КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА

Кваліфікаційна наукова
праця на правах рукопису

Кондратюк Сергій Сергійович

УДК 004.912

ДИСЕРТАЦІЯ
МОДЕЛЮВАННЯ ТА РОЗПІЗНАВАННЯ ЖЕСТІВ УКРАЇНСЬКОЇ ДАКТИЛЬНОЇ АБЕТКИ ЗА ДОПОМОГОЮ КРОСПЛАТФОРМЕНИХ ТЕХНОЛОГІЙ

Спеціальність 01.05.03 – математичне та програмне забезпечення
обчислювальних машин і систем

Подається на здобуття наукового ступеня *кандидата технічних* наук

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело.

С. С. Кондратюк

Науковий керівник: Крак Юрій Васильович,
член-кореспондент НАНУ,
доктор фізико-математичних наук, професор

Київ – 2021

АНОТАЦІЯ

Кондратюк С.С. Розпізнавання та моделювання жестів української дактильної абетки за допомогою кросплатформених технологій. – Рукопис.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 01.05.03 – математичне та програмне забезпечення обчислювальних машин і систем – факультет комп'ютерних наук та кібернетики, Київський національний університет імені Тараса Шевченка, Київ, 2021.

Дисертація присвячена розробці інформаційної технології для моделювання та розпізнавання дактилем (жестів) української дактильної абетки за допомогою кросплатформених засобів та нейронних згорткових мереж із тривимірною згорткою.

За даними Всесвітньої організації охорони здоров'я близько 5% населення мають вади слуху і близько 1-2% дітей народжуються з вадами слуху. Основною формою спілкування таких людей, як між собою так із іншими людьми, є жестова мова. В Україні людей вадами слуху – десятки тисяч, а знання жестової мови для комунікації з такими людьми потребують мільйони. Відзначимо, що в Україні є близько 60 шкіл-інтернатів для навчання глухих дітей, впроваджено інклюзивну освіту, тому надзвичайно актуальною є розробка нових інформаційних технологій, які були б доступні для широкого загалу населення і дозволяли швидко і ефективно вивчати жестову мову. З цією метою вперше запропонована технологія, розроблена за допомогою кросплатформених засобів, для моделювання жестів української дактильної абетки, анімації переходів між станами жестових одиниць та комбінування жестів (слів). Розроблена технологія відтворює послідовність жестів за допомогою віртуальної просторової моделі руки та дозволяє розпізнавання дактилем із вхідного потоку камери за допомогою навченої на зібраному наборі зображень згортковій нейронній мережі, із взятою за основу архітектурою MobileNetv2, та з підібраною оптимальною конфігурацією шарів та параметрів мережі. На зібраному тестувальному наборі даних досягнуто точності у понад 97%.

Запропонована технологія складається з модуля моделювання жестів та модуля розпізнавання жестів (які утворюють технологію навчання української дактильної абетки). Модуль моделювання жестів реалізує об'ємну реалістичну модель руки, а модуль розпізнавання жестів реалізує модель розпізнавання жестів української дактильної абетки за допомогою машинного навчання та згорткових нейронних мереж з архітектурою MobileNetv2, вдосконаленою тривимірними згортками та з оптимальною кількістю шарів. Усі модулі реалізовані у вигляді міжплатформених модулів та об'єднані у формі кросплатформеної технології. Запропонована технологія також забезпечує механізм налаштування для роботи на пристроях з різними технічними характеристиками (наприклад, зменшення кількості багатокутників та крок анімації). У рамках запропонованої технології для підготовки моделі розпізнавання жестів вперше було зібрано набір даних української дактильної абетки, який складається з понад 50 000 зображень, зібраних від 50 осіб у різних умовах навколишнього середовища, умовах освітленості та доповнених методами збільшення даних.

Удосконалено обрану архітектуру згорткової нейронної мережі MobileNetv2 тривимірними згортками, які дозволили навчити модель розпізнавання із якістю 0.97 f1-score на тестовому наборі даних української дактильної абетки, що перевищує, або є на одному рівні, із сучасними розробками в цій області для інших мов.

Результати моделювання і експериментальних досліджень, отримані в роботі, можуть бути використані для розробки перспективних технологій спілкування користувача з комп’ютером, а також в системах альтернативного спілкування мовою жестів.

Ключові слова: кросплатформеність, жестова мова, моделювання дактилем, розпізнавання дактилем, згорткові нейронні мережі, mobilenet

ABSTRACT

Kondratiuk S.S. Ukrainian dactyl alphabet gesture modeling and recognition using crossplatform technologies. - The manuscript.

Thesis for obtaining the scientific degree of the candidate of technical sciences on the specialty 01.05.03 – mathematical and programming software for computational machines and systems. – Faculty of Computer Science and Cybernetics, Taras Shevchenko National University of Kyiv, Kyiv, 2021.

For this purpose, for the first time, a technology developed using cross-platform tools for modeling gestures of the Ukrainian dactyl alphabet, animation of transitions between states of gesture units and combination of gestures (words) was proposed. The developed technology reproduces a sequence of gestures using a virtual spatial model of the hand and performs dactyl recognition from the input stream of the camera using a convoluted neural network trained on the assembled image set, based on MobileNetv2 architecture, and selected optimal layer configuration and parameters. Accuracy of more than 97% was achieved on the collected test data set.

As part of the proposed technology for the preparation of a gesture recognition model, a data set of the Ukrainian dactyl alphabet was collected for the first time, consisting of more than 50,000 images collected from 50 people in different environmental conditions, lighting conditions and supplemented by data augmentation methods. The chosen architecture of the MobileNetv2 convolutional neural network was improved by three-dimensional convolutions, which allowed to teach the recognition model with 0.97 f1-score quality on the test data set of the Ukrainian dactyl alphabet, which is on a par with modern developments in this language in other languages.

The results of simulation and experimental research, obtained in the work, can be used to develop advanced human-computer interfaces.

Keywords: cross platform, sing language, dactyl modeling, dactyl recognition, convolutional neural networks, mobilenet

Список опублікованих праць за темою дисертації

1. Кондратюк С.С., Крак Ю.В. Незалежне програмне забезпечення платформи для розвитку систем зв'язку: моделювання дактилічної мови // Штучний інтелект. - 2016. - т.73 в. 3, с. 36-47. (Особистий внесок здобувача: проаналізовано завдання, розглянуто підходи до впровадження незалежної від платформи технології моделювання мови жестів)

2. Kondratiuk S. Gesture recognition using convolutional neural networks with 3d convolutions // Штучний інтелект. – т.83-84 – 2019 – С.94-100.

3. Kondratiuk S. Gesture recognition using cross platform software and convolutional neural networks // Штучний інтелект. – т. 85-86 – 2019 – С.107-113.

4. Кондратюк С.С., Крак Ю.В. Платформонезалежне програмне забезпечення для розробки систем жестової комунікації // XVIII International Conference “Dynamical Systems Modeling and Stability Investigation (Modeling and Stability)”. Abstracts of conference reports. Kiev, Ukraine. May 24-26. 2017. Вісник Київського національного університету імені Тараса Шевченко. – 2017. – С. 180. (Особистий внесок здобувача: запропоновано кросплатформену технологію для створення системи жестової комунікації)

5. S. Kondratiuk, K. Kruchynin, Iu. V. Krak, S. Kruchinin. Information Technology for Security System Based on Cross Platform Software // NATO Science for Peace and Security Series A: Chemistry and Biology. Editors: Bonca, Janez, Kruchinin, Sergei (Eds.). 2018. Chapter 25. P.331-339. DOI: 10.1007/978-94-024-1304-5_25. Розділ монографії. (Особистий внесок: описано підходи до впровадження незалежних від платформи систем та проаналізовано існуючі технології) (Входить до наукометричної бази Scopus)

6. Serhii Kondratiuk , Iurii Krak , Waldemar Wójcik. Cross platform tools for modeling and recognition of the fingerspelling alphabet of gesture language // Informatica, Automatyka, Pomiare v Gospodarce I Ochronie Środowiska – 2019.

– № 9 (2). – pp. 24-27. DOI: 10.5604/01.3001.0013.2542. Міжнародний науковий журнал. (Особистий внесок: наведено методи та програмні засоби для моделювання та розпізнавання жестів)

7. S. Kondratiuk, Iu. Krak. Dactyl alphabet modeling and recognition using cross platform software // 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP). 21-25 Aug. 2018. Lviv, 2018. – pp. 420-423. doi: 10.1109/DSMP.2018.8478417. (Особистий внесок: описано інфологічну модель кросплатформеної технології) (Входить до наукометричної бази Scopus)

8. I. Krak and S. Kondratiuk. Cross-platform software for the development of sign communication system: Dactyl language modelling // 2017 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), Lviv., 2017, – pp. 167-170, doi: 10.1109/STC-CSIT.2017.8098760. (Особистий внесок здобувача: описаний підхід на основі кросплатформеної технології для вивченяя дактильної абетки) (Входить до наукометричної бази Scopus)

9. Serhii Kondratiuk , Iurii Krak, Olexander Barmak, Anatolii Pashko. Fingerspelling Alphabet 3D Modeling and Recognition Base on CNN Technology for Cross Platform Applications // Proceedings of the Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2019), Zaporizhzhia, Ukraine, April 15-19, 2019. CEUR Workshop Proceedings 2353, CEUR-WS.org 2019. – pp.173-182. (Особистий внесок здобувача: описано методи 3д моделювання дактилем та розпізнавання) (Входить до наукометричної бази Scopus)

10. Крак Ю.В., Кондратюк С.С., Тернов А.С. Інформаційно-комунікаційні технології для імітації мови жестів // InfoCom 2015: Матеріали 1-ї Міжнародної конференції, присвяченої 70-річчю кафедри автоматики та управління в технічних системах, Київ, 24-25 листопада 2015 р. - К.:

«Політехніка», 2015. - С.18-20. (Особистий внесок здобувача: розглянув підходи до моделювання мови жестів)

11. Крак Ю.В., Кондратюк С.С. До розвитку міжплатформених інформаційних технологій для моделювання та розпізнавання дактильної абетки // Міжнародна наукова конференція "Інтелектуальні системи прийняття рішень та обчислення систем (ISDMCI'2016)". 24-27 травня 2016. Матеріали міжнародної наукової конференції. - Херсон: ХНТУ, 2016. - С. 85-86. (Особистий внесок здобувача: пропонується тривимірна модель руки та спосіб встановлення жестів через файли конфігурації)

12. Крак Ю.В., Бармак О.В., Багрій Р.О., Кондратюк С.С. Комп’ютерні технології для спілкування людей з обмеженими можливостями // XXII Всеукраїнська наукова конференція «Сучасні проблеми прикладної математики та інформатики», APAMCS-2016. 5-7 жовтня 2016. - Збірник наукових праць. - Львів: ЛНУ імені Івана Франка. - 2016. - С. 101-103. (Особистий внесок здобувача: завершив аналіз завдання, розглянув підходи до впровадження незалежної від платформи технології моделювання мови жестів)

13. Крак Ю.В., Кондратюк С.С. Платформонезалежні технології моделювання та розпізнавання дактильної жестової мови // XIV Міжнародна науково-практична конференція «Математичне та програмне забезпечення інтелектуальних систем (MPZIS-2016)». Тези доповідей. (16-18 листопада 2016 р.). м. Д.: ДНУ. - 2016. - С. 112-113. (Особистий внесок здобувача: описує підхід до розпізнавання жестів за допомогою нейронних мереж міжплатформеної конвергенції)

14. Кондратюк С.С., Крак Ю.В. До розробки систем жестової комунікації на основі крос-платформених технологій для моделювання дактильної мови // Системи та засоби штучного інтелекту: тези Міжнародної наукової молодіжної школи, 29-30 листопада 2016 р. - Київ: ІПАІ «Наука та освіта». -

2016. - С. 48–52. (Особистий внесок здобувача: проаналізував та розробив крос-платформену технологію для моделювання мови жестів)

15. Кондратюк С.С., Крак Ю.В. Інформаційна технологія моделювання та динамічного відображення конструкцій жестів // Міжнародна наукова конференція "Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту (ISDMCI'2017)". 22-26 травня 2017. Матеріали Міжнародної наукової конференції. - Залізний порт Україна. Херсон: Видавництво ПП Вишемирського В.С., 2017. - С. 70–71. (Особистий внесок здобувача: описаний крос-платформений підхід до впровадження технології моделювання української дактильної абетки)

16. Єфермов М.С., Крак Ю.В., Кондратюк С.С. Моделювання та кросплатформена реалізація жестів дактильної мови // XVII Міжнародна науково-практична конференція «Математичне та програмне забезпечення інтелектуальних систем (MPZIS-2019)». Тези доповідей. (20-22 листопада 2019 р.). м. Д.: ДНУ. - 2019. - С. 99-100. (Особистий внесок здобувача: проаналізована та розроблена міжплатформена технологія для імітації мови жестів за допомогою 3D-моделі руки)

17. Кондратюк С.С., Крак Ю.В., Голік А.О. Моделювання та розпізнавання дактильної інформації про крос-платформені технології // Системи та засоби штучного інтелекту: тези доповідей Міжнародної наукової молодіжної школи. 18 жовтня 2017. - Київ: ІПАІ «Наука та освіта». - 2017. - С.91-93. (Особистий внесок здобувача: міжплатформений підхід до технології моделювання української дактильної абетки на бібліотеці Unity3d)

Зміст

АНОТАЦІЯ **2**
Список опублікованих праць за темою дисертації **5**
Зміст **9**
Перелік умовних скорочень **11**
Список рисунків **12**
Список таблиць **14**
Вступ **15**
Розділ 1. Огляд підходів до моделювання та розпізнавання жестової мови **22**
1.1. Огляд досліджень з побудови кросплатформеної інформаційної технології для моделювання та розпізнавання жестової мови 22
1.1.1. Модель скелету руки для демонстрації жестової інформації 22
1.1.2. Підходи, які базуються на 27 ступенях свободи моделі руки людини 28
1.2. Підходи до моделювання та аналізу рухів руки людини 28
1.2.1. Моделювання руки 32
1.3. Підходи до розпізнавання дактилем 34
1.3.1. Концепції обробки зображень 39
1.3.2. Підходи машинного навчання для класифікації зображень 41
1.4. Використання нейромережевих технологій для задач розпізнавання 42
1.5. Кросплатформені засоби розробки технології моделювання та розпізнавання жестів. 44
1.6. Висновки до Розділу 1 45
Розділ 2. Моделі та методи опису тривимірного скелету та моделі руки, моделювання та розпізнавання жестів за допомогою глибокого навчання **47**
2.1. Моделювання скелету руки людини 47
2.1.1. Модель скелету руки для моделювання жестової інформації 47
2.1.2. Обмеження суглобів скелету 48
2.2. Моделювання руки 50
2.2.1. Моделювання жестів за допомогою просторової моделі руки 50
2.2.2. Параметрична модель руки 52
2.2.3. Модель з текстурою 54
2.3. Адаптивність моделі руки 56
2.4. Математична формалізація процесу подання даних для розпізнавання дактилем 57
2.4.1. Просторові дескриптори 59
2.4.2. Попередня обробка даних 60
2.4.3. Просторово-часове подання даних 61
2.5. Модель розпізнавання на основі глибокої згорткової нейронної мережі 64
2.5.1. Математична формалізація структури нейронної мережі 65
2.5.2. Використання даних у просторово-часовому вимірі 67
2.5.3. Удосконалення моделі нейронної мережі тривимірними згортками для використання даних у просторово-часовому вимірі 67
2.5.4. Тренування нейронної мережі 70
2.5.5. Проблема перенавчання та трансферу навчання 72
2.6. Висновки до Розділу 2 74
Розділ 3. Кросплатформена інформаційна технологія моделювання та розпізнавання жестів за допомогою тривимірних згорток **77**
3.1. Інфологічна модель 77
3.2. Моделювання 81
3.2.1. Аналіз та завантаження жестових конфігурацій 81
3.2.2. Побудова тривимірної сцени з моделлю руки заданої полігональності 82
3.2.3. Анімація жестів 83

3.3. Розпізнавання (тренування) ... 83
3.3.1. Збір набору даних ... 83
3.3.2. Обробка даних ... 83
3.3.3. Розширення (аугментація) даних та розподіл ... 85
3.3.4. Задання набору конфігурацій архітектур ... 87
3.4. Розпізнавання (прогнозування) ... 89
3.4.1. Виділення вхідних даних ... 89
3.4.2. Обробка даних ... 89
3.4.3. Розпізнавання та отримання результатів ... 90
3.5. Кросплатформена розробка ... 91
3.5.1. Моделювання руки ... 91
3.5.2. Розпізнавання жестів ... 92
3.6. Вдосконалена для розпізнавання дактилем архітектура MobileNetv2 ... 93
3.7. Висновки до Розділу 3 ... 94
Розділ 4. Експериментальна програмна реалізація та результати тренувань і тестувань моделей розпізнавання дактилем ... 96
4.1. Користувацький інтерфейс ... 96
4.2. Структура модулів і класів програмної реалізації ... 97
4.2.1. Структура модулів програмної реалізації ... 97
4.2.2. Структура класів програмної реалізації ... 98
4.2.3. Структура бази даних ... 100
4.3. Набір даних ... 101
4.4. Експериментальні тренування та досліди навченої моделі ... 104
4.5. Вибір архітектури моделі ... 105
4.6. Висновки до Розділу 4 ... 110
Висновки по роботі ... 111
Література ... 113
Додатки ... 124

Перелік умовних скорочень

Скорочення, термін, позначення	Пояснення
SVM	(англ. support vector machines) метод опорних векторів
SVD	(англ. Singular value decomposition) сингулярний розклад
DTW	(англ. Dynamic time warping) метод динамічної згортки часової шкали
Crossplatform	Крос платформений, наявний на багатьох платформах
Data mining	(англ.) Інтелектуальна обробка великих масивів даних
Libsvm	(англ. SVM library) відкрита реалізація методу SVM
3d	(англ.) тривимірний
IT	Інформаційні технології
Dlib	Бібліотека алгоритмів обробки зображень
PCA	(англ. Principal component analysis) метод головних компонент
RBF	(англ. Radial basis function) радіальні базисні функції
ANN	(англ. Artificaial neaural network) штучна нейронна мережа
KLT	(англ. Karhunen-Loeve transform) метод Карунена-Лоева
DCT	(англ. Discrete cosine transform) дискретне косинусне перетворення
SDAE	(англ. Stacked denoising autoencoder) знешумлюючий автоенкодер
DBN	(англ. Deep belief network) глибока мережа переконань
CNN	(англ. Convolutional neural network) згорткова (конволюційна) нейронна мережа

Список рисунків

Рис. 1.1. Інтерфейс користувача вебсайту онлайн-бази даних ASL 23
Рис. 1.2. Скелет руки 26
Рис. 1.3. Обмеження та залежність скелета руки людини 27
Рис. 1.4. Кінематична структура та позначення суглобів 30
Рис. 1.5. Діаграма підходів жестового розпізнавання 35
Рис. 2.1. Структура скелету руки 48
Рис. 2.2. Приклад обмеження суглоба 49
Рис. 2.3. (верх) Вказівний палець (середина) Дистальна (ДФ) та проміжна (ПФ) фаланги подані як усічені еліптичні конуси, апроксимуюча фаланга (АФ) — як еліпсоїд. (низ) Меш пальця 53
Рис. 2.4. Шар із нормалями для поверхні руки 55
Рис. 2.5. Шар із текстурою та прозорістю моделі руки 55
Рис. 2.6. Шар із нерівностями поверхні моделі руки 56
Рис. 2.7. Особливості SIFT визначаються з зображень у відтінках сірого та використовуються в якості векторів для створення словника BoW. 60
Рис. 2.8. Послідовність кадрів, розбита на дві підпослідовності розміром 5 кадрів, які перетинаються на 3 кадри. 63
Рис. 2.9. Дві підпослідовності, створені з одного відеопотоку 64
Рис. 2.10. Блок MobileNetv2 65
Рис. 2.11. Загальна архітектура мережі 66
Рис. 2.12. Загальна схема процесу розпізнавання разом із моделлю розпізнавання тривимірними згортками 68
Рис. 2.13. Архітектура MobileNetV2, модифікована, з тривимірними згортками 69
Рис. 2.14. Типова кореляція між місткістю моделі та показниками помилок. Зліва від оптимальної ємності ми могли б збільшити потенціал, щоб знайти краще узагальнення навчального набору. Цей стан називається недостатнім. 73
Рис. 2.15. Шар випадання, який обнулює задану частку нейронів під час кожної ітерації 74
Рис. 3.1. Узагальнена структурна схема ІТ 79
Рис. 3.2. Діаграма компонентів ІТ 81
Рис. 3.3. Аналіз та завантаження жестових конфігурацій 82
Рис. 3.4. Побудова тривимірної сцени з моделлю руки заданої полігональності 82
Рис. 3.5. Анімація жестів 83
Рис. 3.6. Аугментація жестів 86
Рис. 3.7. Приклади підходів до аугментації даних 87
Рис. 3.8. Задання набору конфігурацій нейромереж 88
Рис. 3.9. Виділення вхідних даних 89
Рис. 3.10. Розпізнавання та отримання результатів 90
Рис. 3.11. Архітектура MobileNet 94
Рис. 4.1. Інтерфейс користувача моделювання жестів запропонованої технології 96
Рис. 4.2. Інтерфейс користувача модуля розпізнавання жестів 97
Рис. 4.3. Схема модулів програмної реалізації 97
Рис. 4.4. Структура модулів, бібліотек і класів програмної реалізації модуля моделювання 99
Рис. 4.5. Структура модулів, бібліотек і класів програмної реалізації модуля розпізнавання 100
Рис. 4.6. Приклад таблиці, яка зберігає конфігурацію дактилем для модуля моделювання 100
Рис. 4.7. Інтерфейс програмного забезпечення для запису набору даних дактилів 102
Рис. 4.8. Зразок зібраного набору даних 103
Рис. 4.9. Розподіл зібраного набору даних за якістю освітлення 103
Рис. 4.10. Розподіл зібраного набору даних за якістю зображення 104

