

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет
Кафедра

Комп'ютерної інженерії та управління
Комп'ютерних інтелектуальних технологій та систем

КВАЛІФІКАЦІЙНА РОБОТА **Пояснювальна записка**

рівень вищої освіти другий (магістерський)

Розпізнавання гравців та відстеження об'єктів під час спортивних змагань за
допомогою згорткових штучних нейронних мереж

Виконав:

студент 2 курсу, групи КІТм-20-1

Черній А.О.

Спеціальність 123 Комп'ютерна інженерія

Тип програми освітньо-професійна

Освітня програма Комп'ютерні

інтелектуальні технології

Керівник проф. Безсонов О.О.

Допускається до захисту

(підпис)

Зав. кафедри

(підпис) проф. Руденко О.Г.

2021 р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерної інженерії та управління

Кафедра Комп'ютерних інтелектуальних технологій та систем

Рівень вищої освіти другий (магістерський)

Спеціальність (напрямок) 123 – Комп'ютерна інженерія
(код і назва)

Освітня програма Комп'ютерні інтелектуальні технології
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

2
0 р

« ____ » _____ .

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Чернію Антону Олександровичу
(прізвище, ім'я, по батькові)

1. Тема роботи Розпізнавання гравців та відстеження об'єктів під час спортивних змагань за допомогою згорткових штучних нейронних мереж

затверджена наказом по університету від « 08 » листопада 2021 р. № 1666Ст

2. Термін подання студентом роботи до екзаменаційної комісії 10 грудня 2021 р.

3. Вхідні дані до роботи _____

1) навчальна множина зображень обличчя людей з лицьовими масками та без них;

2) побудова штучної нейронної мережі;

3) Алгоритм визначення пози людини OpenPose

4) OpticalFlow

4. Перелік питань, що потрібно опрацювати в роботі _____

1) огляд предметної області;

2) аналіз предмету дослідження;

3) дослідження методів використання нейронних мереж;

4) розробка штучної нейронної мережі;

5) експериментальні дослідження

6) аналіз якості згорткової нейронної мережі;

7) висновки.

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (слайдів) Демонстраційні матеріали. Плакати – 15 арк. ф. А4

6. Консультанти розділів роботи (заповнюється за наявності консультантів згідно з наказом, зазначеним у п.1)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Видача та узгодження теми проекту	08.11.2021	
2	Огляд стану проблеми та постановка задачі	09.11.2021 – 10.11.2021	
3	Аналіз літератури за напрямком магістерської роботи	10.11.2021 – 14.11.2021	
4	Вибір методів рішення для реалізації та їх обґрунтування	16.11.2021 – 20.11.2021	
5	Експериментальні дослідження	21.11.2021 – 28.11.2021	
6	Оформлення пояснювальної записки	29.11.2021 – 05.12.2021	
7	Підготовка графічного матеріалу	06.12.2021 – 10.12.2021	
8	Перевірка виконаного проекту курівником	10.12.2021	
9	Захист проекту	16.12.2021 – 17.12.2021	

Дата видачі завдання 08 листопада 2021 р.

Студент _____
(підпис)

Керівник роботи _____
(підпис)

проф. Безсонов О.О.
(посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка кваліфікаційної роботи: 84 с., 36 рис., 2 таблиці, 46 джерел.

РОЗПІЗНАВАННЯ ГРАВЦІВ ТА ВІДСТЕЖЕННЯ ОБ'ЄКТІВ ПІД ЧАС СПОРТИВНИХ ЗМАГАНЬ ЗА ДОПОМОГОЮ ЗГОРТКОВИХ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ

Розпізнавання людських дій з цифрового відео є гарячою темою в області комп'ютерного зору. З огляду на експоненційний ріст даних спостереження в Інтернеті в останні роки, як запровадити ефективний та ефективний аналіз відеоданих є надзвичайно важливим. Традиційні методи машинного навчання, які виділяють лише обчислювані ознаки, мають обмеження і не підходять для масивних візуальних даних, в той час як методи глибокого навчання, особливо згорткові нейронні мережі, досягли великих досягнень у цій галузі.

Метою розпізнавання дій людини є класифікація шаблонів, щоб зрозуміти дії людини з візуальних даних та експортувати відповідні теги. На додаток до просторової кореляції, яка існує в 2D-зображеннях, дії людини у відео мають кореляцію у часовій області. Через складність людських дій зміна перспективи, фонові шуми та умови освітлення впливатимуть на розпізнавання.

На основі згорткових нейронних мереж, CNN+LSTM і 3D-CNN використовуються для ідентифікації дій людини. Дослідження, проведені в роботі перевірили алгоритми розпізнавання людських дій.

Ключові слова: розпізнавання дій людини, згортка нейронна мережа, глибоке навчання, LSTM, 3D-CNN, двопотокова CNN

ABSTRACT

Explanatory note of the qualification work: 84 pages, 36 figures, 2 tables, 46 sources.

PLAYER RECOGNITION AND OBJECT TRACKING DURING SPORTS COMPETITIONS USING ROLLED ARTIFICIAL NEURAL NETWORKS

Recognizing human actions from digital video is a hot topic in the field of computer vision. Given the exponential growth of observational data on the Internet in recent years, how to implement effective and efficient video data analysis is extremely important. Traditional machine learning methods, which emit only computational features, are limited and not suitable for massive visual data, while deep learning methods, especially convolutional neural networks, have made great strides in this area.

The purpose of human action recognition is to classify templates to understand human actions from visual data and export relevant tags. In addition to the spatial correlation that exists in 2D images, human actions in video are time-correlated. Due to the complexity of human actions, changing perspectives, background noise and lighting conditions will affect recognition.

Based on convolutional neural networks, CNN + LSTM and 3D-CNN are used to identify human actions. The research conducted in the work tested the algorithms for recognizing human actions.

Keywords: human action recognition, convolutional neural network, deep learning, LSTM, 3D-CNN, two-stream CNN

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерної інженерії та управління _____

Кафедра _____ Комп'ютерних інтелектуальних технологій та систем _____

АНОТАЦІЯ

КВАЛІФІКАЦІЙНОЇ РОБОТИ

рівень вищої освіти другий (магістерський)

Розпізнавання гравців та відстеження об'єктів

під час спортивних змагань за допомогою

згорткових штучних нейронних мереж

Виконав:

студент 2 курсу, групи КІТм-20-1

Черній А.О.

Спеціальність 123 Комп'ютерна інженерія

Тип програми освітньо-професійна

Освітня програма Комп'ютерні інтелектуальні

технології

Керівник проф. Безсонов О.О.

2021 р.

Черній А.О. Розпізнавання гравців та відстеження об'єктів під час спортивних змагань за допомогою згорткових штучних нейронних мереж – Магістерська кваліфікаційна робота.

У магістерській науковій роботі вирішено актуальну проблему створення прототипу моделі розпізнавання, яка дозволяє розпізнавати людей під час спортивних змагань та в повсякденному житті.

Об'єктом дослідження є процеси розпізнавання гравців під час спортивних змагань. Предметом дослідження: моделі та методи нейронних мереж для розпізнавання образів.

Метою кваліфікаційної роботи є дослідження штучних нейронних мереж для розпізнавання образів.

Для цього було проведено дослідження сучасних архітектур штучного інтелекту, машинного зору та декілька видів нейронних мереж. Виконано системний аналіз існуючих технологій та систем. Імітаційне моделювання показало доцільність використання запропонованих підходів.

Актуальність теми дослідження. Розпізнавання людських дій з цифрового відео є гарячою темою в області комп'ютерного зору. Він має чудовий асортимент додатків у безлічі областей, таких як відеоспостереження, взаємодія людини та комп'ютера, пошук візуальної інформації та безпілотне водіння. З огляду на експоненційний ріст даних спостереження в Інтернеті в останні роки, як запровадити ефективний та ефективний аналіз відеоданих є надзвичайно важливим. Традиційні методи машинного навчання, які виділяють лише обчислювані ознаки, мають обмеження і не підходять до масивних візуальних даних, в той час як методи глибокого навчання, особливо згорткові нейронні мережі, досягли великих досягнень у цій галузі. Метою розпізнавання дій людини є класифікація шаблонів, щоб зрозуміти дії людини з візуальних даних та експортувати відповідні теги. На додаток до просторової кореляції, яка існує в 2D-зображеннях, дії людини у відео мають кореляцію у часовій області. Через складність людських дій зміна перспективи, фонові шуми та умови освітлення впливатимуть на розпізнавання. Для вирішення цих складних проблем у цій дипломній роботі висвітлено питання, що допомагають вирішити ці питання.

Розпізнавання людських дій із відео базується на аналізі послідовності відеокадрів за допомогою комп'ютерів, щоб автоматично знаходити дії людини без ручних операцій. В епоху Інтернету з мобільними телефонами повсякденне життя людей було оточене контролем доступу до воріт, датчиків дорожнього руху, камер безпеки та багатьох інших. Повсюдні камери дозволяють відстежувати дії кожного в громадських місцях, ідентифікація дій людини на відеоспостереження має величезне значення у сфері кібербезпеки. Крім того, аналіз і розуміння людських дій у цифрових відео охоплює кілька цікавих тем дослідження, як-от виявлення об'єктів, семантична сегментація, аналіз руху тощо. Отже, аналіз дій людини має широкий спектр застосувань, включаючи інтелектуальне спостереження, інтелектуальний догляд.

Структурна інновація штучних нейронних мереж ефективна та перспективна. Під час безперервної оптимізації методів глибокого навчання візуальні об'єкти на зображенні чітко і чітко ідентифікуються. У галузі біометрії складне середовище, таке як орієнтація, освітлення, кут зору, роздільна здатність, впливатиме на розпізнавання людських обличчя і дій. Це вимагає, щоб моделі глибокого навчання мали сильну здатність до узагальнення для подальшої оптимізації. Відповідна кількість шарів нейронної мережі покращить продуктивність моделі глибокого навчання

Статистика у футболі, включаючи аналіз ефективності окремих гравців та всієї команди, є предметом аналізу спортивних аналітиків протягом понад 40 років.

За останні 20 років футбол сильно змінився через те, що ключові показники гравців такі як: середня швидкість, пройдена дистанція, точність пасів, кількість пасів – сильно зросли. По всьому світу футбольні команди починають використовувати інформаційні системи для кількісної оцінки ефективності гравців, наприклад, використовуючи сенсори на гравцях для відстеження їх показників у реальному часі.

У першому розділі представляється значення цієї роботи, викладаються основні дослідницькі питання, які необхідно досліджувати, та проблеми, які

стосуються подолання труднощів розпізнавання дій людини. Цілі та внесок цієї дипломної роботи також будуть чітко розглянуті.

У розділі 2 розглянуто літературу, досліджені поглиблені опитування, засновані на розвитку розпізнавання людських дій. Описані різні алгоритми, які практикували в інтелектуальному розпізнаванні. Проаналізовані алгоритми серії R-CNN, серії YOLO та OpticalFlow відповідно.

У розділі 3 розглядається метод дослідження. По-перше, пропонуються загальні етапи цього проекту. Були розглянуті алгоритми, засновані на глибокому навчанні, включаючи CNN+LSTM, двопотокове згортання та 3D CNN. Оцінюючи точність результатів – отримуємо модель розпізнавання руху з високою точністю та впевненістю.

У розділі 4 будуть реалізовані обчислювані методи та алгоритми.

У розділі 5 на основі експериментальних результатів обґрунтовуються переваги та недоліки. Результати кожного алгоритму представлені у вигляді таблиць і рисунків.

Деталізуються обчислювальні параметри, обчислювальні методи та обчислювані визначники експериментів, детально аналізуються результати.

В ході роботи обговорено важливу роль глибокого навчання у розпізнаванні дій.

Ключові слова: розпізнавання людських дій, згортка нейронна мережа, глибоке навчання, LSTM, 3D-CNN, CNN.

Список опублікованих робіт за темою магістерської роботи:

Черній А.О. Розпізнавання гравців та відстеження об'єктів під час спортивних змагань за допомогою згорткових штучних нейронних мереж // XXV Міжнародний молодіжний форум «РАДІОЕЛЕКТРОНІКА І МОЛОДЬ У XXI СТОЛІТТІ», 21-23 квітня 2021 р., м. Харків.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ.....	
ВСТУП.....	13
1 ПОСТАНОВКА ЗАДАЧІ.....	15
1.1 Визначення показників ефективності у футболі.....	15
1.2 Огляд показників ефективності.....	16
1.2.1 Фізичні/фізіологічні параметри.....	17
1.2.2 Технічні характеристики.....	19
1.2.3 Тактичні параметри.....	21
1.2.4 Чинники, що впливають на ефективність команди.....	22
1.3 Використання відеоаналітики для аналізу параметрів.....	23
2 ОГЛЯД ІСНУЮЧОГО РІВНЯ ТЕХНІКИ.....	25
2.1 Класифікація подій під час аналізу футбольного матчу.....	25
2.2 Наявні підходи до формування векторного уявлення статичних поз....	29
2.3 Методи побудови 2D скелетної моделі.....	30
2.4 Наявні підходи до формування тривимірного векторного представлення руху спортсменів.....	32
2.5 Наявні підходи до аналізу спортивної гри.....	33
2.6 Згорткові нейронні мережі.....	36
2.7 Рекурентні нейронні мережі.....	37
2.8 Методи оптимізації нейронних мереж.....	39
2.9 Алгоритм детектування об'єктів YOLO.....	41
2.10 Алгоритм визначення пози людини OpenPose.....	42
2.11 Побудова 2D моделей людей, які перебувають у полі зору камери.....	43

3 ВИЗНАЧЕННЯ ПОЗ СПОРТСМЕНІВ ЗА ДАНИМИ ВІДЕОСПОСТЕРЕЖЕННЯ.....	44
3.1 Попередня обробка відео.....	45
3.2 Оцінка параметрів траєкторій 2D скелетів.....	50
3.3 Формування навчальної вибірки.....	52
3.4 Використання Optical Flow.....	54
3.5 Застосування фільтра Калмана для згладжування та уточнення траєкторій 2D скелетів.....	57
4 ПОБУДОВА ТРИВИМІРНОЇ СЦЕНИ ТА ФОРМУВАННЯ ПОЗ СПОРТСМЕНІВ У ТРИВИМІРНОМУ ПРОСТОРИ.....	63
4.1 Алгоритм тривимірної прив'язки гравців до сцени.....	63
4.2 Формування тривимірного уявлення на основі двовимірних скелетних моделей.....	67
5 АНАЛІЗ ВЕКТОРНОГО УЯВЛЕННЯ РУХУ СПОРТСМЕНІВ ТА ПОШУК ТАКТИКО-ТЕХНІЧНИХ ДІЙ.....	69
5.1 Попередня обробка.....	69
5.2 Порівняльний аналіз та вибір структури нейронної мережі.....	70
5.3 Результати аналізу та пошукуTTD.....	74
ВИСНОВКИ.....	79
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	81

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ,
СКОРОЧЕНЬ І ТЕРМІНІВ

DL – deep learning – глибинне навчання;

IR – інформаційний пошук – information retrieval;

ML – machine learning – машинне навчання;

MLP – multilayer perceptron – багатошаровий перцептрон;

БНМ (НМ) – багатошарова нейронна мережа;

ЗНМ – згорткова НМ;

ОРС – оптичне розпізнавання символів;

РНМ – рекурентна НМ;

ШНМ – штучна нейронна мережа;

ШН – штучний нейрон.

ВСТУП

Відеотрансляції футбольних матчів є послідовністю відеозображень з набору відеокамер. При цьому зазвичай зйомка ведеться з декількох камер одночасно. Відеотрансляція редактором розбивається на низку фрагментів, кожен з яких може бути віднесений до одного з кількох планів. Оскільки трансляція зазвичай ведеться в режимі on-line, виникає можливість її використання для аналізу гри в реальному часі на користь підтримки прийняття рішень тренерами команд. При цьому іноді доступні дані з декількох камер. Важливою особливістю є використання відеозйомки високої роздільної здатності. Вже сьогодні стандартом де-факто є використання 4K відео. Наявність такого чудового інструменту, як відео високої роздільної здатності дозволяє перейти від суб'єктивної оцінки гри команди тренером, який знаходиться на лавці біля поля до об'єктивної оцінки як індивідуальних, так і колективних показників команди, що змінюються в процесі гри.

Актуальність цієї роботи обумовлена тим, що у наш час футбол перетворився на самостійну індустрію транснаціонального характеру. При цьому аналіз самої гри перетворився на актив, що прямо впливає на економічну ефективність команд, як учасників глобального ринку. Метою даної є розробка алгоритмічного підходу, що дозволяє отримати об'єктивну оцінку індивідуальних та командних показників безпосередньо в ході футбольного матчу на основі аналізу відеотрансляції високого дозволу з використанням технологій штучного інтелекту, в особливості – згорткових нейронних мереж. Даний алгоритмічний підхід дозволяє проводити аналіз руху спортсменів на основі нейромережевих алгоритмів обробки відео з камер високого дозволу. Для досягнення поставленої мети необхідно було вирішити такі завдання:

- проаналізувати наявні підходи, що застосовуються у спортивній аналітиці для оцінки індивідуальних та групових характеристик спортсменів та виявлення тактико-технічних дій;

- розробити алгоритм формування статичних та динамічних скелетних моделей поз спортсменів у двовимірному та тривимірному просторі;
- сформувати масив даних на основі публічно доступних відеотрансляцій та провести виявлення та аналіз тактико-технічних дій, що виконуються спортсменами.

Основні положення, що виносяться на захист:

- алгоритм формування двовимірних векторних скелетних моделей спортсменів на основі відеотрансляцій високої роздільної здатності;
- алгоритм адаптивного управління областю обробки для відео високої роздільної здатності;
- алгоритм виявлення тактико-технічних дій на основі аналізу руху багатовимірного вектора, що описує двовимірну скелетну модель гравця;
- навчена та інтегрована до складу системи обробки відео модель пошуку та класифікації типових тактико-технічних дій гравців на основі згорткової нейронної мережі.

Наукова та практична значущість підтверджується тим, що розроблені методи та алгоритми дозволяють здійснювати автоматизовану обробку великих обсягів відеоаналітики та формувати набори показників, що описують індивідуальні та групові особливості учасників гри. Ступінь достовірності одержаних результатів забезпечується проведенням тестових випробувань розроблених алгоритмів на реальних даних, як отриманих самостійно, і завантажених з публічних архівів відеоматеріалів.

1 ПОСТАНОВКА ЗАДАЧІ

1.1 Визначення показників ефективності у футболі

Статистика у футболі, включаючи аналіз ефективності окремих гравців та всієї команди, є предметом аналізу спортивних аналітиків протягом понад 40 років.

Роль колективних дій футбольної команди була точно сформульована В. В. Лобановським ще в 1989 [4]:

«Колективний осмислений рух поставлений на чільне місце в сучасному футболі. Ігровий прямокутник став таким же доступним, як баскетболістам – баскетбольний майданчик. Футболістам, підготовленим, природно, як личить, не важко промчатися 50-60 метрів вперед і стільки ж – назад. На полі не має бути пасивних гравців. Сьогоднішній футболіст не є самотнім солдатом, який бореться у своєму кутку із супротивником. Він охоплює все поле битви, а на цьому полі немає клаптиків, вільних від боротьби.»

Таким чином було формально зафіксовано, що оцінка якості команди визначається безліччю індивідуальних та командних показників, управління якими і є завданням тренера.

Завданням спортивного (футбольного) аналітика, у свою чергу, є створення системи безперервного моніторингу індивідуальних та командних показників.

За останні 20 років футбол сильно змінився через те, що ключові показники гравців такі як: середня швидкість, пройдена дистанція, точність пасів, кількість пасів – сильно зросли [5]. По всьому світу футбольні команди починають використовувати інформаційні системи для кількісної оцінки ефективності гравців, наприклад, використовуючи сенсори [6] на гравцях для відстеження їх показників у реальному часі.

Також для оцінки ефективності гравців використовують комп'ютерний зір, який вже добре зарекомендував себе. Воно добре вирішує безліч задач, як, наприклад, відстеження гравців, м'яча, аналіз рухів гравців, які допомагають тренерам, суддям та футбольним аналітикам [7].

В даний час якість телевізійної трансляції для футбольних матчів високого рівня досягла 4К, що відповідає повнокадровій версії роздільної здатності 4096×3072 . На рисунку 1.1 наведено момент гри – кадр з відео.



