НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО

Факультет інформатики та обчислювальної техніки
(назва факультету, інституту)

Кафедра автоматизованих систем обробки інформації і управління
(назва кафедри)

"На правах рукопису"
УДК *004.852*

«До захисту допущено»
В.о.завідувача кафедри

_____ О.А.Павлов
(підпис) (ініціали, прізвище)

“____” _____ 20*19* р.

МАГІСТЕРСЬКА ДИСЕРТАЦІЯ

на здобуття ступеня магістра

за спеціальністю *126 Інформаційні системи та технології* .
(код та назва спеціальності)

ОПП *Інформаційні управляючі системи та технології* .
(код та назва спеціалізації)

на тему: *СИСТЕМА РОЗПІЗНАВАННЯ ПОЗИ ЛЮДИНИ В РЕАЛЬНОМУ ЧАСІ*

Виконав: студент *VI* курсу групи ІС-з81мп
(шифр групи)

Лотоцька Юлія Вікторівна _____
(прізвище, ім’я, по батькові) (підпис)

Науковий керівник *ст. викл.Халус О.А.* _____
(посада, науковий ступінь, вчене звання, прізвище та ініціали) (підпис)

Консультант *к.т.н., доц. Жданова О.Г.* _____
(науковий ступінь, вчене звання, прізвище, ініціали) (підпис)

Рецензент _____ _____
(посада, науковий ступінь, вчене звання, прізвище та ініціали) (підпис)

Засвідчую, що у цій магістерській дисертації немає запозичень з праць інших авторів без відповідних посилань.

Студент _____
(підпис)

Київ – 2019

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ім. ІГОРЯ СІКОРСЬКОГО»

Факультет *інформатики та обчислювальної техніки*
(повна назва)

Кафедра *автоматизованих систем обробки інформації та управління*
(повна назва)

Рівень вищої освіти *другий (магістерський) за освітньо-професійною програмою*

Спеціальність *126 Інформаційні системи та технології*
(код і назва)

ОПП *Інформаційні управляючі системи та технології*
(код і назва)

ЗАТВЕРДЖУЮ
Завідувач кафедри
_____ О.А.Павлов
(підпис) (ініціали, прізвище)
«__» _____ 2019 р.

ЗАВДАННЯ
на магістерську дисертацію студенту
Лотоцька Юлія Вікторівна
(прізвище, ім'я, по батькові)

1. Тема дисертації *СИСТЕМА РОЗПІЗНАВАННЯ ПОЗИ ЛЮДИНИ В РЕАЛЬНОМУ ЧАСІ*

науковий керівник дисертації *Халус Олена Андріївна, старший викладач*
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом по університету від "5" *11* 20*19* р. № 3836-с

2. Строк подання студентом дисертації "2" *12* 20*19* р.

3. Об'єкт дослідження *процес розпізнавання пози людини у режимі реального часу із потокового відео*

4. Перелік завдань, які потрібно розробити *провести аналіз основних алгоритмів та методів розпізнавання пози людини, реалізувати алгоритм попередньої обробки вхідних даних для подачі на нейронну мережу; розробити систему розпізнавання пози людини у відеопотоці у реальному часі; провести дослідження ефективності розроблених алгоритмів та порівняти із*

існуючими аналогами на обраному наборі даних

5. Орієнтовний перелік ілюстративного матеріалу *Діаграма варіантів використання системи, схема роботи алгоритму системи розпізнавання пози людини, архітектура системи розпізнавання пози людини, візуалізація компонентів карт інтенсивності для лівого плеча, візуалізація полів асоціації частин тіла, візуалізація розробленої системи та порівняння із OpenPose на наборі даних nuScene, візуалізація розробленої системи та порівняння із Mask RCNN на наборі даних nuScenes*

6. Орієнтовний перелік публікацій *Дві публікації тез доповідей в матеріалах наукових конференцій*

7. Консультанти розділів дисертації

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

8. Дата видачі завдання “ *2* ” *09* 20___ р.

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Строк виконання етапів магістерської дисертації	Примітка
1	*Отримання завдання*	*15.09*	
2	*Аналіз існуючих методів*	*29.09*	
3	*Вибір інструментів розробки*	*8.10*	
4	*Модифікація існуючих методів розпізнавання пози людини*	*15.10*	
5	*Розробка системи розпізнавання пози людини в реальному часі із відео потоку*	*25.10*	
7	*Тестування системи*	*10.11*	
8	*Оформлення документації*	*17.11*	
9	*Подання роботи на попередній захист*	*20.11*	
10	*Подання роботи на основний захист*	*02.12*	

Студент _____ *Ю.В.Лотоцька*
(підпис) (ініціали, прізвище)

Науковий керівник _____ *О.А. Халус*
(підпис) (ініціали, прізвище)

РЕФЕРАТ

Магістерська дисертація: 98 с., 21 рис., 25 табл., 1 додаток, 38 джерел.

Актуальність теми дослідження. У сучасному світі, з урахуванням зростання безпілотних технологій, технологій віртуальної реальності і доповненої реальності, виникає питання про визначення стану речей у просторі, а саме про визначення положення тіла людини. Можливість визначати позу людини на зображенні чи відео у зазначених областях відіграє ключову роль.

Значні досягнення в цій сфері були зроблені завдяки застосуванню згорткових нейронних мереж(Convolutional neural networks - CNN). Однак, завдання залишається невирішеним для непостановочних сцен: важко визначити точну позу людини по зображенню чи відео в режимі реального часу.

Метою дослідження є поліпшення виявлення та розуміння поведінки людей автомобільними бортовими комп'ютерами за рахунок удосконалення методу розпізнавання пози людини у режимі реального часу. Для досягнення поставленої мети необхідно виконати наступні завдання:

- провести аналіз існуючих алгоритмів та програмних аналогів у предметній області;
- реалізувати алгоритм попередньої обробки вхідних даних для подачі на нейронну мережу;
- розробити систему розпізнавання пози людини у відеопотоці у режимі реального часу;
- провести дослідження ефективності розроблених алгоритмів та порівняти із існуючими аналогами на обраному наборі даних.

Об'єкт дослідження – процес розпізнавання пози людини у режимі реального часу із потокового відео.

Предмет дослідження – методи розпізнавання пози людини за допомогою згорткових нейронних мереж.

Наукова новизна одержаних результатів. Удосконалено метод розпізнавання пози людини шляхом формування карт інтенсивності частин(для детектування точного розташування суглобів) та полів інтенсивності частин(для утворення асоціацій між знайденими частинами) для покращення точності та швидкості розпізнавання у відео потоці у режимі реального часу.

Публікації. Матеріали роботи опубліковані у Всеукраїнській науково-практичній конференції молодих вчених та студентів «Інформаційні системи та технології управління»(ІСТУ-2019) «Аналіз методів автоматичного реферування тексту за допомогою нейронних мереж» та у третій Всеукраїнській науково-практичній конференції молодих вчених та студентів «Інформаційні системи та технології управління» (ІСТУ-2019) «Розпізнавання пози людини у реальному часі».

КОМП’ЮТЕРНИЙ ЗІР, РОЗПІЗНАВАННЯ ПОЗИ ЛЮДИНИ, ЗГОРТКОВІ НЕЙРОННІ МЕРЕЖІ, КАРТИ ІНТЕНСИВНОСТІ ЧАСТИН, ПОЛЯ АСОЦІАЦІЇ ЧАСТИН

ABSTRACT

Master's thesis: 98 pages, 21 figures, 25 tables, 1 appendix, 38 sources.

Relevance of the research topic. In today's world, given the rise of unmanned aerial technology, virtual reality technology and augmented reality, the question arises about determining the state of things in space, namely about determining the position of the human body. The ability to determine the position of a person in an image or video in these areas plays a key role.

Significant advances in this area have been made through the use of Convolutional neural networks (CNN). However, the task remains unresolved for non-stage scenes: it is difficult to determine the exact position of a person in an image or video in real time.

The purpose of the study is to improve the detection and understanding of human behavior by on-board computers by improving the method of recognizing human poses in real time. To achieve this goal, we must perform the following tasks:

- analyze the existing algorithms and software analogues in the subject area;
- implement the algorithm of preliminary processing of input data for submission to the neural network;
- develop a system of recognition of human postures in the video stream in real time;
- research of efficiency of the developed algorithms and compare with existing analogues on the selected data set.

The object of study is the process of recognizing human postures in real time from streaming video.

The subject of the study - methods for recognizing human postures using convolutional neural networks.

Scientific novelty of the obtained results. The method of recognizing human posture has been improved by forming part intensity maps (to detect the exact location of joints) and part intensity fields (to form associations between found parts) to improve the accuracy and speed of video stream recognition in real time.

Publications. The materials are published in the All-Ukrainian Scientific and Practical Conference of Young Scientists and Students “Information Systems and Management Technologies”(ISTU-2019) “Analysis of Methods of Automatic Text Referencing Using Neural Networks” and in the Third All-Ukrainian Scientific and Practical Conference of Young Scientists and Students control systems and technologies ”(ISTU-2019) “Real-time recognition of human postures”.

COMPUTER VISION, HUMAN POSE ESTIMATION, CONVOLUTIONAL NEURAL NETWORK, PART INTENSITY MAPS, PART ASSOCIATION FIELDS

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, СКОРОЧЕНЬ І ТЕРМІНІВ

CNN – Convolutional Neural Network

R-CNN - Recurrent Convolutional Neural Networks

BRNN - Bidirectional Recurrent Neural Network

RMPE - Regional multi-person pose estimation

SSTN - Symmetric Spatial Transformer Network

NMS - Non-Maximum Suppression

ReLU - Rectified Linear Unit

ResNet - Residual Network

КІЧ – Карти Інтенсивності частин

ПАЧ – Поля Інтенсивності Частин

COCO - Common Objects in Context

AP – Average Precision

AR – Average Recall

ВСТУП

Стрімкий розвиток сучасних комп'ютерних технологій призвів до появи потужних і доступних обчислювальних пристроїв, що дозволяють реалізовувати і виконувати складні алгоритми, що вимагають значних обчислювальних потужностей. Завдяки цьому стала можлива практична реалізація різних методів штучного інтелекту, в тому числі - ефективних методів машинного навчання, що дозволяють замінити людину в різних сферах діяльності, що вимагають тривалої монотонної обробки інформації.

З кожним роком зростає кількість безпілотних технологій, технологій віртуальної реальності і доповненої реальності, які використовують інформацію про положення тіла людини у просторі. Можливість визначати позу людини на зображенні чи відео у зазначених областях відіграє ключову роль. На сьогоднішній день, не існує готового універсального рішення таких задач, незважаючи на те, що багато великих компаній IT індустрії, в тому числі Google, Facebook, NVidia, активно ведуть фундаментальні та прикладні дослідження в області машинного навчання.

Значні успіхи були досягненні у розпізнаванні пози людини завдяки використанню нейронних мереж. Одна з найпоширеніших моделей - перцептрон. Однак для вирішення задачі розпізнавання пози людини дана модель не підходить, адже великий розмір вхідних даних(зображення, відео) призводить до значного збільшення кількості синаптичних зв’язків, нейронів у мережі. Як результат, швидкість та обчислювальна складність розпізнавання значно збільшується. Топологія вхідних даних ігнорується у цій моделі нейронної мережі, чітка двомірна структура вхідних зображень не враховується. Для усунення цих недоліків, ми будемо використовувати згорткову нейрону мережу.

Існуючі моделі та методи розпізнавання пози на відео є достатньо складними та займають багато часу при оцінці пози у режимі реального часу. Наразі активно ведеться розробка спрощених алгоритмів, які зможуть досить точно та швидко розпізнати позу людини у відеопотоці, що дозволить їх використовувати у

автопілотних атомобілях у режимі реального часу. Дана робота присвячена поліпшенню існуючих алгоритмів оцінки пози людини у переповнених непостановочних сценах.

В рамках даної роботи був проведений аналіз існуючих методів обробки і класифікації відеопотоків і розпізнавання пози людини у вигляду 2D скелету. За результатами досліджень було запропоновано підхід, який використовує згорткову нейронну мережу для створення карт інтенсивності частин та полів асоціації частин для подальшого декодування та оцінки пози людини.

1 СУЧАСНИЙ СТАН РОЗПІЗНАВАННЯ ПОЗИ ЛЮДИНИ ЗА ДОПОМОГОЮ НЕЙРОННИХ МЕРЕЖ

1.1 Огляд методів розпізнавання пози людини

Один із варіантів вирішення завдання розпізнавання пози людини був представлений ще в 1973 році - Pictorial Structures Model [1]. Основою для даного фреймворка виступала ідея про створення деякої математичної моделі і скелета людини у вигляді графа, де кістки - ребра, а суглоби - вершини даного графа.

Протягом останніх років найсучасніші методи оцінки пози ґрунтуються на конволюційних нейронних мережах. Вони перевершують традиційні методи, засновані на зображувальних структурах та моделях деформованих частин. Популярність глибокого навчання розпочалося з DeepPose [2], який використовує каскад згорткових мереж для оцінки пози цілого тіла. Тоді, замість прогнозування абсолютних місцеположень людських суглобів, деякі автори вдосконалюють оцінку пози, передбачуючи виправлення помилок при кожній ітерації або використовуючи мережу уточнення поз людини для використання залежностей між вхідними та вихідними просторами.

Усі ці підходи до оцінки пози людини можна згрупувати в методи «знизу вгору» та «згори вниз». Перший оцінює спочатку кожен суглоб тіла, а потім групує їх для створення унікальної пози. Останній запускає спочатку детектор людини і оцінює суглоби тіла в межах виявлених обмежувальних діапазонів. Методи «зверху вниз» особливо ефективні, коли пішоходи перекривають інших пішоходів, де стикаються обмежувальні діапазони. Попередні методи «знизу вгору» не обмежують поле, але все ще містять грубу карту функцій для локалізації. Запропонований метод не містить будь-яких обмежень на основі сітки щодо просторової локалізації суглобів і має здатність оцінювати численні пози, що перекриваються один одним.

Детальніше розглянемо існуючі методи розпізнавання пози людини.

OpenPose

OpenPose [3] - це один із найпопулярніших підходів «знизу вгору» для оцінки пози багатьох людей у кадрі. Як і у більшості підходах знизу вгору, OpenPose спочатку виявляє частини(ключові точки), що належать кожній людині на зображенні, а потім призначає частини окремим особам. На рисунку 1.1 показана архітектура моделі OpenPose.

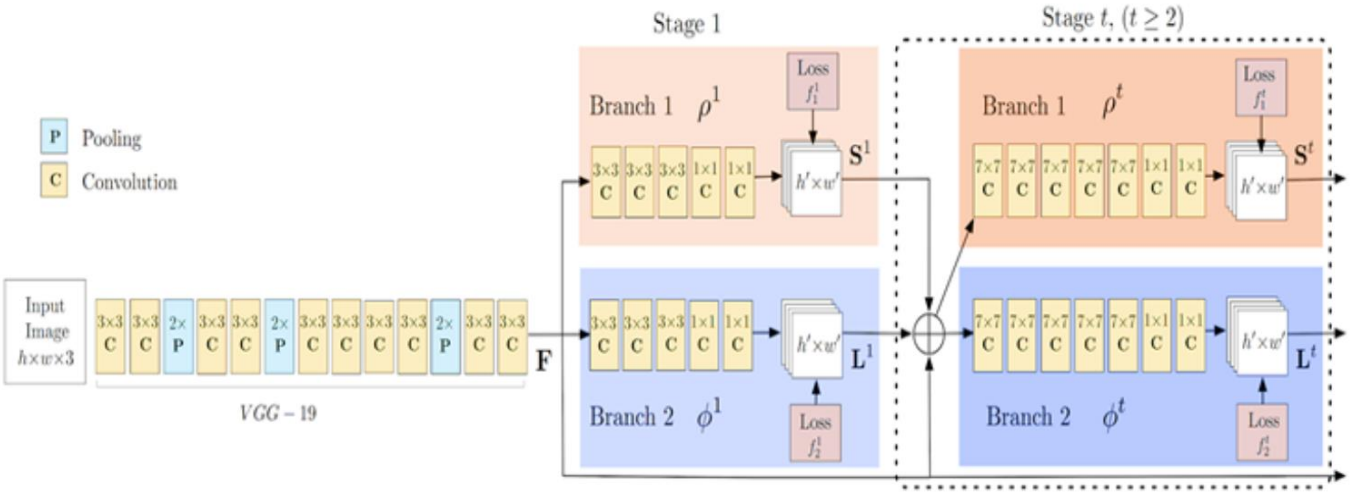


Рисунок 1.1 - Архітектура моделі OpenPose[3].

Мережа OpenPose спочатку витягує функції з зображення за допомогою перших кількох шарів(VGG-19 у вищевказаній блок-схемі). Потім риси подаються у дві паралельні гілки згорткових шарів. Перша гілка прогнозує набір з 18 карт впевненості, при цьому кожна карта представляє певну частину скелета людини. Друга гілка передбачає набір з 38 полів спорідненості частин, який представляє ступінь асоціації між частинами.

Послідовні етапи використовуються для уточнення прогнозів, зроблених кожним шаром. За допомогою карт достовірності частин між двома частинами формуються двосторонні графіки. Використовуючи значення полів спорідненості частин, слабші ланки у двосторонніх графіках обрізаються. Через вищезазначені кроки людські скелети можуть бути оцінені та призначені кожній людині на зображенні.

DeepCut

DeepCut [4] - це підхід знизу вгору для оцінювання пози для людини. Автори при роботі визначили такі завдання:

- підготувати набір кандидатів D частини тіла. Цей набір представляє всі можливі розташування частин тіла для кожної людини на зображенні. Вибрати підмножину частин тіла з вищевказаного набору кандидатських частин тіла
- позначити кожну обрану частину тіла одним із класів частин тіла C. Класи частин тіла представляють типи частин тіла, такі як «рука», «нога», «тулуб» тощо;
- виокремлення частин тіла, які належать одній людині.

Вищезазначені проблеми були вирішені шляхом моделювання їх в задачу Integer Linear Programming(ILP). Моделюється система, розглядаючи триплекси (x, y, z) бінарних випадкових змінних з доменами, як зазначено в формулах нижче [4].

$$x \in \{0,1\}^{D*C},$$

$$y \in \{0,1\}^{\binom{D}{2}},$$

$$z \in \{0,1\}^{\binom{D}{2}*C^2}.$$

Розглянемо двох кандидатів на частину тіла *d* і *d'* з набору кандидатів на частину тіла *D* та класів *c* і *c'* з набору класів *C*. Кандидати на частину тіла були отримані через Faster RCNN або Dense CNN. Тепер ми можемо розробити наступний набір тверджень.

Якщо *x(d, c)* = 1, то це означає, що кандидат частини тіла *d* належить до класу c.

Також *y(d, d ')* = 1 вказує на те, що кандидат частини тіла *d* і *d'* належать одній людині.

Вони також визначають *z(d, d', c, c') = x(d, c) * x(d ', c') * y(d, d ').* Якщо вказане значення рівняння дорівнює 1, то це означає, що кандидат частини тіла *d* належить до класу *c*, кандидат частини тіла *d* належить до класу *c'*, і, нарешті, кандидати в частину тіла *d, d*' належать тій же особі.

Останнє твердження може бути використане для розмежування поз тіла, що належать різним людям. Зрозуміло, що вищезазначені твердження можна сформулювати з точки зору лінійних рівнянь як функції (x, y, z). Таким чином встановлюється ціла лінійна програма, за допомогою якої може бути оцінена поза кількох осіб.

Hierarchical RNN for Skeleton Based Action Detection

У роботі [5] автори використовують поділ людського скелета на п'ять анатомічних частин (тулуб, дві руки і дві ноги) замість того, щоб подавати на вхід нейронної мережі безпосередньо єдиний скелет. В кожної з п'яти частин виділяються кілька значимих точок, рух яких і відслідковуються. Кожна з п'яти частин подається на двосторонню нейронну мережу(Bidirectional Recurrent Neural Network - BRNN).

По мірі зростання кількості шарів образи, виділені підмережами, ієрархічно об'єднуються на вхід для наступних шарів. В кінці повнозв'язний шар і softmax шар завершують архітектуру, видаючи кінцеве уявлення для класифікації рухів. Архітектура даної ієрархічної нейронної мережі з описом частин тіла, за які відповідає кожен шар, представлена на рисунку 1.2.

Точність даного методу є досить високою, але він має істотний недолік: в якості вхідних даних використовуються дані захоплення рухів(motion capture), і це створює необхідність в спеціальному обладнанні для отримання інформації про рухи конкретної людини.

