ЗМІСТ

ВСТУП

У ряді предметних областей, що активно розвиваються, забезпечення безпеки, навігації роботів, автономного керування транспортними засобами та інших затребуваним є вирішення завдань аналізу статичних і динамічних зорових сцен. Дані завдання характеризуються такими особливостями:

- необхідність виділення, розпізнавання та супроводження об'єктів як у статиці, так і в динаміці;
- обробка послідовності кадрів із досить високою точністю в режимі реального часу;
- залежність точності супроводу об'єктів у динаміці від точності виділення та розпізнавання об'єктів у статиці;
- наявність як статичних, так і рухомих камер із різною швидкістю руху;
- наявність сцен з різним освітленням та розміром об'єктів;
- виникнення перекриттів одних об'єктів іншими під час супроводу цих об'єктів у динаміці.

Для вирішення задач виділення об'єктів на статичних зорових сценах може бути застосований підхід на основі «ковзного вікна» спільно з шаблонами, дескрипторами локальних особливостей, такими як HOG, LBP, SIFT, SURF, колірними ознаками, методами кон- турного аналізу. Основними недоліками таких методів, є необходимость перебору досить великої кількості областей, необхідних виділення об'єктів, і навіть специфічність ознак для «ковзного вікна».

Інша група методів ґрунтується на сегментуванні зображень. Такі методи є евристично ненавчаними методами і не залежать від специфіки розв'язуваного завдання, крім того, вони вимагають значних обчислень і навіть за допомогою сучасних обчислювальних засобів не дозволяють реалізувати режим реального часу.

Методи генерації гіпотез про розташування об'єктів на зображенні, засновані на нейромережевому підході, дозволяють усунути зазначені недоліки.

Для того, щоб вирішити проблеми , з об'єктами розпізнавання на статичних візуальних сценах, методи , засновані на використанні математичної статистики і машинного навчання є найбільш часто використовуються . Серед методів машинного навчання досить добре себе зарекомендували штучні нейронні мережі, у тому числі згорткові нейронні мережі. Такі нейронні мережі характеризуються значним збільшенням точності розпізнавання, порівняно з класичними методами. Так, наприклад, вперше застосування «глибоких» згорткових нейронних мереж (CHC) дозволило зменшити середню помилку розпізнавання приблизно в півтора рази в порівнянні з одним з кращих методів, що вирішує задачу класифікації зображень з використанням векторів Фішера та SIFT, запропонованої в.

Для вирішення завдань детектування об'єктів на статичних зорових сценах доцільним є спільне використання методів виділення та розпізнавання об'єктів. Однак для зменшення обчислювальної складності ефективніше використовувати методи, що вирішують ці завдання одночасно. Такі методи досить точні, але, як правило, не дозволяють детектувати об'єкти на зорових сценах в режимі реального часу. Інші методи працюють у режимі реального часу, але не забезпечують необхідну точність. Ще одна група методів забезпечує компроміс між точністю та швидкістю виділення та розпізнавання об'єктів.

Для вирішення завдань детектування об'єктів та об'єднання їх у треки на динамічних зорових сценах потрібно спочатку детектувати об'єкти, потім виділити ознаки окремо для кожного об'єкта, потім на основі отриманих ознак призначити знайдені детекції об'єктів на треки, з використанням методів супроводу об'єктів. Під детекцією об'єкта на зображенні розуміється область зображення, на якій об'єкт виділений за допомогою прямокутника, що обрамляє.

Ознаки об'єктів можуть бути виділені з використанням окремої CHP, однак такий підхід вимагає значних обчислювальних ресурсів.

Аналіз методів супроводу об'єктів показав, що існують досить точні методи супроводу об'єктів, наприклад, але вони не дозволяють виконувати обробку в режимі реального часу, оскільки вирішують завдання глобальної оптимізації і

вимагають наявності одразу всієї послідовності. кадрів. Інші методи виконують обробку в режимі реального часу, але не є достатньо точними, наприклад.

Для дослідження методів аналізу динамічних зорових сцен із застосуванням існуючих баз для тестування пропонується працювати з великою кількістю помилкових детекцій об'єктів та розробляти алгоритми, що дозволяють виділяти дані з шуму. Однак при використанні СНР часто доводиться мати справу не з великою кількістю помилкових детекцій об'єктів, а з їх пропусками.

Дослідження існуючих методів виділення, розпізнавання та супроводу об'єктів на статичних та динамічних зорових сценах виявили такі основні обмеження:

- відсутні методи супроводу та детектування об'єктів, що дозволяють виконати детектування та виділення «глибоких» ознак об'єктів за один прохід СНР;
- відсутні алгоритми, що дозволяють супроводжувати об'єкти за умов невизначеності їх детектування;
- відсутні алгоритми, що дозволяють виконувати підстроювання методу залежно від умов детектування: невизначеності або зашумленості даних.

Таким чином, розробка методів аналізу статичних та динамічних зорових сцен на основі згорткових нейронних мереж, що дозволяють усунути зазначені обмеження, є актуальним завданням .

Мета робочих лежить в розробці нових методів на основі згорткових нейронних мереж для поліпшення точності і ефективності аналізу в статичних і динамічних візуальних сцен в умовах невизначеності їх детектування.

Для цього необхідно вирішити такі завдання :

розробити новий тип СНР, що дозволяє виконати детектування та виділення «глибоких» ознак об'єктів на статичних та динамічних зорових сценах за один прохід СНР;

розробити метод аналізу статичних зорових сцен на основі запропонованого типу СНР;

розробити метод аналізу динамічних зорових сцен на основі пропонованого типу СНР;

розробити алгоритми та програмні засоби, що реалізують запропоновані методи аналізу статичних та динамічних зорових сцен на основі СНР;

виконати оцінку точності та оперативності аналізу статичних та динамічних зорових сцен на основі запропонованих методів та алгоритмів та зіставлення з результатами відомих високоточних та продуктивних методів.

У ході роботи використані такі методи досліджень : теорії розпізнавання образів, теорії штучних нейронних мереж, теорії графів, математичного моделювання, об'єктно-орієнтованого проектування та програмування.

Метод аналізу динамічних візуальних сцен obespechivayu- проводять на високу точність будівництва доріжок, обробки і аналізу даних в режимі реального часу і в умовах невизначеності, як і коли zashum- виявлення ліні даних.

У першому розділі проведено дослідження методів аналізу статичних та динамічних зорових сцен, виділено їх переваги та недоліки та виявлено, що на сьогоднішній день найбільш перспективними є методи, побудовані на основі СНР.

У другому розділі розроблено методи аналізу статичних та динамічних зорових сцен на основі згорткових нейронних мереж. Для реалізації даних методів створено: структуру згорткової нейронної мережі та алгоритм її навчання для аналізу статичних та динамічних зорових сцен, спосіб виділення емпіричного рецептивного поля шару СНР; спосіб обчислення розмірів «якорних» прямокутників багатомасштабної моделі детектування візуальних об'єктів; модель руху об'єкта; спосіб відновлення перепусток об'єктів детектором; спосіб фільтрації детекцій об'єктів

У третьому розділі розроблено алгоритми, що реалізують методи аналізу статичних та динамічних зорових сцен: алгоритм обчислення розмірів «якорних» прямокутників багатомасштабної моделі детектування візуальних об'єктів, алгоритм виділення емпіричного рецептивного поля для кожного шару СНР, алгоритм відновлення пропусків об'єктів детектором, алгоритм фільтрації

детекцій об'єктів, алгоритм виділення «глибоких» ознак детекції об'єкта. Виконано програмну реалізацію розроблених алгоритмів у вигляді бібліотеки програмних функцій.

1 ДОСЛІДЖЕННЯ МЕТОДІВ АНАЛІЗУ СТАТИЧНИХ І ДИНАМІЧНИХ ЗОРОВИХ СЦЕН

1.1 Методи аналізу статичних зорових сцен

1.1.1 Методи виділення об'єктів статичних зорових сцен

В даний час найбільш популярним методом виділення об'єктів на зображенні є підхід, що ґрунтується на ідеї «ковзного вікна».

Двійковий класифікатор послідовно аналізує невеликі області зображення (називаються вікнами), як показано на рис. 1.1, привласнюючи їм мітки «об'єкт» та «не об'єкт». Для виділення об'єктів різного масштабу пошук здійснюється за допомогою побудови зображень різного масштабу.

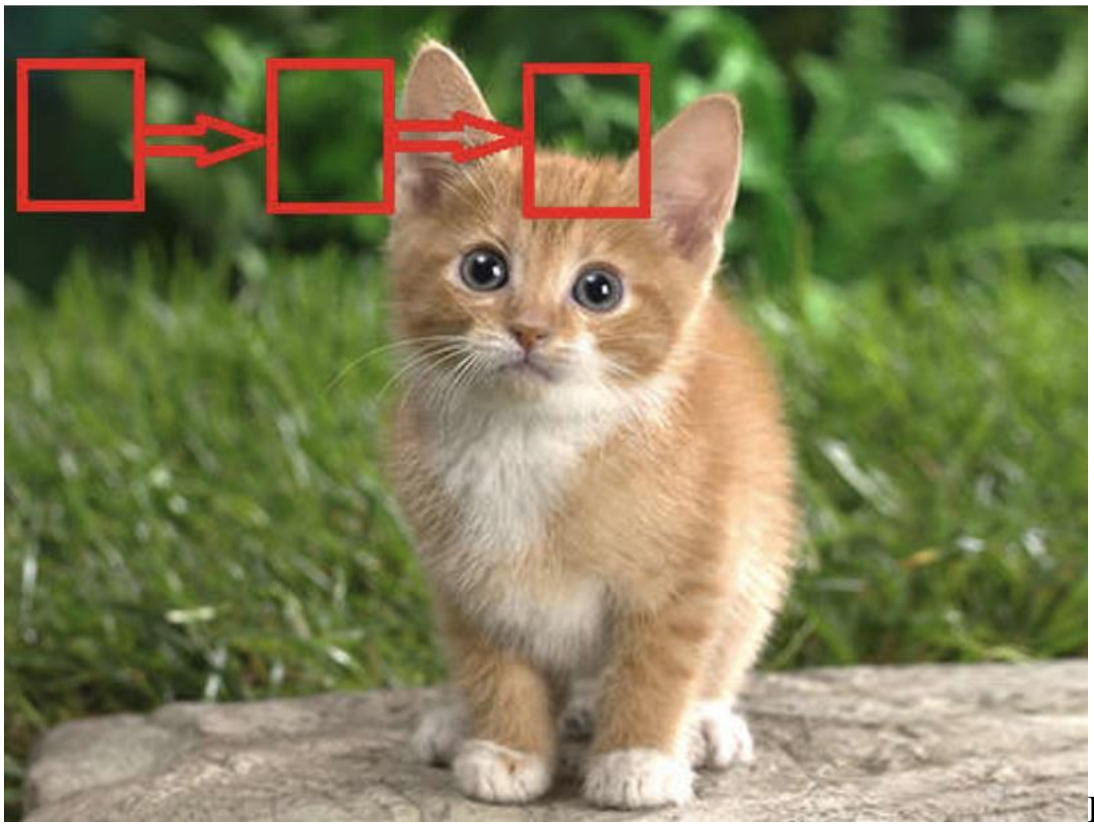

Рисунок 1.1 – Схема роботи методів на основі ковзного вікна

Одна з груп методів з використанням "ковзного вікна" заснована на виділенні об'єктів з використанням шаблонів. Шаблон послідовно накладається на

різні частини зображення та обчислюється кореляція між вихідною областю зображення та шаблоном. Ті ділянки зображення, на яких різницю між двома областями мінімальні, позначаються як шукані. Такі методи не дозволяють з упевненістю сказати , є чи бажаний об'єкт був знайдений або немає, так як результат від методу залежить в значній мірі від масштабу, кута огляду і зображення обертання . Крім того, можливі помилкові спрацьовування, коли вихідного об'єкта на зображенні немає, але є якісь спільні риси.

«Ковзне вікно» може бути використане спільно з методами опису ознак об'єкта, що показано в роботах. Наприклад, як такі ознаки можуть виступати дескриптори локальних особливостей такі як: Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), (Speeded Up Robust Features) (SURF).

У таких дескрипторах ознаки будуються на основі інформації про інтенсивність, колір та текстуру особливої точки. Крім того, особливі точки можуть представлятися кутами, ребрами або навіть контуром об'єкта, тому числення цих ознак виконуються для деякої околиці. Дескриптор є композицією окремих областей зображення (блоків), де для кожного такого блоку розраховуються параметри локальних особливостей.

Переваги таких ознак полягають у тому, що вони дозволяють впоратися із проблемою оклюзій за рахунок своєї локальності. Крім того, ознаки інваріантів до зміни масштабу та орієнтації. До недоліків цих способів є те, що вони НЕ здатні по локалізації в об'єкт в зображенні. Крім того, методи працюють тільки з певним типом локальних здібностей. Наприклад, SURF не працює з об'єктами простої форми та без виражених меж, а SIFT чутливий до зміни освітленості.

В якості атрибутів, що описують об'єкт, також можуть використовуватися кількісно виражені колірні характеристики одного колірного простору (RGB, HSV, LAB). Наприклад, такий підхід може бути використаний для детектування шкіри людини, як описано або для детектування дорожніх знаків. Недоліком цих методів є значний вплив умов освітленості. Крім того, опис об'єкта тільки в

кольорі недостатньо, тому найбільш імовірно , необхідно використовувати додаткові поліпшувалася ознаки.

Інша група методів – методи контурного аналізу [21]. Дані методи базуються на пошуку таких точок зображення, в яких яскравість різко змінюється. Знайдені точки зазвичай поєднуються і утворюють згладжені лінії, які називаються межами. Переваги таких методів у тому, що вони інваріанти щодо обертання, масштабування та зміщення контуру на зображенні. Основні недоліки – однакова яскравість об'єкта та фону або перекриття з іншими об'єктами призводять до того, що контури виділяються неправильно.

Основними недоліками групи методів, що ґрунтуються на використанні «ковзного вікна» є перебір досить великої кількості областей, необхідних виділення об'єктів, і навіть специфічність ознак, із якими «ковзне вікно» найчастіше застосовується. Дані ознаки можна використовувати лише виділення певної групи об'єктів (особи людей, машини, дорожні знаки). Комбінування різних ознак дозволяє застосувати «ковзне вікно» для пошуку об'єктів різних груп, але разом з тим ще більше збільшує обчислювальну складність алгоритму і не дозволяє виконувати обробку в режимі реального часу.

Інша група методів заснована на застосуванні сегментування зображення замість «ковзного вікна», що дозволяє генерувати гіпотези або об'єкти-кандидати «об'єкт», «не об'єкт». Загальна кількість таких гіпотез зазвичай не перевищує кількох тисяч, незалежно від розміру зображення, що знижує обчислювальну складність у порівнянні з «ковзним вікном».

Як такі методи використовуються: метод селективного пошуку (Segmentation as Selective Search), регіональні ознаки для детектування об'єктів (Regionlets for Object Detection), незалежні від категорій гіпотези об'єктів (Category Independent Object Proposals).

Найбільш поширеним методом є селективний пошук (selective search), результат роботи якого показано рис 1.2. В основі даного методу лежить підхід на

основі сегментації зображень на графіках, загальна ідея якого полягає в наступному: кожен піксель зображення є вершиною у графі.

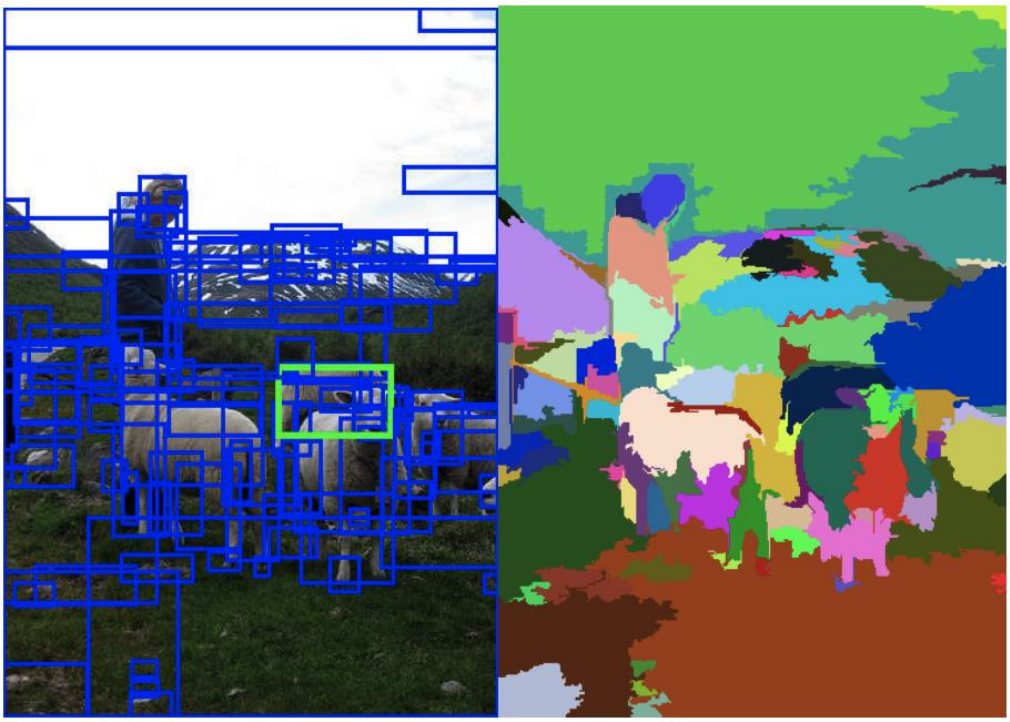

Рисунок 1.2 – Вихідне зображення (ліворуч) та зображення, отримане внаслідок застосування до нього селективного пошуку (праворуч)

Під час виконання сегментації кожен піксель (вершина у графі) поєднується із сусідніми пікселями (вершинами), ребра яких мають найменшу довжину. В результаті такого об’єднання ми отримаємо кілька розсіяних сегментів (підмножин пікселів) з мінімальною загальною вагою всередині. Сегменти об'єднуються між собою, якщо різниця інтенсивностей на їхньому кордоні менша за максимальну різницю всередині кожного з об'єдняних сегментів. В кінцевому підсумку, зображення буде розділений на блоки , відповідні окремим об'єктам.