Рис. 4.11. Розподіл зібраного набору даних за статтю 104
Рис. 4.12. Результати матриці помилок архітектури 1 на тестовому наборі даних 106
Рис. 4.13. Результати матриці помилок архітектури 2 на тестовому наборі даних 106
Рис. 4.14. Результати матриці помилок архітектури 3 на тестовому наборі даних 107
Рис. 4.15. Результати матриці помилок архітектури 4 на тестовому наборі даних 107
Рис. 4.16. Результати матриці помилок архітектури 5 на тестовому наборі даних 108
Рис. 4.17. Діаграми, що показують прогрес навчання моделі MobileNet 108
Рис. 4.18. Діаграма, що показує прогрес навчання моделі з певною архітектурою залежно від кількості ітерацій 109
Рис. 4.19. Діаграма, що показує якість навченої моделі залежно від складності архітектури та наявності в ній тривимірної згортки 109

Список таблиць

Табл. 1.1. Порівняння 3D-рушіїв за функціями та підтримуваними платформами..... 33
Табл. 4.1. Показники навчуваності різних архітектур 105
Табл. 4.2. Середній макробал f1 для порівняння підготовлених архітектур 108
Табл. 4.3. Порівняння навченої моделі з сучасними підходами 109

Вступ

Дисертація присвячена розробці інформаційної технології моделювання дактилем української жестової мови з використанням тривимірної високополігональної моделі руки та розпізнавання дактилем за допомогою глибоких нейронних мереж, обидві частини технології розроблені за допомогою кросплатформених засобів.

Актуальність теми. Жестова мова є одним із основних засобів передачі інформації та спілкування, разом із текстом і мовленням. Жестові мови складаються з окремих знаків, які поєднуються в букви, слова, фрази за допомогою послідовного переходу від одного знаку до іншого. За даними Всеукраїнської громадської організації інвалідів та Українського товариства глухих, в Україні проживають десятки тисяч людей із вадами слуху. Враховуючи статистику ВООЗ, близько 5% населення мають вади слуху, а отже, потенційна громада користувачів жестової мови в Україні становить мільйони.

Це робить українську жестову мову та українську дактильну абетку важливим засобом спілкування в Україні поряд з комунікацією в текстовій та розмовній формах. Вагомий внесок у дослідження проблем жестової інформації зробили закордонні дослідники: Л.С.Димскис, О.Л.Воскресенський, Г.Л.Зайцева, W.C. Stokoe, W. Sandler, R.-H. Liang, M. Ouhyoung, H. Wang, C. Leu, S. Morrissey та ін., та українські науковці, зокрема: В.В. Пасічник, Ю.В. Нікольський, М.В. Давидов, Ю.Г. Кривонос, Ю.В. Крак, О.В. Бармак, С.В. Кульбіда та інші.

Дослідженнями та розробками сучасних систем жестової комунікації зараз займаються в багатьох наукових та навчальних організаціях України і світу, в той же час чимало питань комп’ютерної візуалізації та розпізнавання жестової мови залишаються відкритими.

Отже, розробка нових підходів, методів та їх реалізація на основі сучасних комп’ютерних засобів для моделювання і розпізнавання елементів жестової комунікації є актуальними для соціального та науково-технічного прогресу.

Зв'язок роботи з науковими програмами, планами, темами. Основні дослідження за темою дисертації було виконано на кафедрі теоретичної кібернетики факультету комп'ютерних наук та кібернетики Київського національного університету імені Тараса Шевченка в рамках фундаментальної теми № 16БФ015-04 «Розробка логіко-алгоритмічних методів дослідження формальних моделей природних мов» (номер державної реєстрації 0116U0064780, 2016-2018 рр.), де автор брав участь як виконавець та розробник деяких розділів.

Мета та завдання дослідження. Мета дисертаційної роботи полягає у розробці інформаційної технології для моделювання та розпізнавання жестів української дактильної абетки за допомогою кросплатформених технологій. Дисертаційне дослідження передбачатиме вирішення окремих задач, а саме:

1. Побудова тривимірної моделі руки для моделювання жестів, із високою реалістичністю та адаптивністю до обчислювальних потужностей платформи, на якій відбувається моделювання, що може бути досягнуто адаптацією кількості полігонів та кроком анімації жестових переходів. Модуль із моделювання жестів має бути кросплатформеним та бути єдиною частиною запропонованої технології.

2. Аналіз особливостей зображень різних жестових одиниць (дактилем) із метою побудови модуля для розпізнавання жестів із вхідного потоку зображень з веб-камери без додаткового обладнання (сенсорів, рукавичок і т.ін.).

3. Аналіз алгоритмів розпізнавання жестів, вибір оптимального за сукупністю факторів, як-от: кросплатформеність, швидкодія, адаптивність, висока якість розпізнавання, здатність до вдосконалення. Побудова моделі розпізнавання жестів за допомогою глибоких згорткових нейронних мереж та кросплатформених технологій. Вдосконалення архітектури мережі за допомогою підбору оптимальної конфігурації шарів та параметрів і використанням додаткових механізмів, як-от тривимірна згортка.

4. Розробка засобів та створення достатньо великого набору даних зображень жестів української дактильної абетки з різноманітними умовами

середовища, освітлення, фону, різних людей, що зображують жести, за такими ознаками як вік, стать, розмір руки. Набір даних має бути достатньо великим для якісного навчання глибокої нейронної мережі та для репрезентативної тестової вибірки.

5. Проведення експериментальних досліджень та випробувань побудованої моделі на тестовому наборі даних, на різних платформах і в різних реальних тестових середовищах.

Об'єктом дослідження є процес розпізнавання та моделювання дактилем української дактильної мови.

Предметом дослідження є методи аналізу зображень дактилем та процесу розпізнавання та моделювання дактилем.

Методи дослідження. Для розв'язання задачі дисертаційного дослідження використовувалися алгоритми та методи тривимірного комп'ютерного моделювання, комп'ютерного зору, машинного навчання та глибокого навчання, зокрема методи просторового моделювання руки, засновані на скелеті руки та обмеженнях ступенів свободи, методи анімації жестів дактильної абетки за допомогою покрокового переходу з одного стану скелету в інший, враховуючи наявні обмеження свободи рухів, методи виділення ключових особливостей зображення, методи побудови дескрипторів зображення, методи глибоких згорткових нейронних мереж, методи навчання і тестування моделей розпізнавання, методи пошуку оптимальної архітектури нейронних мереж, методи тривимірних згорток, методи аналізу даних та методи нормалізації зображень.

<u>Наукова новизна отриманих результатів</u>

1. Вперше розроблено кросплатформену інформаційну технологію для української жестової мови. Розроблена технологія складається з кросплатформених модуля моделювання жестів та модуля розпізнавання жестів (які утворюють технологію навчання української дактильної абетки). Реалізовано реалістичну тривимірну модель руки.

2. Вперше модуль розпізнавання жестів української дактильної абетки реалізовано за допомогою машинного навчання та згорткових нейронних мереж з архітектурою MobileNetv2 із тривимірними згортками. Усі модулі об'єднані у формі кросплатформеної технології. Розроблена технологія також реалізує механізм налаштування роботи на пристроях з різними технічними характеристиками (наприклад, зменшення кількості полігонів та кроку анімації).

3. Вперше у рамках технології для розробки модуля розпізнавання жестів було створено статистично значимий масив з 50000 зображень жестів рук 50 осіб у різних умовах навколишнього середовища, умовах освітленості. Масив, доповнений методами аугментації даних, складається з 150000 зображень.

4. Покращено архітектуру MobileNetv2, яку взято за основу для модуля розпізнавання жестів. Архітектура підібрана відповідно до показників натренованої моделі на тестовому масиві даних, із оптимальним співвідношенням складності та якості розпізнавання. Архітектура доповнена механізмом тривимірної згортки, що покращує розпізнавання за допомогою темпоральної інформації з послідовності кадрів. Модуль розпізнавання демонструє якість розпізнавання на рівні або вище сучасних робіт з розпізнавання жестів інших мов за допомогою глибоких нейронних мереж

Практичне значення одержаних результатів полягає у розробці тривимірної моделі руки на основі скелету з обмеженнями ступенів свободи та апробації (реалізації) її у застосуванні до задач аналізу і моделювання дактилем української абетки. Також було розроблено модель із розпізнавання жестів української дактильної абетки за допомогою глибокої нейронної згорткової мережі із тривимірними згортками. Обидві моделі є частинами єдиної інформаційної кросплатформеної технології. Зокрема, було:

- розроблено експериментальну інформаційну технологію розпізнавання дактилем української дактильної абетки та її програмну реалізацію;
- розроблено експериментальну технологію отримання просторової моделі руки людини, що застосовує елементи запропонованого скелету із обмеженнями ступенів свободи.

Моделі, методи і алгоритми моделювання і розпізнавання комунікаційної інформації, запропоновані в даній роботі, використовуються у навчальному процесі, зокрема, при підготовці курсових, бакалаврських і магістерських кваліфікаційних робіт у Київському національному університеті імені Тараса Шевченка на кафедрі теоретичної кібернетики. Результати окремих елементів даного дослідження були використані в наукових дослідженнях, в наукових темах і звітах.

Окрім навчального і педагогічного процесу, результати даного дослідження були впроваджені на виробництві, а саме у ТОВ «Науково-технічна фірма «Інфосервіс»» (м. Хмельницький), зокрема, при розробці програмного забезпечення для: відтворення жестів за допомогою просторової моделі руки людини; розпізнавання жестів української дактильної абетки з відеопотоку з використанням моделей глибоких нейромереж; аналізу просторових рухів руки людини, що підтверджено відповідним актом впровадження.

Значення результатів дослідження полягає у застосовності їх в різних галузях. Найбільш важливим є використання, з залученням технологій розпізнавання і моделювання дактилем, при навчанні жестової мови людей з вадами слуху, для розробки інформаційних технологій для інклюзивної освіти для шкільних і дошкільних закладів із вивченням жестової мови, для самонавчання батьків дітей із вадами слуху або соціальних працівників (медицина, національна поліція та інші служби), працівників сфери обслуговування, що контактують з людьми із вадами слуху.

Окрім соціального спрямування є і технічне спрямування — для формалізації (опису), розпізнавання і моделювання рухів просторових маніпуляційних і робототехнічних систем.

Особистий внесок здобувача

Всі наукові і практичні результати дисертаційної роботи одержано автором самостійно й опубліковано, зокрема, в одноосібно підготовлених працях, а саме: в роботах [2, 3] запропоновано використання глибоких згорткових нейронних мереж для розпізнавання жестів, зокрема взято за основу архітектуру

MobileNetv2 як ту, що виділяється співвідношенням розміру, швидкодії (що є актуальним для кросплатформеної технології) та якості розпізнавання та добре себе зарекомендувала на мобільних пристроях, а також запропоновано використання механізму тривимірних згорток для експлуатації інформації про жести з декількох послідовних кадрів. У друкованих працях, опублікованих у співавторстві, автору належать основні ідеї, теоретична та практична розробка положень, відображених у характеристиці наукової новизни отриманих результатів, а саме: в роботах [5, 6] описано підходи до впровадження незалежних від платформи систем та проведено аналіз існуючих технологій, завершено аналіз завдання, розглянуто підходи до впровадження незалежної від платформи технології моделювання мови жестів, проведено аналіз підходів до розпізнавання зображень, запропоновано та впроваджено кросплатформену технологію розпізнавання жестів української дактильної абетки, запропоновано тривимірну модель руки та спосіб встановлення жестів через файли конфігурації.

Апробація результатів дисертації. Основні результати дисертації доповідались на таких міжнародних наукових і науково-технічних конференціях, семінарах і наукових школах: «НАТО «Наука заради миру та безпеки»» (2017); «Міжнародна конференція IEEE 2018 року щодо видобутку та обробки потоків даних, DSMP» (2018); «Інтелектуальні системи прийняття рішень та обчислення систем (ISDMCI)» (2016, 2017); «Сучасні проблеми прикладної математики та інформатики», APAMCS» (2016); «XIV Міжнародна науково-практична конференція «Математичні та програмні системи для інтелектуальних систем (MPZIS» (2016); «Системи та засоби штучного інтелекту: тези Міжнародної наукової молодіжної школи» (2016); «Штучний інтелект. Інтелектуальні системи» (2018, 2019); «XVIII Міжнародна конференція «Моделювання динамічних систем та винахід для стабілізації (моделювання та стабільність)» (2017); «Комп'ютерні науки та інформаційні технології (CSIT): Міжнародна науково-технічна конференція)» (2017); «Матеріали II міжнародної наукової конференції «Інформатика та прикладна математика»» (2017);

«Системи та засоби штучного інтелекту: Міжнародна наукова молодіжна школа» (2017).

Публікації. Основні наукові положення, висновки та результати дослідження опубліковано у 17 роботах, що включають 11 тез на конференціях, 6 статей, що включають 4 наукових статей, виданих у фахових виданнях України, і 2 видані англійською мовою в іноземних фахових виданнях, включених до міжнародних наукометричних баз даних.

Структура та обсяг дисертаційної роботи. Робота складається зі змісту, вступу, чотирьох розділів основної частини, висновків, списку використаних джерел і додатків. Обсяг роботи: 91 сторінок основного тексту; 50 рисунків на 18 сторінках та 4 таблиці на 4 сторінках; список використаних джерел зі 116 найменувань на 11 сторінках. Загальний обсяг роботи складає 124 сторінки.

Розділ 1. Огляд підходів до моделювання та розпізнавання жестової мови

У першому розділі дисертаційного дослідження проведено огляд моделей, методів та алгоритмів, які застосовуються для задач моделювання і розпізнавання жестів, а також кросплатформених засобів для побудови інформаційної технології для розпізнавання та моделювання жестів. Також розглянуто алгоритми нормалізації, обробки та класифікації зображень.
Основні результати розділу опубліковані автором у працях [10, 11, 12].

1.1. Огляд досліджень з побудови кросплатформеної інформаційної технології для моделювання та розпізнавання жестової мови

Міміка відіграє важливу роль в різних галузях досліджень: в нейрофізіології, лінгвістиці, психології та інших науках, тому багато дослідників намагалися зрозуміти природу і причини її утворення.

1.1.1. Модель скелету руки для демонстрації жестової інформації

Моделювання жестів — це проблема, що розглядається не тільки окремо, але й як частина проблеми моделювання та розпізнавання жестів, і як технологія вивчення та оцінювання мови жестів. Демонстрація жесту чи знаку певного дактиля або мови жестів пов'язана з великою кількістю проблем.

Однією із систем відображення мови жестів є American Sign Language (ASL) Online Dictionary (Інтернет-словник американської мови жестів) [13], який складається з відеобази даних слів і фраз, що відображаються через мову жестів (рис. 1.1). Хоча словник існує вже багато років, проблеми розширення його новими словами все ще ті самі, що і в будь-якої системи навчання на базі даних відео. Більше того, показ жестів з усіх ракурсів є проблематичним, тому більшість відео обмежуються одним чи двома ракурсами. Розширення цієї бази даних мови жестів іншою мовою також вимагає такої ж кількості зусиль.

Рис. 1.1. Інтерфейс користувача вебсайту онлайн-бази даних ASL

Керування за допомогою жестів є актуальною проблемою в розвитку платформонезалежної взаємодії людина-комп'ютер [14]. Ці розробки були залучені низкою комерційних агентств [15, 16], але запропоновані ними системи налаштовані на заздалегідь визначену кількість жестів, а тому не вирішують проблеми моделювання жестової мови.

У роботі [17] аналізуються існуючі підходи до моделювання рук, які поділяються на дві основні групи: просторова та часова. Запропонована система здатна імітувати знакові анімації для заданого тексту. В рамках цієї системи використовується статистична модель для аналізу вхідного тексту, а генеративний алгоритм використовується при створенні відповідної імітаційної кінематики знакових анімацій. У статті [18] запропоновано інструменти ANVIL для введення тексту анотації, генератор жестів NOVA та бібліотеку DANCE, розроблену в роботі [19], яка використовується для анімації жестів. Система побудована на платформі Microsoft Windows та процесорі x86. У дослідженні [20] обговорюється моделювання віртуального персонажа для просторового відтворення жестової мови на платформі Microsoft Windows. Розроблена система вивчення жестової мови, яка складається з двох модулів — модуля демонстрації жестів через відео та модуля розпізнавання жестів (необхідні рукавички), заснованого на Hidden Markov Model (прихована марковська модель).

Метод, запропонований у роботі [21], представляє автоматичну техніку генерації руху підшкірних сухожилків і м'язів шляхом інтеграції традиційного анімаційного конвеєра з новим біомеханічним симулятором. Система обчислює активізацію м’язів для того, щоб здійснювати рух руки. Однак у ній бракує деталей, таких як шкірні згини суглобів та долоні. Метод, що розроблено в роботі [22], з іншого боку, зосереджений на використанні зовнішнього вигляду руки для створення її 3D-моделі. Були включені такі деталі, як згини та особливості шкіри. Це важливо для створення візуально реалістичної моделі руки, але вимагає значної попередньої обробки з використанням методів машинного розпізнавання об’єктів. Метод, наведений в дослідженні [23], подає неявну шкірну обробку, яка використовує геометричний метод для деформації шкіри в режимі реального часу з контактним моделюванням, надавши 3D-модель руки. Інший метод представив індивідуальну модель CAD людини, яка автоматично створює сітку з антропометричних та сканованих даних [24]. Метод, запропонований в роботі [25], дозволяє оцінити тривимірну геометрію та зовнішній вигляд людського тіла за допомогою монокулярного відео RGB-D. Автори методу створили параметричну 3D-модель Delta, якою можна керувати, виходячи з особливостей, отриманих з відео. Тривимірна модель руки з використанням 3D-стереофотограмметрії була розроблена у роботі [26] для створення та відтворювання точних зображень руки, які можна використовувати для аналізу м’яких тканин.

Тим часом метод, розроблений з використанням давача Leap Motion (LM), використовує попередньо завантажене реалістичне загальне зображення руки для виконання завдань і не враховує різні розміри та зовнішній вигляд рук [27]. Цей давач руху LM відслідковує рух руки до 200 кадрів і може визначати розміри та положення кісток руки, значення діапазону та інші геометричні характеристики руки. Програмне забезпечення LM передає цю інформацію до встановленої схеми загального зображення руки та анімує 3D-модель руки безпосередньо шляхом запису руху. Однак давач має лише поле зору (FOV)

шириною 150 градусів і глибиною 120 градусів, що обмежує рух користувача [28].

Відзначимо, що рука людини є унікальною та має тридцять глобальних геометричних особливостей, визначених на основі будови пальців і долоні. До особливостей можна віднести довжину і ширину пальців, товщину долоні, максимальне розчепірення пальців та інші [29]. Руки людини мають достатньо багато анатомічних особливостей, але не вважаються унікальними для повної ідентифікації особи. Однак поєднання з поверхневою анатомією (наприклад, долонею та пальцями) та іншими особливостями робить її унікальною та практичною для біометрики [30].

