

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук
(повна назва)

Кафедра _____ Програмної інженерії
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА **Пояснювальна записка**

рівень вищої освіти _____ другий (магістерський)

Дослідження методів комп'ютерного зору для визначення

положення тіла людини

(тема)

Виконав:

Випускник 2 курсу, групи _____ ІПЗМ-21-1
_____ Калугін В.В.

(прізвище, ініціали)

121- Інженерія програмного
Спеціальність _____ забезпечення

(код і повна назва спеціальності)

Тип програми _____ Освітньо-наукова

(код і повна назва спеціальності)

Керівник _____ проф. Білоус Н.В.

(посада, прізвище)

Допускається до захисту
Зав. кафедри _____

(підпис)

_____ З.В. Дудар
(прізвище, ініціали)

2023 р.

Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук _____
(повна назва)
 Кафедра _____ Програмної інженерії _____
(повна назва)
 Рівень вищої освіти _____ другий (магістерський) _____
(повна назва)
 Спеціальність _____ 121 – Інженерія програмного забезпечення _____
(код і повна назва спеціальності)
 Тип програми _____ освітньо-наукова _____
(освітньо-професійна або освітньо-наукова)
 Освітня програма _____ Інженерія програмного забезпечення _____
(повна назва)

ЗАТВЕРЖДУЮ:

Зав. кафедри _____
(підпис)
 «__» _____ 2023 р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

студента _____ Калугіна Віталія Вячеславовича _____
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження методів комп'ютерного зору для визначення положення тіла людини

затверджена наказом університету від 29.03.2023 № 302ст

2. Термін подання студентом роботи до екзаменаційної комісії " 18 " 05 2023 р.

3. Вихідні данні до роботи теоретичні відомості щодо моделей визначення поз, вимоги до експериментальної програмної системи для проведення досліджень, мова програмування Python, бібліотека BlazePose, середовище розробки Visual Studio Code

4. Перелік питань, що потрібно опрацювати в роботі аналіз та порівняння існуючих моделей визначення пози людини, аналіз існуючих методів порівняння положень тіла, вибір підходящої моделі для експериментального дослідження, моделювання експериментальної програмної системи, створення програмних рішень, проведення експериментів та аналіз отриманих результатів

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Аналіз проблемної області та постановка задачі	22.02 – 05.03.23	Виконано
2	Аналіз та вибір методів та моделей для дослідження	06.03 – 17.03.23	Виконано
3	Аналіз та моделювання предметної області	15.03 – 28.03.23	Виконано
4	Планування експериментів	29.03 – 01.04.23	Виконано
5	Експериментальні дослідження	02.04 – 20.04.23	Виконано
6	Аналіз результатів експериментальних досліджень та розробка рекомендацій	21.04 – 24.04.23	Виконано
7	Програмна реалізація системи навчання технікам ПМР	17.04 – 02.05.23	Виконано
8	Написання та оформлення статті та тез доповіді	20.04 – 30.04.23	Виконано
9	Підготовка пояснювальної записки	15.04 – 05.05.23	Виконано
10	Підготовка презентації та доповіді	30.04 – 09.05.23	Виконано
11	Нормоконтроль		Виконано
12	Рецензування		Виконано
13	Занесення диплома в електронний архів		Виконано
14	Попередній захист		Виконано
15	Допуск до захисту у зав. кафедри		Виконано

Дата видачі завдання _____ «23» січня 2023р.

Студент _____ Калугін В.В.
(підпис) (прізвище, ініціали)

Керівник роботи _____ проф. Білоус Н.В.
(підпис) (прізвище, ініціали)

РЕФЕРАТ

Кваліфікаційна робота містить: 99 с., 32 рис., 3 табл., 24 джерел.

МАШИННЕ НАВЧАННЯ, КОМП'ЮТЕРНИЙ ЗІР, МОДЕЛЬ, ЗГОРТКОВА НЕЙРОННА МЕРЕЖА, ВИЗНАЧЕННЯ ПОЗИ, КЛЮЧОВІ ТОЧКИ, ЕВКЛІДОВА ВІДСТАНЬ, OPEN POSE, BLAZE POSE.

Об'єкт дослідження – моделі машинного навчання для визначення поз тіла людини.

Метою роботи є проведення дослідження щодо визначення ефективності наявних моделей визначення пози та можливості використання цих моделей з метою реалізації програмної системи для навчання технікам прогресивної м'язової релаксації.

Методами використаними у даному дослідженні є аналіз предметної галузі, порівняння наявних моделей визначення пози, вибір моделі та використання її для проведення експерименту, розробка експериментальної програмної системи з метою виявлення найефективнішого методу порівняння поз, аналіз результатів проведених теоретичних та експериментальних досліджень.

У результаті проведених теоретичних досліджень було обрано модель BlazePose для проведення експерименту. За результатами експериментального дослідження метод зваженої відстані показав найкращу ефективність. Дані результати були використані при створенні програмної системи навчання технікам прогресивної м'язової релаксації.

MACHINE LEARNING, COMPUTER VISION, MODEL, CONVOLUTIONAL NEURAL NETWORK, POSE DETECTION, KEYPOINTS, EUCLIDEAN DISTANCE, OPEN POSE, BLAZE POSE.

The object of the research is machine learning models for determining human body positions.

The purpose of the work is to conduct a research on determining the effectiveness of existing models of pose determination and the possibility of using these models for the purpose of implementing a software system for teaching progressive muscle relaxation techniques.

The methods used in this research are analysis of the subject area, comparison of existing posture determination models, selection of a model and its use for conducting an experiment, development of an experimental software system in order to identify the most effective method of pose comparison, analysis of the results of theoretical and experimental studies.

As a result of the conducted theoretical studies, the BlazePose model was chosen for the experiment. According to the results of the experimental study, the weighted distance method showed the best efficiency. These results were used in the creation of a software system for teaching progressive muscle relaxation techniques.

Умови публікації пояснювальної записки

Я, Калугін Віталій Вячеславович, студент гр. ПЗМ-21-1, здобувач вищої освіти на другому (магістерському) рівні кафедри «Програмна інженерія», заявляю: моя кваліфікаційна робота на тему «Аналіз результатів експериментальних досліджень та розробка рекомендацій», що буде представлена в екзаменаційну комісію для публічного захисту, виконана самостійно, в ній не містяться елементи плагіату і вона може бути опублікована в електронному архіві відкритого доступу EIArKhNURE. Всі запозичення з друкованих та електронних джерел мають відповідні посилання.

Я ознайомлений з діючим положенням «Про протидію академічному плагіату в ХНУРЕ», згідно з яким виявлення плагіату є підставою для відмови в допуску кваліфікаційної роботи до захисту та застосування дисциплінарних заходів.

ЗМІСТ

Перелік умовних скорочень.....	8
Вступ.....	6
1 Аналіз предметної галузі.....	11
1.1 Класифікація методів визначення пози людини	19
1.1.1 Метод визначення пози людини на основі регресії.....	20
1.1.2 Метод визначення пози людини на основі теплової карти.....	22
1.1.3 Метод визначення пози людини «згори вниз».....	24
1.1.4 Метод визначення пози людини «знизу вгору».....	26
1.2 Огляд наборів даних та метрик оцінки ефективності моделей визначення положень тіла.....	28
1.2.1 Набір даних FLIC.....	28
1.2.2 Набір даних MPII Human Pose.....	29
1.2.3 Набір даних COCO.....	30
1.2.4 Метрика ефективності OKS.....	32
1.2.5 Метрика ефективності «average precision» (AP) та її варіант «mean average precision» (mAP).....	33
1.2.6 Метрика ефективності PCK та її варіант PCKh.....	34
1.3 Огляд методів порівняння положень тіла.....	35
1.3.1 Порівняння положень тіла за допомогою евклідової відстані.....	36
1.3.2 Порівняння положень тіла за допомогою зваженої відстані.....	39
1.3.3 Порівняння положень тіла за допомогою DTW.....	40
1.4 Огляд моделей глибокого навчання для визначення положень тіла.....	42
1.4.1 Огляд моделі OpenPose.....	42
1.4.2 Огляд моделі PoseNet.....	45
1.4.3 Огляд моделі BlazePose.....	48
2 Постановка науково-технічної задачі.....	51
3 Планування експериментального дослідження	54

3.1	Моделювання експериментальної програмної системи	54
3.2	Підготовка набору даних для проведення експериментального дослідження.....	56
4	Аналіз результатів теоретичного та експериментального досліджень.....	59
4.1	Аналіз результатів теоретичного дослідження.....	59
4.2	Опис розробленої експериментальної програмної системи.....	64
4.3	Аналіз результатів експериментального дослідження.....	67
5	Опис програмної реалізації.....	70
5.1	Опис основного функціоналу та використаних технологій.....	70
5.2	Опис мультимодельного підходу	71
5.3	Майбутні напрямки розвитку.....	72
	Висновки.....	74
	Перелік джерел посилання.....	76
	Додаток А Перелік джерел посилання за науковими напрямами керівника та науковців кафедри програмної інженерії.....	79
	Додаток Б Звіт результатів перевірки на унікальність тексту в мережі інтернет та базі ХНУРЕ.....	80
	Додаток В Слайди презентації.....	81
	Додаток Г Апробація результатів.....	89
	Додаток Д Експертний висновок результатів перевірки кваліфікаційної роботи на відповідність оформлення вимогам ДСТУ 3008:2015.....	99

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

ПМР – прогресивна м'язова релаксація

CNN – згорткова нейронна мережа

HOG – гістограма орієнтованих градієнтів

MPPE – оцінка пози кількох людей

OKS – індекс подібності ключових точок об'єкта

PAF – поля спорідненості частин

SPPE – оцінка пози однієї людини

ВСТУП

Положення тіла відіграє дуже важливу роль у фізичному та психічному благополуччі людини. Погана постава може призвести до розладів опорно-рухового апарату, болю та навіть вплинути на впевненість в собі. З розвитком технологій глибокого навчання відбулася революція в галузі аналізу положення тіла. Ці технології пропонують швидкий і точний інструмент для визначення та порівняння положень тіла та не потребують окремого специфічного обладнання.

Дані технології можна використовувати для створення різноманітних програмних систем, наприклад, дистанційної медичної допомоги, які є дуже актуальними у сучасному світі через їх доступність, зручність, економічну ефективність і т.д. Ці системи можуть надати людям доступ до медичних послуг, які, можливо, було б важко або неможливо отримати інакше, роблячи медичне обслуговування загальнодоступним та зручним.

Метою даного дослідження є оцінка ефективності технологій глибокого навчання визначення та порівняння положень тіла та оцінка можливості використання цих технологій для розробки програмної системи навчання технікам прогресивної м'язової релаксації (ПМР).

Задачами даного дослідження є:

- ознайомитися з наявною літературою щодо технологій глибокого навчання для визначення положень тіла та методів порівняння отриманих положень;
- проаналізувати точність і ефективність доступних моделей машинного навчання у визначенні положень тіла;
- проаналізувати наявні методи для порівняння положень тіла та оцінити їх ефективність;
- розробити програмний застосунок для навчання методикам ПМР з використанням технологій глибокого навчання;

Об'єктом даного дослідження є технології глибокого навчання для визначення пози тіла та методи порівняння отриманих положень.

Предметом дослідження є методи визначення положень тіла з використанням технологій глибинного навчання та методи порівняння отриманих положень.

У цьому дослідженні використовуватиметься комбінація таких кількісних методів дослідження, як аналіз та синтез, методи порівняння. Кількісні методи аналізу та синтезу буде використано для аналізу точності та ефективності обраних моделей машинного навчання у визначенні положень тіла. Метод порівняння буде використано для оцінки найбільш ефективного методу порівняння положень тіла.

Результати цього дослідження з практичної сторони питання сприятимуть розробці програмної системи для навчання методам ПМР, яку можуть використовувати широкий спектр користувачів, як то медичні працівники або люди, зацікавлені в покращенні свого фізичного та психічного стану.

Наукова новизна дослідження полягає в оцінці ефективності методів порівняння положень тіла використовуючи стандартний показник оцінки – індекс F1. Дослідження сприяє ширшому розумінню сильних і слабких сторін різних методів порівняння поз та різних технологій машинного навчання визначення положень тіла, а також допомагає визначити вектор майбутніх досліджень і розробок у цій галузі.

Наукова стаття за темою дослідження була опублікована у науковому журналі «Радіоелектроніка, інформатика, управління» під видавництвом Національного університету "Запорізька політехніка".

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ

Завдання визначення пози людини на вхідному зображенні чи відео можна визначити як завдання пошуку точок з'єднання на тілі людини, також відомих як ключові точки – голова, шия, плечі, лікті, зап'ястки, стегна, коліна та щиколотки.

Значна кількість різноманітних аспектів роблять визначення пози складним завданням в області комп'ютерного зору [1], наприклад:

- неоднозначність: пози людини можуть бути дуже неоднозначними, особливо коли частини тіла накладаються або закривають одна одну;
- варіативність: пози людського тіла можуть сильно відрізнятися залежно від таких факторів, як форма тіла, одяг, кут огляду камери, освітлення та фон;
- складність: людські тіла є складними та шарнірними, з багатьма ступенями свободи та складними рухами в суглобах;
- оклюзії: у багатьох сценаріях частини тіла можуть бути закриті іншими об'єктами або іншими частинами тіла;
- ефективність у реальному часі: у деяких програмах, таких як спортивний аналіз або взаємодія людини з роботом, потрібна оцінка пози в реальному часі, що потребує швидкої обробки вхідних даних та низької затримки.

В залежності від того, в якому просторовому вимірі (2D або 3D) проводиться визначення зазначених ключових точок, завдання визначення пози людини можна класифікувати на двовимірне чи тривимірне.

Визначення 2D пози — це завдання оцінки положення та орієнтації частин тіла людини на плоскому зображенні чи відео. Метою двовимірної оцінки пози є виведення набору двовимірних координат суглобів, які представляють положення кожної частини тіла (див. рис. 1.1).

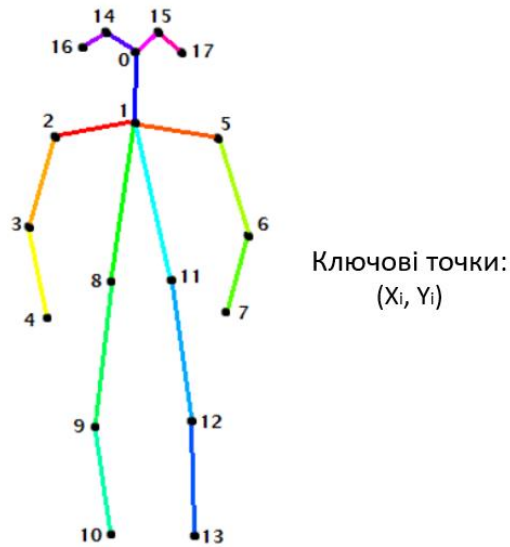


Рисунок 1.1 – Визначення 2D пози

Ці координати зазвичай представлені у вигляді пар (x, y) , де x і y — горизонтальні та вертикальні піксельні координати відповідно.

З іншого боку, визначення 3D-пози — це завдання оцінки положення та орієнтації частин тіла людини в тривимірному просторі. Метою оцінки 3D-пози є виведення набору координат суглобів, які представляють положення кожної частини тіла в тривимірній системі координат (див. рис. 1.2).

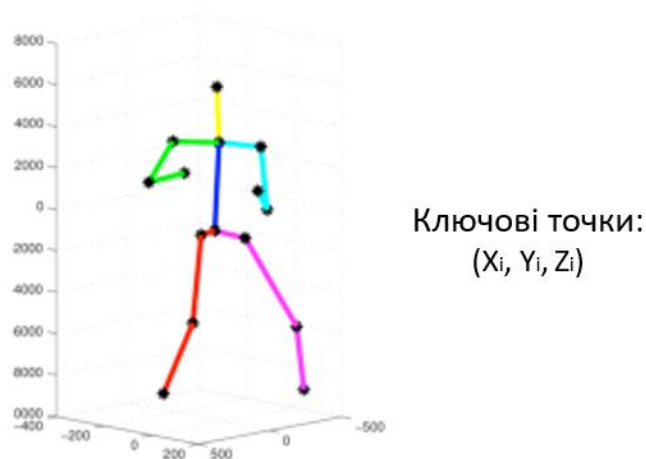


Рисунок 1.2 – Визначення 3D пози

Ці координати зазвичай представлені у вигляді триплетів (x, y, z) , де x , y і z є 3D координатами частин тіла.

У той час як оцінку 2D-пози можна виконати на одному зображенні чи відеокадрі, оцінка 3D-пози зазвичай потребує кількох кутів огляду або послідовностей зображень чи відео, щоб отримати достатню інформацію про глибину. Крім того, для оцінки 3D-пози може знадобитися відкаліброване налаштування камери, щоб забезпечити точну оцінку глибини, тоді як оцінку 2D-пози можна виконувати на некаліброваних зображеннях [2].

Завдання визначення пози також можна класифікувати на основі типу вхідних даних, як то RGB, ToF чи інфрачервоні зображення або відео.

Вхідні дані RGB формату фіксують інформацію про колір у видимому спектрі та зазвичай використовуються в 2D оцінці пози. RGB зображення та відео можуть надавати кольорову інформацію про об'єкт з високою роздільною здатністю, однак вони не надають інформації про глибину чи відстань.

ToF формат (див. рис. 1.3) навпаки надає дані про відстань до сцени та дозволяє точно визначити 3D-позиції суглобів. Датчики глибини використовують інфрачервоне або структуроване світло для вимірювання відстані між камерою та об'єктом [3].