Крім сегментації на графах, метод селективного пошуку враховує різні масштаби зображень, групує окремі частини зображень не лише за яскравістю, але й за іншими ознаками. Крім того, метод селективного пошуку використовує кілька колірних просторів.

Класичні методи виділення гіпотез є евристично ненавченими методами і не залежать від розв'язуваного завдання, крім того, вимагають значних обчислень і навіть на сучасних комп'ютерах не дозволяють вирішувати задачі в режимі реального часу. Тому розширенням даних методів стала група методів, які використовують спеціально навчену нейронну мережу для генерації гіпотез про місцезнаходження об'єктів на зображенні.

Такими методами є:

- Multibox – на вхід даної мережі подається вихідне зображення, а на виході – гіпотеза, а також значення «впевненості», що у цьому вікні дійсно міститься об'єкт.
- OverFeat – на вхід даної мережі подається вихідне зображення, а на виході виходить карта енергій, що показує у яких місцях зображення найімовірніше знаходження об'єкта.

Дані методи дозволяють вирішувати завдання виділення в режимі реального часу і на сьогоднішній день є найбільш перспективними. Якщо потрібно також розпізнавання об'єкта, необхідно використовувати іншу нейронну мережу, на вхід якої буде надходити гіпотеза від першої мережі. Така комбінація нейронних мереж збільшить кількість обчислень і в деяких випадках вимога обробки в реальному часі не буде виконана. Тому досить перспективними є методи, які дозволяють виконувати виділення та розпізнавання об'єктів з використанням однієї нейронної мережі та бажано за один прямий прохід для всіх гіпотез одночасно. Такі методи будуть розглянуті далі.

1.1.2 Методи розпізнавання об'єктів статичних зорових сцен

Під завданням розпізнавання (класифікації) об'єктів на статичній зорової сцені розуміється зіставлення між виділеним об'єктом та однією з міток заздалегідь заданих класів. Припускаємо, що об'єкт вже виділений прямокутником (bounding box), що покриває даний об'єкт цілком або частково, а також у випадку, якщо в даному пункті не описано, яким чином виділяються ознаки об'єкта, то об'єкт вже описаний за допомогою деякого набору ознак представлених дескрипторами. Методи такого опису було розглянуто раніше.

У вигляді методу розпізнавання об'єктів найбільш ймовірно , використовували методи , засновані на використанні математичної статистики і машинного навчання, які будуть розглянуті.

Одним із методів, що базується на застосуванні математичної статистики, є метод головних компонент (Principal Component Analysis, PCA). Хоча метод найчастіше застосовується для зменшення розмірності даних, він може бути використаний, наприклад, для розпізнавання облич. Ідея методу полягає в лінійному ортогональному перетворенні вхідного вектора (як такий вектор може бути використане зображення об'єкта) P розмірності N у вихідний вектор Q (проекція вхідного вектора) розмірності M , M<N . Для застосування даного методу як класифікатора обчислюються відстань від тестового вектора до його проекції та відстань від цієї проекції до усередненого вектора тренувального набору, що дозволяє віднести об'єкт до одного з класів. Основні недоліки методу: метод не використовує інформацію про належність ознаки до певного класу, а також досить чутливий до умови зйомки об'єктів та освітлення.

Інша група методів, що дозволяє частково усунути недоліки методу головних компонентів, заснована на лінійному дискримінантному аналізі. Загальна ідея методів полягає в тому, що вони вибирають проекцію простору зображень на простір ознак таким чином, щоб мінімізувати внутрішньокласову та максимізувати міжкласову відстань у подорожі ознак. У цих методах передбачається, що класи лінійно розділені. Основний недолік полягає в тому, що такі методи вимагають знаходження зворотних коваріаційних матриць класів, що неможливо, якщо матриці є виродженими.

Ще одна група методів ґрунтується на алгоритмах машинного навчання.

Одним з методів класифікації, що широко використовуються, є машина опорних векторів (SVM), яка досить часто застосовується для класифікації ознак зображення об'єкта, побудованих на основі гістограми орієнтованих градієнтів (HOG).

Завдання розпізнавання за методом опорних векторів полягає в знаходженні такої гіперплощини в n- мірному просторі (n – розмірність вектора ознак, з

використанням якого описуються об'єкти зображення, а кожне зображення – точка в n- мірному просторі), яка відокремлює всі точки, що відповідають зображенням даного класу, від інших не належать йому. Оскільки таких гіперплощин може бути досить багато, метод має на меті відшукати так звану оптимальну гіперплощину, відстань до якої від найближчої точки для кожного з класів максимально.

Недолік цього методу полягає в тому, що він нестійкий по відношенню до шуму у вихідних даних. Якщо навчальна вибірка містить шумові викиди, то вони будуть враховані при побудові роздільної гіперплощини .

Інша група методів ґрунтується на ідеї побудови класифікатора з використанням алгоритму бустингу (boosting) – складний класифікатор будується на основі кількох простих «слабких» класифікаторів. Таким чином кожен наступний класифікатор прагне компенсувати недоліки попереднього. Такий підхід використовується, наприклад, у методі Віоли-Джонса для розпізнавання облич. Основні недоліки бустингу полягають у тому, що він вимагає побудову сотень і навіть тисяч базових класифікаторів для композицій, а також підлаштовується під помилки та викиди у вихідні дані.

Як класифікатор може також використовуватися наївний Байесовський класифікатор, який ґрунтується на припущенні, що ймовірність знаходження окремого дескриптора в класі не залежить від ймовірності знаходження інших дескрипторів у цьому класі. При такому підході до каждого дескриптора вихідне зображення шукається найбільш близьке до нього дескриптору в кожному з класів. У результаті найближчим класом вважається той клас, у якого сума відстаней між вихідними дескрипторами зображення та найближчими буде мінімальна. Істотним недоліком даної моделі є нестійкість результатів роботи класифікатора до незначного погіршення навчальної вибірки – навіть незначна зміна вибірки одного класу з «поганими» даними погіршує роботу всієї системи.

Штучні нейронні мережі досить добре зарекомендували себе для використання в якості класифікатора. Найбільш поширеною нейромережевою моделлю є багатошаровий персептрон. Однак він не надто добре підходить для

вирішення реальних завдань розпізнавання образів. Велика розмірність вхідних зображень призводить до різкого збільшення числа нейронів та синапсичних зв'язків такої мережі. В результаті сильно збільшується час і обчислювальна складність процесу навчання, а в ряді випадків досягти збіжності такої мережі взагалі не вдається. Крім того, звичайний персептрон ігнорує топологи вхідних даних, не враховуючи чітку двовимірну структуру зображень.

В даний час досить популярна, є «глибокої» мережею, Ko torye заснований на «глибокі» навчання, то є навчальні ідеї, а НЕ спеціалізований алгоритм для конкретної задачі. «Глибока» мережу obucha- ється незалежно один від одного зібрати більш складними і «глибоким» абстракцію з даних з збільшенням глибини мережі, такими , як «глибокі» особливість зображення.

Розглянемо «глибокі» мережі довіри, що складаються з обмежених машин Больцмана (є породжує стохастичну нейронну мережу). Кожен рівень такої мережі (обмежена машина Больцмана) навчається з використанням методів «без учителя». Потім вся мережа донавчається з використанням алгоритму зворотного поширення помилки, щоб виконати точне підстроювання ваг. Основний недолік таких мереж для класифікації зображень полягає в тому, що вони не забезпечують інваріантність до зсуву та повинні навчатися окремо для кожної позиції зображення. Введення такої надмірності ускладнює масштабування таких моделей для дослідження зображення повністю.

Усунути зазначені недоліки дозволяє згорткова нейронна мережа (СНР). Вона була запропонована відносно недавно Ян ЛеКуном на основі досліджень зорової кори головного мозку тварини. Структура мережі, являє собою чергування згорткових шарів, шарів субдискретизації та повнозв'язних шарів. Детально принципи роботи цих шарів будуть описані окремо. Вперше така мережа була використана для розпізнавання поштових знаків і являла собою два шари згортки, два шари субдискретизації та один повнозв'язний шар.

На відміну від багатошарових нейронних мереж, де шари повнозв'язні, нейрони згорткових шарів та шарів підвиборки СНР пов'язані не з усіма нейронами попереднього шару, а лише з їхньою частиною, що дозволяє значно

спростити модель. Згорткова нейронна мережа досить добре справляється з вирішенням задачі розпізнавання, але при цьому не позбавлена недоліків: навчання такої мережі потребує значних обчислювальних витрат і часу, а також великої навчальної вибірки. Проте, незважаючи на зазначені недоліки, саме згортка нейронна мережа найчастіше використовується для вирішення проблеми розпізнавання .

Ця подія породила інтерес низки дослідників та появою безлічі робіт. Використання подібних архітектур стало можливим в основному за рахунок збільшення обчислювальних потужностей сучасних комп'ютерів (появою графічних прискорювачів та багатопроцесорних систем).

Принципово структура такої мережі практично не відрізнялася від мережі, запропонованої ЛеКуном, і була 5 згорткових шарів (11×11, 5×5, 3×3), 3 шари субдискретизації, 2 повнозв'язних шари, шар локальної нормалізації, а також softmax -куля дозволяє отримати розподіл ймовірностей міток класу.

Інша мережа VGG-16 містить 13 шарів згортки (3×3 або 1×1), 3 повнозв'язних шари, 5 шарів субдискретизації та softmax-шар. Дана мережа хоч і відрізняється більшою кількістю шарів, ніж попередня, але водночас дозволяє скоротити загальну кількість обчислювальних операцій за рахунок використання двох послідовних згорток 3×3 (еквівалентно одному згортку 5×5) та трьох послідовних згорток 3×3 (еквівалентно 7×7). Крім того, щоб не втрачати інформацію при зменшенні дозволу зображення вдвічі (після застосування шару субдискретизації) кількість фільтрів збільшується вдвічі.

Разом з тим, до мережі також додаються фільтри 1×1, що вносить додаткову нелінійність та зменшує кількість обчислень. Крім того, як показано, такі фільтри можуть бути застосовані як «стискаючі», що дозволяє зменшити глибину шару перед застосуванням більш дорогих згорток (наприклад 3×3) без втрати просторового дозволу, зменшуючи тим самим кількість параметрів мережі. Наприклад, застосування такого підходу дозволяє зменшити кількість параметрів порівняно з AlexNet у 50 разів без втрати точності.

Таким чином, найбільш перспективними методами розпізнавання винаходів на сьогоднішній день є методи, засновані на «глибокому» навчанні та згорткових нейронних мережах, оскільки саме вони дозволяють отримати високу точність розпізнавання порівняно з іншими методами.

1.1.3 Методи детектування об'єктів статичних зорових сцен

У цьому підпункті розглядаються комбіновані нейромережеві методи аналізу статичних зорових сцен, які вирішують завдання як виділення, так і розпізнавання об'єктів.

Одним із перших таких методів став R-CNN (Region-based Convolutional Network). Загальна ідея методу полягає у використанні селективного пошуку, який був описаний раніше, для виділення гіпотез (окремих регіонів зображення). Кожна з таких гіпотез масштабується і проганяється через згорткову нейронну мережу, наприклад AlexNet. Основним недоліком даного методу є велика обчислювальна складність (гіпотез близько двох тисяч), через що для обробки одного зображення навіть на сучасних відеокартах витрачається кілька десятків секунд. Тому метод хоч і забезпечує високу точність, але не може бути застосований у системах реального часу.

Щоб прискорити R-CNN було запропоновано SPP (Spatial Pyramid Pooling).

Прискорення досягається за рахунок того, що зображення спочатку повністю проганяється через згорткову нейронну мережу, отримуючи карту ознак, потім для проекції кожної з гіпотез, одержуваних з використанням селективного пошуку на карту ознак, виконується піраміда пулінгів: 1×1, 2×2 , 3×3, 6×6. Внаслідок чого виходить вектор фіксованої довжини для будь-якої гіпотези, яка потім класифікується.

Швидшим методом проти R-CNN і SPP є Fast R- CNN (Fast Region-based Convolutional Network). Цей метод також використовує селективний пошук для формування гіпотез. У даному методі зображення спочатку повністю проганяється через згорткову нейронну мережу, одержуючи карту ознак, потім кожна з гіпотез проектується на цю карту ознак, і, використовуючи ROI-pooling, масштабується до розміру n ×7×7 та класи- фікується (n- кількість ознак) .Метод

хоч і працює швидше, ніж R-CNN, але все ще не може бути застосований в системах реального часу, по-

Оскільки формування гіпотез займає кілька секунд. Шар ROI-Pooling використовує max-pooling для перетворення ознак усередині регіону інтересу (у даному методі обрана гіпотеза) до фіксованого розміру (n ×7×7). Докладніше принципи роботи даного шару будуть описані в другому розділі.

Для роботи в реальному часі було розроблено метод Faster R-CNN.

Замість використання ненавченого селективного пошуку було запропоновано використовувати RPN (Region Proposal Networks) – спеціальну нейронну мережу для формування гіпотез. Спочатку зображення повністю проганяється через згорткову нейронну мережу, отримуючи карту ознак, потім з використанням згортки 3×3 формуються гіпотези (RPN), використовуючи «якорі» – прямокутники з різним співвідношенням сторін та масштабів. Ці прямокутники використовуються для локалізації пошуку об'єктів. У даному методі таких «якорів» 9 для кожної позиції згортки 3×3. Для кожного такого «якоря» передбачається клас та усунення гаданого знайденого об'єкта за координатами щодо «якоря».

Досить швидким методом є YOLO (You Only Look Once), який дозволяє детектувати зображення на сучасних відеокартах зі швидкістю понад 100 кадрів на секунду, але менш точним, порівняно з Faster R-CNN. Метод заснований на використанні згорткової нейронної мережі. Усі зображення покривається сіткою 7×7. Для кожного осередку сітки будується по два «якорі». Потім передбачається клас та зміщення по координатах щодо осередку сітки. Метод хоч і є досить швидким, але через невеликий розмір сітки не дозволяє розпізнавати досить невеликі об'єкти.

Щоб усунути недолік за точністю YOLO було запропоновано YOLO v2. Метод у порівнянні з YOLO використовує ряд додаткових «трюків» при навчанні нейронної згорткової мережі: аугментація даних, навчання на зображеннях більшої розмірності, навчання на зображеннях декількох масштабів. Для формування гіпотез YOLOv2 використовує «якорі», співвідношення сторін яких

формується на основі методу k- середніх. Крім того, у роботі досліджується оптимальна кількість таких «якорів» для кожного осередку карти ознак згорткової нейронної мережі (використовується 5 «якорів»). На відміну від YOLO, де «якорів» 98, загальна кількість «якорів» у YOLOv2 перевищує 1000. Для кожного такого «якоря», як і в YOLO, передбачається клас та зміщення за координатами щодо осередку сітки.

Компромісом між точністю та швидкістю є багатомасштабна модель детектування візуальних об'єктів Single Shot MultiBox Detector (SSD). Така модель виконує детектування зображень у різних масштабах з використанням кількох детекторів. Кожен такий детектор будується на основі однієї з карт ознак CHP (масштаб об'єктів, що детектуються, збільшується зі зростанням глибини мережі). Потім для кожного з детекторів вибирається безліч «якорних» прямокутників («якорів»), що відрізняються один від одного співвідношенням сторін. Розташування центрів збудованих прямокутників залежить від роздільної здатності картки ознак і не залежить від бази зображень, на якій навчається мережа. Крім того, передбачається усунення кожного побудованого прямокутника від його центру, а також ступінь впевненості, що прямокутник покриває об'єкт заданого класу.

Таким чином, досить перспективними є методи, які дозволяють виконувати виділення та розпізнавання об'єктів з використанням однієї згорткової нейронної мережі за один прямий прохід для всіх гіпотез одразу. Крім того, для роботи в режимі реального часу, не втрачаючи точності, бажано виконувати детектування зображення в різних масштабах на основі кількох карт ознак. Крім того, необхідно вибирати для вирішуваного завдання оптимальні розміри «якорних» прямокутників з використанням статистичних методів, наприклад, методу k-середніх.