Отже, рука людини — це складна анатомічна структура, що складається з кісток, м’язів, сухожилків та шкіри, кожен з цих елементів має складний взаємозв'язок з іншим, що суттєво впливає на кінематику та зовнішній вигляд руки. Кістки руки згруповані в три області: кістки пальців кисті, метакарпальна частина (п’ясть) та зап’ястя. Кістки чотирьох пальців складаються з дистальних (DP), середніх (MP) та проксимальних фаланг (PP), тоді як великий палець — лише з дистальної та проксимальної фаланга, а область долоні складається з метакарпалів (MC) [31]. Зап'ясні кістки — це набір з восьми маленьких кісток, а передпліччя — з променевої і ліктьової кістки. Для згину рук передбачені три долонні ділянки шкіри та три поперечні згини. Ці згини пов'язані з основними кістками, які служать орієнтиром при виявленні конкретної будови руки під час операції.

Модель руки, створеної за допомогою цієї процедури, має 4439 вершин. Загальна форма руки програмно відтворюється за допомогою параметричної 3D-моделі руки. Текстура та деформація шкіри не розглядаються, проте даний метод допомагає зменшити зусилля у створенні основної форми індивідуалізованої 3D-руки за допомогою використання математичних зображень у вигляді параметризованих геометричних фігур. Рука може бути побудована за допомогою коригування декількох параметрів:

1. розміри руки (можна отримати від давача Leap Motion);
2. кількість вершин для кожної кістки пальця і долоні;
3. кути розчепірення пальця (можна обчислити за даними Leap Motion).

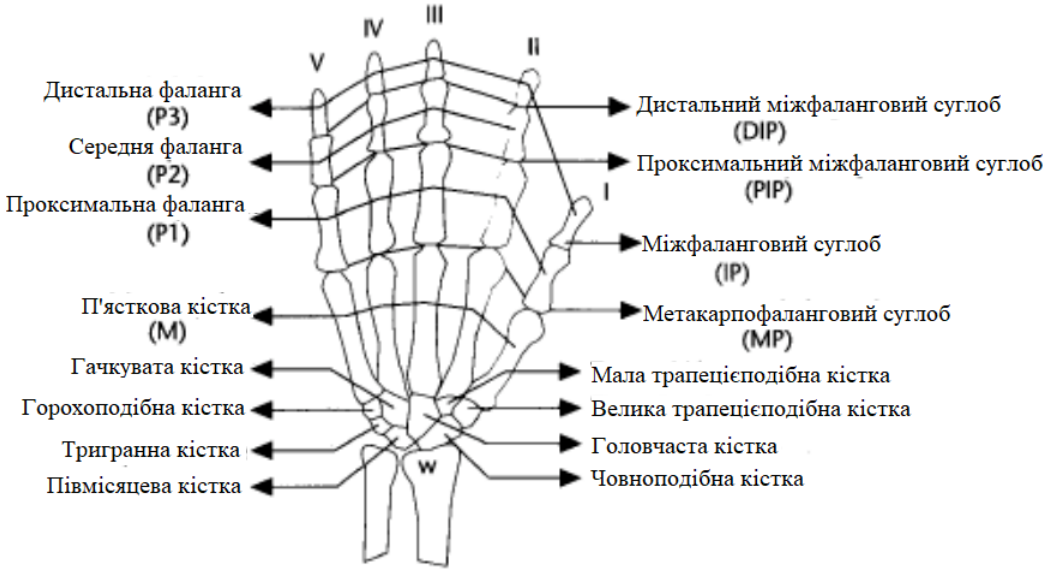


Рис. 1.2. Скелет руки

Скелет руки людини складається з 27 кісток (рис. 1.2). Їх можна розділити на три групи: карпали (кістки зап'ястя — вісім кісток), метакарпали (кістки долоні — п’ять кісток), фаланги (кістки пальця — чотирнадцять кісток).

Для того, щоб створити реалістичний аналіз моделі руки людини, обмеження людської руки є важливими. Більшість суглобів, які з’єднуються з карпалами, мають обмежений ступінь свободи (DOF). Але інші відрізняються, наприклад, великий палець; зазвичай існує два ступені свободи для метакапофалангового (MCP) суглоба та одна ступінь свободи для проксимальних міжфалангових (PIP) суглобів та дистальних міжфалангових (DIP) суглобів. Для великого пальця передбачено дві DOF, для трапеціометакарпального (TM) та метакапофалангеального (МКП) суглоба, а також одна ступінь свободи для міжфалангового (ІП) суглоба. Як результат, є 21 ступінь свободи на 5 пальців і плюс 6 ступенів свободи для озташування та орієнтування руки. Отже, маємо 27 DOF для руки людини.

Інше важливе обмеження — це обмеження руху в суглобах, що є поняттям пов’язаності рухів у сусідніх суглобах. Існує 5 обмежень для людської руки (див. рис. 1.3). По-перше, тому що MCP з чотирьох пальців (крім великого пальця) може виконувати згинання/розгинання та аддукцію/абдукцію (приведення/відведення), а суглоби PIP та DIP цих чотирьох пальців можуть лише згинатися/розгинатися в одному напрямку. В цьому випадку всі кістки цього пальця будуть знаходитися в одній площині. По-друге, кути з'єднання DIP і PIP (відповідно $Qdip$ і $Qpip$) залежні: $Qdip = \frac{2}{3} Qpip$. По-третє, межі кута f з’єднання суглобів MCP залежать від суміжних пальців. По-четверте, суглоб MCP в середині демонструє обмеження аддукції та абдукції.

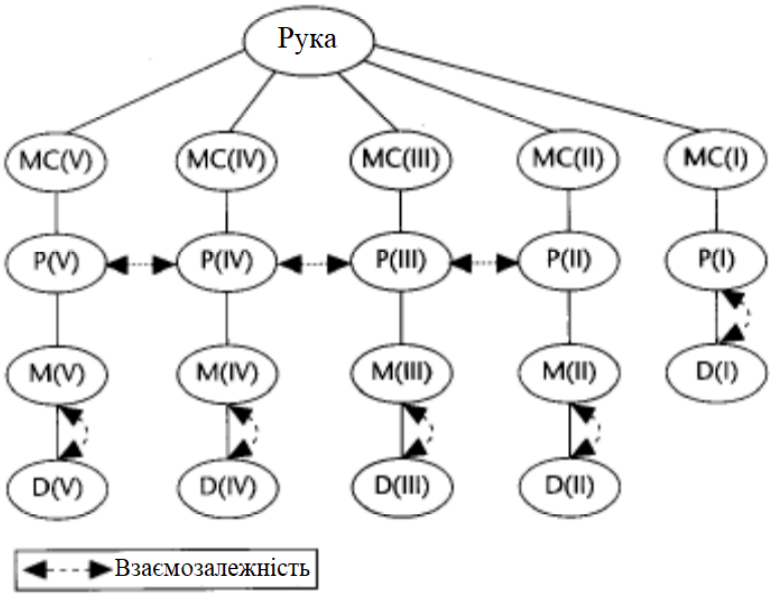


Рис. 1.3. Обмеження та залежність скелета руки людини

На рис. 1.3 кожен вузол представляє сегмент руки. Взаємозалежність в русі між сегментами представлена пунктирними лініями, проведеними між відповідними вузлами. Для руки існує 27 ступенів свободи і різних змін форм завдяки рухам суглоба.

1.1.2. Підходи, які базуються на 27 ступенях свободи моделі руки людини

Модель з 27 ступенями свободи є базою для підходу до аналізу рухів руки людини шляхом синтезу. Якщо, для певного спеціального використання, розробнику не потрібно стільки ступенів свободи, то можна зафіксувати деякі DOF, щоб досягти більш високої продуктивності для системи відстеження рук.

Так, в роботі [32] розроблено 12-DOF базу даних моделей рук на основі 27-DOF моделі. Проаналізовано деякі обмеження 27-DOF моделі та проведено зменшення до 12-DOF моделі без значного погіршення продуктивності. У моделі з роботи [33] використано обмеження, які успішно нехтують трьома DOF на кожен палець (крім великого пальця) та випускають три DOF для великого пальця. Це означає, що є лише одна DOF на чотири пальці і дві DOF для великого пальця, отже, це 6 DOF на всі п’ять пальців, плюс 6 DOF для поступального руху та обертання зап'ястя, — всього лише 12 DOF у моделях руки. У статті відмічається, що ця модель не погіршила показники ефективності.

1.2. Підходи до моделювання та аналізу рухів руки людини

В останні роки було докладено значні зусилля, спрямовані на розпізнавання жестів та пов'язану з цим роботу в аналізі руху тіла через зацікавленість в більш природній та захоплюючій людино-машинній взаємодії (HCI). Оскільки вартість більш потужних комп'ютерів зменшується, а ПК стає дедалі популярнішим, бажано використовувати більш природний інтерфейс, а не традиційні пристрої введення, такі як миша та клавіатура. Використання жестів як одного з найбільш природних способів спілкування людей стає очевидним вибором для більш природного інтерфейсу [34, 35]. Ефективне розпізнавання жестів забезпечить основні переваги не тільки у віртуальному середовищі та інших програмах HCI, але й у таких сферах, як телеконференції, спостереження та анімація персонажів.

Для розпізнавання рухів руки необхідно включати пошук загального руху руки та локального руху кожного пальця таким чином, щоб положення руки

можна було відновити. Один з можливих способів аналізу руху руки — це підхід, заснований на зовнішньому вигляді, який базуєтьсяє на аналізі форм рук на зображеннях [36, 35]. Однак локальний рух руки дуже важко оцінити за допомогою цього засобу. Інший можливий спосіб — модельний підхід [37, 43], в якому за допомогою однієї відкаліброваної камери локальні параметри руху руки можна оцінити, встановивши 3D-модель руки для спостережних зображень.

Один із методів, заснованих на модельному підході, полягає у використанні обмежених нелінійних методів програмування на основі градієнта для одночасного оцінювання глобального та локального руху руки [41]. Недоліком такого підходу є те, що оптимізація часто потрапляє в локальні мінімуми. Інша ідея полягає в моделюванні поверхні кисті та оцінці конфігурацій руки, використовуючи підхід «аналіз за синтезом» [39]. 3D-моделі-кандидати проєктуються на площину зображення, і найкраща відповідність виявляється стосовно деякого вимірювання подібності. По суті, проблема пошуку в дуже високомірному просторі робить цей метод обчислювально складним. Також пропонується метод декомпозиції для аналізу зчленованого руху руки шляхом поділу на її глобальний рух та локальні рухи пальця [43].

Дослідження обмежень руху рук є складною проблемою: хоча вони допоможуть зменшити розмір простору пошуку, але занадто велика їх кількість або занадто складні обмеження також додають труднощів для обчислювальної реалізації. Важливим питанням стає: які обмеження прийняти? Деякі обмеження вже виявлені, вивчені та використані у багатьох роботах [37-40]. До загальних обмежень відносяться обмеження суглобів у межах одного пальця, обмеження суглобів між пальцями та максимальний діапазон рухів пальців. В моделях ці обмеження подано як рівняннями, так і нерівностями. Однак, завдяки високій гнучкості в русі пальця існує ще більше обмежень, які не можуть бути явно представлені рівняннями.

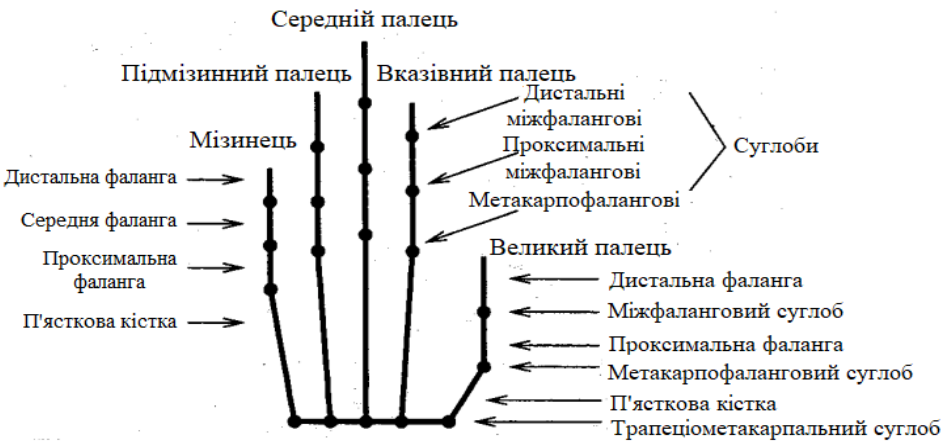


Рис. 1.4. Кінематична структура та позначення суглобів

Рука людини чітко зчленована. Для моделювання з’єднання пальців слід моделювати кінематичну структуру руки. Скелет руки можна зробити абстракцією для кожного пальця як кінематичний ланцюг із базовою рамкою на долоні та кінчиком пальця як кінцевим ефектом. Така кінематична модель руки зображена на рис. 1.4 із назвами кожного суглоба. Ця модель має 27 ступенів свободи. Є 21 DOF, що сприяють суглобам пальців для локального руху, і 6 DOF за рахунок глобального руху [40]. Зчленований локальний рух руки, тобто рух пальця, може бути поданий набором значень кута суглоба.

Деякі обмеження руху можуть мати зображення закритої форми, і вони часто використовуються у сучасних дослідженнях анімації та захоплення візуального руху [37-40, 43]. Однак багато обмежень руху дуже важко подати у закритих формах. Питання, як моделювати такі обмеження, все ще потребує подальшого дослідження.

Обмеження руху рук можна приблизно поділити на три типи. Обмеження I типу — це обмеження рухів пальцями як результат анатомії руки, які зазвичай називають статичними обмеженнями. Обмеження II типу — це обмеження, накладені на суглоби під час руху, які зазвичай в попередніх роботах називали динамічними обмеженнями. Обмеження III типу застосовуються при виконанні природного руху та ще не вивчені.

Обмеження I типу. Цей тип обмеження відноситься до меж діапазону рухів пальцями як результат анатомії руки. Ми розглянемо лише діапазон руху

кожного пальця, якого можна досягти без застосування зовнішніх зусиль, наприклад, згинання пальців назад за допомогою іншої руки. Цей тип обмежень зазвичай представлений такими нерівностями:

$$0° \leq \theta_{MCP\_F} \leq 90°$$
$$0° \leq \theta_{PIP\_F} \leq 110° \quad (1.2.1)$$
$$0° \leq \theta_{DIF\_F} \leq 90°$$
$$-15° \leq \theta_{MCP\_AA} \leq 15°$$

де F позначає згинання, а AA позначає абдукцію/аддукцію.

Обмеження II типу. Цей тип обмеження стосується обмежень, накладених на суглоби під час рухів пальцями. Ці обмеження часто називають динамічними обмеженнями, і їх можна розділити на обмеження внутрішньопальцеві та між пальцями. Внутрішньопальцеві обмеження — це обмеження між суглобами одного пальця. Загальновживане обмеження на основі анатомії рук констатує те, що для вказівного, середнього, підмізинного та мізинця для згинання суглобів DIP також слід зігнути відповідні суглоби PIP. Обмеження між пальцями застосовується для суглобів між сусідніми пальцями. Наприклад, коли вказівний суглоб MCP зігнутий, середній MCP суглоб також змушений згинатися. В роботі [40] проведено вимірювання на кількох людях і отримано набір нерівностей, що наближає межі сусідніх суглобів MCP. Однак є ще багато обмежень, які не можуть бути явно описані рівняннями.

Обмеження III типу. Ці обмеження застосовуються завдяки природності рухів і є більш точними для виявлення та кількісної оцінки. Відзначимо, що дуже мало було зроблено для врахування цих обмежень при моделюванні природного руху руки. Обмеження III типу відрізняються від типу II тим, що вони не мають нічого спільного з обмеженнями, встановленими анатомією руки, а є результатами загальних і природних рухів. Наприклад, найприродніший спосіб для кожної людини зробити кулак з відкритої руки — це згортати всі пальці одночасно, а не закручувати кожен палець по черзі. Цей тип обмежень також складно описати аналітичними залежностями.

Відзначимо, що досить важко явно описати обмеження природних рухів рук у закритому вигляді. Однак їх можна отримати з великого та репрезентативного набору навчальних зразків. Тому в дисертаційному дослідженні запропоновано побудувати конфігураційний простір (тобто простір суглобових кутів) та вивчити обмеження безпосередньо з емпіричних даних, використовуючи описаний нижче підхід.

1.2.1. Моделювання руки

Відзначимо, що використання 3D-рушія [44] для моделювання руки та рендеринг нетривіального завдання є поза рамками впровадження запропонованої технології моделювання та розпізнавання української дактильної мови. Для практичного застосування можна скористатися розробленими кросплатформеними 3D-рушіями, що мають різні умови ліцензування, зокрема, деякі є відкритими, а деякі — патентованими, деякі рушії є безкоштовними на конкретних умовах, таких як академічні та дослідницькі проекти, некомерційні проекти.

Усі тривимірні рушії мають 3 ключові моменти для порівняння: зручність у використанні (інтерфейс користувача, як легко було вивчати та розробляти), функціональність (наскільки складні сцени можуть бути реалізовані), ціна і ліцензування. Розглянемо деякі з рушіїв з урахуванням цих характеристик.

Наприклад, Unreal Engine [45] — ігровий рушій, що розроблений і підтримується компанією Epic Games. Пристосований для створення ігор і завдяки своєму коду, написаному на C++, Unreal Engine має високий рівень портативності та є інструментом, який на сьогодні застосовують багато розробників ігор, він є доступним.

Рушій GameMaker Studio [46] (раніше: Animo до 1999, Game Maker до 2011, GameMaker до 2012, і GameMaker: Studio до 2017) — це ігровий механізм кросплатформених розробок, створений YoYo Games. Цей рушій дозволяє створення багатоплатформних та багатожанрових відеоігор за допомогою користувацької drag-and-drop мови візуального програмування або скриптової

мови, відомої як Game Maker Language. Цей рушій підтримує побудову для Microsoft Windows, macOS, Ubuntu, HTML5, Android, iOS, Amazon Fire TV, Android TV, Microsoft UWP, PlayStation 4, Xbox One, Nintendo Switch.

Ще один рушій Unity3D [47] — це міжплатформений ігровий рушій Mac OS X з підтримкою понад 25 різних платформ, включаючи мобільні, настільні, віртуальної реальності, а також консолі. Його можна використовувати для створення тривимірної, двовимірної, віртуальної реальності та ігор з доповненою реальністю, а також для моделювання та іншого застосування. Рушій використовується у галузях, що не входять до відеоігор, таких як кіно, автомобілебудування, архітектура, інженерія та будівництво. Відзначимо, що Unity — це кросплатформений рушій, редактор якого підтримується на Windows та MacOS, причому версія редактора доступна для платформ Linux, iOS, Android, Tizen, WebGL, PlayStation 4, PlayStation Vita, Xbox One, 3DS, Oculus Rift, Google Cardboard, Steam VR, PlayStation VR, Gear VR, Windows Mixed Reality, Daydream, Android TV, Samsung Smart TV, tvOS, Nintendo Switch, Fire OS, ігрову Facebook, Apple ARKit, ARCore Google, Vuforia та Magic Leap тощо

Табл. 1.1. Порівняння 3D-рушіїв за функціями та підтримуваними платформами

	Unity3d	GameMaker	Unreal Engine 4
Повна підтримка 3D	Доступний	Недоступний	Доступний
Безкоштовна версія	Доступний	Доступний	Недоступний
ПК платформа	Доступний	Доступний	Доступний
Playstation платформа	Доступний	Доступний	Доступний
iOS платформа	Доступний	Доступний	Доступний

Android платформа	Доступний	Доступний	Доступний
Windows платформа	Доступний	Доступний	Доступний
MacOS платформа	Доступний	Доступний	Доступний
Web платформа	Доступний	Доступний	Недоступний

У результаті порівняльного аналізу 3D-рушіїв (див. табл. 1.1) в якості фреймворку для реалізації 3D-моделі руки було обрано двигун Unity3d завдяки наявності в нього всіх необхідних платформ (мобільних, настільних, відеоконсолей та веб), безкоштовної ліцензії на дослідницькі та некомерційні проєкти та повної 3D-підтримки. Відзначимо, що рушій GameMaker не має повної тривимірної підтримки, тому його відхилили, і, незважаючи на те, що Unreal Engine є надзвичайно потужним тривимірним рушієм, має високий рівень вхідного користування, він не підтримує веб-платформу та не має безкоштовної версії, що зробило запропоновану технологію недоступною. Крім того, підтримка веб-версії є вирішальною у випадку пристроїв низького класу, які демонструють низьку продуктивність заради високих показників енергозбереження. Також рушій Unity3d підтримує платформи Magic Leap, які можуть бути використані у подальших розробках запропонованої технології.

