cases} \tag{2.4}$$

де Th – поріг оцінки для фільтрації шумів передбачення;

x – вихідна матриця маски сегментації.

Кожен елемент x_{ij} в матриці вказує на ймовірність категорії. Таким чином, модуль сегментації використаємо для генерації необроблених масок сегментації. Потім семантична маска сегментації та маска сегментації екземплярів переносяться в одну двійкову маску сегментації, яка буде подана в модуль виявлення (рисунок 2.4).

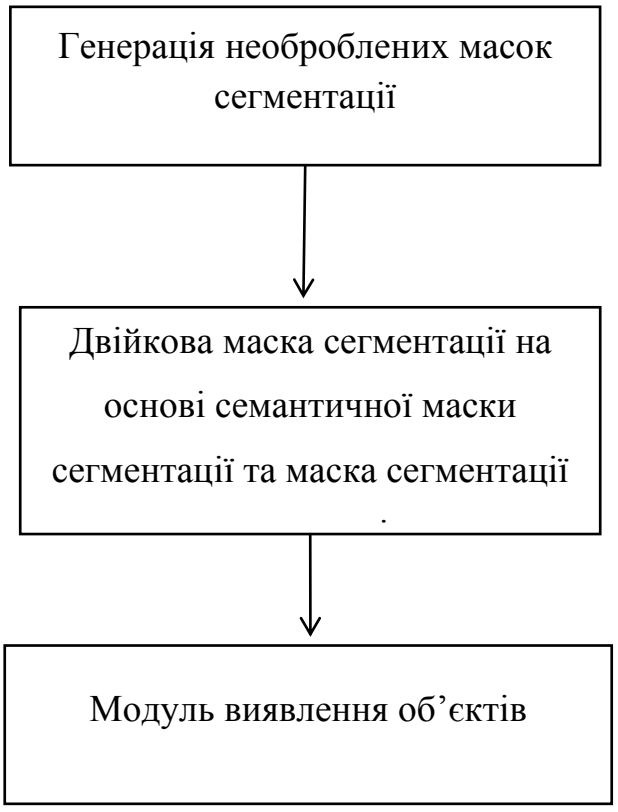


Рисунок 2.4 – Формування бінарної маски сегментації

2.4 Система виявлення об’єктів

Щоб оцінити генеративну здатність методу, реалізуємо метод на двох відомих системах виявлення Faster RCNN та SSD. Під час навчання замінюємо 3-канальні RGB зображення на 4-канальні RGBM зображення:

$$\mathbb{R}^3_{RGB} \xrightarrow{f\left(\mathbb{R}^3_{RGB}, R_{Mask}\right)} \mathbb{R}^4_{RGBM} \tag{2.5}$$

Таким чином розширимо вихідний простір додатковою атрибутивністю.

У системи Faster RCNN спочатку RPN генерує набір пропозицій прямокутних об'єктів, які є областями інтересу (RoI). Потім ознаки RoI витягуються мережею. Далі ознаки подаються в регресор обмежувального поля і класифікатор категорій, щоб передбачити локалізацію цілі і мітку класу. Ці дві задачі навчаються спільно, таким чином визначається функція втрат Faster RCNN:

$$Loss = \frac{1}{N_{bin}} L_{bin} + \lambda \frac{1}{N_{reg}} L_{local} \ , \tag{2.6}$$

де L_{bin} та L_{scr} – втрати для бінарної класифікації та згладжені втрати L_1 для регресії з обмеженою зоною;

N_{bin} та N_{reg} – параметри нормалізації, які визначаються розміром зони обробки та кількістю пропозицій відповідно;

λ – коефіцієнт, що врівноважує ці втрати.

У структурі SSD для регресії цільових обмежувальних рамок генерується низка попередньо визначених "рамок за замовчуванням". Щоб врахувати цільові об'єкти різного масштабу і форми, ці створені за замовчуванням рамки також мають різні співвідношення сторін і розміри. Класифікація та задача локалізації навчаються разом, тому функцію втрат SSD можна подати у вигляді:

$$Loss_s = \frac{1}{N}\left(L_{bin} + \lambda L_{local}\right), \tag{2.7}$$

де L_{conf} та L_{loc} – втрати для класифікації та згладжені втрати L_1 для зони обмеження;і

N – кількість позитивних стандартних клітинок, які збіглися з передбаченими клітинками,

λ – постійний ваговий коефіцієнт для збереження балансу між втратами.

У модулі виявлення RGBM зображення подаються для оброблення у методи визначення об’єктів, після чого прогнозуються межі об'єктів і оцінки категорій.

2.5 Глибоке відслідковування в реальному часі за допомогою корекційної адаптації домену

Візуальне відслідковування – одна з фундаментальних задач комп'ютерного зору. Протягом останнього десятиліття, з розвитком глибокого навчання, все більше алгоритмів відслідковування отримують перевагу від глибоких нейронних мереж, наприклад, згорткові нейронні мережі та рекурентні нейронні мережі. Незважаючи на успіх і цій сфері, все ще існує дилема, яка полягає в тому, що глибоке навчання підвищує точність відслідковування, але ціною високої обчислювальної складності. Як наслідок, більшість добре працюючих трекерів зазвичай страждають від низької ефективності. Вони досягли дуже високої швидкості відслідковування, але не можуть перевершити поверхневі методи.

Використаємо ефективний алгоритм доменної адаптації. Алгоритм відслідковування, названий корекційною доменною адаптацією (Corrective Domain Adaptation – CDA) [30], переносить ознаки з області класифікації в область відслідковування, де окремі об'єкти, а не категорії зображень,

використовуються як навчальні вибірки (рисунок 2.5). Перевага доменної адаптації:

1. Неглибокий візуальний трекер, алгоритм KCF [38], який використаємо в цій роботі, може виділити більш інформативні глибокі ознаки з отриманого зображення.

2. Адаптацію можна також розглядати як процес зменшення розмірності, який видаляє надлишкову інформацію для відслідковування. Це значно зменшує кількість каналів глибинних ознак і призводить до значного збільшення швидкості відслідковування.

3. Адаптація вводить невеликі допоміжні гілки CNN, які можуть легко коригувати прогнози неглибоких візуальних трекерів.

Рисунок 2.5 – Переваги використання доменної адаптації

Використаємо інформацію про категорію цілі відслідковування в CDA наступним чином. Для певної категорії об'єктів "гілки" CNN налаштовуються для корекції квадратів відслідковування, і таким чином досягається вища точність відслідковування.

Експерименти показують, що алгоритм CDA працює зі швидкістю близько 35 кадрів в секунду, досягаючи при цьому точності відслідковування, порівнянної з найсучаснішими трекерами. Крім того, враховуючи інформації про категорію цілі, що відслідковується, коригувальні гілки CNN призводять до значного підвищення точності відслідковування, зберігаючи при цьому швидкість відслідковування майже незмінною.

2.6 Структура мережі з адаптацією домену

В основу алгоритму адаптації покладено алгоритм відслідковування Hierarchical convolutional features – HCF. У HCF глибокі ознаки спочатку витягуються багатошаровою мережею VGG-19. Остаточний прогноз трекінгу отримується методом зваженого голосування. У роботі пропонуємо виконати адаптацію домену простим способом. До кожного шару ознак прикріплюється "гілка відслідковування". Гілка відслідковування фактично є шаром згортки, який зменшує кількість каналів у 8 разів і зберігає розмір карти ознак незмінним. Потім шар згортки навчається шляхом мінімізації функції втрат, пристосованої для відслідковування, як описано нижче.

Навчання параметрів у трекінгу відбувається аналогічно до алгоритму Single Shot MultiBox Detector (SSD). При навчанні вихідні шари VGG-19 фіксуються, і кожна "гілка стеження" навчається незалежно. Для отримання завершеного кола навчання адаптована функція використаємо для регресії розташування об'єктів та їхніх оцінок об'єктності. Для конкретних категорій пропонується використовувати "гілки відслідковування" з корекцією початкових полів відслідковування.

У SSD для регресії прямокутників об'єктів генерується декілька стандартних комірок. Крім того, для розміщення об'єктів у різних масштабах і формах, поля за замовчуванням також відрізняються за розміром та співвідношенням сторін.

Нехай $m_{i,j} \in \{1, 0\}$ – показник відповідності і-го блоку за замовчуванням j-му блоку базової оцінки.

Функція втрат SSD записується таким чином:

$$Loss(m, pred) = \frac{1}{N}\left(L(m,c) + \lambda L(m, pred)\right), \tag{2.8}$$

де c – категорія блоку за замовчуванням;

$pred$ – блок передбачення.

Однак завдання візуального відслідковування суттєво відрізняється від завдання виявлення. Тому адаптуємо функцію втрат для алгоритму KCF, де і розмір об'єкта, і розмір вікна KCF є фіксованими. Вікно KCF відіграє аналогічну роль, що і поля за замовчуванням у SSD, тому потрібно генерувати лише один тип полів за замовчуванням, а функція втрат місцеположення спрощується:

$$Loss_{loc} = \sum_{u \in (x,y,w)} m f_a\left(l_u - g_u\right), \tag{2.9}$$

де u, g_u – геометричні параметри нормалізації.

До уваги береться тільки переміщення {x, y} і немає необхідності для нормалізації поля фону.

Концепція доменної адаптації в цій роботі відрізняється від концепції, визначеної в MD-net , де різні відеопослідовності розглядаються як різні домени і, таким чином, для їх обробки навчаються декілька повністю з'єднаних шарів. Це відбувається тому, що в MD-net вибірка навчальних прикладів відбувається за принципом ковзаючого вікна. Об'єкт, позначений як негативний в одній області, може бути обраний як позитивний зразок в іншій області. Враховуючи,

що номер навчального відео дорівнює C, а розмірність останнього шару згортки дорівнює d_c, MD-net навчає C незалежних $d_c \times 2$ повністю з'єднаних альтернативно шарів. На відміну від MD-net, використаємо домен, який відноситься до загального домену візуального відслідковування KCF. Він призначений для імітації вкладу KCF у візуальне відслідковування. Різні цілі відслідковування розглядаються як одна категорія, тобто як об'єкти.

Під час навчання місцезнаходження об'єкта і передбачення об'єкта регресують для мінімізації згладжених втрат. Складність навчання зменшується, а відповідна збіжність стає більш стабільною.

Використання додаткових каналів, таких як карта глибини, оптичний потік, карта значущості, для покращення оригінальних RGB-зображень є підходом для покращення детекторів об'єктів, який досить широко використовується. Для вивчення ознак каналів, таких як семантична маска сегментації, край і теплова карта каналу для виявлення пішоходів, встановлено, що інтеграція зовнішніх ознак в мережу може підвищити ефективність роботи детекторів на зображеннях як з низькою, так і з високою роздільною здатністю, таким чином підвищуючи точність виявлення.

Гістограма глибини, яка кодує напрямок зміни глибини до вихідних RGB-зображень, з додатковим каналом глибини при використанні в детекторі може відрізнити об'єкти переднього плану від фону. Крім того, багато попередніх досліджень довели, що дії пов'язані з відео, такі як рух і оптичний потік, можуть бути корисними для покращення продуктивності для задач комп'ютерного зору на основі потоку зображень.

Керування оптичним потоком вивчається шляхом введення оптичного потоку в згорткову нейронну мережу. Це може дозволити нейронній мережі виокремлювати часову інформацію. Однак, генерування додаткових функцій, таких як маска сегментації, глибина, карти релевантності та оптичний потік не тільки забирає багато часу, але також вимагає значних обчислювальних ресурсів. Тому компроміс між точністю і швидкістю серед цих методів стає проблемою для більшості застосувань. Хоча люди можуть легко ловити рухомі об'єкти,

навіть якщо об'єкти мають дуже малі розміри. Враховуючи це явище, встановлюємо, що рухи можна розглядати як рухомі ознаки для задачі виявлення об'єктів на відеозаписах. Таким чином, однією з цілей є застосування простого, але ефективного способу використання інформації про рух без значних обчислювальних витрат. Для цього використаємо додаткові канали для керування згортковими нейронними мережами та вивчення попередніх рухів. Ця операція значно покращить здатність детекторів розпізнавати рухомі малі об'єкти.