1.2 Методи аналізу динамічних зорових сцен

Завдання аналізу динамічної зорової сцени полягає в тому, щоб збудувати треки об'єктів на вхідній послідовності кадрів. Таким чином, вирішується завдання супроводу множини об'єктів на множині кадрів.

У даній роботі, виявлення об'єкта в зображенні є в вигляді області зображення, на якому об'єкт обраний з допомогою кадрування прямого прямокутника.

1.2.1 Класифікація моделей супроводу безлічі об'єктів за ознаками

Класифікація моделей супроводу безлічі об'єктів (MOT, Multiple Object Tracking) може бути виконана за трьома ознаками: як ініціалізується модель, як обробляються дані в моделі, що отримується на виході.

За способом ініціалізації моделі виділяють: супровід на основі детектування (Detection-Based Tracking, tracking-by-detection) та супровід без детектування (Detection-Free Tracking).

У супроводі на основі детектування об'єкти спочатку детектуються з використанням методів детектування, які були розглянуті у попередньому пункті, а потім зв'язуються у треки (асоціація). Це завдання часто формулюється як завдання оптимізації з використанням графа.

Кожен детектований об'єкт представляється у вигляді вершини, а вартість переходу від однієї вершини до іншої визначає функцію подібності. Проблема асоціації на графах може бути сформульована з використанням знаходження максимального потоку або знаходження шляху мінімальної вартості задачами глобальної оптимізації, які вирішуються з використанням лінійного програмування. Також існує формулювання з використанням завдання про призначення у дводольному графі, наприклад, вирішуване, наприклад, з використанням угорського алгоритму. Недоліком цих методів є специфічність підбору ваг графа під кожну конкретну модель. Крім того, у разі вирішення задачі глобальної оптимізації відсутня можливість виконати обробку в режимі реального часу, оскільки потрібна вся послідовність кадрів.

Запропонований підхід, заснований на відстеженні множинних гіпотез (МНТ). МНТ сконструйований з усіх можливих призначень графа виявлення об'єкта на контактах, який потім бере до уваги тільки ті призначення, які не перевищували Non. Вартість призначення визначається за допомогою моделі візуального представлення об'єкта на основі R-CNN та моделі руху на основі фільтра Калмана. Завдання супроводу безлічі треків зводиться до пошуку набору треків з максимальним сумарним набором ваг, яке, як показано в еквівалентне знаходженню зваженої незалежної множини графа. Через таку постановку даний метод не дозволяє виконати обробку в режимі реального часу.

У супроводі без детектування об'єкти ініціалізуються вручну на першому кадрі, а потім супроводжуються. Такі методи рідко використовуються для багаторазового виявлення, оскільки нові об'єкти можуть з'являтися і зникати в наступних кадрах, але в той же час можуть використовуватися для супроводу одного об'єкта, наприклад.

Онлайн-методи використовують для обробки лише інформацію з попередніх кадрів та кадру, який обробляється на даний момент. Такі методи є основою для систем реального часу, де важливе реагування відразу після вступу до системи нового кадру. Дана група методів може використовуватися для систем автономного керування, навігації робіт та систем забезпечення безпеки. На відміну від онлайн-методів оф-лайн-методи оперують відразу всією послідовністю кадрів. При цьому обробка може бути виконана як усіх кадрів одночасно, так і окремих відеопослідовностей у разі обмежень обчислювальних ресурсів та обсягу пам'яті, для кожного з яких вирішується своє оптимізаційне завдання.

За способом отримання виходу розрізняють ймовірнісні та детерміністські моделі. Імовірнісні моделі використовують підхід, що ґрунтується на понятті простору станів. Вважається, що об'єкт, що рухається, має певний внутрішній стан, який вимірюється на кожному кадрі. Щоб оцінити такий стан об'єкта, потрібно максимально узагальнити отримані виміри, тобто. визначити новий стан за умови, що отримано набір вимірів станів на попередніх кадрах. Прикладами

таких моделей є методи, побудовані на основі фільтру Калмана або фільтру частинок. Детерміністські моделі використовують якісні евристики руху (невелика зміна швидкості, незмінність відстані в тривимірному просторі між парою точок, що належать об'єкту), по суті, завдання зводиться до мінімізації функції відповідності наборів точок.

У даному пункті було розглянуто класифікацію моделей супроводу об'єктів за ознаками: яким чином ініціалізується модель, як обробляються дані в моделі, що виходить на виході. Виявлено, що існують досить точні офлайн-методи, які обробляють всю послідовність кадрів цілком. Крім того, для завдання супроводу безлічі об'єктів найчастіше використовуються методи на основі детектування, тому для побудови онлайн-методу досить важливим є вибір точного і водночас швидкого детектора.

1.2.2 Класифікація моделей супроводу безлічі об'єктів за компонентами методу супроводу

Класифікація моделей супроводу безлічі об'єктів також виконується відповідно до різних моделей: моделі візуального представлення (Appearance), руху (Motion), поведінки (Interaction), виключення (Exclusion), перекриття (Occlusion).

Моделі візуального представлення використовують для опису об'єкта деякий набір ознак. Моделі руху досліджують динамічну поведінку об'єкта. Моделі поведінки оцінюють вплив інших об'єктів на даний об'єкт (наприклад, якщо кілька людей рухається в натовпі, людина знаходиться в натовпі, то, швидше за все, вона рухатиметься разом з натовпом). Модель виключення заснована на припущенні, що два об'єкти не можуть одночасно займати одне й те саме положення в просторі. Моделі перекриття враховують перекриття частини або об'єкта іншими об'єктами.

Моделі візуального уявлення об'єктів включають в себе дві складові: зображення об'єкта і функція подібності , яка встановлює ступінь подібності між цими двома об'єктами.

Візуальне представлення об'єкта – це його опис за допомогою певного набору ознак. В якості таких ознак можуть виступати, наприклад, вимірює градієнт колірних атрибутів (за допомогою оператора Собеля), руханя (за допомогою оптичного поля або поля оптичного потоку), межі (за допомогою детектора Кенні), спеціальні точки, дескриптори (HOG, SIFT, SURF). Недоліки та переваги цих ознак для опису об'єктів розглянуті у попередньому пункті.

В даний час в основному застосовується підхід, що ґрунтується на описі об'єктів з використанням «глибоких» ознак. При роботі CHP при переході від шару до шару виконується перехід від конкретних особливостей (знак) зображення до абстрактніших. При цьому CHP самоналаштовується та виробляє ієрархію абстрактних ознак – «глибокі» ознаки (у вигляді послідовності карт ознак).

Наприклад, у роботі пропонується використовувати «сіамську» згорткову нейронну мережу для встановлення зв’язку між двома виявленнями об’єктів у сусідніх кадрах. На вхід мережі надходить 4 джерела інформації, які потім перетворюються на одну роздільну здатність: значення пікселів кожної з детекцій об'єктів у колірному форматі LUV та відповідні їм x та y компоненти оптичного потоку. Структура мережі складається з трьох згорткових шарів, трьох повнозв'язних та бінарного класифікатора, на виході якого формується рішення: чи однакові детекції об'єктів надійшли на вхід чи ні. На останньому повнозв'язному шарі формуються локальні ознаки.

Лише локальних ознак недостатньо для множинного супроводу пішоходів, тому крім згорткової мережі додатково використовується інформація про зміну положення об'єкта між кадрами та його відносної швидкості (ознаки контексту). Локальні та ознаки контексту потім надходять на вхід класифікатора з використанням алгоритму бустингу, на виході якого формується значення схожості між ознаками. Потім виконується асоціація з використанням алгоритмів лінійного програмування. Цей метод є офлайн-методом.

Дескриптор будується на основі мережі, що складається з 2 згорткових шарів та 6 шарів, що використовують залишкові блоки та 1 повнозв'язного шару.

В результаті формується дескриптор розміром 128. Дана мережа навчається на базі, яка включає 1,1 млн зображень 1261 пішохода. Недоліком даного методу є неможливість виконати розпізнавання та супроводження об'єктів з використанням однієї згорткової нейронної мережі, що збільшує час, що витрачається на обробку кожного кадру.

В для супроводження одиночного об'єкта використовується згорткова нейронна мережа на основі CaffeNet. На вхід мережі надходить виділений на попередньому кадрі об'єкт, який необхідно супроводжувати регіон пошуку з поточного кадру. Згорткові шари (використовується перші 5 шарів CaffeNet) виділяють ознаки об'єктів, потім ці ознаки надходять на 3 повнозв'язних шари. Дані шари необхідні для порівняння ознак об'єктів двох зображень та знаходження розташування цільового об'єкта. На виході мережі формується положення знайденого об'єкта щодо пошуку.

Для обчислення дескриптора візуального представлення об'єкта використовується 16-шарова нейронна мережа VGG-16, в якій видалено останній повнозв'язний шар і замінено іншим шаром розміром 500. Крім того, для кожного об'єкта на попередніх кадрах враховується передісторія: дескриптор візуального представлення об'єкт кожному попередньому кадрі надходить на вхід рекурентної нейронної мережі на основі LSTM , формуючи дескриптор, який потім з'єднується з дескриптором візуального представлення об'єкта на поточному кадрі з наступним повнозв'язним шаром, формуючи підсумковий дескриптор. Для донавчання такої мережі додається шар «софт-макс» класифікатора з виходом 0 і 1, який визначає, продовжує цей об'єкт траєкторію чи ні. Структура побудови дескриптора показана на рис. 1.6.

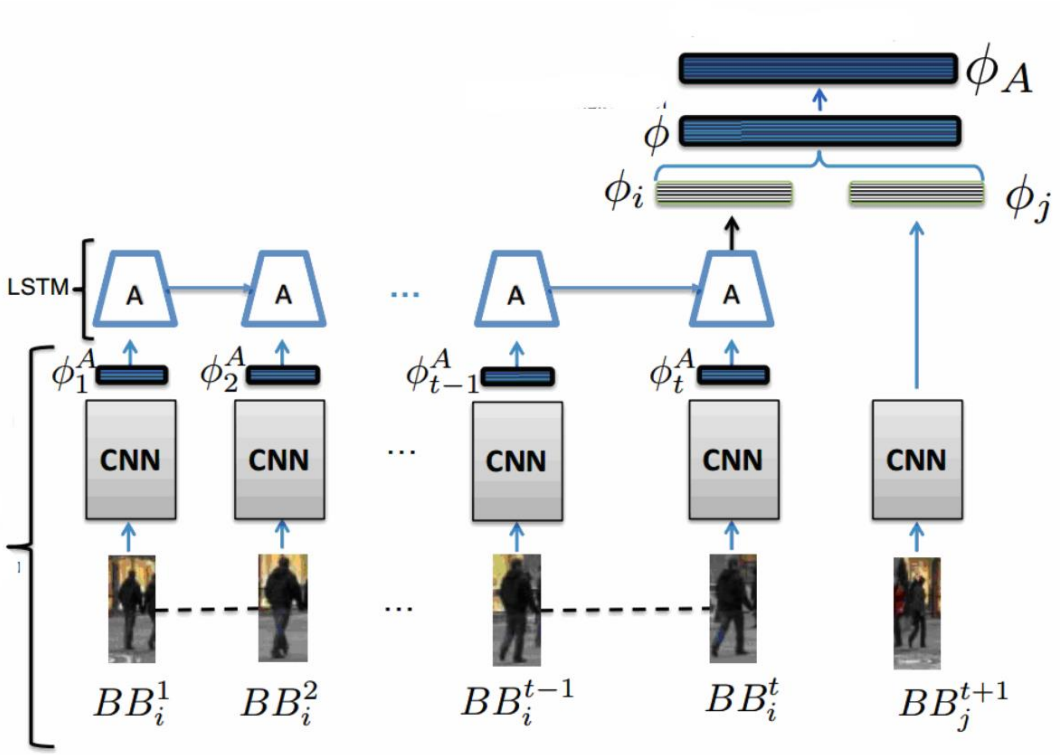

Рисунок 1.6 – Структура побудови дескриптора

Одиночний об'єкт супроводжується з використанням перших 7 шарів згорткової нейронної мережі AlexNet для генерації ознак детекцій об'єктів (формуються на основі останнього шару повного зв'язку), до якої додатково додається наївний байесовський класифікатор. Оскільки ця мережа призначена для класифікації об'єктів, то виконується її навчання у два етапи. На першому етапі мережа донавчається на вибірці з ImageNet 2014 для детектування об'єктів, яка включає частини зображень, на яких присутній тільки даний об'єкт (позитивні приклади) і частини, які мають мале перетин з даним об'єктом (негативні приклади). Після цього етапу донавчання мережа генерує максимальні значення активацій нейронів шару ознак, якщо на ній є один з об'єктів, на якому вона навчалася. На другому етапі мережа навчається генерувати максимальні значення активацій ознак тільки для об'єкта, який необхідно супроводжувати. Для цього на першому кадрі виділяються частини зображення, що мають досить велике перетин з даним об'єктом (позитивні приклади) та частини зображення, на яких відсутній об'єкт (негативні приклади), потім на основі такої вибірки виконується

донавчання. Крім того, додатково донавчання проводиться у процесі супроводу об'єкта.

Крім моделі візуального представлення об'єкта важливим компонентом методу супроводу є модель руху. Така модель досліджує динамічну поведінку об'єкта, визначає, яким чином рухався об'єкт, і дозволяє передбачити, де він перебуватиме в наступний момент часу. В основному використовуються лінійні моделі руху з постійною позицією (об'єкт не рухається), постійною швидкістю або постійним прискоренням.

В Enhancing Linear Programming (ELP) завдання супроводу безлічі об'єктів вирішується з використанням моделі руху на основі лінійної регресії та знаходження шляху мінімальної вартості, причому знаходження даного шляху виконується ітеративно. Спочатку треки формуються без використування моделлю руху – вартості призначення детекцій об'єктів на різних кадрах на треки обчислюються за допомогою відстані між детекціями об'єктів (у пікселях) та часу (вимірюється в кількості кадрів між детекціями об'єктів, що призначаються на один трек). Далі детекції об'єктів, призначені на треки, оцінюються з використанням моделі лінійної регресії. Така оцінка полягає в припущенні, що рух у короткі проміжки часу може бути представлений у вигляді лінійної моделі руху. Зв'язки детекцій об'єктів та треків, що не відповідають лінійній моделі, видаляються. Потім модель додаються потенційні зв'язку між детекціями об'єктів, оцінюється їх вартість. Ці зв'язки передбачаються на основі моделі регресії, якщо вони задовольняють тимчасові обмеження. Такі зв'язки дозволяють поєднувати треки між собою. Цей метод є офлайн-методом.

Завдання супроводу можна розглядати як проблему теорії управління, оцінюючи стан системи на підставі послідовності зашумлених вимірів.

Типовими прикладами таких методів є методи на базі фільтра Калмана та фільтра частинок (particle filter). При використанні фільтра Калмана передбачається, що стан – випадкова величина з нормальним розподілом, а у разі фільтра частинок розподіл задається набором можливих значень стану із зазначенням частот виникнення.

Наприклад, у SORT для передбачення нового стану об'єктів використовується фільтр Калмана. При цьому під станом розуміється набір характеристик: координати центру прямокутника, що обрамляє, його площа та співвідношення сторін. У цій моделі враховується швидкість зміни положення центру прямокутника і площа, причому співвідношення сторін вважається незмінним. Під станом розуміється дещо інший набір характеристик: координати центру прямокутника, що обрамляє, співвідношення сторін і його ширина, при цьому враховується швидкість зміни всіх величин.

Таким чином, для побудови досить точного методу супроводу безлічі об'єктів необхідно використовувати кілька компонент методу, наприклад модель візуального представлення об'єкта та модель руху. При цьому для підвищення точності бажано будувати модель візуального представлення об'єкта на основі СНР, причому для підвищення швидкодії бажано, щоб детектування об'єктів та виділення їх ознак відбувалося за один прохід даної нейронної мережі.

1.3 Дослідження штучних нейронних мереж для аналізу статичних та динамічних зорових сцен

Останні кілька років для вирішення завдань розпізнавання образів як на статичних зображеннях, так і в динаміці все частіше застосовуються «глибокі» згорткові нейронні мережі (СНР). Такі мережі характеризуються значним збільшенням точності розпізнавання, порівняно з класичними методами. Так, наприклад, як показано, вперше застосування «глибокої» СНР дозволило зменшити середню помилку розпізнавання приблизно в півтора рази в порівнянні з одним з кращих методів, що вирішує задачу класифікації зображень з використанням векторів Фішера та Scale-Invariant Feature Transform (SIFT).