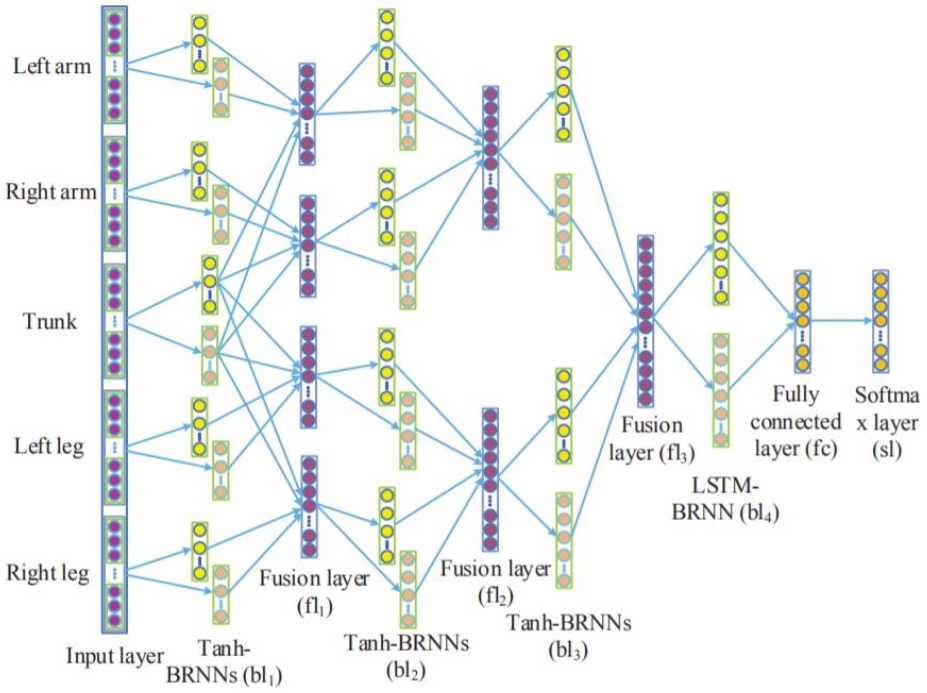


Рисунок 1.2 - Архітектура ієрархічної нейронної мережі [5]

Regional multi-person pose estimation(RMPE або AlphaPose)

RMPE [6] - популярний метод «зверху-вниз» для оцінювання поз людини. Автори стверджують, що методи зверху-вниз зазвичай залежать від точності детектора людини, оскільки оцінка позиції проводиться в області, де знаходиться людина. Отже, помилки в локалізації та передбачувані дублікатів обмежувального поля можуть призвести до того, що алгоритм розпізнавання пози людини створюватиметься неоптимально.

Щоб вирішити цю проблему, автори запропонували використовувати Symmetric Spatial Transformer Network(SSTN) [6] для вилучення окремо взятої особи з неточного обмежувального поля. У цій області використовується Single Person Pose Estimator(SPPE) для оцінки скелета людської пози для цієї людини. A Spatial De-Transformer Network(SDTN) використовується для трансформації оціночної пози людини до початкової системи координат зображень. Нарешті, параметрична Non-

Maximum Suppression(NMS) техніка використовується для вирішення питання надмірно створених поз.

Крім того, автори впроваджують Pose Guided Proposals Generator для розширення зразків навчання, які можуть краще допомогти тренувати мережі SPPE та SSTN. Важливою особливістю RMPE є те, що цю методику можна поширити на будь-яку комбінацію алгоритму виявлення людини та SPPE.

R-CNN (Regions With CNNs)

Архітектура мережі R-CNN [7] була розроблена командою з UC Berkley для застосування Convolution Neural Networks до задачі детектування об’єктів. Процедуру детектування об'єктів мережею R-CNN можна розбити на наступні кроки:

- виділення регіонів-кандидатів за допомогою Selective Search;
- перетворення регіону в розмір, який приймає CNN CaffeNet;
- отримання за допомогою CNN 4096-розмірного вектору ознак;
- проведення N бінарних класифікацій кожного вектору ознак за допомогою N лінійних SVM;
- лінійна регресія параметрів рамки регіону для більш точного охоплення об'єкта.

R-CNN в основному грає роль класифікатора, і він не передбачає межі об'єкта (крім уточнення за допомогою регресії обмежувальною рамкою). Його точність залежить від продуктивності модуля пропозиції регіону. У декількох роботах були запропоновані способи використання глибоких мереж для прогнозування обмежувальних рамок об'єктів.

Fast R-CNN[8] запропонували прискорити процес R-CNN за рахунок пари модифікацій: пропускати через CNN не кожен з 2000 регіоно-кандидатів окремо, а цілком все зображення. Запропоновані регіони потім накладаються на отриману загальну карту ознак. Перетворення ознак, які потрапили в різні регіони, до фіксованих розмірів проводилося за допомогою процедури RoIPooling. Вікно регіону шириною *w* і

висотою *h* поділялося на сітку, що має *HW* комірок розміром *h/H w/W*. По кожній такій клітинці проводився Max Pooling для вибору тільки одного значення, даючи матрицю ознак *HW*.

Після покращень, зроблених в Fast R-CNN, найвужчим місцем нейронної мережі виявився механізм генерації регіонів-кандидатів. У 2015 команда з Microsoft Research змогла зробити цей етап значно швидшим. Вони запропонували обчислювати регіони не по початковому зображенню, а по карті ознак, отриманих з CNN (рисунок 1.3).

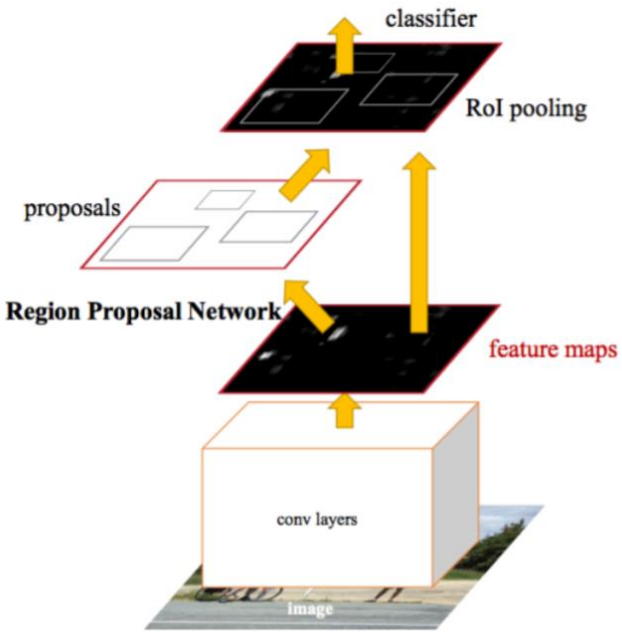


Рисунок 1.3 - Архітектура Faster R-CNN[8]

Для цього було додано модуль Region Proposal Network(RPN). В рамках RPN по витягнутих CNN ознаками проходять «міні-нейромережею» з 3x3 вікном. Отримані з її допомогою значення передаються в два паралельних повнозв’язних шари: box-regression layer(reg) і box-classification layer(cls).

Mask R-CNN

Mask R-CNN[7] продовжує архітектуру Faster R-CNN шляхом додавання ще однієї гілки, яка передбачає положення маски, що покриває знайдений об'єкт, і вирішує вже завдання instance segmentation(рисунок 1.4).

Маска являє собою просто прямокутну матрицю, в якій 1 на деякій позиції означає приналежність відповідного пікселя об'єкту заданого класу, 0 - що піксель об'єкту не належить. Виділення маски відбувається в class-agnostic стилі: маски передбачаються окремо для кожного класу, без попереднього знання, що зображено в діапазоні, і потім просто вибирається маска класу, яка перемогла в незалежному класифікаторі. Стверджується, що такий підхід більш ефективний.

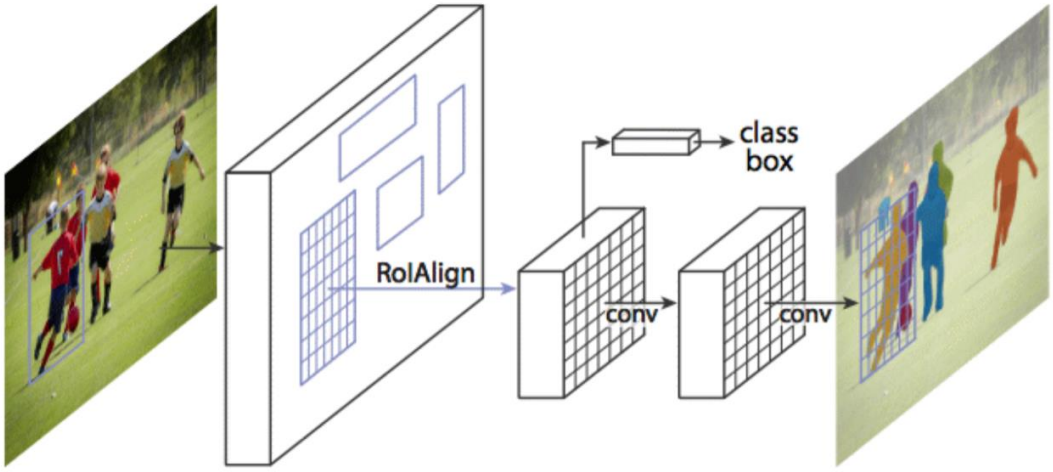


Рисунок 1.4 - Архітектура Mask R-CNN[7]

Одна з основних модифікацій, що виникла через необхідність передбачати маску - зміна процедури RoIPool(обчислює матрицю ознак для регіону-кандидата) на так звану RoIAlign. Справа в тому, що карта ознак, отримана з CNN, має менший розмір, ніж вихідне зображення, і регіон, який охоплює на зображенні цілочислену кількість пікселів, не виходить відобразити в пропорційний регіон карти з цілочисельною кількістю ознак.

RoIPool проблема вирішувалася просто округленням дрібних значень до цілих. Такий підхід нормально працює при виділенні обмежувальної рамки, але обчислена на основі таких даних маска виходить занадто неточною. В противагу цьому, в RoIAlign не використовується округлення, всі числа залишаються дійсними, а для обчислення значень ознак використовується білінійна інтерполяція по чотирьом найближчим цілочисельним точкам.

Паралельно алгоритм виявлення об’єктів може бути навчений для визначення місця розташування осіб. Поєднуючи інформацію про місцезнаходження людини, а також їх набір ключових точок, ми отримуємо скелет-позу людини для кожної людини на зображенні.

Цей метод нагадує підхід «зверху вниз», але етап виявлення людини виконується паралельно етапу виявлення частин тіла. Іншими словами, етап виявлення ключових точок і етап виявлення людини не залежать один від одного.

1.2. Огляд існуючих рішень для реалізації проекту

В якості вхідних даних системи використовується відеофрагмент, але перед подачею цих даних на нейронну мережу необхідно обробити їх, тобто витягти тільки ту інформацію, яка необхідна для вирішення завдання. У нашому випадку ця інформація являє собою координати тіла людини в двовимірному просторі - його скелет.

Щоб отримати скелет, що рухається, на зображенні, спочатку потрібно виділити цей об'єкт, для чого існують відповідні алгоритми.

Методи виділення рухомих об'єктів

Існує два основні класи методів виділення рухомих об'єктів [9]:

- методи виділення границь;
- методи виділення повної області об'єкта.

Основною ідеєю методів першого класу є пошук відмінностей векторів оптичного потоку з наступною побудовою контурів рухомих на зображенні об'єктів. Методи другого класу, навпаки, ґрунтуються на групуванні схожих векторів з оптичного потоку в певній області, які подаються на вихід алгоритму в якості рухомих об'єктів.

Оптичний потік - це зображення видимого руху, що представляє собою зсув кожної точки між двома зображеннями. Для побудови оптичного потоку існує кілька алгоритмів, деякі з яких описані далі.

Метод Лукаса-Канаді[10] - локальний метод, обчислює оптичний потік в кожному пікселі незалежно від значень потоку в інших пікселях. Це диференційний метод, так як використовує похідні для оцінки. Це досить швидкий метод, але, в силу локальності, працює погано на областях з однорідною текстурою.

Алгоритм Хорна-Шанко[11] має дещо більш глобальний характер, ніж метод Лукаса-Канаді. Він спирається на припущення про те, що на всьому зображенні оптичний потік буде досить гладким. Цей алгоритм заснований на гіпотезі про обмеження зміни проекцій векторів оптичного потоку згідно з визначеними рівняннями. Існують також різні варіанти модифікації цих методів, що дозволяють поліпшити їх продуктивність і зменшити кількість помилок, наприклад, поєднання методу Лукаса-Канаді з методом Хорна-Шанко [12].

При нерухомому фоні відеозапису можливе використання більш простого методу - бінаризації або віднімання фону [13]. У ньому рух визначається по пікселям, інтенсивність яких змінюється щодо інтенсивності фонових пікселів. Алгоритм бінаризації для відстеження руху простий в реалізації і швидко обчислюється.

Скелетизація та її застосування

Скелет зображення виділяє геометричні і топологічні властивості форми, наприклад, її зв'язок, топологію, довжину, напрямок, ширину. Разом з відстанню від його точок до границі форми скелет може також служити в якості представлення форми (вони містять всю інформацію, необхідну для відновлення форми) [14].

Скелети широко використовуються в комп'ютерному зорі, аналізі зображень, розпізнаванні образів і цифровій обробці зображень в таких цілях, як оптичне розпізнавання символів, розпізнавання відбитків пальців, візуальний контроль. Існує кілька методів скелетизації, в їх числі найпростіший і найменш дієвий - метод шаблонів [15]. Він полягає в аналізі границь елементів бінаризованого зображення, в результаті якого послідовно знаходять і видаляють крайні елементи об'єктів. Інший метод скелетизації - хвильової метод[16]. Хвильові алгоритми часто використовуються в комп'ютерних іграх для визначення мінімальної відстані від одного об'єкта до іншого в обмеженому дискретному просторі. Для цього, у вихідній точці генерується хвиля, що розповсюджується за певними законами, позначаючи пройдені точки. Одним з найбільш ефективних алгоритмів на даний момент є алгоритм Зонга-Суня, який ґрунтується на діаграмах Вороного.

1.3 Опис розробки системи моделювання процесів визначення пози людини на відеофрагменті

Задача моделювання процесів визначення положення об’єктів на динамічному наборі зображень є досить широкою та має безліч нюансів, які розробник повинен дослідити та узгодити із замовником. У результаті дослідження аналогічних систем, а також вивчення предметної області, було вирішено створити систему, основними функціями якої є розпізнавання положення об’єктів на зображенні, розпізнавання точок скелету, формування пози людини на відео у реальному часі.

Для виконання поставленої задачі було обрано набір засобів реалізації, було спроектовано архітектуру системи, а також її компоненти.

1.4 Постановка цілей та задач дослідження

Призначенням розробки є розпізнавання 2D пози людини у відеопотоці у режимі реального часу.

Основними цілями створення комплексу задач є:

- підвищення швидкості та точності розпізнавання пози людини;
- підвищення ефективності аналізу відео потоку бортовими комп'ютерами автопілотних автомобілів, камер відеонагляду для подальшої обробки для прийняття рішень.

Для досягнення поставлених цілей потрібно вирішити такі завдання:

- провести аналіз існуючих алгоритмів та програмних аналогів у предметній області;
- реалізувати алгоритм попередньої обробки вхідних даних для подачі на нейронну мережу;
- розробити систему розпізнавання пози людини у відеопотоці у режимі реального часу;
- провести дослідження ефективності розроблених алгоритмів та порівняти із існуючими аналогами на обраному наборі даних.

1.5 Висновки до розділу

Проведений огляд існуючих систем показав, що задача автоматичного розпізнавання пози людини сьогодні залишається актуальною. Проаналізувавши існуючі рішення в даній області, ми надалі будемо модифікувати методи «знизу вгору», де спочатку будуть знаходитись суглоби на зображені, а потім – групуватись у пари для конкретної людини на зображені за допомогою нейронної мережі.

2 ЗАСТОСУВАННЯ ЗГОРТКОВОЇ НЕЙРОННОЇ МЕРЕЖІ ДО ЗАДАЧ РОЗПІЗНАВАННЯ ПОЗИ ЛЮДИНИ

2.1 Обґрунтування використання згорткової нейронної мережі до задачі розпізнавання пози людини

На даний момент найкращого результату в методах машинного навчання досягла математична модель нейронної мережі. З розвитком апаратної складової, багато великих компаній мають можливість навчати нейронну мережу на величезних масивах даних за невеликий проміжок часу. Дані інновації вивели згорткові нейронні мережі на перший план, усунувши проблему довгого навчання мережі.

Коли заходить розмова про обробку зображень і, зокрема, розпізнавання об'єктів на фотографіях та відеофрагментах, то виникає проблема великого обсягу даних. Наприклад, найпоширеніша база даних з зображеннями містить мільйони картинок з розмірами 256*256 і записаними в RGB форматі. Звичайна нейронна мережа буде неймовірно довго навчатися навіть на суперпотужних комп'ютерах і, до того ж, ймовірно, прийде до проблеми перенавчання, яка полягає у відсутності можливості сприймати і класифікувати об'єкти в загальному. Згорткова нейронна мережа на вхід отримує вектор значень, перетворюючи його в подальшому через приховані шари.

Ці приховані шари складаються з пулу нейронів, в якому кожний нейрон пов'язаний з нейронами попереднього шару. Останній шар у нейронній мережі - вихідний шар. Зазвичай він являє собою вектор ймовірної приналежності до певного класу. Для прикладу з картинками cImageNet в такій повністю пов'язаній архітектурі кожен нейрон буде мати близько 256*256*3 = 196608 ваг. Для повноцінної класифікації необхідно мати як мінімум кілька таких нейронів.

Рішенням даної задачі є спеціальна архітектура штучної нейронної мережі - так звана згорткова нейронна мережа. Дана назва прямо походить від функції, які виконуються цією мережею-згорткою. Такий тип мереж відноситься до різновиду

глибинного навчання, використовуючи нелінійні перетворення для отримання результату.

В згорткових нейронних більш розумно організована архітектура для роботи з зображеннями. Шари цієї нейронної мережі розкладені в трьох вимірах, які характерні для картинки: висота, ширина і глибина(щодо зображень це канали Red, Green, Blue).

2.2 Математичний опис роботи нейронної мережі

Уоррен Маккаллок(Warren McCulloch) і Уолтер Питтсом (Walter Pitts) запропонували у своїй роботі з моделювання нервової активності модель штучного нейрона [17]. В якості основи для своєї моделі автори використовували біологічний нейрон. Штучний нейрон Маккаллок-Питтса має N вхідних бінарних величин $x1,\ldots,xn$. Ці величини є імпульсами, які надходять на вхід нейрона(рисунок 12). У нейроні імпульси складаються з вагами $w1,\ldots,wn$.

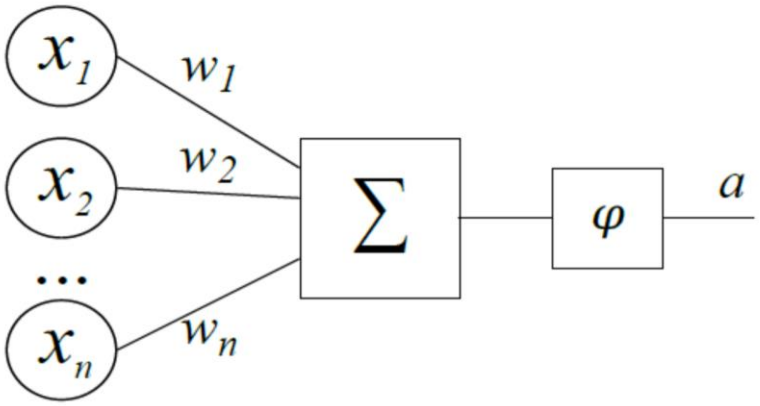


Рисунок 2.1 - Модель штучного нейрона Маккаллок-Піттса

Вихідний сигнал визначається за формулою:

$$\alpha = \varphi(\sum_{i=1}^{n} w_i\, x_i)$$

де нелінійна функція φ(функція активації) перетворює сумарний імпульс в вихідне значення нейрона.

У моделі Маккаллок-Піттса для цього використовувалася функція Хевісайда, в даній роботі буде розглянуто використання напівлінійних функції активації(rectified linear unit).

Об'єднуючи окремі нейрони можна отримати штучну нейронну мережу. Для цього вихідні сигнали нейрона подають на вхід наступного нейрона(рисунок 2.2). Нейронна мережа складається з декількох шарів, на кожному з яких може перебувати кілька нейронів. Шар, який приймає сигнали із зовнішнього світу, називається вхідним. Шар, який видає сигнали до зовнішнього світу - вихідним, інші шари називаються прихованими.