Припустимо, що зображення для поточного відеокадру *Fi*, крок *Step*. Тоді базову карту руху *Mv* можна визначити так:

$$Mv = \frac{\left|I_i - I_{i-step}\right| + \left|I_i - I_{i+step}\right|}{2} \tag{2.10}$$

Просте додавання базової карти руху до RGB зображення може призвести до таких труднощів:

1. Нейронні мережі можуть бути легко переналаштовані на канал руху.

2. Нерухомі об'єкти без руху мають високу ймовірність бути пропущеними.

Тому, щоб уникнути таких проблем, зміщений канал руху генерується як підтримка базової карти руху. Зсув руху між і-м та j-м кадрами можна обчислити за формулою:

$$Shift = \frac{f(side_i)}{4}, i = 1...4. \tag{2.11}$$

Тут $f(side_i)$ усереднене значення зміщення поточного кадру (зображення) відносно попереднього кадру.

Зміщення визначається на основі обробки зміни пікселів на границях зображення.

Після підготовки каналів руху кадри будуть об'єднані з вихідними RGB-каналами, тому вхідні дані складаються з п'яти каналів:

$$R^5\left(R^3{}_{RGB}, Mv, Shift\right) \tag{2.12}$$

Базовий та зсувний шаблони руху кодують попередні рухи у вхідному зображенні, які можуть допомогти нейронній мережі звернути увагу на рухомі об'єкти. Крім того, ці попередні патерни руху дозволяють мережі пропозицій ігнорувати загальні фонові області, такі як небо та будівлі в наборі даних для експериментів.

Було встановлено, що багатозадачне спільне навчання може покращити продуктивність виявлення. Так система Mask RCNN, яка одночасно навчає детектор об'єктів та сегментатор екземплярів, порівняно з Faster RCNN досягає кращої продуктивності завдяки спільному навчанню. Однак маски сегментації на рівні екземплярів недоступні для більшості наборів даних з виявлення об'єктів, що робить цю стратегію складною в реалізації. Деякі дослідження використовують готові методи для генерації семантичної сегментації масок з розв'язання високорівневих задач технічного зору, таких як повторна ідентифікація людини чи виявлення об'єктів.

Зазвичай, маска використовується як зовнішній канал, щоб направляти нейронні мережі для вивчення кращих зображень, що, може розглядатися як певний тип механізму уваги. Однак генерування масок потребує значних обчислювальних витрат для обрахунку масок як на етапі навчання, так і на етапі виведення. Тому безпосереднє використання області обмежувальної рамки як маски сегментації для нагляду за зовнішньою гілкою стає альтернативою. Тому використаємо механізм уваги, який безпосередньо використовує всю область зображення як маску уваги для виявлення пішоходів. Для цього використовуємо зовнішню гілку, яка поділяє ту ж саму магістральну мережу з гілкою виявлення для прогнозування згенерованих масок екземплярів. Важливо зазначити, що ця зовнішня гілка встановлюється лише на етапі навчання для того, щоб

спрямувати магістральну мережу на вивчення піксельного рівня, і не буде активована на етапі тестування. Таким чином, ніяких зовнішніх вхідних даних або додаткових обчислювальних витрат під час виведення не вимагатиметься. Для генерації масок уваги, оскільки не ставляться для мережі вимоги передбачити високоточну маску сегментації на етапі тестування, цього достатньо. Однак, просте використання області обмежувальної рамки як маски екземплярів може призвести до надмірного використання масок уваги та до появи надлишкової фонової інформації. Тому пропонуємо метод генерації масок на основі руху метод генерації масок на основі руху. Базові шаблони руху спочатку витягуються з і-го кадру *Ii* та його відносного кадру *Iistep*. Потім рух на рівні екземпляра обрізається рамкою. Крім того, карта руху екземпляра переноситься у бінарні карти з одночасною фільтрацією шумів. Маска обчислюється на основі набору репрезентативних точок, вибраних з краю чорнової карти. Після доопрацювання можна отримати маски уваги для руху без зайвих обчислень.

Більш кращі маски можна отримати за допомогою людських анотацій, однак в роботі це не буде проведено. Мета запропонованого методу полягає в тому, щоб забезпечити простий базовий підхід до набору даних і дослідити можливі напрямки для задач виявлення малих об'єктів.

Однак не слід ігнорувати такі проблеми:

1. Патерни руху не можна отримати з нерухомих об'єктів, таких як люди, що стоять, припарковані транспортні засоби.

2. Маски руху низької якості можуть негативно вплинути на ефективність виявлення.

Тому використовуємо комбінацію масок – масок на основі руху і масок на основі обмежувальних рамок для навчання зовнішньої маски, тобто маски на основі обмежувальних рамок замінюють маски на основі руху. Така стратегія може значно підвищити надійність детектора, особливо для нерухомих об'єктів.

2.7 Використання масштабу для адаптації домену

Адаптація домену в методі CDA – це шар згортки. При створенні шару відразу виникає питання, як вибрати правильний розмір фільтрів. Важко знайти оптимальний розмір фільтра для всіх шарів особливостей. Тому одночасно навчимо адаптаційні фільтри в різних масштабах. Карти результатів з різними розмірами фільтрів потім об’єднують відповідно, як показано на рисунку 2.6.

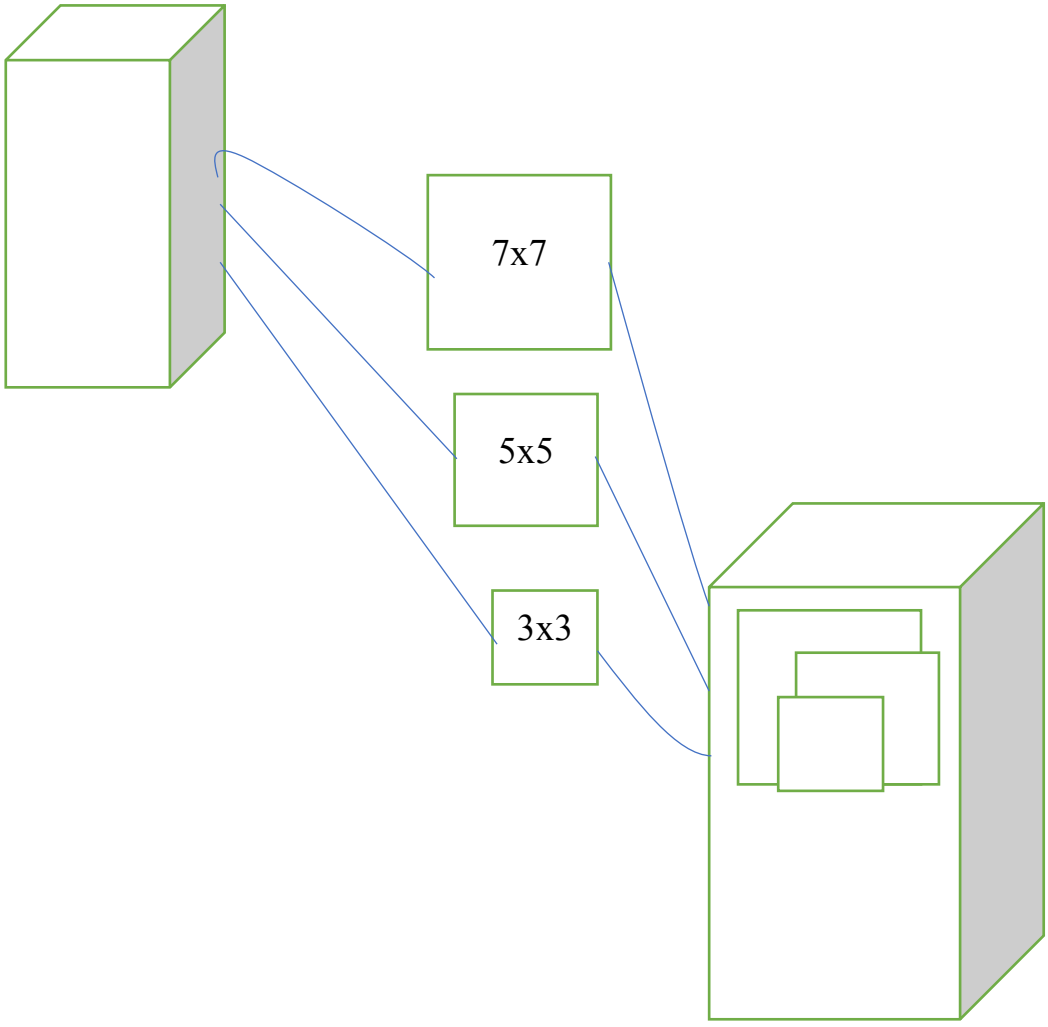


Рисунок 2.6 – Схема об’єднання карт з різних масштабів

Таким чином, на вхід KCF-трекера надходять глибинні ознаки з різних масштабів. На практиці використовуємо фільтри 3 × 3 і 5 × 5 для всіх трьох шарів об'єктів. Приймаючи загальну кількість каналів, що дорівнює K, кожен тип фільтра генерує K каналів і, таким чином, зменшує кількість каналів. Зі

зменшенням кількості каналів швидкість відслідковування значно зростає. Швидкість трекера KCF різко падає зі збільшенням кількості каналів. Після адаптації кількість каналів зменшено у 6 разів, що прискорює роботу трекера у 3 рази.

2.8 Корекційна адаптація домену

У візуальному відслідковуванні достовірна інформація про ціль надається на першому кадрі, тоді як за багатьох обставин ця інформація може бути неоднозначною. Наприклад, якщо у послідовності кадрів потрібно відслідкувати автомобіль.

На першому кадрі може бути видно лише задню частину автомобіля. Таке просте визначення цілі зазвичай призводить до неоднозначності, коли позиція цілі суттєво змінюється, важко оцінити результати відслідковування. Чітко визначена ціль відслідковування зазвичай у візуальному відслідковуванні досить часто відсутня через дуже обмежену інформацію. Тобто через обмежувальну рамку, яка задана на першому кадрі. Тому вводимо категорію об'єкта в задачі візуального відслідковування. Тобто трекер відслідковує об'єкт, знаючи обмежувальну рамку цілі на першому кадрі, а також категорію цілі.

Враховуючи конкретну цільову категорію, використовуємо запропоновану стратегію навчання, з використання визначення об’єкту, для навчання набору гілок CNN на вибірках з цієї категорії.

Потім використовуємо ці гілки для корекції прогнозу глибокого трекера. Високорівнева концепція відслідковування – виявлення – об'єднання проілюстрована на рисунку 2.7.

Допоміжні гілки CNN використовуються для регресійного аналізу об'єктної області. Всі гілки регресії не є обчислювально складними порівняно з усією мережею. Додаткове обчислювальне навантаження не є значним.

Область відслідковування отримується за тією ж стратегією, що і HCF. Водночас, деякі обмежувальні рамки виявлення також генеруються за

допомогою SSD. Після видалення некваліфікованих областей виявлення, середній масштаб і співвідношення сторін результатів виявлення використовуються для корекції поточної області відслідковування.

Враховуючи межі відслідковування та межі виявлення, CDA об'єднує результати у простий, але ефективний спосіб.

Вихідний блок відслідковування коригується блоками виявлення. Корекція зазвичай є корисною завдяки більш чіткому визначенню цільової категорії та добре навченому детектору (рисунок 2.7).