Під час вирішення завдань розпізнавання однією з найбільш значущих проблем є забезпечення необхідної точності, тому більшість робіт присвячена вирішенню саме цього завдання. Наприклад, забезпечення необхідної точності для завдання класифікації об'єктів присвячені роботи, детектуванню об'єктів. В

основному це досягається за рахунок збільшення кількості шарів, через що використовуються значні ресурси як у процесі навчання нейронної мережі, так і під час роботи. Тому сучасні детектори, наприклад, Faster R-CNN хоч і є досить точними, проте не дозволяють розпізнавати об'єкти на зображенні в режимі реального часу. Інші детектори, наприклад, You Only Look Once (YOLO) працюють у режимі реального часу, однак, не є достатньо точними і не дозволяє розпізнати невеликі об'єкти на зображеннях.

Тому іншою проблемою є мінімізація обчислювальної складності та вимога обробки в режимі реального часу [а також забезпечення компромісу між точністю та швидкістю обробки. Таким компромісом є багатомасштабна модель детектування візуальних об'єктів Single Shot MultiBox Detector (SSD) . Однак дана модель має складності налаштування: розміри «якорних» прямокутників залежать від бази зображень, на яких навчається мережа, і вибір цих розмірів є невирішеним питанням цієї моделі. Таке налаштування потребує підбору розмірів «якорних» прямокутників під кожну конкретну навчальну базу, що дещо обмежує застосування даної моделі.

Таким чином, навіть одна проблема є спрощення конфігурації мережі для конкретної тренувальної бази і зменшення необхідної кількості obu- претендентів прикладів. Таким чином, більшість сучасних архітектур SNA stroit- сь на основі вже існуючих, таких як і адаптовані для конкретного завдання , шляхом додавання соляної шарів і додаткове навчання (тонка настройка).

Завдання супроводу об'єктів найчастіше вирішується окремо від детектування і при спільному використанні детектора та методу супроводу потрібно спочатку детектувати об'єкт, потім виділити ознаки окремо для кожного об'єкта, через що зростає складність обробки. Наприклад, у кожна детекція об'єкта окремо подається на вхід CHP для отримання її "глибоких" ознак. Такий підхід дозволяє підвищити точність, але потребує значних обчислювальних ресурсів.

Крім того, існують досить точні методи супроводу об'єктів, наприклад, але вони не дозволяють виконати обробку в режимі реального часу, оскільки

вирішують завдання глобальної оптимізації та вимагають отримання одразу всієї послідовності кадрів. Інші методи виконують обробку в режимі реального часу, але не є достатньо точними, наприклад.

При цьому для дослідження розроблених методів супроводу найчастіше існуючі бази для тестування пропонують працювати з великою кількістю помилкових детекцій об'єктів та будувати алгоритми, що дозволяють виділяти дані з шуму. Однак при використанні СНР іноді доводиться мати справу не з великою кількістю помилкових детекцій об'єктів, а з перепустками.

Таким чином, найбільш перспективними методами для вирішення задач розпізнавання та супроводу об'єктів на сьогоднішній день є методи, засновані на «глибокому» навчанні та згорткових нейронних мережах, оскільки саме вони дозволяють отримати високу точність розпізнавання та супроводу, порівняно з іншими методами.

Тому при вирішенні комплексного завдання детектування та супроводу об'єктів доцільно запропонувати метод, що дозволяє виконати детектування та побудову моделі візуального представлення об'єкта за один прохід СНР, в умовах невизначеності їх детектування.

1.4 Аналіз існуючих способів виділення рецептивного поля СНР

На відміну від багатошарових нейронних мереж, де шари пов'язані, нейрони згорткових шарів і шарів підвиборки СНС пов'язані не з усіма нейронами попереднього шару, а тільки з їх частиною [1]. Тому однією з найважливіших характеристик згорткової нейронної мережі є рецептивне поле (поле сприйняття) і розмір рецептивного поля – кількість зв'язків нейрона на даному шарі з нейронами вхідного шару.

Визначення розміру рецептивного поля на кожному з шарів важливо для правильної роботи згорткової мережі в задачах сегментації та детектування об'єктів, де на відміну від розпізнавання зображення необхідний пошук локального, а не глобального контексту.

У роботах вводиться поняття теоретичного рецептивного поля та його розміру, що визначається як кількість зв'язків нейрона на даному шарі з нейронами вхідного шару.

На основі теоретичного обґрунтування розмір рецептивного поля всієї мережі може бути збільшений лінійно шляхом збільшення кількості згорткових шарів або мультиплікативно за допомогою збільшення шарів вибірки. Наприклад, для мережі ImageNet-CNN (структура якої показано на рис. 1.7) шару pool-5 теоретичний розмір рецептивного поля дорівнює 195 пікселям, при подачі на вхід цієї мережі зображення розміром 224×224 пікселів.

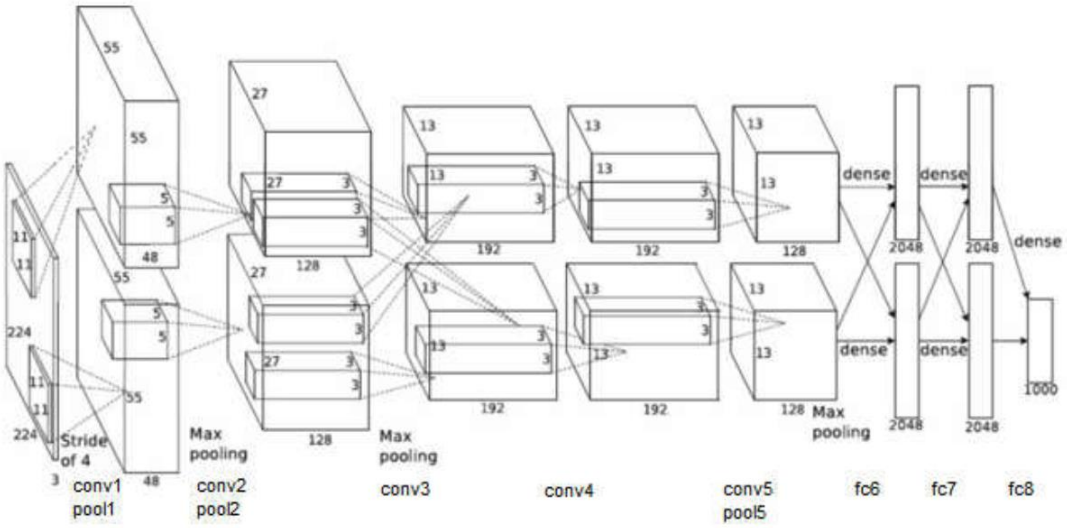

Рисунок 1.7 – Структура згорткової мережі ImageNet- CNN

Однак деякі нейрони перших шарів можуть не впливати на нейрони останніх шарів, при цьому будуть пов'язані з ними.

Досліджується емпіричний розмір рецептивного поля нейрона для згорткових мереж ImageNet-CNN і PlacesNet-CNN (відмінюється від ImageNet-CNN лише базою зображень, на якій навчається СНР). База 200 тисяч зображень буде використовуватися в якості зразка тестування . З цієї бази вибирається 20 зображень, що мають максимальну активацію нейрона заданого шару. Потім з використанням ковзного вікна на кожне зображення накладається фільтр 11×11 з випадковим заповненням та кроком 3 (більше 5000 зображень). Порівнюючи

активацію обраного нейрона оригінального зображення з zashumlen- них зображень , побудованої картою розбіжностями (невідповідність карти) для всього зображення. Потім, щоб зібрати інформацію з 20 зображень, побудована карта центрується для кожного зображення і усереднюється.

Алгоритм не наданий , дозволяє отримати піксельне зображення в емпірично сприйнятливому полі карти невідповідності на основі SNA.

Таким чином, доцільно запропонувати спосіб виділення емпіричного рецептивного поля CHP, що відрізняється від розглянутих адаптивним вибором порога включення частин вихідного зображення в рецептивне поле, а також скороченням кількості зашумлених зображень, необхідних для навчання нейронної згорткової мережі.

1.5 Постановка завдання дослідження

Розгляд існуючих методів аналізу статичних та динамічних зорових сцен виявили такі основні обмеження:

Немає методів, що дозволяють виконувати виявлення об'єктів з сокой точністю в режимі реального часу;

відсутня універсальна процедура налаштування методу детектування, що дозволяє забезпечити компроміс між точністю та швидкістю;

відсутній метод супроводу та детектування об'єктів, що дозволяє виконати детектування та виділення «глибоких» ознак об'єктів за один прохід CHP;

відсутні алгоритми, що дозволяють супроводжувати об'єкти в умовах невизначеності їх детектування;

Там немає алгоритмів , які дозволяють виконувати настройки методи в залежності рії від умов виявлення: невизначеність або шумове забруднення в даних.

Результати проведеного аналізу, а також потреби практики дозволяють конкретизувати та формалізувати завдання дослідження.

Потрібно детектувати та супроводити об'єкти в умовах:

- переміщення камери відсутнє (статичність);
- точність детектування не повинна бути меншою θ ;
- загальна кількість треків, на яких об'єкти повністю не відстежені, не повинна перевищувати ε ;
- обробка повинна бути виконана в реальному часі зі швидкістю не менше кадрів за секунду на сучасній відеокарті з підтримкою CUDA;
- можливі перекриття елементів об'єктів;
- об'єкт НЕ відстежується, якщо вона покрита іншим об'єктом більш докладної , ніж naро- половині програми;
- об'єкт повинен повністю поміщатися у кадр;
- переміщення та зміна розмірів об'єктів між сусідніми кадрами не має перевищувати *∂* від розмірів цього об'єкта;
- переміщення об'єкта, зміна його розмірів на n будь-яких поспіль тих, що йдуть (Послідовних) кадрах підпорядковується лінійному закону руху.

У четвертому розділі будуть дані чисельні значення кожного з параметрів для вирішення конкретного практичного завдання (детектування та супроводження силуетів людей).

Варто зазначити, що існуючі методи вирішують лише одне із завдань (супровід або детектування об'єктів). Відповідно до постановкою завдання, необхідно вибрати в комбінації методів , які зробили б можливим виконати виявлення і відстеження об'єктів в відповідно до необхідними умовами.

Для досягнення поставленої мети необхідно вирішити такі завдання:

- розробка нового різновиду СНР, що дозволяють виконувати детектирування та виділення «глибоких» ознак об'єктів на статичних та динамічних зорових сценах за один прохід СНР;
- розробка методу аналізу статичних зорових сцен;
- розробка методу аналізу динамічних зорових сцен;
- розробка алгоритмів та програмних засобів для реалізації запропонованих методів аналізу статичних та динамічних зорових сцен на основі СНР;

- оцінка точності та оперативності аналізу статичних та динамічних зорових сцен на основі запропонованих методів та алгоритмів.

1.6 Висновки

Проведено аналіз методів виділення об'єктів статичних зорових сцен, виділено їх переваги та недоліки та виявлено, що на сьогоднішній день найбільш перспективними є методи, побудовані на основі CHP.

Розглянуто методи класифікації зображень та зроблено висновок, що методи, засновані на «глибокому» навчанні та згорткових нейронних мережах, дозволяють отримати високу точність розпізнавання в порівнянні з іншими методами.

Виявлено, що для ефективного вирішення задачі детектування об'єктів на зорових сценах можна використовувати комбінування нейронних мереж. Однак це призводить до збільшення кількості обчислень і в окремих випадках вимога обробки в реальному часі не буде виконано.

Здійснено аналіз методів, що дозволяють детектувати об'єкти на зорових сценах з використанням однієї згорткової нейронної мережі за один прямий прохід одразу для всіх гіпотез. Зроблено висновок, що для роботи в режимі реального часу, без втрати точності, бажано виконувати детектування зображення в різних масштабах на основі кількох карт ознак. Крім того, необхідно вибирати для розв'язуваного завдання оптимальні розміри «якорних» прямокутників з використанням статистичних методів, наприклад, методу k- середніх.

Розглянуто методи аналізу об'єктів динамічних зорових сцен та виявлено, що існують досить точні офлайн-методи, які обробляють всю послідовність кадрів цілком. Крім того, аналіз методів супроводу безлічі об'єктів показав, що найчастіше використовуються методи на основі детектування, тому для побудови онлайн-методу досить важливим є вибір точного і водночас швидкого детектора.

Виявлено, що з розробки методу аналізу динамічних зорових сцен потрібно реалізувати моделі візуального уявлення та руху об'єктів. При цьому для

підвищення точності необхідно будувати модель візуального представлення об'єктів на основі СНР, причому для підвищення швидкодії доцільно, щоб детектування об'єктів та виділення їх ознак виконувалося за один прохід даної нейронної мережі.

Здійснено постановку завдання дослідження, що полягає в розробці методів та алгоритмів виділення та детектування об'єктів та їх треків для статичних та динамічних зорових сцен на основі згорткових нейронних мереж.

2 РОЗРОБКА МЕТОДІВ АНАЛІЗУ СТАТИЧНИХ І ДИНАМІЧНИХ ЗОРОВИХ СЦЕН НА ОСНОВІ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ

2.1 Розробка згорткової нейронної мережі для аналізу статичних та динамічних зорових сцен

2.1.1 Структура та опис СНР для аналізу статичних та динамічних зорових сцен

Для аналізу статичних та динамічних зорових сцен пропонується згорткова нейронна мережа, структура якої представлена рис. 2.1. На відміну від відомих рішень (наприклад, Single Shot MultiBox Detector – SSD), структура розробленої СНР, крім багатомасштабної моделі детектування візуальних об'єктів, включає шар виділення «глибоких» ознак знайдених і передбачених детекцій об'єктів – ROI_Pooling, а також використовує пропоновану модель візуального представлення об'єктів, робота якої розглянута у цьому розділі.

Опишемо докладніше структуру багатомасштабної моделі детектування візуальних об'єктів. Дана модель побудована на основі СНР типу VGG-16, яка містить у собі 5 блоків згорткових шарів з шаром субдискретизації (max-pool) після кожного блоку, 3 повнозв'язних шари нейронів та шар класифікатора. У двох перших блоках міститься по два шари, а в інших трьох – по три шари згортки 3×3.

Для використання СНР типу VGG-16 як детектор виключається останній повнозв'язний шар, шар класифікатора та додатково включається кілька блоків детектування, а також шарів, що використовуються для зміни масштабу аналізованого зображення. Так, до карти ознак 4-го блоку (conv4_3) та повнозв'язного шару (FC7) VGG-16 додається по одному детектору. Крім того, до мережі додатково додається ще чотири блоки, що використовуються для зміни масштабу (conv8_2, conv9_2, conv10_2, conv11_2) з детекторами, як показано на рис. 2.1. На виході мережі результати роботи детекторів кожного блоку (всього 6

детекторів) аналізуються з використанням алгоритму придушення немаксимумів (Non-Maximum Suppression) і формується результат детектування.

Блок детектора складається з: шару генерації «якорних» прямокутників, шару локалізації (передбачає координати зміщення щодо центру згенерованого прямокутника, а також зміщення по довжині та ширині) та шару, на виході якого визначається ймовірність того, що у вихідному прямокутнику міститься об'єкт кожного із класів. Наприклад, якщо є 21 клас (20 класів і фон) і генерується 5×5×4 = 100 «якорних» прямокутників (4 «якорні» прямокутники для кожної позиції картки ознак), то розмірність вихідної картки ознак шару локалізації – 5×5×16 (4 прямокутники × 4 координати усунення), шару ймовірності – 5×5×84 (21 клас × 4 прямокутники).

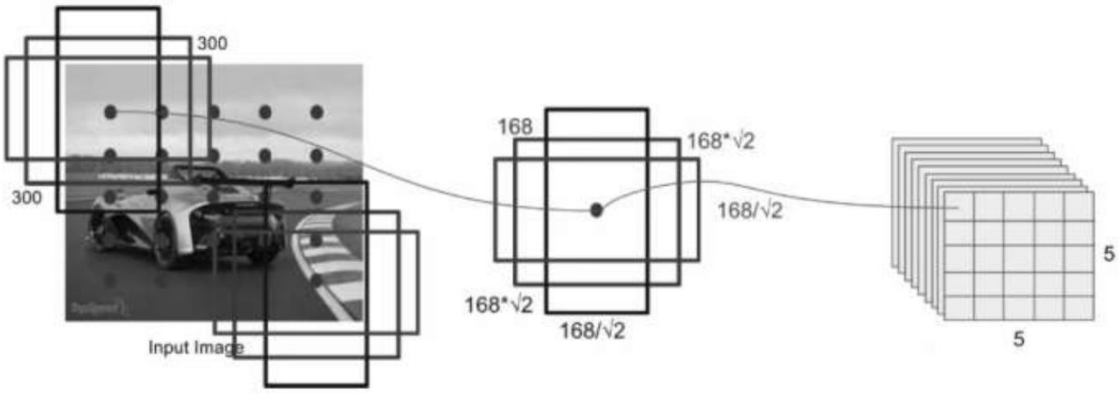

Рисунок 2.1 – Покриття вихідного зображення сіткою «якорних» прямокутників для карт ознак 5×5

Кількість прямокутників, які формуються шаром генерації «якорних» прямокутників, залежить від розмірності карти ознак блоку, на основі якого будується детектор. Такі прямокутники покривають усі зображення сіткою, як показано на рис. 2.1.