1.3. Підходи до розпізнавання дактилем

Розпізнавання жестів — напрям у комп'ютерній науці та мовних технологіях, що має за мету інтерпретацію людських жестів за допомогою математичних алгоритмів. Жести можуть походити з будь-якого руху тіла або стану, але зазвичай походять від обличчя чи рук. Поточні цілі в цій галузі включають розпізнавання емоцій на обличчі та жестів руки. Користувачі можуть використовувати прості жести для керування або взаємодії з пристроями, без контактної взаємодії. Для інтерпретації жестової мови було перебрано значну кількість підходів за допомогою камер та алгоритмів комп'ютерного зору. Однак

ідентифікація та розпізнавання постави, ходи, проксемії та поведінки людини також є предметом вивчення техніки розпізнавання жестів. Розпізнавання жестів можна розглядати як спосіб, завдяки якому комп'ютери почнуть розуміти мову тіла людини, налагоджуючи, таким чином, між машинами та людьми зв'язок більш міцний, ніж прості текстові користувацькі інтерфейси або навіть графічні інтерфейси (GUI), які все ще обмежують більшість вхідних даних клавіатурою та мишею, і взаємодіяти природним шляхом без будь-яких механічних пристроїв. Діаграма підходів розпізнавання жестової інформації наведена на рис. 1.5.

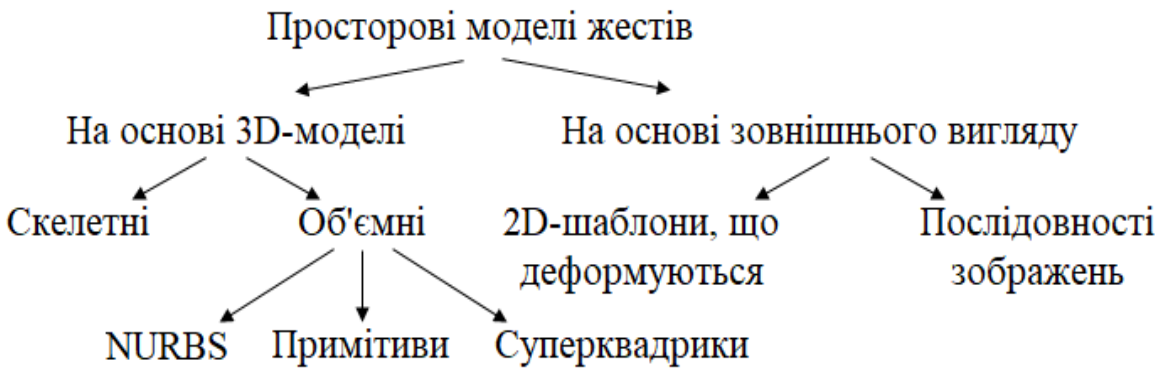


Рис. 1.5. Діаграма підходів жестового розпізнавання

Залежно від типу вхідних даних, підхід до інтерпретації жесту може бути реалізований різними способами, однак більшість методів спираються на ключові покажчики, подані в системі 3D-координат. На основі цього відносного руху жест може бути виявлений з високою точністю, залежно від якості введення та підходу алгоритму.

Щоб інтерпретувати рухи тіла, треба класифікувати їх за загальними властивостями та повідомленнями, які можуть виражати рухи. Наприклад, у мові жестів кожен жест представляє слово або фразу. Таксономія, яка є дуже вдалою для взаємодії людини та комп'ютера [48], демонструє кілька інтерактивних жестових систем, щоб охопити весь жестовий простір:

- маніпулятивна;
- сигнальна;

- комунікативна.

У науковій літературі розрізняють два різних підходи до розпізнавання жестів: на основі 3D-моделі та базуючись на візуальному вигляді. Перший метод використовує 3D-інформацію про основні елементи частин тіла, щоб отримати декілька важливих параметрів, таких як положення долоні або кути суглоба. З іншого боку, системи на основі візуального (зовнішнього) вигляду використовують зображення чи відео для прямої інтерпретації [49].

В роботі [50] запропоновано розпізнавання жестів як нової форми взаємодії людини та комп'ютера. Автор розробив інтерактивне середовище під назвою «комп’ютерно-кероване сприйнятливе середовище», простір, в якому все, що побачив чи почув користувач, відповідало на те, що він зробив. Замість того, щоб сідати і рухати лише пальцями, користувач взаємодіє зі своїм тілом. В одному зі своїх застосувань проєкційний екран стає лобовим склом транспортного засобу, який учасник використовує для навігації у графічному світі. Стоячи перед екраном, простягаючи руки і схиляючись у напрямку, в якому хоче йти, користувач може створити графічний пейзаж. Однак, це дослідження не може розглядатися суто як система розпізнавання жестів руки, оскільки потенційний користувач використовує не тільки руку для взаємодії з системою, а й тіло та пальці.

В роботі [51] розпізнавання жестів було адаптоване під різні застосування, починаючи жестами обличчя і закінчуючи рухами всім тілом людини. Таким чином, виникли різні застосування, що і створили сильнішу потребу в цьому типі системи розпізнавання.

Інтенсивні дослідження розпізнавання жестів руки розпочалися, коли з’явилися перші системи захоплення кадрів для кольорового відеовведення, що дозволило дослідникам захоплювати кольорові зображення в режимі реального часу [52].

Аналіз жестів руки можна виконувати двома основними підходами, а саме: аналіз на рукавичках, аналіз на основі зору [53]. Підхід на основі рукавичок використовує давачі (механічні або оптичні), прикріплені до рукавички, які

діють як ретранслятор процесу згинання пальця в електричні сигнали для визначення стану руки. Відносне положення руки визначається додатковим сенсором. Цей сенсор зазвичай є магнітним або акустичним та кріпиться до рукавички. Програмні інструменти пошуку комплектуються рукавичкою для маніпулятивного застосування. Цей підхід застосовано в роботі [54] для розпізнавання жестів ASL. Рівень розпізнавання становив 75%. Обмеженням цього підходу є те, що користувач зобов’язаний носити громіздкий пристрій і, до того ж, нести багато кабелів, які з'єднують пристрій з комп'ютером [55]. Інша система розпізнавання жестів рукою була запропонована в роботі [56] для розпізнавання чисел від 0 до 10, де кожне число було подане певним жестом руки. Ця система проводить розпізнавання у три основні кроки, а саме: захоплення зображення, початок застосування та безпосереднє розпізнавання чисел. Хоча система досягла рівня розпізнавання 89%, але вона має низку обмежень, оскільки функціонує лише за певних умов, таких як надягання кольорових рукавичок і використання чорного фону.

Другий підхід — аналіз на базі комп’ютерного зору — базується на тому, як люди сприймають інформацію про своє оточення [53]. У цьому підході було використано кілька методик вилучення для розпізнання особливостей зображень жестів. Ці методи включають в себе орієнтаційну гістограму [57, 58], вейвлет-трансформацію [59], коефіцієнти форми Фур’є [60], момент Зерніке [61], фільтр Габора [62, 63, 64], векторне квантування [65], контурні коди [66], моменти Ху [67], геометричну особливість [68], Finger-Earth Mover’s Distance (FEMD) [69] тощо.

Більшість цих методів розпізнавання мають деякі обмеження. Наприклад, в орієнтовній гістограмі, розробленій в роботі [70], алгоритм використовує гістограму локальної орієнтації. Цей простий метод добре працює, якщо зображення однієї і тієї самої жестової карти відображають аналогічні орієнтаційні гістограми, а різні жести відображають істотно різні гістограми [57]. Проблема використання даного методу полягає в тому, що одні й ті самі жести можуть мати різні орієнтаційні гістограми, а різні жести можуть мати аналогічні

орієнтаційні гістограми, що впливає на його ефективність [71]. Цей метод використовувався у роботі [57] для розпізнавання 10 різних жестів руки і використовував метод найближчого сусіда для розпізнавання. Цей же метод розпізнавання був застосований в іншому дослідженні [58] для розв'язання проблеми розпізнавання підгрупи американської мови жестів. На етапі класифікації тут використовується одношаровий перцептор для розпізнавання зображень жестів. Використовуючи той самий метод розпізнавання, а саме гістограму орієнтації, у роботі [53] запропоновано метод розпізнавання жестів, що використовує як статичні жести, так і оригінальні динамічні жести.

У роботі [62] автори використовували Габорів фільтр з PCA для розпізнавання, а потім метод нечіткої кластеризації для розпізнавання 26 жестів алфавіту ASL. Незважаючи на те, що система досягла істотно вищої точності розпізнавання 93,32%, її піддавали критиці за обчислювальну складність, що може обмежити її використання в реальних програмах [64].

Етап виділення ознак зазвичай супроводжується методом класифікації, який використовує витягнутий векторний елемент для класифікації зображення жестів до відповідного класу. Серед методів класифікації, які застосовуються: метод найближчого сусіда [57, 60, 61], штучні нейронні мережі [54], метод опорних векторів (SVM) [63, 67, 64], приховані моделі Маркова (HMM) [72].

Класифікатор за методом найближчого сусіда має недоліки, такі як слабкість узагальнення, чутливість до шумів та вибору міри відстані [73].

Жести рук, які виконуються однією або двома руками, можна класифікувати відповідно до їх застосування в різних категоріях, включаючи розмовні, контрольні, маніпулятивні та комунікативні жести. Загалом, розпізнавання жестів руки має на меті виявлення конкретних людських жестів та використовування їх для передачі інформації. Процес розпізнавання жестів рукою складається в основному з чотирьох етапів:

1) збирання зображень жестів руки;

2) попередня обробка зображень жестів з використанням різних методів, включаючи виявлення меж, фільтрування та нормалізацію;

3) захоплення основних характеристик жестів за допомогою алгоритмів вилучення особливостей;

4) оцінювання (або класифікація), коли зображення відноситься до відповідного йому класу жестів. Є багато методів, які були використані на етапі класифікації в процесі розпізнавання жестів, наприклад, штучні нейронні мережі ANN, сегментація, HMM та динамічне викривлення часу.

1.3.1. Концепції обробки зображень

Розпізнавання жестів — це складне завдання, яке, насамперед, передбачає дві фази: спочатку — вилучення особливостей, що визначають жест на зображенні, потім, використовуючи відповідний класифікатор (у цьому дослідженні це нейронна мережа), кожен жест призначається до відповідного класу на основі вищезгаданого вилучення особливостей. Ці дві процедури містять численні техніки та методи, які підпадають під такі області:

• обробка зображення;

• нейронна мережа.

Розглянемо основні етапи попередньої обробки зображень для задач розпізнавання.

Початковою стадією для будь-якого процесу розпізнавання є сегментація, коли отримане зображення розбивається на змістовні регіони або сегменти. Процес сегментації стосується лише розділення зображення, а не того, що відображають його області. У найпростішому випадку (бінарні зображення) наявні лише дві області: передній план (об’єкт) і область заднього плану. У зображеннях в градації сірого в межах зображення може існувати кілька типів областей або класів. Наприклад, коли природна сцена сегментована, можуть існувати регіони хмар, землі, будівель та дерев [74]. Сегментація поділяє зображення на його складові частини, рівень яких залежить від поставленої задачі. Зазвичай, як показано в дослідженні [75], сегментацію слід припинити, коли об'єкти сцени будуть ізольованими.

У випадку, коли зображення містить об'єкт, що має однорідну інтенсивність та тло з іншим рівнем інтенсивності [76], сегментацію можна здійснити, використовуючи методи порогового визначення, наприклад, розділення гістограми зображення за допомогою одного порогу T. Потім сегментація виконується шляхом сканування зображення попіксельно та позначення кожного пікселя як об'єкта або тла залежно від того, чи рівень сірого на пікселі більше або менше порогового значення T [75] (1.3.1).

$$g(x,y) = \begin{cases} 1, \ if \ g(x,y) > T \\ 0, \text{в іншому випадку} \end{cases} \tag{1.3.1}$$

Ще однією задачею попередньої обробки зображення для задач розпізнавання є виділення контуру об'єкта. Ця інформація може бути використана для аналізу зображення, його фільтрації та ідентифікації об’єкта [76]. Одним з найкращих простих контурних операторів є оператор Собеля, що використовує дві (3×3) маски [74]. Маски Собеля для виявлення контурів шукають горизонтальний і вертикальний напрямки та комбінують цю інформацію в єдиний вимір. Алгоритм використання масок Собеля такий. Кожна маска приводиться у відповідність до зображення. В місці кожного пікселя існують два числа, а саме P_1 і P_2, які відповідають рядку і стовпцеві маски відповідно. Ці числа використовуються для обчислення двох вимірів, а саме: величини ребра та напрямку [77]:

$$\begin{aligned} \text{Величина Краю} &= \sqrt{P_1^2 + P_2^2} \\ \text{Напрямок Краю} &= tan^{-1}(P_1 / P_2) \end{aligned} \tag{1.3.2}$$

Вилучення особливостей є частиною процесу скорочення даних, за яким слідує аналіз. Одним із важливих аспектів аналізу особливостей є визначення таких, що є важливими [77]. Мета вилучення особливостей полягає у пошуку найбільш дискримінуючої інформації у записаному зображенні. Вилучення особливостей оперує двовимірними масивами зображень, але створює перелік описів або ознаковий вектор [74, 78]. Математично ознакою є n-вимірний вектор

зі своїми компонентами, отриманими за допомогою певного аналізу зображення: колір, текстура, форма, просторова інформація та руху відео тощо. Наприклад, як показано в роботах [78, 79], колір може представляти інформацію про колір у зображенні, зокрема, кольорову гістограму, кольорові бінарні набори або кольорові узгоджені вектори.

1.3.2. Підходи машинного навчання для класифікації зображень

Алгоритми класифікації — це системи, які дозволяють відносити вхідні дані до одного з відомих класів. Контрольоване машинне навчання використовує набір позначених навчальних даних, які використовуються для виведення функції класифікатора. Існує велика кількість класифікаторів (див., наприклад, огляд [80]). У відповідності до мети дисертаційних досліджень розглянемо три основні способи класифікації зображень: метод опорних векторів (SVM) [81], метод k-найближчих сусідів (k-NN) та алгоритми лісів випадкових рішень (RDF) [82].

Метод опорних векторів SVM [81] знаходить оптимальну гіперплощину для розділення навчальних даних за їх відомою розміткою. Метод максимально збільшує відстань від розділюючої гіперплощини. Для пошуку оптимальної гіперплощини використовуються методи оптимізації, і такий класифікатор широко використовується для методів аналізу поведінки людини та для багатьох наборів даних [83, 84].

Метод найближчих сусідів (k-NN) — це простий статистичний метод, у якому вхідні дані будуть віднесені до найбільш розповсюдженого класу серед його найближчих сусідів у навчальному наборі. Сусіди визначаються за мірою схожості, яка часто обчислюється з особливостей, отриманих із необроблених даних. У роботі [85] використано метод k-NN на основі дескрипторів HOG та SIFT для класифікації статичних жестів. Однак кілька експериментів щодо порівняння точності k-NN та SVM показали, що продуктивність k-NN порівняно нижча [86, 87, 88].

Метод RDF [82] складається з набору дерев рандомізованої класифікації. Кожне з них пов’язується з випадковим набором вихідних навчальних даних.

Кожне дерево бінарної класифікації будується шляхом рекурсивного розподілу вхідних даних на кожному вузлі, щоб зменшити ентропію розподілу класів. На кожному вузлі обирається випадкова підмножина функцій і обирається поріг, що призводить до найбільшого зменшення ентропії розподілу класів. Відзначимо, що метод лісів випадкових рішень використовується в роботі [108] для статичного розпізнавання жестів на основі функцій фільтра Габора.

Останнім часом у багатьох додатках з методів комп'ютерного зору (CV) (англ. Computer Vision) було показано зміну парадигми: від розпізнавання дій людини — до розпізнавання мовлення, класифікації зображень та маркування, де в області CV спостерігається поява та успішне впровадження технології машинного навчання, що називається глибоким навчанням. Починаючи з 2010 року, дослідники переходять від традиційних функцій ручної роботи до функцій на основі знань, які також називаються алгоритмом керованих даних. Існує багато функціональних методів, заснованих на вивченні завдань візуального розпізнавання, таких як підходи на основі онтологій або генетичне програмування. Тому в дисертаційному дослідженні буде приділено значну увагу методам глибинного навчання, оскільки останніми роками вони стають домінантними в CV.

1.4. Використання нейромережевих технологій для задач розпізнавання

Алгоритми розпізнавання на основі комп'ютерного зору можна розділити на дві групи: без навчання та на основі навчання. Багато CV проблем можна вирішити, використовуючи правильний набір функцій із даних, щоб виконати завдання. Конвеєр алгоритмів на основі ручної роботи часто складається з трьох основних етапів:

1) Генерування даних та попередня обробка даних. Дані збираються з пристроїв, наприклад 2D та/або 3D-давачів, які є входами алгоритмів. Ці дані часто попередньо обробляються. Як правило, етап попередньої обробки складається з виявлення переднього плану, видалення тла та/або фільтрації

необроблених даних для видалення викидів і значень за замовчуванням. Наприклад, в дослідженні [89] набір 3D-стиків — скелет тіла — витягується та обчислюється з послідовності «глибинних» зображень людини і є вхідним алгоритмом розпізнавання жестів. Алгоритми, як правило, використовують комбінацію декількох необроблених даних та/або зображень, щоб отримати максимум відповідної інформації [90].

2) Отримання вручну побудованих ознак. Особливості ручної роботи також називають конструкторською функцією. Вручну побудовані ознаки, як правило, походять від людської інтуїції та попередніх знань, що стосуються конкретних проблем. Для дослідників є певне розуміння проблематики, що випливає з конкретного типу даних, наприклад: як розпізнати різні жести із «глибинних» зображень? Використовуючи різні математичні методи, дослідники виділяють певні особливості (характеристики) із початкових даних, які дають змогу більш правильно вирішити поставлену задачу.

3) Научуваний класифікатор. Машинне навчання дозволяє вирішувати завдання, які надто складно вирішити за допомогою програми дизайну рук. Проблеми з класифікацією — це завдання машинного навчання, коли програмі пропонується вказати, до якої з категорій K або міток належать деякі вхідні дані. Нехай $L = \{(\beta_i, y_i)\}\, i = 1 \ldots N$ — це набір даних, де $\beta_i \in Rn$, y_i— категорійна змінна, $\in 1, \ldots, k$ — сигнатура, причому $yi = f(\beta i)$ — аксіома. Мета тут — знайти функцію математичного наближення $y\hat{}i = h\,(\beta i, \theta)$ з f , яка відображає вхідні дані з її сигнатурою. Ця функція, яка називається класифікатором, повинна мінімізувати функцію витрат, що усуває невідповідність між результатами функції класифікації $y\hat{}$ і аксіомою $y\hat{}$.

Відзначимо, що використання традиційних алгоритмів машинного навчання, таких як Support Vector Machine [81], Random Forest [82] або Hidden Markov Model [91], сильно залежить від обраного способу відображення даних. Відзначимо також, що сучасні алгоритми глибокого навчання дають значні результати щодо вирішення багатьох CV проблем, але вони також мають недоліки. Для належної роботи їм потрібна величезна кількість даних, що

залишається проблемою в певній галузі досліджень, де дані створюються не так легко, як, наприклад, 3D-дані. Крім того, для навчання та параметризації глибоких нейронних мереж потрібно багато обчислювальних ресурсів та експериментів.

1.5. Кросплатформені засоби розробки технології моделювання та розпізнавання жестів

Значення платформи може відрізнятися, але, відповідно до дослідження [6], платформа «може позначати тип процесора та/або іншого обладнання, на якому працює дана операційна система або поєднання типів апаратних засобів, тип операційної системи на комп'ютері або комбінацію типу апаратного забезпечення та типу операційної системи, на якій працюють». У цій роботі платформа розглядається як будь-яке з трьох визначень, наведених вище.