Рисунок 1.1 — Кадр з футбольної трансляції високого дозволу

1.2 Огляд показників ефективності

Розглянемо докладніше показники, які можна використовувати при оцінці гравця чи команди.

Показник ефективності – це скалярна або векторна характеристика, що відповідає деякому аспекту ефективності (успішності) у даному виді спорту. Сформована система показників повинна наводити прямо або опосередковано до успішного результату або результату і дозволяти аналізувати динаміку в результаті тренувального процесу.

Приклади показників ефективності для футбольної аналітики наведено, наприклад, у роботах [8, 9].

Використовуючи це визначення, можна визначити показники, які можна розбити на фізичні, фізіологічні, технічні, тактичні і т.д.

Показники ефективності доцільно поділити на індивідуальні та групові.

У деяких випадках груповий показник може бути обчислений як проста функція від персональних показників гравців команди. Прикладом такого показника є середній час на відстані 100 метрів гравців команди.

У деяких випадках групові показники повинні обчислюватися складним способом.

Є набір показників, які можуть обчислюватися у рамках тренувального процесу. Альтернативу є показники, які можуть бути оцінені тільки в процесі гри. При цьому такі показники можуть суттєво змінюватися від гри до гри.

Розглянемо більш детально такі групи показників:

- фізичні та фізіологічні параметри (включаючи антропометрію);
- технічні показники;
- тактичні показники.

1.2.1 Фізичні/фізіологічні параметри

Перший набір параметрів, що описують окремого гравця – антропометричні та фізіологічні параметри. Вони по суті визначають фізичні можливості гравців.

Може здатися, що фізичні можливості гравців є ключовим фактором, що впливає на результат гри. Але той факт, що в ряді випадків команди, що програли, долали значно більшу дистанцію [10] говорить про те, що існують показники і фактори, відмінні від фізичної готовності, які більш важливі для підсумкового успіху команди.

Огляд антропометричних даних футболістів проведено, наприклад, в роботі [11]. Незначні відмінності було помічено у складі тіла футболістів.

Проте вдається знайти залежність між командами – переможцями та програли у фізичній підготовці гравців. На прикладі різних дій та вправ ці залежності та кореляції між фізичною формою та результатом футбольних матчів наведено у роботах [12–14].

Цікаво, що покращені фізичні можливості гравців впливають на їх технічні здібності. Зокрема, вони можуть впливати на швидкість обробки та передачі м'яча і, отже, можуть опосередковано покращувати технічні та тактичні показники гравця.

Дослідження на основі результатів Ліги А в Італії [15] та англійської Прем'єр ліги [16] показали зв'язок між фізичною віддачею та фінальним результатом команд у рейтингу ліги [10]. Зокрема, була проаналізована на кореляція між швидкістю переміщень гравців та положенням команд у підсумковій таблиці ліги наприкінці ігрового сезону. Однак виявилось, що явної кореляції подоланої дистанції та фінального результату команди немає [17].

Очевидно, що фізичні можливості мають бути основою «готовності до гри». При цьому різні стилі гри можуть впливати на специфічні вимоги до фізичної форми гравців.

В свою чергу, стан гравця «не у формі», «не готовий грати» часто переходить у стан «травмований» і водночас сильно впливає на результат гри. Показано, що у команд, що виграли, було значно менше травм порівняно з командою, що прогнала [12].

З вищесказаного природним чином випливає, що досить універсальним є показник гравця «пройдена за гру відстань» [18].

Однак, як правило, на результат аналізу руху гравця в часі також впливають багато зовнішніх факторів (рахунок матчу, володіння м'ячем, противник, погода та інші).

В рамках роботи не ставилось завдання оцінки фізичного або фізіологічного стану гравців за результатами аналізу рухів.

1.2.2 Технічні характеристики

Другою групою індивідуальних параметрів є технічні.

Подібно до фізичних/фізіологічних параметрів, технічні параметри відіграють у футбольній аналітиці важливу роль. Найпростішим способом вимірювання технічних параметрів (таких як удари по воротах, підкати, фінти) є їх підрахунок. Цей вид аналізу прийнято називати нотаційним аналізом (notational analysis) [8].

Той факт, що існує зв'язок між кількістю забитих голів та пропущених голів та результатом гри, є очевидним. Але щодо інших параметрів, таких, як удари по воротах, передані паси, обробки м'яча, фінти явну кореляцію з результатом гри вибудувати вже складніше.

Тому слід керуватись підходом, запропонованим у роботі [8], де також в аналіз технічних дій запропоновано включати паси (передачі м'яча), боротьбу та удари.

Паси описуються такими можливими показниками:

Кількість отриманих та переданих пасів – характеристика, яка описує кожного гравця та команду загалом.

Довжина ланцюжка пасів – характеристика, що описує команду в цілому в частині довжини ланцюжка передач м'яча між гравцями команди до того моменту, коли ланцюжок буде перервано єдиноборством або ударом по воротах.

Типи пасів – складно формалізована характеристика, що описує особливості пасів (дальність, висота польоту м'яча, закрутка м'яча). Це найскладніша частина аналізу пасу як технічної дії.

Паси в конкретну область – показник формується шляхом розбиття ігрового поля на зони і далі кожному пасу ставиться у відповідність дві зони. Одна, в якій знаходився гравець, який виконує передачу і друга, в якій знаходився гравець, який приймає передачу.

Проникаючі паси до штрафної та воротарської зони, здавалося, вели кількість ударів по воротах і, отже, кількість голів при атаці. Цікаво, що невдалі команди мали тенденцію грати значно більше передач в область, де розташовані захисники.

Точність пасів Немає сумнівів у тому, що точність пасу є важливим показником ефективності у футболі загалом. Наприклад, було показано, що точність пасу була значно вищою за 5 хвилин до гола. Точність передачі не тільки зберігає володіння м'ячем, але також може призвести до можливостей виграшу, а також обмежити можливості м'яча у противника.

Пас противнику, перехоплення – частково відноситься до параметрів пасу.

Боротьба (виграна/програна) є другим типом технічних дій відповідно до класифікації, запропонованої в роботі [8].

Доступно досить обмежену кількість досліджень, що містять аналіз боротьби у футболі та їх впливу на результат гри, робота [19] є досить повним, але не дуже деталізованим дослідженням, що закриває повний погляд на нотаційний аналіз.

Боротьба у свою чергу може бути піддана подальшій класифікації.

Можна окремо розглянути фоли та грубі фоли (за підсумками яких було пред'явлено жовті або червоні картки).

Окремо можна класифікувати боротьбу у захисті чи нападі.

Очевидним результатом боротьби є сторона, яка має м'яч за підсумками боротьби.

Удари по воротах/мимо воріт є досить простим показником, що описує співвідношення сил команд, що грають.

Традиційно їх класифікують на три категорії – гольові удари, удари в площину воріт і удари повз воріт.

Прикладом корисного показника для опису ударів по воротах є класифікація зон, з яких наносився удар. У роботі [20] представлена процентна статистика успішних ударів з певних областей/зон на Чемпіонаті

Європи в Португалії у 2004 році: штрафна площа (44,4%), площа воріт (32,2%) та нестандартні положення (20,4%)).

Проте 90% всіх голів було забито у штрафному майданчику на чемпіонаті світу 1986 року, тоді як у чемпіонаті світу 1990 року цей відсоток становив вже 80% [20], а голи зі штрафної площі для грецького внутрішнього футболу становлять 81,8% [21].

Як бачимо статистикою це назвати складно, оскільки все залежить від рівня турніру, погоди, складу учасників та безлічі інших факторів.

1.2.3 Тактичні параметри

Тактичні параметри виділені в окрему групу але при цьому досить точно схожі на параметри технічні. Основна відмінність у тому, що тактичні параметри визначено лише команди.

Розглянемо деякі приклади:

Володіння м'ячем є добре формалізований параметр. Слід вважати, що м'ячем володіє команда, яка останньою мала контакт із м'ячем.

Якщо запитати, які зовнішні фактори впливають на параметр «володіння м'ячем», то стане зрозуміло, що на параметр впливає бажаний результат («гра на нічию» або «гра на перемогу»), співвідношення рівня команд. [22].

Незважаючи на все перераховане, вважається, що володіння м'ячем має первісне значення для успіху в грі. При цьому деякі питання щодо його впливу на позитивні результати команди у конкретній грі залишаються. Здається, що в домашній грі володіння м'ячем є важливішим показником ефективності (особливо в Англії та Іспанії) [23-25].

Аналітики виявили, що у разі зваженого обліку сил команд та коефіцієнта на «домашнє поле», володіння м'ячем перестає бути стійкою ознакою, що визначає результат.

Тим не менш, команди переможці мають, як правило, більший відсоток володіння м'ячем.

Володіння м'ячем на певній ділянці поля є приватним випадком обговореного раніше параметра «володіння м'ячем».

Неважко помітити, що на володіння м'ячем у певній області поля впливає рахунок на табло під час гри. Команда, що поступається, значно збільшує своє володіння в середині і атакуючої третини поля і зменшує володіння своєї власної оборонної третини поля. Крім того, володіння м'ячем під час матчу показує, що команда, яка забила гол, має протягом деякого часу більше володіння у середині та атакуючої третини поля.

Володіння м'ячем, як послідовність пасів вже згадувалося вище при обговоренні технічних показників.

У роботі [26] проаналізовано кількість пасів, що призвели до голів, у фіналах Кубка Світу ФІФА. Виявилось, що було більше голів забитих при довгих послідовностях пасів, ніж у результаті коротких послідовностей. Більше тривале володіння (більше 5 пасів) призводить до більшої кількості ударів по воротах.

1.2.4 Чинники, що впливають на ефективність команди

Ефективність команди є дуже специфічним похідним показником, на який впливають багато змінних, обговорених вище.

Команда, що виграла, значно відрізнялася від тих, що програли такими параметрами, як :

- голи;
- удари в площину;
- ефективність ударів по воротах (співвідношення голів на удари);
- володіння м'ячем загалом;
- позиційні характеристики (наприклад, контратака);
- володіння з 0-4 пасами;

- успішні передачі.

Як правило, команди – переможці мали більше відсоток володіння м'ячем, ніж команди, що програли при перемозі або нічиєї, і ніякої різниці не було видно під час поразки. Цікаво, що невдалі команди мали менше володіння м'ячем при перемозі порівняно з програшем, проте як успішні, так і безуспішні команди мали більше часу володіння м'ячем, коли вони програвали порівняно з перемогою.

Успішні команди воліли використовувати паси низом у фінальній третині поля, порівняно з невдалими командами, які переважно використовували паси верхи. Успішні команди були також ефективніші, ніж їхні противники, у розіграшах стандартних положень.

1.3 Постановка завдання використання відеоаналітики для аналізу параметрів

Описані вище у цьому розділі показники донедавна, зазвичай, витягувалися з відеофайлів методами ручного розбору.

Відносно недавно з'явилися технології трекінгу гравців та м'яча. При цьому розмітка ТТД і прив'язка ТТД до тимчасової осі однаково існує в експертному режимі.

Разом з тим сучасні технології машинного зору дозволяють обробляти відео в реальному часі.

Відомі реалізації досить складних алгоритмів, які виконуються в реальному масштабі часу на основі обчислювачів на основі GPU.

У роботі запропоновано використовувати методи машинного навчання для визначення параметрів руху гравців у футболі.

При цьому в найпростішому випадку розглядається двомірне уявлення, в рамках якого гравець і м'яч описуються точками на прямокутному двомірному плані.

У складнішому випадку запропоновано описувати гравця як векторного уявлення, у якому гравець описується вектором, компоненти якого відповідають 20-28 суглобам (кисті і стопи описуються без деталізації).

При цьому і в найпростішому і складному векторному випадку на основі аналізу руху гравців проводиться ідентифікація ТТД та визначення показників, що описують особливості виконання ТТД.

Прикладами показників можуть бути як індивідуальні, так і групові показники.

Результатом застосування алгоритмів, що розробляються, має стати розшифровка футбольного матчу у вигляді набору векторних описів ТТД гравців у прив'язці до тимчасової осі.

2 ОГЛЯД ІСНУЮЧОГО РІВНЯ ТЕХНІКИ

2.1 Класифікація подій під час аналізу футбольного матчу

Одним із завдань футбольної аналітики є завдання класифікації футбольних дій. Одна з робіт, присвячена класифікації дій [27], розповідає про розв'язання цього завдання за допомогою рекуррентної нейронної мережі [28]. Для отримання ознак з відео було використано 2 підходи: Bag of Words і SIFT [29]. Як рекуррентна мережа була використана LSTM [30], яка навчалася на даних MICC-Soccer-Actions-4 [31] і показала точність 0.92.

Схожа робота групи Агіємана та ін. [32] також розповідає про застосування рекуррентних нейронних мереж у задачі сумаризації матчу. Автори використовували 3-мірні згорткові мережі та LSTM на 744 фрагментах для класифікації 5 подій: штрафний удар, кутовий удар, гольовий момент, розлучення з центру та кидок з ауту. На даних «Soccer-5» алгоритм показав точність 96,81% за метрикою assuarcy.

Шліпсінг та інші [33] також вирішували завдання класифікації подій. У цій роботі дані були зібрані від двох камер, що містять 6000 прикладів. Для обробки відео футбольне поле було відокремлене від фону. Також було запропоновано класифікувати всіх персон на полі, як гравці, люди поза полем, судді та воротарі. Для польових гравців було розроблено метод відстеження [34] на формування шляху, пройденого на полі. Для класифікації гравців було запропоновано 3 види вилучення ознак: гістограми кольорів, метод головних компонентів [35] та метод spaciograms [36]. А для класифікації було запропоновано 3 алгоритми: алгоритм найближчих сусідів, латентне розміщення Діріхле, метод опорних векторів. Також було вирішено проблему із розпізнаванням конкретного гравця відстеженням шляху цього гравця. Як інші елементи аналітики було запропоновано кілька візуалізацій,

як, наприклад, теплова карта, сітка, що показує найчастішу комірку, і карту шляху по кожному гравцю.

Інше завдання у відеоаналітиці було вирішено групою дослідників з Ватерлоо [37]. Метою цієї роботи було визначити лише корисні для аналізу фрагменти, тобто визначити ігровий час та відокремити його від фрагментів, коли гра зупинена (play-break-segmentation) та визначити вид зйомки: близький, середній, далекий та поза полем (shot-view– classification). Для вирішення завдань група використовувала мережі з паралельним злиттям ознак

У марківській моделі із прихованою передачею. Для першого завдання використовувалася модель Hidden-to-Observable Transferring Markov Model (H2O MM), а для визначення виду зйомки – Parallel Feature Fusion Network (PFF-Net). PFF-Net складалася з 3 частин: локальний перетворювач даних, кілька паралельних автоенкодерів [38] і LG-класифікація, що складалася з шару softmax для байєсовського шару. Навчання відбувалося на даних, що складаються з 9:00 10 футбольних матчів. PFF-Net виявилася точною на 92%, а H2O MM на 96%.

Ще одним завданням відеоаналітики є детекція подій. Однією з робіт, присвячена розв'язанню це завдання, а також вирішенню завдання збагачення даних для відеоаналіки у футбол, є робота групи Джанколи та ін[39]. Для вирішення завдання були зібрані дані з 500 ігор, кількість яких складає 6637 дій. Завданням була детекція трьох подій (жовта картка, червона картка, заміна). Дані були зібрані в 3 етапи: складання відео з відкритих джерел, синхронізація годинника та парсинг звітів з математики та формування статистики. Також потрібно вирішити 2 завдання: детекція події на відрізьку відео довжиною в 1 хвилину і детекція подій на всьому відео. Для першого завдання використовувалася перенавчена нейронна мережа ResNet [40] з різними пулінгами. Для другого завдання було використано рішення першого завдання на всьому відео з вікном в 1 секунду.

Також завдання детекції дій вирішували Хан та ін. [41]. Метою було про гноз 2 подій: штрафного удару та кутового удару. У пайплайн класифікації була включена система детектування (YOLO) [42] як перший ступінь вирішення задачі детекції. Для класифікації подій використали такий інструмент як калькулятор подій (Event Calculus) [43]. Навчання відбувалося на даних із вище висловленої роботи «SoccerNet». Точність предсказання кутового удару виявилася 68.4 % за метрикою precision і 65 % за метрикою recall, тоді як для кутового удару точність виявилася 83.33 % за метрикою precision і 25 % за метрикою recall.

Завдання сумаризації вирішували вчені з університету Каїра [44]. Для цього команда застосовує опорних векторів для виділення важливих сегментів з появою табло з логотипами команд та рахунком у грі. Також колеги використовують алгоритм К-середніх [45] та перетворення Хофа для виявлення воріт та фільтр Габора для виявлення гола. В результаті точність виявлення лого за метрикою recall склала 98.1 %, а за метрикою precision 92.8 %. Точність виявлення голу склала 98.5% по метриці recall та 93.1% по метриці precision.

Вчені з університету Хакайдо вирішували завдання вилучення схожих сцен за допомогою навчання на мультимодальні метричні відстані [46]. Для вирішення вони використовували вилучення ознак з кадрів, аудіо і тексту за допомогою згорткових мереж та Word2Vec [47]. Потім вони обчислюють вагову матрицю для визначення відстані між уявленнями ознак, після цього знаходили максимально схожі сцени. В результаті точність алгоритму за метрикою mAP@10 дорівнювала 0.852.

Вчені з університету Сіань Цзяотун вирішували завдання детекції подій використовуючи приховані умовні випадкові поля [48]. Для початку вони витягли ознаки низького рівня, далі класифікували відрізки відео за змістом і сегментували їх. Наприкінці за допомогою навченої моделі прихованих умовних випадкових полів розпізнавали кожен відрізок відео. В

результаті точність у моделі за метрикою recall виявилася 86.55%, а за метрикою average precision 88.03%.

Група Чоппа займалися детекцією дій використовуючи для цього спеціальну функцію втрат [49]. Вона є лінійною комбінацією двох функцій втрат: функції втрат сегментації та функції втрат виявлення дій. Навчання відбувалося на даних «SoccerNet» і як архітектура нейронної мережі для формування ознак високого рівня колеги використовували архітектуру «ResNet». В результаті точність на тестовій вибірці за метрикою mAP дорівнювала 62.5 %.

Вчені з університету Шаріфа розробили метод швидкої гомографії з її уточненням для лінії штрафної [50] за допомогою точково-лінійної моделі для гомографії та 2-х моделей для уточнення: «random sample consensus» та ітеративний метод. В результаті порівняно з state-of-the-art алгоритмом «Random Sample Consensus» виявився точніше в 2 рази, а ітеративний метод був у 3.9 разів швидше.

Пандья і Завері також займалися сегментуванням подій для детекції жовтих карток і гольових моментів та атак [51]. Для цього вони використовували Optical Flow, колірні ознаки високого рівня та Hidden Markov Model для фільтрації та категоризації подій. Навчалася та тестувалася модель на європейських лігах. Точність моделі на тестовій вибірці дорівнювала 72.4 % за метрикою precision і 88.12 % за метрикою recall.

Команда Ванга представили роботу, присвячену ідентифікації гри, агресивної, використовуючи траєкторії рухів гравців [52].

Група з Амстердама розповіла про сумаризацію непрофесійних ігор [53]. Для цього вони кластеризують траєкторії за допомогою нормалізованих графічних розрізів [54], а потім класифікували дії (ходьба, біг, удар) використовуючи метод опорних векторів. В результаті точність алгоритму на тестовій вибірці дорівнював 79% за метрикою accuracy.

2.2 Наявні підходи до формування векторного уявлення статичних поз

Виявлення та побудова параметрів руху людей є дуже важливим застосуванням алгоритмів машинного зору та мають безліч практичних застосувань. Найбільш перспективними виглядають застосування такого рішення в розподілених системах відеоспостереження, рішеннях для маркетингових досліджень, аналізу руху людей у ході виробничих процесів, рішеннях для спортивної аналітики.

Це завдання має кілька різних рівнів деталізації. Високорівневий підхід покликаний дозволити знаходити в кадрі об'єкти і побудувати описану навколо них прямокутну область у 2-вимірному або 3-вимірному просторі.

Більш складним і цікавим завданням є завдання визначення пози людини у русі. Це завдання є більш складним у зв'язку з тим, що розмірність вектора, що описує позу людини в 3 вимірному просторі навіть при використанні найпростішої моделі порядку 100, що сильно більше розмірності навколишнього паралелепіпеда, описаного навколо об'єкта.