Такий спосіб генерації прямокутників не є універсальним, оскільки для іншої навчальної бази буде потрібний вибір інших масштабів «якорних» прямокутників. Тому необхідна розробка такого способу, який відрізняється від

існуючого вибором розмірів «якорних» прямокутників для конкретної навчальної бази, наприклад, з використанням методу кластеризації k- середніх для кожного блоку детектора.

Крім того, масштаби, на основі яких обчислюються розміри «якорних» прямокутників, можна обчислити на основі виділення емпіричного рецептивного поля кожного блоку детектора.

Згортковий шар характеризується кількістю фільтрів згортки n , шириною w та висотою h фільтра (w×h×n). Ще одним параметром запропонованої СНР є кількість каналів у фільтрі, яка береться дорівнює кількості каналів вхідної матриці. Наприклад, на першому шарі така кількість дорівнює трьом (оскільки зображення представлене у форматі RGB). Згортковий шар застосовує операцію згортки з різним ядром (наприклад, 3×3) до вхідної матриці із заданим кроком (наприклад, 1), тобто накладає фільтр. Параметри такого фільтра налаштовуються при навчанні СНР. Результатом згортки є вихідна матриця (тривимірний тензор), яка називається картою ознак (feature map). Розміри карти ознак, що отримуються після кожного шару в даній моделі, показано на рис. 2.1, наприклад, для шару conv4_3 карта ознак має розмірність 38×38×512.

За кожним згортковим і повнозв'язним шаром у даній моделі (крім згорткових шарів, що входять до блоків детекторів) слід ReLU шар, що дозволяє ввести нелінійність у модель. Вихідні значення для кожного елемента x даного шару визначаються таким чином:

y = m a x (x 0 ,).

Як було описано раніше, в СНС типу VGG-16 після кожного блоку згорткових шарів слід блок субдискретизації (max-pool). Цей шар виконує пошук максимального елемента всередині вікна, що розглядається . Цей фільтр переміщається по вхідній матриці аналогічно згорткового шару.

Повнозв'язний шар можна розглядати як окремий випадок згорткового шару з параметрами w = h = 1 .

Шар ROI-Pooling проектує детекції знайдених та передбачених об'єктів на карту ознак шару conv4_3, з подальшим масштабуванням до розмірності 7×7×512, аналогічно max-pool.

Розглянемо входи та виходи запропонованої структури. На вхід надходить вхідне зображення RGB (кадр), яке перетворюється на роздільну здатність 300×300, а також детекції об'єктів, отримані за допомогою способу відновлення пропусків об'єктів детектором (передбачені детекції об'єктів), які формуються на основі попередніх кадрів (тільки для динаміки) .

На виході формуються детекції об'єктів (клас, упевненість детектуван- ня, координати об'єктів) та їх «глибокі ознаки».

Таким чином, застосовуючи пропонований різновид СНР для аналізу статичних і динамічних зорових сцен, з'являється можливість детекобробка зображень і виділення «глибинних особливостей» виявлення об’єктів за один прохід.

2.1.2 Модель візуального представлення об'єкта та спосіб виділення «глибоких ознак» його детекції

На рис. 2.2 показано структуру моделі візуального представлення об'єкта з використанням багатомасштабної моделі детектування, до якої додатково додається шар ROI-Pooling.

Запропонована модель візуального уявлення є описом об'єкта з допомогою набору «глибоких знаків», утворюється при врожайності де шар ROI_Pooling. Варто зазначити, що існуюча модель візуального представлення – це опис всього зображення з використанням ознак одного з шарів СНР типу VGG-16 (наприклад, шару conv4_3).

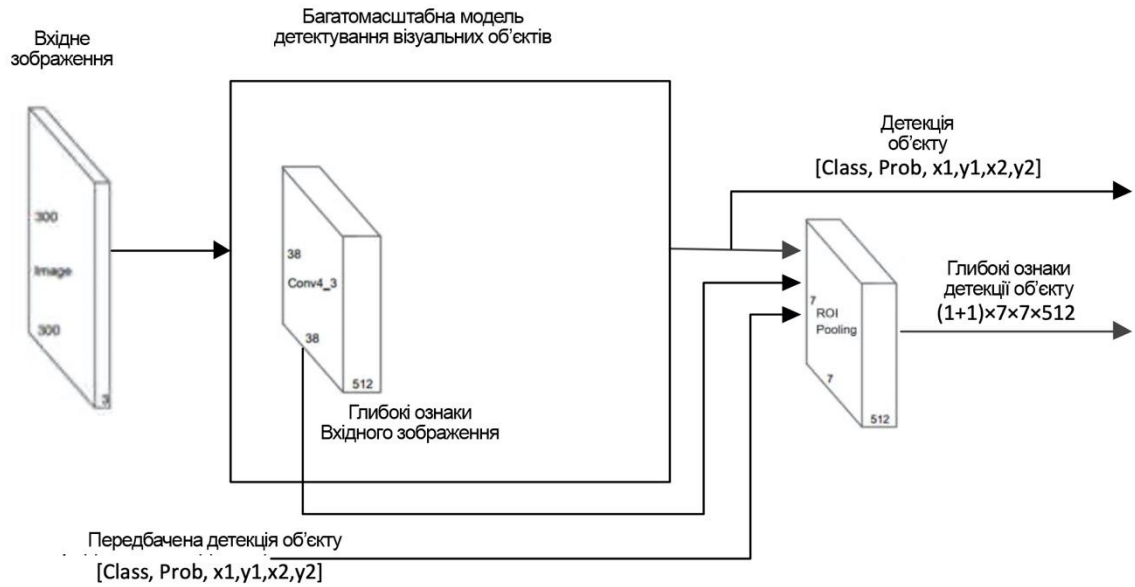

Рисунок 2.2 – Модель візуального представлення об'єкту

Таким чином, виділення «глибоких ознак» кожного з об'єктів за допомогою існуючої моделі виконується шляхом подання зображення об'єкта на вхід мережі та прямого проходження через мережу типу ВГГ-16. Тобто , якщо потрібно виділити ознаки кількох об'єктів, необхідно виконати кілька прямих проходів.

Запропонована модифікація дозволяє виділити ознаки відразу всіх об'єктів за один прямий прохід СНР нового типу.

Як ознаки зображення в даній моделі використовуються ознаки шару conv4_3, які надходять на вхід шару ROI-Pooling. Крім того, на вхід даного шару надходять детекції об'єктів, знайдені з використанням багатомасштабної моделі детектування, а також передбачені детекції об'єктів, отримані за допомогою способу відновлення перепусток. На виході цього шару , «глибокі ознаки» кожен з утворюються виявлень.

2.2.2 Опис методу

Даний метод дозволяє навчити багатомасштабну модель детекту- вання візуальних об'єктів у пропонованій структурі СНР, детектувати об'єкти на статичних зорових сценах та виділяти їх «глибокі ознаки» з використанням даної структури.

Запропонований метод складається з наступних етапів.

Етап 1. Навчання багатомасштабної моделі детектування візуальних об'єктів.

Даний етап призначений для навчання багатомасштабної моделі детектування візуальних об'єктів, після чого можливе детектування об'єктів та виділення їх «глибоких ознак». Проте, розміри формуються

«якорних» прямокутників не є оптимальними для кожної конкретної навчальної вибірки, тому для збільшення точності детектування необхідно виконати етапи 2-4 розробленого методу.

Етап 2. Виділення емпіричного рецептивного поля для кожного шару перед блоком детектора.

Даний етап призначений для визначення мінімальних і максимальних розмірів об'єктів, що детектуються, для кожного шару. Дані розміри обчислюються за допомогою виділення емпіричного рецептивного поля для кожного шару перед блоком детектора, потім розміри цього поля обчислюються.

Під час виконання результату отримано емпіричний рецептивний розмір поля для кожного шару як раз і буде дорівнювати розміру виявлених об'єктів на зображенні.

Етап 3. Розрахунок розмірів «якірних» прямокутників для кожного шару перед блоком детектора.

Цей етап призначений для формування оптимальних розмірів

"якорних" прямокутників, отриманих з використанням статистичного методу кластеризації k- середніх.

Етап 4. Перенавчання багатомасштабної моделі детектування візуальних об'єктів.

Після зміни розмірів «якорних» прямокутників, що генеруються, для кожного шару перед блоком детектора необхідно перенавчання багатомасштабної моделі детектування візуальних об'єктів, яке виконується на даному етапі. В результаті виконання етапу отриману модель можна використовувати для детектування об'єктів та виділення їх «глибоких ознак».

Етап 5. Виділення «глибоких ознак» та детектування об'єктів на зображенні.

Даний етап дозволяє детектувати об'єкти на статичних зорових сценах та виділяти їх «глибокі ознаки» з використанням запропонованої структури СНР.

Таким чином, запропонований метод відрізняється від відомих:

- виділенням емпіричного рецептивного поля;
- збільшує точність способом обчислення розмірів «якорних» прямокутники для кожного шару перед блоком детектора;
- виділенням "глибоких" ознак детекцій об'єктів.

Для виконання етапів 2 і 3 розроблено спосіб виділення емпіричного рецептивного поля та спосіб обчислення розмірів «якорних» прямокутників для багатомасштабної моделі детектування візуальних об'єктів.

тов. Для виконання етапу 5 запропоновано модель візуального представлення об'єкта.

2.2.3 Навчання багатомасштабної моделі детектування візуальних об'єктів

Багатомасштабна модель детектування візуальних об'єктів навчається у два етапи. На першому етапі навчається СНР типу VGG-16 з використанням стохастичного градієнтного спуску за методом зворотного поширення помилки з використанням бази зображень ILSVRC CLS-LOC [82], завдяки чому налаштовуються ваги СНР. Після навчання дана нейронна мережа може працювати як класифікатор зображень.

На другому етапі навчаються шари багатомасштабної моделі детектування, які не входили в СНР типу VGG-16, а також донавчається дана СНР. Для цього ваги шарів, отримані після навчання СНР, копіюються в цю модель.

Донавчання (finetuning) проводиться з наявної основі зображень, тобто. базі тих об'єктів, які необхідно детектувати, наприклад, VOC2007 з використанням стохастичного градієнтного спуску за методом зворотного розповсюдження помилки. Таке донавчання дозволяє не тільки скоротити кількість ітерацій навчання, але й використовувати бази з невеликою кількістю зображень (кілька тисяч), а також підвищити точність детек- тування об'єктів після навчання. Також для скорочення часу навчання набір із бази зображень зберігається у вигляді бази

даних у форматі LMDB. З іншого боку, у процесі навчання (кожні n ітерацій) зменшується швидкість навчання у k раз.

Під час навчання використовується наступна стратегія зіставлення «якорних» і прямокутників, що обрамляють, навчальної вибірки: як позитивні приклади вибираються ті «якірні» прямокутники, які мають перекриття за метрикою IOU з прямокутниками даного класу, що обрамляють, не менше 0,5.

Для кожного «якорного» прямокутника окремо вважається втрата локалізації (помилка положення «якорного» прямокутника) та втрата вибору класу (помилка вибору об'єкта одного класу замість іншого). Загальна функція втрат вважається виходячи з виваженої суми кожної з втрат.

Після зіставлення «якорних» і прямокутників, що обрамляють, навчальної вибірки більшість «якорних» прямокутників є від'ємними прикладами. Тому, щоб зменшити дисбаланс між позитивними та негативними прикладами під час навчання, вибираються позитивні та негативні приклади у співвідношенні 1:3. Крім того, вибираються приклади, які мають найбільше значення функції втрат вибору класу.

В цілях , щоб зробити модель більш надійної, додаткові стратегії аугментации дані використовуються в процесі додаткового навчання. До кожного зображення для донавчання випадково застосовується одне з таких перетворень: вибір зображення повністю, вибір частини зображення, що перекривається з об'єктом, випадковий вибір частини зображення. Крім того, до вибраних частин зображення додатково застосовується різні фотометричні спотворення, наприклад, зміна яскравості та контрастності, а також відображення в одному з напрямків.

2.2.4 Спосіб виділення емпіричного рецептивного поля шару згорткової нейронної мережі

Цей спосіб відрізняється від розглянутих раніше адаптивним вибором порога включення частин вихідного зображення в рецептивне поле, а також скороченням кількості зашумлених зображень, необхідних для навчання CHP.

Завдяки застосуванню даного способу до кожного шару СНР стає можливою діагностика роботи навченої СНР: визначення розмірів частин зображень, що детек- туються. Якщо після застосування даного способу встановлено, що розміри, що детектуються, занадто малі, то слід додати додаткові шари або якщо розміри великі, то виключити шари.

У даній роботі отримані розміри емпіричного рецептивного поля для кожного шару використовуються для визначення розмірів об'єктів, що детектуються на зображенні на даному шарі, що дозволяє виділити межі зміни розмірів «якорних» прямокутників на даному шарі.

Запропонований спосіб складається з наступних етапів.

Етап 1. Вибір n зображень із навчальної вибірки зображень, що мають найбільші значення активації будь-якого з нейронів заданого шару СНР.

Цей етап призначений для вибору таких зображень, які спричиняють найбільші значення активації нейронів заданого шару СНР. Цей етап необхідний, оскільки застосування наступних етапів, описаних в даному способі обчислювально складно для всіх зображень навчальної бази, необхідно вибрати меншу кількість зображень.

Етап 2. Побудова карти диспропорції для кожного з n зображень. Цей етап призначений побудови карти невідповідностей, тобто. та-

ної карти, кожному пікселю якої відповідає значення активації нейрона даного шару, отриманого після зашумлення даного пікселя та пікселів деякої його околиці. Вибирається той нейрон, який мав найбільше значення активації даного зображення цьому шарі.

Даний етап дозволяє визначити вплив кожного пікселя на нейрон заданого шару. Після побудови карти невідповідності кожен піксель зображення буде відповідати значенню активації вибраного нейрона.

Етап 3. Включення пікселів зображення до емпіричного рецептивного поля нейрона СНС та обчислення розміру цього поля для кожного з n зображень.

Даний етап призначений для виділення тих пікселів, які мають найбільший вплив на нейрон заданого шару та подальше включення їх в емпіричне

рецептивне поле даного нейрона. Після виконання етапу формується емпіричне рецептивне поле нейрона CHP, а також його розмір, отриманий на підставі кількості пікселів, включених до цього поля.

Етап 4. Вибір мінімального та максимального розміру емпіричного рецептивного поля шару CHP з n зображень.

Цей етап призначений для визначення розміру емпіричного рецептивного поля шару, яке обчислюється на основі отриманих максимальних значень активацій нейронів n зображень. Виконання даного етапу дозволяє визначити розміри частин зображення, що виділяються на цьому шарі.

Порівняно з роботою, розглянутою раніше, запропонований спосіб відрізняється виконанням етапів 2 і 3. Етап 2 відрізняється від існуючого скороченням необхідної кількості зашумлених зображень при навчанні згорткової нейронної мережі. Етап 3 визначає процедуру включення пікселів вихідного зображення в рецептивне поле, що дозволяє підібрати поріг такого включення адаптивно.

Етап вибору n зображень із навчальної вибірки зображень, що мають найбільші значення активації будь-якого з нейронів заданого шару CHP

На даному етапі для кожного зображення вибирається нейрон із максимальним значенням активації на вибраному шарі та зберігається. Потім отримані збережені значення сортуються в порядку зменшення та вибираються n зображень, які мають найбільше значення активації нейрона (кожне зображення може мати свій нейрон з максимальною активацією).

Етап побудови карти невідповідностей для кожного з n зображень

Якщо час навчання не є критичним, то карту невідповідності можна будувати наступним чином: з використанням ковзного вікна на зображення накладати фільтр шуму w×h із заповненням RGB(220, 220, 220) та кроком s (ковзне вікно) та фіксувати значення активації вилайкового нейрона noise i , j для кожного пікселя (i, j) зображення.

Однак якщо час навчання критичний, то перед застосуванням ковзного вікна до вихідного зображення можна застосувати це вікно не до всього

зображення, а лише до його частини. Процедуру визначення такої частини зображення можна проводити так:

Крок 1. На вхід нейронної мережі подається вихідне зображення та вирахується активація вибраного нейрона

Крок 2. Вихідне зображення поділено вздовж кожної осі на дві частини, як показано на рис. 2.5 (всього 4 області).

Крок 3. На кожну область накладається фільтр шуму із заповненням RGB(220, 220, 220), що збігається із заданою областю. Потім чотири отримані зображення подаються на вхід нейронної мережі та обчислюється активація вибраного нейрона

Крок 4. Вихідна область включається до результуючої області

Крок 5. Кроки 1 – 4 повторюються з тією різницею, що вихідне зображення вздовж кожної осі ділиться не на дві області, а на чотири (виходить 16 областей). При цьому якщо якась отримана область не потрапила в результуючу область кроку 4, то дана область не розглядається.

Крок 6. Аналогічно крокам 1 – 5 вихідне зображення поділяється на 8 та 16 областей вздовж кожної осі.