Рисунок 1.3 – Приклад ToF зображення

Інфрачервоні камери вловлюють інфрачервоне світло, що випромінюється від сцени або відбивається об'єктом, тож дані такого формату можуть забезпечити надійне відстеження контурів тіла навіть в умовах слабого освітлення. Вони

зазвичай використовуються для монокулярної оцінки пози, але також можуть використовуватися і для мультимодальної оцінки в поєднанні з іншими датчиками, такими як датчики глибини.

В залежності від типу вхідних даних, обраної системи координат, в яких проводиться визначення положення тіла людини, необхідного рівня деталізації та складності вирішення тієї чи іншої задачі можна побудувати певні моделі тіла людини (див рис. 1.4).



Рисунок 1.4 – Класифікація моделей тіла людини

Кінематична, або ж модель на основі скелета, представляє тіло людини як набір взаємопов'язаних кісток або суглобів, де кожен суглоб відповідає певній частині тіла, наприклад голові, тулубу, руках і ногах (див. рис. 1.5) [4]. В такому разі завдання оцінки пози передбачає оцінку позицій і орієнтацій цих суглобів у 3D або 2D просторі.



Рисунок 1.5 – Кінематична модель тіла людини

Ця модель часто використовується в програмах реального часу, таких як захоплення руху, розпізнавання дій і взаємодія людини з роботом через її простоту та ефективність. Вона також широко використовується в дослідженнях комп'ютерного зору як еталон для оцінки алгоритмів визначення пози.

Однією з переваг моделі на основі скелета є те, що вона забезпечує просте та інтуїтивно зрозуміле представлення людського тіла, що полегшує інтерпретацію та візуалізацію. Вона також є досить ефективною з точки зору обчислення, що робить її придатною для програм реального часу з обмеженими обчислювальними ресурсами.

Однак ця модель має й деякі обмеження. Наприклад, вона не забезпечує детальне представлення форми тіла, що може ускладнити точну оцінку пози тіла, особливо в сценаріях із оклюзіями або складними позами. Вона також не фіксує варіації форми тіла, що може обмежити її узагальнення для різних індивідів і популяцій.

Для подолання цих обмежень існують різні розширення та модифікації кінематичних моделей, наприклад додавання додаткових суглобів або використання кількох моделей для представлення різних частин тіла. Також можливе поєднання моделей на основі скелета з іншими представленнями тіла, такими як площинні або об'ємні моделі, для досягнення балансу між точністю та ефективністю.

Площинна модель часто використовуються в дослідженнях комп'ютерного зору як альтернатива моделі на основі скелета, особливо коли потрібне точне моделювання форми тіла (див рис. 1.6).

Основною перевагою даної моделі є те, що вона забезпечує компактне та виразне зображення людського тіла, яке може зафіксувати варіації форми тіла та пози окремих людей і груп населення. Вона також дозволяє використовувати геометричні обмеження, такі як гладкість або симетрія, які можуть підвищити точність і надійність алгоритмів оцінки пози [5].

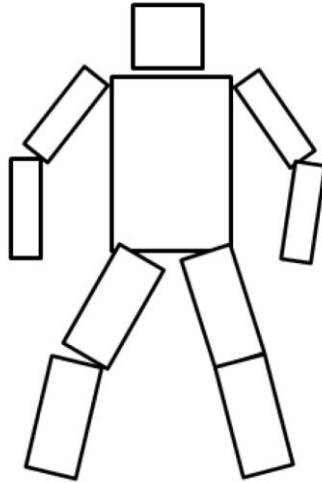


Рисунок 1.6 – Площинна модель тіла людини

До недоліків даної моделі можна віднести те, що її оптимізація та використання можуть бути дорогим з точки зору обчислень, особливо коли кількість плоских сегментів велика. Вона також може бути чутливою до змін у формі тіла та пози, що може зробити її менш придатною для додатків у реальному часі з обмеженими даними для навчання.

Останнім типом моделей тіла є об'ємна модель. Вона використовує 3D-об'єм для відображення пози та форми тіла (див. рис. 1.7).



Рисунок 1.7 – Об'ємна модель тіла людини

У цій моделі кожен воксель представляє невелику кубічну область простору, яка відповідає зовнішньому вигляду або зайнятості цієї області.

Об'ємна модель зазвичай використовуються в програмах, де потрібна детальна та реалістична 3D-модель людського тіла. Вона дозволяє створювати реалістичні та виразні анімації або зображення тіла, а також може фіксувати складні та непомітні варіації форми та пози тіла.

Використання даної моделі може дозволити оцінити дрібні деталі поверхні тіла та впоратися з оклюзіями, які можуть виникнути, коли різні частини тіла стикаються або накладаються.

Однак навчання та оцінка об'ємної моделі є дорогими операціями з точки зору обчислень, особливо коли роздільна здатність об'єму висока. Вона також може страждати від порожнього простору або пустот, що призводить до неефективності та надмірності в процесі моделювання та логічного висновку [6].

Для визначення пози людини протягом багатьох років використовувалася низка класичних підходів, таких як Picture Structure Models (PSM), гістограма орієнтованих градієнтів (HOG) і Edgelets. Ці підходи представляють людське тіло як набір з'єднаних частин або невеликих відрізків ліній і кодують просторові відносини між ними, щоб визначити позу тіла на зображенні [1].

PSM представляють тіло як структуру дерева, де кожен вузол представляє частину, а ребра між вузлами представляють просторові відносини між частинами. Модель кодує попередні знання про відносне положення та орієнтації частин тіла у формі набору шаблонів, отриманих із даних навчання. Ці шаблони визначають зовнішній вигляд кожної частини та просторові співвідношення між ними.

Враховуючи вхідне зображення, PSM шукають найкращу відповідну конфігурацію частин тіла, яка максимізує подібність між зображенням і вивченими шаблонами. Це робиться шляхом ітеративного уточнення оцінок положення та орієнтації кожної частини на основі локальних особливостей зображення навколо кожної частини.

PSM є досить ефективним методом в визначенні пози людського тіла, але залишається чутливим до оклюзій, змін освітлення та одягу та складних поз тіла.

Цей метод також вимагає значної кількості навчальних даних для вивчення точних шаблонів.

Гістограма орієнтованих градієнтів (HOG) — ще один класичний підхід до визначення пози людини, що широко використовується в комп'ютерному зорі. HOG — це метод виділення ознак, який обчислює щільне представлення локальних напрямків градієнта на зображенні.

HOG представляє тіло як набір локальних ділянок зображення, і кожна ділянка описується гістограмою напрямків градієнта (див. рис. 1.8). Ці гістограми обчислюються для набору клітинок у межах кожного фрагмента, і клітинки групуються разом, щоб утворити більші просторові блоки. Потім гістограми в кожному блоці нормалізуються для врахування варіацій освітлення та контрасту.

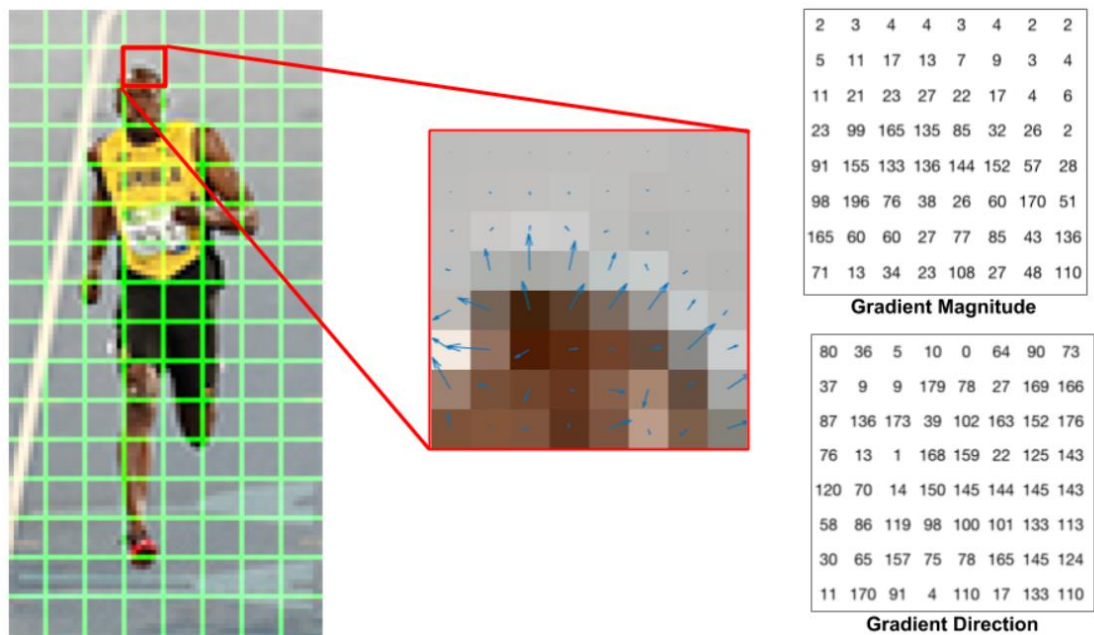


Рисунок 1.8 – Приклад роботи HOG методу

Отримавши вхідне зображення, HOG виділяє ці локальні особливості та використовує їх для навчання класифікатора, який може розрізняти різні частини тіла. Під час тестування HOG обчислює локальні особливості тестового зображення та використовує класифікатор для прогнозування наявності та розташування частин тіла [7].

Основна відмінність між HOG і PSM полягає в тому, що HOG є технікою виділення ознак, тоді як PSM є структурованою моделлю, яка фіксує просторові відносини між частинами тіла. HOG зазвичай використовується як зовнішній екстрактор функцій для більш складних моделей, які фіксують просторову структуру тіла, таких як PSM або моделі на основі глибокого навчання [8].

Ці класичні підходи широко використовувалися для вирішення задачі визначення поз та показували прийнятні результати у певних сценаріях, але новітні методи з використанням глибоко машинного навчання, що можуть впоратися з вже описаними проблемами (оклюзії, варіації в освітленні та одязі, а також складність самої пози людини) на сьогодні є більш популярними та ефективними.

Залежно від кількості людей, які відстежуються, дані методи також можна класифікувати на методи для оцінки пози однієї людини або для кількох осіб. Оцінка пози однієї людини (SPPE) є найпростішою з двох, із гарантованою присутністю лише однієї людини в кадрі. З іншого боку, оцінка пози кількох людей (MPPE) повинна вирішити додаткову проблему оклюзії між людьми. Початкові підходи до оцінки пози були здебільшого зосереджені на SPPE, однак з наявністю величезних наборів даних для кількох осіб проблема MPPE останнім часом привертає все більше уваги [5].

1.1 Класифікація методів визначення пози людини

Як вже зазначалося, методи оцінки пози людини в глобальному плані можуть бути поділені на дві основні категорії – це оцінка пози однієї людини (SPE) і оцінка пози кількох людей (MPE).

В свою чергу в рамках SPE існує дві основні підкатегорії методів: на основі теплової карти та на основі регресії. Метод на основі теплової карти створює теплову карту для кожного з'єднання тіла, яка відображає ймовірність розташування з'єднання в кожному пікселі зображення. Метод заснований на

регресії на пряму регресує спільні координати з вхідного зображення як безперервні значення [1].

MPE також можна розділити на дві підкатегорії: метод «згори вниз» і «знизу вгору» (див. рис. 1.9). Метод «згори вниз» спочатку виявляє і виділяє кожну людину на зображенні, а потім застосовує метод оцінки пози однієї людини до кожного виділеного об'єкту. Метод «знизу вгору» виявляє усі суглоби тіла на зображенні, а потім групує їх для окремих осіб [9].



Рисунок 1.9 – Класифікація методів оцінки поз

Більш детально наведені підкатегорії методів оцінки пози будуть розглянуті далі.

1.1.1 Метод визначення пози людини на основі регресії

Ключова ідея методу на основі регресії полягає в тому, щоб оцінити координати суглобів безпосередньо із зображення без явного створення теплової карти для кожного суглоба (див. рис. 1.10) [10]. Цей підхід має кілька переваг, включаючи можливість обробки прихованих суглобів і потенціал для більш точної

оцінки пози шляхом прямого прогнозування координат суглоба як безперервних значень.

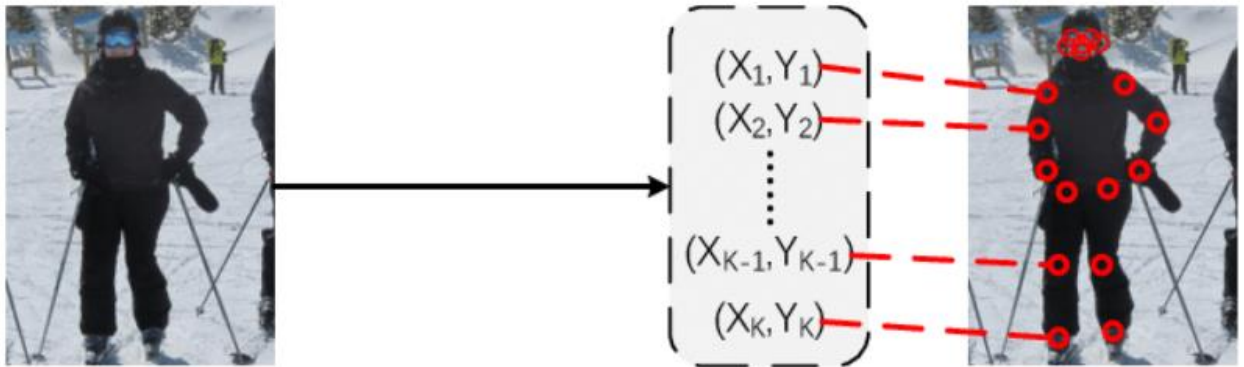


Рисунок 1.10 – Метод визначення пози на основі регресії

В даному випадку зазвичай використовують архітектуру згорткової нейронної мережі (CNN) з метою вилучення ознак із вхідного зображення, а потім один або більше пов'язаних шарів прогнозують координати ключових точок. Вихідні дані мережі – це набір прогнозованих координат для кожного суглоба, які можна безпосередньо порівняти з істинними координатами для обчислення ефективності.

Крім того, у деяких моделях використовується багатозадачний підхід до навчання, коли мережа навчена передбачати як координати суглобів, так і показники достовірності для кожного суглоба. Це дозволяє мережі навчитися розрізняти видимі та приховані суглоби та генерувати точніші прогнози щодо ключових точок.

Загалом, методи на основі регресії показали багатообіцяючі результати в останні роки та досягли високої ефективності на контрольних наборах даних для оцінки пози як однієї людини, так і кількох людей. В той же час вони можуть бути досить дорогими з обчислювальної точки зору та вимагати великих обсягів анотованих даних для навчання.

1.1.2 Метод визначення пози людини на основі теплової карти

Метод оцінки пози людини на основі теплової карти генерує теплову карту для кожного суглоба тіла, вказуючи на ймовірність того, чи цей суглоб розташований у кожному пікселі вхідного зображення (див. рис. 1.11).

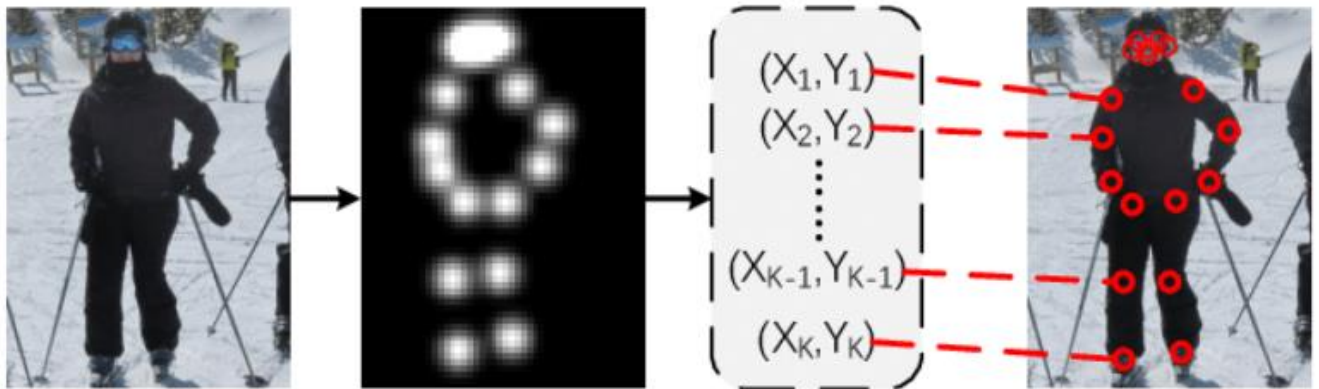


Рисунок 1.11 – Метод визначення пози на основі теплової карти

Щоб передбачити позу людини на вхідних даних, теплові карти для кожного суглоба спочатку генеруються за допомогою CNN. Потім ці теплові карти використовуються для оцінки двовимірних або тривимірних координат кожного з'єднання, як правило, з використанням алгоритму постобробки, наприклад немаксимального придушення (NMS) або жадібного алгоритму. Розраховані координати ключових точок потім використовуються для обчислення остаточної пози.

Існує також кілька варіацій методів на основі теплової карти, які були запропоновані в останні роки. Наприклад, деякі методи використовують багат шарову Stacked Hourglass архітектуру для підвищення точності теплових карт (див. рис. 1.12).