Останні досягнення в області глибоких нейромереж дозволили створити досить ефективні алгоритми виявлення поз на окремому зображенні [55], [56].

Традиційне векторне уявлення для скелета людини у двовимірному вигляді наведено рисунку 2.1.

Послідовно застосовуючи їх до послідовності кадрів відеопотоку, ми можемо отримати пучок траєкторій для точок, які відповідають суглобам скелетної моделі. Але, оскільки ці алгоритми аналізують кожен кадр незалежно, траєкторії, що отримуються при цьому, можуть включати помилки і артефакти різної природи (викиди вимірювань, пропущені точки траєкторії, переплутані точки сусідніх траєкторій).

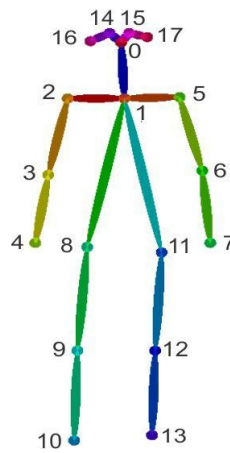


Рисунок 2.1 – Приклад подання персони у вигляді вектора

2.3 Методи побудови 2D скелетної моделі

Наведемо короткий огляд існуючих технологій виявлення та формування векторів параметрів для 2D поз. Насамперед варто згадати роботу дослідників із Carnegie Mellon University, які створили стенд CMU Panoptic Studio, що дозволив сформувати масив розмічених відео даних [57].

Станом на початок 2020 року система включала:

- 480 VGA камер з роздільною здатністю 640 x 480, 25 fps, з апаратно реалізованою синхронізацією;
- 31 HD камер з роздільною здатністю 1920 x 1080, 30 fps, з апаратно реалізованою синхронізацією;
- 10 сенсорів Kinect II з роздільною здатністю 1920 x 1080 (RGB), 512 x 424 (хмара точок), 30 fps, з апаратно реалізованою синхронізацією;
- 5 DLP проекторів синхронізованих із РВ камерами.

Завдання оцінки параметрів поз людей на відео може розглядатися як окремий випадок завдання траєкторної обробки безлічі цілей (multiple object tracking – MOT). У цьому ми відмовляємося від розгляду пози, як цілого, і проводимо обробку безлічі незалежних точкових об'єктів – суглобів.

При розв'язанні задачі аналізу пучка траєкторій використовують одну з двох парадигм – побудову траєкторій на основі вже виконаного виявлення позначок або з детектування позначок «на льоту» [58].

У першому випадку алгоритми машинного зору вирішують задачу виявлення об'єктів незалежно для кожного кадру у послідовності. І вже потім виявлені для кожного кадру об'єкти поєднуються в траєкторії.

У другому підході немає алгоритму виявлення, натомість він покладається на тимчасові зміни у відео потоці для виявлення об'єктів. За останні роки, з розвитком високоефективних алгоритмів виявлення об'єктів у реальному часі, підхід, заснований на детекції, став переважати в літературі. Тим не менш, незалежний аналіз кадрів призводить до неминучої втрати інформації про тимчасові зміни у відео. Це знання може бути корисним у області виявлення об'єктів та допомогти покращити ефективність стеження. Щоб врахувати і окремі кадри, і тимчасові зміни, було запропоновано різноманітні підходи.

Наприклад, в [59] описаний новий підхід по об'єднанню тимчасових та простих властивостей, заснований на додаванні рекурентної тимчасової компоненти до мережі згортки, зробленої для виявлення об'єктів в одному кадрі. Результати виявлення об'єктів з мережі в послідовних кадрах подавалися до RNN. Отримана архітектура потім вчиться пророкувати уточнені положення відстеження.

У [60] пропонується використовувати трекер, який використовує попередню інформацію про можливу динаміку пози людини.

В роботі [61] описано застосування рішення з відкритим кодом OpenPose для вирішення задачі виявлення та визначення параметрів 2D пози у разі наявності в кадрі безлічі персон. Запропонований метод використовує непараметричне уявлення, відоме як Part Affinity Fields (PAFs), для аналізу при належності виявлених частин тіла до персоналій, що знаходяться в кадрі. Цей bottom-up підхід забезпечує досягнення високої точності в реальному часі незалежно від кількості осіб у кадрі.

У статті [62] для уточнення неправильних передбачень положень суглобів автори використовують нову схему доповнення даних для навчання, використовуючи «жорсткі» конструкції пози людини.

У статті [56] представлена оригінальна архітектура MultiPoseNet оцінки поз для безлічі персоналій у кадрі, що поєднує багатопараметричну модель з новим методом асоціації суглобів з персонами.

MultiPoseNet поєднує виявлення фігур, їх сегментацію та оцінку поз. Новий метод асоціації суглобів з персонами реалізований на основі Pose Residual Network (PRN), яка на основі ключових точок та виявлених персоналій у кадрі забезпечує уточнення поз на основі зв'язування ключових точок та персоналій.

У статті [63] проаналізовано підходи до аналізу ходи людей на основі обробки зображень на користь застосування у складі інтелектуальних транспортних систем.

У статті [64] авторами запропонований новий критерій «Occluded Human (OSHuman)» який дозволяє оцінити якість відновлення об'єкта у вигляді структури даних, що включає рамку, що обмежує, позу людини і маски маски фігури. Цей набір даних містить 8110 докладних анотованих фігур людей на 4731 зображенні. З середнім значенням 0,67 для кожної людини OSHuman є на думку авторів найбільш релевантним з наборів даних, пов'язаних із розпізнаванням поз людей.

2.4 Наявні підходи до формування тривимірного векторного представлення руху спортсменів

Ряд робіт присвячено оцінці параметрів поз у тривимірному просторі. Робота [65] пропонує використовувати CNN для перекладу 2D-зображення в 3D.

Автори [66] запропонували два нових підходи до багатопозиційного відновлення пози людини на основі тріангуляційного підходу забезпечує формування 3D об'єкта з безлічі 2D.

Перший підхід представляє використання алгебраїчної тріангуляції з урахуванням ваг точок, одержуваних з вихідних зображень. Другий, більш

складний підхід, заснований на тривимірному агрегуванні 2D теплових карток об'єктів з подальшим уточненням за допомогою тривимірних згорток, які дають остаточні 3D теплові картки для суглобів.

Для більш складного випадку при фільтрації може враховуватися максимально наближена до реальності біокінематична модель руху людини. Для реалізації біокінематичної моделі може, наприклад, використовуватися комерційне рішення AnyBody Technology або рішення з відкритим кодом OpenSim. Коротке обговорення переваг та недоліків наведено в [67].

Найпростіший і витончений підхід наведено у роботі [68].

Пропонується використовувати нейронну мережу для відновлення стереоскопічного зображення (тобто зображення об'єкта, отриманого з двох камер). Далі справжнє зображення та отримане передбачене друге зображення інтерпретуються як стереоскопічна пара. На ній знаходяться об'єкти (люди та їх скелети), а потім за допомогою геометричних алгоритмів обчислюються тривимірні координати за двовимірними координатами з двох зображень через стандартні підходи до обробки стереопар.

У роботі [69] наведено опис підходу до відновлення 3D поз людини для багатооб'єктних відеопослідовностей.

2.5 Найвні підходи до аналізу спортивної гри

Огляд застосування технологій машинного зору та відеоаналітики у спорті наведено у роботі [7]. Також робота містить огляд наборів даних, що використовуються у спортивній індустрії.

Перша стаття «SoccerNet» [39] розповідає про модель детекції дій. Для нього були зібрані дані з 500 ігор кількість яких складає 6637 дій. Алгоритм навчався класифікувати 3 класи (жовта картка, червона картка, заміна). Дані збиралися в 3 етапи: складання відео з відкритих джерел, синхронізація годинника та парсинг звітів по матчу у формування статистики. Також потрібно вирішити 2 завдання: класифікацію події на відрізку відео

довжиною в 1 хвилину і детекція подій на всьому відео. Для першого завдання використовувалася передбачена нейронна мережа ResNet [40] з різними пулінгами. Для другого завдання було використано вирішення першого завдання на всьому відео з вікном в 1 секунду.

Ще одна робота, присвячена класифікації дій [70], використовує для неї рекуррентну нейронну мережу [28]. Для отримання ознак із відео використовувалося 2 підходи: Bag of Words і SIFT [29]. Як рекуррентна мережа була використана LSTM [30], яка навчалася на датасеті MICC-Soccer Actions-4 [31] і показала точність 0.92

Також відносно недавно опублікувалася робота, присвячена аналізу відео футбольних матчів [33]. У роботі дані збиралися від 2 камер, містять 6000 прикладів. Для обробки відео футбольне поле було відокремлено від фону. Також було запропоновано класифікувати всіх персон на полі як гравці, люди поза полем, судді та воротарі. Для польових гравців був розроблений метод відстеження [34] для формування шляху, пройденого на полі. Для класифікації гравців було запропоновано 3 види вилучення ознак: гістограми кольорів, метод головних компонентів [35] та метод spaciograms [36].

Для класифікації було запропоновано 3 алгоритми: NN, LDA та SVM, який показав себе краще за перші два (помилка класифікації вийшла 0.03). Також було вирішено проблему з розпізнаванням конкретного гравця від стеження цього гравця. Як інші елементи аналітики було запропоновано кілька візуалізацій як, наприклад, теплова карта, сітка, що показує найчастішу комірку і карту шляху по кожному гравцю.

Ще одна робота з університету Ватерлоо [37], присвячена футбольній аналітиці розповідає про використання мережі з паралельним злиттям при знаках та марківській моделі із прихованою передачею. Метою цієї роботи було визначити тільки корисне для аналізу фрагменти, тобто визначити гровий час і відокремити його від фрагментів (play-break-segmentation), коли гра зупинена та визначити вид зйомки: близький, середній, далекий і поза

полем (shot-view -classification). Для першого завдання використовувалася модель Hidden-to Observable Transferring Markov Model (H2O MM), а для визначення виду зйомки –Parallel Feature Fusion Network (PFF-Net). PFF-Net складалася з 3 частин: локальний перетворювач даних, кілька паралельних автоенкодерів [38] та LG-класифікація, що складалася з шару softmax та байєсовського шару. Навчання відбувалося на датасеті, що складається з 9 годин 10 футбольних матчів. PFF-Net виявилася точною на 92%, а H2O MM на 96%.

Також внесок у відеоаналітику, а саме внесок у розв'язання задачі детектування подій зробила робота, представлена на конференції GCAI у 2019 році [41]. Метою був прогноз 2 подій: штрафного удару та кутового удару. У пайплайн класифікації було включено систему детектування (YOLO) [42] як перший ступінь до вирішення класифікації. Для класифікації подій використали такий інструмент як калькулятор подій (Event Calculus) [43]. Навчання відбувалося на датасеті SoccerNet. Точність передбачення кутового удару виявилася 68.4 % за метрикою precision і 65 % за метрикою recall, тоді як для кутового удару точність виявилася 83.33 % за метрикою precision і 25 % за метрикою recall.

Інша стаття [7] розповідає про досягнення, застосування комп'ютерного зору та труднощі у вирішеннях найближчих завдань. Наприклад, зараз активно використовуються системи трекінгу гравців, м'яча, системи аналізу рухів гравців, які допомагають тренерам, суддям, футбольним аналітикам.

Ще одна стаття[32] розповідає про сумаризацію відео на основі глибокого навчання, а саме 3-х мірні згорткові мережі та LSTM на 744 фрагментах для класифікації 5 подій: штрафний удар, кутовий удар, гольовий момент, розлучення з центру та кидок з ауту . На даних Soccer-5 алгоритм показав точність 96,81% за метрикою assigasy.

Наступна стаття [5] описує, як змінилися ключові показники гравців у футболі з часом у англійській прем'єр-лізі. Порівнювалися сезон 2006-2007

року та 2012-2013 року. Як макропоказники використовувалися дистанція, пройдена в спринті, кількість пробіжок, кількість пасів і кількість успішних пасів.

У статті [6] описано програмне забезпечення «Metrix розроблене спеціально для тренерів для відстеження ефективності гравців команди, а саме для моніторингу бігових моментів та моментів з прискоренням гравців.

Також вклад внесла стаття [25], метою якої було скласти дані про позицію на полі та швидкості, які знімалися з сенсорів на гравцях. У результаті на виході вийшли дані: відео та дані із сенсорів.

Стаття [72] пропонує вирішення завдання виявлення та визначення пози у сценах. Метод ідентифікує закриті частини тіла та прибирає неоднозначності частин тіла між людьми у безпосередній близькості один до одного. Результат експерименту виявився на рівні state-of-the-art серед усіх інших рішень.

2.6 Згорткові нейронні мережі

Згорткові нейронні мережі використовують у аналізі послідовностей зображень. Вона набула широкої популярності завдяки стійкості до змін масштабу, повороту, спотворень та іншого зміни в даних. Це досягається завдяки операції згортки, яка в одновимірному випадку виражається так:

$$y_{i,j} = \sum_p \sum_{m=-M}^M k_{m,j,p} x_{i+m,p}$$

де $y_{i,j}$ – значення нейронів на виході, j – індекс вихідного каналу, $x_{i,p}$ – значення нейронів на вході, p – індекс вхідного каналу, M – півширина ядра, $k_{m,p}$, – Коефіцієнти ядра згортки. За своєю суттю згорткові мережі перетворюють вихідні ознаки низького рівня (наприклад набори пікселів у зображенні або якась послідовність) в абстрактні ознаки високого рівня шляхом зменшення розмірності. Цього можна досягти застосуванням

операції згортки і після цього – застосуванням іншої операції до цієї згортки, наприклад взяття мінімуму або середнього.

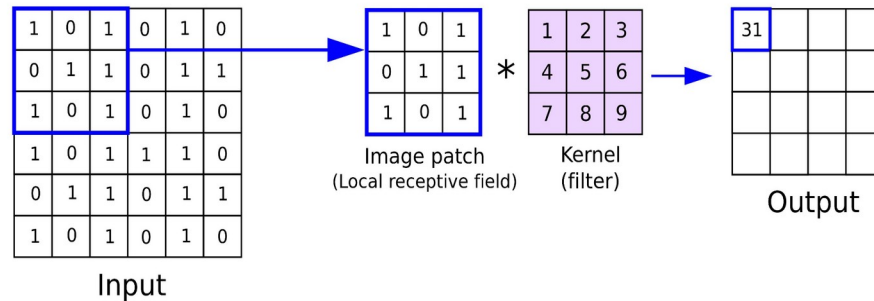


Рисунок 2.2 – Ілюстрація роботи згорткової нейронної мережі

2.7 Рекурентні нейронні мережі

Рекурентна нейронна мережа (Recurrent Neural Network) – вид нейронної мережі, де зв'язки між елементами утворюють спрямовану послідовність, тобто. значення елементів нейрона у поточному стані залежить від значення нейронів у попередньому стані [73]. Завдяки цій моделі з'явилася можливість аналізувати та обробляти серії часові ряди та інші послідовності.

Одним із простих прикладів рекурентної мережі наведено на рисунку 2.3, де на вході мережі подається послідовність векторів \vec{x}_i , а на виході – послідовність \vec{y}_i . Значення вихідних векторів \vec{y}_i обчислюються через вектори входу \vec{x}_i та вектори прихованих станів \vec{h}_i , які формуються на основі попередніх $\vec{h}_i = f(\vec{x}_i, \vec{h}_{i-1})$. Можна інтерпретувати \vec{h}_i як пам'ять мережі. Але на практиці прихований шар не може містити в собі велику кількість попередніх кроків.

Незважаючи на інтуїтивно зрозумілу архітектуру та властивості рекурентної нейронної мережі, вона має низку проблем при навчанні. Вона перестає справлятися із встановленням залежностей через великі проміжки часу. Це може виникати через проблеми зникаючих градієнтів або через

вибух градієнтів. Справа в тому, що на кожному кроці вихідний елемент \vec{y}^i виходить по суті примноженням на матрицю прихованого шару \vec{y}^{i-1} . Тому якщо послідовність виходить досить довгою, а власні числа > 1 , то можна градієнти будуть дуже швидко зростати (проблема вибуху градієнта), а якщо власні числа < 1 , то, навпаки, градієнти дуже швидко зменшуватимуться до нуля (проблема зникаючих градієнтів). Це призводить до втрати зв'язків у прихованому шарі між досить віддаленими подіями.

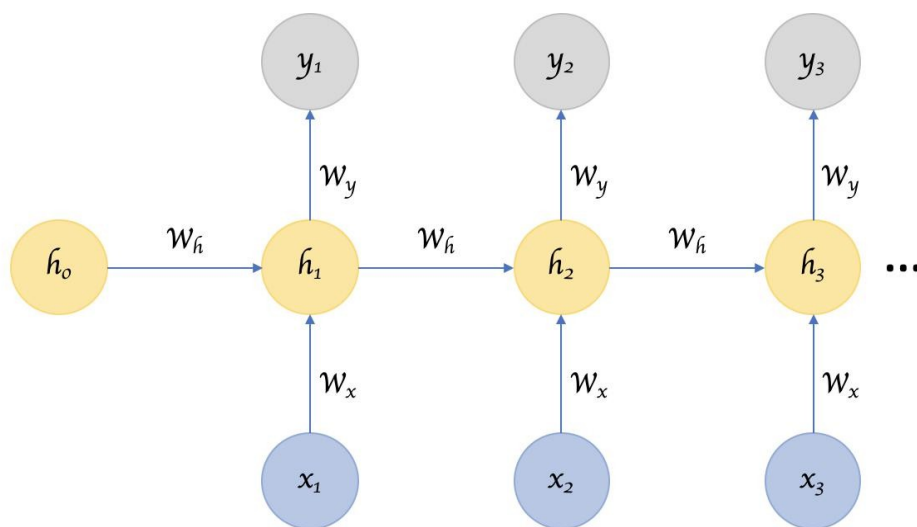


Рисунок 2.3 – Приклад роботи рекурентної нейронної мережі з одним прихованим шаром

Цю проблему вирішила спеціальна модифікація рекурентних нейронних мереж – мережа з довгою короткостроковою пам'яттю (Long Short Term Memory) [74]

Вид осередку LSTM представлений рисунку 2.4. Відмінною частиною модуля lstm від звичайної рекурентної нейронної мережі є наявність вузла стану модуля. Він і вирішує проблему згасання та вибуху градієнта.

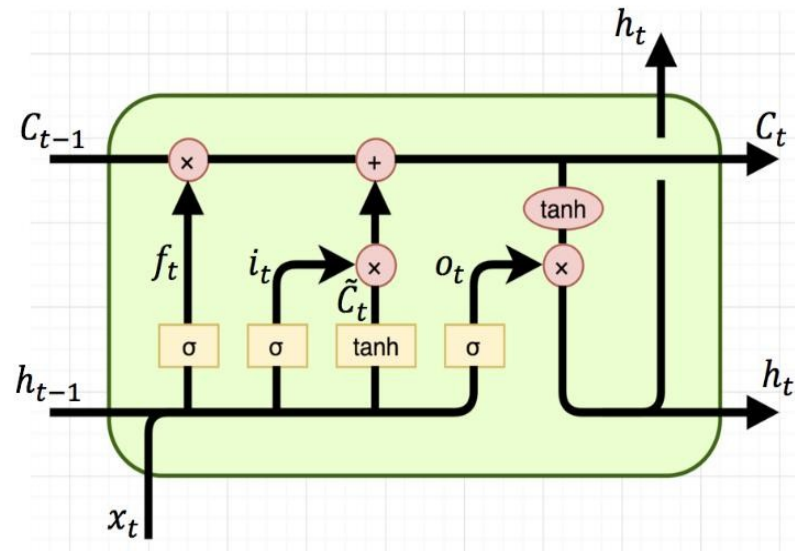


Рисунок 2.4 – Схема модуля LSTM

На відміну від традиційних рекурентних мереж вузол стану проходить по шануй прямо з входу модуля до його виходу майже минаючи примноження на матриці ваг. Входи – використовується для занулення інформації і допомагає при обробці довгих послідовностей, а – використовується для запам'ятовування нової інформації та пропускання її у вузол стану. Вихід використовується для пропускання комбінації інформації з вузла стану, елемента прихованого шару -1 і елемента вхідної послідовності.