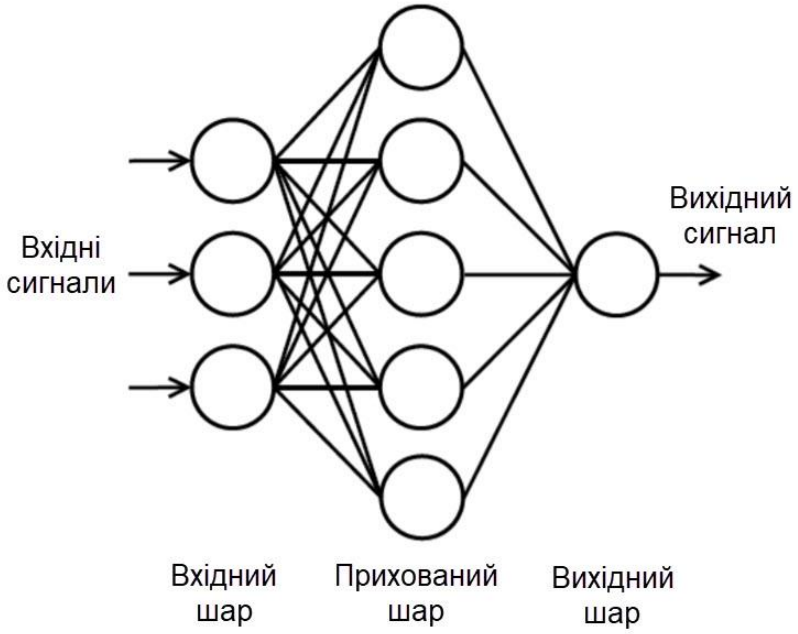


Рисунок 2.2 - Штучна нейронна мережа

Штучні нейронні мережі діляться на мережі прямого поширення сигналу(feedforward networks), в яких немає циклів, і рекурентні мережі (recurrent networks), в яких цикли дозволені.

2.2.1 Опис шарів згорткової нейронної мережі

Згорткова нейронна мережа - це нейронна мережа, архітектура якої заснована на чергуванні згортальних шарів і субдискретизуючих шарів. Вперше була запропонована в роботі[18]. Структура мережі - односпрямована(без зворотних зв'язків), багатошарова.

Шар згортки (convolution layer)

Шар згортки являє собою набір карт ознак(набір матриць). У кожної карти є ядро(так зване скануюче ядро або фільтр).

Кількість таких карт відрізняється від завдання до завдання, наприклад, якщо взяти велику кількість карт, то якість розпізнавання покращиться, але збільшиться обчислювальна складність. У більшості випадків співвідношення кількості карт ознак пропонується вибирати рівним один до двох[19]. Приклад організації зв'язку наведені на рисунку 2.3.

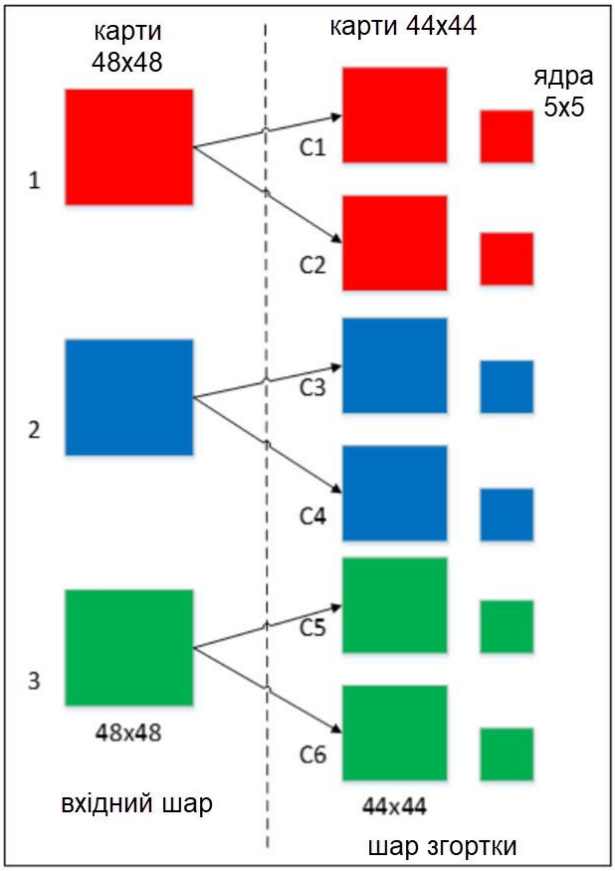


Рисунок 2.3 - Зв'язок карт ознак

Розміри карт згортального шару є однаковими і обчислюються за формулою:

$$(w, h) = (mW - kW + 1, mH - kH + 1),$$

де (w, h) - обчислюється розмір згорткової карти;

mW - ширина попередньої карти;

mH - висота попередньої карти;

kW - ширина ядра;

kH - висота ядра.

Ядро являє собою фільтр або вікно, яке ковзає по всій області попередньої карти і знаходить певні ознаки об'єктів. Наприклад, якщо мережу навчали на великій кількості фігур людей, то одне з ядер могло б в процесі навчання видавати найбільший сигнал в області руки або голови, інше ядро могло б виявляти інші ознаки. Розмір ядра зазвичай беруть в межах від 3x3 до 7x7. Якщо розмір ядра маленький, то воно не зможе виділити будь-які ознаки, якщо занадто велике, то збільшується кількість зв'язків між нейронами. Також розмір ядра вибирається таким чином, щоб розмір карт згортального шару був парний. Це дозволяє не втрачати інформацію при зменшенні розмірності в субдискретизуючому шарі.

Ядро також являє собою систему ваг або синапсів, це одна з головних особливостей згорткової нейронної мережі. У звичайній багатошаровій мережі дуже багато зв'язків між нейронами, тобто синапсів, що вельми уповільнює процес розпізнавання. У згортковій мережі - навпаки, загальні ваги дозволяють скоротити число зв'язків, що дає можливість знаходити одну і ту ж саму ознаку по всій області зображення.

Спочатку значення кожної карти згортального шару рівні 0. Значення ваг ядер задаються випадковим чином в області від мінус 0,5 до 0,5. Ядро ковзає по попередній карті і виконує операцію згортки, яка описується наступною формулою:

$$(f * g)[m, n] = \sum_{k,l} f[m - k, n - l] * g[k, l]$$

де f - вихідна матриця зображення;

g - ядро згортки.

Описати цю операцію можливо таким чином – проходимо вікном розміру ядра g із певним кроком(у даному випадку 1) все зображення f. На кожному кроці відбувається по елементне множення вмісту вікна на ядро g, і результат підсумовується і записується в матрицю результату. Візуально цей процес представлений на рисунку 2.4.

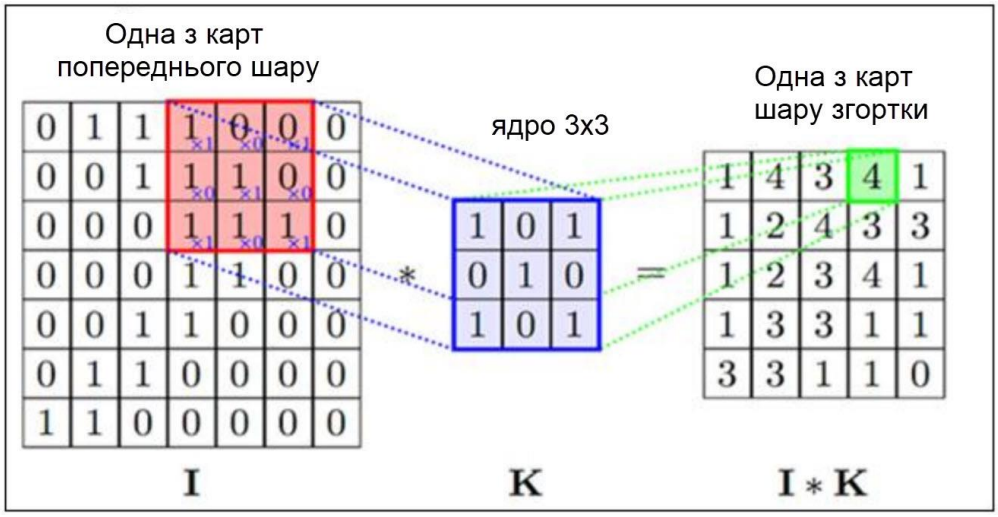


Рисунок 2.4 - Операція згортки з картою ознак

В кінцевому підсумку шар згортки можна описати формулою:

$$x^l = f(x^{l-1} * k^l + b^l)x^l,$$

де x^l - вихід шару l;

f () - функція активації;

b^l- коефіцієнт зсуву шару l;

* - операція згортки входу x з ядром k.

Варто зазначити, що через крайові ефекти розмір матриць зменшується і формула приводиться до вигляду:

$$x_j^l = f(\sum_i x_i^{l-1} * k_j^l + b_j^l),$$

де x_j^l - карта ознак j (вихід шару l);

f () - функція активації;

b_j^l - коефіцієнт зсуву шару l для карти ознак j;

k_j^l - ядро згортки j карти, шару l;

* - операція згортки входу з ядром k.

Шар активації

Скалярний результат кожної згортки потрапляє на функцію активації, яка представляє собою якусь нелінійну функцію. Вона потрібна для того, щоб додати в мережу нелінійність.

Однією з популярних функцій активації є функція ReLU(rectified linear unit)[20]. Вона набуває вигляду такої формули:

$$f(x) = \max(0, x)$$

Її графік наведено на рисунку 2.5.

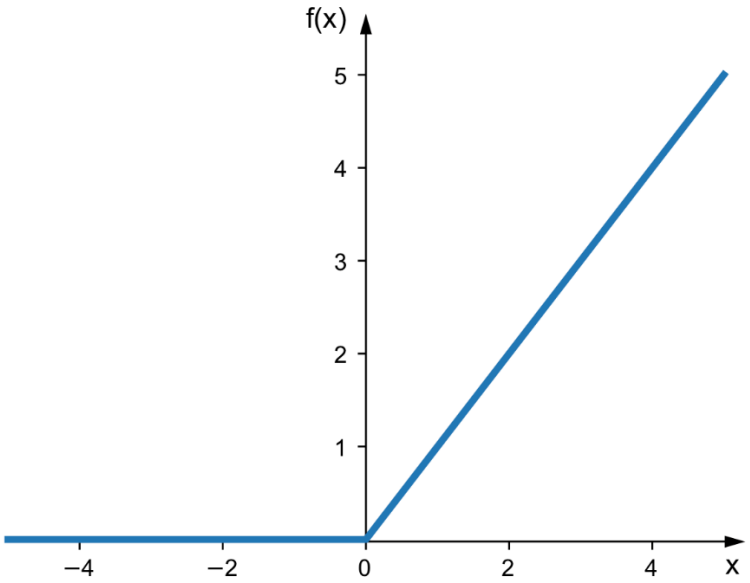


Рисунок 2.5 - Графік ReLU

Позитивними сторонами цієї функції є:

- реалізація ReLU за допомогою порогового перетворення матриці активацій в нулі, це не вимагає ресурсномістких обчислень. Крім того, ReLU не схильний до перенасичення;
- використання ReLU у порівнянні із використанням сигмоїду і гіперболічного тангенсу, значно підвищує швидкість стохастичного градієнтного спуску. Це обумовлено, насамперед, лінійним характером функції та відсутністю насичення.

Негативною стороною ReLU є те, що функції не завжди достатньо надійні. У процесі навчання можуть вони можуть виходити з ладу. Інколи, коли великий градієнт проходить ReLU, це може призвести до оновлення ваг і як результат - даний нейрон більше не активується. Якщо відбудеться така подія, то градієнт, що проходить через даний нейрон буде дорівнювати нулю. Тобто, цей нейрон буде виведений з ладу і це необоротно. При великій швидкості навчання може виявитися, що велика кількість нейронів не активується. Для вирішення цієї проблеми потрібно вибрати оптимальні параметри швидкості навчання.

Також були розроблені модифікації цієї функції, їх графіки наведені на рисунку 2.6.

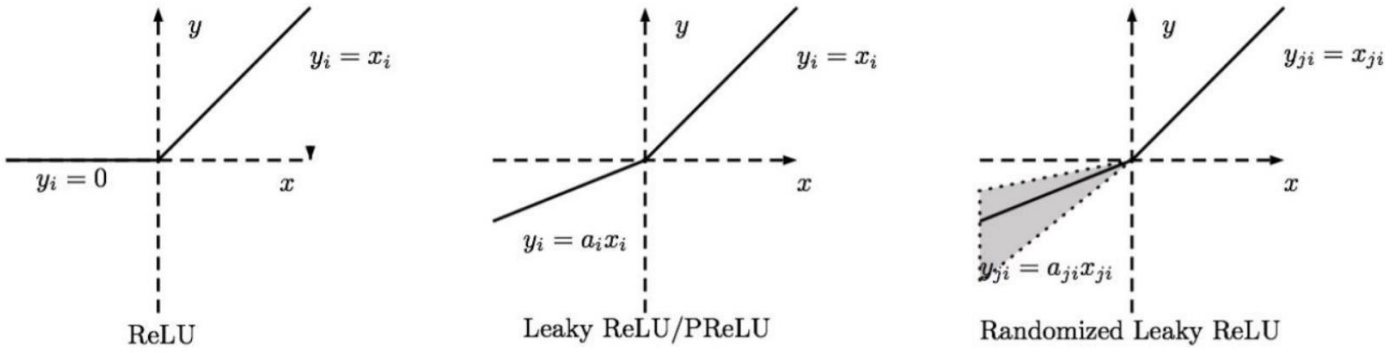


Рисунок 2.6 - Модифікації ReLU

Слід зазначити, що за результатами дослідження[20] дані модифікації перевершили стандартний ReLU і оптимальними для вибору вважаються параметричний ReLU (PReLU) і рандомізований ReLU (RReLU).

Субдискретизуючий шар (subsampling, pooling layer)

Субдискретизуючий шар (підвибірковий шар) має карти, але їх кількість співпадає з попереднім(згортальним) шаром. Мета шару - зменшення розмірності карт попереднього шару. Якщо на попередній операції згортки вже були виявлені деякі ознаки, то для подальшої обробки настільки докладне зображення вже не потрібно, і воно ущільняється до менш докладного. До того ж фільтрація вже непотрібних деталей перешкоджає перенавчанню.

В процесі обробки ядром підвибіркового шару карт попереднього шару, скануюче ядро не перетинається на відміну від згортального шару. Зазвичай, кожна карта має ядро розміром 2x2, що дозволяє зменшити попередні карти згортального шару в 2 рази. Вся карта ознак поділяється на осередки 2x2 елемента, з яких вибираються максимальні за значенням.

Зазвичай, в підвибірковому шарі застосовується функція активації ReLU. На рисунку 2.7 наведена операція підвибірки.

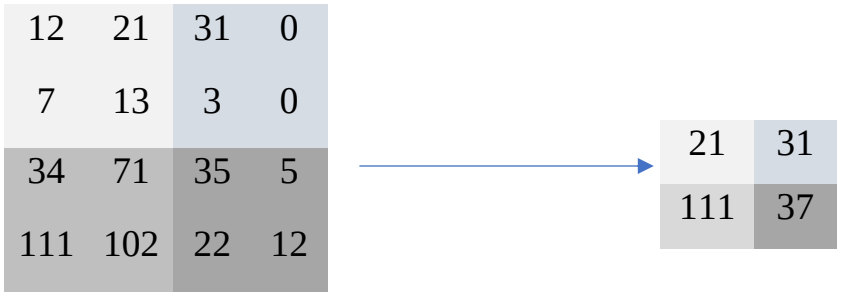


Рисунок 2.7 - Операція підвибірки, формування нового шару(2x2 Max-Pool)

Дану операцію можна описати наступною формулою:

$$x^l = f(a^l * subsample(x^{l-1}) + b^l),$$

де x^l - вихід шару l;

f () - функція активації;

a^l, b^l - коефіцієнти зсуву шару l;

$subsample$ () - операція вибірки.

Повнозв'язних шар

Повнозв'язний шар можна розглядати як звичайний шар багатошарового перцептрона(рисунок 2.8). Мета шару - класифікація, моделює складну нелінійну функцію, оптимізуючи яку, поліпшується якість розпізнавання.

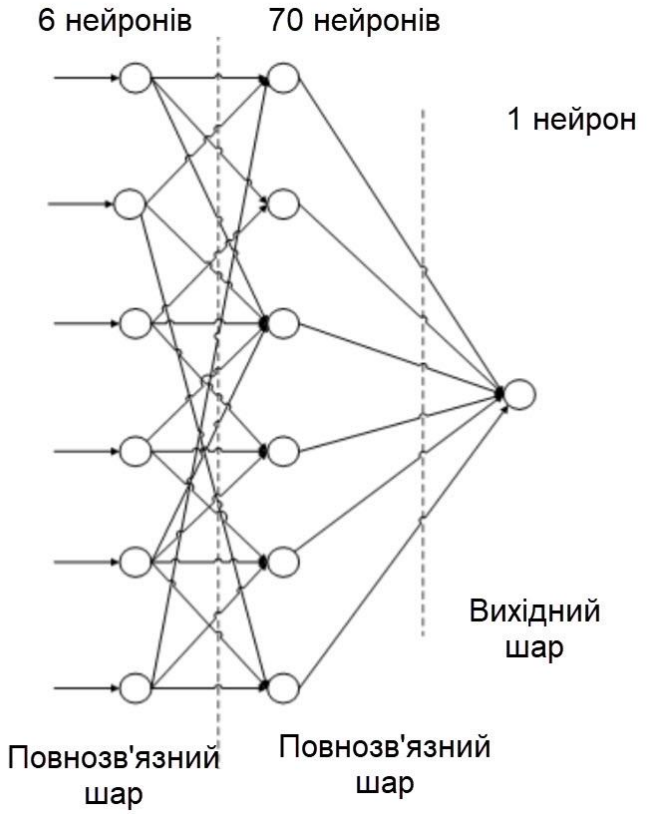


Рисунок 2.8 - Схема повнозв'язного шару

Нейрони кожної карти попереднього субдискретизуючого шару пов'язані з одним нейроном прихованого шару. Таким чином число нейронів прихованого шару дорівнює числу карт субдискретизуючого шару, але зв'язки можуть бути не обов'язково такими. Наприклад, тільки частина нейронів будь-якої з карт субдискретизуючого шару може бути пов'язана з першим нейроном прихованого шару, а частина, що залишилася з

другим, або всі нейрони першої карти пов'язані з нейронами 1 і 2 прихованого шару. Обчислення значень нейрона можна описати формулою:

$$x_j^l = f\left(\sum_i x_i^{l-1} * w_{i,j}^{l-1} + b_j^{l-1}\right),$$

де x_j^l - карта ознак j (вихід шару l);

f () - функція активації;

b^l - коефіцієнт зсуву шару l;

$w_{i,j}^l$ - матриця вагових коефіцієнтів шару l.

2.2.2 Навчання згорткової мережі

На початковому етапі нейронна мережа є ненавченої(неналаштована). У загальному під навчанням розуміють послідовне подання образу на вхід нейронної мережі з навчального набору, потім отриману відповідь порівнюють з бажаним виходом. В нашому випадку це 1 - образ представляє людину, мінус 1 - образ представляє фон або якийсь інший об'єкт(не людину), отримана різниця між очікуваною відповіддю і отриманим результатом і є функцією помилки(дельта помилки). Потім цю дельту помилки необхідно поширити на всі пов'язані нейрони мережі,

Алгоритм зворотного поширення помилки

Для навчання нейронних мереж дуже часто використовується алгоритм зворотного поширення помилки(backpropagation)[21]. Цей алгоритм є першим і наразі основним для навчання багатошарових нейронних мереж.

За цим алгоритмом ваги прихованого нейрона повинні змінюватися прямо пропорційно помилці тих нейронів, з якими даний нейрон пов'язаний. Ось чому зворотне поширення цих помилок через мережу дозволяє коректно налаштовувати ваги зв'язків між усіма шарами.

У цьому випадку величина функції помилки зменшується і мережа навчається. Величина помилки обчислюється за такою формулою:

$$E_p = \frac{1}{2}\sum_j (t_{pj} - y_{pj})^2,$$

де E_p- величина функції помилки для образу p;

t_{pj} – бажаний вихід нейрона j для образу p;

y_{pj} - активований вихід нейрона j для образу p.

Неактивований стан кожного нейрона j для образу p записується у вигляді зваженої суми за формулою:

$$S_{pj} = \sum_i w_{ij}\, y_{pi},$$

де S_{pj} - зважена сума виходів пов'язаних нейронів попереднього шару на вагу зв'язку, по-іншому ще позначається як неактивований стан нейрона j для образа p;

w_{ij} - вага зв'язку між i і j нейронами;

y_{pi} - активований стан нейрона j від попереднього шару для образа p.