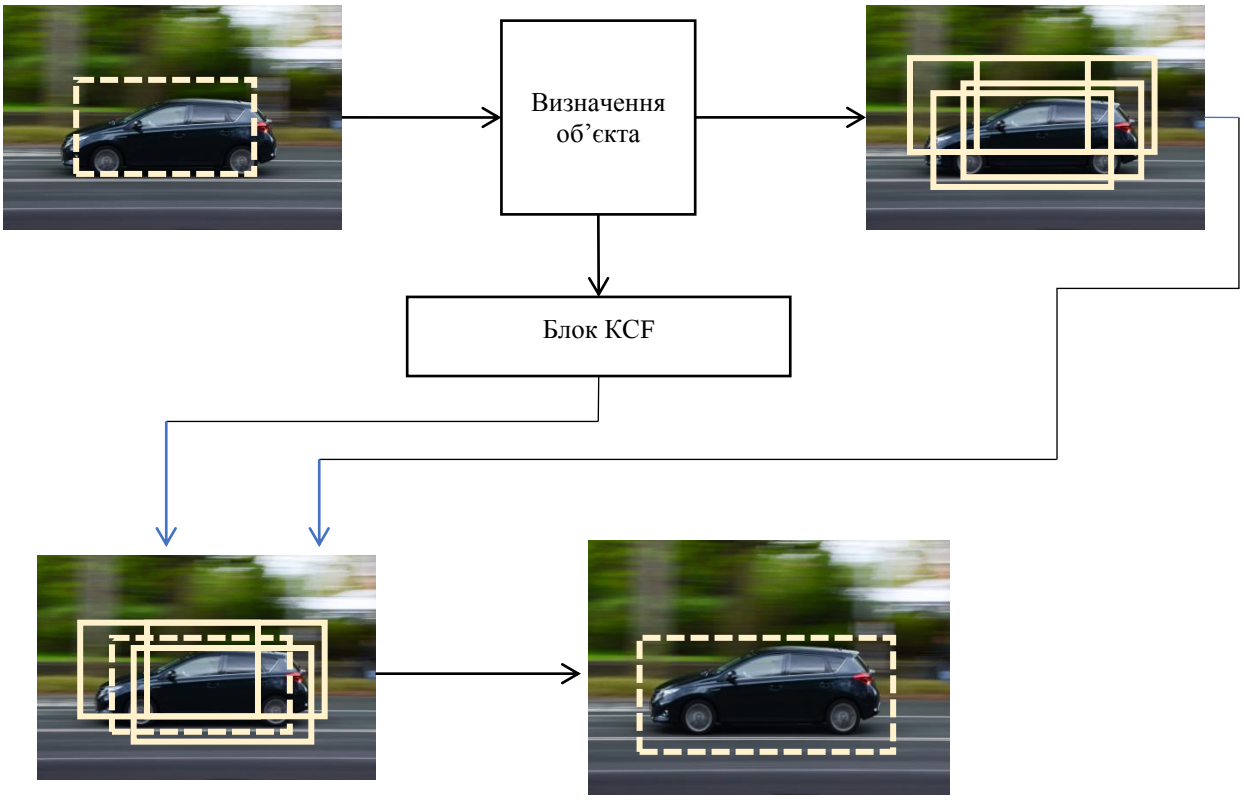


Рисунок 2.7 – Схема відслідковування – виявлення – об'єднання

Для визначення об’єктів розглянемо анотації, які використовуються в корпусах даних.

Для анотації особи використовуємо протокол анотації для виявлення пішоходів. Пішоходи анотуються шляхом проведення лінії від верхівки голови

до низу ступні, а потім генерується обмежувальна рамка з фіксованим співвідношенням сторін.

Анотуючи екземпляри у такий спосіб із фіксованим співвідношенням сторін, можуть виникнути деякі проблеми:

1. Особи з різними позами мають різне співвідношення сторін, тому просте фіксування співвідношення сторін може включити непотрібний фон, або виключити людські частини з обмежувальних рамок (рисунок 2.8). Крім того, іноді такий підхід може призвести до поганого вирівнювання.

Рисунок 2.8 – Протокол анотації людини

2. Фіксоване співвідношення сторін не може застосовуватися до "сидячих людей", "велосипедистів" і т.д., тому ці люди позначені як ігноровані, що не дозволяє використовувати фіксоване співвідношення сторін. Це шкодить різноманітності набору даних і може негативно вплинути на продуктивність детектора (рисунок 2.9).

Як наслідок, не будемо дотримуватись правила фіксованого співвідношення сторін у вибраному наборі даних. Вся частина об'єкта екземпляра включена в прямокутну обмежувальну рамку .

Рисунок 2.9 – Ігноровані області в наборах даних

Анотація транспортного засобу. Оскільки транспортні засоби є жорсткими об'єктами, протоколи анотацій для такого типу цілей майже ідентичні в різних наборах даних.

Ігнорування анотації регіону. Окрім справжніх позитивних навчальних вибірок, у багатьох наборах даних часто можна побачити деякі області, позначені як ігноровані, що пов'язано з наявністю низької роздільної здатності, фальшивих об'єктів на плакатах або занадто переповнених, щоб їх можна було анотувати. Однак, зважаючи на різноманітність різних протоколів анотування, деякі ігноровані регіони є непотрібними в існуючих наборах даних. Анотації було побудовано на основі масок семантичної сегментації наданих, та успадкованих від визначень на рівні пікселів і не підходять для задач розпізнавання. Часто існує велика кількість ігнорованих областей у наборі даних включаючи дорожні знаки, світлофори, люди в транспортних засобах, яких важко розпізнати, тощо. Також у наборі даних, більшість транспортних засобів малих або навіть середніх розмірів не позначені. Області, що включають такі випадки, розглядаються як ігноровані області.

Така політика анотування викликає обмеження:

1. Повинна бути зроблена обробка для ігнорованих областей.

2. Різноманітність набору даних буде порушено, і це може в подальшому вплинути на продуктивність детектора.

3. Не можна використовувати для оцінки здатності CNN виявляти малі та скупчені цілі.

Тому в наборі даних, які будуть використанні для експериментів випадки, що не відповідають меті експерименту, наприклад із сильним затіненням або фальшивими об'єктами на знімках будуть позначені як об'єкти ігнорування.

Висновок до розділу 2

Запропоновано ефективний алгоритм адаптації домену для візуального відслідковування об'єктів. Алгоритм відслідковування, переносить ознаки з

області класифікації в область відслідковування, де окремі об'єкти, а не категорії зображень, використовуються як навчальні зразки. Крім того, адаптація також природно використовується для введення концепції об'єктності у візуальне відслідковування. Це усуває невизначеність мети в задачах візуального відслідковування, і показує емпіричну перевагу більш чітко визначеної задачі. Відповідно це дозволило отримати такі результати:

1. Використано ефективний метод адаптації домену для візуального відслідковування. Адаптація не тільки призводить до збільшення швидкості відслідковування в реальному часі, але також зберігає високу точність відслідковування, яку можна порівняти з найсучаснішими трекерами.

2. Для певного типу цілей, що відслідковуються, використано гілки CNN, які спочатку навчені адаптувати глибоку ознаку до області візуального відслідковування, для корекції початкових блоків відслідковування. У рамках складної системи виведення точність відслідковування значно підвищується.

3. Перевага коригувальної адаптації емпірично доводить, що більш чітко визначена ціль відслідковування, а не просто обмежувальна рамка, може суттєво покращити процес відслідковування. Іншими словами, ця дозволило знайти шлях до вирішення проблеми погано поставленої задачі візуального відслідковування.

Розділ 3 Експериментальна перевірка способу захоплення та трасування об’єктів на зображеннях за допомогою нейронних мереж

Задача виявлення об’єктів має за мету передбачити обмежувальні рамки всіх об’єктів на зображенні, в приміщенні чи на вулиці, з частин об’єктів чи цілого об’єкту. Таким чином, оцінюємо використаний метод на наборі даних CrowdHuman.

Представляємо набори даних і протоколи оцінювання, які використовуємо в цій роботі, а потім наводимо деякі деталі реалізації. Покажемо кількісні та якісні результати методу на основі двох систем виявлення, Faster RCNN та SSD на наборі даних CrowdHuman. Експериментально проаналізуэмо ефективність масок сегментації.

Набір даних CrowdHuman складається з 15 тис. навчальних та 4 тис. тестових зображень. З точки зору щільності, в середньому на одне зображення в наборі даних припадає ~22.64 01 об’єктів у наборі даних CrowdHuman. Анотації CrowdHuman містять як видиму частину так і повну частину для об’єктів. Таким чином, під час навчання використовуємо лише видиму область для CrwodHuman. Для сегментації використовуються DeepLabv3+ та Mask RCNN, які генерують семантичні маски сегментації. Обидва алгоритми навчаються на наборі даних CrowdHuman. Моделі сегментації, навчені на наборі даних, безпосередньо використовуються для генерації масок сегментації для експериментів без будь-якого подальшого налаштування.

Для систем відслідковування наявні таки типи сценаріїв:

1. Виявлення рухомих об'єктів на зображеннях.

Очікується, що система буде навчати формувати моделі, визначати об’єкти та перевіряти продуктивність на наборі перевірки. Потім, остаточна модель використовується для генерації результатів виявлення на тестовий набір. Остаточна продуктивність автоматично оцінюється за набором об'єктивних кількісних показників.

2. Відслідковування одного об'єкта за зображеннями.

З огляду на початкові обмежувальні анотації конкретного об'єкту, це завдання вимагає оцінки місця розташування об'єкта через різні кадри. Для цього завдання візьмемо підмножину зі 100 високоякісних відео (відео з 1 по 100) із загальною кількістю 32000 кадрів. Зокрема, відео з 1 по 80 будуть використовуватися як навчальний набір і відео з 81 по 90 будуть використовуватися як набір для перевірки. Надані обмежувальні анотації конкретних об'єктів кожного кадру в тренувальному наборі та наборі перевірки. Тестовий набір складається з відео з 91 по 100, і для ініціалізації буде надано тільки анотацію першого кадру. Остаточна модель використовується для генерації результатів відслідковування на тестовому наборі.

3. Відслідковування декількох об'єктів на зображеннях.

Це завдання спрямоване на пошук декількох об'єктів, що становлять інтерес, підтримання їхньої ідентичності та визначення їхніх індивідуальних траєкторій. Для цього завдання в корпусі даних надано 100 послідовностей (відео з 1 по 100) із загальною кількістю 32000 кадрів з набору даних. Зокрема, відео з 1 по 80 будуть використовуються як навчальний набір, а відео з 81 по 90 будуть використовуються як набір для перевірки. Надані обмежувальні анотації та ідентифікатор екземпляра кожного об'єкта в кожному фреймі. Тестовий набір складається з відео від 91 до 100. Навчають моделі на тренувальному наборі та перевіряють продуктивність на наборі перевірки. Потім завершена модель використовується для генерації результатів відслідковування на тестовому наборі.

В межах проведених експериментів розглядається перший та другий тип сценаріїв.

3.1 Реалізація способу визначення об’єктів на зображеннях

Оцінюємо метод як на системі Faster RCNN, так і на системі SSD. Як методи Faster RCNN, так і методи SSD використовують ResNet як магістральну мережу. Для осіб CrowdHuman ініціалізуємо моделі за допомогою попередньо

навченої моделі ImageNet. Слід зазначити, що методи потребують 4-канального входу RGBM, який несумісний з оригінальними моделями, навченими за допомогою ImageNet. Використовуємо випадково ініціалізований фільтр для додаткового каналу.

Для швидших методів на основі RCNN навчали мережі протягом 180 тис. ітерацій на наборі даних CrowdHuman з базовою швидкістю навчання 0,2 і зменшеною в 20 разів після 20 тис. ітерацій.

Для оптимізації мережі було використано метод стохастичного градієнтного спуску. Пакет для оброблення містить по 4 зображення на графічний процесор.

Потім для набору даних CrowdHuman налаштовуємо моделі для 30 тис. ітерацій. Початкову швидкість навчання встановлено на 0,001 і зменшено після 20 тис. ітерацій

Для методів на основі SSD мережі навчано протягом 50 тис. ітерацій на наборі даних CrowdHuman з базовою швидкістю навчання 0.01 і зменшеною в 20 разів після 50 тис. ітерацій. Оскільки розмір вхідного зображення менший, ніж у швидкого RCNN, блок даних для методів SSD включає 6 зображень на один графічний процесор.