2.8 Методи оптимізації нейронних мереж

Сама по собі нейронна мережа не має жодної цінності для задач машинного навчання. Вони починають мати значення тоді, коли їх параметри підібрані так, що за даними, що подаються на вхід моделі, модель могла наблизити передбачення моделі максимально точно до істинних передбачень. Цього якраз домагаються зміною параметрів моделі, а завдання підбором параметрів називається завдання оптимізації нейромережі. Завдання

формулюється в такий спосіб. Існує справжні відповіді y та передбачення моделі $f(x, p)$. Задається функціонал втрат:

$$Loss = \sum_{i=1}^N L(y_i, f(x_i, w))$$

де N – розмір вибірки, f – функція відповідна архітектурі нейромережі, w – набір параметрів нейромережі. Цей функціонал буде мінімальним при

$$w_{opt} = \arg \min_w \lim Loss$$

Це і буде оптимізаційним завданням. Функція втрат собою відбиває ступінь розбіжності справжніх відповідей із передбаченими і може задаватися по-різному. У задачі бінарної класифікації переважно використовується бінарна перехресна ентропія чи логістична функції втрат.

$$Logloss = -\frac{1}{N} \sum_{i=1}^N (y_i * \log(P_i) + (1 - y_i) * \log(1 - P_i))$$

, де y_i – істинна відповідь, P_i – передбачена можливість приналежності до одного з класів.

Одним із простих методів оптимізації є метод градієнтного спуску:

$$w_i = w_{i-1} - \alpha \sum_{i=1}^N \nabla_w L(y_i, f(x_i, w))$$

де α є темпом навчання. Суть методу ось у чому. Обчислюється градієнт функції втрат за ваги в точці цих ваг. Цей градієнт є напрямом якнайшвидшого зростання функції. Тому для оптимізації береться цей градієнт зі знаком мінус і цей градієнт, помножений на коефіцієнт складається зі старими вагами. Так відбувається певну кількість разів, поки метод не зійдеться. Обмеженням цього є те, що це метод здатний оптимізувати лише гладку функцію, і навіть якщо функція буде гладкою, то досягнення глобального мінімуму не гарантується.

Найефективнішою оптимізації нейронних мереж є алгоритм стохастичного градієнтного спуску (stochastic gradient descent) [75]. Навчання

проводиться шматково. Послідовність пар (x_i, y_i) у навчальній вибірці переміщується випадковим чином, після чого ваги оновлюються ітеративно для кожного шматка (батча):

$$w' = w - \alpha \nabla_w L(y_i, f(x_i, w))$$

, де i приймає значення від 0 до N.

Цей алгоритм на відміну від звичайного градієнтного спуску сходиться до глобального мінімуму при деяких обмеженнях. Існує кілька різновидів алгоритму SGD, що відрізняються параметрами, що стосуються темпу зростання.

2.9 Алгоритм детектування об'єктів YOLO

Одним із методів детектування об'єктів є метод YOLO (You Only Look Once) [42], який на даний момент перевершує інші моделі детектування об'єктів у реальному часі. Даний алгоритм заснований на згорткових нейронних мережах і відрізняється від інших алгоритмів детектування об'єктів тим, що інші алгоритми застосовують згорткові мережі кілька разів до різних частин зображення. У YOLO згорткова мережа застосовується один раз до всього зображення.

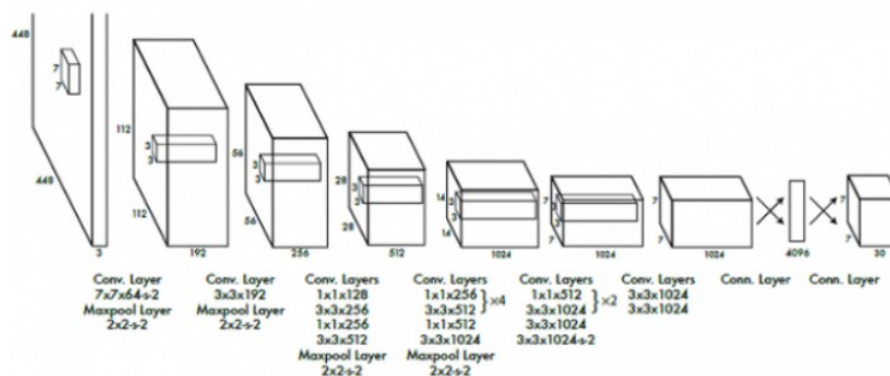


Рисунок 2.5 – Архітектура YOLO

З цього випливає те, що YOLO обробляє зображення в рази швидше за інші алгоритми. Модель поділяє зображення на сітку і прогнозом даного алгоритму буде координати об'єкта (зазвичай це прямокутник у якому даний об'єкт укладено) і ймовірність приналежності до класу об'єкта.

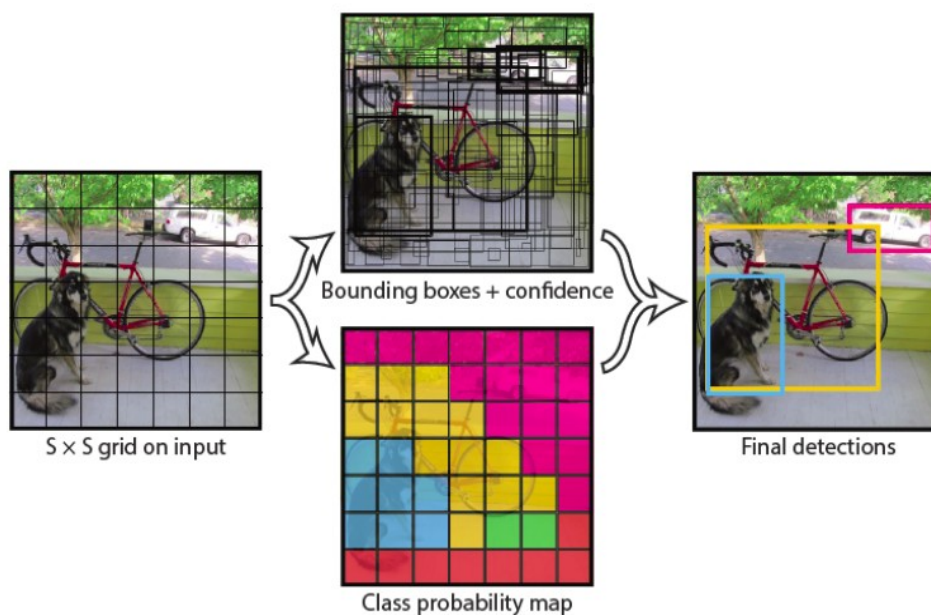


Рисунок 2.6 – Схема роботи YOLO

У задачі детектування як метрика використовується коефіцієнт перекриття вікон (IoU, Intersection over Union) для уникнення того, щоб один і той же об'єкт не визначався багаторазово. Дана модель також має високу стійкість до викидів (робастність).

2.10 Алгоритм визначення пози людини OpenPose

Одним з алгоритмів визначення 2-мірної пози людини є алгоритм OpenPose [61]. Як вхідні дані алгоритм приймає зображення персони, а на виході алгоритм видає локалізацію об'єкта (координати лівої верхньої точки координати правої нижньої точки) та вектор самого зображення, елементи якого є координатами суглобів. Може статися так, що на зображенні персони

відсутня ряд кінцівок (наприклад, якщо персону видно у профіль). Тоді ці суглоби ігноруються моделлю а як координат ставиться константне число.

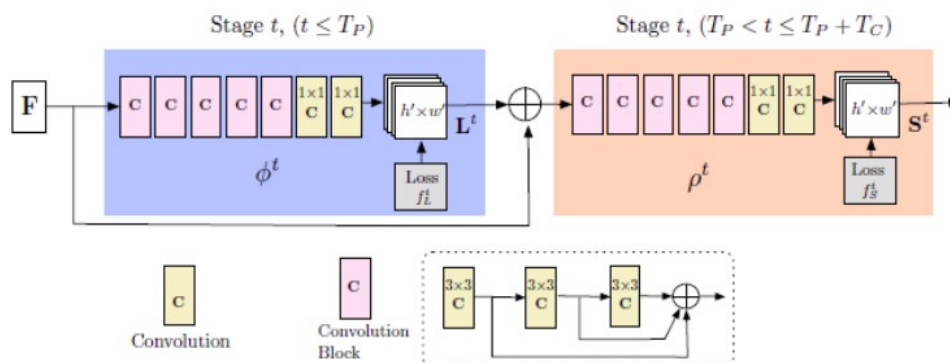


Рисунок 2.7 – Архітектура OpenPose

2.11 Побудова 2D моделей людей, які перебувають у полі зору камери

Відстеження рухів людини є важливим прикладом застосування алгоритмів машинного зору, який можна використовувати у ділових цілях. Найвідомішими прикладами подібного використання в цифровій індустрії є розподілена система відеоспостереження, цифровий маркетинг, відстеження людини у промисловому середовищі.

Вирішення цього завдання можуть відрізнятися різним ступенем деталізації. Загальним підходом до виявлення об'єкта є визначення положення людини як цілого об'єкта і наступна оцінка рамки, що обмежує її 2D або 3D.

Цікавішим підходом було б виявлення людської пози в русі. Це завдання складніше, т.к. поза людини має значно більше вимірів порівняно з рамкою, що обмежує.

Останні успіхи у глибокому навчанні призвели до появи ефективних однокадрових алгоритмів відстеження пози, таких як, наприклад, [55] [56]. Застосовуючи їх послідовно до потоку відео, можна отримати набір траєкторій для суглобів. Тим не менш, оскільки ці алгоритми, як правило,

аналізують вхідні кадри незалежно, отримані траєкторії зазвичай мають різні артефакти, такі як розриви або пропущені точки.

3 ВИЗНАЧЕННЯ ПОЗ СПОРТСМЕНІВ ЗА ДАНИМИ ВІДЕОСПОСТЕРЕЖЕННЯ

У цьому розділі ми вирішуємо завдання визначення поз спортсменів у статистиці та динаміці для кількох людей у кадрі.

Наша мета – запропонувати новий алгоритм надійного відстеження поз кількох людей у відеопотоці, використовуючи як тимчасові, так і просторові особливості даних. Для цього ми об'єднуємо прогнози, зроблені однокадровим алгоритмом виявлення людини/поз (наприклад, OpenPose та YOLO) з оцінками на основі оптичного потоку через фільтр Калмана.

Блок – схема алгоритму побудови 2D скелетних моделей людей, що знаходяться в полі зору камери, наведена на рисунку 3.1.

1. Попередня обробка відео призначена для нормалізації вихідних зображень у складі відео в термінах яскравість/контраст і сформувати значення вектора Gf для кожного кадру у відео.

2. Ідентифікація фігур людей за допомогою алгоритму YOLO дозволяє сформувати безліч прямокутних областей, що задаються координатами верхнього лівого та нижнього правого кутів.

3. Формування зменшених кадрів пов'язаних з виявленими об'єктами дозволяє перейти від одного великого зображення до множини зображень, на основі яких далі здійснюється виявлення та оцінка параметрів поз.

4. Побудова 2D скелетних моделей та оцінка їх параметрів призводить

до формування вектора $\vec{B}^{fp} = \{u_1^{fp}, v_1^{fp}, u_2^{fp}, v_2^{fp}, \dots, u_N^{fp}, v_N^{fp}\}$, де $N = 20$ відповідає числу суглобів у вибраній моделі для людської постаті.

5. Обчислення Optical Flow для виявлених точок розташування суглобів скелетних моделей використовувалися для уточнення параметрів траєкторій суглобів у складі виявлених 2D скелетів.

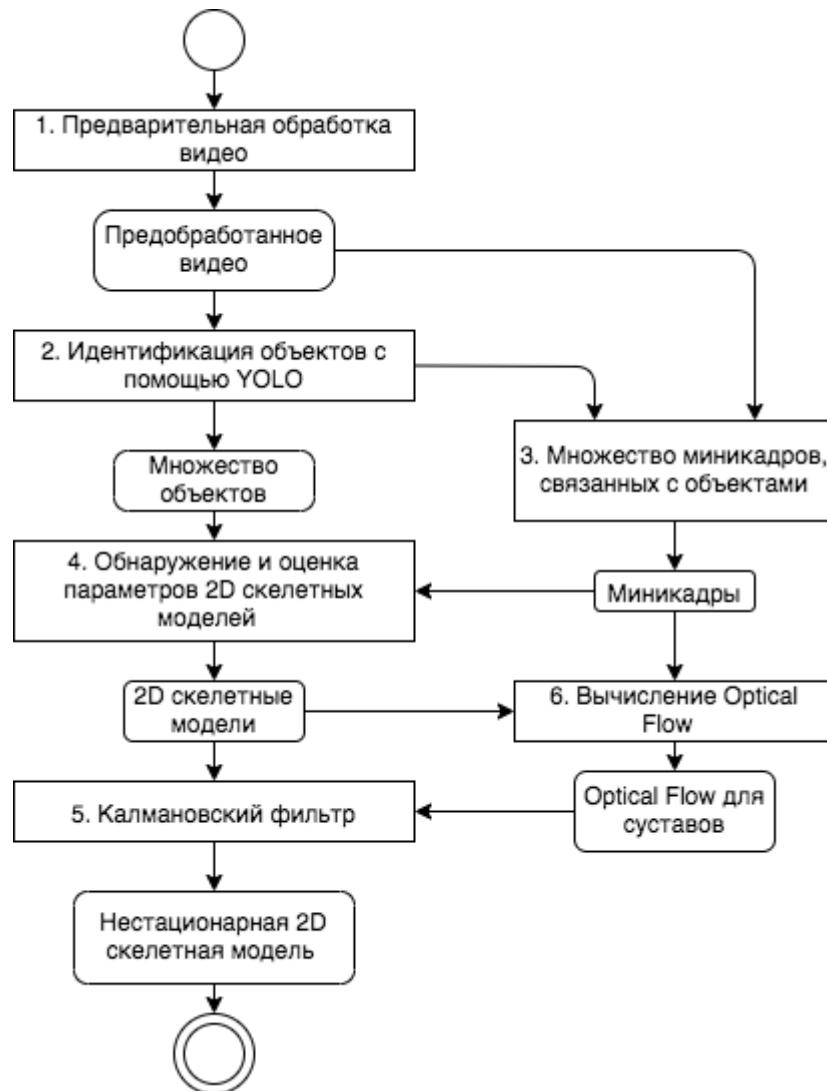


Рисунок 3.1 – Блок – схема алгоритму побудови 2D моделей людей

3.1 Попередня обробка відео

При переході до обробки відео великих форматів, класичні підходи щодо застосування неймережових методів виявлення та ідентифікації об'єктів та скелетів для багатофігурних сцен у полі зору камери перестають працювати.

Нами було запропоновано застосувати для пошуку об'єктів YOLO на зменшеному розмірі кадру. Результатом застосування YOLO є безліч прямокутників, відповідних виявленим об'єктам і заданих верхнім лівим та нижнім правим кутам у системі відліку кадру $O_f c, U_f c, V_f c$.

На стадії попередньої обробки виробляються:

- коригування параметрів яскравість/контраст для більш точної обробки.
- фільтрація об'єктів (глядачів), що знаходяться за межами футбольного поля, для того, щоб система детектування не захоплювала їх.

Варто також зазначити, що фрагменти відео відбиралися в ті моменти, коли камера нерухома або повільно повертається. Це зроблено для того, щоб не було розмиття та системи детектування та визначення пози працювали максимально точно.

Знизу на рисунках 3.2 та 3.3 представлено ілюстрацію попередньої обробки відео перед детектуванням футболістів.

Наступним кроком є етап детектування зображень футболістів для подальшої роботи з самими зображеннями футболістів. Це дозволить спростити процесу аналізу переходом від цілого кадру з великим дозволом до сукупності об'єктів (футболістів).

Як алгоритм детектування футболістів було взято YOLO[42]. Він працює набагато швидше за інші алгоритми детектування, що є досить важливим фактором через те, що витрачений час на детектування пропорційно кількості кадрів. Також цей алгоритм досить стабільний у нашому завданні і майже завжди здатний детектувати всіх гравців на поле з високим ступенем точності. Ілюстрація роботи YOLO представлена на рисунку 3.4.

Таким чином, завдання аналізу цілого кадру зводиться до аналізу сукупності зображень футболістів на цьому кадрі.



Рисунок 3.2 – Кадр фрагмента відео

$$f : \bigcup_{\substack{i \in [0, N] \\ j \in [0, K]}} \{A_{ij}\} \mapsto M_{ij}, M = 0, 1; A_{ij} = \bigcup_{\substack{x \in [0, X] \\ y \in [0, Y]}} \{A_{xy}(i, j)\}$$

, де N – кількість кадрів, K – кількість гравців на кадрі i , X – детектована область j -го футболіста в i -му кадрі по горизонталі, Y – детектована область j -го футболіста в i -у кадрі по вертикалі.

Після виконання необхідної попередньої обробки відео, вибір якої залежить від специфічних характеристик відеообладнання, використаного при його отриманні, наступним кроком є вилучення скелетних моделей фігур людей з одиночних кадрів у всіх випадках, коли це можливо. Як інструмент для детектування двовимірних скелетних моделей нами було обрано рішення OpenPose. Це рішення забезпечує детектування суглобів на зображенні, що містить кілька фігур людей і вперше було представлено в роботі [61]. Нами використано рішення з відкритим кодом, що забезпечує формування скелетної моделі у форматі BODY-25. Спрощена модель, що містить 18 точок, зображена на рис.3.5.

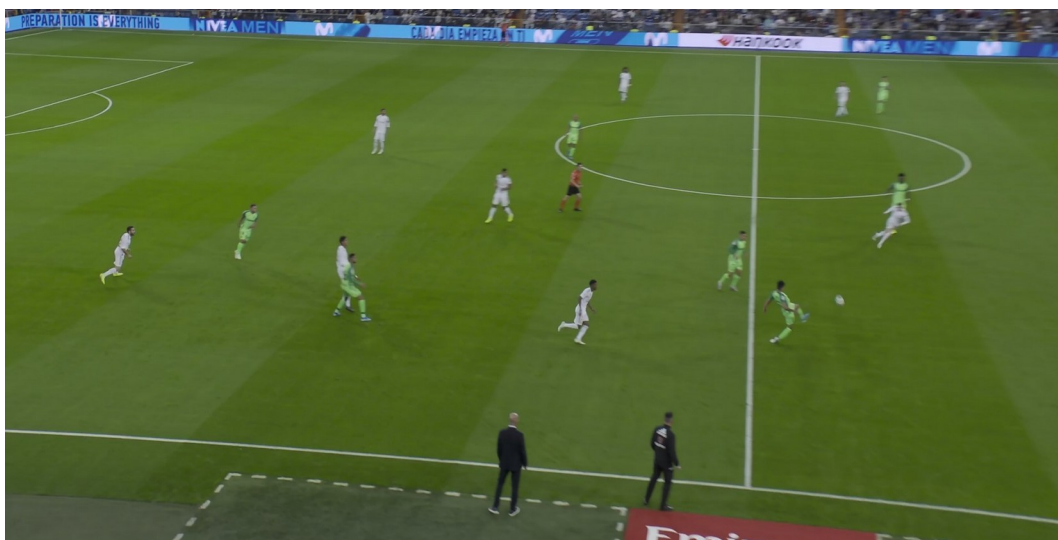


Рисунок 3.3 – Передобрацьований кадр

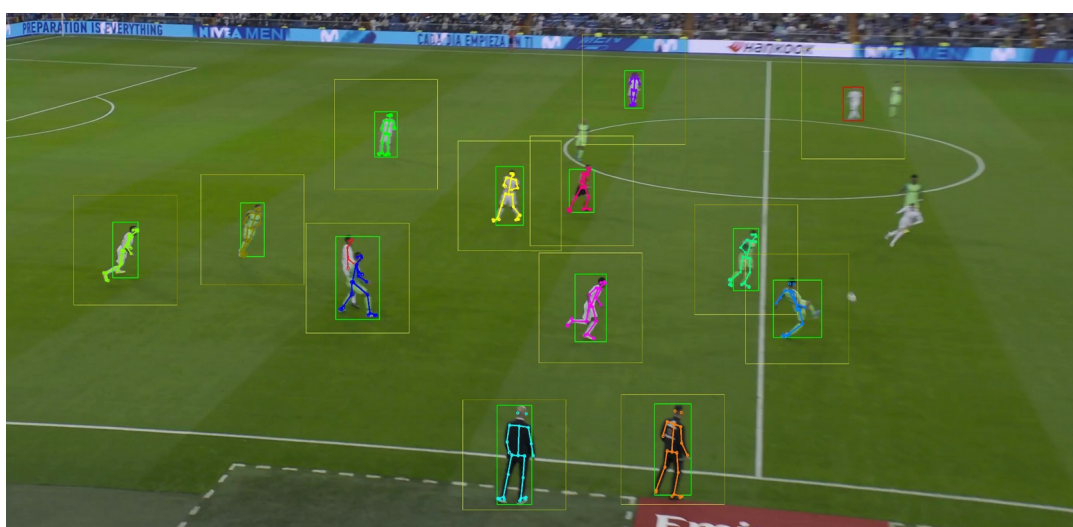


Рисунок 3.4 – Кадр, оброблений за допомогою YOLO та OpenPose

Однак необхідно помітити, що спроба прямого застосування бібліотеки OpenPose до обробки кадрів з відео високої роздільної здатності 4K формату не призводить до успіху. Це пов'язано з тим, що вимоги до алго ритмом пам'яті лінійно пов'язані з площею оброблюваного зображення і становить приблизно 3Gb для зображення розміром 656x656. Таким чином для обробки зображень 4K формату буде потрібно розподілений вичілювальний кластер та оптимізація алгоритму для проведення обчислень на такому кластері.