Вихід кожного нейрона j є значенням активаційної функції fj, яка переводить нейрон в активований стан. Як функції активації може використовуватися будь-яка безперервно диференціююча монотонна функція. Активований стан нейрона визначається за формулою:

$$y_{pj} = f_j(S_{pj}),$$

де y_{pj}- активований стан нейрона j для образу p;

f_j - функція активації;

S_{pj} – неактивований стан нейрона j для образа р.

Для мінімізації помилки найчастіше використовується градієнтний спуск і його модифікації. Суть цього методу зводиться до пошуку мінімуму або максимуму функції за рахунок руху вздовж вектора градієнта.

Загальноприйнятим записом помилки нейрона є символ δ, тобто:

$$\delta = \frac{\partial E}{\partial y_j}.$$

Для вихідного шару помилка обчислюється взяттям похідної від формули середньоквадратичної помилки:

$$E_p^f = \sum_j (t_{pj} - y_{pj}).$$

Для розрахунку помилки прихованого шару використовується алгоритм зворотного поширення помилки, він полягає в послідовному обчисленні помилок прихованих шарів за допомогою значень помилки вихідного шару, тобто значення помилки поширюються по мережі в зворотному напрямку від виходу до входу. Розрахунок помилки прихованого шару:

$$\delta_i = \frac{\partial y_j}{\partial S_j} * \sum_j \delta_j * w_{ij},$$

де $\frac{\partial y_j}{\partial S_j}$ - значення похідної функції активації по її аргументу для нейрона j;

δ_i - помилка нейрона i прихованого шару;

δ_j - помилка нейрона j наступного шару;

w_{ij} - вага зв'язку між нейроном i поточного (прихованого) шару і нейроном j вихідного або теж прихованого шару.

Поетапно цей алгоритм представлений так:

- пряме поширення сигналу по мережі, обчислення стану нейронів;
- обчислення значення помилки δ для вихідного шару;
- зворотне поширення: послідовно від кінця до початку для всіх прихованих шарів обчислюємо δ за формулою;
- оновлення ваг мережі на розрахункову раніше δ помилки.

Розрахунки помилок для субдискретизуючого і згортального шарів відбуваються по-іншому.

Розрахунок помилки на субдискретизуючому шарі представляється в декількох варіантах. Перший випадок, коли субдискретизуючий шар знаходиться перед повнозв'язним, тоді він має нейрони і зв'язки такого ж типу, як в повнозв'язному шарі, відповідно обчислення δ помилки нічим не відрізняється від обчислення δ прихованого шару. Другий випадок, коли субдискретизуючий шар знаходиться перед згортальним, обчислення δ відбувається шляхом зворотної згортки.

Для розрахунку помилки на згортальному шарі необхідно обчислити дельти поточного шару за рахунок знань про дельти субдискретизуючого шару. Насправді дельта похибка не обчислюється, а копіюється. При прямому поширенні сигналу нейрони підвибіркового шару формувалися за рахунок неперекриваючих слоїв сканування по згортальному шарі, в процесі якого вибиралися нейрони з максимальним значенням, при зворотному поширенні, ми повертаємо дельту помилки раніше обраному максимальному нейрону, інші ж отримують нульову дельту помилки.

2.3 Удосконалення методу розв’язання задачі розпізнавання пози людини

Існує кілька способів визначення пози людини на зображенні. У перших роботах під позою розуміли положення множини частин тіла людини, таких як гомілку, передпліччя, тулуб, голова і ін. Положення кожної частини тіла описувалося прямокутником, сторони якого могли не бути паралельні осям координат. Вирішення завдання в такій постановці виявилося складним оскільки розміри частин тіла на зображенні можуть сильно змінюватися в залежності від пози людини. Тому в роботі [23] автори визначають позу як положення суглобів, що з'єднують частини тіла людини, на зображенні. В даний час ця постановка використовується для визначення пози.

Завдання визначення пози людини полягає в знаходженні положення фіксованої множини точок на зображенні людини. Деякі з цих точок відповідають суглобам тіла людини і утворюють його віртуальний скелет. Від завдання детектування оцінка пози

відрізняється передбаченням структурованого виходу, тобто положення різних суглобів на зображенні залежить один від одного. Ця залежність визначається фізичними розмірами тіла людини. Існує два основні методи визначення пози людини на зображенні: використання моделі з набору деформованих частин і регресія положення суглобів.

Модель з набору деформованих частин

Перший з розглянутих підходів об'єднує етапи локалізації окремих суглобів скелета і вибору найбільш ймовірної конфігурації пози людини в рамках загальної задачі мінімізації. Особливість даного методу полягає в можливості оцінити правдоподібність будь пози на даному зображенні. Іншими словами, модель з набору деформованих частин задає ймовірнісний розподіл пози людини на зображенні.

Модель з набору деформованих частин є розширенням стандартних методів локалізації для кращої обробки об'єктів, що складаються з декількох частин. Особливістю даного підходу є можливість враховувати допустимі зміни взаємного положення частин об'єкта відносно один одного. Вперше цей підхід був застосований для виявлення об'єктів в сцені, але в подальшому його застосували для визначення пози людини на зображенні.

Об'єкт моделюється марківської мережею, вершини якої відповідають потрібним суглобам, а ребра задають обмеження на їх взаємне розташування. У роботах [22] автори обмежуються розглядом тільки парних потенціалів, заданих на ребрах. Для того, щоб виведення в графічній моделі був точним і ефективним, автори обмежуються розглядом тільки моделей у вигляді дерева.

У загальному вигляді модель з набору частин визначає позу P людини як мінімум функції енергії, описуваної двома типами потенціалів (факторів):

$$E(P) = \sum_{i \in V} \varphi_i(p_i, s) + \sum_{(i,j) \in E} \psi^8_{(i,j)}(p_i, p_j, s)$$

де φ_i - унарний потенціал суставу pi,

$\psi^8_{(i,j)}$ - задає парний потенціал для суставів p_i і p_j на зображення,

s - дискретний параметр розмірів людини.

В роботі [23] визначено, що парний потенціал ψ_i не залежить від вхідного зображення. Завдяки використанню глобального прихованого параметру розміру s тіла людини, вдається уникати ситуацій, коли в найденій позі одні частини непропорційно більші інших.

Одним із недоліків моделей із набору деформованих частин, висвітлений у роботі [23], є деревовидна структура залежностей між суглобами. Наприклад, положення колін не мають прямої залежності, і пов'язані лише через положення суглобів тулуба. Це призводить до того, що алгоритм може розташувати суглоби обох ніг людини на зображенні однієї ноги. Для вирішення цієї проблеми в роботі [4] пропонується розширити модель людини набором позлетів(в англійській версії poselet), які обмежують взаємне розташування деякої підмножини суглобів тіла людини. Незважаючи на те, що отримана графічна модель більше не є деревом, алгоритм виведення залишається ефективним, оскільки допускає перебір по невеликій множині стану позлетів.

Базова модель [23] передбачає, що всі суглоби тіла людини видно на зображенні. Це стає серйозною проблемою в ситуаціях часткової видимості тіла людини на зображенні через перекриття об’єкта і частковим виходом людини за діапазон зображення. У роботах [24] автори визначають чи є суглоб перекритим зображенням іншої людини.

В роботі [25] вказується важливість моделювання зовнішності людини для більш точної оцінки її пози. Автори розширили модель пози людини гістограмою кольорів зображення і запропонували метод спільної оцінки параметрів, який підвищує точність визначення пози.

У роботах [26] автори використовують евристичні ознаки для опису зображення. Навчання параметрів графічної моделі відбувається за допомогою структурного методу опорних векторів [27]. В роботі [28] автори показують, що виведення в моделі з набору деформованих частин може бути представлений у вигляді згорткової нейронної

мережі. Це дозволило навчати ознаки зображення спільно з параметрами графічної моделі.

В роботі [27] автори показали, що модель з набору деформованих частин дозволяє знаходити не тільки одну позу, що мінімізує $E(P)$, але і шукати інші мінімуми функціоналу, що відрізняються положенням хоча б одного суглоба від попередніх. Велика кількість таких мінімумів описує велику кількість гіпотез по виявленню найкращих поз людини на зображенні. Коли кількість необхідних гіпотез значно менше кількості можливих положень одного суглоба на зображенні, обчислювальна складність побудови таких гіпотез не більше ніж в два рази перевершує складність побудови кращої гіпотези пози людини.

Можливості побудови гіпотез пози людини на зображенні і розширення мінімізованого функціоналу дозволили застосувати його для визначення пози людини на відео.

Регресія положення суглобів

Альтернативним підходом до визначення пози людини на зображенні є метод регресії положення суглобів з зображення. На відміну від моделі з набору деформованих частин, він не дозволяє оцінити якість довільної пози на зображенні, але може враховувати положення груп суглобів.

В роботі [28] автори побудували відображення вхідного зображення в координати кожного суглоба. Запропонований ними алгоритм являє каскад з двох нейронних мереж, які послідовно уточнюють положення кожного суглоба на зображенні. Автори показали, що перший регрес вказує наближене положення суглобів всього тіла, в той час як другий уточнює положення окремих суглобів. В роботі [28] автори вказали, що пророкування теплової карти положення суглобів дозволяє домогтися кращих результатів в рамках того ж підходу.

В роботі [29] автори розширили попередніх підхід. Для уточнення локалізації кожного суглоба вони запропонували використовувати також теплові карти положення

інших суглобів. Такий підхід найбільш близький до моделі з набору деформованих частин, але він дозволяє неявно враховувати положення всіх суглобів на зображенні при їх уточненні. Найбільш складною для даного методу є ситуація наявності на одному зображенні декількох людей, зображення яких частково перекривають один одного.

Визначення пози людини на відео

Модель з набору деформованих частин допускає розширення на випадок послідовності кадрів. Для цього вводиться модель руху, що описує зміну пози між кадрами. Найбільш простим її варіантом є завдання незалежних моделей руху для кожного суглоба:

$$E(\{P_t\}_t^T) = \sum_{t=1}^{T} E(P) + \sum_{i \in V} \sum_{t=1}^{T-1} \psi_i^8 \left(p_i^{t+1}, p_i^t, s^t\right)$$

де $E\ (P)$ - модель пози людини на зображенні.

В роботі [27] пропонується проста модель зміни пози, що припускає її незначну зміну між кадрами. В якості парного потенціалу вибирається квадратична функція зміни положення суглобів між кадрами.

Пошук мінімуму функціоналу виявляється складним завданням, так як відповідна графічна модель містить цикли. Точний висновок виявляється неможливим через його високу обчислювальну складність. Тому автори [27] використовували побудову найкращих гіпотез пози людини на зображенні, щоб зменшити кількість допустимих поз і звести задачу визначення оптимального стану в марківському ланцюгу.

У дисертації ми пропонуємо узагальнення мінімізуючого функціоналу, що враховує швидкість руху кожного суглоба скелета людини. Також пропонується алгоритм пошуку його локального мінімуму, як по множині положень суглобів, так і за параметрами їх швидкості.

Позою людини в момент часу t є послідовність точок $P^t \in \{\{p_{i=1}^t | p_i^t \in R^2\}$ на зображені.

У даній роботі розглядається алгоритм визначення пози людини в якості наступного етапу обробки відеоданих після супроводу об'єкта. Тому постановка задачі визначення пози в відеопослідовності має наступний вигляд:

Вхід:

- відеопослідовність $I = \{I_t\}_{t=1}^N$
- траєкторія руху людини $T = \{b^t\}_{t=1}^N$

Вихід: поза людини на кожному кадрі $P = \{P^t\}_{t=1}^N$.

Відома траєкторія руху людини обмежує область зображення, на якій проводиться визначення його пози.

Розглядається розпізнавання пози людини на відео як задачу мінімізації функції енергії $E(P, \theta | I, T)$, де θ – приховані параметри моделі. Параметр θ може включати як приховані параметри пози людини на одному кадрі (наприклад, параметр розміру), так і глобальні параметри моделі людини (кольорова модель). Функція енергії $E(P, \theta | I, T)$ задає ненормовану функцію правдоподібності пози на відео у вигляді $p\tilde{}(P, \theta | I, T) =$ учз$(-E(P, \theta | I, T))$. В подальшому, для спрощення викладок будемо припускати залежність функції енергії від початкового відео кадру і положення тіла людини в кожному кадрі, тобто $E(P, \theta) = E(P, \theta | I, T)$.

Модель спостережуваних даних є узагальненням моделі пози людини на зображені на випадок відеопослідовності. Для цього базова модель розширюється припущенням про залежність пози людини на різних кадрах. Стандартний підхід до узагальнення базової моделі відповідає наступній функції енергії:

$$E(P, \theta) = \sum_{t=1}^{T} E_I\left(P^{t+1}, P^t, \theta\right)$$

Де $E_I(P^{t+1})$ – модель пози людини в кадрі,

а $E_I(P^{t+1}, P^t)$ – модель зміни пози між кадрами.

Таку модель можна розглядати як марківський ланцюг першого порядку, де стан в кожен момент часу є багатовимірною величиною і описує позу людини. Вона складається з двох частин:

- базова модель пози людини на зображенні $E_I(P^t, \theta))$
- модель руху суглобів $E_T(P^{t+1}, P^t, \theta)$

Найбільш простим способом задання моделі зміни пози є припущення про незалежність руху суглобів:

$$E_T(P^{t+1}, P^t, \theta) = \sum_{i=1}^{K} \psi_i^t \, (p_i^{t+1}, p_i^t, \theta)$$

Так як моделі руху різних суглобів схожі, то досить розглянути її лише для одного з них. Для спрощення позначень в даному підрозділі опускається індекс розглянутого суглоба. Наприклад, P^t використовується для позначення стану розглянутого суглоба на кадрі $I^t t$. Таке розширення моделі пози людини на зображенні на випадок відеопослідовності використовувався також в попередніх роботах. Наприклад, в роботі [26] пропонувалася модель руху, яка передбачає слабку зміну пози людини між кадрами:

$$\psi^t(P^{t+1}, P^t, \theta) = \frac{1}{2s^{t2}} (P^{t+1} - P^t)^{T-1} (\sum\nolimits_p^p)^{-1} (P^{t+1} - P^t)$$

Таким чином, оптимальне значення такої моделі руху досягається при сталості пози людини в відео. Зміна пози при русі виявляється «допустимим шумом».

У даній роботі ми розширюємо цю модель руху. Використовується лінійна динамічна система для опису руху суглобів тіла людини. Для цього прихований стан θ моделі розширюється характеристикою руху кожного суглоба. У роботі розглядається лінійна модель руху суглобів, тобто стан кожного суглоба описується його положенням P^t на кадрі і миттєвою швидкістю руху $v^t \in R^2$. За аналогією пози людини, позначається швидкість всіх суглобів у відеопослідовності через V. Якщо

позначити через $h^t = [P^t, \mathrm{v}^t]$ стан розглянутого суглоба пози людини на кадрі t, то запропонована модель руху набуває вигляду:

$$\sum_{t=1}^{T-1} \psi\,(P^{t+1},\, P^t, \theta) = \sum_{i=1}^{K} \psi_i^0$$

$$\sum_{t=1}^{T-1} \Psi(P^{t+1}, P^t, \Theta) = \sum_{i=1}^{K} \left(\psi_i^0(v_i^1) + \sum_{t=1}^{T-1} \psi_i^t(h_i^{t+1}, h_i^t, \Theta) \right) + \sum_{t=1}^{T-1} \eta^t(s^{t+1}, s^t)$$

2.4 Висновки до розділу

В ході аналізу було сформульовано послідовність алгоритмів для попередньої обробки вхідних даних для подачі на нейронну мережу. Були детально описані алгоритми, а також принципи роботи шарів згорткових нейронних мереж, які використовуються для роботи із зображеннями і послідовностями зображень: згортальних шарів і шарів підвибірки. Наведено методи передачі даних між шарами. Наведено опис функцій активації і методів, які будуть використовуватися в розробці архітектури власної системи для розпізнавання пози людини на зображеннях та відео. У розділі був запропонований алгоритм оцінки пози людини в відеопослідовності, що враховує положення кожного суглоба тіла людини на кадрі.

3 ОПИС СИСТЕМИ РОЗПІЗНАВАННЯ ПОЗИ ЛЮДИНИ

3.1 Опис функціональності системи

Функціональні і нефункціональні вимоги

Для вирішення задачі розпізнавання пози людини повинно бути створено систему, що забезпечує можливість автоматизованого розпізнавання пози однієї або декількох людей у відеофрагменті. Користувач повинен мати можливість вказати шлях доступу до даних для розпізнавання. Розроблювана система повинна відповідати таким функціональним вимогам:

- система повинна приймати на вхід потік даних користувача (зображення, відеофрагменти, відеопотік);
- система повинна виділяти на відео рухомі об'єкти та детектувати людину;
- система повинна розбивати послідовність кадрів із відеофрагментів;
- система повинна розпізнавати пози людини на відеофрагменті у реальному часі.

Оскільки існують готові бібліотеки та надбудови для створення нейронних мереж, система повинна відповідати таким нефункціональним вимогам:

- система повинна бути реалізована на мові Python;
- система повинна використовувати фреймворк для створення нейронних мереж TensorFlow;
- система повинна бути єдиною і не вимагати від користувача додаткових дій для роботи, крім запуску та задання вхідних даних.

3.2 Варіанти використання системи

Для проектування програми було використано мову графічного опису для об'єктного моделювання UML. Була побудована модель взаємодії актора «Користувач» з системою. Діаграма варіантів використання представлена на рисунку 3.1.

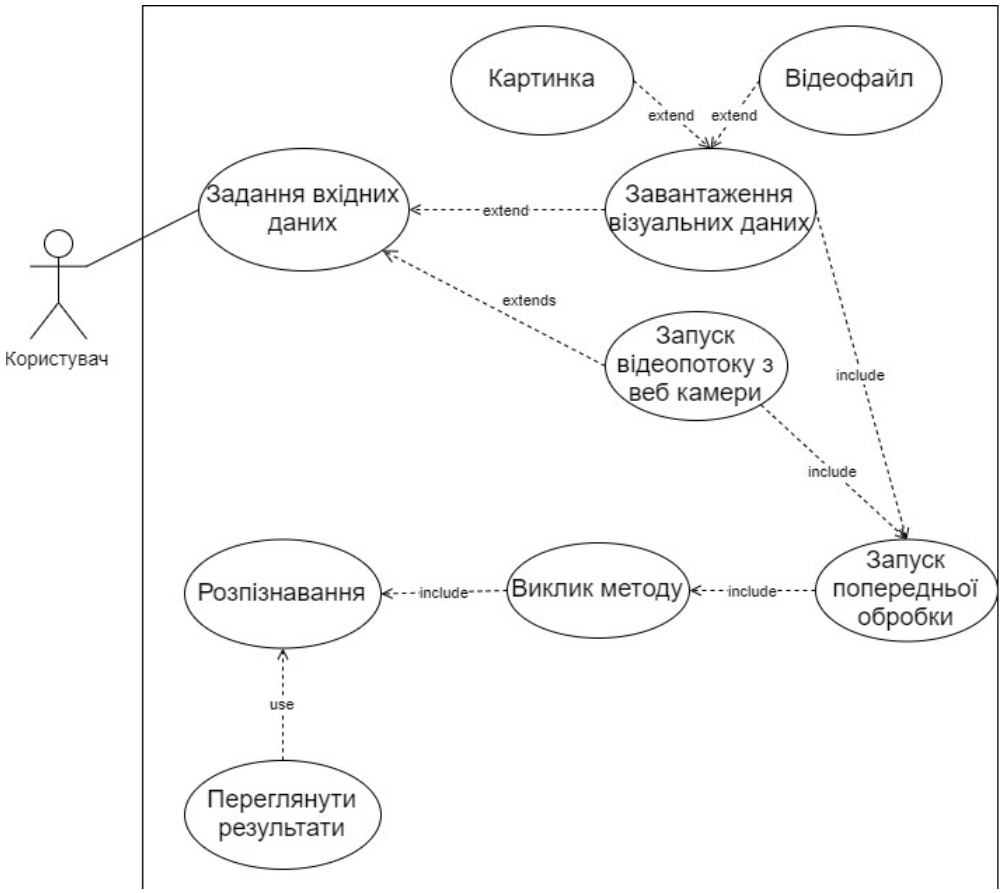


Рисунок 3.1 - Модель взаємодії актора «Користувач» з системою

Актори, які взаємодіють з системою

Користувач є єдиним актором і використовує систему для розпізнавання пози людини у режимі реального часу.