Для навчання моделі Faster RCNN використовуємо максимум 30 епох з ранньою зупинкою щоб уникнути перенавчання. Модель навчається з розміром партії 4 та швидкістю навчання 0.01. Було застосовано оптимізацію стохастичного градієнтного спуску з наступними параметрами: momentum = 0.7, weight-decay = 0.001. Крім того, використовували планувальник швидкості навчання, який зменшує швидкість навчання на 0.2 кожні 15 епізодів та на 0.1 кожні 10 епох. Функція втрат моделі Faster RCNN є комбінацією двох різних функцій втрат шляхом їх додавання. Перша функція втрат – це втрати для двох класів (є об'єкт чи ні). Друга функція – це регресійні втрати обмежувальних рамок, а саме робастна функція втрат (згладжена L_1). Ця функція приймає на вхід координати виявлених об’єктів та передбачень на вході, а потім обчислює *x*, яка є відстанню L_1 між двома цими об’єктами. Для функції втрат

використовували потрійні втрати, в якій еталонний вхід порівнюється з входом, що співпадає, та входом, що не співпадає. Відстань від опорного входу до входу, що відповідає, мінімізується, а відстань до входу, що не відповідає, максимізується.

Таблиця 3.1 – Модулі сегментації та модулі виявлення.

N	Модуль сегментації	Бінарна маска
1	DeepLabv3+	-
2	DeepLabv3+	+
3	RCNN	+
4	RCNN	-
Модуль виявлення		Магістральна мережа
SSD		ResNet
Faster RCNN		ResNet

Для оцінки результатів використовуємо такі стандартні метрики як точність (SA) та влучність (SR) .

Проведено кілька експериментів з різними налаштуваннями. Використовуємо опорну мережу ResNet для Mask RCNN для генерації масок сегментації та DeepLabv3+ для генерації бінарних масок сегментації (таблиця 3.1). Для більш якісних моделей екземплярів використовуємо ResNext як опорну мережу для семантичних моделей з оцінками зберігаємо інформацію про оцінки, тобто кожен піксель у масці вказує на ймовірність об’єкта з класу У випадку з маскою використовуємо ResNet як магістральну мережу для методів Faster RCNN та SSD.

В таблиці 3.2 та таблиці 3.3 показано продуктивність методу на наборі даних CrowdHuman з використанням системи Faster RCNN та SSD .

Для системи Faster RCNN, як показано в таблиці 3.2, порівнюємо метод з кількома детекторами. FCIS, MaskRCNN, які навчаються разом із задачами

сегментації екземплярів, інші використовують лише метод виявлення об’єктів без обмежувальних рамок.

Таблиця 3.2 – Порівняння ефективності виявлення Faster RCNN системи з базовим детектором та детекторами з використанням масок.

Номер моделі	Вхідні дані	Маска	SA	SP
1	RGB	-	0.510	0.565
2	RGB	-	0.523	0.586
3	RGB	-	0.527	0.589
4	RGB	-	0.534	0.586
5	RGB	-	0.563	0.564
6	RGB	-	0.539	0.648
7	RGB	-	0.543	0.652
8	RGBM	DeepLabv3+	0.560	0.635
9	RGBM	DeepLabv3+	0.565	0.657
10	RGBM	RCNN	0.569	0.679
11	RGBM	RCNN	0.681	0.685

Таблиця 3.3 – Порівняння ефективності виявлення системи SSD з базовим детектором та детекторами з використанням маски

№	Вхідні дані	Маска	SA	SP
1	RGB	-	0.356	0.515
2	RGBM	DeepLabv3+	0.456	0.526
3	RGBM	RCNN	0.442	0.536
4	RGB	-	0.418	0.524
5	RGBM	DeepLabv3+	0.423	0.563
6	RGBM	Маска RCNN	0.442	0.558

Як показано в таблиці 3.2, дві бінарні моделі мають приблизно 2% покращення порівняно з маскою RCNN, навченою однією задачею, і 1% покращення порівняно з маскою RCNN, навченою спільно.

Крім того, модель з використанням маски досягає значного покращення порівняно з базовим детектором, яке становить 4% та 3% відповідно. При цьому варто зазначити, що моделі, навчені на більш якісній масці, показують кращі результати, ніж моделі, навчені на менш якісній масці.

3.2 Виявлення об’єктів за допомогою глибоко вивчених семантичних масок

Також було застосовано використання масок до системи SSD. Порівнюємо лише бінарні моделі для системи SSD обчислювальні ресурси. Як показано в таблиці 3.3 метод використанням маски досягає кращої продуктивності на наборі даних.

Крім того, кожен шар згортки в SSD примусово фокусується на прогнозуванні об'єктів певного масштабу. Ознаки, що використовуються для прогнозування малих об'єктів, витягуються з поверхневих шарів, які містять лише кілька семантичних контекстів.

Тому досліджуємо, чи може додатковий семантичний контекст покращити здатність SSD виявляти малі об’єкти. У таблиці 3.4 показано середню точність малих цілей (площа наближено 32^2) на CrowdHuman. Відповідно маска сегментації дійсно може покращити ефективність виявлення SSD на малих і середніх об'єктах.

Експерименти можна розділити на дві групи: експерименти з виявлення експерименти з виявлення та експерименти з відслідковування. У першій групі експериментів, тобто в експериментах з виявлення навчаємо модель Faster RCNN двома різними методами доповнення даних. Для першого експерименту використовували те ж саме доповнення, що і в підході для відслідковування пішоходів, а саме випадковий вертикальний зсув. Назвемо цей експеримент

тестом на виявлення базової лінії. У другому експерименті застосовуємо розширену стратегію доповнення згадану вище, до набору даних. Називемо цей експеримент тестом розширеного виявлення.

У другій групі експериментів, тобто експериментах з повторної ідентифікації, застосовуємо як базовий детектор так і розширений детектор для отримання виявлень у контексті маски. Крім того, використовуємо мережу з різними моделями переідентифікації, тобто без жодної моделі переідентифікації та з моделлю повторної ідентифікації на основі ResNet, яка також використовувалася в трекерному підході для відслідковування пішоходів. Далі поєднуємо кожен детектор з кожним з моделями повторної ідентифікації в експериментах. Результати експерименту представлено на рисунку 3.1.

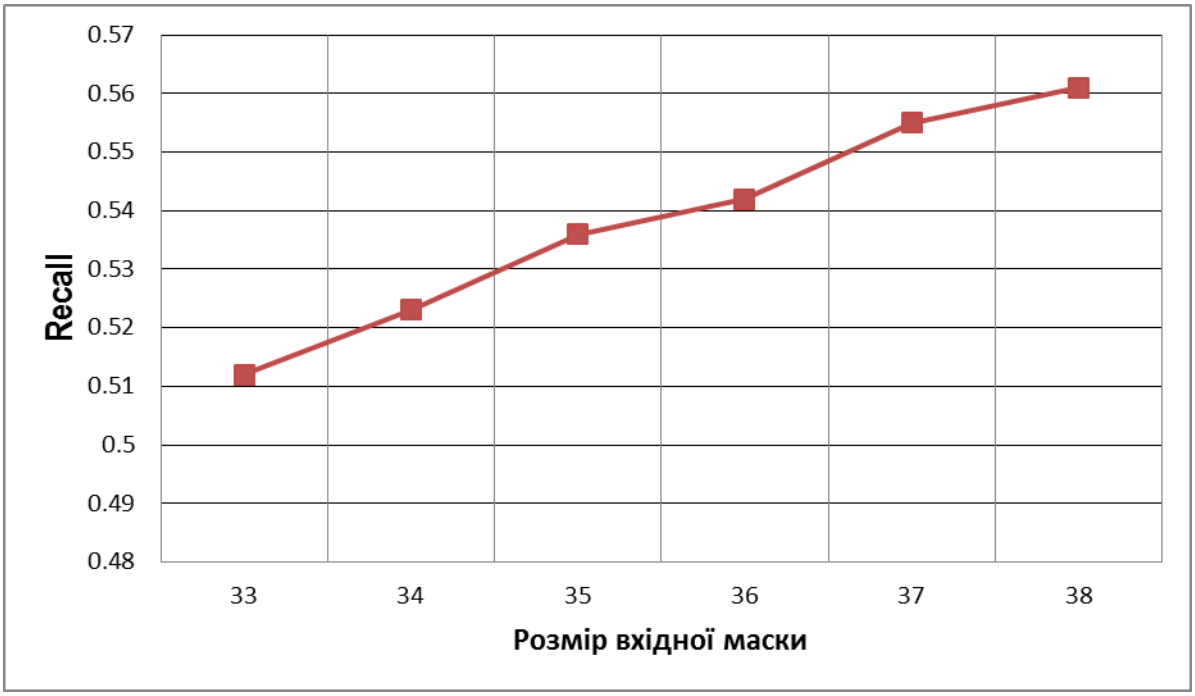


Рисунок 3.1 – Ефективність масок сегментації з різною точкою доступу. Лінія показує залежність між ефективністю детектора та якістю вхідних масок.

Для подальшого дослідження ефективності масок сегментації, оцінили моделі, навчені на масках, згенерованих різними магістральними мережами ResNet (34, 50, 101).

Продуктивність детектора значно покращилася, коли подали на нього якісніші маски сегментації. Однак все ще існує розрив між прогнозованою маскою сегментації та реальною маскою.

Крім того, варто зазначити, що поріг оцінки *Th* , який використано в рівнянні 2.4, також може впливати на ефективність виявлення. Проаналізуємо чутливість порогу *Th* (таблиця 3.4).

Таблиця 3.4 – Аналіз порогового значення *Th* чутливості

Th	SA	SP	mSA	mSP
0.0	0.546	0.364	0.618	0.723
0.1	0.554	0.368	0.625	0.726
0.2	0.552	0.375	0.635	0.721
0.3	0.563	0.395	0.645	0.734
0.4	0.572	0.410	0.643	0.736
0.5	0.571	0.415	0.648	0.742
0.6	0.558	0.402	0.632	0.746
0.7	0.553	0.393	0.628	0.741
0.8	0.550	0.362	0.626	0.735
0.9	0.532	0.358	0.623	0.732

Нижчий *Th* може призвести до високого середнього показника прогнозування, тоді як вищий *Th* може забезпечити кращу точність виявлення. Відповідно *Th* може фільтрувати шум у прогнозах і водночас зберігати багатий семантичний контекст, що може бути корисним для продуктивності виявлення.

У таблиці 3.5 та таблиці 3.6 наведено результати оцінювання на наборі даних CrowdHuman. Маски сегментації згенеровано за тією ж моделлю, що була навчена без жодних додаткових налаштувань, тому що маски сегментації генеруються.

Налаштування експерименту ідентичні до CrowdHuman, як семантична модель, так і модель екземплярів отримують покращення порівняно з базовим SSD-детектором.

Для перевірки того, що додатковий семантичний контекст може покращити роботу SSD-методів на малих і середніх об'єктах, також оцінили середню точність на малих і середніх цілях. Знову побачили, що метод справді досягає кращих результатів на цих малих і середніх об'єктах.

Таблиця 3.5 – Порівняння ефективності виявлення системи RCNN з базовим детектором та запропонованими детекторами, керованими маскою, на CrowdHuman

Модель	Вхідні дані	Маска	SA	SP
1	RGB	-	0.384	0.465
2	RGBM	DeepLabv3+	0.395	0.473
3	RGBM	RCNN	0.393	0.474
4	RGBM	RCNN	0.425	0.504

Таблиця 3.6 – Порівняння ефективності виявлення системи SSD з базовим детектором та на основі маски на CrowdHuman

Модель	Вхідні дані	Маска	AP	AR
1	RGB	-	0.285	0.365
2	RGBM	DeepLabv3+	0.310	0.389
3	RGBM	RCNN	0.323	0.392

Крім того, у таблиці 3.7 показано порівняння обчислювальних витрат на одному графічному процесорі Nvidia GTX 1070 між запропонованим методом та базовими детекторами, розмір партії було встановлено рівним 1 для Faster RCNN та SSD.