Нами було запропоновано рішення, що полягає в розбитті зображення на безліч частково перетинаються областей та застосування детектора до

кожної з них з подальшим об'єднанням отриманих векторних скелетних моделей.

Таким чином, ми можемо суттєво підвищити ефективність роботи алгоритму в тому випадку, коли кількість людей у полі зору камери відносно невелика, що часто спостерігається у штатному режимі роботи систем відеоспостереження.

Ми застосували YOLO (You Only Look Once) [42] і з його допомогою вирішили задачу пошуку фігур у кадрі 4K роздільної здатності. Також було виявлено, що правильне застосування виявлення об'єктів за допомогою YOLO дозволило усунути частину помилково – позитивних помилок у роботі OpenPose.

Для кожного кадру у вихідній послідовності запропоноване рішення спочатку застосовує детектор об'єктів, навчений на навчальній вибірці COCO [76].

Результатом роботи YOLO є набір прямокутних меж кожного з детектованих об'єктів, верхній лівий та нижній правий кути кожної з областей при цьому задаються координатами у системі відліку a, Ofc, Ufc, Vfc .

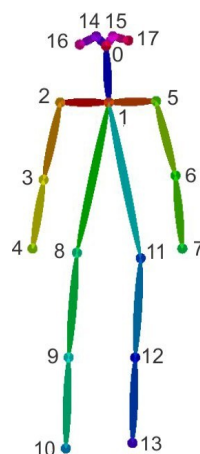


Рисунок 3.5 – Суглоби скелетної моделі людини, що використовуються в OpenPose

Для цих виявлених прямокутних областей формуються покривають їх квадратні області з роздільною здатністю, оптимальним для застосування бібліотеки Open Pose.

Результатом стадії детектування є таким чином безліч прямокутних областей кожного з виявлених об'єктів. Кожен такий об'єкт задається координатами 1,1,2,2, достовірністю виявлення та безліччю суглобів, виявлених алгоритмом OpenPose. Сформований таким чином скелет відповідає прямокутному об'єкту, виявленому алгоритмом YOLO.

Додатково ми обчислюємо оцінку похибки для кожного суглоба скелетної моделі шляхом інтегрування теплової картки повернутої OpenPose. Пізніше саме ці похибки будуть використані для оцінки грішностей при калманівській фільтрації.

3.2 Оцінка параметрів траєкторій 2D скелетів

Після детектування футболістів та переходу до іншого формулювання завдання наступним завданням є визначення 2-мірної пози. Це знизить розмірність даних, тим самим спростить процес обробки даних і прискорить навчання моделі та покращить її якість. І тому використовувався алгоритм OpenPose [72]. В результаті після обробки кожному детектованого футболіста відповідатиме 25-мірний вектор \vec{x} , елементами якого будуть координатами суглоба. Звичайно, можуть виникнути ситуації, коли суглоб закритий тілом футболіста і система визначення пози може не розпізнати суглоб. Тоді цьому суглобу ставиться координата (-1, -1) і означатиме відсутність суглоба. Щоб це врахувати, було вирішено додати до координат 3 елемент, який буде пов'язаний з фактором відсутності суглоба:

$$z = \begin{cases} 0, & x = -1 \\ 1, & x > 0 \end{cases}$$

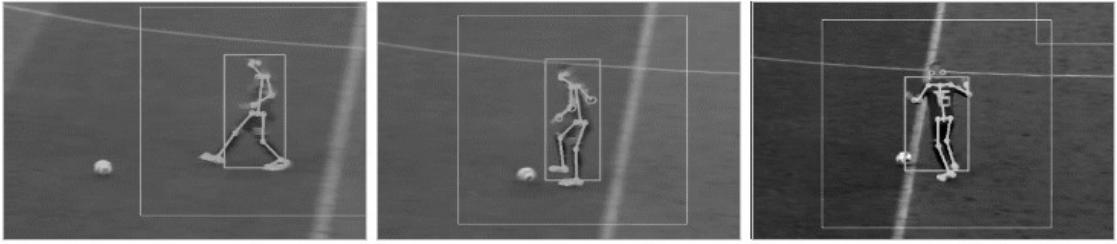


Рисунок 3.6 – Ілюстрація роботи YOLO та OpenPose на зображенні одного футболіста

Після обробки всього фрагмента відео за допомогою YOLO та OpenPose кожному детектованому футболісту відповідає послідовність векторів, довжина якої відповідає числу кадрів, у яких було детектировано зображення футболіста.

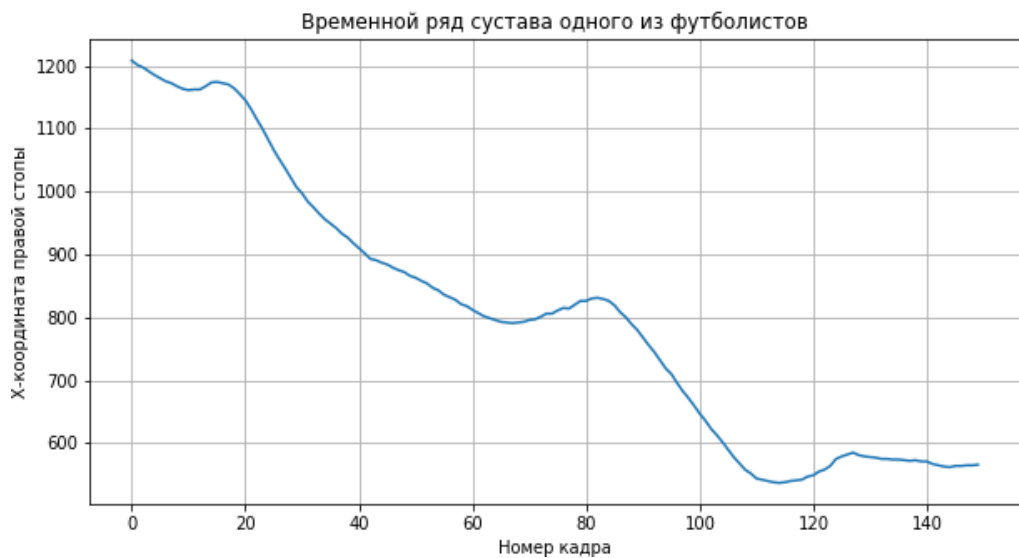


Рисунок 3.7 — Приклад тимчасової послідовності одного із суглобів футболіста

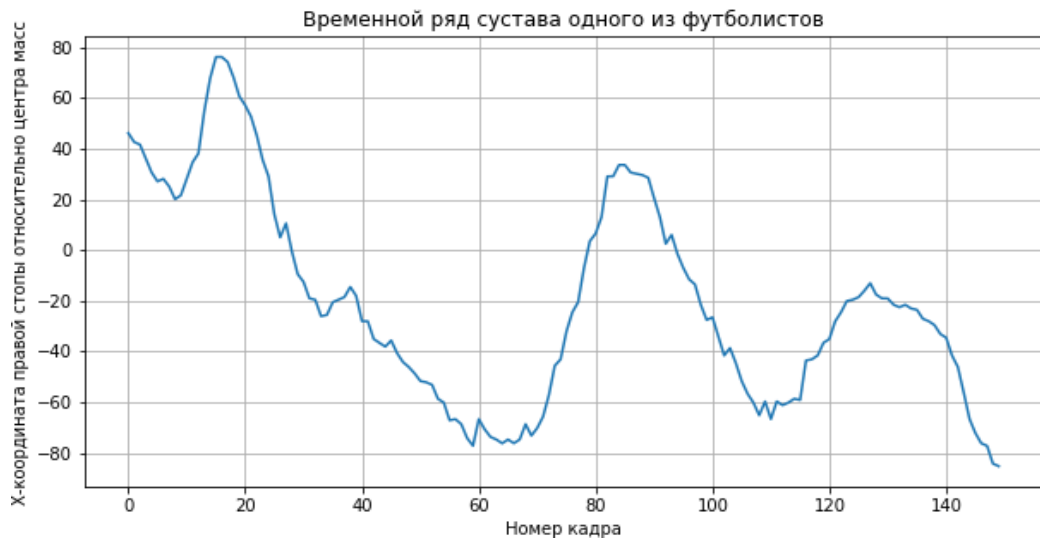


Рисунок 3.8 — Приклад тимчасової послідовності одного із суглобів футболіста щодо центру мас

Один із прикладів результату роботи YOLO та OpenPose на прикладі одного з суглобів представлений на рисунку 3.7

Далі було здійснено перехід до центру мас футболіста для поліпшення точності класифікації, результат якого представлено рисунку 3.8.

Таким чином, завдання аналізу послідовностей зображень футболістів зводиться до аналізу послідовностей векторів поз футболістів.

$$f : \bigcup_{\substack{i \in [0, N] \\ j \in [0, K]}} \{\vec{P}_{ij}\} \mapsto M_{ij}, M_{ij} = 0, 1$$

де $P_{ij} = ((x_0(i, j), y_0(i, j)))^T$, N – кількість кадрів у відеофрагменті, K – кількість футболістів.

3.3 Формування навчальної вибірки

Після формування послідовності векторів кожного виявленого футболіста наступним кроком є розмітка даних. Інтерес представляють футболісти, які саме віддали пас. Тому тільки вони будуть використовуватись для аналізу. Розмітка проводилася наступним чином: на

тимчасовому відрізку, який відповідає часу появи футболіста у фрагменті відео, кожному кадру присвоювалося число, яке дорівнювало 0 якщо цей кадр не є кадром у момент пасу, і дорівнювало 1 в іншому випадку. Моментом пасу був час самого дотику з м'ячем (0.1 сек.) та його околиця (± 0.2 сек.), що відповідає підготовці до пасу та переходу скелета у стан перед пасом. Приклад моменту пасу наведено рисунку 3.9

Під час розмітки було розмічено 55 футболістів, котрі віддавали паси ногою.

Наступним кроком було формування послідовностей для навчання. Це необхідно для того, щоб на вхід модель надходили послідовності однакової довжини. Для цього було обрано метод ковзного вікна. Цей метод є досить популярним у завданнях обробки послідовностей. За допомогою нього послідовність розбивається на відрізки фіксованої довжини – вікна. Вікно було обрано довжиною $L = 80$ кадрів і кроком 5 кадрів. Через війну формувався тензор розмірністю $(N, 80, 25, 3)$, де N – кількість послідовностей, отриманих шляхом ковзного вікна. Варто також зазначити, що послідовності, в яких розмічені класи починалися з одиниць, а не з нулів виключалися з аналізу через те, що модель таких даних навчається некоректно. Приклад роботи методу ковзного вікна представлено рисунку 3.10

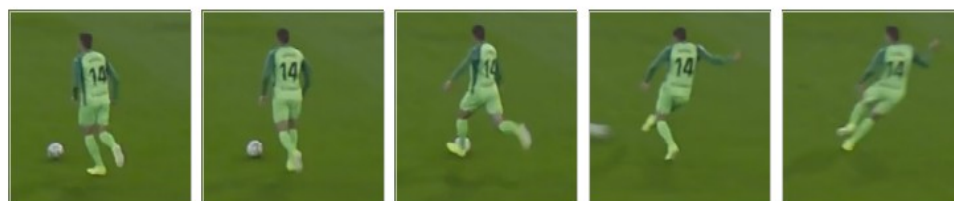


Рисунок 3.9 – Приклад моменту передачі м'яча

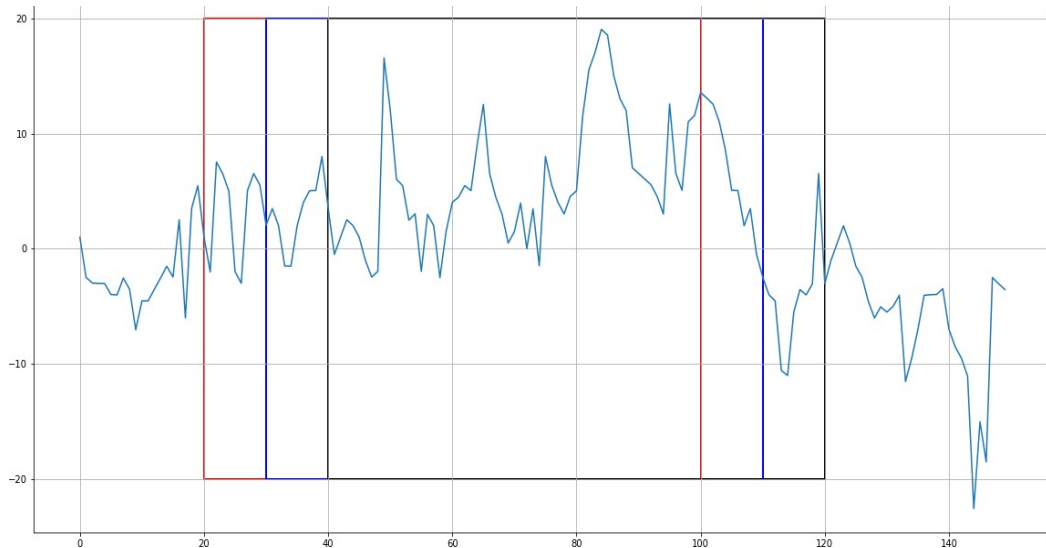


Рисунок 3.10 – Ілюстрація роботи ковзного вікна

В результаті було зібрано 7818 послідовностей із балансом класів 0.92:0.08 (негативний до позитивного). Таким чином, задача зведена від безперервного передбачення до завдання передбачення послідовностей.

3.4 Використання Optical Flow

Ще одним запропонованим новим підходом для покращення точності скелетних моделей сформованих на основі окремих кадрів є використання optical flow (оптичного потоку). Optical Flow є технологією, що широко використовується в даний час для вирішення широкого спектру завдань, пов'язаних з обробкою відео. Основна ідея полягає в тому, щоб відстежувати переміщення окремих точок на зображенні між двома послідовними кадрами у відеопослідовності.

Як правило, оптичний потік описується 2D векторним полем $\{v_{flow} = (dx, dy) T\}$ для кожної точки n -го кадру $In(x, y)$. Відповідний піксель $n+1$ го кадру описується усуненням $In+m(x + dx, y + dy)$.

Для обчислення optical flow у літературі описано безліч підходів. У нашій роботі ми використовували похід, запропонований у роботі [77], що отримав назву Dense Inverse Search (Пошук зворотної густини). Він

відноситься до безлічі алгоритмів «щільного оптичного потоку» і відрізняється низькою обчислювальною складністю при збереженні хороших показників при проведенні стандартних тестів оптичного потоку. Шляхом застосування цього алгоритму до кожного кадру відеопослідовності і взявши точкове значення швидкості струму в точках, які алгоритм Open Pose визначив, як суглоби скелетної моделі, ми, таким чином, отримуємо ще один опосередкований спосіб вимірювання швидкості руху суглобів двовимірної скелетної моделі.

Однією із завдань, яку ми поставили є завдання побудови індивідуальних траєкторій кожного суглоба двовимірної скелетної моделі з метою уточнення алгоритму віднесення точки, до якої прив'язаний сустав у – м кадрі до того чи іншого двовимірного скелета.

Це завдання частково розглядалося на роботах [78; 79], які були взяті за основу в нашій роботі.

Приклад кадру з трансляції футбольного матчу, використаної для порівняльного розрахунку Optical Flow з використанням згаданих вище алгоритмів, наведено на рисунку 3.11.

Для візуального представлення результатів обчислення Optical Flow ми використали колірну модель HSV. Відтінок (Hue) визначає напрямок вектора руху, а насиченість (Saturation) визначає модуль вектора швидкості.

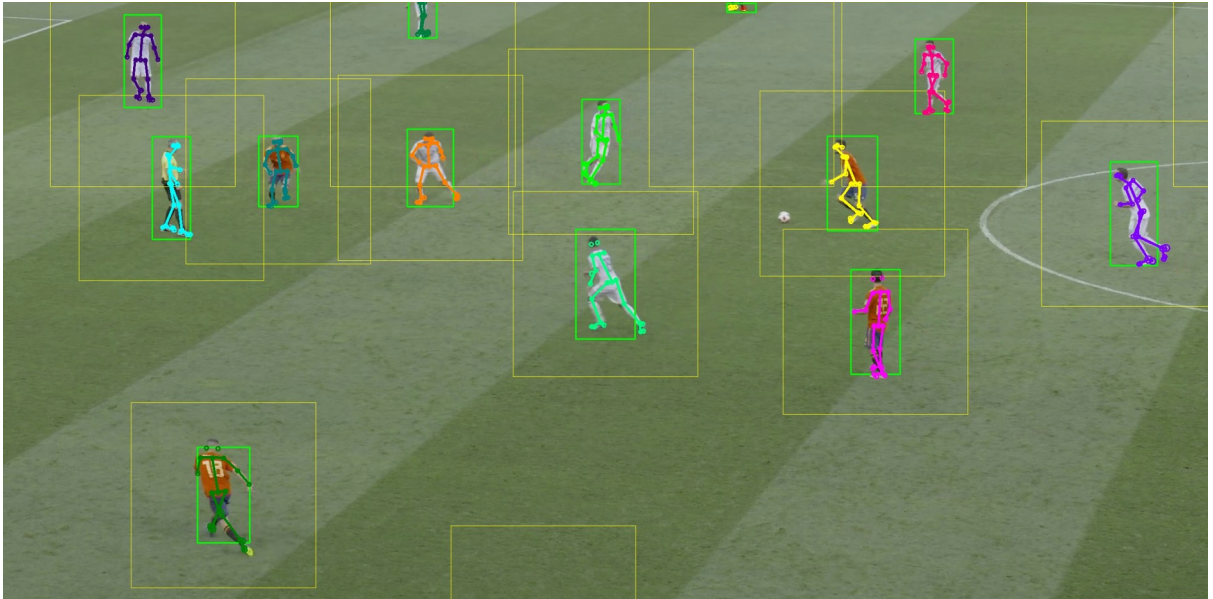


Рисунок 3.11 – Приклад кадру з трансляції футбольного матчу, використаної для порівняльного розрахунку Optical Flow з використанням двох різних алгоритмів

Результат роботи алгоритму Dense-Inverse-Search для обчислення Optical Flow наведено рисунку 3.12.