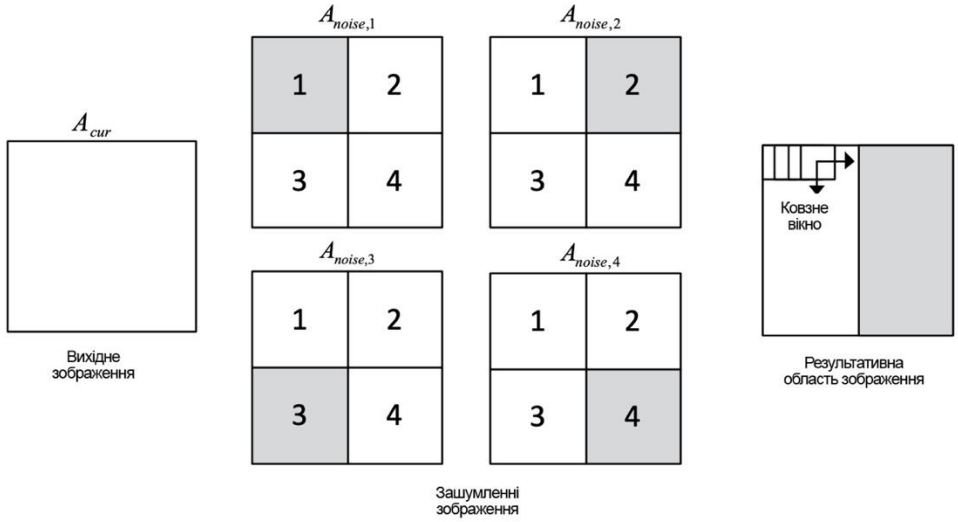

Рисунок 2.3 – Поділ вихідного зображення на 4 області та застосування до результуючої області ковзного вікна (біла частина зображення справа входить у результуючу область, сіра не включається)

Запропонований алгоритм включення пікселів для кожного з n зображень у рецептивне поле на основі побудованої карти складається з наступних кроків і показаний на рис. 2.4:

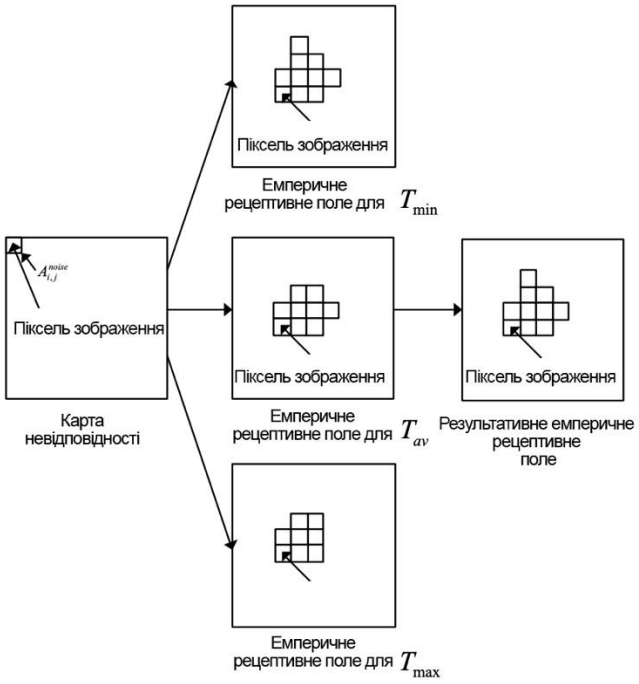

Рисунок 2.4 – Отримання з карти невідповідностей результуючого емпіричного рецептивного поля (показано одну останню ітерацію)

А після скорочення кількості зашумлених зображень, з використанням запропонованого алгоритму для області зображення, що залишилася, накладається фільтр шуму w×h з заповненням RGB(220, 220, 220) і кроком s і фіксується значення активації обраного нейрона кожного пікселя (i , j) зображення.

При цьому якщо хоча б один піксель зображення входить в область, що розглядається , а решта ні, то на аналізоване зображення все одно накладається фільтр.

Після побудови карти невідповідності кожен піксель (i , j) зображення

А.ження буде відповідати значення активації вибраного нейрона noise i , j

Етап включення пікселів зображення в емпіричне рецептивне поле нейрона CHP та обчислення розміру цього поля для кожного з n зображень

Вибір мінімального та максимального розміру емпіричного рецептивного поля шару CHP з n зображень

Вибирається мінімальний та максимальний розмір рецептивного поля шару CHP серед розмірів полів n зображень, отриманих на попередньому етапі.

2.2.5 Спосіб обчислення розмірів «якорних» прямокутників для багатомасштабної моделі детектування візуальнихоб'єктів

Даний спосіб дозволяє з використанням методу кластеризації k - середніх обчислити розміри «якорних» прямокутників для конкретної бази на кожному шарі блоку детектора багатомасштабної моделі детектування об'єктів.

2.3 Метод аналізу динамічних зорових сцен

2.3.1 Опис методу

Розроблений метод відрізняється від існуючих використанням запропонованого різновиду CHP, процедурою відновлення пропусків (передбачення) детектування, використанням результату передбачення, у разі, якщо показання детектора відсутні, роботою в умовах невизначеності або зашумленості даних детектування.

Даний метод дозволяє на кожному кадрі виділити об'єкти з використанням детектора та зіставити їх із наявними треками (послідовності координат об'єктів на попередніх кадрах) або видалити трек, якщо на послідовності кадрів об'єкт, пов'язаний із треком відсутній.

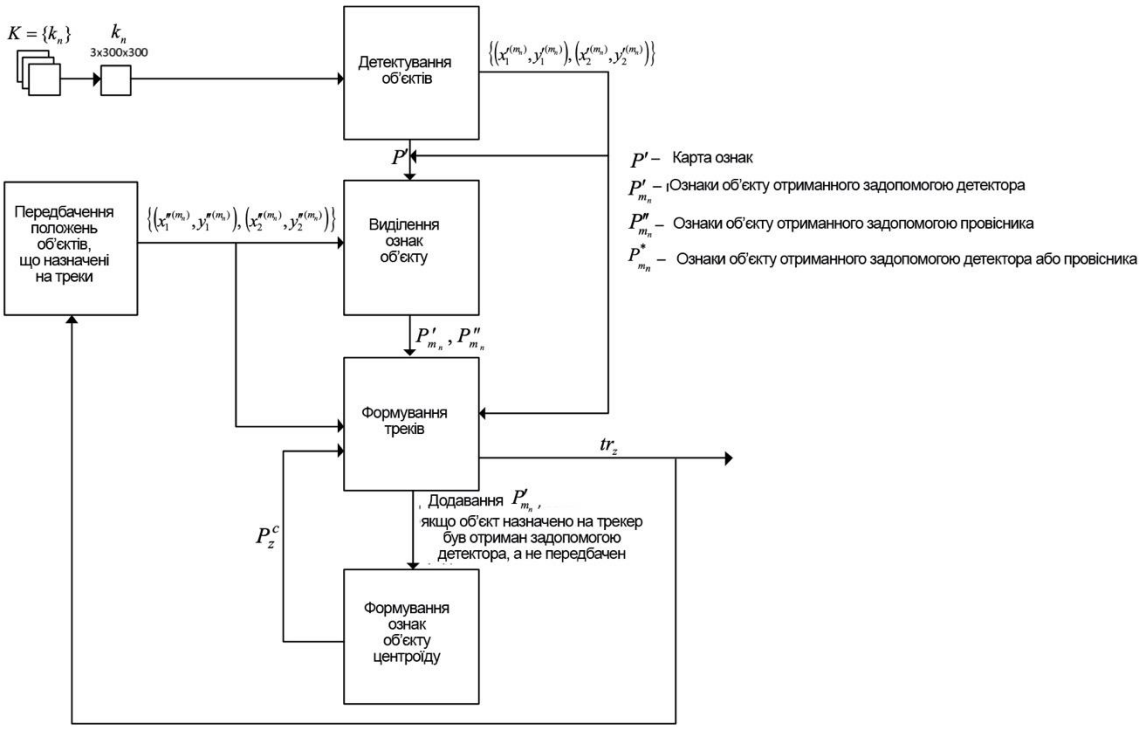

Рисунок 2.5 – Структура методу аналізу динамічних зорових сцен

На рис. 2.5 показано структуру розробленого методу. Детектування та виділення ознак об'єктів виконується з викованням запропонованої структури CHP на основі моделі візуального представлення об'єкта. Пророцтво положень об'єктів, призначених на треки, виконується на основі розробленої моделі руху об'єктів. Формування ознак об'єкта-центроїду для кожного треку та формування треків виконується на відповідних етапах розробленого методу.

2.3.2 Модель руху об'єкту

Моделі руху досліджують динамічну поведінку об'єкта і на основі передісторії дозволяють передбачити, в якому положенні простору буде цей об'єкт в наступні моменти часу.

У даній роботі використовується лінійна модель руху об'єкта з постійною швидкістю (у разі застосування моделі іншого типу не потрібно значно змінювати розроблений метод). Для такої моделі достатньо знати положення об'єкта в просторі тільки в двох попередніх точках (моментах часу), однак для нівелювання похибок показань детектора слід використовувати передісторію руху на основі

більшого числа точок, наприклад 5–10 і виконувати апроксимацію на основі методу - Менших квадратів (МНК).

Рисунок 2.6 – Позначення положення об'єкта у просторі

2.3.3 Спосіб фільтрації детекцій об'єктів

Як було зазначено раніше, під час роботи з існуючими базами для тестування методів аналізу динамічних зорових сцен, наприклад 2D MOT15, доводиться працювати з великою кількістю помилкових детекцій об'єктів і розробляти алгоритми, дозволяють виділяти дані з шуму. Приклад великої кількості хибних виявлення об’єктів, взятих із бази даних, наведено на рисунку 2.9. Крім пішоходів, детектор виділяє також їх частини, інші об'єкти та частини інших об'єктів, які не є пішоходами.

Вхідні дані : вхідне зображення, набір детекцій об'єктів

O = { o n } = [n , x 1 , y 1 , x 2 , y 2 , P], де 1 n , = ...,N – кількість детекцій об'єктів, P –можливість правильного детектування,P min – мінімальна ймовірність правильного детектування, ціями об'єктів.

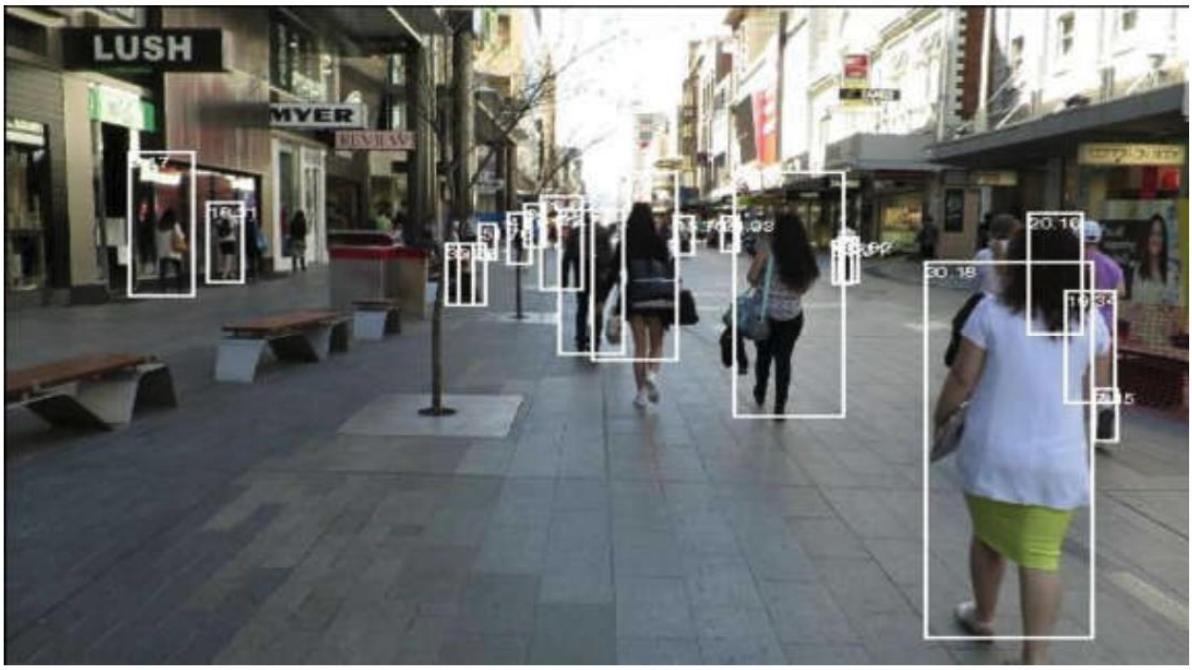

Рисунок 2.9 – Детекції об'єктів із бази 2D MOT15

2.4 Висновки

Розроблено новий тип згорткової нейронної мережі, який відрізняється тим, що дозволяє виконувати аналіз статичних та динамічних зорових сцен за один прохід CHP. Це досягається за рахунок введення шару виділення «глибоких» ознак знайдених та передбачених детекцій об'єктів, а також використання модифікації моделі візуального представлення об'єктів для їх опису.

Запропоновано метод аналізу статичних зорових сцен на основі розробленого нового типу згорткових нейронних мереж. Даний метод у порівнянні з відомими, дозволяє досягти більш високої точності та оперативності детектування об'єктів за рахунок виділення «глибоких» ознак детекцій цих об'єктів. Це досягається завдяки наступним нововведенням: адаптивному вибору порога включення частин вихідного зображення в рецептивне поле, скорочення

кількості зашумлених зображень, необхідних для навчання СНР, виділення меж зміни розмірів «якорних» прямокутників та обчислення їх розмірів для конкретної навчальної вибірки.

Запропоновано метод аналізу динамічних зорових сцен на основі розробленого нового типу згорткових нейронних мереж, який, порівняно з відомими, дозволяє досягти більш високої точності побудови треків, здійснювати обробку в режимі реального часу, виконувати аналіз як в умовах невизначеності, так і зашумленості. даних детектування. Це досягається за рахунок використання модифікації моделей візуального представлення та руху об'єктів, обчислення центроїдів ознак об'єкта для кожного треку, оригінальних способів відновлення пропусків об'єктів детектором та фільтрації детекцій об'єктів.

Для реалізації методів аналізу статичних та динамічних зорових сцен розроблено:

спосіб виділення емпіричного рецептивного поля СНР, що відрізняється від існуючого адаптивним вибором порога включення частин вихідного зображення в рецептивне поле, а також скороченням кількості зашумлених зображень, необхідних для навчання СНР.

спосіб обчислення «якорних» прямокутників для багатомасштабної моделі детектування візуальних об'єктів, що дозволяє з використанням методу кластеризації k- середніх обчислити розміри «якорних» прямокутників для конкретної навчальної бази.

модель руху об'єкта дозволяє передбачити в якому положенні простору буде об'єкт у наступні моменти часу;

метод відновлення відсутніх об'єктів детектором дозволяє відновити виявлення об'єктів, які були пропущені детектором;

метод фільтрації виявлення об'єктів, що забезпечує можливість спільної діяльності суб'єкта в детектуванні шумових даних;

спосіб виділення «глибоких» ознак детекції об'єкта, що дозволяє виділити набір ознак детекції об'єкта, за допомогою якого кілька детекцій різних об'єктів можуть порівнюватися між собою.

3 РОЗРОБКА АЛГОРИТМІВ І ПРОГРАМНИХ ЗАСОБІВ ДЛЯ РЕАЛІЗАЦІЇ МЕТОДІВ АНАЛІЗУ СТАТИЧНИХ І ДИНАМІЧНИХ ЗОРОВИХ СЦЕН НА ОСНОВІ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ

3.1 Алгоритми для реалізації методів аналізу статичних та динамічних зорових сцен на основі CHP

У цьому розділі розглянуто розроблені алгоритми: аналізу статичної зорової сцени; обчислення розмірів "якорних" прямокутників для багатомасштабної моделі детектування візуальних об'єктів; виділення емпіричного рецептивного поля для кожного шару CHP. Крім того, розглянуто розроблений алгоритм аналізу динамічної зорової сцени.

Більшість процедур, що використовуються в даних алгоритмах, описані у 2-му розділі, тому тут будуть вказані назви цих процедур, а основна увага буде приділена питанням отримання вхідних та вихідних даних для виконання цих алгоритмів.

3.1.1 Алгоритм обчислення розмірів «якорних» прямокутників для багатомасштабної моделі детектування візуальнихоб'єктів

На рис. 3.1 показано схему даного алгоритму. Користувач вводить шлях до файлу TXT, в якому зберігається список усіх шляхів до файлів XML навчальної бази. Кожному зображенню бази відповідає один XML файл. У кожному такому файлі зберігається набір координат прямокутників, що обрамляють, знайдених детекцій об'єктів для конкретного зображення. Такий формат є стандартним для навчальної бази VOC 2007.

Крім того, користувач вводить початкові параметри: кількість блоків детектора та розміри мінімального та максимального розміру рецептивного поля для кожного шару перед блоком детектора.