Короткий опис варіантів використання

Користувач може:

- задати вхідні дані системи - шлях до відеофрагментів і параметри програми;
- запустити попередню обробку відеопотоку;
- запустити розпізнавання пози людини за тими даними, які він поставив;
- переглянути результати роботи програми у реальному часі.

3.3 Етапи роботи системи розпізнавання пози людини

Для того, щоб розпізнати позу людини на зображені та відео необхідно провести підготовку вхідних даних та застосувати алгоритми машинного навчання для розпізнавання. Детальний опис роботи системи поданий у Додатку А.

На вхід подається зображення чи відеопотік. Для подальшої обробки, відеопотік потрібно розбити на окремі кадри. Тому перший етап – дискретизація вхідного відео потоку. На другому етапі – виокремлення точок розташування суглобів на зображені. Для цього етапу було використано загорткову нейрону мережу та архітектуру ResNet. На даному етапі отримуємо масив точок суглобів та з'єднання найближчих суглобів у пари за допомогою алгоритму полів інтенсивності частин.

На третьому етапі за допомогою лінійної регресії знайдені пари частин тіла з'єднуються у пози конкретної людини.

Розглянемо обробку вхідних даних:

- елемент вхідного зображення – зображення, відео файл, відеопотік;
- коефіцієнт масштабування зображення - число від 0.2 до 1. За замовчуванням дорівнює 0.5. Необхідний для збільшення або зменшення зображення перед тим, як подавати його на мережу. Встановлюємо цей параметр нижче, щоб зменшити зображення і збільшити швидкість при подачі через мережу за рахунок точності;
- відобразити по горизонталі. За замовчуванням використовується значення false. Параметр необхідний, якщо пози необхідно перевернути/дзеркально відобразити по горизонталі. Встановлюємо значення true для відеороликів, де відео за замовчуванням перевертається горизонтально(наприклад, веб-камера), щоб пози поверталися в правильній орієнтації;
- крок виходу. Значення за замовчуванням 16. У внутрішньому представленні цей параметр впливає на висоту і ширину шарів в нейронної мережі. На високому рівні він впливає на точність і швидкість оцінки пози. Чим нижче значення

вихідного кроку, тим вище точність і тим менше швидкість, чим вище значення, тим швидше швидкість, але і знижується точність.

Поріг для оцінки впевненості в позі - від 0 до 1. За замовчуванням 0.5. На верхньому рівні параметр контролює мінімальний показник впевненості в повертаються позах.

На виході алгоритму отримуємо:

- позу, яка містить оцінку впевненості і масив з 17 ключових точок;
- позицію і оцінку впевненості для кожної ключової точки. Знову ж таки, всі позиції ключових точок мають координати x і y у вхідному просторі і можуть відображатися прямо на зображенні.

Компоненти, що входять в систему.

Main - головний клас, що забезпечує роботу системи. Він має атрибут parameters, що зберігається в деякій структурі Parameters і містить параметри, задані користувачем. Метод startPretreatment() запускає скрипт попередньої обробки даних, звертаючись до класу Pretreatment. Метод startNN() запускає роботу нейронної мережі, звертаючись до класу NeuralNetwork.

Pretreatment - клас, який відповідає за попередню обробку даних. Він має атрибут parameters, що зберігається в деякій структурі Parameters і містить параметри, передані з головного класу. Метод cutVideo() обрізає відео до заданої довжини, frame() розбиває відео на кадри, morph починає роботу модуля Morph і scale() ініціює роботу класу Scaling.

NeuralNetwork - клас, який відповідає за роботу нейронної мережі. Він містить такі атрибути: model - програмна модель нейронної мережі, trainX - відеоряд навчальної вибірки, trainY - класи навчальних відео, testX - контрольна вибірка, testY - «відповіді» на контрольну вибірку. Метод readData() дозволяє рахувати вхідні дані, buildModel(), trainModel() виконують побудову та навчання моделі відповідно. Метод testModel() виконує перевірку контрольної вибірки і виводить результат.

Клас Scaling відповідає за підготовку кадрів до масштабування. Метод findBorders() виконує пошук крайніх точок зображеного скелета і записує їх в цілочисельні атрибути up, down, right, left. Після цього метод cut() обрізає зображення до цих крайніх точок.

Компонент MainProgram є головним модулем програми, який здійснює контроль над усіма діями системи і ініціює роботу інших компонентів, передаючи їм вхідні параметри і запускаючи їх.

NeuralNetwok - компонент, що виконує розпізнавання пози людини по відеофрагменті із записом рухів за допомогою штучної нейронної мережі і інкапсулює роботу цієї мережі.

PreprocessingScript - компонент, що забезпечує попередню обробку вхідних даних системи. Параметри попередньої обробки, включаючи розташування вхідних даних, отримані на вхід від компонента MainProgram.

Попередня обробка вхідних даних

Самим неформалізовані етапом застосування нейронних мереж для вирішення прикладних завдань, є попередня обробка даних. Іноді попередньої обробки можна уникнути або звести її до мінімуму, але при роботі з зображеннями вона грає важливу роль. З огляду на вимоги до високої швидкості розпізнавання, алгоритм попередньої обробки не повинен вимагати великих обчислювальних витрат.

В якості вхідних даних системи використовується відеофрагмент. Попередня обробка цього відеофрагменту складається із декількох етапів:

ЕТАП 1. Framing(розбиття на кадри). Оскільки нейронна мережа обробляє послідовність зображень, відеофрагмент необхідно розбити на кадри.

ЕТАП 2. Scaling (масштабування). Для оптимізації роботи нейронної мережі потрібно зменшити кількість ознак, що подаються на неї, тобто вхідні дані повинні бути масштабовані до меншого розміру. Оскільки фігура людини на різних відеофрагментах може перебувати на абсолютно різній відстані від камери, масштаб має бути таким, щоб зображення завжди займало весь розмір

рамки, і відстань до камери не вплинула на результати розпізнавання. Для цього зображення спочатку обрізається з усіх 4-х сторін для видалення порожнього фону, а після цього розтягується або стискається до необхідних розмірів.

Для попередньої обробки вхідних даних була реалізована утиліта командного оболонки bash, що послідовно виконує етапи попередньої обробки. На вхід утиліти подаються вхідне відео і параметри, що дозволяють регулювати розмір відеофрагменту і масштабування. На виході отримуємо відмаштабовані кадри, готові для подачі на нейронну мережа. У даній роботі були застосовані наступні параметри: 1 секунда тривалості фрагмента, а також розмір вхідного зображення 500x500.

3.4 Архітектура згорткової нейронної мережі для вирішення задачі розпізнавання пози людини

Для вирішення завдання з розпізнавання пози людини на зображеннях та відео потокіах була обрана згорткова нейронна мережа ResNet (Residual Network). Архітектура даної мережі реалізує ідею передачі значень виходу і входу двох послідовно розташованих згорткових шарів для наступних шарів (рисунок 3.6)[31].

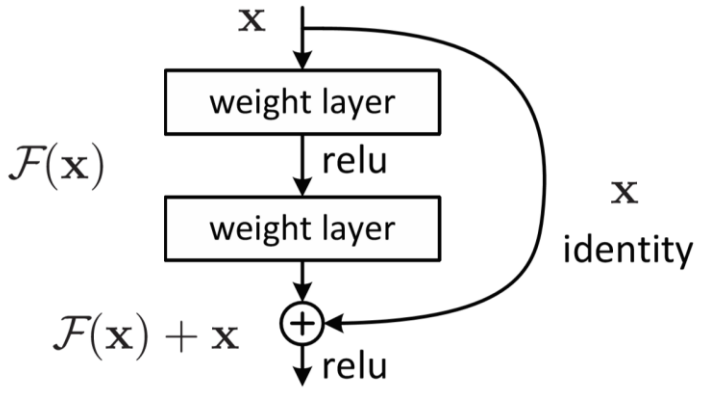


Рисунок 3.6 - Будівельний блок ResNet[31].

Це перший вид нейронний мереж, в якому кількість шарів може бути дуже великим без ризику деградації мережі. Так, розміри деяких мереж можуть досягати 1000 шарів, і це не призводить до зростання помилки класифікації. Таким чином, дана архітектура дозволяє впоратися з так званої деградацією нейронної мережі, коли на певному етапі збільшення шарів більше не дає позитивного результату, а лише збільшує помилку (дане погіршення якості роботи ніяк не пов'язане з перенавчанням).

Архітектура ResNet має велику кількість варіантів: ResNet-50, ResNet-101, ResNet-152 та інші. Число в назві позначає кількість шарів даної нейронної мережі. Для вирішення спеціалізованого завдання розпізнавання пози людини була обрана архітектура ResNet-50. Такий вибір був зроблений на підставі того, що даний вид нейронної мережі вимагає найменших обчислювальних витрат для навчання.

Ми пропонуємо розширити поняття полів в оцінці поз, щоб вийти за рамки скалярних та векторних полів до складених полів. Ми представляємо нову архітектуру нейронної мережі з двома головними мережами. Для кожної частини тіла або суглоба одна головна мережа передбачає показник достовірності, точне розташування та розмір цього суглоба, який ми називаємо картою інтенсивності частин (КІЧ). Інша головна мережа передбачає асоціації між частинами, що називається полем асоціації частин (ПАЧ), що має нову композиційну структуру. Система має можливість зберігати дрібнозернисту інформацію на картах низької роздільної здатності. Точний регрес до місця розташування суглобів є критичним, і в даній роботі використовуємо втрату типу L1 на основі Лапласа [33]. Експерименти показують, що даний розроблений метод перевершує методи «знизу вгору», як і встановлені методи «зверху вниз» для зображень із низькою роздільною здатністю.

На рисунку 3.7 представлена архітектура розробленої системи. Це спільна базова мережа ResNet з двома головними мережами: одна головна мережа передбачає надійність, точне розташування та розмір суглоба, які ми називаємо картою інтенсивності частин (КІЧ або англ. MIF), а інша головна мережа прогнозує асоціації між частинами, що називається полем асоціації частин (ПАЧ або англ. PAF).

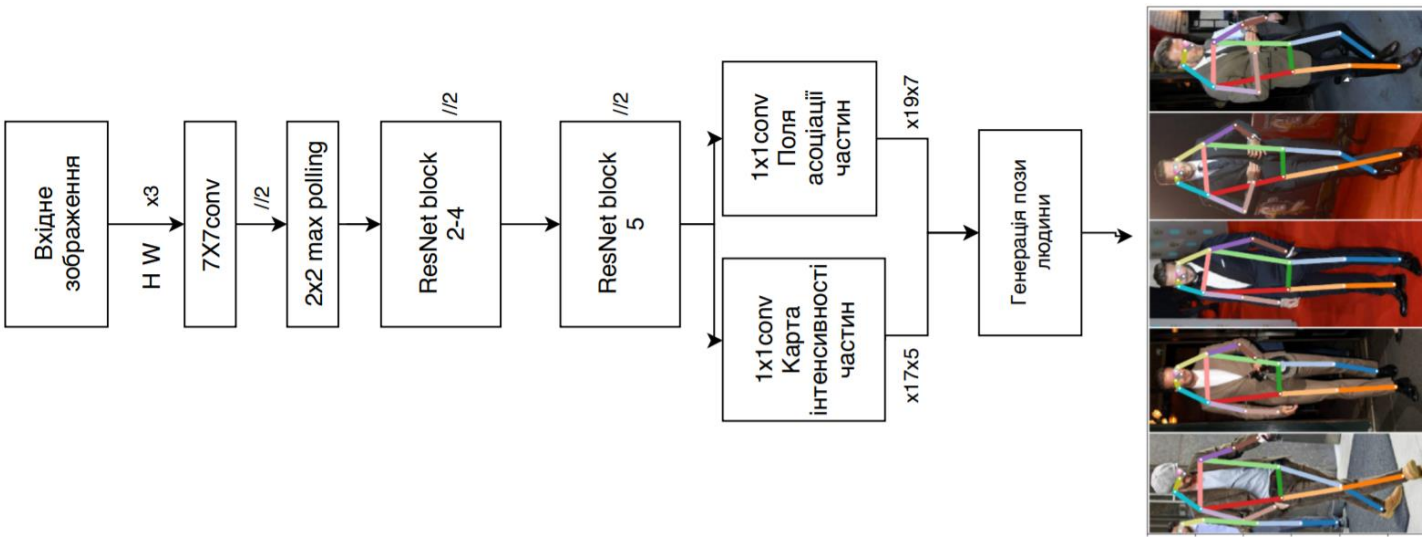


Рисунок 3.7 - Архітектура моделі розпізнавання пози людини

На вході отримуємо зображення із певними параметрами довжини та ширини та 3-ма кольоровими каналами. Encoder на основі нейронної мережі виробляє карти КІЧ та ПАЧ з каналами 17 × 5 та 19 × 7. Декодер - це програма, яка перетворює поля КІЧ та ПАЧ в оцінку пози, що містять 17 стиків. Кожен суглоб представлений координатою x та y та показником достовірності.

Алгоритм визначення декількох поз здатний виділити більше однієї пози на зображенні. Він трохи повільніший і складніший, ніж алгоритм визначення однієї пози, але має перевагу: якщо на зображенні кілька людей, виявлені для них ключові точки з меншою ймовірністю пов'язані з неправильною позою. З цієї причини, навіть якщо необхідно виявити позу тільки одну людину, цей алгоритм може бути більш підходящим.

Більш того, перевагою алгоритму є те, що кількість людей на вхідному зображення не впливає на продуктивність. Час обчислення буде однаковим незалежно від того, чи буде виявлено 15 осіб або 5.

Карти інтенсивності частин

Карти інтенсивності частин виявляють та точно локалізують частини тіла. У роботі[3] було введено синтез карти достовірності з регресією для виявлення

ключових точок. У даній статі ми резюмуємо цю техніку мовою складених полів і додаємо шкалу σ як новий компонент для формування нової карти інтенсивності частин.

КІЧ мають складну структуру. Вони складаються зі скалярного компонента для достовірності, векторного компонента, який вказує на найближчу частину тіла конкретного типу та іншого скалярного компонента за розміром суглоба. Більш формально, у кожному вихідному місці *(i, j)* КІЧ прогнозує достовірність c, вектор *(x, y)* з розворотом b та шкалу σ і може бути записаний як:

$$p^{ij} = \{p_c^{ij}, p_x^{ij}, p_y^{ij}, p_b^{ij}, p_\sigma^{ij}\}$$

Коли мережа обробляє зображення, повертається теплова карта разом із векторами зсуву, які можна декодувати так, щоб знаходити в зображенні області з високим показником впевненості для ключових точок пози. Карта довіри MIF дуже груба. На рисунку 3.8(а) показана карта впевненості для лівого плеча на прикладі зображення. Щоб покращити локалізацію цієї карти довіри, ми з'єднаємо її з векторіальною частиною MIF, показаною на рисунку 3.8(б), у карту довіри з високою роздільною здатністю.

Створюється мапа довіри з високою роздільною здатністю *f(x, y)* із згортковим ненормалізованим ядром Гаусса N шириною $p\sigma$ над регресуючими цілями з карти інтенсивності частин, зваженого рівнем довіри *pc:*

$$f(x, y) = \sum_{ij} p_c^{ij} N(x, y | p_x^{ij}, p_y^{ij}, p_\sigma^{ij})$$

Це рівняння підкреслює безмережевий характер локалізації. Вивчається просторова міра σ суглоба, як частина поля. Приклад показаний на рисунку 3.8(в). Отримана карта локалізованих суглобів використовується для генерації поз та оцінки місця новостворених суглобів.

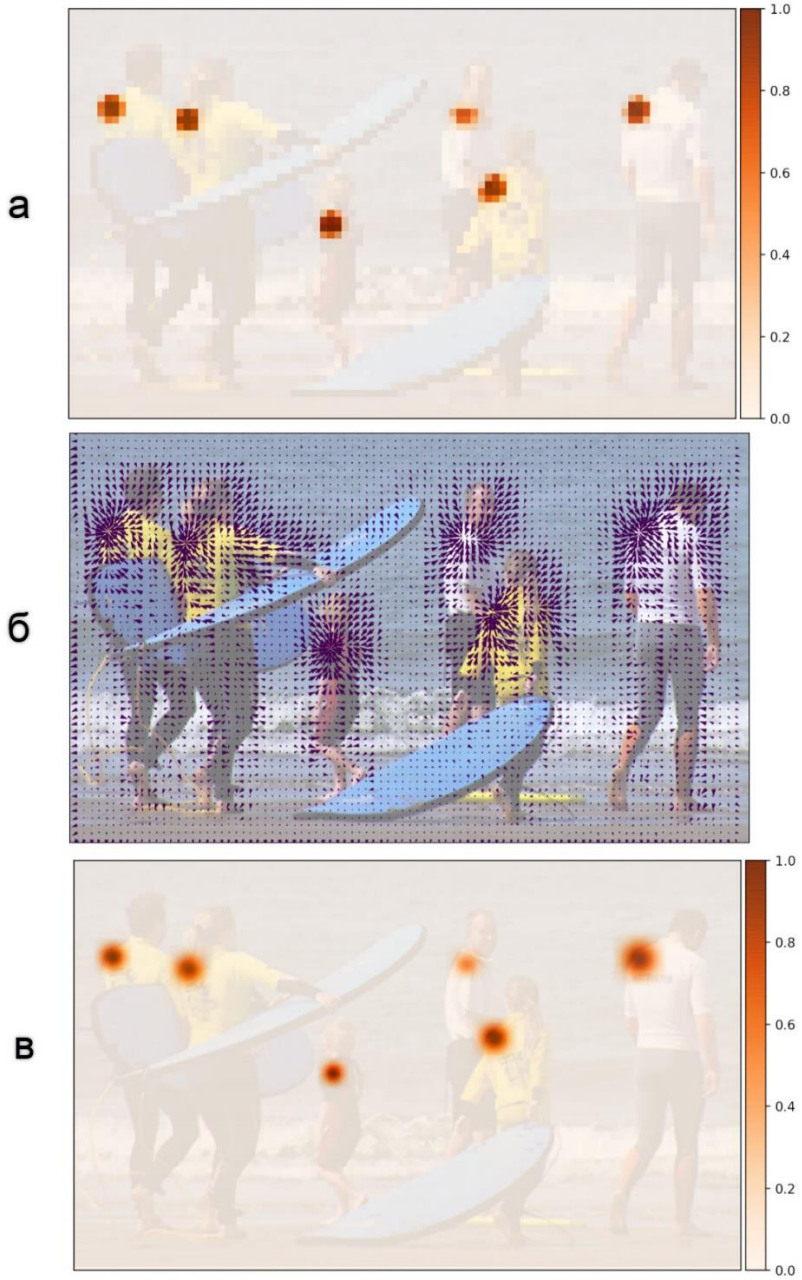


Рисунок 3.8. - Візуалізація компонентів КІЧ для лівого плеча. Це один із 17 складових КІЧ. Карта достовірності показана в (а), а векторне поле показано в (б). Поєднання карти достовірності, компонентів вектору і масштабу показані у (в).

Поля асоціації частин

Поєднання суглобів у декілька поз є складним завданням у багатолюдних сценах, де люди частково перекривають один одного. Двокрокова обробка методами

«зверху вниз» ефективна в цій ситуації: спочатку виявляються особи, що обмежують діапазони, а потім іде процес знаходження одного типу стику для кожного обмежувального поля.

Ми пропонуємо застосування поля асоціації частин(ПАЧ), щоб з'єднати спільні розташування частин разом у пози. Ілюстрована схема ПАЧ показана на рисунку 3.9.

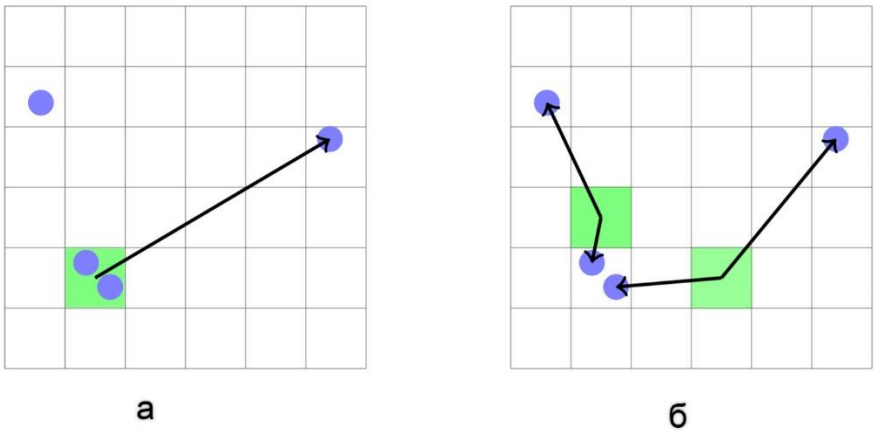


Рисунок 3.9 - Візуалізація різниці між зміщення середнього діапазону PersonLab(а) та ПАЧ(б) на сітці карт функції.