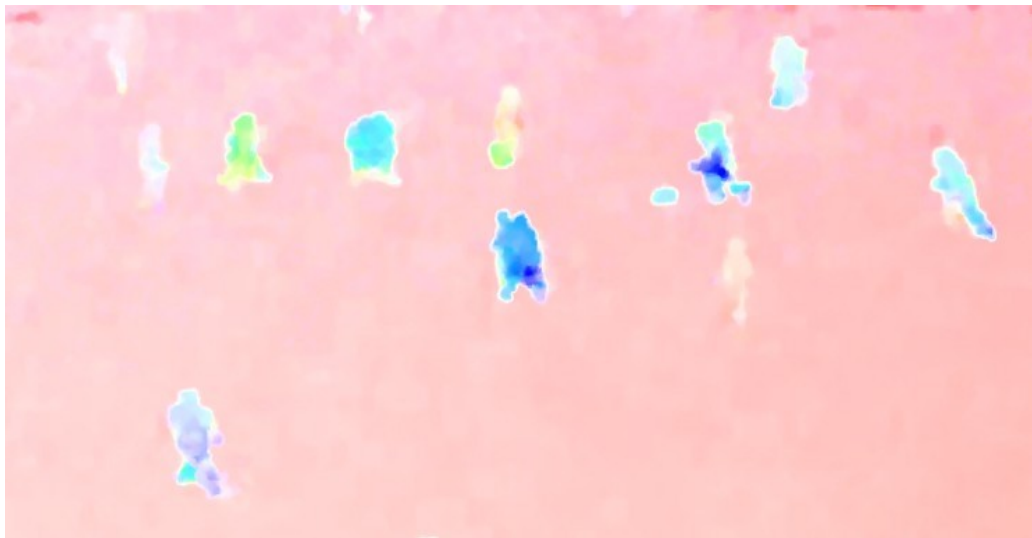


Рисунок 3.12 – Приклад Optical Flow, розрахований за допомогою алгоритму Dense-Inverse-Search

Результат використання алгоритму DeepFlow до розрахунку Optical Flow представлений рисунку 3.13.

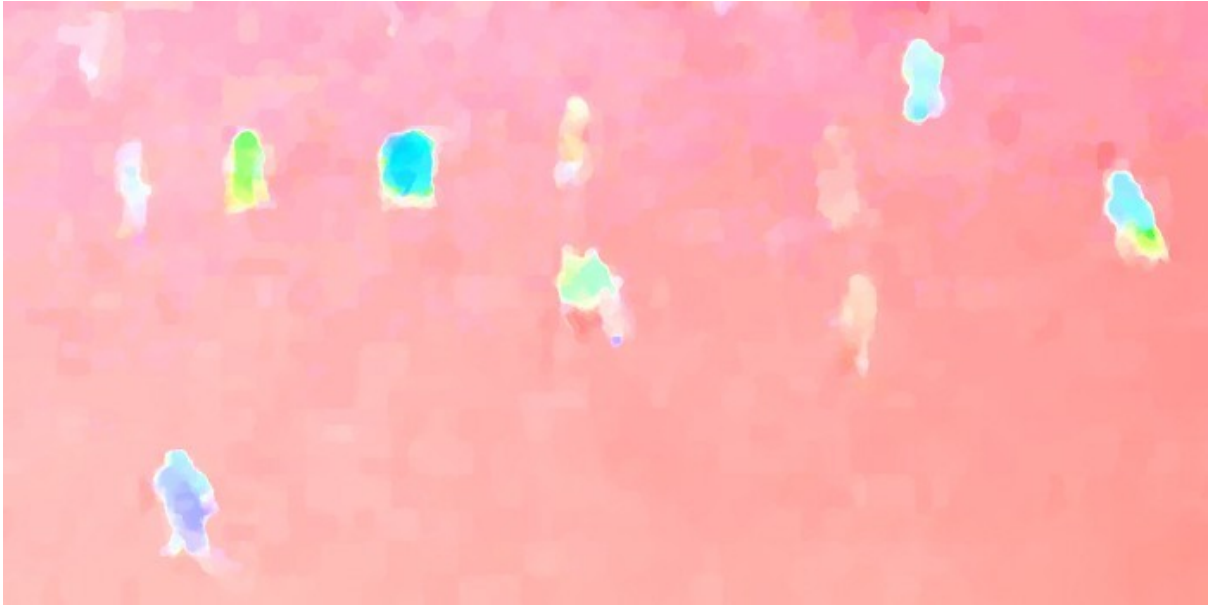


Рисунок 3.13 – Результат використання алгоритму DeepFlow для розрахунку Optical Flow

Шляхом використання цих алгоритмів для кожного кадру відеопослідовності ми отримуємо значення швидкості у двовимірному просторі для кожного суглоба, виявленого алгоритмом Open Pose. Фактично ми отримуємо альтернативне джерело інформації про швидкість в особливих точках зображення.

При цьому одночасно ми розв'язуємо задачу побудови траєкторій руху векторних скелетних моделей для кожного з людей, які перебувають у кадрі. Для цього в найпростішому випадку можна використовувати результати, описані в роботах [78] та [79].

3.5 Застосування фільтра Калмана для згладжування та уточнення траєкторій 2D скелетів

Для того, щоб виконати спільну обробку оцінок розташування суглобів двовимірної скелетної моделі, отриманих незалежно алгоритмами OpenPose і Optical Flow, нами був використаний фільтр Калмана.

Фільтр обробки тимчасових послідовностей, який отримав ім'я свого творця був запропонований в 1960 році в статті [80]. Методично вивіреним опис застосування фільтра Калмана можна знайти у пізніших статтях та книгах (див., наприклад [81; 82]).

Застосування калманівської фільтрації передбачає наявність моделі руху модельованого об'єкта. Існує кілька загальноприйнятих підходів

до побудови моделі у разі коли рух неформалізується або формалізується складним неявним чином (див., приклад [83]).

Популярним підходом для двовимірного випадку є побудова моделі, що описує рух з постійною швидкістю з накладеною на нього нейронною моделлю в полярних координатах, що використовує розширений фільтр Калмана [78].

Альтернативним і значно складнішим підходом є побудова тривимірної скелетної моделі та фільтрація руху в тривимірному просторі на основі гіпотез про динаміку руху суглобів у тривимірному просторі.

Для наших чисельних експериментів ми використали постійну декартову швидкість, незалежно оцінювану для кожного із суглобів моделі у системі відліку, пов'язаної з кадром. В цьому випадку вектор стану включає

Для наших чисельних експериментів ми використовували постійну швидкість декарту, незалежно оцінювану для кожного з суглобів моделі в системі відліку, пов'язаної з кадром. В цьому випадку вектор стану включає 4 компоненти, що представляють оцінку положення і швидкості суглобів скелетної моделі: $s(t) = (\hat{u}t, \hat{v}t, \hat{u}'t, \hat{v}'t) T$. Стан у момент часу $t + 1$ може бути пов'язаний зі станом у момент часу t за допомогою наступного виразу:

$$s(t + 1) = \begin{pmatrix} 1 & 0 & \delta t & 0 \\ 0 & 1 & 0 & \delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \hat{u}_t \\ \hat{v}_t \\ \hat{u}'_t \\ \hat{v}'_t \end{pmatrix} + \begin{pmatrix} 0.5\delta t^2 & 0 \\ 0 & 0.5\delta t^2 \\ \delta t & 0 \\ 0 & \delta t \end{pmatrix} \begin{pmatrix} \nu_x(t) \\ \nu_y(t) \end{pmatrix}$$

з коваріаційною матрицею шумів процесу:

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \bar{a}_u & 0 \\ 0 & 0 & 0 & \bar{a}_v \end{pmatrix}$$

Для кожного проведеного чисельного експерименту ми будували дві траєкторії кожного суглоба. Першу з урахуванням незалежного виміру швидкості на основі optical flow та другу без його обліку.

Для отримання оцінки обчислювальної складності та якості одержуваного результату в різних додатках, нами був підготовлений набір відео фрагментів для подальшої обробки:

- фрагменти трансляції футбольного матчу в 4k роздільній здатності;
- відео із системи відеоспостереження супермаркету;
- відео із зовнішньої системи спостереження.

Для трансляції футбольного матчу в рамках додаткової попередньої обробки було проведено відрізання трибун із уболівальниками.

На рисунку 3.14 представлені сформовані двовимірні скелети для прикладу футбольної трансляції та системи відеоспостереження супермаркету.

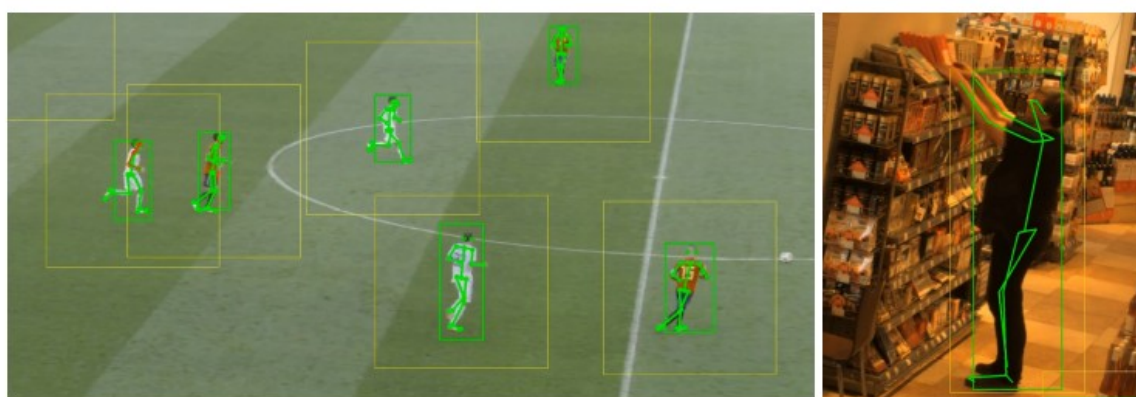


Рисунок 3.14 – Отримані двовимірні скелети для футбольної трансляції та системи відеоспостереження супермаркету. Жовті прямокутники відповідають об'єктам, виявленим OpenPose. Зелені прямокутники відповідають об'єктам, які виявлені за допомогою YOLO.

Приклади результатів фільтрації траєкторій наведено на рисунках 3.15 та 3.16.

На рисунку 3.15 а) представлений графік координати траєкторії голеноstopного суглоба людини, що рухається в поле зору зовнішньої камери на дотримання. Періодично виникає неузгодженість між виміряною і відфільтрованим значенням пояснюється тим, що модель фільтра Калмана не включала особливостей руху людини, коли на кожному кроці при дотику стопи з землею напрям швидкості голеноstopного суглоба змінювалося на протилежне.

На рисунку 3.15 б) представлений графік координати траєкторії колінного суглоба людини, що рухається в поле зору зовнішньої камери спостереження. Короткі сплески шуму викликані помилками виявлення OpenPose, що виникають через суміщення колінного суглоба з іншими суглобами під час руху.

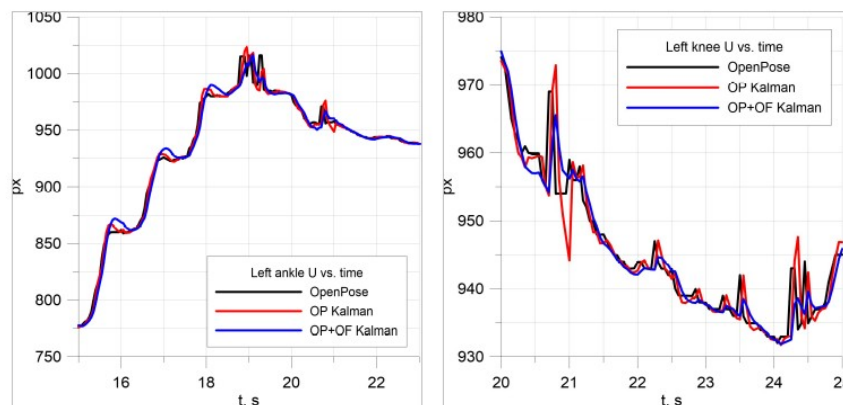


Рисунок 3.15 — а) Координата U виявленої та відфільтрованої траєкторії голеноstopного суглоба людини, що рухається, у полі зору зовнішньої камери відеоспостереження б) Координата U виявленої та відфільтрованої траєкторії колінного суглоба людини, що рухається, у полі зору зовнішньої камери

На рисунку 3.16 а) представлений графік координати U вимірної і від фільтрованої траєкторії зап'ястя людини з частковою видимістю, коли в деякі моменти часу OpenPose втрачає цей суглоб. Грубо спрямлені лінії відповідають моментам, коли суглоб виявлявся невидимим, в результаті чого OpenPose не давав позначки про місцезнаходження суглоба і фільтр просто забезпечував формування нових позначок виходячи з гіпотези про рівномірний і прямолінійний рух.

На рисунку 3.16 б) представлений графік координати U вимірної та від фільтрованої траєкторії ліктя футболіста.

Слід зазначити, що фільтр, який використовує для уточнення траєкторій вимірювання Optical Flow, демонструє більш стабільну поведінку для деяких прикладів відео, що містять помилки оцінки координат суглобів за допомогою алгоритму OpenPose. Щоб оцінити цей ефект кількісно, ми запровадили показник якості.

Щоб підтвердити коректність застосування фільтра Калмана у відсутність об'єктивно отриманої контрольної траєкторії, ми надійшли наступним чином. Були обчислені величини неузгодженостей та проведено перевірку, чи є їх розподіл нормальним із нульовим середнім значенням та постійною дисперсією. Такий підхід використовувався, наприклад, [84].

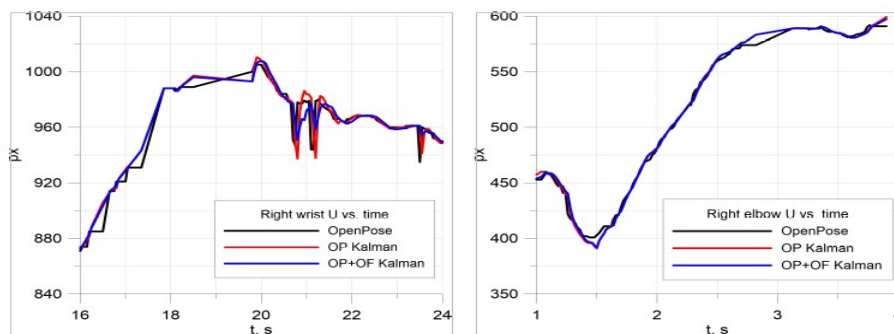


Рисунок 3.16 — а) Графік координати U вимірної та відфільтрованої траєкторії зап'ястя людини б) Графік координати U вимірної та відфільтрованої траєкторії ліктя футболіста.

Необхідно визнати, що в нашому випадку модель Калмана є дуже наближеною, і цей метод не можна використовувати для надійної оцінки впливу вимірювань швидкості. Натомість ми обчислюємо простіші метрики, виходячи з досить інтуїтивної ідеї – прогнозовані значення, що надаються «кращим» фільтром, повинні в середньому менше відрізнятися від фактичних виміряних значень. Це може призвести до неправильного результату, якщо помилка прогнозу може корелювати з помилкою вимірювання на наступному кроці. Однак у нашому випадку у досліджуваного виміру швидкості джерело помилок відрізняється від помилок вимірів, що забезпечуються OpenPose.

З урахуванням наведених міркувань, ми розрахували середньоквадратичну різницю між прогнозованими та виміряними координатами для фільтра Калмана з вимірюваннями швидкості оптичного потоку та без них

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N |\vec{B}^{fp} - \widehat{\vec{B}^{fp}}|^2$$

де прогноз значення на виході фільтра Калмана виконується перед будівництвом вимірювання на цьому етапі. Сума береться по всіх виявлених ключових точках знайдених у відео та віднесених до обраної траєкторії.

4 ПОБУДОВА ТРИВИМІРНОЇ СЦЕНИ ТА ФОРМУВАННЯ ПОЗ СПОРТСМЕНІВ У ТРИВИМІРНОМУ ПРОСТОРИ

4.1 Алгоритм тривимірної прив'язки гравців до сцени

Алгоритм побудови сцени виглядає так:

У кадрі, що отримується з нерухомої камери C , фіксується опорна площина P .

Далі на площині задається перспективне перетворення, здатне переводити координати з системи відліку в двовимірну систему відліку, заданої на площині. Оскільки на площині система відліку може бути обрана будь-який, ми поставимо її так: точка з координатами $(0,0)$ відповідає проєкції координати розташування камери на площину P .

Осі обрані таким чином:

O – вісь, паралельна орієнтації камери,

O – вісь, перпендикулярна осі O , що йде «зліва направо» при веденні відліку від камери.

Далі для системи відліку, пов'язаної з площиною P можна знайти калібрувальні коефіцієнти для визначення відстані від камери (або від її проєкції на площині P) до об'єктів на площині P у вибраних одиницях виміру.

Далі в кадрі, що отримується з камери C , за допомогою нейронної мережі (YOLO) детектуються люди – кожній знайденій людині відповідає описаний навколо нього прямокутник (bounding box), який відповідає габаритним розмірам знайдених гравців у системі відліку UV . Далі на зображенні формується областей (ROI), де – це кількість знайдених людей. Кожна область подається на вхід алгоритму OpenPose, який видає як вихід скелетні моделі для знайдених гравців.

На рисунку 4.1 алгоритм виявив трьох гравців і далі алгоритм OpenPose для них побудовані скелетні моделі.

Для знайдених скелетних моделей визначаються координати в системі відліку, пов'язаної з площиною.

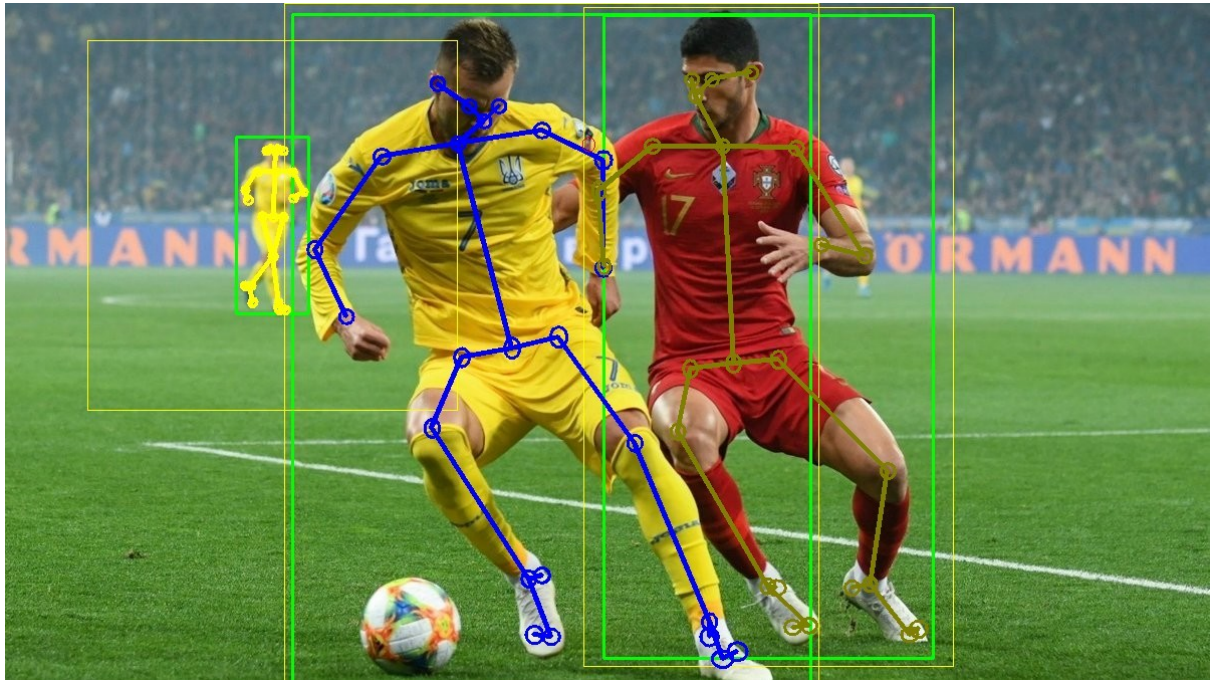


Рисунок 4.1 – Визначення ROI для виявлених фігур гравців

Для цього використовуються координати скелетної моделі в системі відліку UV , а також апріорне припущення, що знайдені в кадрі гравці знаходяться на площині P . Це припущення необхідне при перерахунку ординат знаходження цих людей із системи відліку, пов'язаної з кадром, до системи відліку, пов'язаної з площиною.

Для знайдених на поточному кадрі людей перевіряється, чи були ці люди вже раніше (на попередніх кадрах) виявлені. Якщо так, то отримані координати додаються в масив, що містить траєкторію людини з поточним ключем. Якщо ні – створюється новий унікальний ключ і масив, йому відповідний. Цей ключ фактично описує нового гравця у полі зору камери.