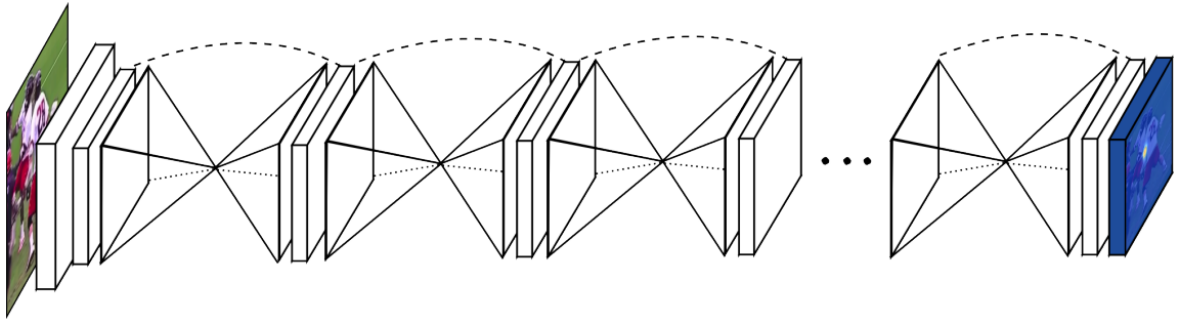


Рисунок 1.12 – Stacked Hourglass архітектура

Основна ідея архітектури Stacked Hourglass полягає у використанні послідовності модулів (див. рис. 1.13) для захоплення та уточнення просторової інформації в різних масштабах.

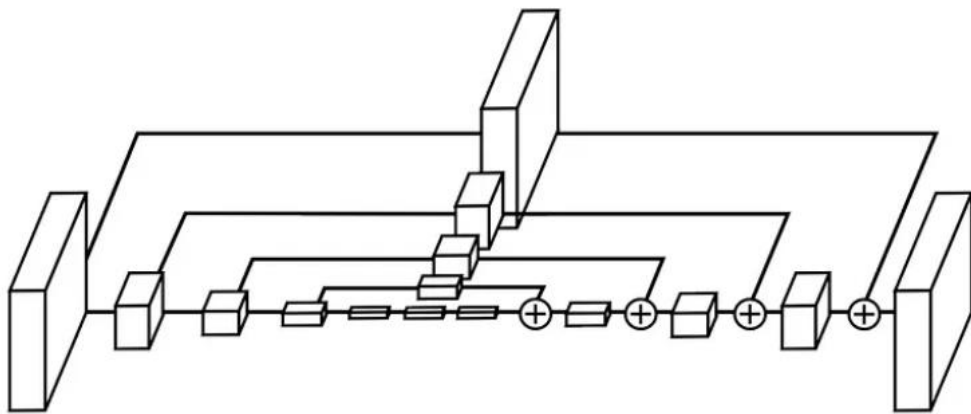


Рисунок 1.13 – Окремий модуль у Stacked Hourglass архітектурі

Кожен модуль складається з послідовності згорткових шарів і шарів об'єднання, які зменшують дискретизацію вхідного зображення, за якими йде послідовність шарів підвищення дискретизації та згорткових шарів, які реконструюють вихідну роздільну здатність. З'єднання пропуску в кожному модулі дозволяють мережі об'єднувати інформацію з різних масштабів і отримувати як локальну, так і глобальну інформацію [11].

Загалом одним із основних обмежень методу на основі теплової карти є його обчислювальна складність, що може бути проблемою для додатків реального часу. Щоб вирішити цю проблему, в деяких випадках використовують спрощену

мережеву архітектуру або методи скорочення, щоб зменшити кількість параметрів у мережі. Крім того, деякі методи використовують методи дистиляції знань при їх передачі від великої до меншої мережі.

1.1.3 Метод визначення пози людини «згори вниз»

Конвеєр підходу «згори вниз» (див. рис. 1.14) складається з двох важливих частин: детектор людського тіла для отримання обмежувальних рамок особи та оцінювач пози однієї людини для прогнозування розташування ключових точок у цих обмежувальних рамках.

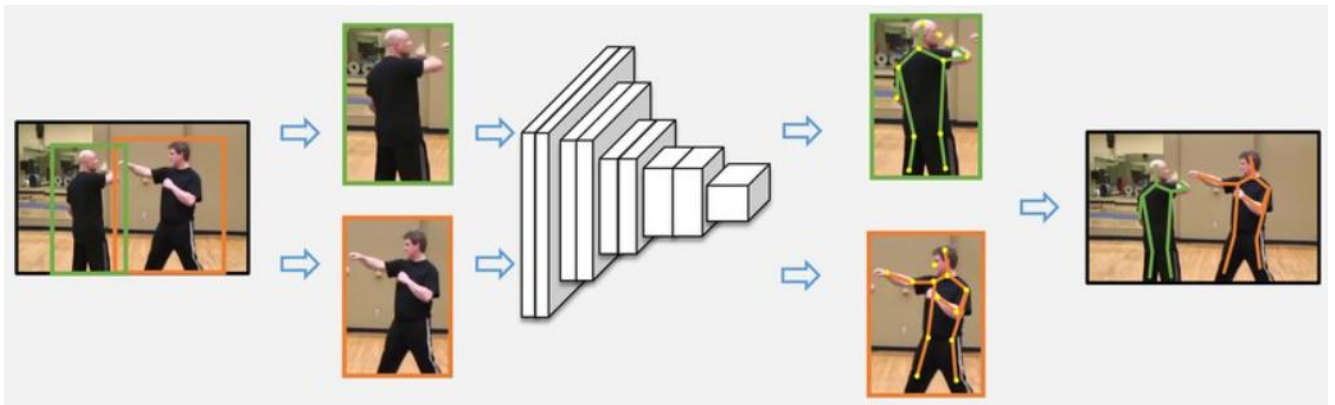


Рисунок 1.14 – Конвеєр підходу «згори вниз»

У цьому підході вхідне зображення спочатку обробляється детектором тіла, який зазвичай використовує алгоритми виявлення об'єктів, такі як Faster R-CNN або YOLO, щоб визначити місцезнаходження людини на зображенні [12].

Faster R-CNN – це двоетапний алгоритм виявлення об'єктів, який використовує мережу пропозицій регіону (RPN) для створення пропозицій об'єктів і окрему мережу для класифікації пропонованих регіонів (див. рис. 1.15).

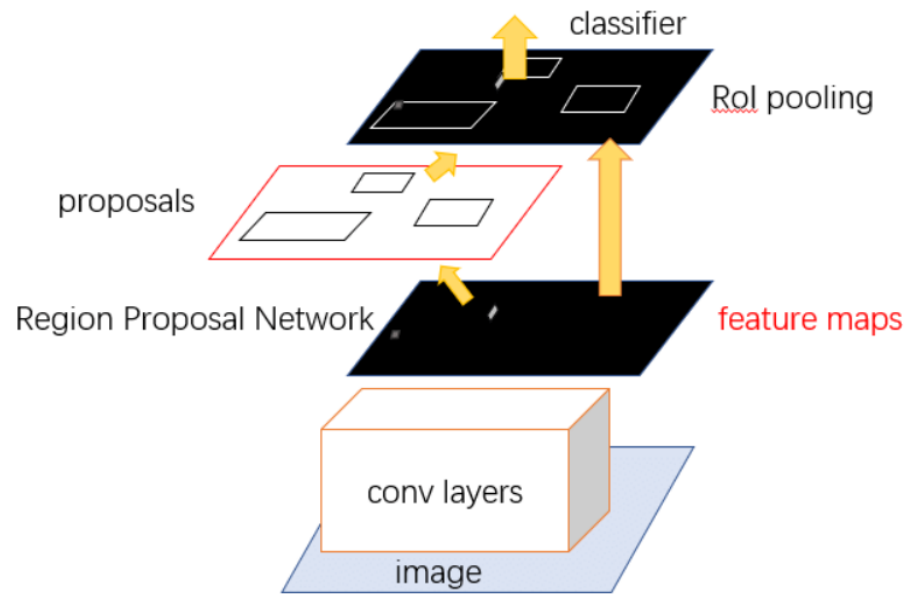


Рисунок 1.15 – Архітектура Faster R-CNN алгоритму

RPN – це повністю згорточна мережа, яка приймає зображення як вхідні дані та генерує набір пропозицій об'єктів, ковзаючи маленьким віконцем по зображенню в різних масштабах і пропорціях. Потім пропозиції передаються через мережу класифікації, яка виводить розподіл ймовірностей за різними класами об'єктів [13].

Перевага Faster R-CNN полягає в тому, що він є високоточним і може виявляти об'єкти в різних масштабах і співвідношеннях сторін. Однак це може потребувати значних витрат при обчисленні через його багатоступеневу архітектуру.

З іншого боку, YOLO – це одноетапний алгоритм виявлення об'єктів, який безпосередньо передбачає обмежувальні прямокутники та ймовірності класу для кожного об'єкта на зображенні. YOLO ділить зображення на сітку та передбачає обмежувальні прямокутники та ймовірності класу для кожної клітинки сітки. Перевагою YOLO є те, що він швидкий і може виявляти об'єкти в режимі реального часу. Однак він може бути не таким точним, як Faster R-CNN, особливо для невеликих об'єктів або об'єктів іншого масштабу.

Отже, після виявлення тіла одним із розглянутих алгоритмів область зображення, що містить людину, обрізається та пропускається через мережу оцінки

пози, яка локалізує окремі суглоби. Ця мережа зазвичай являє собою архітектуру на основі теплової карти, яка виводить карту ймовірностей для кожного місця з'єднання на зображенні.

Підходу «зверху вниз» часто віддають перевагу перед іншими стратегіями, наприклад підходом «знизу вгору», оскільки він забезпечує більш цілісне уявлення про тіло, дозволяючи точніше локалізувати суглоб. Однак це також може бути дорожчим з точки зору обчислень, особливо якщо детектор тіла складний і потребує багато обчислень. Одна з ключових переваг підходу «зверху вниз» полягає в тому, що він стійкий до оклюзій і захаращеного фону, оскільки детектор тіла часто може точно ідентифікувати особу, навіть якщо деякі частини тіла закриті.

1.1.4 Метод визначення пози людини «знизу вгору»

Підхід «знизу вгору» — це ще одна поширена стратегія для оцінки пози людини, яка передбачає спочатку виявлення окремих суглобів на зображенні, а потім групування їх разом для формування повної пози для кожної окремої людини (див. рис. 1.16).



Рисунок 1.16 – Конвеєр підходу «знизу вгору»

У цьому підході зображення спочатку обробляється мережею виявлення з'єднань, яка виводить місцезнаходження та оцінку достовірності для кожного окремого з'єднання на зображенні [9].

Мережа виявлення з'єднань зазвичай використовує архітектуру на основі теплової карти, яка виводить карту ймовірностей для кожного місця з'єднання на зображенні. Кожна карта ймовірностей вказує на ймовірність наявності суглоба в цьому місці на зображенні. Потім ці ймовірнісні карти використовуються для визначення розташування окремих суглобів на зображенні.

Після того, як суглоби виявлені, вони групуються разом, щоб сформувати повну позу. Цей процес групування може бути досягнутий за допомогою різних методів, таких як кластеризація, підходи на основі графів або поля спорідненості частин (PAF).

PAF є популярним підходом, який використовує набір векторів для кодування зв'язку між різними суглобами. Потім ці вектори використовуються для групування суглобів, щоб сформувати повну позу.

Перевага підходу «знизу вгору» полягає в тому, що він може точно виявляти складні пози та працювати з кількома людьми на зображенні. Він також є досить ефективним з точки зору обчислень, оскільки для виявлення всіх ключових точок на зображенні потрібна лише одна мережа.

Однак підхід «знизу вгору» може бути менш точним, ніж підхід «зверху вниз», особливо для виявлення прихованих суглобів або поз, які частково видно на зображенні.

1.2 Огляд наборів даних та метрик оцінки ефективності моделей визначення положень тіла

При створенні моделі для визначення пози людини в основному використовують наступні набори даних, які є у вільному доступі: FLIC, LSP, MPII Human Pose та COCO. Ці анотовані набори даних містять істинні положення ключових точок, які можна використовувати для навчання та розрахунку різних метрик ефективності моделі.

Деякі найбільш популярні набори даних та метрики ефективності будуть розглянуті далі.

1.2.1 Набір даних FLIC

Набір даних FLIC (Frames Labeled In Cinema) – це широко використовуваний набір даних для оцінки алгоритмів визначення пози людини. Він містить 5003 кадри, витягнуті з 235 оригінальних відеокліпів, із широким спектром людських поз у природних умовах [14] (див. рис. 1.17). Цей набір даних призначений для оцінки алгоритмів визначення природних, складних поз людини з реалістичним фоном, і вже став стандартним еталоном.

Кожен кадр у FLIC супроводжується координатами 10 суглобів верхньої частини тіла у двовимірному просторі.



Рисунок 1.17 – Приклад даних із набору FLIC

Цей набір даних використовувався в численних дослідженнях, і було показано, що він є досить складним набором даних. Він й досі використовується для оцінки широкого діапазону підходів, включаючи методи глибокого навчання та традиційні методи комп'ютерного зору. Набір даних часто використовується в поєднанні з такими метриками, як точність і середня похибка положення суглоба (MPJPE), щоб оцінити ефективність різних моделей.

1.2.2 Набір даних MPII Human Pose

Набір даних MPII – це ще один широко використовуваний еталонний набір даних для оцінки алгоритмів визначення пози людини. Він містить понад 25 000 зображень людей у природному середовищі з анотованими двовимірними та тривимірними положеннями суглобів. Набір даних було зібрано з відео YouTube, і в ньому представлені люди різного віку, розміру, в різноманітному одязі, а також широкий спектр видів діяльності та точок зору (див. рис. 1.18).



Рисунок 1.18 – Приклад даних із набору МРІІ

Кожне зображення в наборі даних МРІІ супроводжується двовимірними положеннями 16 суглобів тіла, включаючи голову, шию, плечі, лікті, зап'ястки, стегна, коліна та щиколотки [9]. Крім того, набір даних також надає тривимірні позиції ключових точок для підмножини зображень. Анотації були отримані за допомогою напівавтоматичного інструменту анотування, який поєднував ручне анотування з автоматичним виявленням.

1.2.3 Набір даних СОСО

Набір даних СОСО став популярним вибором для дослідження оцінки пози людини завдяки своєму великому розміру, різноманітності зображень і поз, а також високій якості анотацій. Зображення в наборі даних СОСО були зібрані з різних

джерел, включаючи Flickr, і охоплюють широкий діапазон сценаріїв, наприклад сцени в приміщенні та на вулиці, різні умови освітлення та фон (див. рис. 1.19).



Рисунок 1.19 – Приклад даних із набору COCO

Цей набір даних містить понад 200 000 зображень та більше 250 000 поз людей, а кожна поза супроводжується координатами 17 ключових точок [15].

Анотації для набору COCO були створені за допомогою підходу краудсорсингу, коли працівників просили анотувати розташування ключових точок людей на зображеннях. Потім анотації були переглянуті та виправлені експертами для забезпечення їх точності. Ці анотації часто використовуються для навчання згорткових та рекурентних нейронних моделей.

Для оцінки ефективності алгоритмів визначення пози людини набір даних COCO зазвичай поєднується з такими метриками, як «average precision» (AP) і «mean average precision» (mAP).

1.2.4 Метрика ефективності OKS

Індекс подібності ключових точок об'єкта (OKS) – це показник, який використовується в задачах виявлення об'єктів і локалізації ключових точок, наприклад, у завданнях СОСО. Він вимірює подібність між передбачуваними ключовими точками (наприклад, орієнтирами на обличчі, розташуванням суглобів) і базовими ключовими точками істинності для певного об'єкта на зображенні.

Показник OKS враховує як відстань між передбачуваним і істинним розташуванням ключових точок, так і масштаб людини на зображенні. Зокрема, він розраховує оцінку подібності як функцію евклідової відстані між передбачуваним і базовим розташуванням ключових точок істинності, нормалізованого за масштабом людини на зображенні згідно з формулою (1). Масштаб визначається як середнє значення ширини та висоти обмежувальної рамки людини.

$$OKS = \exp\left(-\frac{d_i^2}{2s^2k_i^2}\right), \quad (1)$$

У наведеній вище формулі d_i – евклідова відстань між реальною ключовою точкою та оціненою ключовою точкою, s – масштаб: площа граничного поля, поділена на загальну площу зображення, k – константа, яка визначається окремо для кожного контрольного пункту.

Більш високе значення OKS вказує на те, що прогнозовані ключові точки більш схожі на істинні ключові точки, а виявлення об'єкта та локалізація ключових точок вважаються більш точними. Індекс OKS коливається від 0 до 1, причому значення 1 вказує на ідеальне перекриття.

1.2.5 Метрика ефективності «average precision» (AP) та її варіант «mean average precision» (mAP)

«Average precision» – це ще один загальнозживаний показник ефективності, який вимірює точність передбачуваних розташувань ключових точок порівняно з розташуваннями ключових точок істинності на зображенні.

Щоб обчислити AP, прогнозовані ключові точки зіставляються з наземними ключовими точками істинності з використанням певного порогового значення відстані. Передбачувана ключова точка вважається істинно позитивною (TP), якщо її відстань до найближчої наземної контрольної точки менша за порогове значення. Передбачувана ключова точка вважається хибно позитивною (FP), якщо в межах порогу немає відповідної базової ключової точки істинності. Основна істинна ключова точка вважається хибно негативною (FN), якщо немає відповідної прогнозованої ключової точки.

Використовуючи ці підрахунки TP, FP і FN, можна обчислити «precision» і «recall» для різних порогів відстані. «Precision» – це відношення TP до суми TP і FP, а «recall» – відношення TP до суми TP і FN.

Крива «precision-recall» будується шляхом побудови графіка залежності «precision» від «recall» для різних порогових значень відстані. Потім AP обчислюється як площа під цією кривою.