Ознаки як базового визначення, так і за допомогою масок значно перетинаються з вхідною маскою сегментації, навіть якщо базовий метод виявлення об’єктів навчався без жодних додаткових даних. Ці ознаки несуть багатий семантичний контекст і можуть бути корисними для методів розрізнення об'єктів на передньому плані та відшарування його від фону.

Крім того, завдяки використанню маски сегментації, ознаки, отримані методом на основі масок, отримують сильнішу реакцію як на фон, так і на передній план. Таким чином, зовнішня вхідна маска сегментації може допомогти детекторам дізнатися більше дискримінаційних ознак, що є ключем до покращення ефективності виявлення.

Таблиця 3.7 – Порівняння обчислювальних витрат між базовими детекторами та запропонованим методом

Ім'я	Сегментація	Виявлення
Faster RCNN	-	0.152c
Faster RCNN	0.735c	0.153c
Faster RCNN + маска	0.286c	0.154c
SSD	-	0.023c
SSD	0.689c	0.024c
SSD + маска	0.210c	0.021c

Щоб наглядно представити, як маски сегментації покращують роботу методів об’єктів, візуалізували прогнозовані обмежувальні рамки та вхідні маски сегментації на CrowdHuman порівняно з базовими детекторами.

Метод з масками показує кращі результати у випадках важкої оклюзії та менших об'єктів (рисунок 3.2). Це може бути пов'язано з тим, що маски сегментації забезпечують додатковий семантичний контекст і грають роль механізму уваги, який може допомогти детекторам зосередитися на областях, де можуть з'явитися потенційні об'єкти-кандидати на розпізнавання.

а)

б)

в)

г)

д)

е)

Рисунок 3.2 – Візуалізація робастності методу. В метод виявлення об’єктів подаються менш точні маски сегментації, що може прогнозувати задовільні результати (г, д); метод виявлення об’єктів може добре працювати на малих і важких закритих цілях (а, б); вихідні зображення для визначення об’єктів (в, д).

Водночас, метод також демонструє високу стійкість до поганих масок сегментації. Оскільки сегментація зображення є задачею зору на рівні пікселів, то маска сегментації може зазнавати перешкод, коли цілі закриті іншими об'єктами, при її формуванні.

У цьому випадку маска сегментації може бути розрізана на кілька нерегулярних частин. Однак метод також може надійно обробляти такі розділені маски сегментації.

Для подальшого дослідження ефективності масок сегментації визначимо та порівнюємо ознаки, які виділені базовим детектором.

3.3 Реалізація способу відслідковування об’єктів на зображеннях

Оцінюємо CDA-трекер у двох сценаріях. Перший CDA для типових об'єктів, в яких відмовляються від коригувальних гілок CNN. Другий, CDA для конкретних цільових категорій. Результати порівняно з деякими трекерами. CDA-трекер базується на мережі VGG-19, яка ініціалізується за допомогою класифікаційного набору даних ILSVRC, а потім навчає 2 шари адаптації доменів, які переносять глибинні ознаки з домену класифікації в домен відслідковування.

Використаємо трекери реального або напівреального часу та порівнюємо алгоритм з HCF, Struck, MD-net .

Для типових об'єктів коригувальні гілки CNN відкидаються і використовуються лише результати відслідковування KCF.

Використаємо Object Tracking Benchmark (OTB). Це корпус для візуального відслідковування, який широко використовується для оцінки продуктивності алгоритмів візуального відслідковування. Набір даних містить 100 послідовностей, кожна з яких анотована покадрово за допомогою обмежувальних рамок та 11 атрибутів .

Імена послідовностей подаються у форматі CamelCase без пробілів та підкреслень (_).

Якщо існує декілька цілей, кожна з них позначається як точка+ідентифікатор_номер (наприклад, xxxxxx.1 та xxxxxx.2).

Кожен рядок у файлах представляє обмежувальну рамку цілі в цьому кадрі – (x, y, ширина рамки, висота рамки).

У більшості послідовностей перший рядок відповідає першому кадру, а останній – останньому.

Оцінка ефективності базується на двох метриках: похибка визначення центру та коефіцієнт перекриття обмежувальної рамки (рисунок 3.3). Оцінку за один прохід використаємо для порівняння алгоритму з HCF, Struck, MD-net.

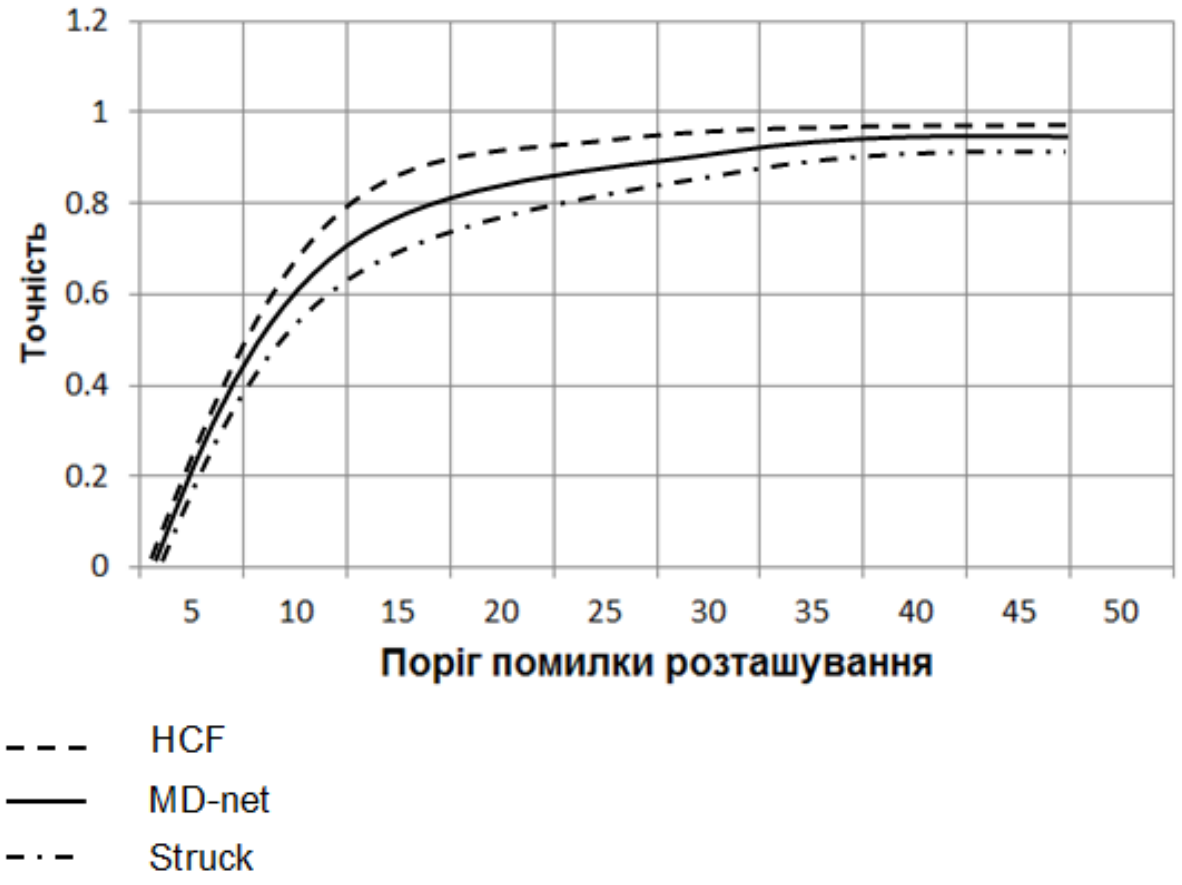


Рисунок 3.3 – Графіки похибок визначення місцезнаходження

Алгоритм CDA зберігає свою перевагу над усіма іншими трекерами реального часу і має схожу точність з HCF. MD-net мережа має значний розрив у продуктивності з усіма іншими трекерами, хоча працює зі швидкістю близько 1 кадр/с. Щоб ще більше проілюструвати порівняння між трекером CDA та

іншими трекерами реального часу, показано результати відслідковування порівняльних трекерів реального часу на деяких ключових кадрах репрезентативних відеопослідовностей OTB-100 (таблиця 3.8). Для порівняння також показано результати HCF.

Таблиця 3.8 – Точність та швидкість відслідковування порівнюваних трекерів на OTB

	HCF	MD-Net	Struck
Точність	90.4	86.6	81.3
AUC	0.657	0.631	0.586
Швидкість (к/с)	9	2	32

Аналогічний експеримент був проведений для категорії пішоходів, щоб оцінити ефективність системи для коригування із змінними об'єктами. Вибрали всі 40 відеопослідовностей пішоходів з OTB-100 в якості тестового набору і показали ефективність відслідковування трекерів, що порівнювана з отриманими раніше результатами.

Висновок до розділу 3

Представлено простий метод покращення визначення об’єктів з додатковими семантичними ознаками шляхом агрегування вихідних RGB-зображень за допомогою масок сегментації. Реалізували метод на двох популярних системах виявлення – Faster RCNN та SSD, і оцінили метод на наборі даних CrowdHuman.

Експерименти показують, що зовнішні маски сегментації можуть значно покращити ефективність виявлення об’єктів. Крім того, експериментально проаналізували ефективність масок сегментації, згенерованих різними методами, і виявили важливість додаткового семантичного контексту.

Крім того, щоб отримати уявлення про те, як маски сегментації можуть допомогти згортковій нейронній мережі отримати більше дискримінаційних ознак, візуалізували вивчені ознаки як базових методів, так і методів з використанням масок.

Слід сказати, що маски сегментації можуть значно підвищити ефективність виявлення об’єктів. Однак у цій роботі модуль сегментації та модуль розпізнавання працюють окремо один від одного, а маски сегментації генеруються заздалегідь.

Одним з можливих майбутніх напрямків може бути інтеграція цих двох процедур разом і спільне навчання декількох завдань. Крім того, протестували метод на задачах оцінки визначення об’єктів. Метод отримав невелике покращення. Таким чином, вважаємо, що додатковий семантичний контекст може також покращити виконання інших задач технічного зору.

Запропоновано ефективний алгоритм перенесення ознак з області класифікації в область візуального відслідковування. Отриманий візуальний трекер, CDA, працює в режимі реального часу і досягає точності відслідковування, порівнянної з глибинними трекерами.

Для конкретної цільової категорії CDA коригує візуальне відслідковування за результатами виявлення. Оскільки трекер і метод виявлення об’єктів визначають більшу частину глибинної мережі, не потрібно багато додаткових обчислень. Зазначене покращення означає, що відсутність цільової категорії може призвести до низької ефективності відслідковування, тоді як вирішення цієї невизначеності спосіб може дати набагато кращі глибинні трекери.

Слід зазначити, що оновлення нейронної мережі в режимі онлайн може значно підвищити точність відслідковування. Однак, існуюча схема оновлення в режимі онлайн призводить до значного зниження швидкості. Одним з можливих майбутніх напрямків може бути одночасне оновлення KCF-моделі та певної частини нейронної мережі. Таким чином, можна досягти балансу між точністю та ефективністю і отримати кращий трекер.

Іншим можливим напрямком є залучення декількох категорій об'єктів до коригувальних гілок CNN для певного типу сценарію відслідковування. Наприклад, для відслідковування дорожніх сцен можна враховувати пішоходів, автомобілі, велосипеди та мотоцикли. Це може призвести до ще більшої надійності відслідковування, ніж CDA для однієї категорії.