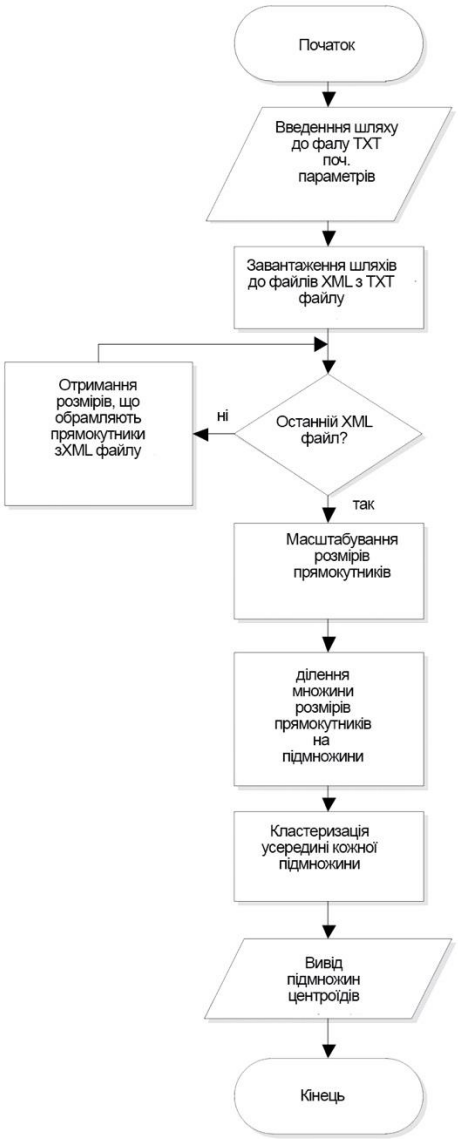

Рисунок 3.1 – Схема алгоритму обчислення розмірів «якорних» прямокутників для багатомасштабної моделі детектування візуальних об'єктів

Після того, як всі XML-файли будуть прочитані, проводяться операції, описані в п. 2.2.4: масштабування розмірів прямокутників, що обрамляють , розподіл безлічі розмірів прямокутників на підмножини, кластеризація всередині кожного підмножини. В результаті виконання даного алгоритму буде сформовано набір підмножин, що складаються з центроїдів.

Кожні такі медіани і будуть бути «якір» прямокутником для різномасштабних моделей виявлення візуальних об'єктів.

3.1.2 Алгоритм виділення емпіричного рецептивного поля для кожного шару CHP

На рис. 3.2 показано схему даного алгоритму. Користувач вводить шлях до файлу TXT, в якому зберігається список усіх шляхів до файлів зображень JPG навчальної бази.

Крім того, користувач вводить початкові параметри: розмір фільтра шуму, крок з яким накладається фільтр шуму, кількість (n) зображень з максимальною активацією нейрона, список шарів для виділення емпіричного рецептивного поля (l).

Потім завантажується модель CHP, обчислюються значення максимальної активації нейрона шару CHP для кожного зображення з навчальної бази, серед яких потім вибираються n зображень для кожного шару з максимальною активацією нейрона заданого шару. Після чого для кожного з N зображень проводяться операції, описані у п. 2.2.3: скорочення кількості зашумлених зображень, побудова карти невідповідностей, обчислення емпіричного рецептивного поля.

Після чого для кожного шару виводиться мінімальний та максимальний розмір рецептивного поля CHP серед усіх розмірів полів n зображень.

3.1.3 Алгоритм аналізу динамічної зорової сцени

Перед початком аналізу динамічної зорової сцени, як було зазначено у розділі 2, важливо визначити умови роботи детектора: невизначеність або зашумленість даних детектування. У разі невизначеності детектування необхідно відновлювати детекції об'єкта для кожного треку, де вона відсутня. При зашумленості даних такої необхідності немає, але додатково застосовується фільтрація детекцій об'єктів з використанням способу, запропонованого в розділі 2. Оскільки інших відмінностей немає, то розглянемо алгоритм аналізу динамічної зорової сцени тільки в умовах невизначеності детектування. На рисунку 3.3 показана схема даного алгоритму.

Користувач вводить початкові параметри: пороги та початкові значення, зазначені в розділі 2, джерело введення даних (веб-камера або набір зображень із

заданої папки), джерело виведення даних (на екран, текстовий файл, набір графічних файлів).

Рисунок 3.3 – Схема алгоритму аналізу динамічної зорової сцени в умовах невизначеності детектування

Потім завантажується модель CHP і послідовно відбувається завантаження кожного кадру. Далі відбувається детектування об'єктів, одержання їх ознак, а також ознак передбачених об'єктів з використанням способу відновлення

перепусток об'єктів детектором. Дані детекції об'єктів та ознаки виходять з використанням моделі завантаженої СНР. Потім відбувається призначення детекцій об'єктів на треки з використанням угорського алгоритму. Детекції об'єктів, які вдалося призначити на треки, ініціалізують нові треки. Далі для кожного треку передбачається положення об'єкта на наступному кадрі (це необхідно для формування історії для способу відновлення пропусків об'єктів детектором). Якщо детекція об'єкта для якогось треку відсутня, то виконується її відновлення (у разі можливості) з використанням способу відновлення пропусків об'єктів детектором. Крім того, для кожної призначеної або передбаченої детекції об'єкта додатково перевіряється допустимість призначення: порівнюється ступінь відмінності об'єкта-центроїду треку та даної детекції об'єкта із заданим порогом. Далі відбувається видалення тих треків, у яких кількість таких кадрів, на яких рівень відмінності між об'єктом-центроїдом треку та призначеним об'єктом перевищує допустимий поріг занадто великий (більше порога). Після виконання цього алгоритму відбувається виведення набору треків у вказаний користувачем засіб виведення.

3.2 Розробка бібліотеки програмних функцій, що реалізують методи аналізу статичних та динамічних зорових сцен

3.2.1 Структура програмних засобів, що реалізують методи аналізу статичних та динамічних зорових сцен

Оскільки при реалізації навчання та тестування СНР потрібні великі обчислювальні потужності, досить важливо реалізувати ефективну взаємодію з ресурсами комп'ютера: пам'яттю, процесором, графічним процесором. Даного результату можна досягти, реалізувавши процедури навчання та тестування з використанням мови C++, оскільки він дозволяє ефективно поводитися з пам'яттю безпосередньо, використовуючи покажчики.

Для взаємодії модулів, написаних мовою C++, зручно використовувати мову Python, оскільки вона дозволяє прискорити розробку програмних засобів

завдяки наявності великої кількості бібліотек та широких можливостей роботи з різними типами даних.

Структура програмних засобів, що реалізують методи аналізу статичних та динамічних зорових сцен, представлена на рис. 3.4.

Підсистема налаштування СНР складається з: модуля кластеризації розмірів прямокутників, що обрамляють, і модуля виділення емпіричного рецептивного поля.

Підсистема аналізу статичних та динамічних зорових сцен складається з: модуля видачі результатів, модуля взаємодії з базами, модуля взаємодії з Caffe, модуля передбачення, модуля обчислення метрик, модуля фільтрації даних, модуля формування треків.

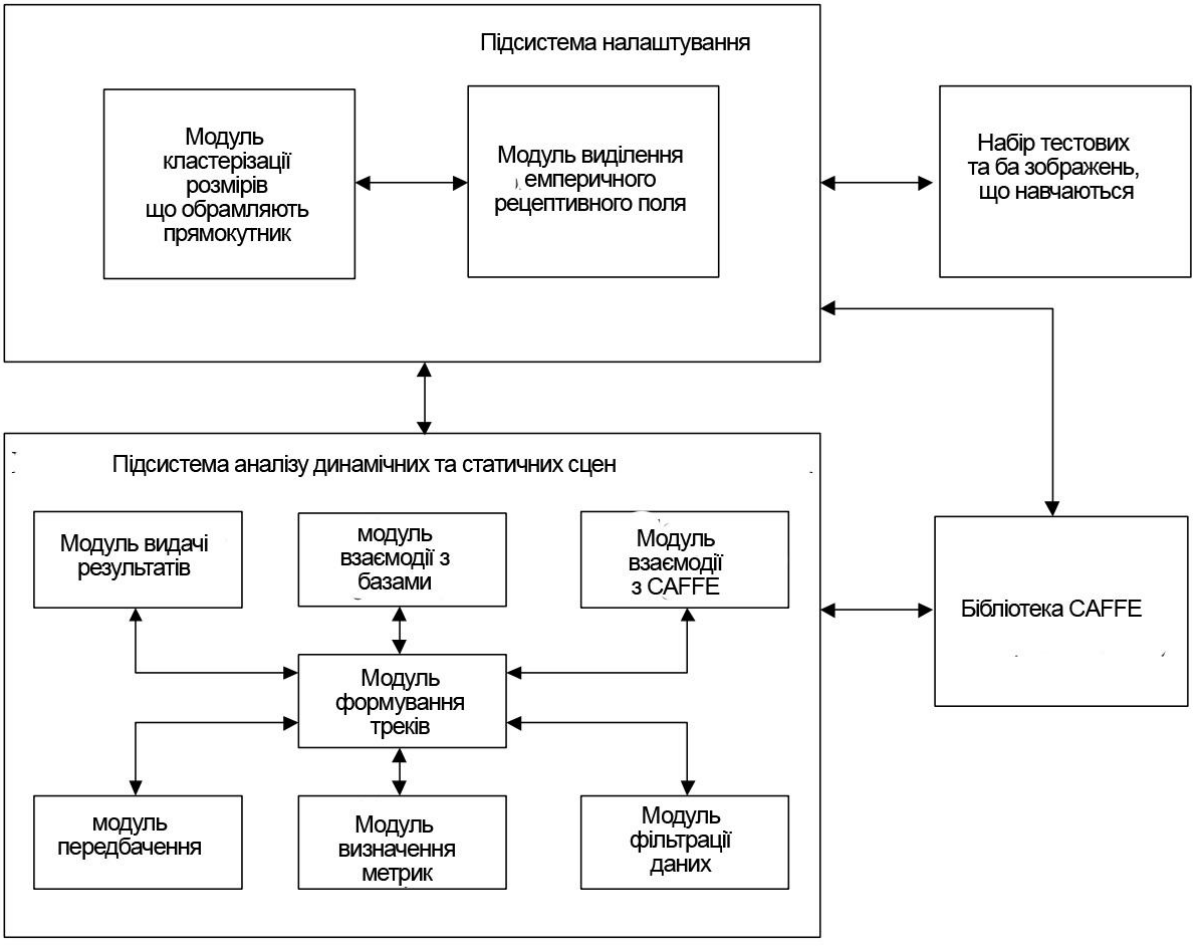

Рисунок 3.4 – Структура програмних засобів, що реалізують методи аналізу статичних та динамічних зорових сцен

Модуль видачі результатів дозволяє вивести знайдені та супроводжені на кадрі об'єкти (текстовий файл, зображення). Модуль взаємодії з базами дозволяє зчитувати детекції об'єктів з текстового файлу бази. Модуль взаємодії з Caffe дозволяє працювати з моделлю на основі СНР, побудованої в Caffe. Модуль передбачення реалізує передбачення положення об'єкта на кадрах.

Модуль фільтрації даних дозволяє виконати фільтрацію отриманих детекцій об'єктів. Модуль формування треків реалізує процедури створення треків, асоціації треків з детекціями об'єктів, видалення треків.

Набір тестових та навчальних баз зображень містить дані, необхідні для навчання та тестування СНР та розроблених методів. Можливості та принципи роботи бібліотеки Caffe будуть розглянуті у цьому розділі

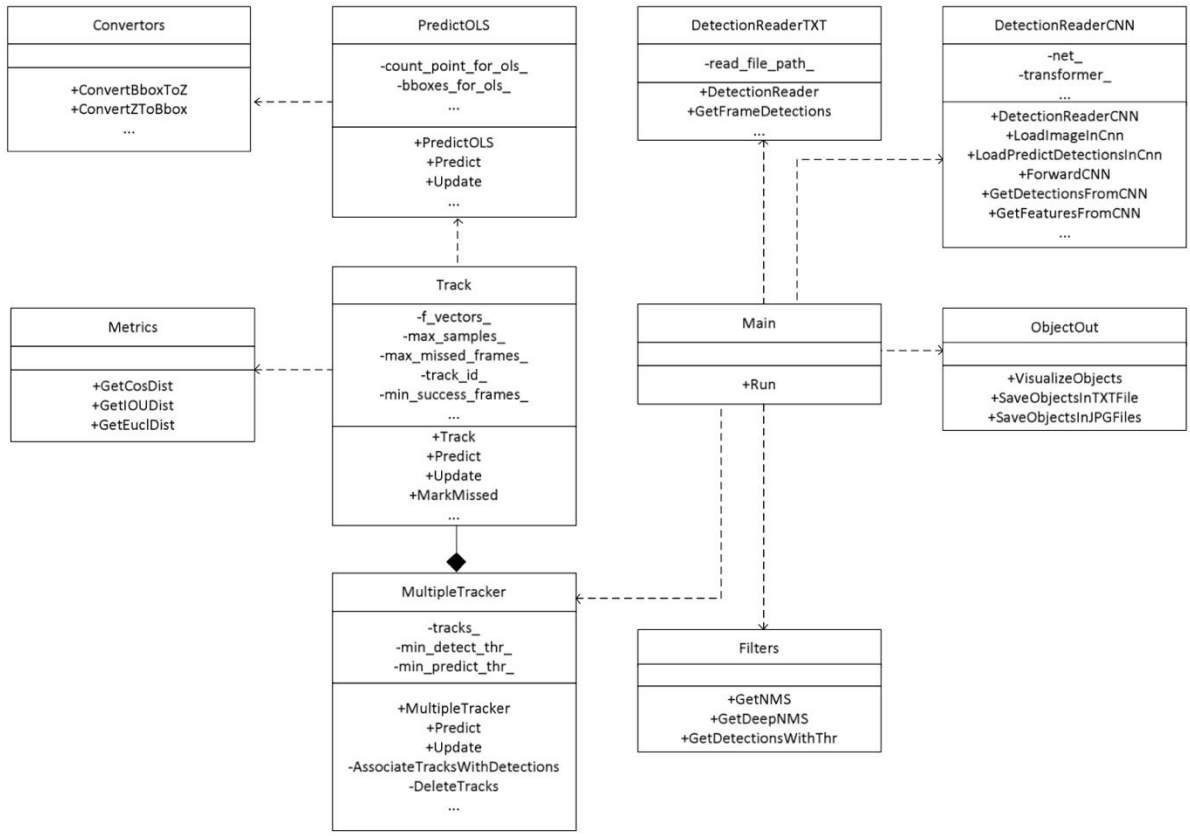

Рисунок 3.5 – Діаграма класів, що реалізує бібліотеку програмних функцій

3.2.2 Бібліотека програмних функцій, що реалізують налаштування СНР для аналізу статичних та динамічних зорових сцен

Бібліотека програмних функцій реалізована у вигляді набору модулів.

Опишемо розроблені модулі, їх призначення та виконувані функції.

3.2.3 Бібліотека програмних функцій, що реалізують аналіз статичних та динамічних зорових сцен за допомогою навченої СНР

Бібліотека функцій програми реалізована у вигляді набору класів. Діаграма класів, що реалізує бібліотеку програмних функцій, показано рис. 3.5.

3.3 Робота зі СНР з використанням бібліотеки Caffe

3.3.1 Опис можливостей та принципів роботи Caffe

Caffe – це бібліотека з відкритим вихідним кодом для навчання та тестування глибоких нейронних мереж.

Ця бібліотека має такі переваги:

- моделі та опис обчислень, що проводяться з цими моделями, визначаються у вигляді файлів конфігурації замість вихідного коду;
- висока швидкість обробки даних;
- має хорошу розширюваність, так як виконана у вигляді окремих модулів;
- відкритість бібліотеки;
- наявність спільнот для спільного обговорення проектів;
- дозволяє реалізовувати шари як із використанням gpu так і cpu;
- наявність готових моделей навчання, що розповсюджуються .

Реалізація бібліотеки виконана з використанням мови програмування C++, крім того, для зручності використання є надбудови на Python та MATLAB. Для прискорення обчислень Caffe може бути запущена на

GPU із використанням базових можливостей технології CUDA.

Топологія нейронної мережі в caffe представляється як набору шарів (тренувальних і тестових) і зберігається у файлі з розширенням .prototxt. Кожен наприклад шар має наступні параметри: ім'я (назва), тип (тип), вхідні дані шару (внизу), вихідні дані (вгорі), додаткові параметри (які є індивідуальними для кожного конкретного шару, наприклад convolu- tion_param) .

Параметри навчання нейронної мережі задаються в іншому файлі з розширенням .prototxt. До таких параметрів відносяться: шлях до файлу з топологією мережі (net), параметри градієнтного спуску (base_lr, weight_decay та інші), максимальна кількість ітерацій (max_iter), архітектура, на якій будуть проводитися обчислення (solver_mode), шлях для збереження навченої мережі (snapshot_prefix).

Крім того, крім стохастичного градієнтного спуску підтримуються й інші методи навчання, наприклад алгоритм з адаптивною швидкістю навчання або прискорений градієнтний спуск Нестерова та ін.

Caffe оперує блобами – це N- мірні масиви (як правило 4-мірні N × K × H × W). Наприклад, якщо подати на вхід мережі масив зображень 10×3×227×227, то N – кількість зображень, поданих на вхід (10), K – кількість каналів (3) , H – висота (227) , W – ширина (227).