Існує багато підходів для використання такої технології на широкому спектрі платформ і пристроїв. Один з можливих підходів — це впровадження окремої програми на кожній платформі чи типі пристроїв. Хоча такий підхід легше здійснити (немає необхідності підтримувати єдину базу коду, яка б працювала на всіх платформах, та наявний ширший набір розробки фреймворків та інструментів для кожної платформи), головним недоліком є необхідність великої кількості різних реалізацій, що збільшує витрати на розробку, не забезпечує однакову функціональність на різних пристроях та апаратних засобах, і, можливо, призведе до того, що деякі платформи взагалі не матимуть відповідної реалізації технології через конкретні причини (наприклад, відсутність фреймворку розробки з необхідною функціональністю на цій платформі, типі апаратури конкретно цього пристрою).

Іншим підходом до впровадження технології вивчення української дактильної абетки та української жестової мови є розробка платформ, запропонованих в роботі [8]. Відповідно до [8], кросплатформа «відноситься до здатності програмного забезпечення працювати на більш ніж одній платформі з однаковою (або майже однаковою) функціональністю». Оскільки технологія має

на меті роботу на декількох платформах(у будь-якому сенсі платформи), це означає, що технології працюють з однаковим функціоналом незалежно від будь-якого чинника окремо або комбінації цих чинників: операційної системи (наприклад, FreeBSD, Linux, Mac OS X, Solaris і різні системи Microsoft Windows), типу процесора (наприклад, x86, PowerPC, SPARC або Alpha) та типу апаратної системи (наприклад, мейнфрейм, робоча станція, настільний ПК, портативний ПК або вбудований). Це визначення схоже на термін незалежність платформи. Незалежність платформи передбачає, що програмне забезпечення буде працювати на будь-якій платформі, тоді як сенсом кросплатформи є те, що програмне забезпечення буде працювати як мінімум на двох платформах. У дисертаційній роботі кросплатформа розглядається як незалежна платформа, яка, можливо, не зважає на деякі менш розширені платформи. Міжплатформені технології можна використовувати замість віртуальних машин [8] або як набір моноплатформених технологій. Використання цих технологій дозволяє розробити єдину базу коду для різних типів платформ, незалежно від типу процесора, операційної системи продуктивності обладнання, та безперешкодно використати її на всій платформі.

1.6. Висновки до Розділу 1

З огляду на тематику дисертаційного дослідження, в розділі 1:

- показано, що моделювання та розпізнавання жестів використовується в багатьох галузях (комунікації, людинно-компютерних інтерфейсах тощо);
- розглянуто класичні та сучасні підходи щодо моделювання жестів за допомогою кросплатформених технологій та тривимірних моделей;
- розглянуто класичні та сучасні підходи щодо розв’язання задачі розпізнавання жестів за допомогою глибокого навчання та кросплатформених технологій.

На основі аналізу існуючих систем моделювання жестів та систем розпізнавання жестів сформульовано перелік задач для дисертаційного дооослідження:

- розробити скелет тривимірної моделі руки, на базі якого розробити високополігоональну модель руки, за допомогою якої відтворити дактилеми української дактильної абетки та анімацію переходів між ними;
- розробити систему з розпізнавання дактилем із зображення на базі методів глибокого навчання із застосуванням згорткових нейрнних мереж із тривимірними згортками;
- зібрати та використати для навчання, тестування та підбору оптимальної архітектури та її гіперпараметрів базу даних зображень жестів української дактильної абетки.

Розділ 2. Моделі та методи опису тривимірного скелету та моделі руки, моделювання та розпізнавання жестів за допомогою глибокого навчання

В даному розділі розглянуто питання розробки математичної моделі скелету руки і запропоновано підхід до визначення структури та параметрів просторової моделі руки. Також розглянуто підхід до анімації жестів та окремо розглянуто адаптивність параметрів просторової моделі руки. Наведено приклади застосування математичної моделі для задачі моделювання жестів української дактильної абетки. Розглянуто підхід подання даних, обробки даних та побудови моделі для розпізнавання жестів української дактильної абетки за допомогою глибокого навчання, тривимірних згорток та просторово-часових даних.

2.1. Моделювання скелету руки людини

2.1.1. Модель скелету руки для моделювання жестової інформації

При моделюванні тривимірної моделі руки, в рамках виконаної роботи, першим кроком розроблено модель скелету, яка лежить в основі моделювання жестів, адже саме завдяки цій моделі, яка включає в себе 31 DOF та модель обмежень суглобів, було створено систему реалістичного моделювання жестів та переходів між жестами. Таким чином, рух руки в цілому та всіх її частин відбувається лише у рамках реалістичних анатомічних обмежень. Особливо це важливо під час моделювання та анімації жестових переходів, коли руку з позиції першого жесту необхідно трансформувати у позицію наступного жесту шляхом зміни конфігурації скелету.

У роботі запропоновано математичну модель скелету руки (рис. 1), яка складається з 21 вузлів (суглобів). Кожен вузол має певну кількість ступенів свободи (DOF), позначену відповідним числом у кожному вузлі, зокрема:

кінчики пальців не мають DOF, а кореневий вузол (кисть) має 6 DOF, інші вузли мають 3 та 1 DOF.

Лінією з точок позначено дистальні інтрафаланги (ДІФ), штрихованою лінією позначено проміжні інтрафаланги (ПІФ), штрих-пунктирною лінією з двома точками позначено апроксимуючі інтрафаланги (АІФ), штрих-пунктирною лінією позначено п’ястки (ПФ). Відповідні кути та DOF позначені біля вузлів великого пальця.

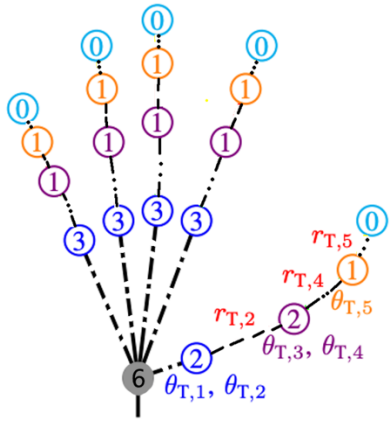


Рис. 2.1. Структура скелету руки

Скелет являє собою ієрархічну структуру, де кожна кістка є батьківською або дочірньою до іншої. Таким чином, скелет визначається як незалежний набір пов’язаних кісток (ланцюг), який має лише відношення батько-дитина. У моделі подано п’ять підскелетів $(S0, \dots, S4)$, по одному для кожного пальця, і всі вони починаються з одного вузла. У скелеті кожного пальця подано чотири кістки $(B0, B1, B2, B3)$:

$$\forall S_i : S_i = \{B_{i0}, B_{i1}, B_{i2}, B_{i3}\}. \quad (2.1.1)$$

Запропонована модель скелету відповідає реальній анатомічній системі кісток руки.

2.1.2. Обмеження суглобів скелету

У розробленій моделі скелету подано 3 види обмежень у кожному суглобі: статичні, внутрішні (на рівні одного пальця) та зовнішні (на рівні між пальцями). На рисунку 2.2 зображено схему обмеження суглобів, що забезпечує нульову

відносну швидкість у точці кріплення, яка з'єднує кістки. Обмеження також розділяються за пальцями на дві категорії: існує модель обмежень для великого пальця та загальна модель обмежень для всіх інших пальців.

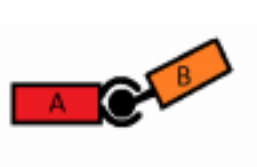


Рис. 2.2. Приклад обмеження суглоба

Внутрішня загальна модель обмежень пальця:

$$\theta_{\text{ДІФ}} \approx \frac{2}{3}\,\theta_{\text{АІФ}}$$

$$\theta_{\text{АІФ}} \approx \frac{3}{4}\,\theta_{\text{ПІФ (згинання/розгинання)}} \qquad (2.1.2)$$

Внутрішня модель обмежень великого пальця:

$$\theta_{\text{ПрІФ}} \approx \frac{1}{2}\,\theta_{\text{ПІФ (згинання/розгинання)}}$$

$$\theta_{\text{ПІФ(згинання/розгинання)}} \approx \frac{5}{4}\,\theta_{\text{ВеликийПалецьІФ (згинання/розгинання)}} \qquad (2.1.3)$$

У моделі обмежень також подано зовнішні обмеження, які корелюють із вузлами (суглобами) двох різних пальців.

2.1.2.1. Парний рух

У моделі розроблено парний рух під час згинання підмізинного пальця. Це згинання викликає також еквівалентне згинання середнього пальця та мізинця, як показано у формулі (2.1.4):

$$\theta_{\text{ПІФ(згин./розг.) Середній палець}} \approx \frac{1}{2}\,\theta_{\text{ПІФ (згин./розг.) Мізинець}} \qquad (2.1.4)$$

Також парний рух відбувається під час відведення та доведення мізинця та підмізинного пальця, як показано в:

$$\theta_{\text{ПІФ(відведення/доведення) Підмізинний палець}} \approx \theta_{\text{ПІФ (відв./дов.) Мізинець}} \qquad (2.1.5)$$

2.1.2.2. Відношення кутів

У моделі розроблено два види відношень кутів:

1) Під час згинання у ПІФ (проміжный інтрафаланзі) мізинця, як показано в:

$$\theta_{\text{ПІФ(згин./розг.) Підмізинний палець}} \approx \frac{7}{12}\,\theta_{\text{ПІФ (згин./розг.) Мізинець}}$$
$$\theta_{\text{ПІФ(згин./розг.) Підмізинний палець}} \approx \frac{2}{3}\,\theta_{\text{ПІФ (згин./розг.)Середній палець}} \quad (2.1.6)$$
$$\theta_{\text{ПІФ(згин./розг.) Підмізинний палець}} - \theta_{\text{ПІФ (згин./розг.)Середній палець}} < 60°$$
$$\theta_{\text{ПІФ(згин./розг.) Підмізинний палець}} - \theta_{\text{ПІФ (згин./розг.)Мізинець}} < 50°$$

2) Під час згинання у ПІФ вказівного пальця, яке виникає між середнім та вказівним пальцями, як показано в:

$$\theta_{\text{ПІФ(згин./розг.) Середній палець}} \approx \frac{1}{5}\,\theta_{\text{ПІФ (згин./розг.) Вказівний палець}} \quad (2.1.7)$$

2.2. Моделювання руки

2.2.1. Моделювання жестів за допомогою просторової моделі руки

Системи моделювання жестів мають багато застосувань у різних сферах людино-комп’ютерних інтерфейсів [92], таких як: медицина, розпізнавання жестів, тривимірне моделювання, доповнена реальність тощо. Для запропонованої технології та зокрема задачі вивчення жестів спеціально розроблений модуль моделювання жестів. Для зручного вивчення жесту він повинен бути доступним у вигляді тривимірної моделі руки, з можливістю обертання в усіх напрямках, щоб отримати найкращий огляд усіх можливих деталей жесту й представити його максимально точно. Існує декілька способів моделювання жестів за допомогою:

- створення бази відеозаписів із жестами [93];
- створення моделі скелету руки;
- створення реалістичної моделі руки.

Створення моделі руки — це перший крок у завданні моделювання мови жестів у рамках розробленої роботи.

У дисертації розроблено параметричну 3D-модель руки, яка використовує кусково-задані геометричні фігури для програмного створення основних форм руки. Використовується скелет із заданими обмеженнями та DOF відповідних суглобів. Зважаючи на те, що рука людини має відомі основні анатомічні складові, такі як пальці й долоня, геометрія руки гарно досліджена та може бути використана для створення її загальної форми.

Під час моделювання жестових переходів (а також під час моделювання динамічних жестів для певних дактилем) перетворення жесту в рамках роботи було розроблено у вигляді часового ряду скелетів (їх конфігурацій) моделі руки. Для кожного кадру t із послідовності кадрів, які зображують динамічний жест, позиція у просторі сцени кожного вузла скелету подана у вигляді трьох координат:

$$j_i(t) = [\, x_i(t) y_i(t) z_i(t)] \tag{2.2.1}$$

Таким чином, скелет у кадрі t поданий $3Nj$-вимірним вектором

$$s(t) = [\, x_1(t) y_1(t) z_1(t), \dots\ x_{Nj}(t) y_{Nj}(t) z_{Nj}(t)] \tag{2.2.2}$$

, де Nj є кількістю вузлів, які утворюють скелет та Nf є номером кадру у послідовності.

Остаточним відображенням є матриця розміру $Nf \times\ 3Nj$, де кожен рядок t є вектором $s(t)$

$$M = \begin{pmatrix} s(1) \\ \dots \\ s(N_f) \end{pmatrix} \tag{2.2.3}$$

2.2.2. Параметрична модель руки

В рамках роботи створена реалістична модель руки, яка складається зі скелету з заданими обмеженнями та DOF вузлів (суглобів) і побудованої на основі скелету високополігональної тривимірної моделі. Для побудови тривимірної моделі використано меші (mash), дискретизоване зображення об’єкта простими фігурами, такими як тетраедри й трикутники [94]. Меш [95] складається з вершин і трикутників (полігонів), специфічних і достатніх для визначення форми моделі руки.

У роботі розроблено параметричну модель руки, яка лягла в основу фотореалістичної тривимірної моделі. Вона складається з двох наборів параметрів:

- геометрії руки;
- кількості точок мешів.

Запропонована модель є сегментованою та поділена на частини на основі розташування кісток скелету руки: усічений еліпсоїд для дистальних фаланг і долоні, усічений еліптичний конус для решти кісток (рисунок 2.3) та еліпсоїд для долоні.

Зрізаний еліптичний конус має п’ять параметрів: базові піввісі (a, b), довжина (l) та висота (h). Використовуючи подібні трикутники, параметр l для усіченого еліптичного конуса з піввісями кістки $(b1)$ та наступна сполучена кістка $(b2)$, та висота $(h1)$ визначаються як:

$$\frac{l_1}{l-h_1} = \frac{b_1}{b_2} \tag{2.2.4}$$

$$l_1 = \frac{b_1 h_1}{b_1 - b_2} \tag{2.2.5}$$

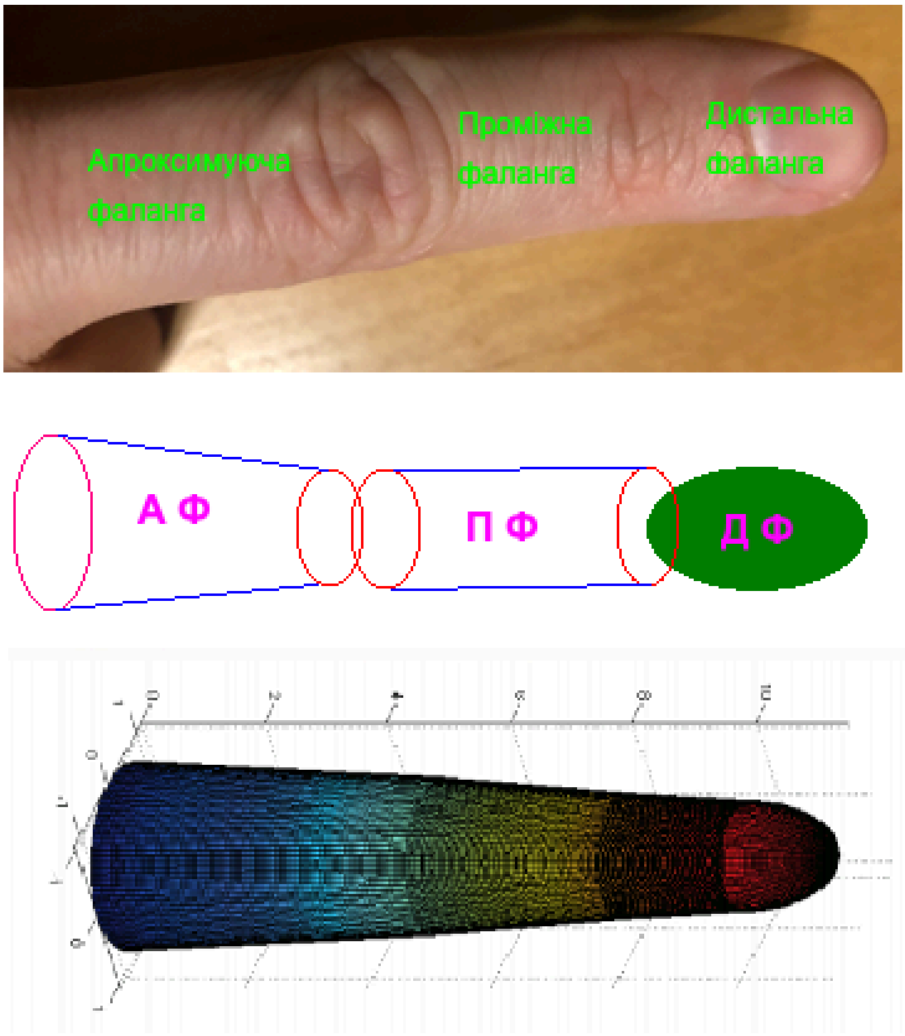


Рис. 2.3. (верх) Вказівний палець (середина) Дистальна (ДФ) та проміжна (ПФ) фаланги подані як усічені еліптичні конуси, апроксимуюча фаланга (АФ) — як еліпсоїд. (низ) Меш пальця

Параметричні рівняння для усіченого еліптичного конуса, в яких використовуються довжина кістки (h), ширина кістки (a), приблизна товщина кістки (b) та відповідна довжина конуса, визначаються наступним чином:

$$x = a\,\frac{l-h}{l}\cos v$$
$$y = b\,\frac{l-h}{l}\sin v$$
$$z = u \qquad (2.2.6)$$
$$v \in [0, 2\pi]$$
$$u \in [0, h]$$

Еліпсоїд має три параметри: його напіввісі (a, b, c). Враховуючи довжину (c) і ширину (a) дистального відділу фаланги та приблизної товщини кістки (b), його параметричними рівняннями є:

$$x = a \cos u \sin v$$
$$y = b \sin u \sin v$$
$$z = c \cos v \quad (2.2.7)$$
$$v \in [0, 2\pi]$$
$$u \in [0, \pi]$$

Кожен меш пальця було створено з заданими наборами кутів можливого розведення пальців. Для утворення долоні використано еліпсоїд [96].

2.2.3. Модель з текстурою

У роботі виконано реалістичне текстурування та освітлення моделі. У розробленій моделі руки нормальний вектор подається до кожної вершини, щоб бути правильно затемненим або освітленим. Нормалі в кутах трикутників мешу перпендикулярні до площини їх трикутника.

Рівномірне затемнення кожного трикутника створює ефект наявності гострих країв, застосованих до таких частин руки, як кінчики пальців, стики пальців та долоні. Потім нормалі інтерполюються через трикутники, щоб створити плавне затемнення до приблизних вигнутих поверхонь.

Шар із заданими нормалями для побудованої моделі руки зображено на рисунку 2.4.

Окрім нормалей, для відтворення більш точних деталей на поверхні руки було додано текстури. Текстура схожа на зображення, накладене та натягнуте на меш об’єкта. Для кожного трикутника зображення текстури, що відповідає тоншим деталям об'єкта, відображається на меші. Таким чином, замість імітації більш дрібних деталей на сітці, яке потребує тисяч трикутників, текстура

зменшує обчислення у візуалізації, використовуючи зображення для заміни дрібних деталей на сітці меншою кількістю трикутників.

Рис. 2.4. Шар із нормалями для поверхні руки

Рис. 2.5. Шар із текстурою та прозорістю моделі руки

Для деформованих об'єктів також важливо враховувати характер їх поверхні відносно кінематики, щоб правильно імітувати їх деформацію. Якість

сітки може впливати на точність та оптимізацію генерації та моделювання. Під час побудови реалістичної поверхні моделі руки було створено та використано зокрема шар із звичайною текстурою (див. рис. 2.5).

Рука є рухомим та деформованим предметом, тому для її зображення потрібно побудувати сітку, яка враховує її складну форму, пов’язану з DOF скелету, деформацію шкіри, пов’язану з її базовим скелетом, згинами руки та кольором шкіри. Відображення відмінностей в поверхні руки було досягнуто за допомогою шару з нерівностями текстури, що краще передає реалістичність структури її поверхні (рис. 2.6).