В той же час «mean average precision» – це підсумкова статистика, яка усереднює значення AP для всіх частин тіла або поз. Іншими словами, це середнє значення AP для кожної частини або пози, і воно часто використовується для порівняння загальної ефективності різних моделей оцінки пози [16].

У задачах визначення пози людини mAP зазвичай використовується для оцінки продуктивності моделей глибокого навчання на великих наборах даних, таких як вже розглянуті COCO або MPE.

1.2.6 Метрика ефективності РСК та її варіант РСКh

РСК (відсоток правильних ключових точок) – це загальнозживаний показник оцінки в задачах комп'ютерного зору та оцінки поз. Він вимірює точність алгоритму оцінки пози шляхом порівняння передбачуваних ключових точок із істинними ключовими точками в наборі даних.

РСК визначається як відсоток передбачуваних ключових точок, які знаходяться в межах певної порогової відстані від істинних ключових точок. Порогова відстань зазвичай встановлюється як частка розміру зображення або розміру частини тіла, що оцінюється.

Наприклад, припустімо, що проводиться оцінка алгоритму для визначення пози на наборі даних зображень людського тіла. Набір даних містить базові ключові точки істинності для кожного зображення, наприклад положення суглобів або частин тіла. Для оцінки точності алгоритму спочатку використовують його для оцінки ключових точок для кожного зображення в наборі даних. Потім порівнюються передбачувані ключові точки з основними ключовими точками правдивості за допомогою РСК.

Для обчислення РСК підраховується кількість передбачуваних ключових точок, які знаходяться в межах порогової відстані відповідних ключових точок істинності, і ділиться на загальну кількість ключових точок. Отримане значення є відсотком правильних ключових точок, або РСК.

РСК є корисним показником, оскільки він забезпечує кількісну міру точності алгоритму оцінки пози. Його можна використовувати для порівняння продуктивності різних алгоритмів на одному наборі даних або для оцінки продуктивності одного алгоритму з часом у міру його вдосконалення чи модифікації.

Основна відмінність між РСК і РСКh полягає в тому, як вони обробляють варіації розміру частини тіла. РСК вимірює точність алгоритму оцінки пози на основі фіксованої порогової відстані, яка зазвичай є часткою розміру зображення.

Цей фіксований поріг не є ідеальним, якщо розмір голови чи інших частин тіла може суттєво відрізнятись на різних зображеннях, що може ускладнити порівняння результатів між наборами даних або відстеження прогресу з часом.

Щоб усунути це обмеження, PCKh нормалізує порогову відстань на основі розміру голови на кожному зображенні. Це означає, що порогова відстань пропорційна розміру голови, що допомагає врахувати варіації розміру та положення тіла на різних зображеннях. Зокрема, порогова відстань обчислюється як фіксована частка розміру голови на кожному зображенні, а не як фіксована частка розміру зображення [14].

Наприклад, якщо порогову відстань встановлено на 0,1 розміру голови, оцінена ключова точка вважатиметься правильною, якщо її відстань до основної точки правдивості на землі становить менше 10% від розміру голови на зображенні. Ця нормалізація дозволяє більш узгоджено оцінювати різні зображення та набори даних і може допомогти забезпечити більш точне вимірювання продуктивності алгоритму.

Підсумовуючи, і PCK, і PCKh є мірами точності алгоритмів оцінки пози, а PCKh враховує варіації розміру тіла та положення на різних зображеннях шляхом нормалізації порогової відстані на основі розміру голови на кожному зображенні.

1.3 Огляд методів порівняння положень тіла

Порівняння поз – це процес порівняння двох або більше положень людського тіла для визначення їх подібності чи відмінності.

Загалом існує кілька методів порівняння поз, починаючи від простих вимірювань відстані до більш складних методів машинного навчання. Одним із найпоширеніших підходів є представлення пози у вигляді набору ключових точок, а потім порівняння цих ключових точок між позами. Це можна зробити за

допомогою мір відстані, таких як евклідова відстань, або мір подібності, таких як косинус – подібність [17].

Інші методи порівняння поз включають використання статистичних моделей, таких як «Principal Component Analysis» (PCA) або «Dynamic Time Warping» (DTW), які можуть охопити основну структуру даних поз і дозволити більш точне порівняння.

Кожен із цих методів має свої сильні та слабкі сторони, і вибір методу залежить від конкретного завдання.

Окрім того, підвищити точність і надійність результатів порівняння шляхом зменшення шуму та неузгодженості може попередня обробка даних – серія кроків очищення та нормалізації даних, які застосовуються до необроблених даних пози перед тим, як вони будуть використані в алгоритми порівняння.

До основних методів попередньої обробки можна віднести нормалізацію, згладжування, вирівнювання та відбір ознак.

У наступних підрозділах будуть більш детально розглянуті деякі з наведених методів порівняння положень тіла.

1.3.1 Порівняння положень тіла за допомогою евклідової відстані

Евклідова відстань є мірою прямолінійної відстані між двома точками в евклідовому просторі. У контексті порівняння двох положень тіла людини представлених у вигляді наборів ключових точок евклідова відстань може бути використана для вимірювання подібності між цими двома наборами та розраховується згідно з формулою (2).

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

У наведеній вище формулі d – евклідова відстань, n – розмірність векторів, x , y – відповідні координати двох векторів у площині вимірювання i .

Одним із недоліків використання евклідової відстані для порівняння ключових точок пози є те, що вона не враховує різницю в орієнтації чи масштабуванні пози. Наприклад, якщо одна людина знаходиться ближче до камери, ніж інша, її ключові точки виглядатимуть більшими на зображенні та будуть далі одна від одної на евклідовій відстані. Щоб врахувати це, ми можемо нормалізувати відстані, наприклад, за допомогою L2 нормалізації, щоб гарантувати, що вектори мають одиничну довжину, а косинусна відстань чітко визначена.

Нормалізація L2 передбачає ділення кожного вектора на його норму L2, яка є квадратним коренем із суми квадратів елементів у векторі (див. рис. 1.20). Це масштабує кожен вектор до одиничної довжини та гарантує, що косинусна відстань між двома векторами коливається від 0 до 1, де 0 означає, що вектори ідентичні, а 1 означає, що вони ортогональні або зовсім несхожі.

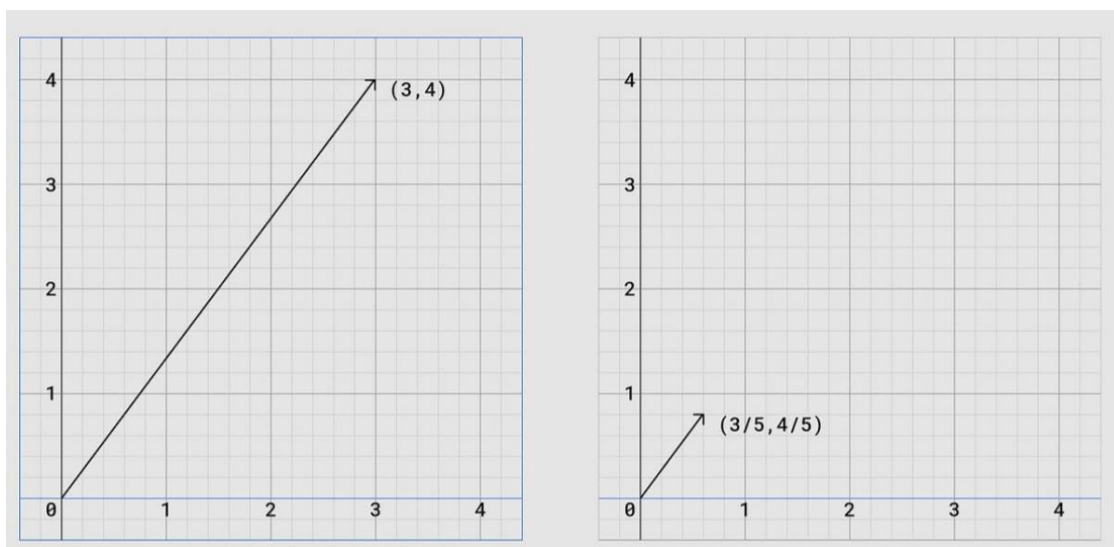


Рисунок 1.20 – приклад L2 нормалізації вектора

У випадку порівняння ключових точок пози людини ми можемо представити кожен позу як вектор, об'єднавши координати ключових точок. Щоб переконатися, що вектори мають однакову розмірність, нам може знадобитися доповнити вектори нулями або скоротити їх до фіксованої довжини.

Коли у нас є вектори, що представляють пози, ми можемо виконати нормалізацію L2 для кожного вектора, а потім обчислити косинусну подібність між нормалізованими векторами. Косинусна подібність визначається як косинус кута між двома векторами, і її можна обчислити за допомогою скалярного добутку двох векторів, поділеного на їхні величини згідно з формулою (3).

$$\cos \theta = \frac{\vec{A} \times \vec{B}}{\|\vec{A}\| \times \|\vec{B}\|} \quad (3)$$

Використовуючи величину косинусної подібності, можна обчислити величину відстані між векторами згідно з формулою (4).

$$D(F_{xy}, G_{xy}) = \sqrt{2 * (1 - cosineSimilarity(F_{xy}, G_{xy}))} \quad (4)$$

В результаті отримуємо значення, через яке можна оцінити схожість позицій. Чим менше це значення, тим більше схожі пози. Отримана подібність коливається від -1, що означає повну протилежність, до 1, яка означає однакові пози, а 0 вказує на ортогональність або декореляцію, тоді як значення між ними вказують на проміжну подібність або відмінність.

1.3.2 Порівняння положень тіла за допомогою зваженої відстані

Значним недоліком попередньо розглянутого методу порівняння поз за допомогою евклідової відстані є те, що він ніяк не враховує під час обчислень оцінку достовірності, пов'язану з кожною ключовою точкою.

У більшості сучасних моделей для визначення пози людини кожна ключова точка пов'язана з оцінкою достовірності, яка відображає надійність виявлення ключової точки. Показники достовірності коливаються від 0 до 1, де 1 означає високу достовірність виявлення, а 0 означає низьку достовірність або відсутність ключової точки.

Таким чином, попередній підхід можна вдосконалити за допомогою оцінки достовірності ключових точок, де вищий показник достовірності матиме більший вплив на відстань, а нижчий показник достовірності матиме менший ефект відповідно [18].

Отже, відстань з урахуванням показника достовірності може бути підрахована згідно з формулою (6).

$$D(F, G) = \frac{1}{\sum_{k=1}^n F_{c_k}} * \sum_{k=1}^n F_{c_k} * |F_{x_{y_k}} - G_{x_{y_k}}| \quad (5)$$

У наведеній вище формулі G і F – два вектори поз після $L2$ нормалізації, n – кількість визначених контрольних точок, F_{c_k} – значення ймовірності знаходження правильного з'єднання для елемента з номером k вектора F , $F_{x_{y_k}}$ і $G_{x_{y_k}}$ – положення x і y k -ї ключової точки для кожного вектора.

1.3.3 Порівняння положень тіла за допомогою DTW

Наведені у попередніх підрозділах метрики можуть бути легко використані для порівняння кількох статичних поз людей, але якщо ми хочемо порівняти відмінності між двома людьми, що виконують певну дію, нам потрібно порівняти послідовності поз для визначення схожості чи відмінності виконання дії. Послідовності зазвичай мають різну довжину, тож лише певні періоди, де виконується дія, повинні бути використані у порівнянні. Таким чином головною проблемою в цій задачі є вирівнювання початкових поз для обох послідовностей.

Просте вирішення цієї проблеми полягає у використанні підходу ковзаючого вікна, коли вікно захоплює кадри виконання дії з першої послідовності та знаходить вікно в другій послідовності. Вікно буде ковзати по другій послідовності, щоб знайти вікно кадрів, де загальна відстань між позами буде мінімальна (див. рис. 1.21).

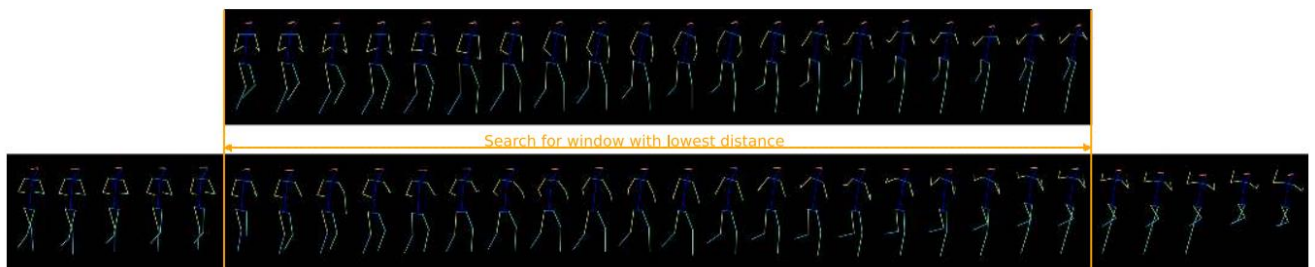


Рисунок 1.21 – Приклад роботи ковзаючого вікна

За таким принципом працює Dynamic Time Warping (DTW) – метод для порівняння послідовностей даних, які можуть бути схожими, але мають різну довжину або часові варіації. Він широко використовується в області комп’ютерного зору для порівняння дій людини, де його можна використовувати для порівняння послідовностей ключових точок пози людини, отриманих з різних відео або кадрів.

Щоб застосувати DTW для порівняння, кожна дія представляється у вигляді послідовності координат ключових точок у часі. Таким чином кожен дію можна представити у вигляді матриці, де кожен рядок представляє координати ключової точки на певному часовому етапі.

Потім ми порівнюємо дві послідовності за допомогою DTW, що передбачає пошук оптимального вирівнювання між двома послідовностями шляхом деформації їхніх осей часу. Оптимальне вирівнювання визначається шляхом знаходження шляху через матрицю витрат, який мінімізує загальну вартість викривлення.

Ціна викривлення зазвичай визначається як евклідова відстань між координатами ключових точок у кожній точці шляху вирівнювання. У контексті порівняння поз функція вартості може також включати додаткові обмеження, такі як гладкість або просторова когерентність, залежно від конкретного застосування.

Однією з переваг використання DTW для порівняння дій людини є те, що він може обробляти варіації в часі та тривалості, які можуть виникнути через різницю в частоті кадрів відео або швидкості руху людини. DTW здатний вирівнювати послідовності, деформуючи їх часові осі, що дозволяє точно порівнювати послідовності поз, які мають різну часову динаміку.

Порівняно з евклідовою або косинусною відстанню (див. рис. 1.22), DTW може бути кращим вибором для порівняння дій, коли послідовності мають різну довжину або часові варіації.

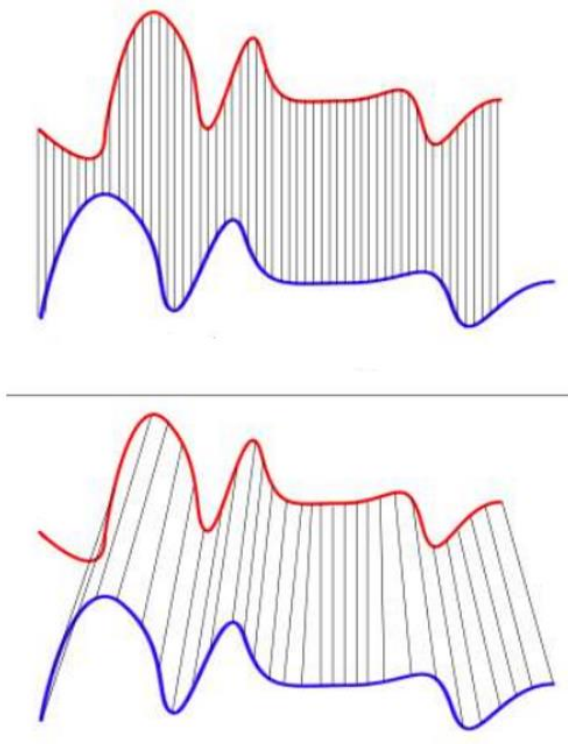


Рисунок 1.22 – Приклад порівняння двох послідовностей за допомогою евклідової відстані (вгорі) та DTW (внизу)

DTW може фіксувати часову структуру послідовності і обробляти часові спотворення або варіації, які можуть бути недоступні за допомогою простих показників відстані.

Однак DTW може бути обчислювально дорогим і потребувати значних обчислювальних ресурсів для великомасштабних завдань порівняння поз.

1.4 Огляд моделей глибокого навчання для визначення положень тіла

1.4.1 Огляд моделі OpenPose

OpenPose – це бібліотека програмного забезпечення з відкритим вихідним кодом для виявлення ключових точок у режимі реального часу для кількох осіб і багатопотокової обробки, розроблена Університетом Карнегі – Меллона та Центром бачення та навчання Берклі. Вона використовує методи глибокого

навчання та комп'ютерного зору, щоб визначити 2D або 3D позу однієї чи кількох людей на зображенні чи відео.

Основою в архітектурі цієї моделі (див. рис. 1.23) є глибока нейронна мережа VGG-19, що створює оригінальну карту ознак вхідних даних та розділяє їх на дві гілки для подальшої обробки. Перша гілка використовує CNN для прогнозування теплової карти суглобів тіла людини, тоді як друга використовує CNN для отримання полів часткової спорідненості (PAF) для кожного суглоба. PAF представлено двовимірним вектором, який зберігає орієнтацію та положення кінцівки. На кожному етапі враховується зіставлення PAF із вхідним шаром ознак і тепловою картою ключових точок, що відоме як S^t і L^t . Отримана тепла карта та PAF на етапах прогнозування дають інформацію про положення та вектор напрямку всіх ключових точок. Для оптимізації та узгодження кінцівок використовується Угорський жадібний алгоритм [19].