На вхід мережі подається вхідний двійковий об'єкт, Caffe його рукоятки з довідковому шаром, внутрішня реалізація яких описана в окремому модулі на мові C ++, в результаті якого виходить вихідний двійковий об'єкт.

Серіалізація (передача) структурованих даних файлу конфігурації мережі (.prototxt) до класів C++ виконується за допомогою технології Google Protobuf. Такі структури описуються у файлі caffe.proto.

Наприклад, можна описати у файлі caffe.proto структуру RP-серіалізованих даних з параметром pooled_h:

```
message RP {
optional uint32 pooled_h = 1 [default = 0];
}
```

Після цього дана структура з параметром може бути скомпільована клас C++ і з'являється можливість її використання.

Кожен шар в Caffe на мові C ++ описаний в вигляді окремого класу, який включає в себе наступні методи: LayerSetUp (настройка параметрів шару, перевірка введених параметрів), Reshape (зміна розмірів оброблюваного блоба), Forward_cpu (описує, яким чином змінюється блоб при прямому проході мережі),

Backward_cpu (описує, яким чином змінюється блоб при зворотному проході мережі) (для CPU реалізації) або Forward_gpu, Backward_gpu (для GPU реалізації).

Таким чином, Caffe є досить зручною бібліотекою для роботи зі СНР: тензори СНР представляються у вигляді блобів, а структури згорткових шарів зручно описувати у файлі конфігурації. Крім того, при необхідності додавання свого шару його можна реалізувати в окремому модулі на C++, після чого додати файл конфігурації.

3.3.2 Опис структури шару ROI-Pooling з використанням Caffe

У файлі з розширенням .prototxt описується розмірність вхідного блоку для подачі на вхід шару ROI-Pooling передбачених положень об'єктів так:

input: "rois_predict" input_shape {

dim: 10 #count_rois dim: 4 # [x1, y1, x2, y2]

}

У цьому прикладі описано, що на вхід буде подано 10 передбачених положень об'єктів. Кількість передбачених положень об'єктів змінюється в процесі роботи на кожному кадрі (Reshape) завдяки можливості доступу до структури за допомогою обгорток, наприклад, з використанням Python.

Опис в шарі структури знаходиться в наступному:

layer {

name: "roi_pool" type: "ROIPooling" bottom: "conv4_3"

bottom: "detection_out" bottom: "rois_predict" top: "roi_pool"

top: "roi_pool_predict" roi_pooling_param { pooled_w: 7

pooled_h: 7

spatial_scale: 38 # 38/300

}

}

В даному випадку на вхід шару надходить: блоб з шару conv4_3, детекції СНР detection_out, а також передбачені положення об'єктів. На виході формуються ознаки детекцій roi_pool та ознаки передбачених детекцій об'єктів

roi_pool_predict. Також описуються параметри шару: розмірність шару (pooled_w, pooled_h), коефіцієнт масштабування (spatial_scale).

Опис структури параметрів шару у файлі caffe.proto:

message ROIPoolingParameter {optional uint32 pooled_h = 1 [default = 0]; optional uint32 pooled_w = 2 [default = 0]; optional float spatial_scale = 3 [default = 1];

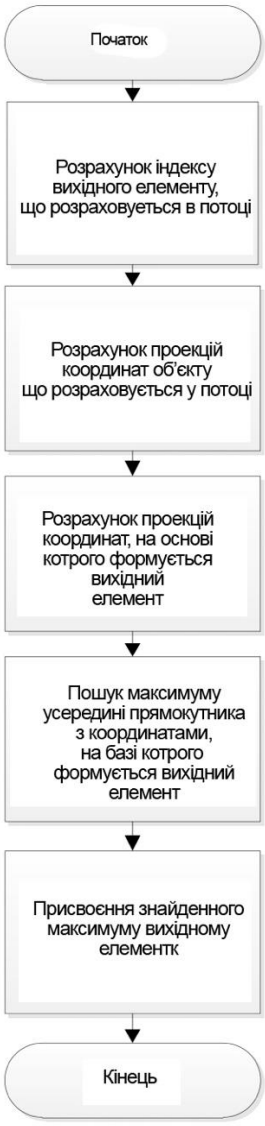

Рисунок 3.7 - Схема алгоритму, що запускається в потоці

Програмна реалізація роботи шару є перетворенням вхідного блоба у вихідний і описується з використанням мови C++ окремо для архітектури CPU та GPU.

Спосіб знаходження глибоких ознак детекції об'єкта вже був описаний у розділі 2. Тут зупинимося на деяких обчислювальних особливостях. По- перше, індекс вихідного двійкового об'єкта елемента 1 або 2 (один з елементів ROI-Акумулювання особливість карти) є обчисленої , який буде обчисленої в потоці. Потім обчислюються проекції координат об'єкта на карту ознак (індекси у вхідному блобі), який обчислюватиметься в потоці (номер об'єкта однозначно визначається на основі індексу вихідного елемента). Далі всередині прямокутника, що обрамляє об'єкт на карті ознак, обчислюються координати, на основі яких формується вихідний елемент (тобто обчислюються координати нового прямокутника всередині даного). Потім усередині отриманого прямокутника шукається елемент із максимальною активацією, який присвоюється вихідному елементу блоку.

3.4 Висновки

Розроблено алгоритми, що реалізують метод аналізу статичних зорових сцен: алгоритм обчислення розмірів «якорних» прямокутників для багатомасштабної моделі детектування візуальних об'єктів, алгоритм виділення емпіричного рецептивного поля для кожного шару CHP.

Розроблено алгоритми, що реалізують метод аналізу динамічних зорових сцен: алгоритм відновлення перепусток об'єктів детектором, алгоритм фільтрації детекцій об'єктів, алгоритм виділення «глибоких» ознак детекції об'єкта.

Виконано програмну реалізацію розроблених алгоритмів у вигляді бібліотеки програмних функцій. Розроблено бібліотеку програмних функцій, що реалізують налаштування CHP для аналізу статичних і динамічних зорових сцен, що включає два модулі: модуль кластеризації розмірів обрамляючих прямокутників навчальної вибірки CHP і модуль виділення емпіричного рецептивного поля шару CHP. Також розроблено бібліотеку програмних функцій, що реалізує аналіз статичних та динамічних зорових сцен за допомогою навченої CHP, розглянуто діаграму класів даної бібліотеки, описано основні функції.

Розглянуто бібліотеку Caffe для роботи зі CHP, описано її основні можливості та принципи роботи. Було виявлено, що Caffe є досить зручною бібліотекою для роботи зі CHP: тензори CHP представляються у вигляді блобів, а структури згорткових шарів зручно описувати у файлі конфігурації. Крім того, при необхідності додавання свого шару його можна реалізувати в окремому модулі C++, після чого додати у файл конфігурації.

Розглянуто опис структури шару ROI-Pooling із використанням

Caffe та виконана його програмна реалізація для CPU та GPU.

ВИСНОВОК

В результаті дослідження вирішено наукове завдання, що полягає у розробці методів та алгоритмів аналізу статичних та динамічних зорових сцен на основі згорткових нейронних мереж.

Виконано дослідження методів аналізу статичних зорових сцен. Проведено аналіз методів виділення, розпізнавання, а також детекти- рування об'єктів статичних зорових сцен. Висвітлено переваги та недоліки методів і встановлено, що на сьогоднішній день найбільш перспективними є методи, які базуються на CHP.

Виконано дослідження методів аналізу динамічних зорових сцен. Подано класифікацію методів супроводу безлічі об'єктів за ознаками та компонентами методу супроводу.

Розроблено новий тип згортки нейронної мережі, що дозволяє виконуватися nyat аналізу статичних і динамічних візуальних сцен для одного проходу CHC. Це досягається за рахунок введення шару виділення «глибоких» ознак знайдених та передбачених детекцій об'єктів, а також використання модифікації моделі візуального представлення об'єктів для їх опису.

Запропоновано та обґрунтовано метод аналізу статичних зорових сцен, що дозволяє досягти більш високої точності та оперативності детектування об'єктів за рахунок виділення «глибоких» ознак їх детекцій. Це досягається завдяки наступним нововведенням: адаптивному вибору порога включення частин вихідного зображення в рецептивне поле, скорочення кількості зашумлених зображень, необхідних для навчання CHP, виділення меж зміни розмірів «якорних» прямокутників та обчислення їх розмірів для конкретної навчальної вибірки.

Ми розробили метод аналізу динамічних візуальних сцен, забезпечується vayuschy з високою точністю і будівництво доріжок, обробки і аналізу даних в режимі реального часу і в умовах невизначеності, а й при даних детектування. Це досягається за рахунок використання модифікації моделей візуального

представлення та руху об'єктів, обчислення центроїдів ознак об'єкта для кожного треку, оригінальних способів відновлення пропусків об'єктів детектором та фільтрації детекцій об'єктів.

Розроблено алгоритми, що реалізують методи аналізу статичних та динамічних зорових сцен: обчислення розмірів «якорних» прямокутників для багатомасштабної моделі детектування візуальних об'єктів, виділення емпіричного рецептивного поля для кожного шару CHP, виділення "глибоких" ознак детекції об'єкта, відновлення перепусток детектора на основі МНК, фільтрації даних детектора. Також розроблено програмні засоби, що реалізують зазначені алгоритми.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1. Аведьян Э.Д., Галушкин А.И., Селиванов С.А. Сравнительный анализ структур полносвязных и сверточных нейронных сетей и их алгоритмов обучения // Информатизация и связь. – 2017. – № 1.

2. Антощук С.Г. Отслеживание объектов интереса при построении автома- тизированных систем видеонаблюдения за людьми // Электротехнические и компьютерные системы. – 2012. – №8(84). – С. 151–156.

3. Программа для анализа динамических зрительных сцен в режиме реаль- ного времени / В.В. Борисов, О.И. Гаранин // Свидетельство о государственной регистрации программы для ЭВМ № 2018126210 от 24.05.2018.

4. Борисов В.В., Синявский Ю.В., Гаранин О.И., Коршунова К.П. Сверточ-

ная нейро-нечеткая сеть для исследования гидромеханических процессов в условиях неопределенности // Сборник тезисов докладов XVI Всероссийской научной конференции«Нейрокомпьютеры и их применение» 13 марта 2018 года. Москва. – С. 140-141.

5. Буй Тхи Тху Чанг, Фан Нгок Хоанг, Спицын В.Г. Распознавание лиц на основе применения метода Виолы–Джонса, вейвлет-преобразования и метода главных компонент // Известия Томского политехнического университета. – 2012. – № 5. – С.54–59.

6. Беляев Е.А., Тюрликов А.М. Алгоритмы оценки движения в задачах сжа- тия на низких битовых скоростях // Компьютерная оптика. – 2008. – №4.

7. Вапник В.Н. Восстановление зависимостей по эмпирических данным. –

М.: Наука, 1979.

8. Графический редактор макромодели транспортной сети с возможностью нахождения максимального потока/ О.И. Гаранин // Свидетельство о го- сударственной регистрации программы для ЭВМ № 2014660475 от 08.10.2014.

9. Гаранин О.И. Модель регулятора транспортных потоков // Сборник тру-

дов 10-ой международной научно-технической конференции студентов и аспирантов «Информационные технологии, энергетика и экономика» 17-18 апреля 2013 года. Филиал МЭИ в г. Смоленске. Смоленск. – 2013.

10. Гаранин О.И. Классификация моделей транспортных средств // Сборник тру- дов 11-ой международной научно-технической конференции студентов и ас- пирантов «Информационные технологии, энергетика и экономика» 17-18 ап- реля 2014 года. Филиал МЭИ в г. Смоленске. Смоленск. 2014. – С. 223-226.

11. Гаранин О.И. Способ моделирования транспортной сети на основе мно- гомодельного подхода // Сборник трудов 11-ой международной научно- технической конференции студентов и аспирантов «Информационные технологии, энергетика и экономика» 17-18 апреля 2014 года. Филиал МЭИ в г. Смоленске. Смоленск. 2014. С. 227-229.

12. Гаранин О.И. Способ моделирования транспортных сетей с использова- нием многомодельного подхода // Сборник материалов областного кон- курса студенческих научных работ 2014 г. Смоленск, 2014.

13. Гаранин О.И., Зернов М.И. Анализ алгоритмов выделения и идентифика- ции лиц на изображениях // Сборник трудов 12-ой международной науч- но-технической конференции студентов и аспирантов «Информационные технологии, энергетика и экономика» 16-17 апреля 2015 года. Филиал МЭИ в г. Смоленске. – 2015. – С. 195-198.

14. Гаранин О.И., Зернов М.И. Исследование возможностей алгоритмов распознавания лиц для решения задачи классификации на изображе- ниях // Сборник трудов пятой международной научно-практической конференции «Информатика, математическое моделирование, эконо- мика». Смоленский филиал Российского университета кооперации. – 2015. – С. 44-48.

15. Гаранин О.И. Применение алгоритмов распознавания лиц и

логики про- странства для распознавания зрительной сцены. // Естественные и техни- ческие науки. – 2016. – №9(99). – С. 105–108.

16. Гаранин О.И. Способ выделения эмпирического рецептивного поля свер- точной нейронной сети // Нейрокомпьютеры разработка и применение. – 2017. – № 3. С. 63-69.

17. Гаранин О.И. Способ настройки многомасштабной модели детектирова- ния визуальных объектов в сверточной нейронной сети // Нейрокомпью- теры разработка и применение. – 2018. – № 2. С. 50-56.

18. Золотых Н.Ю., Кустикова В.Д., Мееров И.Б. Обзор методов поиска и со- провождения транспортных средств на потоке видеоданных // Вестник Нижегородского университета им. Н.И. Лобачевского. – 2012. – № 5. – С. 348–358.

19. Поспелов Д.А. Ситуационное управление: теория и практика – М.: Наука

– Физ. мат. лит. – 1986.

20. Тассов К.Л., Бекасов Д.Е. Обработка перекрытий в задачах отслеживания объектов в видеопотоке // Инженерный журнал: наука и инновации. – 2013. – № 6.

21. Сакович И.О., Белов Ю.С. Обзор основных методов контурного анализа для выделения контуров движущихся объектов // Инженерный журнал: наука и инновации. – 2014. – № 12.

22. Тимошенко Д.М. Методы автоматической идентификации личности по изображениям лиц, полученным в неконтролируемых условиях: Диссер- тация на соискание ученой степени кандидата технических наук. – Санкт- Петербург, 2014. – 140 с.

23. Филатов И.Ю. Алгоритмы совместной обработки информации от борто- вых источников летательного аппарата на основе логики взаимного рас- положения объектов: Автореферат диссертации на соискание ученой сте- пени кандидата технических наук // РГРТУ. Рязань. – 2006. – 22 с.

24. Ющенко А.С. Методы нечеткой логики в управлении

мобильными мани- пуляционными роботами // Вестник МГТУ им. Н.Э.Баумана. Приборо- строение. – 2012. – № 7. – С. 29–43. Ahonen T., Hadid A., Pietikainen M. Face Recognition with Local Binary Pat- terns // Proc. 8th European Conference on Computer Vision (ECCV). – 2004. P. 469–481.

25. Babenko A., Lempitsky V. Additive Quantization for Extreme Vector Com- pression // Proceedings of the IEEE Computer Society Conference on Com- puter Vision and Pattern Recognition. – 2014. P. 931–938.

26. Babenko A., Lempitsky V. Aggregating Deep Convolutional Features for Im- age Retrieval // Proceedings of the IEEE International Conference on Com- puter Vision. – 2015. P. 1269–1277.

27. Bay H., Tuytelaars T., Gool L.V. SURF: Speeded Up Robust Features // Proc. 10th European Conference on Computer Vision (ECCV). – 2006. P. 404–417.

28. Berclaz J., Fleuret F., Fua P. Robust people tracking with global trajectory Computer Vision and Pattern Recognition. – 2006. – P. 744–750.

29. Bewley A., Ge Z. Simple online and realtime tracking // arXiv.org [Электрон- ный ресурс]. 2016. – URL: https://arxiv.org/abs/1602.00763 (дата обращения: 14.03.2018).

30. Borisov V.V., Garanin O.I. A Method of Dynamic Visual Scene Analysis Based on Convolutional Neural Network // In Proc. of 16th Russian Confer- ence on Artificial Intelligence, RCAI-2018, Moscow, Russia, in September 24–27, 2018. – Springer: Communications in Computer and Information Sci- ence. Vol. 934. – PP. 60–69. https://doi.org/10.1007/978-3-030-00617-4.

31. Brunelli, R. Template Matching Techniques in Computer Vision: Theory and Practice // Wiley. – 2009.

32. Cai B., Xu X. BIT: Biologically Inspired Tracker // IEEE Transactions on Im- age Processing. – 2016. – 25(3). – P. 1327–1339.

33. Chi Z., Li H., Dual Deep Network for Visual Tracking // arXiv.org[Электронный ресурс]. 2016. : https://arxiv.org/abs/1612.06053 (дата обращения: 14.03.2018).