Висновок

У роботі представлено дослідження двох напрямків з теми комп'ютерного зору – виявлення та відслідковування об’єктів. Виявлення має за мету локалізувати набір об'єктів-кандидатів і класифікувати їх у певну категорію, тоді як відслідковування має за мету передбачити положення цілі на основі достовірної інформації з першого кадру у відеопослідовності.

У частині виявлення представили простий метод покращення методів об’єктів за допомогою додаткових семантичних ознак шляхом агрегування вихідних RGB зображень з масками сегментації. Реалізували метод на двох відомих системах виявлення – Faster RCNN та SSD. Оцінили метод на наборі даних – CrowdHuman. Водночас використано базовий метод виявлення об’єктів який використовує особливості каналу руху. Показано, що злиття додаткових ознак дозволяє досягти більш точних і надійних результатів виявлення об'єктів.

Для подальшої роботи одним з можливих напрямків є використання технології пошуку архітектури нейронної мережі для автоматичного пошуку базової мережі для задачі виявлення об'єктів.

У частині відслідковування пропонується простий, але ефективний алгоритм для перенесення ознак з області класифікації в область візуального відслідковування. Крім того, вводимо об'єктність у візуальне відслідковування об'єктів. Для конкретної цільової категорії візуальний трекер орієнтується на результати виявлення. Можливий майбутній напрямок полягає у залученні декількох категорій об'єктів до коригувальних гілок для певного типу сценарію відслідковування. Наприклад, для відслідковування дорожньої обстановки можна враховувати пішоходів, автомобілі, велосипеди та мотоцикли. Це може призвести до збільшення надійності відслідковування, ніж метод використання методу для однієї категорії

Перелік посилань

1. Badue C., Guidolini R., Carneiro R. V., Azevedo P., Cardoso V. B., Forechi A., Jesus L., Berriel R., Paixao T. M., Mutz F. Self-driving cars: A survey. *Expert Systems with Applications*. 2021. Vol. 165. Pp. 113816.

2. Benjdira B., Khursheed T., Koubaa A., Ammar A., Ouni K. Car detection using unmanned aerial vehicles: Comparison between faster r-cnn and yolov3: *2019 1st International Conference on Unmanned Vehicle Systems-Oman (UVS)*, IEEE, 2019. Pp.1–6.

3. Bharati P., Pramanik A. Deep learning techniques—R-CNN to mask R-CNN: a survey. *Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2019*. 2020. Pp. 657–668.

4. Biswas D., Su H., Wang C., Stevanovic A., Wang W. An automatic traffic density estimation using Single Shot Detection (SSD) and MobileNet-SSD. *Physics and Chemistry of the Earth, Parts A/B/C*. 2019. Vol. 110. Pp. 176–184.

5. Bizjak M., Peer P., Emeršič Ž. Mask R-CNN for ear detection: *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, IEEE, 2019. Pp.1624–1628.

6. Chan S., Tao J., Zhou X., Bai C., Zhang X. Siamese implicit region proposal network with compound attention for visual tracking. *IEEE Transactions on Image Processing*. 2022. Vol. 31. Pp. 1882–1894.

7. Chang A., Dai A., Funkhouser T., Halber M., Niessner M., Savva M., Song S., Zeng A., Zhang Y. Matterport3d: Learning from rgb-d data in indoor environments. *arXiv preprint arXiv:1709.06158*. 2017.

8. Chen L.-C., Papandreou G., Kokkinos I., Murphy K., Yuille A. L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*. 2017. Vol. 40, No. 4. Pp. 834–848.

9. Cheng L., Li J., Duan P., Wang M. A small attentional YOLO model for landslide detection from satellite remote sensing images. *Landslides*. 2021. Vol. 18, No. 8. Pp. 2751–2765.

10. Cheng T., Wang X., Huang L., Liu W. Boundary-preserving mask r-cnn: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV 16*, Springer, 2020. Pp.660–676.

11. El-Rewaidy H., Fahmy A. S., Pashakhanloo F., Cai X., Kucukseymen S., Csecs I., Neisius U., Haji-Valizadeh H., Menze B., Nezafat R. Multi-domain convolutional neural network (MD-CNN) for radial reconstruction of dynamic cardiac MRI. *Magnetic Resonance in Medicine*. 2021. Vol. 85, No. 3. Pp. 1195–1208.

12. Fan H., Ling H. Siamese cascaded region proposal networks for real-time visual tracking: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019. Pp.7952–7961.

13. Ge Z., Liu S., Wang F., Li Z., Sun J. Yolox: Exceeding yolo series in 2021. *arXiv preprint arXiv:2107.08430*. 2021.

14. Gowayyed M. A., Torki M., Hussein M. E., El-Saban M. Histogram of oriented displacements (HOD): Describing trajectories of human joints for action recognition: *Twenty-third international joint conference on artificial intelligence*, 2013.

15. Guan D., Cao Y., Yang J., Cao Y., Yang M. Y. Fusion of multispectral data through illumination-aware deep neural networks for pedestrian detection. *Information Fusion*. 2019. Vol. 50. Pp. 148–157.

16. Gupta S., Girshick R., Arbeláez P., Malik J. Learning rich features from RGB-D images for object detection and segmentation: *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VII 13*, Springer, 2014. Pp.345–360.

17. Jaeger P. F., Kohl S. A., Bickelhaupt S., Isensee F., Kuder T. A., Schlemmer H.-P., Maier-Hein K. H. Retina U-Net: Embarrassingly simple exploitation of segmentation supervision for medical object detection: *Machine Learning for Health Workshop*, PMLR, 2020. Pp.171–183.

18. Jiang P., Ergu D., Liu F., Cai Y., Ma B. A Review of Yolo algorithm developments. *Procedia Computer Science*. 2022. Vol. 199. Pp. 1066–1073.

19. Joze H. R. V., Shaban A., Iuzzolino M. L., Koishida K. MMTM: Multimodal transfer module for CNN fusion: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. Pp.13289–13299.

20. Kanazawa A., Sharma A., Jacobs D. Locally scale-invariant convolutional neural networks. *arXiv preprint arXiv:1412.5104*. 2014.

21. Kattenborn T., Leitloff J., Schiefer F., Hinz S. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS journal of photogrammetry and remote sensing*. 2021. Vol. 173. Pp. 24–49.

22. Kumar A., Zhang Z. J., Lyu H. Object detection in real time based on improved single shot multi-box detector algorithm. *EURASIP Journal on Wireless Communications and Networking*. 2020. Vol. 2020, No. 1. Pp. 1–18.

23. Lan X., Zhang S., Yuen P. C. Robust Joint Discriminative Feature Learning for Visual Tracking.: *IJCAI*, 2016. Pp.3403–3410.

24. Law H., Teng Y., Russakovsky O., Deng J. Cornernet-lite: Efficient keypoint based object detection. *arXiv preprint arXiv:1904.08900*. 2019.

25. Law S., Seresinhe C. I., Shen Y., Gutierrez-Roig M. Street-Frontage-Net: urban image classification using deep convolutional neural networks. *International Journal of Geographical Information Science*. 2020. Vol. 34, No. 4. Pp. 681–707.

26. Li H., Li Y., Porikli F. Deeptrack: Learning discriminative feature representations online for robust visual tracking. *IEEE Transactions on Image Processing*. 2015. Vol. 25, No. 4. Pp. 1834–1848.

27. Li H., Zhang J., Li Z., Liu J., Wang Y. Improvement of Min-entropy evaluation based on pruning and quantized deep neural network. *IEEE Transactions on Information Forensics and Security*. 2023. Vol. 18. Pp. 1410–1420.

28. Li J., Liang X., Shen S., Xu T., Feng J., Yan S. Scale-aware fast R-CNN for pedestrian detection. *IEEE transactions on Multimedia*. 2017. Vol. 20, No. 4. Pp. 985–996.

29. Li K., Wan G., Cheng G., Meng L., Han J. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS journal of photogrammetry and remote sensing*. 2020. Vol. 159. Pp. 296–307.

30. Li S., Liu C. H., Lin Q., Wen Q., Su L., Huang G., Ding Z. Deep residual correction network for partial domain adaptation. *IEEE transactions on pattern analysis and machine intelligence*. 2020. Vol. 43, No. 7. Pp. 2329–2344.

31. Li X., Ma C., Wu B., He Z., Yang M.-H. Target-aware deep tracking: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019. Pp.1369–1378.

32. Liu R., Ao B., Wen Q., Wu X., Yin J., Li K. Combining ExtremeNet with Shape Constraints and Re-Discrimination to Detect Cells from CD56 Images: *2022 26th International Conference on Pattern Recognition (ICPR)*, IEEE, 2022. Pp.4587–4593.

33. Long J., Shelhamer E., Darrell T. Fully convolutional networks for semantic segmentation: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015. Pp.3431–3440.

34. Magalhães S. A., Castro L., Moreira G., Dos Santos F. N., Cunha M., Dias J., Moreira A. P. Evaluating the single-shot multibox detector and YOLO deep learning models for the detection of tomatoes in a greenhouse. *Sensors*. 2021. Vol. 21, No. 10. Pp. 3569.

35. Maity M., Banerjee S., Chaudhuri S. S. Faster r-cnn and yolo based vehicle detection: A survey: *2021 5th international conference on computing methodologies and communication (ICCMC)*, IEEE, 2021. Pp.1442–1447.

36. Pang J., Chen K., Shi J., Feng H., Ouyang W., Lin D. Libra r-cnn: Towards balanced learning for object detection: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019. Pp.821–830.

37. Pang Y., Xie J., Khan M. H., Anwer R. M., Khan F. S., Shao L. Mask-guided attention network for occluded pedestrian detection: *Proceedings of the IEEE/CVF international conference on computer vision*, 2019. Pp.4967–4975.

38. Shao J., Du B., Wu C., Zhang L. Can we track targets from space? A hybrid kernel correlation filter tracker for satellite video. *IEEE Transactions on Geoscience and Remote Sensing*. 2019. Vol. 57, No. 11. Pp. 8719–8731.

39. Su Y., Li D., Chen X. Lung nodule detection based on faster R-CNN framework. *Computer Methods and Programs in Biomedicine*. 2021. Vol. 200. Pp. 105866.

40. Tesema F. B., Wu H., Chen M., Lin J., Zhu W., Huang K. Hybrid channel based pedestrian detection. *Neurocomputing*. 2020. Vol. 389. Pp. 1–8. URL: https://doi.org/10.1016/j.neucom.2019.12.110.

41. Tian Z., Shen C., Chen H., He T. Fcos: A simple and strong anchor-free object detector. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2020. Vol. 44, No. 4. Pp. 1922–1933.

42. Wang C.-Y., Liao H.-Y. M., Wu Y.-H., Chen P.-Y., Hsieh J.-W., Yeh I.-H. CSPNet: A new backbone that can enhance learning capability of CNN: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020. Pp.390–391.

43. Wang S.-Y., Wang O., Zhang R., Owens A., Efros A. A. CNN-generated images are surprisingly easy to spot... for now: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020. Pp.8695–8704.

44. Wu M., Yue H., Wang J., Huang Y., Liu M., Jiang Y., Ke C., Zeng C. Object detection based on RGC mask R-CNN. *IET Image Processing*. 2020. Vol. 14, No. 8. Pp. 1502–1508.

45. Wu X., Sahoo D., Hoi S. C. Recent advances in deep learning for object detection. *Neurocomputing*. 2020. Vol. 396. Pp. 39–64.

46. Xie X., Cheng G., Wang J., Yao X., Han J. Oriented R-CNN for object detection: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021. Pp.3520–3529.

47. Xu B., Wang W., Falzon G., Kwan P., Guo L., Chen G., Tait A., Schneider D. Automated cattle counting using Mask R-CNN in quadcopter vision system. *Computers and Electronics in Agriculture*. 2020. Vol. 171. Pp. 105300.

48. Yang R., Zhang F., Xia J., Wu C. Landslide extraction using Mask R-CNN with background-enhancement method. *Remote Sensing*. 2022. Vol. 14, No. 9. Pp. 2206.