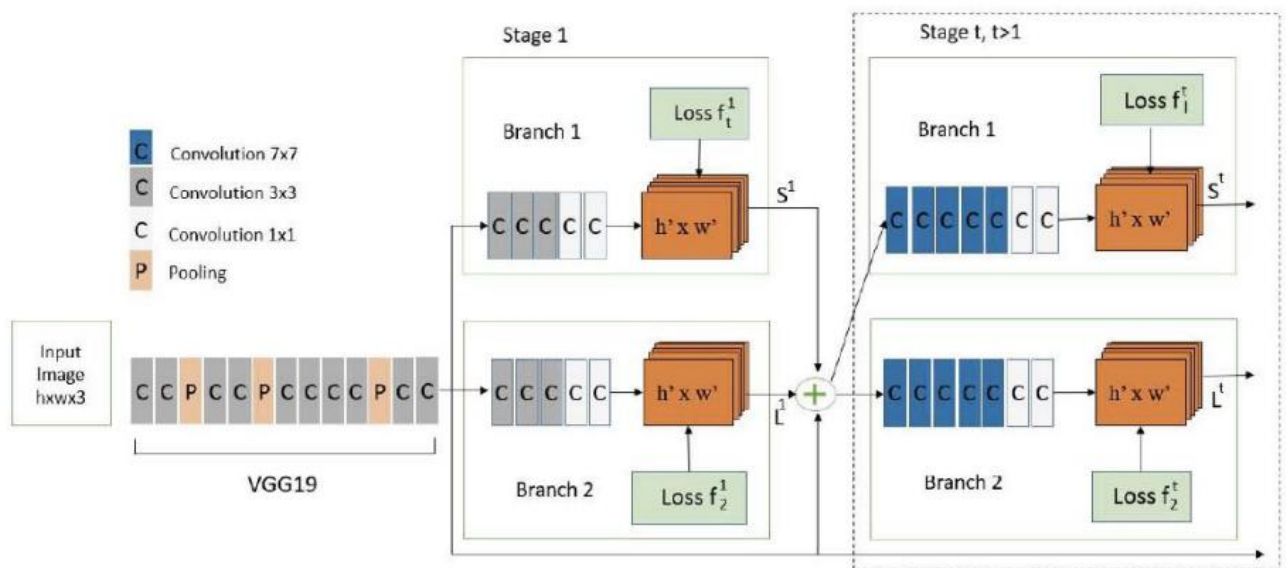


Рисунок 1.23 – Архітектура моделі OpenPose

Таким чином, конвеєр моделі OpenPose (див. рис. 1.24) можна поділити на наступні етапи:

- вхідний: необроблене зображення або відеокادر попередньо обробляються шляхом зміни розміру, обрізання та нормалізації даних, щоб підготувати їх для подальшої обробки;

– магістральний: CNN вилучає характеристики високого рівня з вхідних даних. Цей етап складається з кількох згорткових шарів, за кожним із яких іде нелінійна функція активації та шар об'єднання, щоб поступово зменшувати просторову роздільну здатність даних і збільшувати їх глибину;

– етап багатоступеневої агрегації ознак: на цьому етапі відбувається агрегація ознак, отриманих на етапі магістралі, у кількох масштабах і роздільній здатності. Це робиться за допомогою кількох паралельних підмереж, які обробляють карти ознак у різних масштабах і роздільній здатності та поєднують їх для створення єдиного набору.

– етап виявлення ключових точок: цей етап використовує набір підмереж для виявлення ключових точок на тілі людини. Він складається з кількох паралельних підмереж, які генерують набір ключових точок – кандидатів для кожної частини тіла, а потім підмереж уточнення, які уточнюють оцінки ключових точок на основі карт ознак.

– етап оцінки поля спорідненості частин: цей етап оцінює зв'язки між різними частинами тіла за допомогою PAF.

– етап оцінки пози: етап оцінки пози поєднує виявлені ключові точки та оцінені поля спорідненості частин, щоб створити граф людського тіла та оцінити 2D або 3D позу кожної людини на зображенні чи відео. На цьому етапі використовуються методи оптимізації графа для коригування зв'язків між ключовими точками на основі оціненої пози.

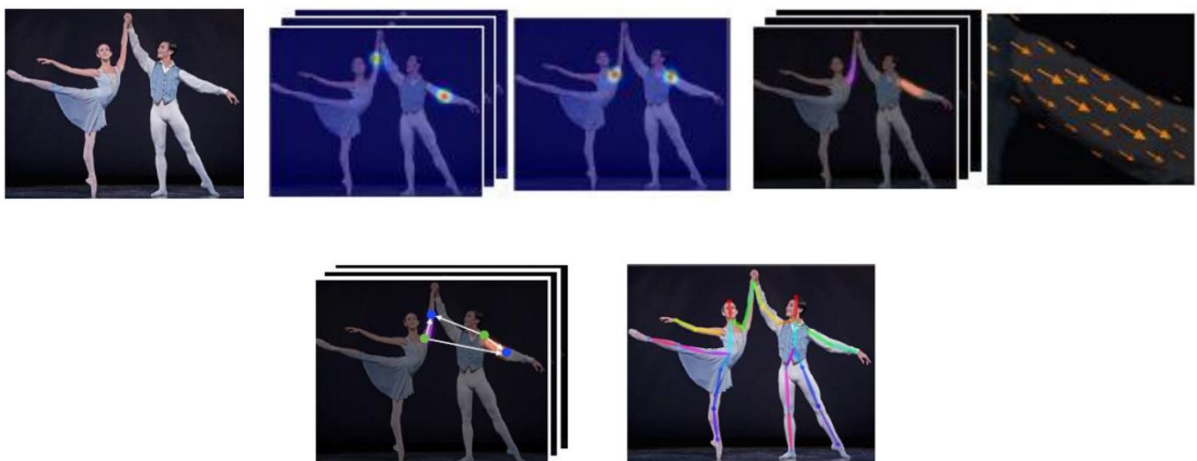


Рисунок 1.24 – Конверс моделі OpenPose

Модель OpenPose оптимізована для досягнення високої продуктивності в реальному часі та може працювати на різноманітних апаратних платформах, включаючи CPU, GPU та FPGA.

Бібліотека має широкі можливості налаштування та дозволяє користувачам змінювати різні параметри, такі як поріг виявлення, кількість ключових точок для виявлення та архітектуру нейронної мережі. Вона також надає різноманітні вихідні формати, включно з положеннями ключових точок, балами достовірності оцінки пози та графічним представленням виявленої пози, накладеної на вхідне зображення чи відео [20].

OpenPose може виявляти та відстежувати кількох людей одночасно, навіть коли вони взаємодіють один з одним, закривають один одного або їх частково видно на зображенні чи відео. Вона також може оцінювати орієнтацію та рух кожної частини тіла, що робить її придатною для широкого спектру програм.

1.4.2 Огляд моделі PoseNet

PoseNet – це модель глибокого навчання, розроблена дослідниками Google, яка може визначати позу як однієї, так і багатьох людей на зображеннях чи відео в режимі реального часу [20].

Модель GoogleNet спочатку була розроблена для класифікації та виявлення об'єктів. Вона є глибокою нейронною мережею на основі модуля, відомого як Inception (див. рис. 1.25).

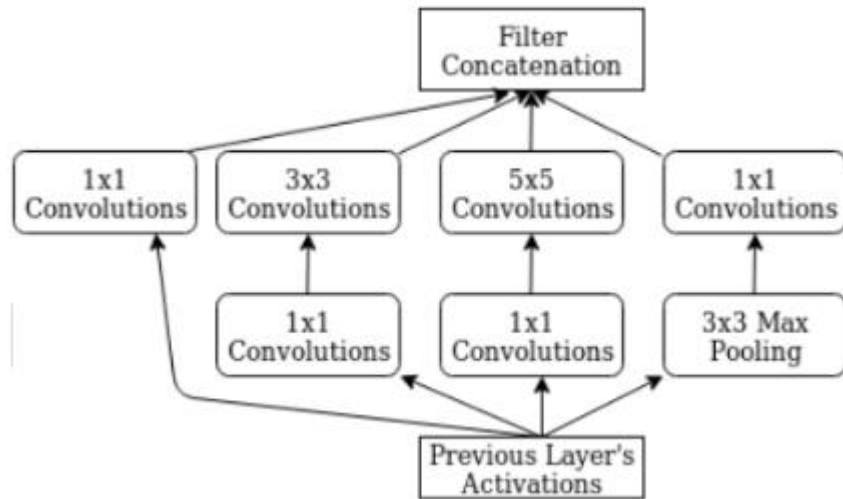


Рисунок 1.25 – Inception модуль

GoogleNetта приймає як вхідне зображення 224×224 пікселів і поширює його через 9 початкових модулів, розташованих один на одному, використовуючи Rectified Linear Units (ReLU) як функцію активації. Кожен рівень у такій мережі вивчає подальшу абстракцію вхідних даних. Абстракція найвищого рівня, яка знаходиться на останньому рівні мережі, разом із двома проміжними абстракціями передається на максимальний рівень для прогнозування класів об'єктів.

PoseNet замінює цей рівень класифікації на два паралельні рівні (див. рис. 1.26). Вони регресують до пози (представленої (x, y, z) координатами) та орієнтації (представленої у вигляді кватерніону).

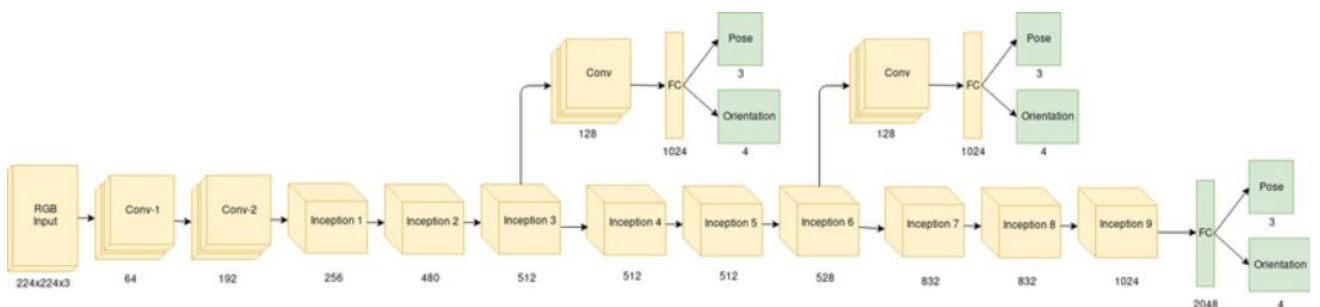


Рисунок 1.26 – Архітектура моделі PoseNet

Жовті блоки представляють попередньо навчені модулі, які PoseNet успадковує від GoogleNet. Зелені блоки показують спеціальні модулі PoseNet, які потрібно навчити з нуля [21].

Один із підходів до підвищення точності PoseNet полягає в поєднанні його з комірками LSTM.

LSTM (Long Short-Term Memory) – це тип рекурентної нейронної мережі, яка добре підходить для моделювання даних часових рядів. Комірki LSTM призначені для фіксації довготривалих залежностей у вхідній послідовності та можуть допомогти підвищити точність моделі шляхом включення інформації з минулих кадрів.

Щоб удосконалити PoseNet із комірками LSTM, можна використовувати вихідний вектор ключових точок пози, створений PoseNet для кожного кадру, як вхідні дані для комірки LSTM. Потім комірка LSTM може обробляти послідовність ключових точок пози в кількох кадрах, щоб зафіксувати тимчасові залежності та створити більш точну оцінку пози.

Одним із підходів до реалізації PoseNet з осередками LSTM є використання архітектури TCN. TCN – це тип згорткової нейронної мережі, яка розроблена для ефективної обробки даних часових рядів. TCN поєднує згорткові шари з розширеними звивинами, і їх можна навчити наскрізно за допомогою стандартних методів зворотного поширення.

В архітектурі TCN для PoseNet із комірками LSTM вихідний вектор ключових точок пози для кожного кадру спочатку обробляється набором згорткових шарів для вилучення ознак. Потім об'єкти подаються в стек комірок LSTM, які обробляють послідовність ключових точок пози в кількох кадрах. Потім вихідні дані комірок LSTM подаються в набір повністю з'єднаних шарів для отримання остаточної оцінки пози.

Включивши комірki LSTM у PoseNet, можна підвищити її точність у задачах оцінки пози, особливо в ситуаціях, коли часова інформація важлива, наприклад у програмах захоплення руху або розпізнавання дій. Однак важливо ретельно оцінити продуктивність моделі та налаштувати гіперпараметри, щоб забезпечити оптимальну продуктивність для конкретного завдання.

1.4.3 Огляд моделі BlazePose

BlazePose – це модель оцінки пози в реальному часі, розроблена Google, яка здатна точно оцінювати 33 2D і 3D ключові точки людського тіла з одного RGB-зображення або відеопотоку. Вона базується на легкій та ефективній архітектурі, яка дозволяє працювати на широкому спектрі пристроїв, включаючи смартфони, планшети та ноутбуки.

Конвеєр моделі BlazePose складається з детектора осіб, за яким слідує мережа оцінки пози (див. рис. 1.27).

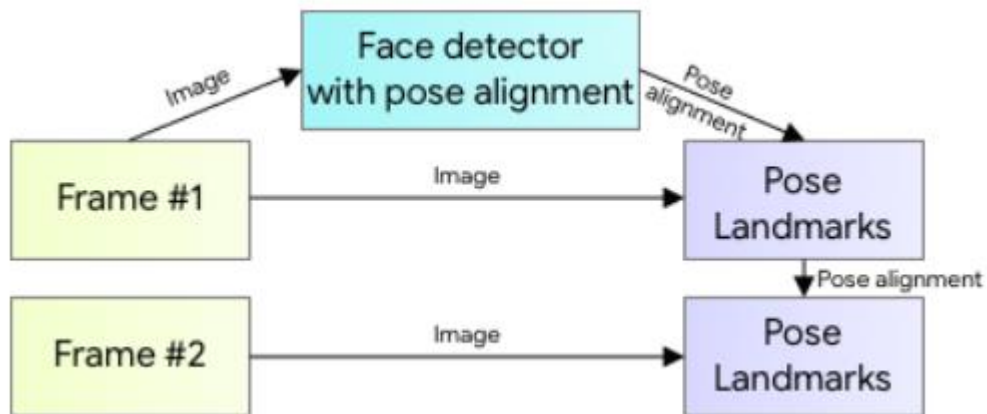


Рисунок 1.27 – Конвеєр моделі BlazePose

Мережа оцінки пози передбачає координати ключових точок, присутність людини на поточному кадрі та уточнену область інтересу для поточного кадру. Коли мережа показує, що людини немає, детектора осіб повторно запускається на наступному кадрі [22].

Компонент оцінки пози моделі прогнозує розташування 33 ключових точок людини (див. рис. 1.28) і використовує пропозицію щодо вирівнювання осіб, надану на першому етапі конвеєру.

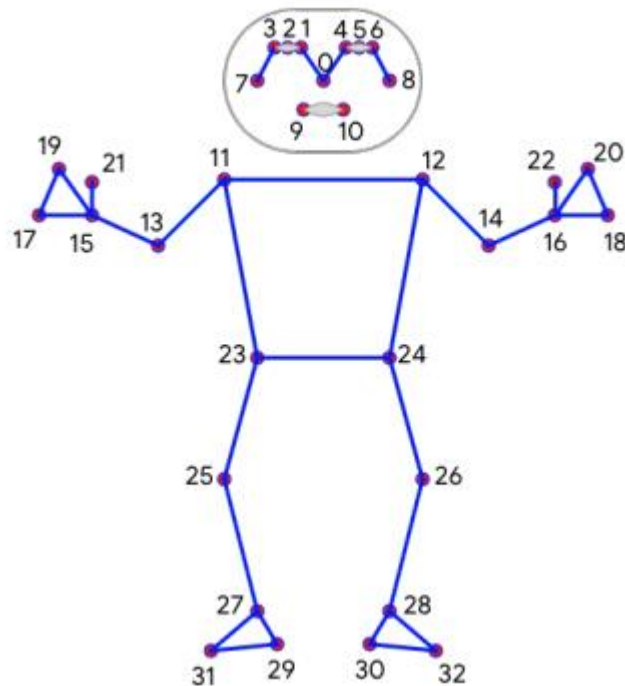


Рисунок 1.28 – Топологія моделі BlazePose з 33 ключовими точками

Крім того, детектор осіб у моделі BlazePose також має певні відмінності у порівнянні з іншими моделями глибокого навчання. Так, більшість сучасних рішень для виявлення об'єктів покладаються на алгоритмі немаксимального придушення (NMS) на останньому етапі постобробки.

Це добре працює для жорстких об'єктів з невеликим ступенем свободи. Однак цей алгоритм погано працює для сценаріїв зі складними позами людей, коли вони, наприклад, махають або обіймаються.

Це пояснюється тим, що кілька неоднозначних блоків задовольняють порогове значення перетину через об'єднання (IoU) для алгоритму NMS. Щоб подолати це обмеження, детектор моделі BlazePose зосереджується на виявленні обмежувальної рамки відносно стійкої частини тіла, як то обличчя або тулуб. Дослідники визначили, що в багатьох випадках найкращим варіантом сигналізувати нейронній мережі про наявність людини у кадрі є наявність обличчя людини (оскільки воно має висококонтрастні риси та має менше варіацій зовнішнього вигляду).