У цей масив записується отримана координата. Для порівняння нами використали угорський алгоритм. Далі, аналізуючи траєкторію, робляться

висновки про характер руху (біг/ходьба тощо). Дані висновки використовуються для побудови траєкторії руху даної людини у трьох мірних системах відліку, пов'язаної з площиною P (третя вісь відраховується від точки $(0,0)$ у напрямку до камери перпендикулярно площині P). Як точки, для якої буде побудована 3D траєкторія запропоновано вибрати координату тазового суглоба в скелетній моделі.

Як характерна точка, що описує тривимірний рух людини як матеріальну точку, можна використовувати дві точки:

- центральну точку прямокутника, описаного навколо знайденого у кадрі гравця;
- точку, що відповідає положенню тазового суглоба гравця у кадрі.

Перший підхід має перевагу перед другим у швидкодії, оскільки для його реалізації не потрібні додаткові алгоритми, потрібні виключно вихідні дані алгоритму YOLO. Нестача підходу полягає в тому, що алгоритм нестійкий до різних змін поз людини. Другий підхід вимагає запровадження алгоритму OpenPose для послідовного аналізу всіх знайдених у кадрі людей визначення координат їх суглобів. Таким чином, швидкодія залежатиме від кількості знайдених у кадрі людей. Тим не менш, у такому випадку нам необхідний мінімальний набір апріорних знань, оскільки виведенням алгоритму буде координата необхідного

Для оцінки висоти тазового суглоба над площиною P пропонується використати алгоритм, описаний у [85]. Цей алгоритм вимагає мінімальну кількість апріорних даних (в даному випадку потрібна одна вимога: координата таза по X, Y повинна співпадати з координатами деякої точки на площині P і ці координати можуть бути в будь-який час отримані. Такою точкою може бути, наприклад, центр мас двох нижніх суглобів).

Знаходження висоти тазового суглоба над площиною P є, по суті, достатньою умовою для отримання 3D траєкторії руху людей у заданій сцені.

Було проведено оцінку швидкодії запропонованого підходу з урахуванням класифікатора YOLO. Для обробки одного кадру роздільною здатністю 408x408 на відеокарті Nvidia RTX 2060 Super знадобилося 20мс.

Тести на продуктивність проводилися за допомогою бібліотеки OpenCV (модуль Deep Neural Network).

Запропонований підхід відмінно працює для ігрових видів спорту, де є апріорна інформація про те, що майданчик є горизонтальним та плоским.

Для деяких прикладних застосувань запропонований підхід буде досить складно застосовний. Як приклад такого застосування можна навести завдання відстеження людей на ескалаторах. Приклад зображень з камер, що відстежують поведінку людей на стрічці ескалатора аеропорту Шереметьєво, наведено на рисунках 4.2. Визначення прив'язки фігур людей до точки у тривимірному просторі ускладнюється рухом ескалатора, що призводить до зміни точки перетину вертикальної проекції з поверхнею в часі та наведені вище міркування перестають працювати.



Рисунок 4.2 – Відстеження людей на стрічці ескалатора

Також проблеми можуть виникати під час верхнього розташування камери. Точка проекції суглоба на горизонтальну площину при цьому захована під самим об'єктом і алгоритм прив'язки об'єкта до сцени в цьому випадку має дещо відрізнятись. На рисунку 4.3 наведено приклад такого ракурсу.

Важливо відзначити що у разі верхнього розташування камери ми хоч і втрачаємо можливість використовувати наведений вище ланцюжок міркувань, у нас з'являється можливість визначати координати об'єкта у прив'язці до площини ігрового поля або іншого робочого простору простішим методом – просто через гіпотезу про те, що для кожної точки об'єкта є простий спосіб порахувати точку проєкції на площину в термінах центральної проєкції з віссю, відхилення якої від нормалі до поверхні є малим параметром.

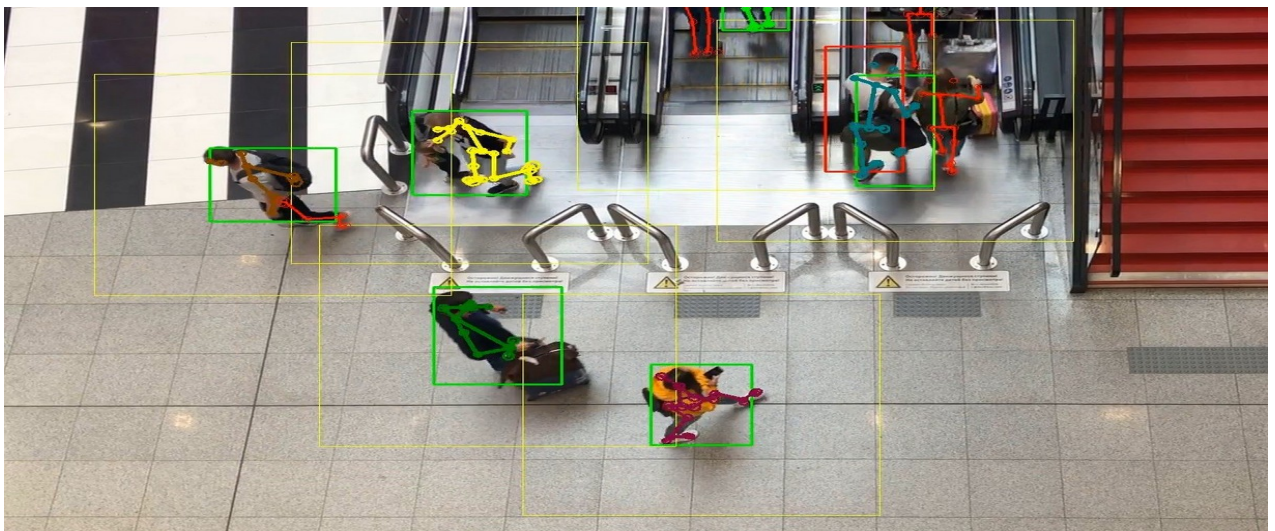


Рисунок 4.3 – Верхнє розташування камери

4.2 Формування тривимірного уявлення на основі двовимірних скелетних моделей

Важливим окремим випадком тривимірної сцени є сцена, що являє собою рух безлічі людей у полі зору камери та їх взаємодія один з одним і предметами та об'єктами, що знаходяться в кадрі.

Ідея представлення фігур людей у формі скелетних моделей не є чимось новим.

У роботі [86] запропоновано скелетне механістичне уявлення фігури людини.

У 1982 році А.М. Формальський формалізував тривимірне рух для скелетної моделі у роботі [87].

Побудова 2D скелетних моделей та оцінка їх параметрів призводить до формування вектора $\vec{S}^{fp} = \{u_1^{fp}, v_1^{fp}, u_2^{fp}, v_2^{fp}, \dots, u_N^{fp}, v_N^{fp}\}$ де $N = 20$ відповідає числу суглобів у вибраній моделі для людської фігури.

Завданням фільтрації є формування кожного об'єкта p тривимірного вектора $\vec{B}^{fp} = \{\hat{x}_1^{fp}, \hat{y}_1^{fp}, \hat{z}_1^{fp}, \hat{x}_2^{fp}, \hat{y}_2^{fp}, \dots, \hat{x}_N^{fp}, \hat{y}_N^{fp}, \hat{z}_N^{fp}\}$

Індекс f в даному випадку має сенс часу $T = f\Delta$

У найпростішому випадку модель має враховувати основні гіпотези:

- кістки є стисливими;
- є обмеження на швидкості та прискорення суглобів;
- рух відбувається на площині.

Як модель руху в більш складному випадку використовується програмний симулятор OpenSim [88].

Для класифікації сцен, що відбуваються на полі, ми використали модель SoccerNet [89], а також підхід до визначення місцезнаходження м'яча, запропоновану в [90].

Навчання 3D конвертера здійснюється так:

На першій фазі проводиться оцінка 3D пози на основі 2D теплових карток [91].

Далі здійснюється тріангуляція 3D скелета на основі машинного чення [92]. І в результаті проводиться уточнення координат 3D скелета на основі CNN нейромережі за допомогою методу, описаного в [93].

Найближчим аналогом запропонованого рішення є підхід, описаний у роботі [94] використовує інформацію про тимчасову зміну параметрів скелета, одержувану за допомогою нейромережевого алгоритму на основі RNN.

5 АНАЛІЗ ВЕКТОРНОГО УЯВЛЕННЯ РУХУ СПОРТСМЕНІВ ТА ПОШУК ТАКТИКО-ТЕХНІЧНИХ ДІЙ

5.1 Попередня обробка

Розв'язання задачі пошуку та класифікації тактико-технічних дій усередині фрагмента відео розбивається на етапи:

- Формуємо покадрові скелети учасників сцени.
- Здійснюємо прив'язку скелетів до координат у системі координат сцени.
- Здійснюємо виявлення траєкторії м'яча.
- Здійснюємо класифікацію поз гравців у динаміці (виявлення типових рухів).
- Здійснюємо виявлення сцен на основі логічних правил. Класифікація поз гравців у динаміці (виявлення типових рухів) здійснюється наступним чином:
 - здійснюють згладжування шумів траєкторій суглобів шляхом гауссівської фільтрації;
 - визначають точки екстремумів швидкості суглобів;
 - переходять на нормоване уявлення скелетів на фазовому просторі координата / швидкість;
 - здійснюють розрахунок вектора показників руху гравця;

здійснюють вирішення задачі класифікації (використаний алгоритм K-means, але цей вибір не є оптимальним з точки зору продуктивності)

5.2 Порівняльний аналіз та вибір структури нейронної мережі

На блок-схемах 5.1 – 5.5 зображені перевірені як можливі варіанти архітектури нейронних мереж.

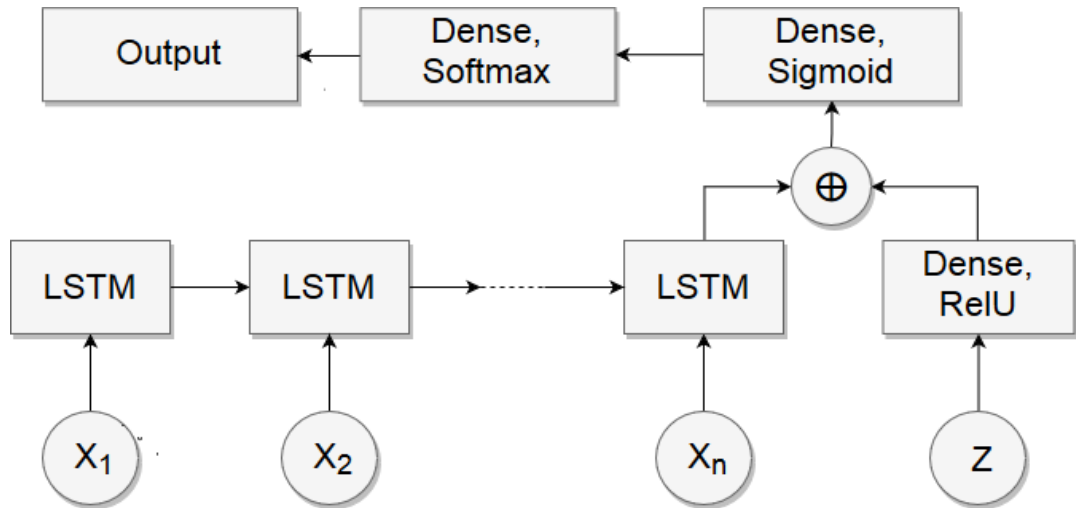


Рисунок 5.1 – DNN – нейронна мережа, що складається тільки з повнозв'язних шарів

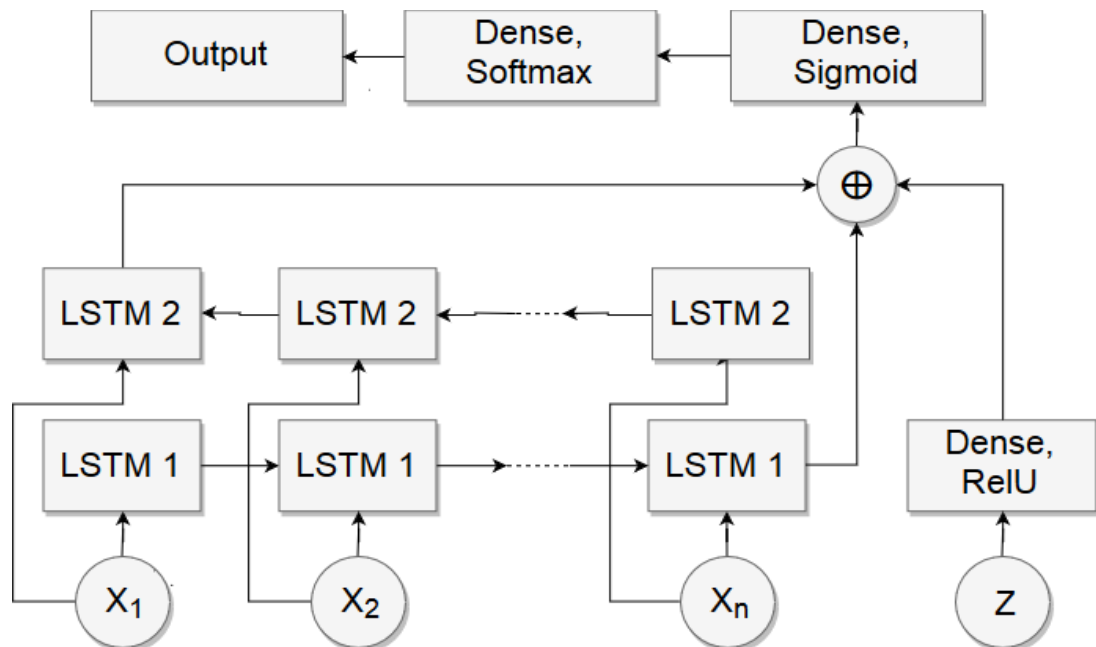


Рисунок 5.2 – Рекурентна нейронна мережа на основі модулів LSTM

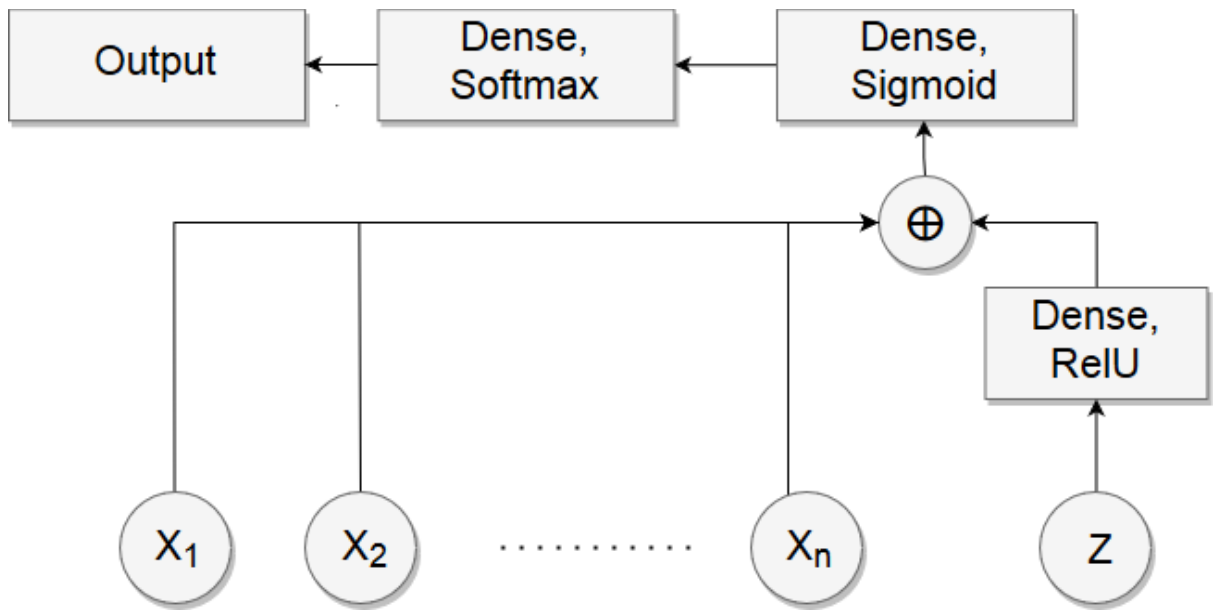


Рисунок 5.3 – Двонаправлена рекурентна нейронна мережа на основі модулів LSTM

Входом мережі у всіх випадках є послідовність із N перетворених векторів \vec{x}_i та вектор ознак виділених на етапі передобробки \vec{z} , виходом – масив ймовірностей. Розшифровка елементів:

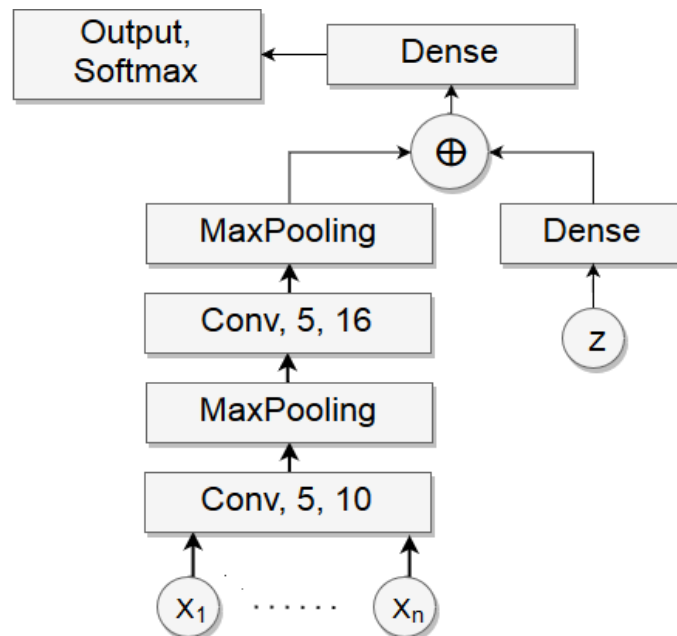


Рисунок 5.4 – Згорткова нейронна мережа, що складається з двох згорткових шарів і двох шарів пулінгу

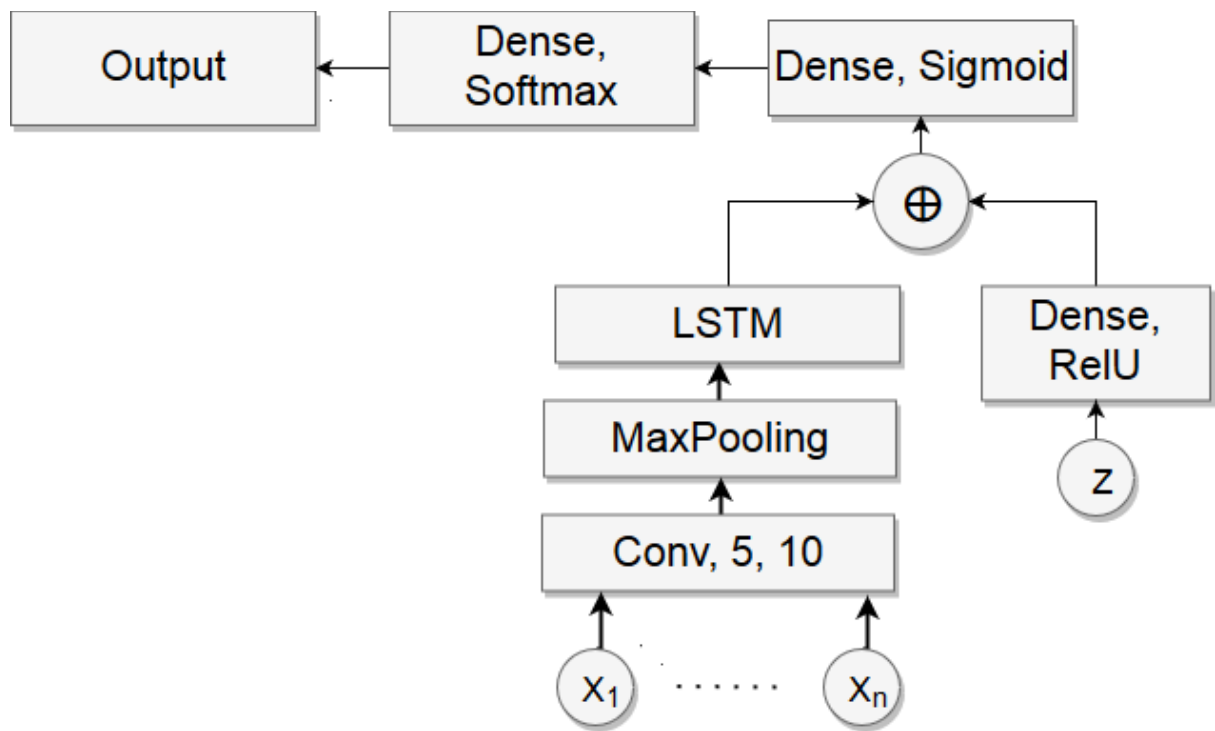


Рисунок 5.5 — Комбінація із згорткового шару та шару LSTM

Dense – елемент нейронної мережі, кожен нейрон якого отримує дані на вхід від кожного з нейронів, що входять до складу попереднього шару. Результатом роботи даного шару є формування добутку вхідного вектора значень на матрицю, що описує коефіцієнти шару, що обговорюється, з подальшим застосуванням функції активації.