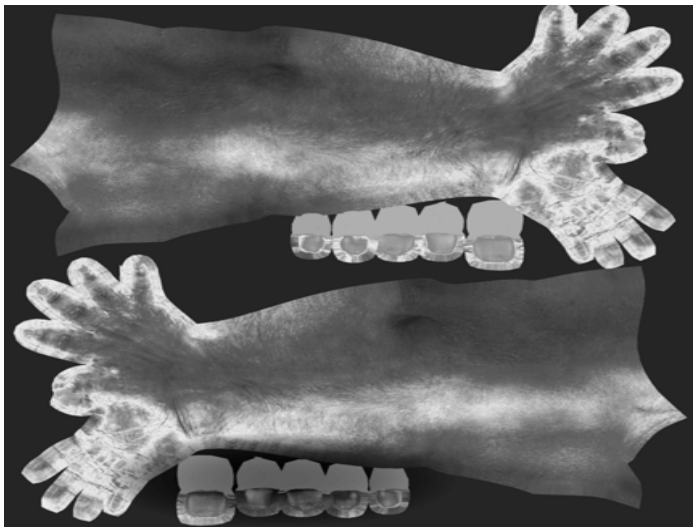


Рис. 2.6. Шар із нерівностями поверхні моделі руки

2.3. Адаптивність моделі руки

В рамках кросплатформеного аспекту виконаної роботи було запропоновано адаптивний підхід у моделюванні руки. Під час створення фотореалістичної 3D-моделі виконується оптимізація кількості полігонів моделі для швидшої візуалізації та адаптація моделі під платформу. Хоч 3D-модель руки потребує декілька тисяч полігонів, однак, із збільшенням кількості полігонів збільшується обчислювальна складність побудови моделі, що робить реальну візуалізацію в режимі реального часу повільнішою. В рамках роботи придатним рішенням визнано два підходи, а саме: використання меншої кількості полігонів та використання різних технік тривимірного моделювання для копіювання

дрібних деталей, які можуть бути відсутніми завдяки оптимізації від високополігональної до низькополігональної; модифікація анімацій жестів.

Тривимірна модель руки складається з понад 10 000 полігонів для забезпечення реалістичного зображення на високопродуктивних платформах і пристроях. Однак мобільні пристрої та платформи мають різний рівень обчислювальної потужності, також вони не завжди задовольняють вимогам зображувати руку з однаковим рівнем реалістичності через менший розмір екрана. Мікрозупинки та відставання анімацій на низькопродуктивній платформі можуть зробити запропоновану технологію непридатною до використання. Для розв'язання цієї проблеми кількість полігонів тривимірної моделі руки регулюється відповідно до обладнання, яке використовується операційною системою. При першому запуску технологія проводить детальне тестування наявного обладнання (виявляє платформу, рівень процесора та наявність дискретного відеоприскорювача), і, виходячи з цих характеристик, кількість полігонів рівномірно зменшується для усіх частин моделі руки, що робить її менш реалістичною, але збільшує обчислювальну продуктивність системи і плавність анімацій жестів.

Другий спосіб — коригування кроку анімації під час переходів між жестами. Зниження кроку анімації робить рух руки менш плавним, але усуває затримку кадрів та як наслідок відставання анімацій.

В рамках дисертаційної роботи запропоновано відтворення тривимірної моделі руки у веббраузері, що можливо завдяки використаним кросплатформеним засобам розробки моделі руки на платформах, які не забезпечують задовільного рівня продуктивності навіть після регулювання кількості полігонів моделі руки та зменшення кроку анімації.

2.4. Математична формалізація процесу подання даних для розпізнавання дактилем

Для запропонованого підходу до розпізнавання жестів етапи попередньої обробки включають набір технік і методів, які використовуються для локалізації

області руки та відокремлення її від фону, намагаючись зменшити шум у необроблених даних. Якість попередньої обробки даних має певні складнощі, і це може суттєво вплинути на продуктивність алгоритму. У дослідженні [97] показано, що сегментація руки може бути дуже складною залежно від сценарію використання, складного тла, і що неточний результат цього кроку спричиняє погане розпізнавання жестів.

Як правило, методи локалізації рук на основі RGB-зображень використовують розпізнавання кольору шкіри. Вони здебільшого дають правильний результат у простих контекстах, оскільки тон шкіри зазвичай відрізняється від кольору тла. Однак методи, засновані на RGB, залишаються вкрай чутливими до освітлення, індивідуальних відмінностей та тла. У праці [98] виконано огляд методів розпізнавання шкіри на основі RGB-зображень.

Поява датчиків глибини, що надають інший вимір у зображенні, дозволила подолати певні проблеми. Сегментація рук на основі карти глибини може бути здійснена шляхом встановлення граничних показників карти глибини, як продемонстровано у роботах [99, 100], та зростання або скорочення області, як показано у дослідженні [101]; іноді це супроводжується ітераційним етапом уточнення, як показано у роботі [102]. У таких випадках рука зазвичай вважається найближчим предметом до камери.

В рамках роботи, враховуючи запропонований підхід до розпізнавання жестів, немає необхідності у вручну створених ознаках та особливостях руки, оскільки визначення таких ознак та особливостей виконується глибокою нейронною мережею самостійно. Однак локалізація руки дозволяє полегшити задачу розпізнавання жесту, пришвидшивши пошук та зменшити можливість помилкових розпізнавань.

В рамках роботи запропоновано та розроблено модель, яка одночасно локалізаціє та ідентифікує жести, що разом визначається як «розпізнавання» жестів. Такий підхід дозволяє уникнути розробки окремо моделі для локалізації жестів (або руки) та для ідентифікації жесту у локалізованій області.

В рамках роботи було розглянуто декілька алгоритмів, наведених в роботах [103, 104] в якості базисного рівня у порівнянні з нейронною мережею. За їх допомогою була виконана бінарна класифікація кожного пікселя на зображенні для визначення приналежності пікселя до руки або тла. В цих алгоритмах було використано певні класичні ознаки та особливості зображень із жестами, засновуючись на даних із RGB-зображення. В рамках дисертаційної роботи розглянуто глибинні зображення, отримані з камер із сенсорами глибини в якості додаткового джерела вхідних даних та покращення результатів нейронної мережі. Однак такий підхід потребує додаткового обладнання, що накладає певні обмеження на кросплатформений аспект розроблених моделей.

2.4.1. Просторові дескриптори

Існує декілька дескрипторів, які обчислюються з зображення RGB або «глибинного» зображення та дозволяють отримати відповідні особливості з даних і таким чином розв’язати задачу розпізнавання жестів. Ці дескриптори можна розділити на три групи: дескриптори даних у матричній формі, дескриптори, засновані на формі руки, спеціально створені для аналізу жестів, і скелетні ознаки, обчислені за допомогою систем визначення позиції руки. В рамках роботи було вирішено зупинитися на дескрипторах у матричній формі, оскільки вони не потребують додаткових кроків (таких як визначення ключових точок або скелету руки на зображенні) та обладнання (у випадку «глибинних» зображень).

У роботі обрано підхід Bag Of Words (BoW) — «торба слів», який використовує ознаки SIFT, визначені з зображень жестів з відтінками сірого, разом з векторною квантифікацією, яка відображає ключові точки в об’єднану гістограму векторного квантування після етапу кластеризації алгоритмом K-середні. Було застосовано аналіз основних компонентів (PCA) для зменшення розмірності особливостей, в результаті чого було отримано новий дескриптор, стійкий до масштабу та обертання рук. Метод зображений на рисунку 2.7.

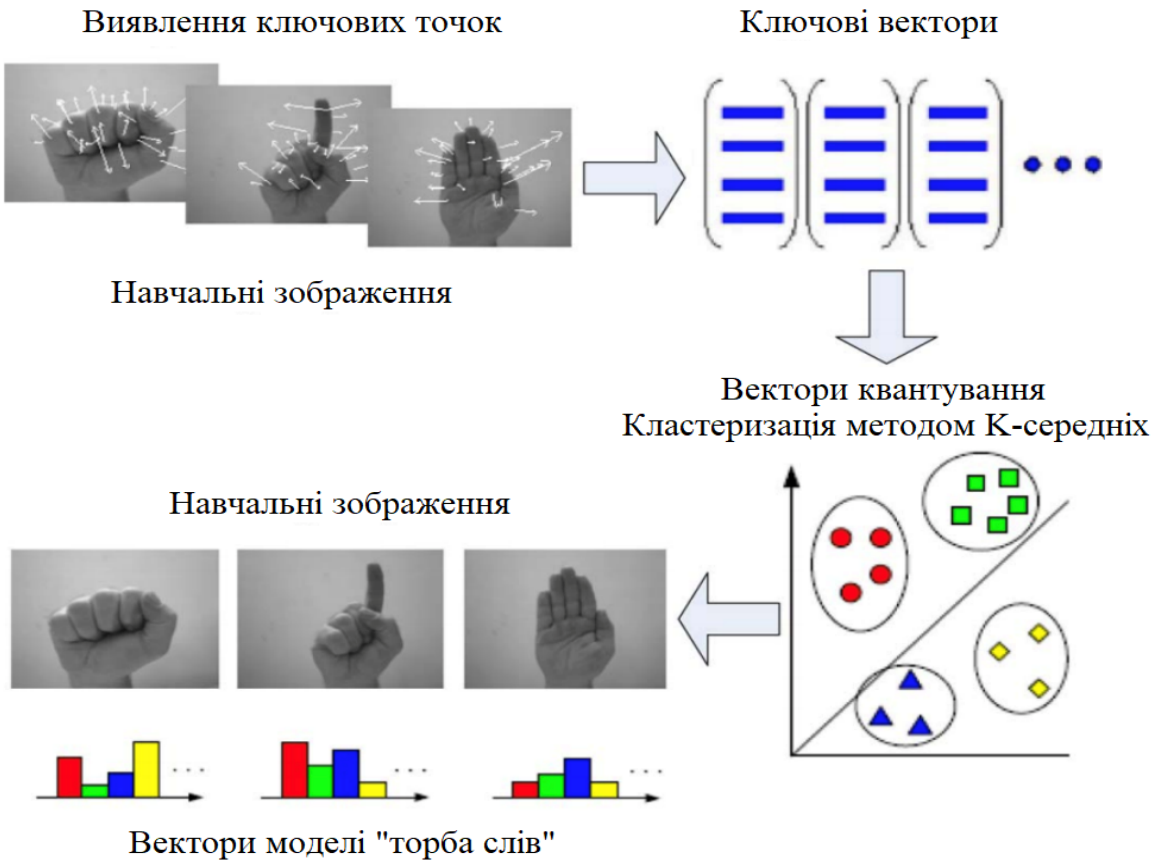


Рис. 2.7. Особливості SIFT визначаються з зображень у відтінках сірого та використовуються в якості векторів для створення словника BoW.

Серед переваг такого підходу:

- немає необхідності у зборі великої кількості навчальних зображень для побудови класифікатора жестів;
- додавання нового жесту не вимагає перенавчання всієї моделі.

Серед недоліків:

- якість розпізнавання жестів не зростає із збільшенням набору даних;
- є необхідність розробки додаткового кроку з визначення локалізації руки;
- чутливість до особливостей конкретної руки, особливостей зображення, середовища, в якому знаходиться рука.

2.4.2. Попередня обробка даних

При отриманні відеопослідовностей та окремих кадрів ті ж самі дактилеми можуть відрізнятися як за зовнішніми ознаками руки (задача розпізнавання таких відмінностей покладається на модель розпізнавання), так і за параметрами самих даних (розмір, якість, фокусна відстань, освітлення, тло, артефакти, розмиття і т.ін.). Для подальших обчислень в рамках обраної моделі розпізнавання була розроблена уніфікована процедура обробки даних для зведення їх до загального вигляду, як на етапі тренування моделі, так і на етапі розпізнавання.

Таким чином, сукупність даних (кадр з відео, окреме зображення), можна представити у вигляді:

$$D = \{aug\ (d_i)\}, i = \overline{1, n} \quad (2.4.1)$$

, де aug – функція аугментації даних, яка повертає вихідний тензор розмірності такої ж як і у вхідного, де n – кількість елементів у наборі даних, $n > 0$, причому d_i це тензор, що відображає дані з певного зображення.

Розмірність d_i - $M \times n \times 3$, де m – фіксована ширина зображення, n – фіксована висота зображення, 3 — три канали кольорового зображення, які передаються на вхід моделі під час тренування або розпізнавання.

Для подальшого використання у моделі розпізнавання необхідно:

- привести вибірку даних D до однакового розміру за висотою та шириною вхідних данних і кількістю каналів;
- аугментувати отримані зображення (у випадку етапу тренування моделі);
- здійснити видалення шумів та нормалізацію зображення.

2.4.3. Просторово-часове подання даних

Завдання динамічного розпізнавання жестів включає моделювання часового аспекту жестів на додаток до визначення особливостей. В роботі було розглянуто дві стратегії динамічного часового моделювання жестів:

- створення дескрипторів, які несуть просторову та часову інформацію;

• розпізнавання послідовностей просторових дескрипторів або зображень за допомогою просторо-часових класифікаторів.

В роботі було використано другий підхід. В якості вхідних даних нейронної мережі можна подавати як одиничні зображення жестів, так і послідовність зображень, у випадку якщо аналізується відеопотік. Таким чином було досягнуто декілька вдосконалень порівняно існуючими підходами розпізнавання жестів:

- аналіз кількох сусідніх зображень одночасно дозволяє навчати мережу, беручи до уваги часовий аспект, тобто динаміку зміни жестів на декількох зображеннях, для більшої стійкості до змін у такому динамічному об’єкті як рука;
- можна згладжувати невдале розпізнавання жесту на одному зображенні у послідовності з відеопотоку. Одне з зображень може бути із артефактами, надмірним чи недостатнім освітленням чи розмите, або об’єкт буде перекрито якоюсь перешкодою — всі ці проблеми можуть бути згладжені розпізнаванням жестів, використовуючи сусідні кадри у послідовності.

Для підвищення ефективності такого підходу було запропоновано та виконано використання часового плаваючого вікна (рисунок 2.8). Для цього запропоновано поділити вхідну послідовність на n підпослідовностей із мінімальною довжиною m, які взаємно перетинаються (на певну частину, від 10% до 50% довжини підпослідовності).

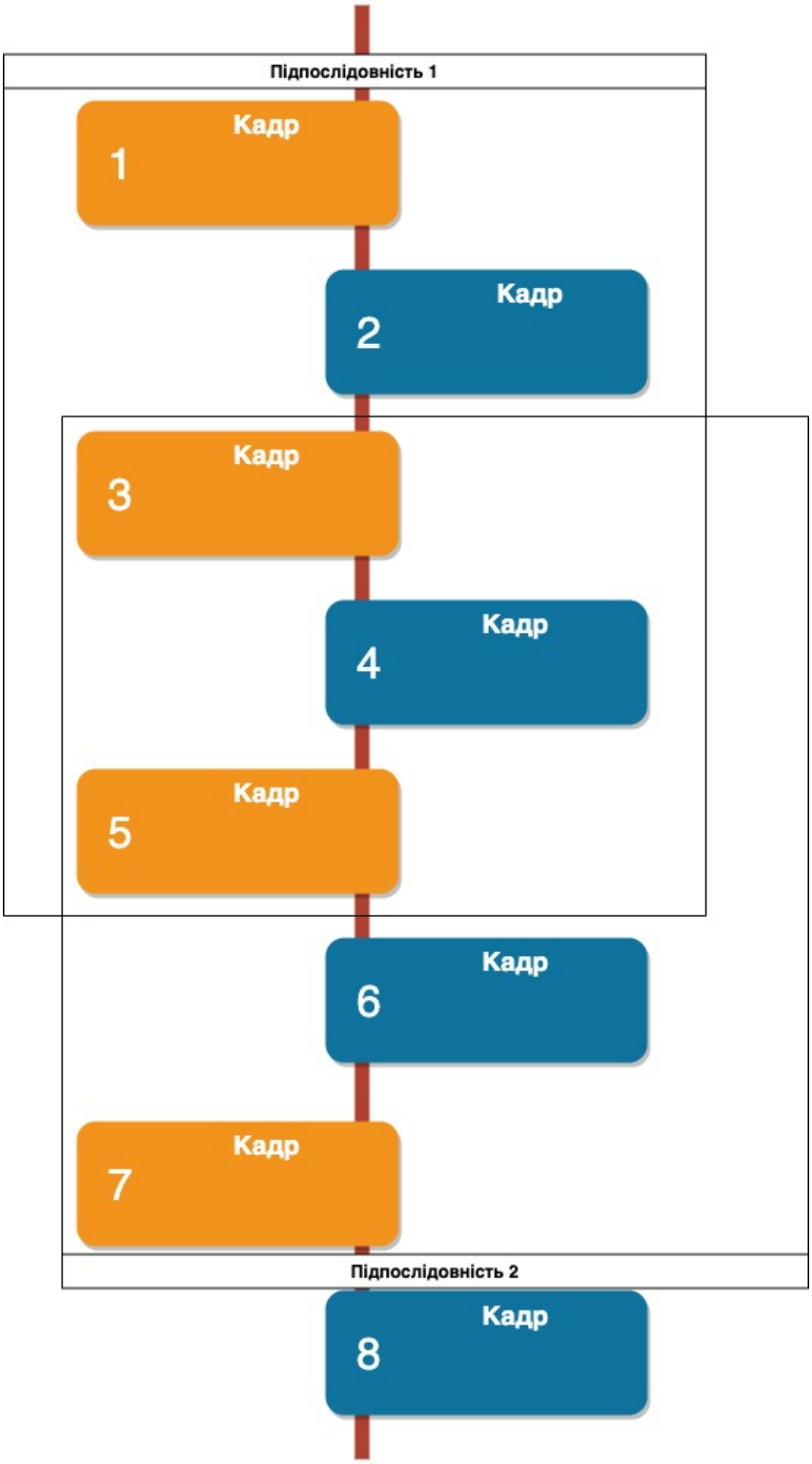


Рис. 2.8. Послідовність кадрів, розбита на дві підпослідовності розміром 5 кадрів, які перетинаються на 3 кадри.

Таким чином, з одного вхідного відеопотоку з жестом можна отримати n відеопотоків даного жесту меншого розміру.

Отже, дані можна подати у вигляді:

$$D = \{ (d_{i-k}, \dots, d_i, \dots, d_{i+k})\}, i = \overline{1, n-k} \quad (2.4.2)$$

, де k – кількість попередніх та наступних кадрів від поточного, з яких формується послідовність зображень (Рис. 2.9).

Рис. 2.9. Дві підпослідовності, створені з одного відеопотоку

2.5. Модель розпізнавання на основі глибокої згорткової нейронної мережі

В якості моделі розпізнавання запропоновано та реалізовано модель глибокої згорткової нейронної мережі. Серед переваг даного підходу можна виділити такі:

- виконується повна процедура із розпізнавання жестів, а саме: локалізація руки на кадрі та ідентифікація жесту;
- якість розпізнавання стійка до змін зовнішнього вигляду об’єкту, а також до змін у зовнішньому середовищі, зокрема тлі, якості зображення, фокусній відстані, освітленні тощо;
- не потребує додаткового обладнання окрім камери, яка видає триканальні кольорові зображення;
- не потребує калібрування;
- якість розпізнавання зростає зі збільшенням набору тренувальних даних;
- не потребує специфічної попередньої обробки даних (окрім базових операцій, таких як зміна розміру, видалення шуму і нормалізація).

Серед основних недоліків даного підходу слід відзначити необхідність збору репрезентативного тренувального набору зображень жестів, що вимагає значної роботи.

Архітектура запропонованої в роботі нейронної мережі виконана на основі архітектури MobileNetv2, що забезпечує високу швидкодію, зокрема на мобільних платформах, через спеціальні шари та методики, використані у нейромережі (рисунок 2.10).