49. Younis A., Shixin L., Jn S., Hai Z. Real-time object detection using pre-trained deep learning models MobileNet-SSD: *Proceedings of 2020 the 6th international conference on computing and data engineering*, 2020. Pp.44–48.

50. Yuan D., Zhang X., Liu J., Li D. A multiple feature fused model for visual object tracking via correlation filters. *Multimedia Tools and Applications*. 2019. Vol. 78. Pp. 27271–27290.

51. Zhang J., Sun J., Wang J., Yue X.-G. Visual object tracking based on residual network and cascaded correlation filters. *Journal of ambient intelligence and humanized computing*. 2021. Vol. 12. Pp. 8427–8440.

52. Zhang Y., Gao J., Zhou H. Breeds classification with deep convolutional neural network: *Proceedings of the 2020 12th International Conference on Machine Learning and Computing*, 2020. Pp.145–151.

53. Zhang Z., Peng H. Deeper and wider siamese networks for real-time visual tracking: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019. Pp.4591–4600.

54. Zheng D., Xiao J., Huang K., Zhao Y. Segmentation mask guided end-to-end person search. *Signal Processing: Image Communication*. 2020. Vol. 86. Pp. 115876.

55. Zhu H., Xue M., Wang Y., Yuan G., Li X. Fast visual tracking with siamese oriented region proposal network. *IEEE Signal Processing Letters*. 2022. Vol. 29. Pp. 1437–1441.

56. Zlocha M., Dou Q., Glocker B. Improving RetinaNet for CT lesion detection with dense masks from weak RECIST labels: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22*, Springer, 2019. Pp.402–410.

ДОДАТКИ

Додаток А

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

ЗАХОПЛЕННЯ ТА ТРАСУВАННЯ ОБ'ЄКТІВ НА ЗОБРАЖЕННЯХ В СИСТЕМАХ НАВЕДЕННЯ ЦІЛІ ЗА ДОПОМОГОЮ НЕЙРОННИХ МЕРЕЖ

Виконав:
студент групи КН-19-1
Горохольський Станіслав В'ячеславович
Керівник:
д.т.н., доцент кафедри КН
Манзюк Е.А.

Актуальність

Виявлення та відслідковування об'єктів є важливими задачами комп'ютерного зору, які стали ключовими завданнями для багатьох практичних застосувань, таких як відеоспостереження, інтелектуальні транспортні системи, охоронні системи. Завдяки технологіям глибокого навчання, таким як згорткові нейронні мережі, сучасні системи виявлення та відслідковування об'єктів досягають значно кращої точності у практичних застосуваннях.

Сучасна система виявлення об'єктів в основному складається з двох етапів: локалізація набору об'єктів-кандидатів і класифікація цих об'єктів за певною категорією.

Візуальне відслідковування має за мету слідувати за переміщенням конкретного об'єкта, позначеного на першому кадрі відеопослідовності. До появи глибокого навчання, традиційні алгоритми відслідковування приділяли найбільшу увагу розробці надійної моделі з точки зору стратегії оновлення моделі, ансамблевого постпроцесора, моделі спостереження та інші. Деякі з них досягли значних успіхів як у точності, так і у швидкості.

Мета і задачі роботи

Метою кваліфікаційної роботи бакалавра є розробка способу захоплення та трасування об'єктів на зображеннях в системах наведення цілі за допомогою нейронних мереж:

Для досягнення поставленої мети визначені наступні задачі дослідження:

- визначити послідовність застосування способу захоплення об'єктів на зображеннях в системах наведення цілі за допомогою нейронних мереж;
- визначити послідовність застосування способу трасування об'єктів на зображеннях в системах наведення цілі;
- реалізувати програмну систему захоплення та трасування об'єктів на зображеннях в системах наведення цілі за допомогою нейронних мереж;
- провести експериментальне тестування розробленого способу.

Розробити програмну реалізацію методу автоматизованого підбору відповідей на запитання за семантичною подібністю, провести її тестування.

В роботі розроблена система виявлення та відслідковування об'єктів на основі глибокого навчання.

Для покращення виявлення об'єктів використано семантичну контекстну інформація для виявлення об'єктів, та семантичні ознаки, які отримані із застосуванням семантичних масок сегментації. Ці маски сегментації діють як механізм отримання областей і дозволяють детекторам зосереджуватися на тих ділянках зображення, де найімовірніше з'являться потенційні кандидати в об'єкти.

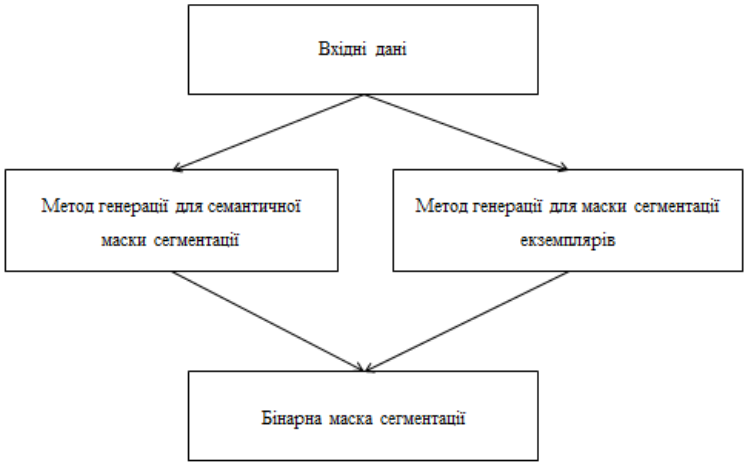
Аналіз відомих підходів

На основі проведеного аналізу було визначено, що результати сегментації зображень можуть бути корисними з точки зору продуктивності глибоких згорткових нейронних мереж.

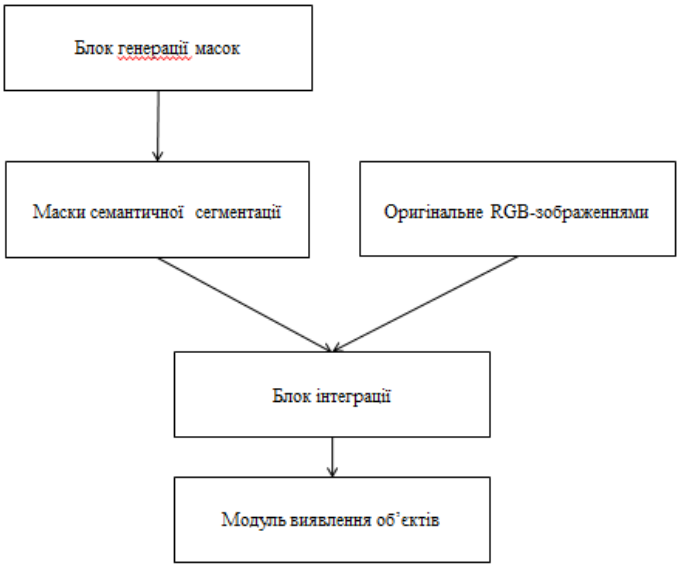
Відомі методи мають ряд обмежень:
1. Обмеженість даних: Для навчання декількох задач, наприклад, виявлення обмежувальних рамок та сегментації екземплярів, використаємо процедура спільного навчання. Однак для більшості наборів даних виявлення об'єктів попіксельні анотації недоступні. Таким чином, така процедура навчання навряд чи може бути адаптована до тих наборів даних, які не мають міток пікселів.
2. Стійкість: Для видалення глибоких ознак декілька задач використовують одну і ту ж саму магістральну мережу. Ця стратегія значно підвищує ефективність мережі. Однак ознаки, отримані різними задачами, можуть не вплинути на продуктивність інших задач.

Модуль сегментації

Для модуля сегментації використано два готові методи генерації масок сегментації: DeepLabv3+ для семантичної маски сегментації та Mask RCNN для маски сегментації екземплярів. Потім і семантична маска сегментації, і маска сегментації екземплярів, згенеровані модулем сегментації, переносяться у бінарну маску сегментації



Структурна схема блоку виявлення об'єктів



Щоб дослідити ефективність вхідних масок сегментації, використаємо декілька налаштувань для DeepLabv3+ та Mask RCNN, щоб згенерувати маски сегментації різної якості.
Для DeepLabv3+ розробимо два типи масок сегментації:
– бінарна маска семантичної сегментації;
– оціночна маска семантичної сегментації з балами.

Модуль сегментації

Бінарна семантична маска сегментації визначається як:

$$Mask_{binary} = f\left(\frac{e^{X_i}}{\sum_{j=1}^{T} e^{X_i}}\right), \quad i=1,\ldots,T$$

Оціночна маска семантичної сегментації визначається

$$Mask_{score} = f\left(\frac{e^{X_i}}{\sum_{j=1}^{T} e^{X_i}}\right), \quad j=1,\ldots,T$$

де X – матриця згенерована моделлю DeepLabv3+;

$Mask$ – елемент в матриці сегментації;

T – кількість елементів матриці сегментації.

Виявлення об'єктів

Під час навчання замінюємо 3-канальні RGB зображення на 4-канальні RGBM зображення. Далі локалізуємо ціль і мітку класу.

Ці дві задачі навчаються спільно, таким чином визначається функція втрат Faster RCNN

$$Loss = \frac{1}{N_{bin}} L_{bin} + \lambda \frac{1}{N_{reg}} L_{local}$$

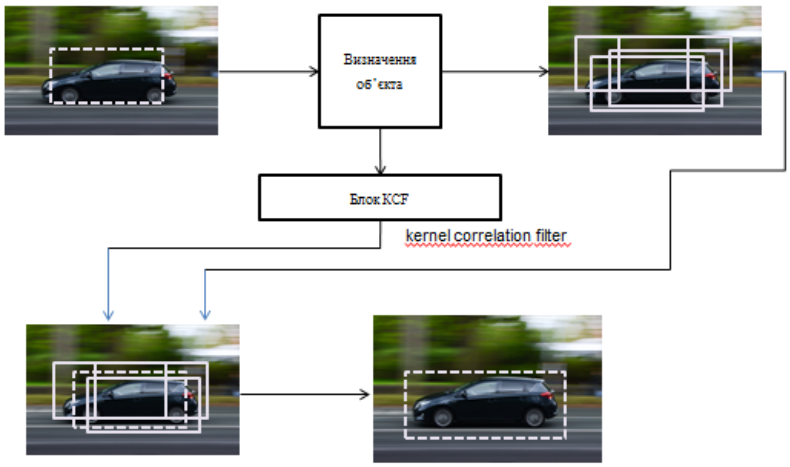
де L_{bin} та L_{scr} – втрати для бінарної класифікації та згладжені втрати L_1 для регресії з обмеженою зоною;

N_{bin} та N_{reg} – параметри нормалізації, які визначаються розміром зони обробки та кількістю пропозицій відповідно;

λ – коефіцієнт, що врівноважує ці втрати.