Як наслідок, модель використовує швидкий детектор обличчя як проксі для детектора осіб. Цей детектор обличчя передбачає додаткові параметри вирівнювання, характерні для людини: середня точка між стегнами людини, розмір кола, що охоплює всю особу, і нахил (кут між лініями, що з'єднують дві точки середини плечей і середини стегон).

Окремо увагу слід приділити архітектурі даної моделі (див. рис. 1.29). В BlazePose використовується комбінований підхід теплової карти, зміщення та регресії. Теплова карта та зсув втрат використовуються лише на етапі навчання моделі та відповідні вихідні шари видалені з моделі.

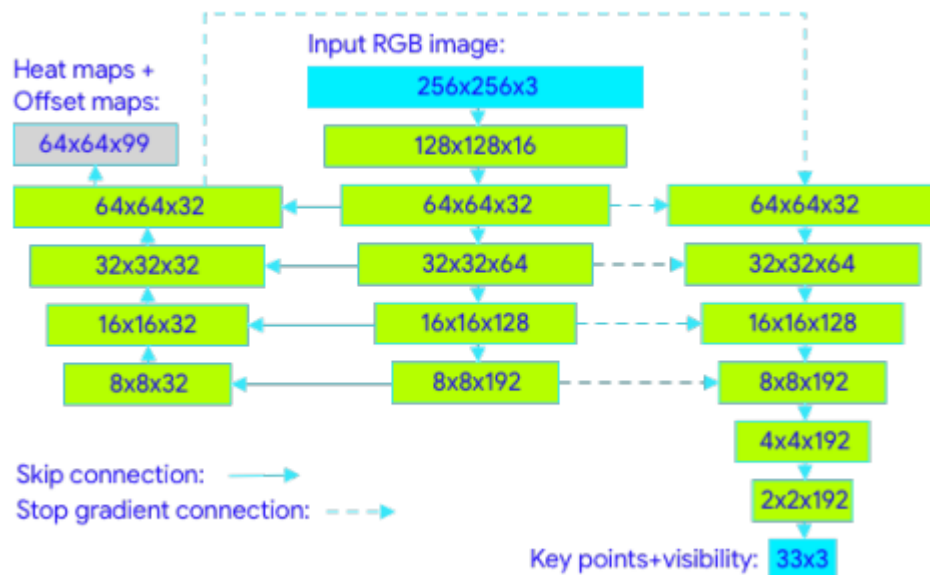


Рисунок 1.29 – Архітектура моделі BlazePose

Таким чином, теплова карта використовується максимально ефективно. Цей підхід частково натхненний "stacked hourglass" архітектурою, але в випадку BlazePose розробники складають крихітну мережу на основі теплової карти кодера-декодера та подальшу мережу регресійного кодера.

Крім того skip-connections активно використовуються між усіма етапами мережі, щоб досягти балансу поміж ознаками високого та низького рівня. Однак градієнти від регресійного кодера не поширюються назад на навчені теплові карти. Це не лише покращує передбачення теплової карти, а також значно підвищує точність регресії координат.

2 ПОСТАНОВКА НАУКОВО-ТЕХНІЧНОЇ ЗАДАЧІ

Проблема, що розглядається в даному дослідженні, є комплексною та потребує поступового вирішення ряду задач.

По-перше, необхідно проаналізувати наявні теоретичні відомості та публікації, що містять інформацію про точність і ефективність існуючих моделей визначення пози тіла людини, та вибрати для подальших досліджень найкращу за наведеними показниками.

Порівнюючи різні моделі слід враховувати наступні фактори:

- точність, з якою модель визначає положення ключових точок;
- здатність моделі працювати з різними умовами навколишнього середовища;
- обчислювальні ресурси, необхідні для запуску моделі;
- кількість ключових точок, що можуть бути визначені;
- швидкість визначення ключових точок.

При аналізі точності обраних для аналізу моделей необхідно звернути увагу на такі метрики, як:

- OKS;
- PCK або PCKh;
- AP або mAP.

Використовуючи кілька показників, можна отримати більш повну оцінку точності моделі.

Як результат найбільш ефективна модель визначення поз повинна бути обрана для подальшого дослідження.

Наступною задачею є визначення найбільш точного методу порівняння отриманих за допомогою моделі машинного навчання поз.

Як показник ефективності для кожного методу порівняння поз, оцінка F1 повинна бути розрахована.

Оцінка F1, також відома як F-міра, є загальноживаним показником для оцінки точності та відображає гармонійне середнє значення між precision (часткою справжніх позитивних результатів (TP) серед усіх прогнозованих позитивних результатів) та recall (часткою TP серед усіх фактичних позитивних результатів).

Задачу розрахунку оцінки F1 також можна поділити на наступні етапи:

- підготовка набору даних;
- реалізація методів порівняння положень тіла;
- реалізація програмної системи для підрахунку оцінки F1 на базі підготовленого набору даних та реалізованих методів порівняння положень тіла;
- аналіз результатів.

Набір даних має бути достатньо великим і різноманітним, щоб отримати статистично значущі результати.

Дослідження буде зосереджено на аналізі та порівнянні 2D моделей визначення пози тіла людини та методів порівняння отриманих поз.

Визначення 2D пози є добре відомою технікою з великим набором досліджень і розробок. Існує багато моделей і методів визначення двовимірної пози, які були розроблені, протестовані та оптимізовані. Зосереджуючись на визначенні 2D пози, дослідження може спиратися на наявні знання та досвід, а також використовувати загальноживані методи та інструменти для оцінки та порівняння.

Крім того, 2D визначення пози зазвичай виконується за допомогою стандартних камер і методів обробки зображень, тоді як для 3D визначення часто потрібне спеціальне обладнання, наприклад датчики глибини або системи захоплення руху.

Зосередившись на двовимірній оцінці пози, дослідження може вивчити потенціал використання проаналізованих методів та моделей для більш широкого спектру програмних систем у різних областях, які не будуть вимагати спеціального обладнання чи інфраструктури.

В підсумку перелік задач даного дослідження можна визначити наступним чином:

- провести аналіз доступної технічної літератури та наукових праць щодо наявних моделей визначення поз тіла людини;
- за результатами проведеного аналізу літератури обрати модель з найвищими показниками точності та ефективності для подальших досліджень;
- провести моделювання програмної системи для оцінки методів порівняння положень тіла людини;
- реалізувати змодельовану програмну систему, що в якості показника точності методів порівняння положень тіла людини буде використовувати оцінку F1;
- проаналізувати отримані результати;
- провести моделювання та розробку програмної системи навчання технікам ПМР, що буде слугувати доказом практичної цінності одержаних результатів теоретичного та експериментального дослідження.

В якості останнього етапу даної роботи необхідно сформулювати висновки та визначити вектор подальших досліджень.

3 ПЛАНУВАННЯ ЕКСПЕРИМЕНТАЛЬНОГО ДОСЛІДЖЕННЯ

3.1 Моделювання експериментальної програмної системи

Згідно до постановки науково-технічної задачі даного дослідження, одним із питань, що вимагають вирішення, є визначення найбільш ефективного методу порівняння положень тіла людини отриманих як результат роботи моделей машинного навчання визначення положення тіла людини.

Згідно з цим в ході цієї роботи пропонується провести експериментальне дослідження, що передбачає створення експериментальної програмної системи для оцінки точності методів порівняння положень тіла людини. Дана програмна система повинна надавати наступний функціонал:

- завантаження набору даних для проведення дослідження;
- налаштування порогових значень;
- конфігурація моделі визначення положення тіла;
- підрахунок оцінки F1 для кожного реалізованого методу порівняння положень тіла та конфігурації моделі;
- представлення результатів у вигляді графіку або таблиці.

Для більшого розуміння роботи експериментальної програмної системи була створена діаграма послідовності (див. рис. 3.1), що відображає процес взаємодії між користувачем, програмною системою та моделлю визначення пози тіла.

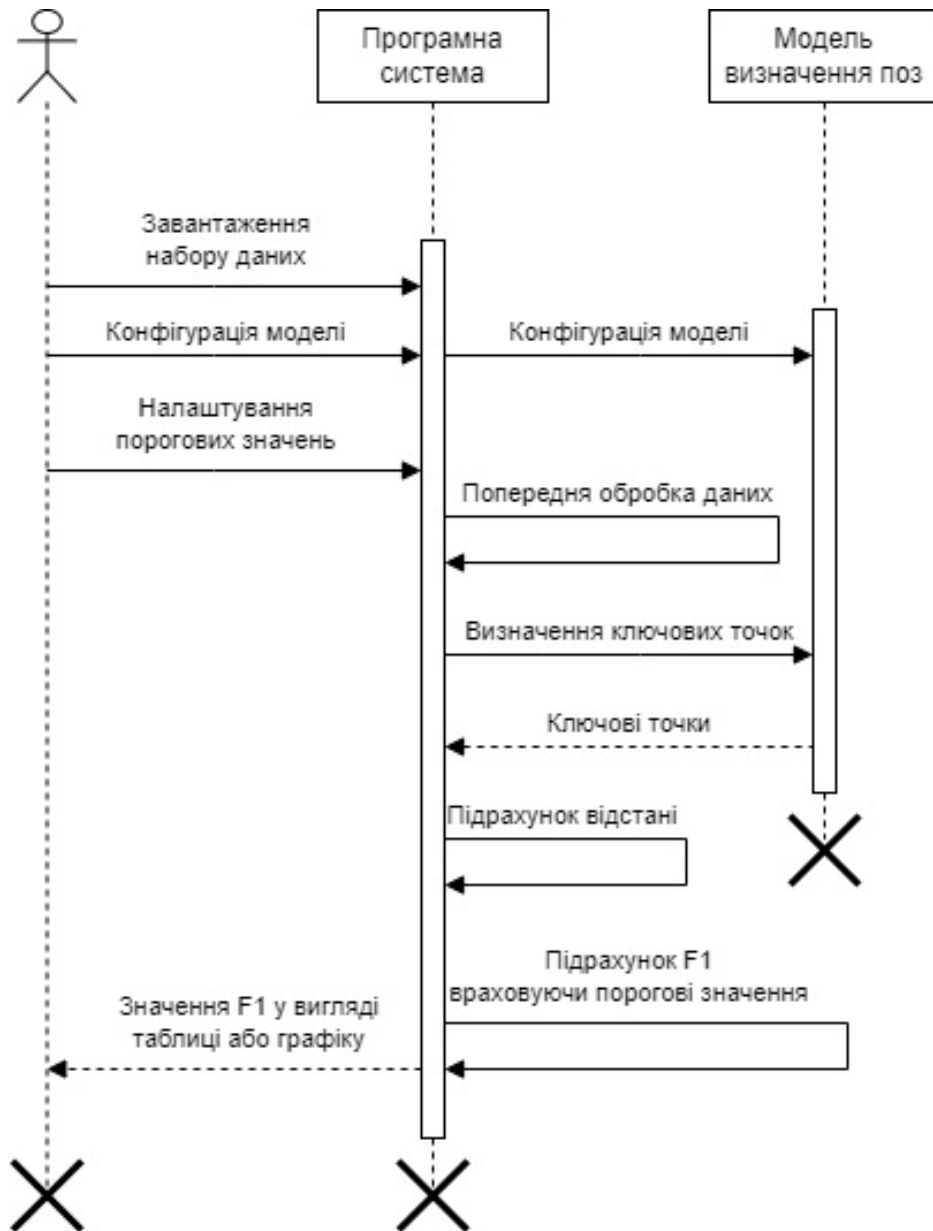


Рисунок 3.1 – Діаграма послідовності експериментальної програмної системи

Розглянемо більш детально основний функціонал, до якого має доступ кінцевий користувач даної системи.

Користувач може завантажити набір даних поз у визначеному форматі, який повинен містити анотації щодо того, які пози є однаковими, а які ні. Програма має мати можливість обробляти різні формати зображень (наприклад, .png, .jpeg тощо).

Користувач повинен мати можливість регулювати порогове значення подібності, яке визначатиме, які пари визначених поз вважаються схожими чи несхожими. Програмна система має надавати можливість налаштовувати порогове

значення у широкому діапазоні, щоб користувач міг побачити вплив різних значень на значення оцінки F1.

Користувач повинен мати можливість налаштувати модель визначення пози, яку використовує програмна система. Це може включати налаштування таких параметрів, як роздільна здатність вхідного зображення, кількість потоків або архітектура моделі.

Після налаштування моделі визначення пози програмна система повинна визначити ключові точки для кожної людини в наборі даних. Цей етап може також включати попередню обробку даних, щоб переконатися, що визначені ключові точки придатні для порівняння. Це можуть бути такі завдання, як зміна розміру, обрізка або нормалізація. Конкретні етапи попередньої обробки залежатимуть від обраного методу порівняння поз, що використовується.

Програмна система має обчислити оцінку F1 для кожного порогового значення, налаштованого користувачем, кожної обраної конфігурації моделі визначення поз та кожного реалізованого методу порівняння.

Як висновок програмна система має представляти результати розрахунку оцінки F1 у зручному форматі, наприклад у вигляді графіка чи таблиці.

3.2 Підготовка набору даних для проведення експериментального дослідження

Одним із важливих етапів експериментального дослідження є збір даних. Якість результатів напряму залежить від якості набору даних, який використовується. Ретельно відбираючи та готуючи набір даних, можна значно підвищити статистичну значимість отриманих результатів.

Як вже зазначалося, набір даних повинен містити зображення поз людей з відповідними мітками, чи є ці пози однаковими.

Існуючі загальнодоступні набори даних, такі як СОСО або МРІІ, часто використовуються для задач комп'ютерного зору, а саме виявлення об'єктів і визначення пози тіла людини. Однак ці набори даних не можна використовувати для конкретного експериментального завдання, оскільки вони не містять міток, які вказують на те, чи є пари поз схожими чи несхожими [23].

Набори даних СОСО та МРІІ містять ключові точки, які ідентифікують розташування різних частин тіла на зображенні, але вони не надають інформації про подібність між різними позами. Ця інформація є обов'язковою для експериментальної програмної системи, оскільки вона використовується для обчислення оцінки F1 для різних методів порівняння положень тіла.

Створення анотованого набору даних для даного експериментального дослідження може бути складним і трудомістким завданням, але воно є важливим для отримання коректних результатів. Набір даних має включати різноманітний діапазон поз і типів тіла.

Загалом рекомендується, щоб набір даних для подібного роду досліджень створювався експертами, що мають глибокі знання предметної області і можуть точно анотувати дані. Однак в рамках даного дослідження отримати такий набір даних неможливо через різні обмеження, такі як час, вартість та відсутність експертів.

Для вирішення даної проблеми пропонується використати інструменти моделювання поз тіла людини. Ці інструменти моделювання можна використовувати для створення 3D-моделей людських фігур (див. рис. 3.2) у різних позах і положеннях. Вони також надають ряд додаткових функціональних можливостей, як то коригування положення окремих частин тіла, зміну орієнтації та кута камери, зміну освітлення та затінення сцени.

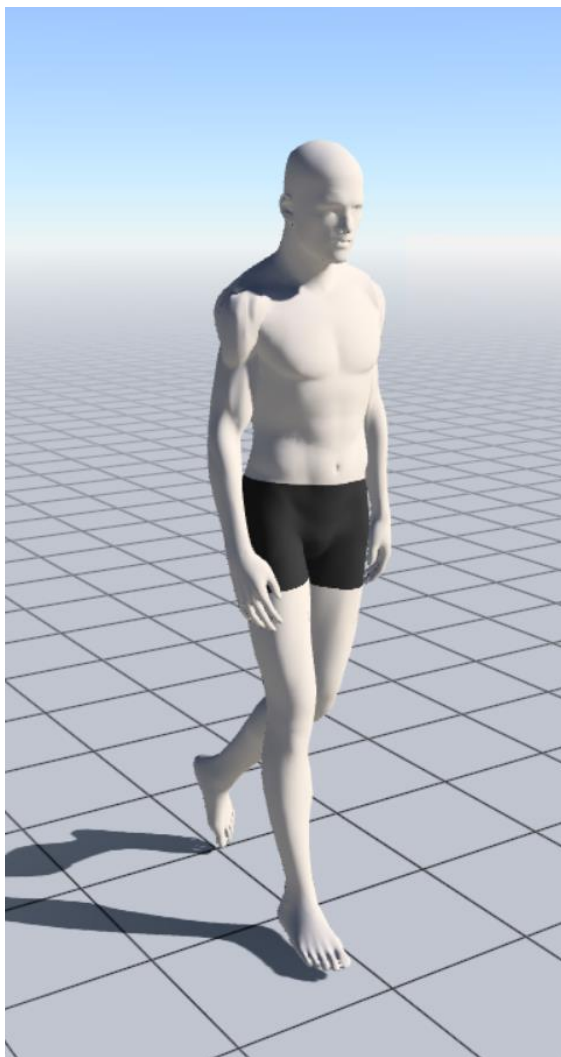


Рисунок 3.2 – приклад 3D-моделі тіла людини

Використовуючи ці інструменти, було створено набір даних зображень з різноманітними позами та типами фігур та анотаціями, які вказують на подібність поз. Даний набір даних був використаний для проведення подальшого дослідження.

4 АНАЛІЗ РЕЗУЛЬТАТІВ ТЕОРЕТИЧНОГО ТА ЕКСПЕРИМЕНТАЛЬНОГО ДОСЛІДЖЕНЬ

4.1 Аналіз результатів теоретичного дослідження

Згідно до постановки задачі перед проведенням експериментального дослідження необхідно проаналізувати наявні теоретичні відомості про різні моделі визначення пози тіла людини та обрати найбільш точну та ефективну для подальшої роботи.

Для цього потрібно порівняти наявні моделі за наступними характеристиками:

- продуктивність;
- точність;
- підтримка різних типів вхідних даних;
- простота використання;
- підтримка різних мов програмування.