Сині кола являють собою стики, а достовірність позначена зеленим кольором. Зсуви середнього діапазону (а) починаються із центра комірок карт. Поля асоціації частин (б) мають точність з плаваючою точкою.

На виході для кожного такого розташування, ПАЧ прогнозують ступінь довіри, два вектори до двох частин, з якими з'єднується ця асоціація, і дві ширини b для просторових точок регресії. ПАЧ представляються як:

$$a^{ij} = \{a_c^{ij}, a_{x1}^{ij}, a_{y1}^{ij}, a_{b1}^{ij}, a_{x2}^{ij}, a_{y2}^{ij}, a_{b2}^{ij},\}$$

Візуалізація асоціацій між лівим плечем і лівим стегном показані на рисунку 3.10.

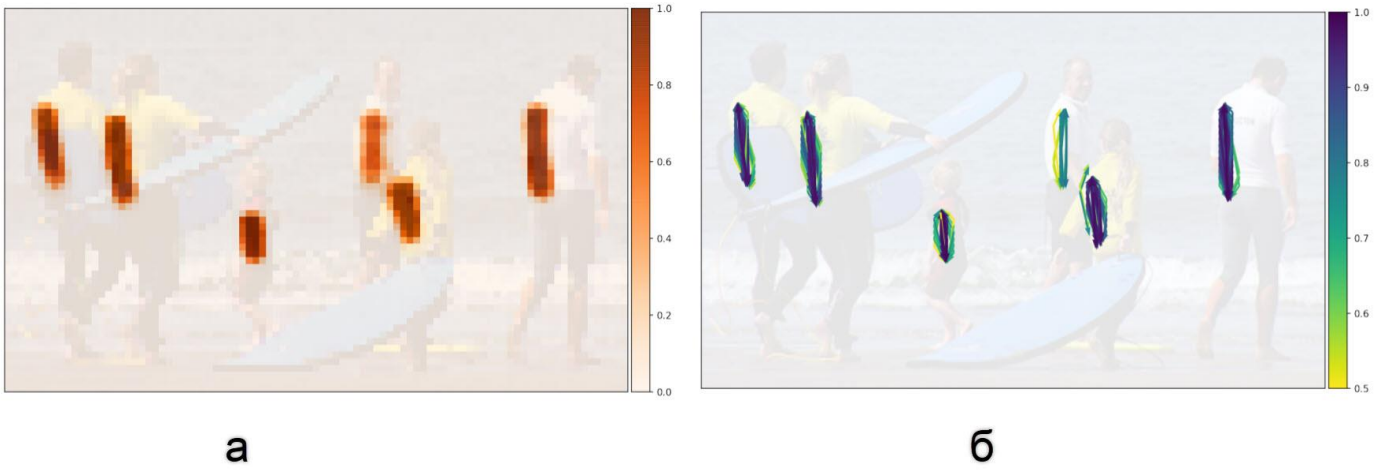


Рисунок 3.10 - Візуалізація компонентів ПАЧ, що асоціює ліве плече з лівим стегном.

Це один із 19 полів ПАЧ. Кожне розташування карти зображень - це походження двох векторів, які вказують на плечі та стегна для асоціювання. Початкова довіра асоціацій *ac* показана в частині (а) та векторні компоненти для *ac> 0,5* показані в частині(б).

Обидві кінцеві точки локалізовані з регресіями, на яких не впливає дискретизація, оскільки вони будуються в методах на основі сітки. Це допомагає точно знаходити спільні розташування осіб, які близько знаходяться один від одного, та розбирати їх на чіткі примітки для подальшого аналізу.

У наборі даних COCO є 19 з'єднань для класу person, кожен з яких з'єднує два типи стиків. Наприклад, існує асоціація праве-коліно-до-правої-щиколотки. Алгоритм побудови компонентів ПАЧ у конкретному розташуванні карти функцій складається з двох етапів. Спочатку знаходиться найближчий стик будь-якого з двох типів, який визначає одну з векторних складових.

Далі визначається інший векторний компонент, який представляє асоціацію. Другий суглоб не обов’язково є найближчим і може знаходитись далеко.

Під час навчання компоненти поля повинні вказувати на частини, які мають бути пов’язані. Аналогічно тому, як *x* компонент векторного поля завжди повинен

вказувати на ту саму ціль, що й компонент *y*, компоненти поля ПАЧ повинні вказувати на одне об'єднання частин.

Adaptive Regression Loss

Алгоритми оцінки пози людини повинні враховувати різноманітність масштабів, які може мати людська поза на зображенні. Хоча помилка локалізації суглоба великої людини може бути незначною, ця сама абсолютна помилка може бути головною помилкою для параметрів маленької людини. Ми використовуємо втрати типу L1 для тренування та підготовки виходів регресії. Ми вдосконалюємо здатність до локалізації нашої мережі, вводячи залежність масштабу в цю втрату регресії за допомогою SmoothL1 або втрати Лапласа[33].

Втрата SmoothL1 дозволяє налаштувати радіус r навколо початку координат, де він створює більш м'які градієнти. Для обмежувальної області особи A_i та розміру ключової точки σ_k, $r_{i,k}^{smooth}$ можна задати формулою $\sqrt{A_i \sigma_k}$.

Втрата Лапласа - це ще одна втрата типу L1, яка розріджена за рахунок передбачуваного поширення *b*:

$$L = \frac{|x - \mu|}{b + log\ (2b)}$$

Це не залежить від будь-яких оцінок A_i і σ_k і ми будемо використовувати його для всіх векторних компонентів.

Greedy Decoding

Розшифровка - це процес перетворення вихідних карт функцій нейронної мережі в набори з 17 координат, які проводять оцінку пози людини. Нова поза з’єднується із векторами ПІЧ з найвищими значеннями в довільній карті високої роздільної здатності f(x,y). Далі додаються з'єднання до інших суглобів за допомогою полів ПАЧ. Після того, як встановлено з'єднання з новим суглобом, це рішення рахується остаточним.

Кілька асоціацій ПАЧ можуть утворювати з'єднання між поточним та наступним суглобом. З огляду на розташування початкового суглоба ~ *x*, розрахунки *s* асоціацій ПАЧ *a* обчислюються як:

$$s(a, \overrightarrow{x}) = a_c \exp(-\frac{||\overrightarrow{x} - \overrightarrow{a_1}||_2}{b_1}) f_2(a_{x2}, a_{y2})$$

що враховує впевненість у цьому з'єднанні a_c, відстань до місця розташування першого вектору, відкалібровану з ймовірністю розподілу Лапласа та високу роздільну здатність частини в цільовому розташуванні другого вектору *f2*.

Для підтвердження запропонованої позиції нового стику(суглобу) ми проводимо процес зворотної відповідності. Цей процес повторюється до отримання повної пози. Ми застосовуємо не максимальне стиснення на рівні ключових точок. Радіус стиснення динамічний і заснований на передбачуваній шкалі компонента ПІЧ. Ми не уточнюємо жодних полів ні під час тренувань, ні під час тестування.

Використані бібліотеки для навчання: OpenCV

Розглянемо open-source бібліотеку OpenCV[34]. Дана бібліотека в основному направлена на знаходження образів в режимі real-time. Широке використання даного інструмента обумовлене його відкритістю, кросплатформеністю та великою кількістю літератури та навчальних посібників. Вона написана на C/C++, її вихідний код відкритий. Бібліотека включає близько 1000 функцій і алгоритмів. Розробкою займались ще з 1998 року такі компанії як Intel, Itseez при активній участі спільноти. Число завантажень бібліотеки перевищує 6000000 разів, що свідчить про її популярність.

Бібліотека розповсюджується за ліцензією BSD. Тобто, її можливо використовувати абсолютно безкоштовно у проектах з відкритим кодом та в комерційних проектах. Бібліотеку не потрібно переносити повністю у свій проект, можливе використання лише відповідної частини коду. Для застосування на безоплатній основі потрібно дотримуватись вимозі ліцензії – в супроводжуючих матеріалах потрібно вказувати копії ліцензії OpenCV.

В основаному, OpenCV – бібліотека алгоритмів та функцій по обробці зображень та відео. Приведемо приклад доступних модулів:

- opencv_video – займається аналізом та відстежуванням об'єктів на відео, також усуває фон;
- opencv_gpu – модуль, який підключає обчислення на GPU за рахунок CUDA (Nvidia);
- opencv_imgproc – обробка зображень (фільтри, перетворення);
- opencv_ml – інструменти машинного навчання (SVM, дерево прийняття рішень).

Навчальні вибірки

Для того, щоб заощадити часовий ресурс для навчання нейронної мережі, було прийнято рішення про використання наявних навчальних вибірок. COCO(Common Objects in Context)[35] видає 18 ключових точок (19 точка це фон).

Кожна така точка має свої координати на зображенні для кожного кадру(фрейму). З цього випливає, що для кожного кадру можна побудувати скелет, а також впровадити свою функцію постобробки.

3.5 Експерименти та оцінка роботи системи

Камери в автомобілях, що керуються безпілотно, мають широке поле зору і повинні знаходити усіх пішоходів у межах цього поля зору. Ми хочемо наслідувати розподіл пішоходів за висотою пікселів із загальнодоступним набором даних та протоколом оцінки пози людей.

Крім того, щоб продемонструвати можливість широкого застосування нашого методу, ми також досліджуємо оцінку пози в контексті завдання реідентифікації людини (Re-Id). Тобто, отримуючи зображення людини, ототожнення цієї людини на інших зображеннях.

Ми кількісно оцінюємо запропонований нами метод використовуючи COCO датасет для розпізнавання людей на зображеннях із низькою роздільною здатністю. З початкового набору даних COCO, ми обмежуємо максимальну довжину сторони

зображення до 321 пікселів, щоб імітувати обрізання зображення 4-к камери. Ми отримуємо обмежувальні діапазони, які мають висоту 66 ± 65 пікселів. Показники COCO містять розподіл для образів людей середнього розміру у AP^Mта AR^M, які мають початкову площу обмежувального поля між $(32\ \text{px})^2$ та $(96\ \text{px})^2$. Після зміни розміру для зображення низької роздільної здатності це відповідає обмежувальним полям висотою 44 ± 19 px.

Ми ретельно вивчаємо ефективність нашого методу на знімках, знятих автомобілями, що рухаються самостійно, а також на випадкових сценаріях з великою кількістю людей.

У контексті Re-Id ми досліджуємо популярний та загальнодоступний набір даних Market-1501. Він складається із 64 × 128 піксельних crops пішоходів. Ми застосовуємо ту саму модель, яку ми навчали за даними COCO.

Завдання виявлення ключових точок COCO оцінюється як завдання виявлення об'єкта, при цьому основні показники є варіантами середньої точності (average precision - AP) та середньою межею відкликання (average recall - AR). COCO передбачає фіксоване відношення розміру ключової точки до обмежувальної площі поля для кожного типу ключових точок для визначення подібності об'єкта з ключовими точками (object keypoint similarity – OKS). Для кожного зображення оцінювач поз повинен передбачати розташування 17 ключових точок для кожної пози та здійснити підрахунок для кожної пози. Для оцінювання беруться лише 20 кращих підрахункових підсумків з балами.

Для опорних значень Mask R-CNN та OpenPose ми застосували реалізацію по роботам [3, 7], модифікованій для забезпечення максимальної довжини зображення. Маска R-CNN була модифікована для застосування при низькій роздільній здатності зображення. Результат нашого методу отримали на основі ResNet50.

Зокрема, ми використовуємо 64115 зображень у навчальному наборі COCO за 2017 рік, який має анотацію для тренінгу. Наша перевірка проводиться на наборі валідації COCO 2017 року з 5000 зображень. Базові мережі - це модифіковані мережі

ResNet50/101/152. Головні мережі - це одношарові 1x1 пікселеві згортки, що подвоюють просторову роздільну здатність. Довірча складова поля нормалізується сигмоподібною нелінійністю.

Ми застосовуємо лише декілька незначних доповнень даних. Для створення рівномірних груп ми обрізаємо зображення до квадрату, де сторона квадрата становить від 95% до 100% від короткого краю зображення, а місце розташування вибирається випадковим чином. Ми застосовуємо обрізання фото, щоб зберегти якомога більше даних про навчання мережі. Половину часу все зображення використовується необрізаним, а додаються смуги, щоб воно було квадратним. Подальша зміна розміру використовує бікубічну інтерполяцію. Навчальні зображення та примітки випадковим чином перевертаються горизонтально.

Компоненти полів, що утворюють карти довіри, навчаються з незалежними бінарними перехресними втратами ентропії. Ми використовуємо втрати L1 для компонентів масштабу полів КІЧ та використовуємо втрати Лапласа для всіх векторних компонентів.

Під час тренінгу ми фіксуємо статистику запущених операцій нормування партії до їх попередньо перевірених значень. Ми використовуємо оптимізатор SGD зі швидкістю навчання 10^{-3}, імпульсом 0,95, розміром партії 8 та відсутністю ваг. Ми використовуємо усереднення моделей для вилучення стабільних моделей для перевірки. На кожному кроці оптимізації ми оновлюємо експоненціально зважену версію параметрів моделі. Наша константа розпаду дорівнює 10^{-3}. Час навчання 75 епох ResNet101 на двох GTX1080Ti становить приблизно 95 годин.

Ми порівнюємо запропонований нами метод з відтворюваними сучасними методами OpenPose та методом Mask R-CNN. Хоча наша мета - перевершити підходи «знизу вгору», ми все ще аналізуємо результати підходу «зверху вниз», щоб оцінити роботу нашого методу. Оскільки це емуляція маленьких людей на великому зображенні, ми модифікували існуючі методи, щоб запобігти збільшенню масштабів невеликих зображень.

У таблиці 3.1 представлені наші кількісні результати з використанням датасету COCO. Ми перевершуємо OpenPose метод і навіть підхід Mask R-CNN по всім показникам. Ці цифри в цілому нижчі, ніж їх аналоги для зображень із більшою роздільною здатністю. Два концептуально дуже різні базові методи демонструють подібну ефективність, тоді як наш метод явно випереджає більш ніж на 18% в AP.

Таблиця 3.1 - Застосування методу оцінки пози для зображень низької роздільної здатності з довгою стороною, що дорівнює 321 пікселів

	AP	$AP^{0.50}$	$AP^{0.75}$	AP^{M}	AP^{L}	AR	$AR^{0.50}$	$AR^{0.75}$	AR^{M}	AP^{L}
Mask R-CNN	42.6	67.1	41.5	27.2	58.9	48.1	75.0	49.9	36.7	66.5
OpenPose	38.6	61.5	36.1	24.1	55.7	44.0	64.8	45.7	25.4	67.5
Наш метод	48.3	74.1	51.7	36.1	68.6	56.0	76.4	57.9	38.9	75.7

Наші кількісні результати імітують розподіл людей на міських вуличних сценах, використовуючи загальнодоступний, анотований набір даних. Використовуючи наш метод вдалось досягнути мінімальних показників помилково виявлених поз. На рисунку 3.12 показані якісні результати на вуличних сценах. Ми також виявляємо пішоходів, які частково перекривають один одного. Критичний жест, такий як «махання» до машини, виявляється лише за допомогою нашого методу. І Mask-RCNN, і OpenPose не точно оцінили жест руки на зображені. Така різниця може бути основоположною при розробці безпечних автомобілів, що рухаються самостійно.

Рисунок 3.12. Візуалізація нашого алгоритму (праворуч) проти OpenPose [3] (перша колонка) на наборі даних nuScenes

Рисунок 3.13 - Візуалізація нашого алгоритму (праворуч) проти Mask RCNN [4] (перша колонка) на наборі даних nuScenes

Система виділяє прямокутниками всіх людей, яких інші методи не виявивили, а з колами всі помилково визначені пози. Зауважимо, що наш метод правильно оцінює жест «маханання» (перший ряд, перший прямокутник) людини, тоді як інші методи цього не роблять.

Для кількісної оцінки ефективності на датасеті Market-1501 ми створили спрощену метрику точності. Точність становить 43% для Mask R-CNN і 96% для нашого методу. Оцінка базується на кількості зображень із правильним розпізнаванням пози із 202 випадкових зображень. У правильній позі розміщено до трьох суглобів.

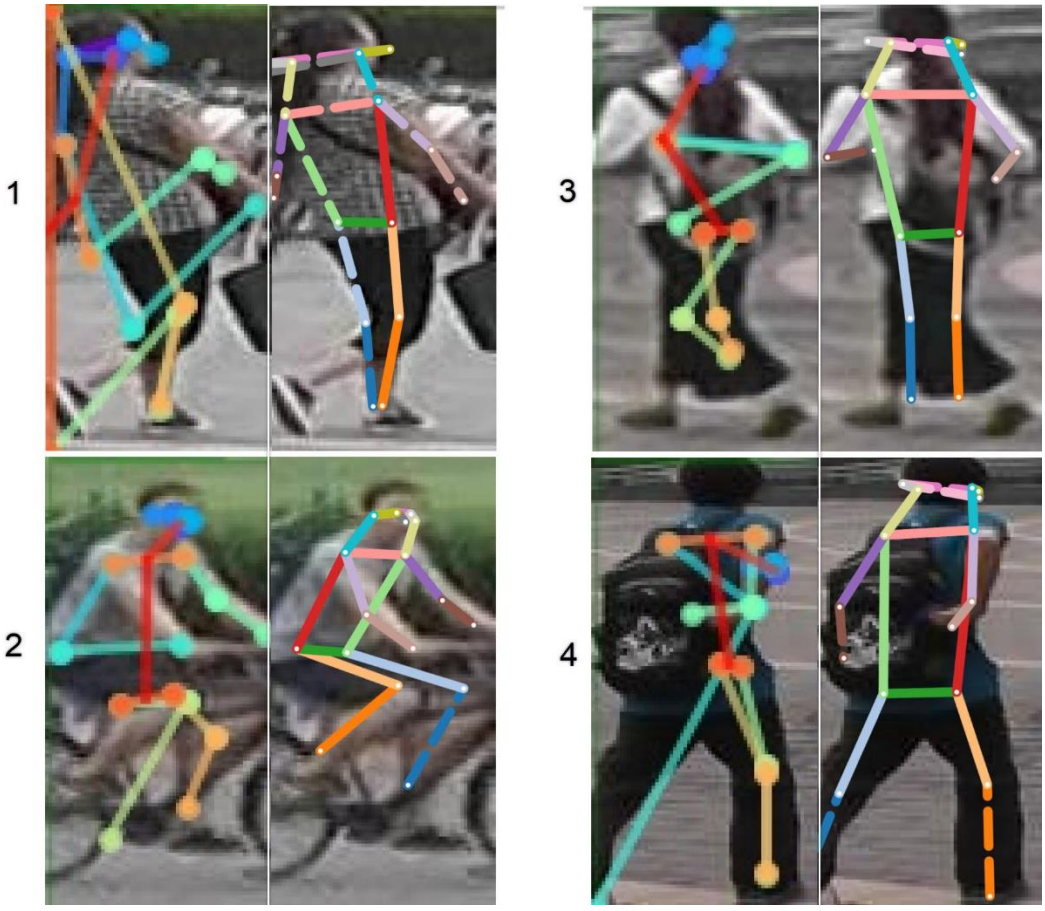


Рисунок 3.14 - Зображення ліворуч на кожному із 4 прикладів - вихід із Mask R-CNN.

Щоб покращити результат Mask R-CNN, ми задали параметри щоб метод передбачав рівно одну позу в обмежувальному діапазоні, яке охоплює ціле зображення.

Праве зображення - це результат роботи нашої системи, яка не обмежувалась однією людиною та могла обрати на вихід жодну чи декілька поз, що є складнішим завданням(рисунок 3.14).

Таблиця 3.2 - Показники у відсотках, оцінені на датасеті COCO 2017, встановлені при оптимальній роздільній здатності

	AP	AP^M	AP^L
Mask R-CNN	62.1	58.1	70.6
OpenPose	60.9	56.8	67.1
PersonLab	64.4	62.3	72.1
Наш метод	66.5	62.7	72.9

Інші методи оптимізовані для зображення з більшою роздільною здатністю. Для порівняння ми показуємо кількісне порівняння на датасеті COCO 2017 з високою роздільною здатністю, показаному в таблиці 3.2. Як видно із результатів, система працює нарівні з найкращими існуючим методами знизу вгору.