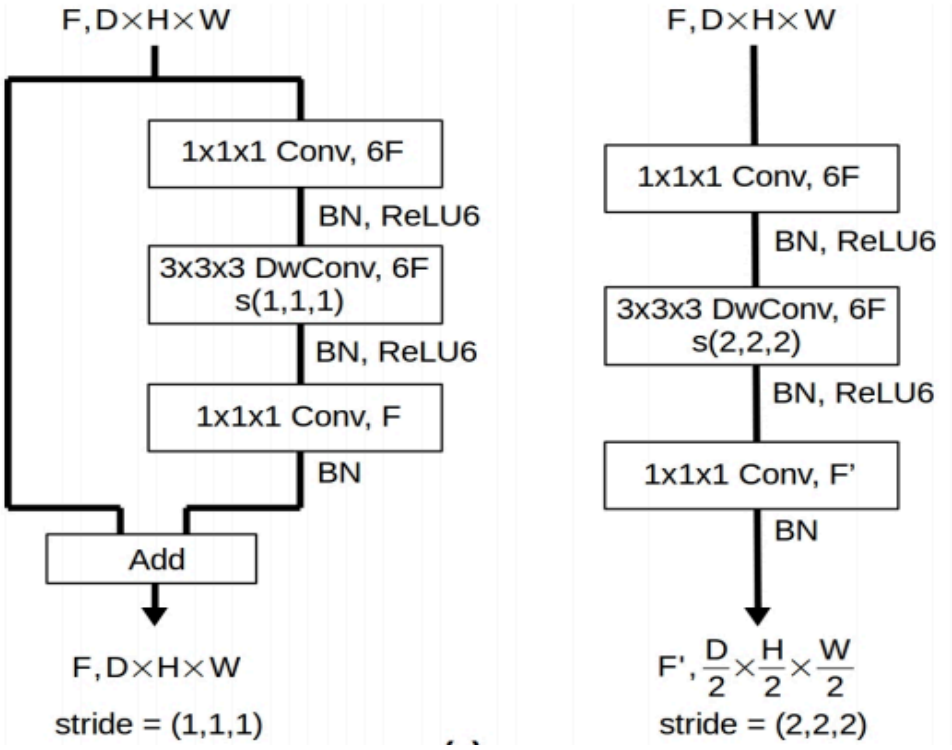


Рис. 2.10. Блок MobileNetv2

2.5.1. Математична формалізація структури нейронної мережі

У запропонованій нейронній мережі, стандартні згорткові шари можна подати у вигляді:

$$G_{k,l,n} = \sum_{i,j,m} K_{i,j,m,n} * F_{k+i-1,l+j-1,m} \quad (2.5.1)$$

, де входом згорткового шару є відображення ознак F розмірності $Df \times Df \times M$ і виходом є відображення ознак G розмірності $Df \times Df \times N$, де Df – це просторова ширина та висота, M – кількість вхідних каналів, N – кількість вихідних каналів.

Стандартний згортковий шар задається параметром згорткового ядра K розміності $D_K \times D_K \times M \times N$, де D_K є просторовою розмірністю ядра, та M – кількість вхідних каналів, N – кількість вихідних каналів, визначених раніше. Обчислювальна складність такого шару:

$$D_K \times D_K \times M \times N \times D_F \times D_F \quad (2.5.2)$$

У виконаній роботі в архітектурі нейромережі використано згорткові шари у вигляді:

$$G_{k,l,n} = \sum_{i,j,} \widehat{K}_{i,j,m,} * F_{k+i-1,l+j-1,m} \quad (2.5.3)$$

, де $\widehat{K}$ це поглибинне згорткове ядро розмірності $D_K \times D_K \times M$ де, m -ий фільтр у $\widehat{K}$ застосовується до m -ого каналу із F щоб видати m -ий канал відфільтрованого вихідного відображення ознак $\hat{G}$.

Обчислювальна складність такого шару

$$D_K \times D_K \times M \times D_F \times D_F \quad (2.5.4)$$

Input	Operator	t	c	n	s
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d 1x1	-	1280	1	1
$7^2 \times 1280$	avgpool 7x7	-	-	1	-
$1 \times 1 \times 1280$	conv2d 1x1	-	k	-	

Рис. 2.11. Загальна архітектура мережі

2.5.2. Використання даних у просторово-часовому вимірі

Враховуючи динамічну природу деяких жестів та рухливість такого об’єкту як рука, у роботі виконано узагальнення згорткової нейронної мережі із 2D-зображень на 3D-відеопослідовності.

Виходячи із запропонованого просторово-часового подання даних, поданого у формулі 2.4.2, на вхід нейронній мережі надається підпослідовність зображень із жестами з вхідного відеопотоку.

Однією з основних складностей є асиметрія відеопотоку щодо часу та простору. На відміну від зображень, які можуть бути обрізані та перероблені у фіксований розмір, відео можуть відрізнятися у часовому вимірі. В дослідженні здійснено виділення підпослідовності зображень фіксованої довжини з відео, та подальша робота ведеться вже з ними.

2.5.3. Удосконалення моделі нейронної мережі тривимірними згортками для використання даних у просторово-часовому вимірі

В роботі виконано адаптацію нейронної мережі до просторово-часового формату вхідних даних. Для кращого використання просторово-часових ознак вхідних даних було запропоновано вдосконалити архітектуру згорткової нейромережі тривимірними згортками (формула 2.5.3). Таким чином, тривимірна згорткова нейронна мережа здатна виконувати операцію згортання не лише у просторі зображення, але й у часі.

Загальна схема процесу зображена на рис. 2.12.

Мережа MobileNetv2 [159] є іншою ресурсно-ефективною 2D-архітектурою. Вона побудована на основі MobileNet, використовує відокремлені глибинні згортки, але вносить два нових компоненти:

1) лінійні звуження між шарами;

2) короткі зв’язки між звуженнями.

Ідея першого компоненту полягає в одночасному збереженні малого розміру моделі через зменшення кількості каналів та витягнення якомога більшого обсягу інформації, використовуючи глибинні згортки після декомпресії даних. Цей згортковий модуль дозволяє зменшити використання пам’яті протягом проходу нейронної мережі. Другий модуль дозволяє пришвидшити тренування та конструювати глибші моделі, такі як ResNet [162].

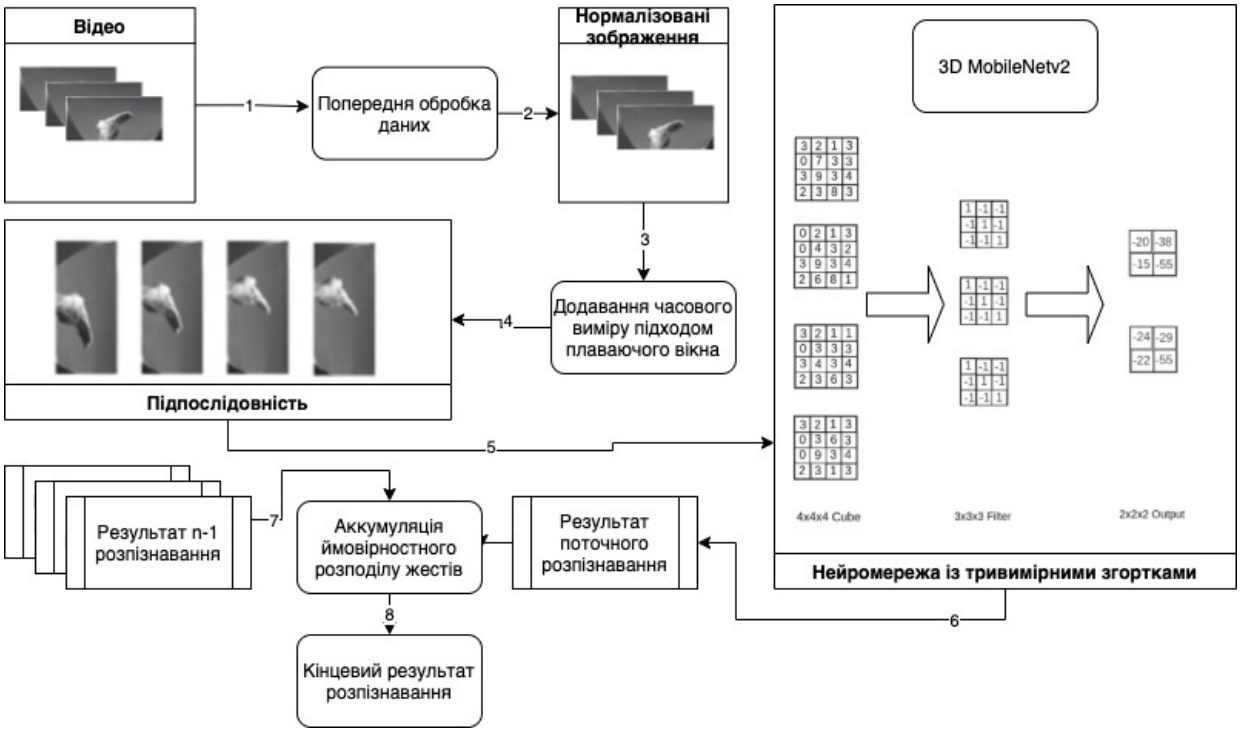


Рис. 2.12. Загальна схема процесу розпізнавання разом із моделлю розпізнавання тривимірними згортками

Архітектура вдосконаленої нейронної мережі зображена на рис. 2.13

Layer / Stride	Repeat	Output size
Input clip		3x16x112x112
Conv(3x3x3)/s(1,2,2)	1	32x16x56x56
Block/s(1x1x1)	1	16x16x56x56
Block/s(2x2x2)	2	24x8x28x28
Block/s(2x2x2)	3	32x4x14x14
Block/s(2x2x2)	4	64x2x7x7
Block/s(1x1x1)	3	96x2x7x7
Block/s(2x2x2)	3	160x1x4x4
Block/s(1x1x1)	1	320x1x4x4
Conv(1x1x1)/s(1,1,1)	1	1280x1x4x4
AvgPool/s(1,1,1)	1	1024x1x1x1
Linear	1	*NumCls*

Рис. 2.13. Архітектура MobileNetV2, модифікована, з тривимірними згортками

В роботі у нейронній мережі використовується функція витрат, яку можна подати у вигляді:

$$\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} [(x_i - \hat{x_i})^2 + (y_i - \hat{y}_i)^2]$$

$$+ \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} \left[\left(\sqrt{w}_i - \sqrt{\widehat{w_i}} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] \quad (2.5.5)$$

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} \left(C_i - \widehat{C_i} \right)^2 + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{nooobj} \left(C_i - \widehat{C_i} \right)^2$$

$$+ \sum_{i=0}^{S^2} 1_{ij}^{obj} \sum_{c \in classes} (p_i(c) - \hat{p}_i(c))^2$$

де,

- xi, yi є положенням центроїд прямокутника з жестом
- $wi, hi,$ є шириною та висотою прямокутника з жестом
- Ci , є оцінкою впевненості наявності об’єкта на зображенні
- $pi(c)$, функція витрат класифікації

Всі функції втрат є середньоквадратичними помилками, крім функції втрат класифікації, яка використовує перехресну ентропію. Це дозволяє одночасно оптимізувати пошук позиції жесту та його клас. Таким чином, немає необхідності в окремому кроці локалізації руки.

2.5.3.1. Відображення результатів розпізнавання з підпослідовностей

У роботі запропонований, як подальший розвиток технологій розпізнавання, механізм згладжування аномальних результатів розпізнавання завдяки реалізованому способу акумуляції ймовірностей з попередніх підпослідовностей.

Оскільки підпослідовності формуються за принципом плаваючого вікна з перетинами, результати розпізнавання мають плавно зменшувати максимальну ймовірність поточного жесту, і з плином часу (тобто, підпослідовностей) має зростати максимальна ймовірність наступного жесту.

Модель, запропонована та реалізована в роботі, полягає в акумулюванні прогнозів з попередніх підпослідовностей та оновленні поточного результату розпізнавання, лише коли акумульована сума ймовірностей перетинає певний поріг $thresh$, що можна подати в такому вигляді:

$$\sum_{t=t-n}^{t+n} \sum_{i=t-k}^{t+k} p_i > thresh, \qquad (2.5.6)$$

де:

- p_i – ймовірність певного жесту на кадрі,
- i – номер кадру на поточній підпослідовності,
- t – номер поточної підпослідовності,
- k – розмір підпослідовності в обох напрямках,
- n – кількість акумульованих підпослідовностей.

2.5.4. Тренування нейронної мережі

Завдання нейронної мережі f полягає у використанні набору розмічених даних, з метою мінімізувати різниці між її вихідним $y\hat{}$ та міткою y заданого вводу x, через функцію витрат та процедуру оптимізації. Для виконання цього завдання тренується модель, шляхом оновивлення параметрів $f(\theta)$, щоб отримати найкращу функцію наближення $y\hat{} = f(x, \theta)$. У дисертаційній роботі

оптимізація для підготовки мережі проводиться за алгоритмом зворотного розповсюдження. Цей алгоритм працює у два етапи:

1. Поширення. Коли вхідний вектор надходить до мережі, то розповсюджується вперед пошарово, доки не досягне вихідного шару. Потім вихід мережі порівнюється з відповідною сигнатурою за допомогою функції втрат, і обчислюється значення помилки. Потім ця помилка поширюється назад, починаючи з вихідного шару, до тих пір, поки кожен нейрон не матиме асоційованого значення помилки, яке приблизно відображає його внесок у помилку.

2. Оновлення значень. Зворотнє розповсюдження використовує ці значення помилок для обчислення градієнта функції втрат. Цей градієнт піддається дії методу оптимізації, який використовує його для оновлення значень, щоб мінімізувати функцію втрат.

Крок зворотного розповсюдження повторюється, поки мережа кілька разів не обробить весь набір даних. Повний прохід через весь набір даних називається епохою. Отже, до кінця першої епохи модель обробить кожен зразок у навчальному наборі один раз.

Під час тренування для пришвидшення процесу збіжності нейромережі було використано алгоритм Stochastic Gradient Descent (SGD), який оптимізує градієнтний спуск і мінімізує функцію втрат під час мережевих тренувань. Він називається стохастичним, оскільки в ньому під час ітерації замість аналізу всіх даних виконується випадковий вибір декількох прикладів із вибірки. Коефіцієнт тренування λ було задано для встановлення оптимального часу тренування нейромережі. Цей параметр було зменшено з метою більш повільного, але стабільного наближення до локального мінімума. Параметр B (батч) визначає кількість навчальних зразків, які будуть поширюватися через мережу при кожній ітерації. Використання батчів у SGD дозволяє зменшити дисперсію оновлень градієнта (застосовуючи середнє значення градієнтів у батчі) та прискорити оптимізацію моделі.

Відмітимо, що в процесі досліджень всі ці гіперпараметри було підібрано оптимальним чином під час тренування моделі.

2.5.5. Проблема перенавчання та трансферу навчання

Мета процесу класифікації — це створити класифікатор, який правильно працює на нових, раніше небачених вхідних даних або узагальнених масивах даних.

Зазвичай набір даних складається з двох наборів, що не перетинаються. Перший, який називається навчальним набором, складається з даних, за якими модель навчається. Другий, тестовий набір, складається з даних, які алгоритм не бачить під час навчального етапу. Класифікатор повинен мінімізувати показники помилок між його результатами та еталонними даними шляхом процедури оптимізації. Цей показник помилки називається помилкою навчання при обчисленні на навчальному наборі та помилкою узагальнення або помилкою тесту при обчисленні на тестовому наборі. Ефективність алгоритму навчання визначає його здатність зменшувати помилку навчання та зменшувати різницю між навчальною та тестовою помилкою, що називається розривом узагальнення.

Потенціал моделі глибокого навчання — це її здатність адаптуватись до певної задачі. Два основні гіперпараметри визначають місткість моделі: її глибину та ширину. Модель з низькою місткістю може не відповідати складності навчального набору. Модель з високою місткістю може мати надто багато параметрів, внаслідок чого вивчатиме конкретні властивості навчального набору замість їх узагальнення. Модель з високою ємністю може вирішувати складні завдання, але їй потрібно більше даних, щоб уникнути перенавчання. На рис. 2.14 показані взаємозв'язки між місткістю моделі та показниками помилок. Знайти найкращу мережеву структуру, яка б ідеально узагальнювала навчальний набір, надто складно, оскільки існує значна кількість різних рішень.

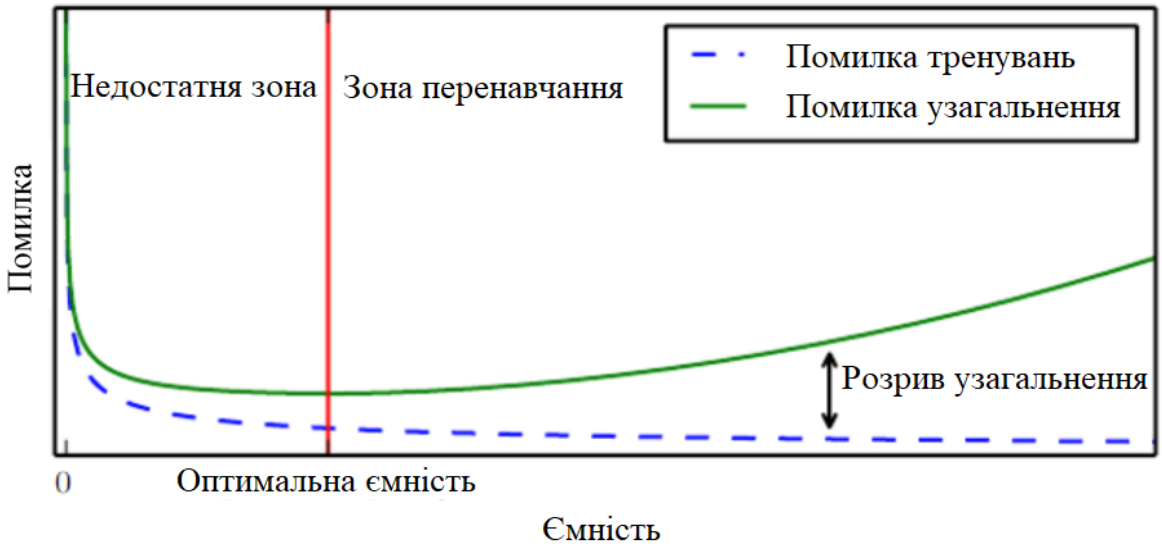


Рис. 2.14. Типова кореляція між місткістю моделі та показниками помилок. Зліва від оптимальної ємності ми могли б збільшити потенціал, щоб знайти краще узагальнення навчального набору. Цей стан називається недостатнім.

Для завдань класифікації зображень існують великі набори даних, такі як набір даних Open Images [107], який складається з 10 000 000 розмічених зображень.

У модулі розпізнавання жестів запропонованої технології було використано такі підходи регуляризації з метою уникнення перенавчання моделі:

• використання меншої архітектури моделі;

• використання шарів випадання (рис. 2.15). Випадання випадковим чином змушує вузли в нейронній мережі «випадати», встановлюючи їх значення таким, що дорівнює нулю, через що мережа змушена покладатися на інші особливості. Це призводить до більш узагальненого подання даних;

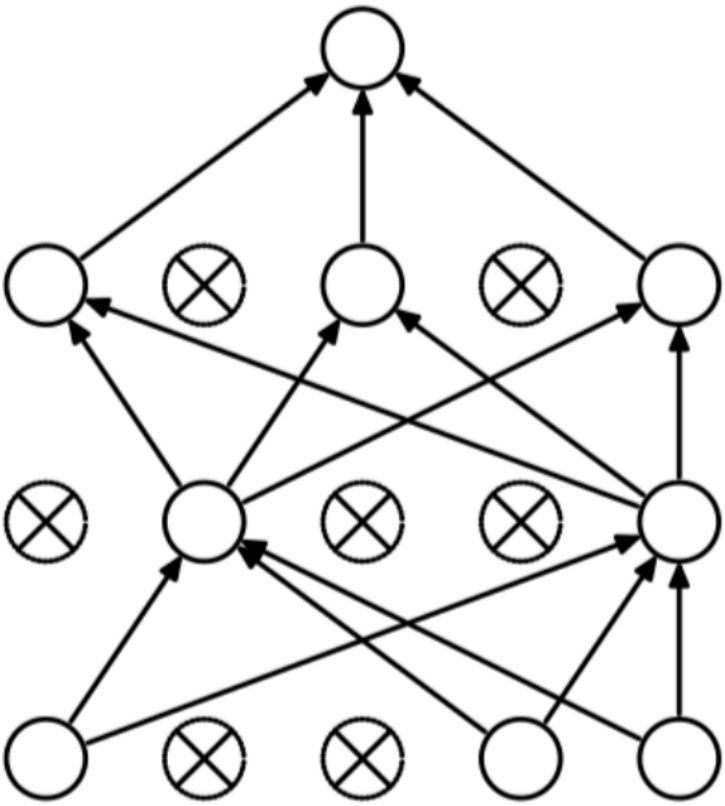


Рис. 2.15. Шар випадання, який обнулює задану частку нейронів під час кожної ітерації

• використання більшого набору даних. Оскільки глибоким мережам потрібна велика кількість даних при навчанні для досягнення високих результатів, було зібрано великий набір даних української дактильної абетки та розширено автоматичними змінами (аугментаціями) в зображеннях.

Одним з можливих підходів тренування моделі також було розглянуто трансфер (передачу) навчання від інших, вже натренованих моделей [108].