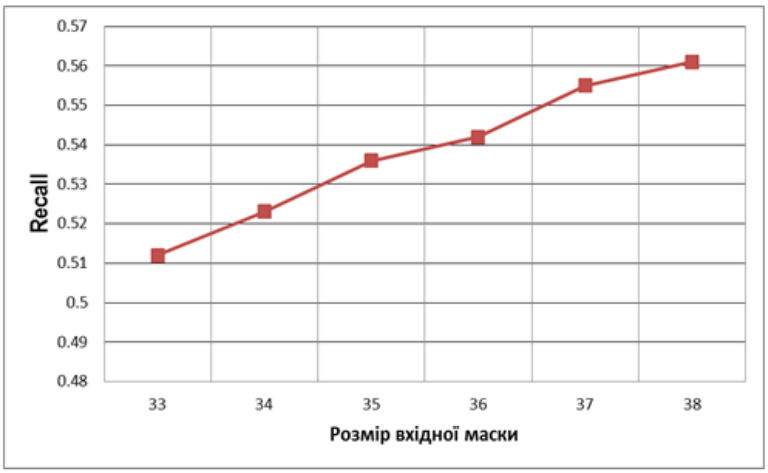
Переваги алгоритму доменної адаптації

Схема відслідковування – виявлення – об'єднання



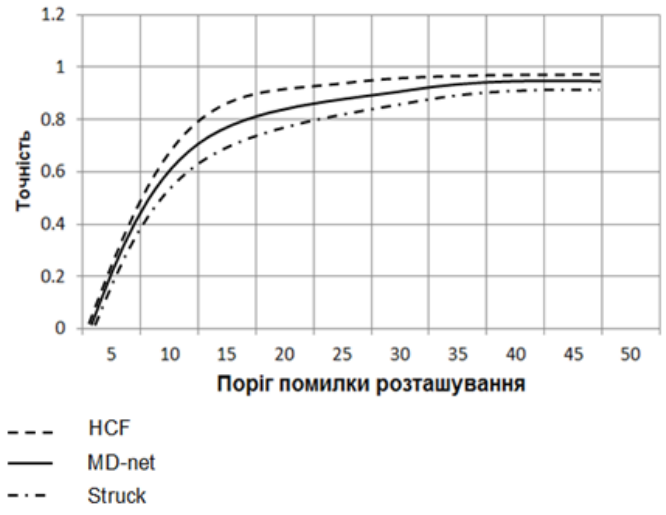
Вихідний блок відслідковування коригується блоками виявлення. Корекція зазвичай є корисною завдяки більш чіткому визначенню цільової категорії та добре навченому детектору

Ефективність масок сегментації з різною точкою доступу.



Лінія показує залежність між ефективністю детектора та якістю вхідних масок

Графіки похибок визначення місцезнаходження



Оцінка ефективності базується на двох метриках: похибка визначення центру та коефіцієнт перекриття обмежувальної рамки

Точність та швидкість відслідковування порівнюваних трекерів

	HCF	MD-Net	Struck
Точність	90.4	86.6	81.3
AUC	0.657	0.631	0.586
Швидкість (кадр/с)	9	2	32

Висновки

У роботі представлено дослідження двох напрямків з теми комп'ютерного зору – виявлення та відслідковування об’єктів. Виявлення має за мету локалізувати набір об'єктів-кандидатів і класифікувати їх у певну категорію, тоді як відслідковування має за мету передбачити положення цілі на основі достовірної інформації з першого кадру у відеопослідовності.

У частині виявлення представлено метод покращення методів об’єктів за допомогою додаткових семантичних ознак шляхом агрегування вихідних RGB зображень з масками сегментації.

Використано базовий метод виявлення об’єктів, який використовує особливості каналу руху. Показано, що злиття додаткових ознак дозволяє досягти більш точних і надійних результатів виявлення об'єктів

Mon May 29 09:38:28 EEST 2023, Петровський Сергій Степанович, Хмельницький національний університет, ХНУ

Anti-Plagiarism v-15.257

Максимальне співпадіння з одним документом 1.0%

Словники перевірки: en_US, ru_RU, ua_UA. **Помилок в документах: 10%**

	Документ		Сумарний збіг по Базі Даних	
ID: 114212 Назва: КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА Додано в БД: 2023-05-29 Автора: С.В. Горохольський Керівники: Е.А. Манзюк Консультанти: Опоненти:	Символи	Лексеми	Символи	Лексеми
	63829	999	891 (1%)	12 (1%)

Джерело плагіату

ID	Опис	Наявність плагіату в документі	
		Символи	Лексеми

UNICHECK

Ім'я користувача:
Кафедра КН

ID перевірки:
1015296385

Дата перевірки:
29.05.2023 10:39:24 EEST

Тип перевірки:
Doc vs Internet + Library

Дата звіту:
29.05.2023 10:45:07 EEST

ID користувача:
100005671

Назва документа: **КН-19-1 Горохольський**

Кількість сторінок: **59** Кількість слів: **10524** Кількість символів: **79060** Розмір файлу: **1.73 MB** ID файлу: **1014968338**

1.55%
Схожість

Найбільша схожість: **0.96%** з джерелом з Бібліотеки (ID файлу: **1011266678**)

1.33% Джерела з Інтернету	116	Сторінка 61
1.46% Джерела з Бібліотеки	77	Сторінка 62

0% Цитат

Не знайдено жодних цитат

Посилання	1	Сторінка 62

0%
Вилучень

Немає вилучених джерел

Модифікації

Виявлено модифікації тексту. Детальна інформація доступна в онлайн-звіті.

Замінені символи	3

РІШЕННЯ ЕКСПЕРНОЇ КОМІСІЇ КАФЕДРИ КОМП'ЮТЕРНИХ НАУК
ПРО ДОПУСК КВАЛІФІКАЦІЙНОЇ РОБОТИ ДО ЗАХИСТУ

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

Назва: Захоплення та трасування об'єктів на зображеннях в системах наведення цілі за допомогою нейронних мереж

Автор: *студент групи КН-19-1 Горохольський Станіслав В'ячеславович*

Спеціальність: *122 – Комп'ютерні науки*

Освітня програма: *освітньо-професійна*

Науковий керівник: *д.т.н., доцент Манзюк Е.А.*

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 Горохольського С.В., не є плагіатом, оскільки: запозичення розміщені в розділі огляду існуючих підходів, не описують безпосередньо авторську роботу і не стосуються її результатів; усі запозичення фрагментарні; до запозичень входять фрагменти програмного коду, що не мають авторства і містять поширені конструкції; серед запозичень знаходяться загальновідомі терміни, скорочення.

Обсяг запозичень, визначений системами виявлення збігів/ ідентичності/схожості, складає:

- за системою Anti-Plagiarism: 1%;

- за системою Unicheck: 1.55%.

Сумарний обсяг всіх запозичень, визначений системою виявлення збігів/ ідентичності/схожості є допустимим.

Керівник роботи _____ *Едуард МАНЗЮК*

Гарант ОП _____ *Олександр МАЗУРЕЦЬ*

Завідувач кафедри КН _____ *Олександр БАРМАК*

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МОН УКРАЇНИ

Кафедра комп'ютерних наук

ВІДГУК НАУКОВОГО КЕРІВНИКА

на кваліфікаційну роботу бакалавра

студента *гр. КН-19-1 Горохольського Станіслава В'ячеславовича*

за темою *Захоплення та трасування об'єктів на зображеннях в системах наведення цілі за допомогою нейронних мереж*

1. Актуальність теми

Виявлення та відслідковування об'єктів є важливими задачами комп'ютерного зору, які стали ключовими завданнями для багатьох практичних застосувань, таких як відеоспостереження, інтелектуальні транспортні системи, охоронні системи. Завдяки технологіям глибокого навчання, таким як згорткові нейронні мережі, сучасні системи виявлення та відслідковування об'єктів досягають значно кращої точності у практичних застосуваннях. В роботі розроблена система виявлення та відслідковування об'єктів на основі глибокого навчання, яка є актуальною задачею комп'ютерних наук.

2. Відповідність роботи предметній області Стандарту спеціальності 122 Комп'ютерні науки

За стандартом, а саме описом предметної області, об'єктами вивчення та діяльності є математичні, інформаційні, імітаційні моделі реальних явищ, об'єктів, систем і процесів та методи і технології отримання, зберігання, обробки, передачі та використання інформації. Метою роботи є розробці способу захоплення та трасування об'єктів на зображеннях в системах наведення цілі за допомогою нейронних мереж. При вирішенні поставленої задачі використано математичні моделі, методи та алгоритми розв'язання теоретичних і прикладних задач, що виникають при розробці методів машинного навчання. Результати виконання кваліфікаційної роботи бакалавра відповідають стандарту бакалавра спеціальності 122 – Комп'ютерні науки.

3. Професійні та особистісні якості бакалавра

При роботі над кваліфікаційною роботою бакалавра Горохольський Станіслав В'ячеславович продемонстрував належні знання та вміння, своєчасно реалізовував поставлені завдання. Під час написання пояснювальної записки, розробки прикладного програмного забезпечення продемонстрував наявні компетентності та результати навчання. Опанував професійні вміння за напрямком «Комп'ютерні науки».

4. Ступінь самостійності під час виконання кваліфікаційної роботи

Одержані в роботі результати є наслідком особистої діяльності студента, який самостійно виконував усі поставлені задачі.

5. Ступінь оволодіння методами дослідження

При реалізації кваліфікаційної роботи показав належний рівень компетентностей та володіння необхідними методами, методиками та технологіями предметної області комп'ютерних наук.

6. Повнота та якість розкриття теми роботи

Тема роботи повною мірою обґрунтована та розкрита належним чином. Проведено аналіз відомих досліджень відповідно до обраної теми. Поставлені завдання, реалізовані та розроблено програмне забезпечення для реалізації запропонованого метода.

7. Логічність, послідовність, аргументованість, літературна грамотність викладення матеріалу

Структура роботи та послідовність викладення логічні та відповідають поставленій меті. Викладення матеріалу послідовне, аргументоване, літературно грамотне.

8. Можливість практичного застосування кваліфікаційної роботи бакалавра, окремих її частин

Розроблений у роботі метод може бути використаний в системах захоплення цілі та слідування за її переміщенням..

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

Враховуючи належний рівень виконання та забезпечення усіх необхідних вимог, робота може бути допущена до захисту. Рекомендована оцінка «добре».

Керівник _____ *д.т.н., доцент каф. КН Едуард МАНЗЮК*

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МОН УКРАЇНИ

Кафедра комп'ютерних наук

РЕЦЕНЗІЯ

на кваліфікаційну роботу бакалавра

студента *гр. КН-19-1 Горохольського Станіслава В'ячеславовича*
за темою: *Захоплення та трасування об'єктів на зображеннях в системах наведення цілі за допомогою нейронних мереж*

1. Актуальність обраної теми

Виявлення та відслідковування об'єктів є важливими задачами комп'ютерного зору, які стали ключовими завданнями для багатьох практичних застосувань, таких як відеоспостереження, інтелектуальні транспортні системи тощо. Тому розробка систем штучного інтелекту для систем відслідковування є актуальним завданням.

2. Повнота розкриття мети та завдань роботи

Мета та завдання, які сформульовані в кваліфікаційній роботі розкрито належним чином. Проведено ґрунтовний аналіз предметної області, визначено актуальність проведення розробки, проведено дослідження відомих методів. Здійснено опис кроків розробленого методу та реалізовано його в програмній системі. Отримані експериментальні підтвердження ефективності практичного застосування розробленого методу.

3. Зміст кожного розділу роботи

Записка кваліфікаційної роботи бакалавра містить три розділи. У першому розділі проведено аналіз предметної області, досліджено відомі роботи та визначено актуальність теми. У другому розділі представлено метод захоплення та трасування об'єктів на зображеннях. Третій розділ присвячено практичній реалізації розробленого метода та експериментальній перевірці його ефективності.

4. Оцінка розробленої інформаційної системи, її практична цінність

Розроблений метод захоплення та трасування об'єктів на зображеннях дозволяє визначати об'єкти та здійснювати слідування за їхнім переміщенням, що може бути застосовано в системах спостереження, охорони тощо.

5. Якість оформлення кваліфікаційної роботи бакалавра

Записка оформлена відповідно до вимог та правил. Викладення матеріалу логічне та послідовне.

6. Недоліки кваліфікаційної роботи бакалавра

Рекомендовано вдосконалити систему додавши можливість здійснювати відслідковування за декількома об'єктами на зображені одночасно.

7. Загальний висновок (допускається чи не допускається до захисту), та оцінка на яку заслуговує кваліфікаційна робота.

Враховуючи рівень виконання та забезпечення усіх необхідних вимог, робота може бути допущена до захисту. Рекомендована оцінка "добре".

Рецензент *к.ф.-м.н., доц. каф. ВМКЗ* Романюк А.О.