В якості моделей для порівняння було вирішено обрати такі бібліотеки, як OpenPose та BlazePose, як ті, що за результатами їх попереднього огляду на даний момент надають найбільш широкий функціонал кінцевому користувачу та займають лідируючі позиції в контексті точності та ефективності.

Порівнюючи продуктивність OpenPose і BlazePose, необхідно враховувати кілька факторів, зокрема апаратне забезпечення, що використовується, вхідні дані та конкретну конфігурацію кожної моделі.

OpenPose і BlazePose можуть досягти продуктивності в реальному часі на сучасних CPU та GPU. Однак BlazePose оптимізовано для швидшої роботи на недорогому апаратному забезпеченні. У тестуванні, проведеному Google, BlazePose досягає продуктивності в реальному часі зі швидкістю більше 30 FPS навіть на мобільних пристроях, як показано на табл. 4.1. З іншого боку, OpenPose

не завжди може підтримувати продуктивність у реальному часі на недорогому апаратному забезпеченні.

Таблиця 4.1 – швидкість роботи BlazePose на різних пристроях і середовищах виконання

	MacBook Pro 15 (FPS)	iPhone 11 (FPS)	Pixel 5 (FPS)	Desktop Intel i9-10900K. Nvidia GTX 1070 GPU. (FPS)
MediaPipe	92 81 38	N/A	32 22 N/A	160 140 98
TensorFlow.js	48 53 28	34 30 N/A	13 11 5	44 40 30

Тип вхідних даних також може впливати на продуктивність цих моделей.

BlazePose в основному розроблений для роботи з 2D відео, але також підтримує й вхідні дані у 3D форматі. Навпаки, OpenPose підтримує різні типи введення, включаючи зображення, відео та 3D хмари точок. Загалом обробка вхідного 2D відео потребує менше обчислень, ніж обробка 3D даних, тому BlazePose може мати перевагу з точки зору продуктивності під час обробки вхідного 2D відео.

OpenPose і BlazePose пропонують різні параметри конфігурації, які можуть вплинути на продуктивність. Наприклад, кількість ключових точок, виявлених кожною бібліотекою, може впливати на час обробки. OpenPose виявляє 25 ключових точок на людину, тоді як BlazePose виявляє 33 ключові точки на людину. Загалом виявлення більшої кількості ключових точок вимагає більше обчислювальних ресурсів і може вплинути на продуктивність.

Іншим параметром конфігурації, який може вплинути на продуктивність, є модель, яку використовує кожна бібліотека. OpenPose пропонує кілька різних моделей, у тому числі легку модель, призначену для швидшої роботи на

апаратному забезпеченні нижчого рівня. BlazePose також пропонує кілька різних моделей, включаючи легку модель, призначену для мобільних пристроїв.

Загалом BlazePose оптимізовано для швидшої роботи на апаратному забезпеченні нижчого класу, тоді як OpenPose пропонує більшу гнучкість щодо вхідних даних і параметрів конфігурації. Проте конкретна продуктивність кожної бібліотеки залежатиме від апаратного забезпечення, що використовується, вхідних даних і обраної конфігурації.

Порівнюючи точність OpenPose і BlazePose, необхідно враховувати кілька факторів, зокрема тип оцінюваної пози, ступінь оклюзії або перекриття даних і конкретну конфігурацію кожної бібліотеки.

OpenPose, і BlazePose здатні точно оцінювати пози людини, але конкретні пози, які кожна бібліотека оцінює найкраще, можуть відрізнятися. Загалом OpenPose краще оцінює складні пози з кількома частинами тіла, які накладаються один на одного, наприклад, коли руки схрещуються перед тілом. BlazePose, з іншого боку, краще оцінює пози, у яких руки витягнуті від тіла, наприклад під час виконання стрибків або інших вправ.

На точність кожної бібліотеки також може вплинути ступінь перекриття або накладання даних. OpenPose, як правило, краще справляється з оклюзіями та перекриванням даних, оскільки вона може використовувати інформацію з інших частин тіла, щоб оцінити розташування закритих або накладених частин тіла.

Специфічна конфігурація кожної бібліотеки також може вплинути на точність оцінки пози. Загалом виявлення більшої кількості ключових точок може підвищити точність, але також може потребувати більше обчислювальних ресурсів.

Іншим параметром конфігурації, який може вплинути на точність, є модель, яку використовує кожна бібліотека. OpenPose пропонує кілька різних моделей, кожна з яких має різні компроміси між точністю та обчислювальними ресурсами. BlazePose також пропонує кілька різних моделей, що відрізняються необхідними ресурсами для обчислень та точністю результатів.

Загалом OpenPose вважається більш точним, ніж BlazePose, особливо в складних ситуаціях із перекриттям даних. Однак конкретна точність кожної бібліотеки залежатиме від типу оцінюваної пози, ступеня оклюзії і конкретної обраної конфігурації.

Порівнюючи простоту використання OpenPose і BlazePose, слід враховувати кілька факторів, зокрема процес встановлення, наявність документації та підтримки, а також легкість інтеграції бібліотек у програми.

Як OpenPose, так і BlazePose може бути складно встановити, особливо для користувачів з обмеженим досвідом у створенні середовищ глибокого навчання. OpenPose вимагає інсталяції кількох залежностей, включаючи Caffe, структуру глибокого навчання та бібліотеку комп'ютерного зору OpenCV. BlazePose вимагає інсталяції TensorFlow, фреймворку глибокого навчання, і TensorFlow Object Detection API. Однак обидві бібліотеки надають докладні інструкції зі встановлення, які можуть допомогти спростити процес.

Обидві бібліотеки мають повну документацію та ресурси підтримки. OpenPose надає детальну документацію на своєму веб-сайті, включаючи інформацію про те, як використовувати бібліотеку та як налаштувати різні параметри. Крім того, OpenPose має активну спільноту користувачів із форумами та репозиторіями GitHub, де користувачі можуть ставити запитання та ділитися своїм досвідом.

BlazePose також містить детальну документацію на своєму веб-сайті, включаючи інформацію про те, як використовувати бібліотеку та як навчати власні моделі. BlazePose також має репозиторій GitHub із прикладами коду та форум спільноти для вирішення різних проблем та питань.

OpenPose, і BlazePose можна відносно легко інтегрувати в програми. OpenPose надає C++ API, а також оболонки для Python і MATLAB, які можуть спростити процес інтеграції. Крім того, OpenPose має кілька готових додатків, у тому числі демонстрацію оцінки пози в реальному часі, яка може стати відправною точкою для інтеграції бібліотеки в інші програми.

BlazePose надає API на основі TensorFlow, який можна використовувати в Python або інших мовах програмування, які підтримують TensorFlow. BlazePose також надає попередньо підготовлені моделі, які можуть спростити процес інтеграції.

Загалом і OpenPose, і BlazePose мають гарну документацію та ресурси підтримки, які можуть допомогти користувачам подолати будь-які труднощі, з якими вони стикаються під час встановлення чи процесу інтеграції. Проте встановити OpenPose може бути дещо складніше, особливо для користувачів із обмеженим досвідом налаштування середовищ глибокого навчання.

Виходячи з розглянутих характеристик, BlazePose може бути кращою бібліотекою для створення програми, яка навчає користувачів вправам прогресивної м'язової релаксації.

По-перше, BlazePose краще оцінює пози, коли руки витягнуті від тіла, що актуально для вправ прогресивної м'язової релаксації, які часто передбачають розгинання рук і ніг.

По-друге, BlazePose надає попередньо підготовлені моделі, які можуть спростити процес інтеграції та дозволити зосередитися на розробці інтерфейсу користувача та навчального вмісту програми.

Нарешті, BlazePose використовує TensorFlow, популярну структуру глибокого навчання з великою документацією та ресурсами підтримки, що полегшує усунення будь-яких проблем, які виникають у процесі розробки.

Підсумовуючи, хоча і OpenPose, і BlazePose є потужними бібліотеками для оцінки пози, BlazePose є більш універсальним інструментом з кращими показниками точності та ефективності, що означає можливість його застосування у ширшому ряді практичних задач.

4.2 Опис розробленої експериментальної програмної системи

Після того, як була обрана модель визначення пози, згідно до постановки задачі необхідно провести експериментальне дослідження з метою визначення найбільш ефективного методу порівняння поз. Для вирішення даної задачі експериментальна програмна система була розроблена.

В якості платформи для даної програмної системи була обрана мова програмування Python. Python має велику екосистему бібліотек та інструментів машинного навчання, які можна легко інтегрувати та використовувати для вирішення різного роду задач.

Згідно до результатів теоретичного дослідження модель BlazePose є оптимальним вибором в якості моделі визначення пози тіла людини, тож саме ця модель була використана під час розробки експериментальної програмної системи.

Ця модель також пропонує кілька конфігурацій, які дозволяють знайти компроміс між швидкістю та точністю. Для вибору в рамках програмної системи доступні наступні конфігурації:

- BlazePose-Light: найшвидша та найлегша конфігурація BlazePose, яка підходить для програм у режимі реального часу на мобільних пристроях або веб-камерах. Вона досягає високої точності, використовуючи лише частину обчислювальних ресурсів;

- BlazePose-Full: стандартна конфігурація BlazePose, яка забезпечує хороший баланс між швидкістю та точністю. Вона використовує глибшу архітектуру нейронної мережі, ніж BlazePose-Light, що дозволяє фіксувати складніші варіації пози та оклюзії;

- BlazePose-Heavy: найточніша, але й найповільніша конфігурація BlazePose, яка використовує набагато глибшу та складнішу архітектуру нейронної мережі, ніж дві інші конфігурації. Вона розроблена для застосувань, які вимагають найвищої можливої точності.

В якості методів, що використовуються для порівняння отриманих положень тіла, були реалізовані метод евклідової відстані та зваженої відстані.

Однак перед тим, як один з методів порівняння положень буде застосовано до отриманих ключових точок, кожне зображення проходить етап попередньої обробки.

Спочатку за допомогою вже зазначеної моделі BlazePose виконується визначення ключових точок (див. рис. 4.1), які будуть використані для розрахунку обмежувальної рамки навколо тіла людини на зображенні.

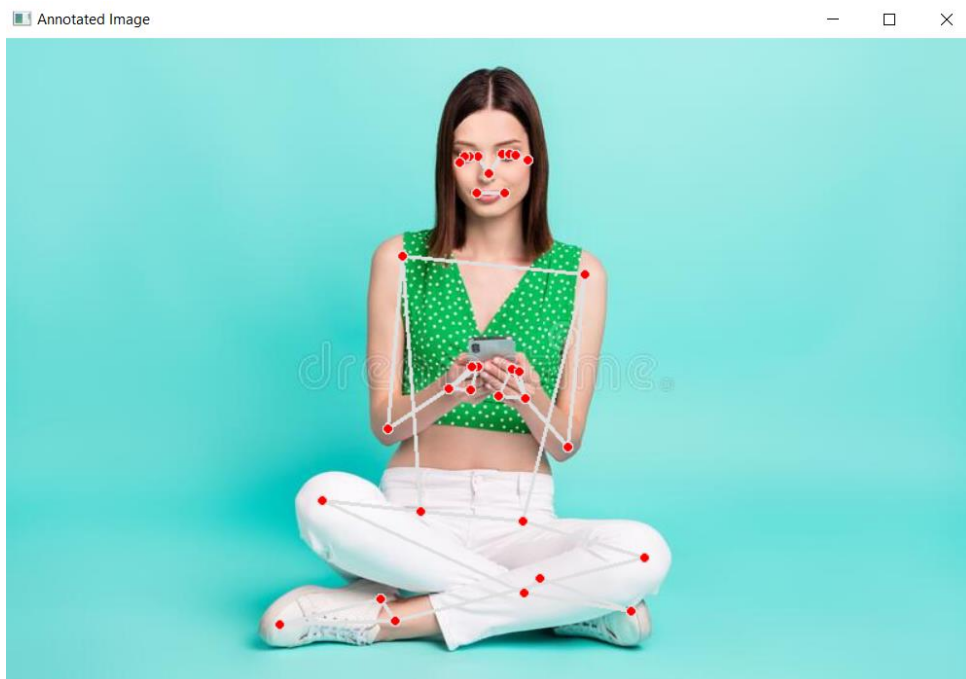


Рисунок 4.1 – Приклад визначення ключових точок для розрахунку обмежувальної рамки

Після того, як обмежувальна рамка була визначена, виконується обрізка та масштабування вхідного зображення. Це дозволяє правильно порівняти ключові точки у майбутньому.

Під час масштабування вхідного зображення було використано такий підхід, як padding. Якщо ми хочемо стандартизувати розмір зображень для обробки, ми можемо використовувати відступи, щоб гарантувати, що всі зображення мають однакові розміри. Процес заповнення передбачає додавання нулів до країв зображення, щоб воно відповідало бажаному розміру чи формі, та центрування

оригінального зображення. Приклад коду використаного у експериментальній програмній системі для масштабування зображень наведено далі:

```
def resize_image(image, size=(256,256)):
    h, w = image.shape[:2]
    c = image.shape[2] if len(image.shape)>2 else 1
    if h == w:
        return cv2.resize(image, size, cv2.INTER_AREA)
    dif = h if h > w else w
    interpolation = cv2.INTER_AREA if dif > (size[0]+size[1])//2
else cv2.INTER_CUBIC
    x_pos = (dif - w)//2
    y_pos = (dif - h)//2
    if len(image.shape) == 2:
        mask = np.zeros((dif, dif), dtype=image.dtype)
        mask[y_pos:y_pos+h, x_pos:x_pos+w] = image[:h, :w]
    else:
        mask = np.zeros((dif, dif, c), dtype=image.dtype)
        mask[y_pos:y_pos+h, x_pos:x_pos+w, :] = image[:h, :w, :]
    return cv2.resize(mask, size, interpolation)
```

Після проведення попередньої обробки даних, нові ключові точки визначаються за допомогою моделі BlazePose та порівнюються за допомогою реалізованих методів евклідової та зваженої відстані.

Під час порівняння ключових точок застосовується нормалізація L2, щоб переконатися, що величина окремих ключових точок не сильно впливає на розрахунок відстані. Без нормалізації вектори ключових точок з більшими величинами можуть мати непропорційно великий вплив на загальний розрахунок відстані.

Ця нормалізація гарантує, що кожен вектор ключових точок має одиничну довжину, що робить розрахунок відстані більш стійким до варіацій у величині векторів.

Отже приклад коду для розрахунку евклідової відстані виглядає наступним чином:

```
def compute_cosine_distance(results_a, image_a, results_b,
image_b):
    keypoints_a = get_keypoints_as_array(results_a, image_a)
    keypoints_b = get_keypoints_as_array(results_b, image_b)
    cosine =
np.dot(keypoints_a, keypoints_b) / (norm(keypoints_a) * norm(keypoints_b)
)
    distance = 2 * (1 - cosine)
    return math.sqrt(distance)
```

Таким чином експериментальна програмна система дозволяє підрахувати оцінку F1 для різних конфігурацій моделі BlazePose та різних порогових значень на основі заздалегідь підготованого набору даних.

4.3 Аналіз результатів експериментального дослідження

Запропонований експеримент мав на меті оцінити продуктивність різних комбінацій методів порівняння пози та конфігурацій бібліотеки BlazePose за показником F1.

Експеримент проводився на наборі даних із 50 пар зображень людських поз, причому кожна пара поз містила анотацію чи є ці пози однаковими.

В експерименті використовувалися три різні конфігурації бібліотеки BlazePose: BlazePose-Light, BlazePose-Full і BlazePose-Heavy. Кожна конфігурація була використана у парі з 2 різними методами порівняння поз – евклідовою відстанню та зваженою відстанню.

Оцінка F1 була розрахована для кожної комбінації методу порівняння пози та конфігурації бібліотеки BlazePose з використанням різних порогових значень від 0.0 до 0.40 з кроком 0.5, як наведено у табл. 4.2 та табл. 4.3, де ED – метод евклідової відстані, WD – метод зваженої відстані

Таблиця 4.2 – результати визначення F1 оцінки для середовища MediaPipe, згрупованих за конфігураціями бібліотеки та методами порівняння поз

	MediaPipe Runtime					
	BlazePose-Light		BlazePose-Full		BlazePose-Heavy	
	ED	WD	ED	WD	ED	WD
$F_1^{0.05}$	50.53	52.33	54.63	56.03	55.35	57.34
$F_1^{0.10}$	60.18	64.47	63.58	65.24	64.72	66.94
$F_1^{0.15}$	63.73	70.12	66.28	69.93	66.98	70.23
$F_1^{0.20}$	67.46	75.25	72.56	75.36	73.05	77.73
$F_1^{0.25}$	71.38	77.33	75.73	79.03	76.43	82.67
$F_1^{0.30}$	66.36	70.02	69.42	72.37	70.82	74.51
$F_1^{0.35}$	59.98	63.35	62.47	67.62	64.62	66.27
$F_1^{0.40}$	53.26	55.14	57.49	59.07	58.26	61.18

Таблиця 4.3 – результати визначення F1 оцінки для середовища Tensorflow, згрупованих за конфігураціями бібліотеки та методами порівняння поз

	Tensorflow Runtime					
	BlazePose-Light		BlazePose-Full		BlazePose-Heavy	
	ED	WD	ED	WD	ED	WD
$F_1^{0.05}$	43.68	44.24	44.05	46.42	50.46	52.54
$F_1^{0.10}$	48.46	49.76	49.46	51.76	54.78	57.42

	Tensorflow Runtime					
	BlazePose-Light		BlazePose-Full		BlazePose-Heavy	
	ED	WD	ED	WD	ED	WD
$F_1^{0.15}$	52.32	54.32	54.38	55.72	59.28	62.68
$F_1^{0.20}$	54.46	57.62	56.86	59.62	63.65	66.37
$F_1^{0.25}$	56.89	59.09	59.42	62.09	67.73	69.87
$F_1^{0.30}$	52.41	53.56	54.94	56.39	65.84	66.61
$F_1^{0.35}$	49.42	51.64	51.30	52.84	60.32	63.72
$F_1^{0.40}$	44.06	48.59	47.18	49.15	59.46	62.73

Загалом результати показали, що BlazePose з середою Mediapipe перевершив Tensorflow за показником F1 для всіх комбінацій алгоритмів порівняння поз і порогових значень. Серед методів порівняння поз зважена відстань виявилася найкращою.