Таблиця 3.3 - Вивчення залежності від типу втрати L1

	AP	AP^M	AP^L
Vanilla L1	36.5	22.9	55.6
SmothL1, r = 0.2 $\sqrt{A_i}\ \sigma_k$	36.7	23.4	55.5
SmothL1, r = 0.5 $\sqrt{A_i}\ \sigma_k$	36.2	22.8	55.2
SmothL1, r = 1.0 $\sqrt{A_i}\ \sigma_k$	34.9	21.8	54.6
Laplace	41.8	29.3	57.4
Laplace (використовуючи b в декодері)	43.0	28.9	60.1

Показники у таблиці 3.3 відображаються у відсотках. Усі моделі мають за основу ResNet-50 і пройшли навчання протягом 10 епох.

Ми вивчали залежності від типу втрати L1, які узагальнені в таблиці 3. Ми виявили, що можемо налаштувати виконання по відношенню до менших або більших об'єктів, змінюючи загальну шкалу r-smooth і тому ми проаналізували вплив цього параметра. Однак справжнє покращення результатів одержуємо за допомогою використання втрати на основі Лапласа. Доданий компонент масштабу σ до поля ПІЧ покращив AP нашої моделі ResNet101 від 65% до 66,1%.

Таблиця 3.4 - Взаємозв'язок між точністю розпізнавання та часом виводу одиночного зображення t з різною кількістю шарів ResNet на датасеті COCO val.set.

	AP [%]	t [мс]	t(dec) [мс]
ResNet50	61.8	237	191
ResNet101	66.1	251	186
ResNet152	67,9	276	185

Остання колонка – це час декодування t(dec). Показники для різних списків ResNet наведенні у таблиці 4. За тією ж основою ми перевершуємо PersonLab на 10,0% в AP з одночасним 29% приростом у швидкості. Дані для PersonLab на ResNet101 60.0% та 355 мс.

3.6 Висновки до розділу

Розроблено модифікований метод «знизу вгору» для оцінки 2D-пози людини на зображеннях із кількома людьми, який може бути ефективно застосований в транспортній галузі, особливо в автомобілях з автопілотом та соціальних роботах. Продемонстровано, що спроектована система перевершує попередні методи в режимі

низької роздільної здатності та працює нарівні з існуючими методами в режимі з високою роздільною здатністю.

4 РОЗРОБКА СТАРТАП ПРОЕКТУ

У розділі проведено маркетинговий аналіз стартап проекту для визначення можливості ринкового впровадження системи розпізнавання пози людини та можливих напрямів реалізації. Проведення маркетингового аналізу передбачає виконання нижченаведених кроків[36].

4.1 Опис основної ідеї проекту

В межах підпункту послідовно проаналізовано та подано у вигляді таблиці:

- зміст ідеї – система розпізнавання пози людини у відеопотоці;
- можливі напрямки застосування;
- основні вигоди, що отримає користувач;
- чим відрізняється розроблена система від існуючих аналогів та замінників.

Перші три пункти подані у вигляді таблиці (таблиця 4.1), що дає цілісне уявлення про зміст ідеї та потенційні ринки.

Таблиця 4.1- Опис ідеї стартап-проекту

Зміст ідеї стартап-проекту	Напрямки застосування	Вигоди для користувача
Розпізнавання пози людини на відео	Системи спостереження	Покращення точності та швидкості розпізнавання пози для реагування
	Система керування автопілотних автомобілів	Швидше і більш точне детектування людей на зображенні для автоматичного прийняття критичних рішень
	Масові події	Швидке розпізнавання та реагування на надзвичайні події

Аналіз потенційних техніко-економічних переваг ідеї стартап-проекту у порівняні із пропозиціями конкурентів передбачає:

а) визначення техніко-економічних властивостей та характеристик ідеї стартап-проекту [36];

б) визначення попереднього кола конкурентів, що існують на ринку, та проведення збору інформації щодо техніко-економічних показників для власного стартап-проекту та проектів-конкурентів;

в) проведено порівняльний аналіз показників: для власної ідеї визначені показники, що мають:

1) гірші значення (слабкі - W);
2) аналогічні значення (нейтральні - N);
3) кращі значення (сильні S) (таблиця 4.2).

Таблиця 4.2 - Визначення сильних, слабких та нейтральних характеристик

	Потенційні товари/концепції конкурентів		
	Мій проект	OpenPose	Mask R-CNN
W слабка сторона	Високі вимоги до точності розпізнавання	Високі апаратні вимоги, погано працює на зображеннях із низькою роздільною здатністю	Високі апаратні вимоги, погано працює на зображеннях із низькою роздільною здатністю
N нейтральна сторона	Мультиплатформність	Мультиплатформність	Мультиплатформність

Продовження таблиці 4.2

	Потенційні товари/концепції конкурентів		
	Мій проект	OpenPose	Mask R-CNN
S сильна сторона	Точність та швидкість розпізнавання точок скелету, робота із зображеннями низької роздільної здатності, низькі апаратні вимоги, розпізнавання поз великої кількості людей у відеопотоці	Швидкодія	Швидкодія для визначення пози однієї людини

4.2 Технологічний аудит ідеї проекту

В межах підрозділу проведено аудит технології, за допомогою якої можна реалізувати дану ідею проекту з розпізнавання пози людей бортовими комп'ютерами автопілотними транспортними засоби пересування[36].

Визначення технологічної складової для реалізації ідеї проекту передбачає аналіз складових, (таблиця 4.3):

а) за якою технологією буде виготовлено товар згідно ідеї проекту?

б) чи існують такі технології, чи їх потрібно розробити/доробити?

в) чи доступні такі технології авторам проекту?

За результатами аналізу аналогів системи можна зробити висновок, що технологічна реалізація проекту за допомогою існуючих технологічних засобів можлива.

Таблиця 4.3 - Технологічна здійсненність ідеї проекту

№ п/п	Ідея стартап-проекту	Технології її реалізації	Наявність технологій	Доступність технологій
1	Розпізнавання пози людини у вигляді скелету	Набір даних з розміченими частинами тіла людини	Наявна	Доступна безкоштовно
2		TensorFlow	Наявна	Доступна безкоштовно
3		ResNet	Наявна	Доступна безкоштовно
Висновок: реалізувати проект можливо. Обрана технологія реалізації ідеї проекту: вікно з відеопотоком користувача, розпізнавання пози людини у реальному часі				

4.3 Аналіз ринкових можливостей запуску стартап-проекту

Визначення ринкових можливостей, які можна використати під час ринкового впровадження проекту, та ринкових загроз, які можуть перешкодити реалізації проекту, дозволяє спланувати напрями розвитку проекту із урахуванням стану

ринкового середовища, потреб потенційних клієнтів та пропозицій проектів-конкурентів.

Проводимо аналіз попиту: наявність попиту та його обсяг, динаміка розвитку ринку (таблиця 4.4).

Таблиця 4.4 - Попередня характеристика потенційного ринку стартап-проекту

№ п/п	Показники стану ринку (найменування)	Характеристика
1	Кількість головних гравців, од.	5
2	Загальний обсяг продаж, грн/ум.од	1000
3	Динаміка ринку (якісна оцінка)	Зростає
4	Наявність обмежень для входу	Немає
5	Специфічні вимоги до стандартизації /сертифікації	Немає

Середня норма рентабельності в галузі (або по ринку) порівнюється із банківським відсотком на вкладення. За умови, що останній є вищим, можливо, має сенс вкласти кошти в інший проект. За результатами аналізу можемо зробити висновок, що ринок є привабливим для входження.

Далі визначимо потенційні групи клієнтів та їх характеристики, зформуємо орієнтовний перелік вимог до товару (таблиця 4.5).

Таблиця 4.5 - Характеристика потенційних клієнтів стартап-проекту

No п/ п	Потреба, що формує ринок	Цільова аудиторія (цільовий сегмент ринку)	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до системи
1	Розпізнавання пози людини у реальному часі у відеопотоці	Транспортна сфера, медичні заклади, охорона сфера	Особливості купівлі системи: компанії заключають довготривалі договори, в свою чергу, стартапи віддають перевагу випробувальному терміну Використання: компанії вимагають точну та швидку роботу програмного продукту	Стабільність роботи Наявність випробувального періоду Наявність документації Точність та швидкість роботи Підтримка різних платформ

Після визначення потенційних груп клієнтів проведемо аналіз ринкового середовища: складемо таблиці факторів, що будуть сприяти ринковому впровадженню проекту чи перешкоджати (таблиця 4.6).

Таблиця 4.6 - Фактори загроз

No п/ п	Фактор	Зміст загрози	Можлива реакція компанії
1	Низький попит	Продукт досить вузькоспрямований	Проведення активної рекламної компанії
2	Обмеженість функцій	Інструмент обмежений наявними функціями	Додавання нових функцій при виникненні потреби

Таблиця 4.7 - Фактори можливостей

No п/ п	Фактори	Зміст можливості	Можлива реакція компанії
1	Збільшення технологій доповненої реальності та автопілотних систем керування транспортними засобами	Індустрія зростає з кожним роком	Вихід на глобальний ринок
2	Відсутність повноцінних альтернатив	Існуючі аналогічні системи не надають таких результатів по розпізнаванню пози людини у реальному часі	Розширення набору існуючих функцій

Надалі проводиться аналіз пропозиції: визначаються загальні риси конкуренції на ринку. Аналіз пропозиції необхідно виконати аналізуючи існуючі види конкуренцій[36]. Пропозиції повинні відповідати на питання “Як просувати продукт”. Аналіз пропозицій зображено на таблиці.

Таблиця 4.8 Ступеневий аналіз конкуренції на ринку

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність компанії (можливі дії компанії, щоб стати конкурентоспроможною)
1. Вказати тип конкуренції - монополія/олігополія/чиста /монополістична	чиста	Укладання прямих договорів з компаніями, просування продукту на виставках на конференціях
2. За рівнем конкурентної боротьби - локальний/національний	національний	Публікація статей на міжнародних сайтах
3. За галузевою ознакою - міжгалузева/ внутрішньогалузева	внутрішньогалузева	Покращувати наявні функції
4. Конкуренція за видами товарів: - товарно-видова - товарно-родова - між бажаннями	товарно-видова	Покращувати наявний функціонал
5. За характером конкурентних переваг - цінова та нецінова	нецінова	Надання функціоналу та параметрів розпізнавання, які не надають конкуренти
6. За інтенсивністю - марочна/немарочна	марочна	Надання функціоналу, що не надають конкуренти

Таблиця 4.9 Аналіз конкуренції в галузі за М. Портером

Складові аналізу	Прямі конкуренти у галузі	Потенційні конкуренти	Постачальники	Клієнти	Товаризамінники
	OpenPose	DeepCut	Мінімізація витрат часу постачальників	Контроль якості	Лояльність споживачів
Висновки	Визначити ступінь конкурентної боротьби з боку прямих конкурентів	Можливість виходу на ринок, тому що існуючі рішення не надають аналогічнихпереваг	Постачальники підлаштовуються під ринок	Клієнти диктують вимоги опираючись на умови експлуатації	Обмеження для роботи через товари замінники

На основі аналізу конкуренції, проведеного у таблиці 4.8, а також із урахуванням характеристик ідеї проекту(таблиця 4.9), вимог споживачів до товару (таблиця 4.6) та факторів маркетингового середовища(таблиця 4.7) визначаємо та обґрунтовуємо перелік факторів конкурентоспроможності[36].

Таблиця 4.10 - Обґрунтування факторів конкурентоспроможності

No п/ п	Фактори конкурентоспроможності	Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проектів значущим)
1	Швидкість обробки	Існуючі конкуренти не мають таких алгоритмів, які здатні так швидко оцінювати пози людини на відео в режимі реального часу
2	Точність обробки	Існуючі конкуренти не мають таких алгоритмів, які здатні так точно оцінювати пози людини на відео в режимі реального часу

За визначеними факторами конкурентоспроможності(таблиця 4.10) проводимо аналіз сильних та слабких сторін даного стартап-проекту(таблиця 4.11).

Таблиця 4.11 - Порівняльний аналіз слабких та сильних сторін

№ п/п	Фактори конкурентоспроможності	Бали 1-20	Рейтинг товару-конкурентів в порівнянні з даним продуктом						
			–3	–2	–1	0	+1	+2	+3
1	Швидкість обробки	10	+						
2	Точність обробки	20			+				

Фінальним кроком ринкового аналізу можливостей впровадження проекту є SWOT-аналіз (матриця аналізу сильних(Strength), слабких(Weakness) сторін, загроз(Threds) та можливостей(Opportunities)(таблиця 4.11) на основі раніше виділених ринкових загроз, можливостей, сильних та слабких сторін.

Перелік ринкових загроз та можливостей складаємо на основі аналізу факторів загроз та можливостей маркетингового середовища.

Ринкові загрози, ринкові можливості є прогнозованими результатами впливу факторів, які ще не реалізовані на ринку та мають деяку ймовірність здійснення [36].

Таблиця 4.12 - SWOT-аналіз стартап-проекту

Сильні сторони: Швидкодія Розпізнавання пози людини на відео з частково перекритими людьми	Слабкі сторони: Високі вимоги до системи
Можливості: Популярність індустрії Відсутність альтернативних систем	Загрози: Обмеженість функціоналу

Перелік ринкових загроз та ринкових можливостей складаємо на основі аналізу факторів загроз та можливостей маркетингового середовища[36].

Таблиця 4.13 - Альтернативи ринкового впровадження стартап-проекту

No п/ п	Альтернатива (орієнтовний комплекс заходів) ринковоїповедінки	Ймовірність отримання ресурсів	Строки реалізації
1	Орієнтація поточної моделі на ринок транспорту	60%	600 год
2	Орієнтація поточної моделі на ринок медичних закладів	20%	200 год
3	Орієнтація поточної моделі на ринок спостереження та охорони людей	20%	200 год

4.4 Розроблення ринкової стратегії проекту

У даному підрозділі проаналізовано ринкові стратегії, визначено стратегії охоплення ринку: описані цільові групи потенційних споживачів(таблиця 4.14).

За результатами аналізу були обрані цільові групи потенційних споживачів, для яких пропонується дана система, та визначаються стратегія охоплення ринку [36].

Таблиця 4.14 - Вибір цільових груп потенційних споживачів

No п/п	Опис профілю цільової групи потенційних клієнтів	Готовність споживачів сприйняти продукт	Орієнтовний попит в межах цільової групи (сегменту)	Інтенсивність конкуренції в сегменті	Простота входу у сегмент
1	Ринок транспорту	Потребують переговорів	Високий	Середня	Складно
2	Медичні заклади	Готові	Високий	Висока	Просто
3	Ринок спостереження та охорони людей	Готові	Високий	Середня	Просто
Які цільові групи обрано: ринок транспорту та спостереження та охорони людей					

Для роботи в цих сегментах ринку сформуємо базову стратегію розвитку (таблиця 4.15).

Таблиця 4.15 - Визначення базової стратегії розвитку

No п/п	Обрана альтернатива розвитку проекту	Стратегія охоплення ринку	Ключові конкурентоспроможні позиції відповідно до обраної альтернативи	Базова стратегія розвитку*
1	Орієнтація поточної моделі на ринок транспорту	Стратегія концентрованого маркетингу	Корпорації, які займаються розробкою машин з автопілотним управлінням	Стратегія спеціалізації (спирається на диференціацію)
2	Орієнтація поточної моделі на ринок спостереження та охорони людей	Стратегія концентрованого маркетингу	Медичні заклади потребують точності розпізнавання	Стратегія спеціалізації (спирається на диференціацію)

Розроблення ринкової стратегії першим кроком передбачає визначення стратегії охоплення ринку: опис цільових груп потенційних споживачів.

Перелік ринкових загроз та ринкових можливостей складається на основі аналізу факторів загроз та факторів можливостей маркетингового середовища. Після визначення потенційних груп клієнтів проводиться аналіз ринкового середовища: складаються таблиці факторів, що сприяють ринковому впровадженню проекту.

Наступний крок - вибір стратегії конкурентної поведінки на ринку (таблиця 16).

Таблиця 4.16 - Визначення базової стратегії конкурентної поведінки

No п/п	Чи є стартап-проект «першопрохідцем» на ринку?	Чи буде дана компанія шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде дана компанія копіювати основні характеристики товару конкурента, і які?	Стратегія конкурентної поведінки*
1	Ні	шукати нових споживачів	Так. Потрібно збільшити набір вбудованого в систему функціоналу, потрібно зменшити кількість необхідних ресурсів для задач розпізнавання пози, для охоплення нового ринку користувачів потрібно збільшити підтримку інших платформ та мобільних пристроїв	Стратегія заняття конкурентної ніші

З обраних сегментів до по стартап-компанії та до продукту розробляється стратегія позиціонування(таблиця 4.17), що полягає у формуванні ринкової позиції, за яким споживачі мають ідентифікувати проект.

Таблиця 4.17 - Визначення стратегії позиціонування

No п/п	Вимоги до товару від цільової аудиторії	Базова стратегія розвитку	Ключові конкурентоспроможні позиції стартап-проекту	Вибір асоціацій, які мають сформувати комплексну позицію власного проекту
1	Швидкість та точність роботи Невисока ціна Підтримка декількох платформ Наявність документації	Стратегія спеціалізації (спирається на диференціацію)	Медичні заклади та системи охорони потребують швидкості розробки, яку надає підтримка багатьох платформ даним продуктом	Підтримка декількох платформ, пришвидшення роботи систем та точності розпізнавання

4.5 Розроблення маркетингової програми проекту

У таблиці 4.18 відображені результати попереднього аналізу конкурентоспроможності товару на ринку.

Таблиця 4.18 - Визначення ключових переваг концепції потенційного товару

No п/п	Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)
1	Пришвидшення розробки ПЗ	Підтримка декількох платформ, точніше визначення пози	Більшість конкурентів підтримують одну платформу
2	Робота на зображених з низькою роздільною здатністю	Можливе використання у відеопотоках при поганій видимості чи низькій роздільній здатності веб камери	Конкурент DeepCut погано розпізнає пози при поганій видимості чи низькій роздільній здатності веб камери

Далі розроблена трирівнева маркетингова модель товару(таблиця 4.19).

Таблиця 4.19 - Опис трьох рівнів моделі товару

Рівні товару	Сутність та складові		
I. Товар за задумом	Розпізнавання пози людини в режимі реального часу		
II. Товар у реальному виконанні	Властивості/характеристика	М/Нм	Вр/Тх /Тл/Е/Ор
	можливість оптимізації витрат часу	М	Тл
	можливість оптимізації витрат коштів	М	Вр
	відповідність актуальним технологіям	М	Тх

Продовження таблиці 4.19

Рівні товару	Сутність та складові
II. Товар у реальному виконанні	Відповідає вимогамДСТУ ISO/IEC 25030:2015 Програмна інженерія. Вимоги щодо якості та оцінювання програмного продукту (SQuaRE). Вимоги щодо якості
	Пакування: готовий до використання інсталятор
	Марка: PoseSkeleton
III. Товар із підкріпленням	Потенційний користувач може ознайомитись з роботою системи та її алгоритмами з наукових конференцій, наукових вісників на яких було представлено дану систему
За рахунок чого потенційний товар буде захищено від копіювання: Назва і контент захищені ліцензією MIT; захист інтелектуальної власності	

М/Нм – монотонні або немонотонні;

Вр/Тх/Тл/Е/Ор – вартісні, технічні, технологічні, ергономічні або органолептичні(останній – для продуктів харчування)

Після формування маркетингової моделі товару слід особливо відмітити – чим саме проект буде захищено від копіювання. Захист може бути організовано за рахунок захисту ідеї товару (захист інтелектуальної власності), або ноу-хау, чи комплексне поєднання властивостей і характеристик, закладене на другому та третьому рівнях товару.

Наступним етапом є визначення оптимальної системи збуту(таблиця 4.20):

- проводити збут власними силами чи залучати сторонніх посередників;
- вибір і обґрунтування оптимальної глибини каналу збуту;
- вибір і обґрунтування посередників.

Таблиця 4.20 - Формування системи збуту

No п/п	Специфіка закупівельної поведінки цільових клієнтів	Функції збуту, які має виконувати постачальник товару	Глибина каналу збуту	Оптимальна система збуту
1	Доступ до системи повинен надаватися в режимах «тріал» та «повний»	Легкість в підключені, легкість у сплаті послуг	Розробник даного продукту – компанія - Користувач	Проводити збут за допомогою посередників

Останньою складової маркетингової програми є розроблення концепції маркетингових комунікацій, що спирається на попередньо обрану основу для позиціонування, визначену специфіку поведінки клієнтів (таблиця 4.21).