LSTM – Long Short Term Memory шар мережі, котрий як і оригінальної роботі [95], розмір прихованого стану точно збігається з розміром входу наступного шару.

Conv – шар згортки, аналогічний використуваним CNN мережах. У прямокутнику на схемі наведено розмірність ядра та кількість виходів шару.

MaxPooling – шар пулінгу, що забезпечує зменшення розмірності завдання шляхом застосування до вхідних сигналів функції *Ma*.

Там де це застосовно, після коми наведена функція активації даного шару.

У ході попередніх експериментів перевірялася ефективність розпізнавання наступних варіантів побудови архітектури використовуваної нею ронної мережі:

- Нейронна мережа, що складається лише з повнозв'язних шарів. (DNN – глибока нейронна мережа прямого поширення) Структура мережі представлена на рисунку 5.1.
- Рекурентна нейронна мережа на основі безлічі модулів LSTM наведена на рисунку 5.2.
- Двонаправлена рекурентна нейронна мережа на основі модулів LSTM наведена на рисунку 5.3.
- Згорткова нейронна мережа, що складається з двох згорткових шарів та двох шарів зменшення розмірності представлена на рисунку 5.4.
- На рисунку 5.5 представлена гібридна спеціалізована мережа, що складається з згорткового шару та шару LSTM.

У таблиці 2 наведено порівняльний аналіз продуктивності описаних вище архітектур нейронних мереж.

Для аналізу нестационарних векторних описів рухів гравців була обрана рекурентна мережа, наведена на рисунку 5.6. Кількість нейронів останнього шару мережі відповідає кількості технічних дій, що розпізнаються. Така рекурентна мережа є ефективною у випадку, коли кількість розпізнаваних тактико-технічних дій гравців відносно мала.



Рисунок 5.6 – LSTM мережа

На N виходах нейронної мережі формуються ймовірності p_n , $n \in [0, N]$, де N – число тактико – технічних дій гравців, що виявляються нами. p_0 –

ймовірність того, що на досліджуваній ділянці відео тактико – технічних дій з довідника, що використовується, не було.

5.3 Результати аналізу та пошуку ТТД

Проведено експеримент для чотирьох найпростіших видів рухів:

- фінт вліво (ТТД G1);
- фінт праворуч (ТТД G2);
- вод м'яча з ауту (ТТД G3);
- підкат (ТТД G4).

Також результатом може бути рішення «жоден із перелічених» якому поставимо у відповідність позначення G0.

Як приклад на рисунку 5.7 наведені кадри з відеоролика з виконанням ТТД G1 «Фінт вліво» з накладеною скелетною моделлю.

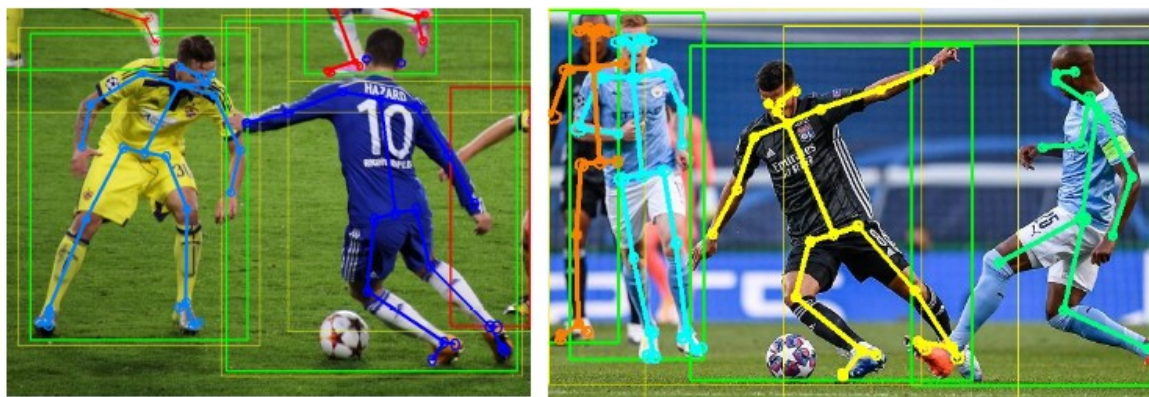


Рисунок 5.7 – Виконання ТТД G1 «Фінт вліво»

На рисунках 5.8 наведені кадри з відеоролика з виконанням ТТД G3 «Введення м'яча з ауту» з накладеною скелетною моделлю та на рисунках 5.9 – для виконання ТТД G4 «Підкат» відповідно.



Рисунок 5.8 – Виконання ТТД G3 «Введення м'яча з ауту»

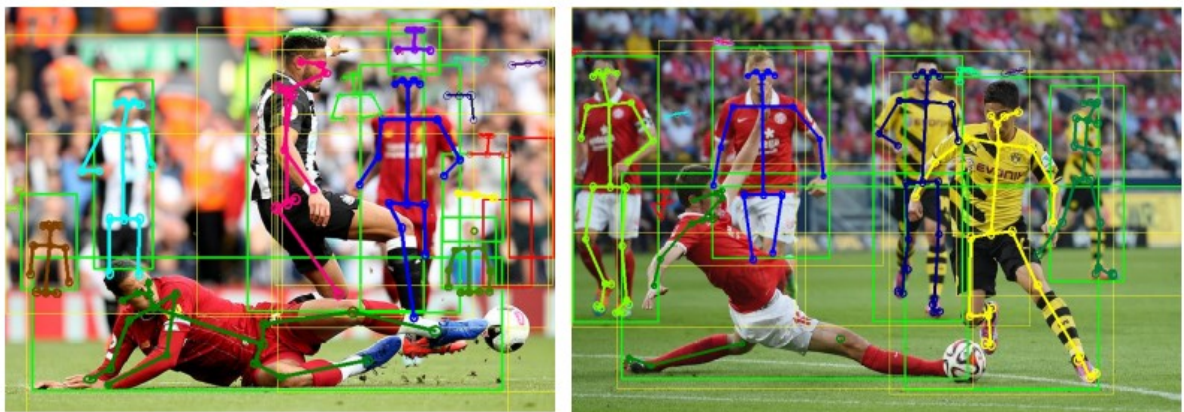


Рисунок 5.9 – Виконання ТТД G4 «Подкат»

ля навчання нейромережі була використана частина масиву даних, що включає фрагменти футбольних матчів. Розмітка навчальної вибірки нами здійснювалася у ручному режимі з допомогою експертної оцінки.

Базовою метрикою, на підставі якої приймається рішення про застосування запропонованого підходу до класифікації тактико-технічних дій є точність (Accuracy): $A = P/N$ де, P – кількість ТТД в експериментальній вибірці, за якими класифікатор у результаті прийняв правильне рішення про віднесення ТТД до деякого класу, а N – розмір навчальної вибірки.

Однак ця метрика не дає вичерпного уявлення про ефективність алгоритму у разі, коли частота появи різних тактико-технічних дій нерівномірна, що є досить типовою ситуацією для ТТД, що виконуються в ході футбольного матчу.

Таблиця 3 – Матриця неточностей для навчальної вибірки у разі чотирьох типів ТТД.

	none	G1	G2	G3	G4
none	2365.0	13.0	34.0	5.0	5.0
G1	46.5	1384.0	2.0	1.0	0.2
G2	32.0	0.0	1345.0	0.0	0.0
G3	39.0	0.5	1.5	1765.0	0.0
G4	67.0	0.5	0.5	0.0	2198.0

Таблиця 4 – Матриця неточностей для експериментальної вибірки у разі чотирьох типів ТТД.

	none	G1	G2	G3	G4
none	676.0	5.0	12.0	0.0	0.0
G1	23.0	355.0	0.0	0.0	0.0

G2	12.0	0.0	498.0	0.0	0.0
G3	9.0	1.0	0.0	389.0	0.0
G4	15.0	0.0	0.0	0.0	450.0

За результатами експерименту для навчальної вибірки було отримано метрики, наведені у таблиці 5.

Розраховані значення дозволяють зробити висновок про те, що запропонований підхід до виявлення ТТД гравців дає хороший результат навіть без урахування спільної обробки алгоритмів розпізнавання дій для групових сцен.

Таблиця 5 – Значення метрик для навчальної та експериментальної вибірки

	Навчальна вибірка	Експериментальна вибірка
Accuracy	0.95	0.96
Precision	0.98	0.95
Recall	0.98	0.96
F-Measure	0.95	0.97

Були також розраховані матриці сполученості для кожного класу про ТТД. Недіагональні елементи матриці відповідають імовірностям помилково – позитивних і помилково – негативних спрацьовувань алгоритму, відповідно.

Для класу G_0 число помилково-позитивних рішень відповідає кількості випадків помилкового пропуску ТТД.

Аналогічним чином, число помилково-негативних спрацьовувань винахідника описує ситуацію, коли ТГД не виконувалося і виявлення руху сталося випадково. Керуючи заданою числовою характеристикою кількості хибно-негативних спрацьовувань, ми можемо управляти «чутливістю» алгоритму.

При додаванні нових ТГД систему буде необхідно необхідно перенавчити, доповнивши навчальну вибірку таким чином, щоб у результаті параметри розпізнавання були краще мінімально прийнятних.

Ця робота може і має бути виконана під час проведення НДДКР з метою розробки повнофункціонального вирішення аналізу відеотрансляцій футбольних матчів.

ВИСНОВКИ

У роботі поставлено та вирішено завдання аналізу відеотрансляції ігрових видів спорту високого дозволу на прикладі футбольних матчів з використанням нейромережевих технологій з метою виявлення тактико-технічних дій гравців.

Основні результати роботи полягають у наступному:

- розроблено та реалізовано алгоритм формування двовимірних векторних скелетних моделей спортсменів на основі відеотрансляцій високої роздільної здатності;
- розроблено та реалізовано алгоритм адаптивного управління областю обробки для відео високої роздільної здатності;
- розроблено та реалізовано алгоритм пошуку типових патернів руху на основі аналізу руху багатовимірного вектора, що описує двовимірну скелетну модель гравця;
- розроблено навчену та інтегровану до складу системи обробки відео модель пошуку та класифікації типових тактико-технічних дій гравців на основі згорткової нейронної мережі.
- запропоновано використовувати алгоритм обчислення оптичного потоку для уточнення координат суглобів (joints), при цьому показано, що розрахунок оптичного потоку для суглобів у 2D не призводить до різких змін результату, однак він здатний зменшити флуктуацію у випадках, коли виникають помилки виявлення суглобів (перепустки від міток у термінах теорії локації).
- формалізовано завдання пошуку скелетних моделей у кадрах високої роздільної здатності з використанням процедури адаптивної зміни вирішення;
- запропоновано використання методів фільтрації для уточнення координат суглобів двовимірної скелетної моделі.

- показано, що фільтр Калмана дозволяє вирішити завдання фільтрації траєкторій 2D і 3D скелетів.
- показано, що LSTM мережа є зручним інструментом для аналізу за векторного часового ряду, що описує двовимірне і тривимірне.

Розроблений підхід до вирішення задачі пошуку тактико– технічних дій на основі аналізу двовимірної та тривимірної динамічної скелетної моделі з використанням LSTM нейромережі.

Експериментальна перевірка результатів дослідження проводилася на основі відеофайлів із записами матчів Ліги Чемпіонів УЄФА високого (4К) дозволу.

Результати роботи можуть використовуватися в тренувальній практиці як команд майстрів, так і в масовому тренувальному процесі.

Запропоновані алгоритми аналізу тактико-технічних дій дозволяють швидко виконувати анотування відеоматеріалів футбольних матчів у близькому до реального часу.

Видається можливим створення аналогічних рішень для інших ігрових видів спорту та реалізація дослідно-конструкторської розробки для програмного комплексу аналізу тактико-технічних процесів футболістів.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Adaptive distributed video surveillance system / A. V. Khelvas // 2020 International Conference on Technology and Entrepreneurship (ICTE) (ICTE 2020). San Diego, USA, apr. 2020.
2. Improved 2D Human Pose Tracking Using Optical Flow Analysis / A. Khelvas // Intelligent Systems and Applications; / ed. by K. Arai, S. Kapoor, R. Bhatia. Cham : Springer International Publishing, 2021. P. 10-22.
3. Разработка архитектуры программной AI платформы для анализа тактико – технических действий и функционального состояния футболистов в процессе игры по данным видеотрансляции / А. Хельвас // Сборник трудов XXII Международной конференции «Цифровая обработка сигналов и ее применение DSPA-2020». — ИПУ им.Трапезникова, 2020. С. 10-22.
4. Лобановский В. В. Бесконечный матч. 3-е изд. Москва : Физкультура и спорт, 1991.
5. The evolution of physical and technical performance parameters in the English Premier League / C. Barnes // International journal of sports medicine. — 2014. Т.35, P. 1095-1100.
6. Real-time Analysis of Physical Performance Parameters in Elite Soccer / K. Andreassen // 2019 International Conference on Content-Based Multimedia Indexing (CBMI). IEEE. 2019. P. 1-6.
7. Computer vision for sports: Current applications and research topics / G. Thomas // Computer Vision and Image Understanding. 2017. Apr. Т. 159.
8. Hughes M., Bartlett R. The use of performance indicators in performance analysis // Journal of sports sciences. 2002. Nov. Т. 20.
9. Team’s Performance on FIFA U17 World Cup 2011: Study based on Notational Analysis / F. Clemente // Journal of Physical Education and Sport. 2012. Mar. Т. 12. P. 13—17.

10. Carling C. Interpreting Physical Performance in Professional Soccer MatchPlay: Should We be More Pragmatic in Our Approach? // Sports medicine (Auckland, N.Z.) 2013. May. T. 43.

11. Anthropometric and Physiological Characteristics of Young Soccer Players According to Their Playing Positions: Relevance for Competition Success / C. Peñas // Journal of strength and conditioning research / National Strength and Conditioning Association. 2011. Nov. T. 25.

12. Physical Fitness, Injuries, and Team Performance in Soccer / A. Arnason // Medicine and science in sports and exercise. 2004. Mar. T. 36.

13. Isokinetic Strength and Anaerobic Power of Elite, Subelite and Amateur French Soccer Players / G. Cometti // International journal of sports medicine. 2001. Jan. T. 22. P. 45—51.

14. Small-sided games in soccer: amateur vs. professional players' physiological responses, physical and technical analysis / D. A // Jan. 2012.

15. Technical performance during soccer matches of the Italian Serie A league / E. Rampinini // Journal of science and medicine in sport / Sports Medicine Australia. Dec. T. 12. 2007.

16. Analysis of High Intensity Activity in Premier League Soccer / V. Di Salvo // International journal of sports medicine. Feb. T. 30. 2009.

17. Influence of certain tactical attacking patterns on the result achieved by the teams participants of the 2010 FIFA World Cup in South Africa / A. Jankovic [и др.] // Fizicka kultura. 2011. Jan. T. 65.P. 34—45.

18. High-Intensity Activity Profiles of Elite Soccer Players at Different Performance Levels / P. Bradley // Journal of strength and conditioning research / National Strength & Conditioning Association. 2009. Nov. T. 24.

19. James N. Notational analysis in soccer: past, present and future. // International Journal of Performance Analysis in Sport. 2006. Nov. T. 6, P. 67—81.

20. Yiannakos A., Armatas V. Evaluation of the goal scoring patterns in European Championship in Portugal 2004 // *International Journal of Performance Analysis in Sport*. 2006. June. T. 6, P. 178—188.

21. GOAL SCORING PATTERNS IN GREEK TOP LEVELED SOCCER MATCHES / V. Armatas // *Citius Altius Fortius*. 2009 June. T. 6.

22. Peñas C., A D. Ball Possession Strategies in Elite Soccer According to the Evolution of the Match-Score: the Influence of Situational Variables // *Journal of Human Kinetics*. 2010. Jan. T. 25, P. 93—100.

23. The influence of match location, quality of opposition, and match status on technical performance in professional association football / J. Taylor // *Journal of sports sciences*. 2008. Aug. T. 26.

24. Penas C., Acero R. Determinants of possession of the ball in soccer // *Journal of sports sciences*. 2007. Aug. T. 25.

25. Taylor J., Mellalieu S., James N. A Comparison of Individual and Unit Tactical Behaviour and Team Strategy in Professional Soccer // *International Journal of Performance Analysis in Sport*. 2005. Nov. T. 5, P. 87— 101.

26. Hughes M., Franks I. Analysis of passing sequences, shots and goals in soccer // *Journal of sports sciences*. 2005. June. T. 23, P. 509—14.

27. Q. Dang, J. Yin, B. Wang, and W. Zheng, «Deep learning based 2d human pose estimation: A survey,» *Tsinghua Science and Technology*, vol. 24, no. 6, P. 663–676, 2019.

28. C. Gao, Y. Zou, and J.-B. Huang, «ican: Instance-centric attention network for human-object interaction detection,» *arXiv preprint arXiv:1808.10437*, 2018.

29. R. Girshick, J. Donahue, T. Darrell, and J. Malik, «Rich feature hierarchies for accurate object detection and semantic segmentation,» in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, P. 580–587.

30. G. Gkioxari, R. Girshick, P. Dollár, and K. He, «Detecting and recognizing humanobject interactions,» in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, P. 8359–8367.
31. Liu, J., Shahroudy, A., Xu, D., & Wang, G. (2016). Spatio-temporal lstm with trust gates for 3D human action recognition. In European Conference on Computer Vision
32. Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing*, 234, P. 11-26
33. Ning, C., Zhou, H., Song, Y., & Tang, J. (2017). Inception single shot multibox detector for object detection. In IEEE International Conference on Multimedia & Expo Workshops
34. Pan, C. Yan, W. (2020) Object detection based on saturation of visual perception. *Multim. Tools Appl.* P.27-28.
35. Plagemann, C., Ganapathi, V., Koller, D., & Thrun, S. (2010). Real-time identification and localization of body parts from depth images. In IEEE International Conference on Robotics and Automation
36. Qian, R., Liu, Q., Yue, Y., Coenen, F., & Zhang, B. (2016). Road surface traffic sign detection with hybrid region proposal and Fast R-CNN. In International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery P. 555-559.
37. Rahmani, H., & Bennamoun, M. (2017). Learning action recognition model from depth and skeleton videos. In IEEE International Conference on Computer Vision
38. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*
39. Sadoughi, F., Kazemy, Z., Hamedan, F., Owji, L., Rahmanikatifari, M., & Azadboni, T. T. (2018). Artificial intelligence methods for the diagnosis of breast cancer by image processing: a review. *Breast Cancer: Targets and Therapy*.

40. Shao, Z., Cai, J., & Wang, Z. (2017). Smart monitoring cameras driven intelligent processing to big surveillance video data. *IEEE Transactions on Big Data*, P. 105-116
41. Song, C., He, L., Yan, W., Nand, P. (2019) An improved selective facial extraction model for age estimation. In *IVCNZ*
42. Strong, A. (2016). Applications of artificial intelligence & associated technologies
43. Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational Intelligence and Neuroscience*
44. Wong, S. Y., Yap, K. S., Zhai, Q., & Li, X. (2019). Realization of a hybrid locally connected extreme learning machine with DeepID for face verification. *IEEE Access*.
45. Xie, C., Li, P., & Sun, Y. (2019). Pedestrian Detection and Location Algorithm Based on Deep Learning. In *International Conference on Intelligent Transportation, Big Data & Smart City*.
46. Zheng, K., Yan, W., Nand, P. (2018) Video dynamics detection using deep neural networks. *IEEE Trans. Emerg. Top. Comput. Intell*.