При пороговому значенні 0.25 найефективнішою була комбінація BlazePose-Heavy із комбінацією зваженої відстані, яка досягла оцінки F1 82.67. При такому ж пороговому значенні найгіршою була комбінація BlazePose-Light з косинусною відстанню, яка досягла оцінки F1 71.38.

Загалом результати свідчать про те, що вибір моделі визначення пози та методу порівняння пози може мати значний вплив на продуктивність завдань визначення та порівняння пози.

BlazePose, зокрема, є кращим вибором для високоточних програм оцінки пози. Зважена відстань, з іншого боку, може бути найефективнішим методом порівняння пози для цього завдання, оскільки вона постійно була серед найефективніших комбінацій у всіх порогових значеннях і конфігураціях оцінки пози.

5 ОПИС ПРОГРАМНОЇ РЕАЛІЗАЦІЇ

5.1 Опис основного функціоналу та використаних технологій

Програма розроблена, щоб допомогти користувачам практикувати прогресивну м'язову релаксацію, техніку релаксації, яка передбачає напруження та розслаблення груп м'язів тіла. Додаток використовує BlazePose для відстеження тіла та MediaPipe Hands для відстеження рук, щоб надати користувачам зворотний зв'язок у режимі реального часу щодо їхньої точності пози та прогресу під час виконання вправ ПМР.

Додаток містить кілька вправ ПМР, кожна з яких націлена на певні групи м'язів тіла. Користувачі можуть вибирати з ряду вправ на основі своїх уподобань і рівня досвіду. Додаток надає вказівки щодо виконання кожної вправи та відстежує прогрес користувача з часом.

Під час вправ BlazePose відстежує орієнтири тіла користувача, включаючи голову, тулуб, руки та ноги, і забезпечує зворотний зв'язок у режимі реального часу щодо його пози та форми. MediaPipe Hands відстежує рухи рук користувача під час виконання вправ, забезпечуючи зворотний зв'язок щодо положення та рухів рук.

Додаток також містить функцію оцінки пози, яка надає користувачам відгук про точність пози та форму під час вправ ПМР. Ця функція використовує алгоритм зваженої відстані для порівняння пози користувача з правильною позою для кожної вправи та надає вказівки щодо вдосконалення техніки.

Загалом додаток створено, щоб надати користувачам повний та персоналізований практичний досвід ПМР. Використовуючи сильні сторони BlazePose і MediaPipe Hands, програма може надати користувачам точний і детальний зворотний зв'язок щодо рухів тіла під час вправ, що призводить до покращення техніки та кращих результатів релаксації.

5.2 Опис мультимодельного підходу

Додаток для навчання ПМР використовує багатомодельний підхід, щоб надати користувачам точні та вичерпні відгуки про їхню поставу та рух під час вправ. Цей підхід передбачає використання двох окремих моделей машинного навчання, BlazePose і MediaPipe Hands, які разом аналізують і відстежують рухи тіла й рук користувача в реальному часі.

BlazePose використовується для відстеження пози та рухів тіла користувача, включаючи положення та орієнтацію голови, плечей, тулуба та кінцівок. Модель тренується на великому наборі даних людських поз і використовує передові методи комп'ютерного зору для точного виявлення та відстеження рухів користувача. BlazePose має високу точність і може відстежувати рухи користувача в режимі реального часу, забезпечуючи миттєвий відгук про його поставу під час вправ РМР.

З іншого боку, MediaPipe Hands використовується для відстеження позиції та рухів рук користувача. Модель може відстежувати рухи рук користувача також в режимі реального часу та забезпечує зворотний зв'язок щодо положення та руху рук під час вправ ПМР.

Вихідні дані обох моделей потім об'єднуються за допомогою алгоритму зваженої відстані, щоб забезпечити повне уявлення про форму користувача під час вправ. Алгоритм порівнює рухи користувача з еталонною позою та обчислює оцінку для кожної вправи на основі точності й акуратності його рухів.

Використовуючи багатомодельний підхід, програма може надавати користувачам дуже точний і вичерпний відгук про коректність виконання вправ.

На додаток до переваг підвищеної точності та повного зворотного зв'язку, використання багатомодельного підходу в додатку також забезпечує більш надійну та адаптовану систему. Наприклад, якщо одна з моделей стикається з труднощами під час відстеження певної пози чи руху, інша модель може надати додаткову інформацію, щоб допомогти компенсувати будь-які неточності.

5.3 Майбутні напрямки розвитку

Майбутні напрямки розвитку для додатку навчання ПМР можуть включати розширення спектру вправ і рухів, які відстежуються, шляхом включення додаткових моделей машинного навчання. Наприклад, можна інтегрувати моделі для відстеження виразу обличчя або дихання, щоб забезпечити більш вичерпний зворотний зв'язок щодо загального розслаблення та рівня стресу користувачів під час вправ.

Крім того, додаток можна інтегрувати з іншими додатками та пристроями для здоров'я та оздоровлення, такими як розумні годинники або фітнес-трекери, щоб забезпечити більш цілісний підхід. Використовуючи дані з багатьох джерел, додаток міг би надавати користувачам більш персоналізовані відгуки та рекомендації, допомагаючи їм покращити своє здоров'я фізичне та ментальне.

Одним із можливих майбутніх напрямків розвитку програми PMR є використання методів оцінки 3D-поз, щоб забезпечити ще більш точний і детальний зворотний зв'язок щодо рухів і пози користувачів. Оцінка тривимірної пози передбачає реконструкцію тривимірних положень суглобів тіла з послідовності двовимірних зображень або відео. Використовуючи оцінку 3D-поз, програма може фіксувати ще більш детальну інформацію про рухи користувачів, таку як глибина та орієнтація суглобів, яку можна використовувати для надання більш точного зворотного зв'язку та вказівок.

Крім того, включення 3D оцінки пози в додаток може дозволити більш складне відстеження та аналіз рухів користувачів з часом. Аналізуючи моделі рухів і пози користувачів, програма може надавати персоналізовані рекомендації щодо вдосконалення техніки та зниження ризику травм. Це може бути особливо корисним для користувачів, які відновлюються після травми чи операції, оскільки це дозволить точніше відстежувати прогрес і реабілітацію.

Іншим потенційним майбутнім напрямком для програми є включення динамічної деформації часу (DTW) як алгоритму порівняння поз. DTW – це потужна техніка для порівняння даних часових рядів, яку можна застосувати до проблеми порівняння поз людини, знятих програмою. На відміну від інших заходів подібності, DTW дозволяє порівнювати два часові ряди, які можуть мати різну довжину або не синхронізуватися, що робить його потужним інструментом для порівняння складних рухів, зафіксованих програмою.

Включивши DTW у додаток, програма потенційно може забезпечувати ще точніший зворотний зв'язок щодо рухів і пози користувачів, оскільки вона може враховувати варіації часу чи темпу, які можуть виникати під час вправ ПМР. Крім того, DTW можна використовувати для виявлення моделей або варіацій у рухах користувачів з часом, що може надати цінну інформацію про фізичне здоров'я та самопочуття користувачів.

Крім того, завдяки поєднанню DTW з іншими алгоритмами порівняння поз, такими як косинусна відстань і зважена відстань, програма може запропонувати користувачам ряд варіантів для відстеження та порівняння їхніх рухів і поз, дозволяючи отримати більш персоналізований зворотний зв'язок і вказівки на основі їхніх індивідуальних потреб і уподобань.

ВИСНОВКИ

Підсумовуючи, проведене дослідження із вивчення різних методів визначення та порівняння поз дало цінну інформацію про сильні сторони та обмеження цих інструментів для аналізу поз людини.

У рамках нашого дослідження ми провели широкий огляд різних моделей і підходів до оцінки пози людини. Це дозволило нам прийняти обґрунтовані рішення про те, які бібліотеки та алгоритми має сенс тестувати в подальшому експерименті.

Огляд різних моделей показав, що існує багато різних підходів до оцінки пози, починаючи від 2D оцінки ключових точок до 3D оцінки пози за допомогою датчиків RGB-D. Також було виявлено, що різні моделі оптимізовані для різних випадків використання, таких як програми в реальному часі, високоточне відстеження або обробка з низькою затримкою.

Провівши цей огляд, було визначено, що BlazePose є найперспективнішою та універсальною моделлю з метою використання у додатки навчання технікам ПМР. Ця модель забезпечує точне та надійне відстеження пози тіла, а також пропонує продуктивність у реальному часі та ефективне використання ресурсів.

Окрім огляду різних моделей визначення поз, також було проведено експеримент, щоб визначити найефективніші методи порівняння поз. Зокрема, наступні методи були протестовані: евклідова відстань і зважена відстань.

Експеримент включав порівняння поз, оцінених різними конфігураціями моделі BlazePose, на наборі даних зображень із мітками. Потім була розрахована оцінка F1 для кожного методу порівняння при різних порогових значеннях.

Результати показали, що метод зваженої відстані стабільно перевершує метод евклідової відстані за різними пороговими значеннями. Це вказує на те, що метод зваженої відстані забезпечує надійніший і точніший показник схожості поз, що має вирішальне значення для виявлення та виправлення неправильних поз у додатку навчання ПМР.

Загалом експеримент підкреслює важливість ретельного вибору та оцінки різних методів порівняння поз для різних додатків, беручи до уваги такі фактори, як точність, надійність і порогова чутливість. Це може допомогти переконатися, що вибраний алгоритм добре підходить до конкретних потреб і обмежень цільової програми, як це було у випадку з нашим додатком.

Дослідження продемонструвало потенціал цих інструментів для таких програм, як навчання ПМР, де точне відстеження та порівняння поз людини є важливими.

В якості вектору подальших досліджень, можна, наприклад, дослідити можливість використання DTW як методу порівняння пози, що ще більше підвищити точність і ефективність програми для навчання ПМР, забезпечуючи більш персоналізований відгук і вказівки для користувачів.

Підсумовуючи, дослідження забезпечує цінну основу для майбутніх розробок з використанням бібліотек визначення і методів порівняння поз для відстеження та аналізу поз людини.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Dubey S., Dixit M. A comprehensive survey on human pose estimation approaches. *Multimedia Systems*. 2022. URL: <https://doi.org/10.1007/s00530-022-00980-0> (date of access: 15.04.2023).
2. Bilous N. V., Svidin O. E. ПОШУК ТА ПОРІВНЯННЯ ПОЗ НА ОСНОВІ 3D КООРДИНАТ СУГЛОБІВ ЛЮДИНИ. *Proceedings of the 9th International scientific and practical conference "Priority directions of science and technology development"*. Kyiv, Ukraine. 2021. URL: <https://sci-conf.com.ua/ix-mezhdunarodnaya-nauchno-prakticheskaya-konferentsiya-priority-directions-of-science-and-technology-development-16-18-maya-2021-goda-kiev-ukraina-arhiv/> (date of access: 15.04.2023).
3. Linna M., Kannala J., Rahtu E. Real-time Human Pose Estimation with Convolutional Neural Networks. *International Conference on Computer Vision Theory and Applications*, Funchal, Madeira, Portugal, 27–29 January 2018. 2018. URL: <https://doi.org/10.5220/0006624403350342> (date of access: 15.04.2023).
4. Fast and Lightweight Human Pose Estimation / H. Ren et al. *IEEE Access*. 2021. Vol. 9. P. 49576–49589. URL: <https://doi.org/10.1109/access.2021.3069102> (date of access: 11.05.2023).
5. Human Pose Estimation from Monocular Images: A Comprehensive Survey / W. Gong et al. *Sensors*. 2016. Vol. 16, no. 12. P. 1966. URL: <https://doi.org/10.3390/s16121966> (date of access: 15.04.2023).
6. Deep Learning Methods for 3D Human Pose Estimation under Different Supervision Paradigms: A Survey / D. Zhang et al. *Electronics*. 2021. Vol. 10, no. 18. P. 2267. URL: <https://doi.org/10.3390/electronics10182267> (date of access: 11.05.2023).
7. Li B., Huo G. Face recognition using locality sensitive histograms of oriented gradients. *Optik*. 2016. Vol. 127, no. 6. P. 3489–3494. URL: <https://doi.org/10.1016/j.ijleo.2015.12.032> (date of access: 15.04.2023).
8. Shcherbakova G. Y., Krylov V. N., Bilous N. V. Methods of automated classification based on wavelet-transform for automated medical diagnostics. *2015*

Information Technologies in Innovation Business Conference (ITIB), Kharkiv, Ukraine, 7–9 October 2015. 2015. URL: <https://doi.org/10.1109/itib.2015.7355048> (date of access: 23.04.2023).

9. Deep learning based 2D human pose estimation: A survey / Q. Dang et al. *Tsinghua Science and Technology*. 2019. Vol. 24, no. 6. P. 663–676. URL: <https://doi.org/10.26599/tst.2018.9010100> (date of access: 17.04.2023).

10. Toshev A., Szegedy C. DeepPose: Human Pose Estimation via Deep Neural Networks. *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, OH, USA, 23–28 June 2014. 2014. URL: <https://doi.org/10.1109/cvpr.2014.214> (date of access: 18.04.2023).

11. Newell A., Yang K., Deng J. Stacked Hourglass Networks for Human Pose Estimation. *Computer Vision – ECCV 2016*. Cham, 2016. P. 483–499. URL: https://doi.org/10.1007/978-3-319-46484-8_29 (date of access: 18.04.2023).

12. Nguyen T. D., Kresovic M. A survey of top-down approaches for human pose estimation. *Norwegian University of Science and Technology, Norway*. 2022. URL: <https://doi.org/10.48550/arXiv.2202.02656> (date of access: 18.04.2023).

13. 2D Human pose estimation: a survey / H. Chen et al. *Multimedia Systems*. 2022. URL: <https://doi.org/10.1007/s00530-022-01019-0> (date of access: 21.04.2023).

14. A Survey on Human Pose Estimation / V. Pandel et al. *International Research Journal of Engineering and Technology*. 2022.

15. The Progress of Human Pose Estimation: A Survey and Taxonomy of Models Applied in 2D Human Pose Estimation / T. L. Munea et al. *IEEE Access*. 2020. Vol. 8. P. 133330–133348. URL: <https://doi.org/10.1109/access.2020.3010248> (date of access: 21.04.2023).

16. Convolutional Pose Machines / S.-E. Wei et al. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 27–30 June 2016. 2016. URL: <https://doi.org/10.1109/cvpr.2016.511> (date of access: 21.04.2023).

17. Rakova A. O., Bilous N. V. Research on Methods for Development of Software System for Face Orientation Vector Determining in the Image. *Radio*

Electronics, Computer Science, Control. 2020. URL: <https://doi.org/10.15588/1607-3274-2020-3-11> (date of access: 05.05.2023).

18. PersonLab: Person Pose Estimation and Instance Segmentation with a Bottom-Up, Part-Based, Geometric Embedding Model / G. Papandreou et al. *Computer Vision – ECCV 2018*. Cham, 2018. P. 282–299. URL: https://doi.org/10.1007/978-3-030-01264-9_17 (date of access: 25.04.2023).

19. Abdulghani M. M., Ghazal M. T., M. Salih A. B. Discover human poses similarity and action recognition based on machine learning. *Bulletin of Electrical Engineering and Informatics*. 2023. Vol. 12, no. 3. P. 1570–1577. URL: <https://doi.org/10.11591/eei.v12i3.4930> (date of access: 25.04.2023).

20. Jo B., Kim S. Comparative Analysis of OpenPose, PoseNet, and MoveNet Models for Pose Estimation in Mobile Devices. *Traitement du Signal*. 2022. Vol. 39, no. 1. P. 119–124. URL: <https://doi.org/10.18280/ts.390111> (date of access: 25.04.2023).

21. Seifi S., Tuytelaars T. How to Improve CNN-Based 6-DoF Camera Pose Estimation. *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, Seoul, Korea (South), 27–28 October 2019. 2019. URL: <https://doi.org/10.1109/iccvw.2019.00471> (date of access: 28.04.2023).

22. BlazePose: On-device Real-time Body Pose tracking / V. Bazarevsky et al. *CVPR Workshop on Computer Vision for Augmented and Virtual Reality*, Seattle, WA, USA, 2020. URL: <https://doi.org/10.48550/arXiv.2006.10204> (date of access: 28.04.2023).

23. Дослідження методів для розробки програмної системи розпізнавання емоцій та визначення стану здоров'я людини / N. V. Bilous et al. *Біоніка інтелекту*. – Харків, ХНУРЕ, 2020. URL: [https://doi.org/10.30837/bi.2020.1\(94\).10](https://doi.org/10.30837/bi.2020.1(94).10) (date of access: 18.04.2023).

24. Bilous N. V., Krasov A. I., Vlasenfo V. P. Deletion method of image low-frequency components using fast median filter algorithm. *Journal of Engineering Sciences*. 2016. URL: <https://doi.org/10.21272/jes> (date of access: 20.04.2023).