Таблиця 4.21 - Концепція маркетингових комунікацій

No п/п	Специфіка поведінки цільових клієнтів	Канали комунікацій, що використовують цільові клієнти	Ключові позиції, обрані для позиціонування	Завдання рекламного повідомлення	Концепція рекламного звернення
1	Купують програмний продукт	Системи відеонагляду	Підтримка різних платформ прискорення розробки прогрманих продуктів	Довести, що система пришвидшить роботу систем, в яких буде працювати дана система	-

4.6 Висновки до розділу

Розроблена система має переваги над існуючими конкурентами, що робить її конкурентноздатною на ринку. Система має шляхи для подальшого розвитку, проаналізовано шляхи збуту та маркетингова стратегія. Основна цільова аудиторія — це компанії, що займаються розробкою автопілтоних автомобілів, охоронні служби, медичні заклади для яких важливі спрощені алгоритми розпізнавання пози людини, що дозволяє швидко та точно розпізнавати пози людини у режимі реального часу.

ВИСНОВКИ

Була розроблена система розпізнавання пози людини. Були проаналізовані існуючі методи розпізнавання пози, а також проекти, технічно близькі до даного. Було прийнято рішення проводити розпізнавання на скелетізованних даних, попередньо оброблених алгоритмами морфологічних перетворень, а також використовувати CNN архітектуру нейронної мережі.

На основі вимог до системи були виділені основні класи та компоненти системи, встановлені зв'язки між ними. Була детально розглянута реалізація окремих прецедентів.

В ході розробки системи були отримані наступні результати:

- проведено аналіз програмних аналогів і наукової літератури предметної області;
- розроблений і реалізований алгоритм попередньої обробки даних;
- сформовані навчальна і тестова вибірки;
- реалізована, навчена і протестована модель нейронної мережі для розпізнавання пози людини у режимі реального часу.

Розроблено модифікований метод «знизу вгору» для оцінки 2D-пози людини на зображеннях із кількома людьми, який може бути ефективно застосований в транспортній галузі, особливо в автомобілях з автопілотом та соціальних роботах. Продемонстровано, що спроектована система перевершує попередні методи в режимі низької роздільної здатності та працює нарівні з існуючими методами в режимі з високою роздільною здатністю.

В майбутньому планується провести роботу по підвищенню точності розпізнавання. Для цього необхідно істотно розширити навчальну вибірку і, можливо, скоригувати топологію нейронної мережі.

Етапи роботи були опубліковані у статтях: «Аналіз методів автоматичного реферування тексту за допомогою нейронних мереж»[37] та «Розпізнавання пози людини у реальному часі»[38].

ПЕРЕЛІК ПОСИЛАНЬ

1. Poselet conditioned pictorial structures / L. Pishchulin [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2013. — С. 588—595

2. Toshev A., Szegedy C. Deeppose: Human pose estimation via deep neural networks // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2014. — С. 1653—1660.

3. Zhe Cao, Tomas Simon, Shih-En Wei Yaser, Sheikh The Robotics Institute, Carnegie Mellon, University: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields // – Режим доступу: https://arxiv.org/pdf/1611.08050.pdf

4. L. Pishchulin, E. Insafutdinov, S. Tang, B. Andres, M. Andriluka, P. V. Gehler, and B. Schiele. Deepcut: Joint subset partition and labeling for multi person pose estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4929–4937, 2016

5. Du Y., Wang W., Wang L. Hierarchical recurrent neural network for skeleton based action recognition. // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2015. – Vol. 7. – No. 12. – P. 1110–1118.

6. H.-S. Fang, S. Xie, Y.-W. Tai, and C. Lu, “RMPE: Regional multi-person pose estimation,” in ICCV, 2017.

7. K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask r-cnn,” in *ICCV*, 2017.

8. Girshick R. Fast R-CNN. Режим доступу: https://arxiv.org/pdf/1504.08083

9. Козлов В.А., Потапов А.С. Анализ методов выделения движущихся объектов на видеопоследовательностях с шумами. // Научнотехнический вестник информационных технологий, механики и оптики. Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики», 2011. – Т. 3. – № 73. – С. 39–43.

10. Lucas B.D., Kanade T. An iterative image registration technique with an application to stereo vision. // Proc. 7th Int. Jt. Conference on Artificial Intelligence, 1981. – Vol. 2. – P. 674–679.

11. Horn B.K.P., Schunck B.G. Determining optical flow. // Artif. Intell. 44 Elsevier, 1981. – Vol. 17. – No. 1–3. – P. 185–203.

12. Bruhn A., Weickert J., Schnörr C. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. // International Journal of Computer Vision, Kluwer Academic Publishers, 2005. – Vol. 61. – No. 3. – P. 1–21.

13. Отслеживание движения и алгоритмы сопровождения ключевых точек. НОУ Интуит. Режим доступу: www.intuit.ru/studies/courses/10622/1106/lecture/18022

14. Abeysinghe S.S. Segmentation-free skeletonization of grayscale volumes for shape understanding. S.S. Abeysinghe, M. Baker, W. Chiu, T. Ju. // IEEE International Conference on Shape Modeling and Applications, 2008. – P. 63–71.

15. Крашенникова Ю.С. Использование процедуры «скелетизации» для выделения линий на спутниковых изображениях. Ю.С. Крашенникова, Е.А. Лупян, Т.А. Немченко, М.Ю. Захаров. // Исследование Земли из космоса, 1994. – Т. 6. – С. 43–51.

16. Применение волнового алгоритма для нахождения скелета растрового изображения. Распознавание образов и искусственный интеллект. Режим доступу: ocrai.narod.ru/vectory.html

17. McCulloch W.S., Pitts W. A Logical Calculus of the Ideas Immanent in Nervous Activity. The Bulletin of Mathematical Biophysics. 1943. vol. 5, no. 4. pp. 115–133. DOI: 10.1007/BF02478259

18. Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard and L.D. Jackel: Backpropagation Applied to Handwritten Zip Code Recognition, Neural Computation, 1(4):541-551, Winter 1989

19. Habibi, Aghdam, Hamed. Guide to convolutional neural networks: a practical application to traffic-sign detection and classification. Heravi, Elnaz Jahani, Cham, Switzerland.

20. Bing Xu, Naiyan Wang, Tianqi Chen, Mu Li. Empirical Evaluation of Rectified Activations in Convolutional Network (2015).

21. Backpropagation In Convolutional Neural Networks Режим доступу: http://www.jefkine.com/general/2016/09/05/backpropagation-inconvolutional-neural-networks/

22. Felzenszwalb P., McAllester D., Ramanan D. A discriminatively trained, multiscale, deformable part model // Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. — IEEE. 2008. — С. 1—8.

23. Yang Y., Ramanan D. Articulated pose estimation with flexible mixtures-ofparts // Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. — IEEE. 2011. — С. 1385—1392.

24. Parsing occluded people / G. Ghiasi [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2014. — С. 2401—2408.

25. Modeling Instance Appearance for Recognition–Can We Do Better Than EM / A. Chou [і ін.] // International Workshop on Structured Prediction: Tractability, Learning, and Inference. — 2013.

26. Yang Y., Ramanan D. Articulated pose estimation with flexible mixtures-ofparts // Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. — IEEE. 2011. — С. 1385—1392.

27. Park D., Ramanan D. N-best maximal decoders for part models // 2011 International Conference on Computer Vision. — IEEE. 2011. — С. 2627—2634

28. Toshev A., Szegedy C. Deeppose: Human pose estimation via deep neural networks // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2014. — С. 1653—1660.

29. Bulat A., Tzimiropoulos G. Human pose estimation via convolutional part heatmap regression // European Conference on Computer Vision. — Springer. 2016. — С. 717—732.

30. YOLO: Real-Time Object Detection. https://pjreddie.com/darknet/yolo/

COCO 2018 Keypoint Detection Task // Режим доступу: http://cocodataset.org/#keypoints-2018

31. Kaiming He, Xiangyu Zhag, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition Режим доступу: https://arxiv.org/abs/1512.03385

32. Reddit MachineLearning: How does DenseNet compare to ResNet and Inception? Режим доступу: https://www.reddit.com/r/MachineLearning/comments/67fds7/d_how_does_densenet_compare_to_resnet_and/

33. A. Kendall and Y. Gal. What uncertainties do we need in bayesian deep learning for computer vision? In Advances in neural information processing systems, pages 5574–5584, 2017. 2, 4

34. Офіційний ресурс бібліотеки OpenCV // Режим доступу: https://opencv.org/

35. COCO 2018 Keypoint Detection Task // Режим доступу: http://cocodataset.org/#keypoints-2018

36. РОЗРОБЛЕННЯ СТАРТАП-ПРОЕКТУ – Режим доступу: http://kaf-pe.kpi.ua/wp-content/uploads/2015/04/roz_startap_proektiv_met_vk.pdf

37. Лотоцька Ю.В., Халус О.А. Аналіз методів автоматичного реферування тексту за допомогою нейронних мереж // II Всеукраїнська науково-практична конференція молодих вчених та студентів – 2019 – с. 9 – 12

38. Лотоцька Ю.В., Халус О.А. Розпізнавання пози людини у реальному часі // III всеукраїнська науково-практична конференція молодих вчених та студентів – 2019

ДОДАТОК А. Графічний матеріал

Схема структурна варіантів використання системи

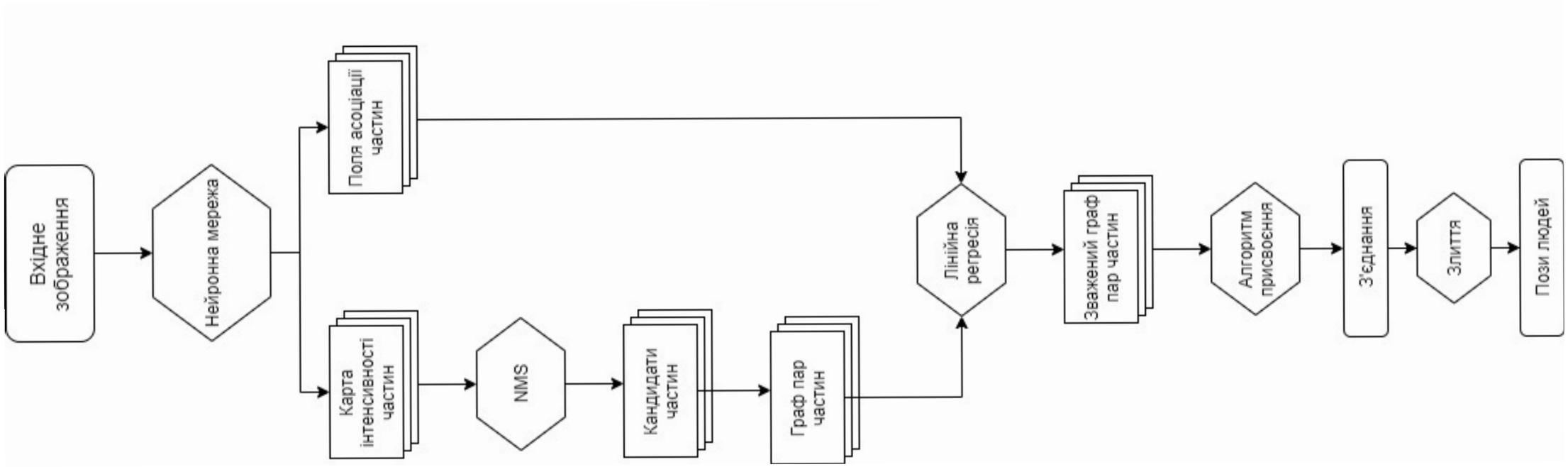
Демонстраційний плакат до магістерської дисертації

на тему «*СИСТЕМА РОЗПІЗНАВАННЯ ПОЗИ ЛЮДИНИ В РЕАЛЬНОМУ ЧАСІ*»

Виконав студент гр. ІС-81 ПІБ Лотоцька Ю.В.

Керівник ПІБ Халус О.А.

Схема роботи алгоритму системи розпізнавання пози людини



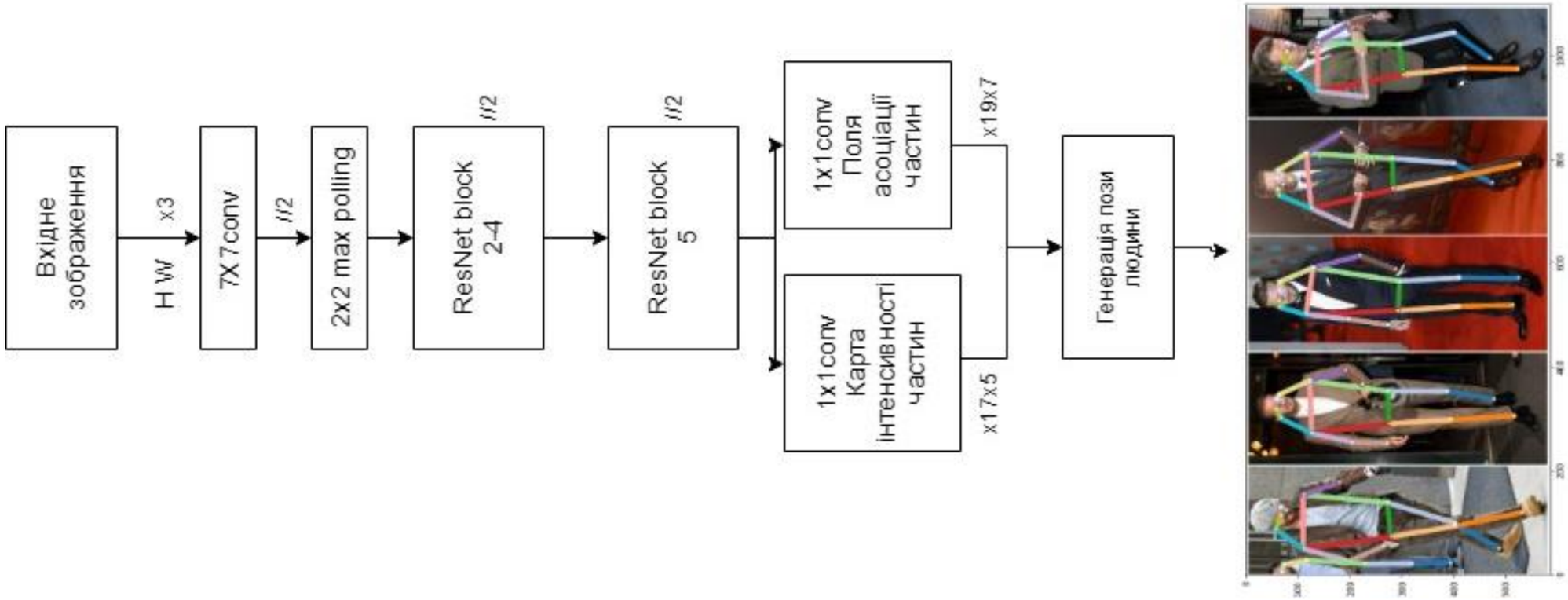
Демонстраційний плакат до магістерської дисертації

на тему «*СИСТЕМА РОЗПІЗНАВАННЯ ПОЗИ ЛЮДИНИ В РЕАЛЬНОМУ ЧАСІ*»

Виконав студент гр. ІС-81 ПІБ Лотоцька Ю.В.

Керівник ПІБ Халус О.А.

Архітектура системи розпізнавання пози людини



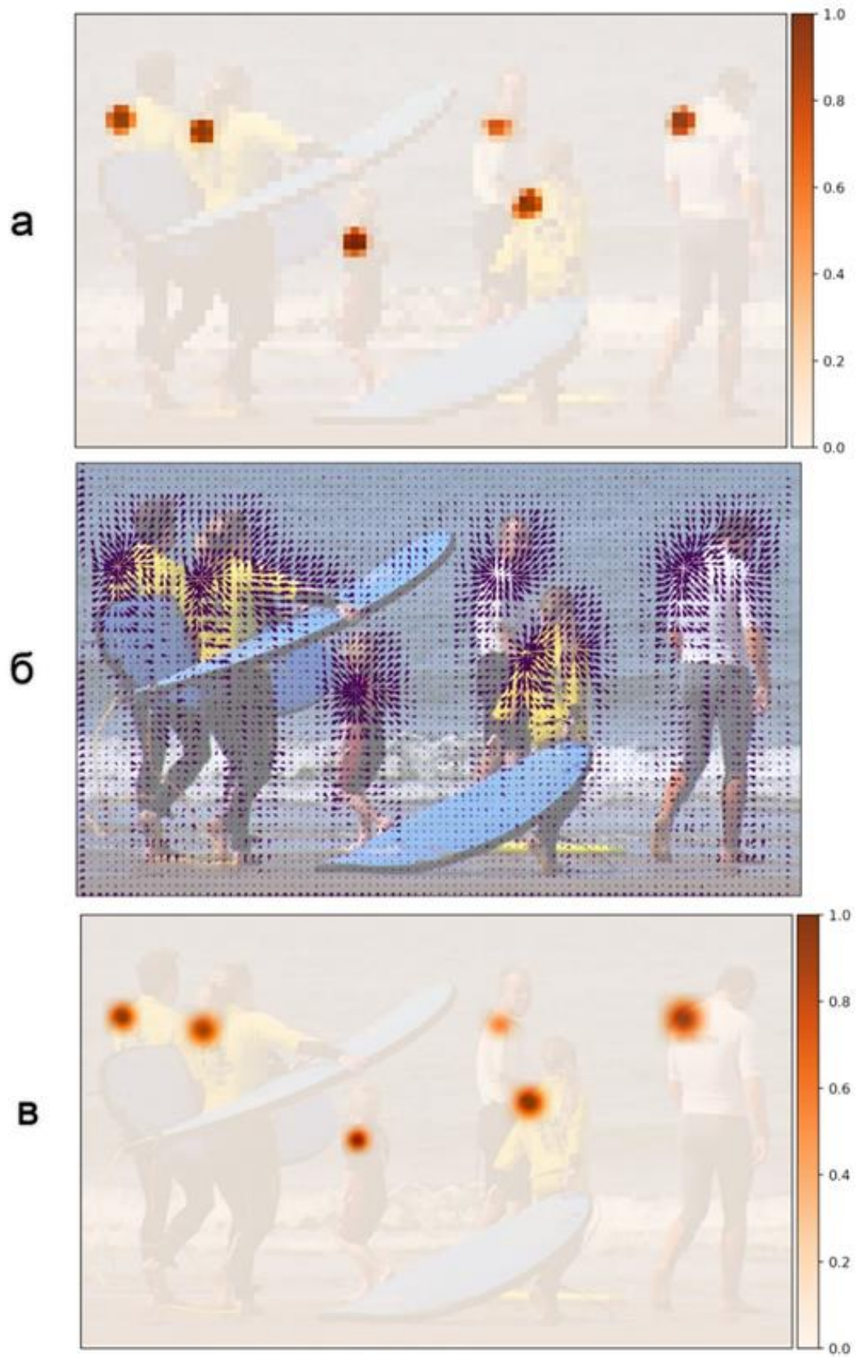
Демонстраційний плакат до магістерської дисертації

на тему «*СИСТЕМА РОЗПІЗНАВАННЯ ПОЗИ ЛЮДИНИ В РЕАЛЬНОМУ ЧАСІ*»

Виконав студент гр. ІС-81 ПІБ Лотоцька Ю.В.

Керівник ПІБ Халус О.А.

Візуалізація компонентів карт інтенсивності для лівого плеча



Демонстраційний плакат до магістерської дисертації

на тему «*СИСТЕМА РОЗПІЗНАВАННЯ ПОЗИ ЛЮДИНИ В РЕАЛЬНОМУ ЧАСІ*»

Виконав студент гр. ІС-81 ПІБ Лотоцька Ю.В.

Керівник ПІБ Халус О.А.

Візуалізація полів асоціації частин(ліве плече – ліве стегно)

а

б

Демонстраційний плакат до магістерської дисертації

на тему «*СИСТЕМА РОЗПІЗНАВАННЯ ПОЗИ ЛЮДИНИ В РЕАЛЬНОМУ ЧАСІ*»

Виконав студент гр. ІС-81 ПІБ Лотоцька Ю.В.

Керівник ПІБ Халус О.А.

Візуалізація розробленої системи та порівняння із OpenPose на наборі даних nuScenes

OpenPose

Розроблена система

Демонстраційний плакат до магістерської дисертації

на тему «*СИСТЕМА РОЗПІЗНАВАННЯ ПОЗИ ЛЮДИНИ В РЕАЛЬНОМУ ЧАСІ*»

Виконав студент гр. ІС-81 ПІБ Лотоцька Ю.В.

Керівник ПІБ Халус О.А.

Візуалізація розробленої системи та порівняння із Mask RCNN на наборі даних nuScenes

OpenPose

Розроблена система

Демонстраційний плакат до магістерської дисертації

на тему «*СИСТЕМА РОЗПІЗНАВАННЯ ПОЗИ ЛЮДИНИ В РЕАЛЬНОМУ ЧАСІ*»

Виконав студент гр. ІС-81 ПІБ Лотоцька Ю.В.

Керівник ПІБ Халус О.А.