2.6. Висновки до Розділу 2

Розробка систем і алгоритмів, що реалізують задачу аналізу (розпізнавання) та моделювання (синтезу), вимагають широкого розуміння процесу чи явища, що породжує цю інформацію. Дактилеми (жести руки) можна розглядати як зовнішні вияви зміни скелетно-м’язової структури, виражені у вигляді відповідних деформацій поверхні руки, зміни положення пальців та долоні у відповідності до обмежень, які накладені на скелет. Модель, що описує скелет руки, має бути

обрана виходячи із критеріїв простоти та реалістичної поведінки моделі руки. Для того, щоб пов’язати зовнішні і внутрішні прояви, необхідна тривимірна модель руки, яка визначає величину деформацій поверхні руки як функцію від деформацій скелетної структури.

Доповненням до моделі моделювання руки є модель розпізнавання жестів, що пов’язує особливості зображень з дактилемами на різних рівнях із конкретним жестом з української дактильної абетки. Така модель зазвичай складається з двох концептуальних кроків: визначення особливих точок та особливостей на зображенні об’єкта (руки, що зображує певну дактилему), пошук цих точок та особливостей з метою визначення дактилеми, яка зображена на вхідному зображенні. Перший крок у класичних підходах комп’ютерного зору зазвичай задавався людиною, на відміну від більш сучасних підходів глибоких нейронних мереж, які самостійно, під час тренування, визначають, які особливості на зображенні є важливими для визначення класу об’єкту. У роботі розглянуто обидва підходи та використано моделі глибоких нейронних мереж.

Моделі згорткових нейронних мереж якісно визначають за допомогою операції згортання важливі особливості на різних зображеннях одного й того самого об’єкту, що є особливо важливим для моделі руки, яка є об’єктом, здатним до суттєвих змін, та значно відрізняється залежно від ракурсу, в якому відображається. Згорткові нейронні мережі також стійкі (інваріантні) до змін масштабу, завдяки чому відпадає необхідність у фіксації оптимальної фокусної відстані. Також вони стійкі до деформацій зображення, шуму, різниці в освітленні, різнорідності тла та середовища.

Модель розпізнавання також слугує засобом верифікації якості моделі моделювання в єдиному процесі навчання людини через зображення жесту та верифікації успішного навчання через розпізнавання змодельованого жесту, який тепер зображується людиною.

Виходячи із вищенаведеного, у розділі 2 були отримані такі результати:

- запропонована модель скелету руки та обмеження DOF, що пов’язує деформації поверхні руки із зміною положення скелету;

- запропоновано підхід щодо досягнення адаптивності моделі руки за допомогою зміни кількості полігонів та кроку анімації, з якою відбувається перехід з одного жесту в інший;
- запропоновано метод попередньої обробки зображень для моделі розпізнавання дактилем, враховуючи просторово-часові особливості інформації відеоряду та особливості окремого зображення;
- розглянуто та використано методи нейронних згорткових мереж та їх шарів, особливості навчання та важливі етапи при зборі та поданні необхідного навчального та тестувального наборів даних.

Розділ 3. Кросплатформена інформаційна технологія моделювання та розпізнавання жестів за допомогою тривимірних згорток

У третьому розділі запропонована кросплатформена технологія моделювання та розпізнавання жестів української дактильної абетки. Розглянуто головні складові та етапи технології, такі як початковий етап збору даних, обробка даних, організація структури жестів для моделювання, підготовка даних для моделі розпізнавання жестів, попередній етап тренування та оцінювання якості моделі. Наведено основні кроки технології розробки моделі розпізнавання жестів.

3.1. Інфологічна модель

Запропонована інформаційна технологія складається з двох головних компонентів, кожен з яких, в свою чергу, складається з певних етапів. Загалом у процесі використання всі компоненти складаються в єдину систему подання вихідної та аналізу вхідної інформації.

В узагальненому вигляді перший крок моделювання послідовності жестів (слів) складається з таких етапів:

- *аналіз платформи*, вибір відповідної полігональності моделі руки та кроку анімації переходу між жестами;
- *завантаження відповідного набору жестів для заданої мови* (у випадку запропонованої технології — української дактильної абетки) з бази даних у вигляді конфігураційних файлів;
- *побудова тривимірної сцени відповідно до першого жесту* у заданій користувачем послідовності (слові або реченні);
- *анімований перехід між жестами* (дактилемами) у словах із визначеним кроком анімації.

Другий крок — розпізнавання жестів складається з 2 частин (тренування та розпізнавання), які поділяються на етапи.

1.Тренування:

- *збір набору даних* відповідно до заданих дактилем;
- *обробка даних* (нормалізація, зменшення шуму);
- *аугментація* (розширення) зібраного набору даних за допомогою таких підходів як гаусівський шум, афінні перетворення, обрізання та здвиг, розгортання та спотворення перспективи;
- *розподілення* на тренувальний та тестовий набір даних;
- задання *набору конфігурацій* для тренування декількох нейромереж із різною архітектурою для визначення оптимальної за точністю на тестовому наборі даних;
- *тренування та тестування* згорткової нейромережі;
- *вибір оптимальної архітектури* згорткової нейромережі.

2. Розпізнавання (прогнозування):

- *виділення послідовності зображень* із вхідного відеоряду;
- *обробка даних;*
- *отримання результату* розпізнавання згорткової нейромережі;
- *визначення дактилеми* з урахуванням попередніх результатів.

На узагальненій структурній схемі ІТ (рис. 3.1) подані укрупнені етапи технології, а також наведені операції на даних етапах та вхідні й вихідні дані. На основі даної схеми пропонується деталізувати кожен з етапів ІТ.

Основою технології є композиція трьох кросплатформених модулів: тривимірна модель руки (яка реалізована за допомогою кросплатформеної системи Unity3D), інтерфейс користувача (реалізований також з Unity3D) та модуль розпізнавання жестів (реалізований з кросплатформеною бібліотекою Tensorflow) Основна функціональність реалізована на C# і Python і працює на настільних ОС (MacOS, Linux, Windows) та на мобільних ОС (Android, iOS).

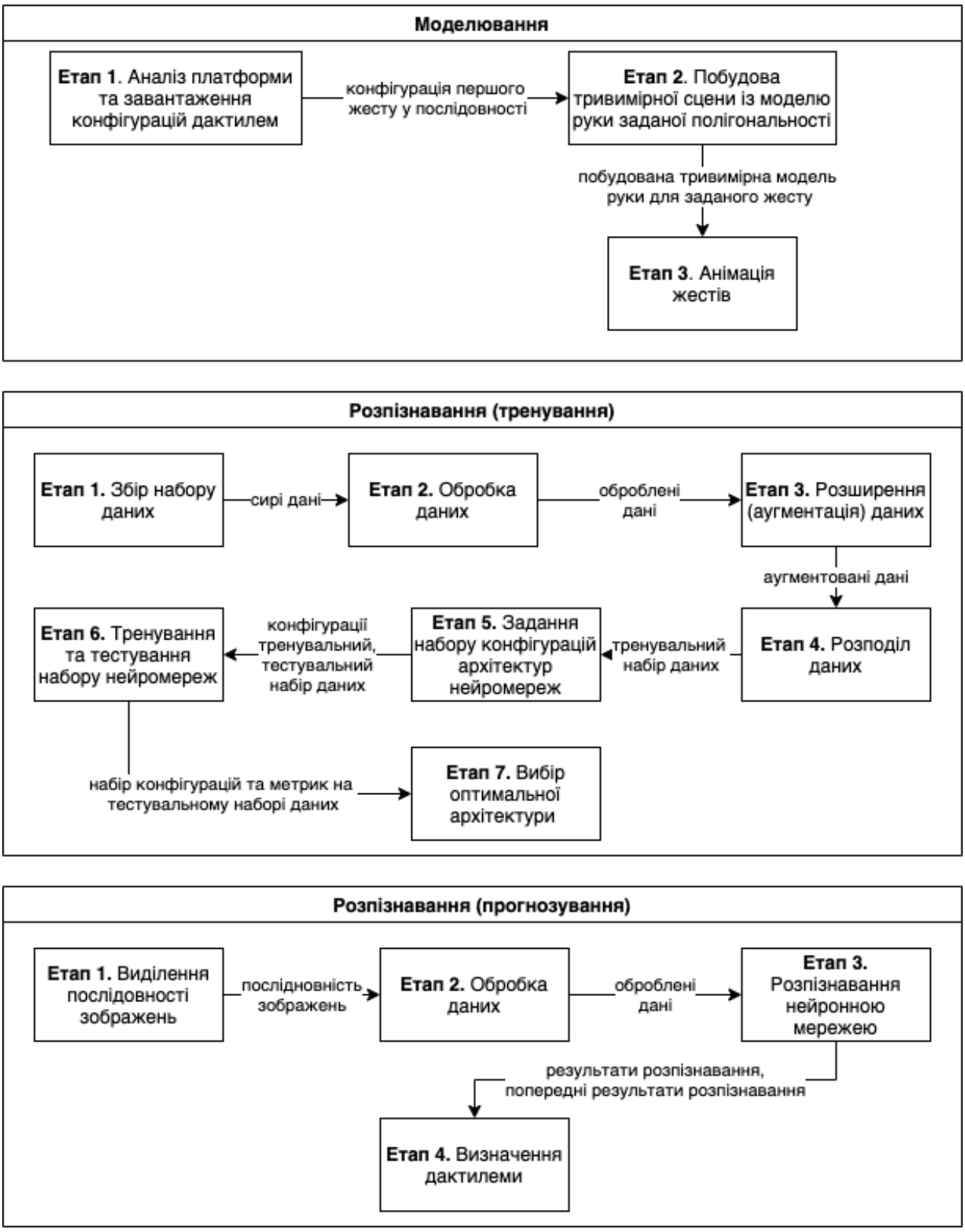


Рис. 3.1. Узагальнена структурна схема ІТ

Діаграма архітектури технології (рис. 3.2) демонструє взаємодію основних компонентів запропонованої системи. Жести для моделювання жестів зберігаються у визначеному форматі (YAML) в базі даних та використовуються механізмом моделювання жестів для встановлення конфігурації просторової

тривимірної моделі руки з використанням заданих параметрів жесту з відповідного запису бази даних. Модуль моделювання жестів працює над базою даних жестів і є частиною програми, яка складається з компоненту, який безпосередньо здійснює моделювання жестів, та компонентів інтерфейсу, які розробляються в рамках Unity3D за допомогою мови програмування C#. Віртуальна модель руки задається скелетом і набором параметрів та їх обмеженнями для кожного скелетного з'єднання. Модуль розпізнавання жестів реалізований фреймворком Tensorflow з використанням мови програмування Python. Модуль розпізнавання жестів працює незалежно від модуля моделювання жестів і використовує базу даних жестів. Основними компонентами модуля розпізнавання жестів є модель, яка використовується для розпізнавання жестів, і обгортка, яка перетворює дані з камери у відповідний формат для моделі.

Модуль моделі руки є кросплатформеним та забезпечує надходження вхідних даних про модель руки для модуля розпізнавання жестів. Модуль візуалізації руки отримує ці дані та технічні характеристики жестів від модуля зберігання жестів і надає високополігональну модель руки. Модуль вивчення жестів та модуль модифікації жестів реалізовані за допомогою платформи Unity3D, обидва беруть в якості вхідних результатів візуалізацію моделі руки. Модуль модифікації жестів забезпечує оновлені специфікації жестів і передає їх у сховище жестів.

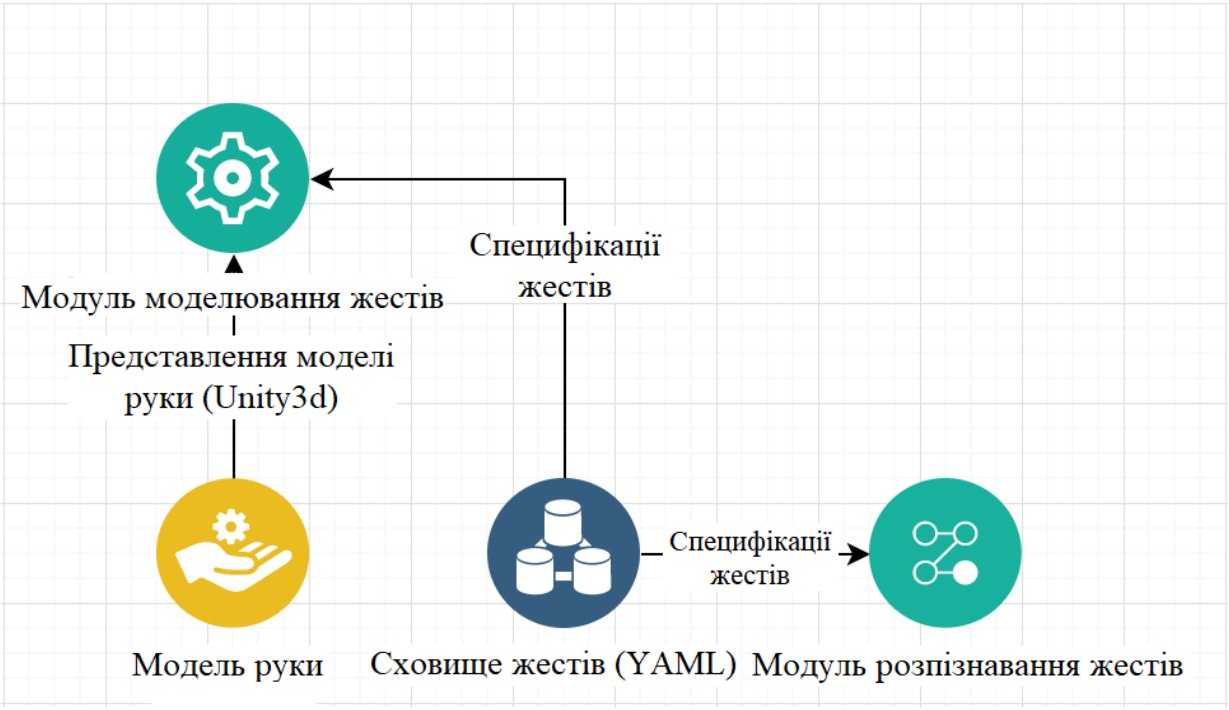


Рис. 3.2. Діаграма компонентів IT

3.2. Моделювання

3.2.1. Аналіз та завантаження жестових конфігурацій

Даний крок передбачає завантаження конфігурацій дактилем української абетки, які зберігаються у базі даних у форматі YAML, у память. Жест зберігається в такому вигляді:

$$\%YAML\colon 1.0\ rotation\colon [\ -9.5845758914947510e-02, 8.3791027449819921e-09,$$
$$-9.9539619684219360e-01, -7.5690525770187378e-01,$$
$$-6.4944797754287720e-01, 7.2881683707237244e-02,$$
$$-6.4645802974700928e-01, 7.6040601730346680e-01, 6.2246840447187424e-02\ ]$$
$$hand\_joints\colon\ finger1joint1\colon [\ 0.,0.,0.\ ]\ \ finger1joint2\colon [\ 0.,0.,0.\ ]\ \ \ldots$$

Також проводиться аналіз наявної платформи, та за умови, що платформа мобільна або не має дискретного відеомодуля, обирається для завантаження модель руки з меншою полігональністю та виставляється менший крок анімації,

тобто відбувається менше проміжних переходів між двома жестами у послідовності.

- Вхід: інформація про платформу
- Проводиться аналіз платформи, відбувається завантаження конфігураційних файлів жестів української дактильної абетки
- Вихід: завантажені жести, адаптована конфігурація адаптованої моделі руки

Рис. 3.3. Аналіз та завантаження жестових конфігурацій

3.2.2. Побудова тривимірної сцени з моделлю руки заданої полігональності

На даному кроці відбувається завантаження моделі руки заданої полігональності з попереднього кроку. З послідовності жестів, які необхідно відтворити, та переходів між ними обирається перший жест. На скелет завантаженої тривимірної моделі руки накладається набір обмежень ступенів свободи відповідно до обраного підходу. Завантажується конфігурація для скелету першого жесту.

- Вхід: послідовність жестів які, необхідно відтворити, конфігурація моделі руки
- Проводиться побудова скелету руки з заданими обмеженнями та в залежності від конфігурації першого жесту з відтворюваної послідовності завантажується тривимірна модель
- Вихід: побудована тривимірна модель руки у стані першого жесту з відтворюваної послідовності

Рис. 3.4. Побудова тривимірної сцени з моделлю руки заданої полігональності

3.2.3. Анімація жестів

Анімація жестів конфігурується кроком анімації, тобто кількістю проміжних інтерпольованих станів скелету моделі руки під час переходу від одного жесту до іншого. Інтерполяція відбувається за допомогою движку Unity3D з урахуванням обмежень на ступені свободи скелету руки.

- Вхід: побудована тривимірна модель руки в стані першого жесту з відтворюваної послідовності, послідовність жестів, які необхідно відтворити, конфігурація моделі руки
- Проводиться інтерполяція станів скелету між поточним та наступним жестом у послідовності дактилем, відбувається побудова проміжних станів з урахуванням кроку анімації та обмежень ступенів свободи різних елементів скелету руки.
- Вихід: побудована тривимірна модель руки у стані останнього жесту з відтворюваної послідовності

Рис. 3.5. Анімація жестів

3.3. Розпізнавання (тренування)

3.3.1. Збір набору даних

Збір даних складається з таких кроків:

- отримання відео (запис відео) або зображення;
    - у випадку відео — сегментація відео на окремі зображення;
- обробка даних;
- збереження даних для подальшого використання у тренуванні або тестування нейромережі.

3.3.2. Обробка даних

Даний крок полягає в обробці даних з метою уніфікації вхідних даних для нейромережі та приведення їх до вигляду, який дозволить уникнути

перенавчання нейромережі на певних особливостях набору даних, які можуть виникнути під час їх збору.

Обробка даних скаладється з трьох кроків:

- нормалізація;
- зменшення шуму;
- приведення до єдиного розміру.

3.3.2.1. Зменшення шуму

Просторові фільтри можуть ефективно використовуватися для видалення різних типів шуму на цифрових зображеннях. Такі фільтри зазвичай працюють у невеликих областях від (3 × 3) до (11 × 11). Численні просторові фільтри реалізовані з конволюційними масками, тому що така операція дає результат — зважену суму значень конкретного пікселя та кількох сусідніх до нього пікселів. Цей результат називається лінійним фільтром. Медіанні фільтри — це по суті усереднюючі фільтри; вони оперують локальними групами пікселів, що називаються квартали, та заміняють центральний піксель пікселем з середнім показником у цьому кварталі. Ця заміна виконується за допомогою конволюційної маски [90]. Медіанний фільтр є нелінійним фільтром. Нелінійний фільтр дає результат, який не може бути досягнений за допомогою зваженої суми сусідніх пікселів, як це досягалось за допомогою конволюційної маски [90].

Однак медіанний фільтр виконує операції над локальною ділянкою після визначення її розміру. Центральний піксель замінюється медіаною або мірою центральної тенденції, що береться за сусідніми пікселями, а не за середнім значенням [90].

Середній (ковзаючий) фільтр не враховує крайніх значень (високих або низьких) і не дозволяє таким значенням впливати на вибір значення пікселя, яке насправді є репрезентативним для ділянки. Тому середній фільтр добре знімає крайні ізольовані шумові пікселі (часто відомі як сольовий або перцевий шум), при цьому істотно зберігається просторова деталь. Однак його продуктивність

погіршується при великій кількості шумних пікселів, більшій за половину від всієї кількості пікселів у вікні [87].

3.3.2.2. Нормалізація координат

Ідея нормалізації координат полягає у відображенні координат масштабованого зображення руки у стандартному розмірі між [−1, +1] [109]. Мета цього кроку – зберегти область координат зображень, зафіксованих незалежно від оригінального розміру. За умова збереження області координат в обмежених границях буде забезпечено збігання моментів вищого порядку. Таким чином, масштабовані координати зображень (X, Y) перетворяться на нормалізовану множину $(X\ n, Y\ n)$, яка може бути розглянута як стандартний варіант вихідної координати (X, Y). Використовуючи центр зображення, значення координат кожного пікселя (X, Y), відображеного в області [−1, +1], можна обчислити за допомогою наступних рівнянь:

$$X_n = \left(\frac{2}{W-1}\right) * X \quad (3.1.1)$$

$$Y_n = \left(\frac{2}{H-1}\right) * Y \quad (3.1.2)$$

де H, W - висота і ширина масштабованого зображення відповідно [91].

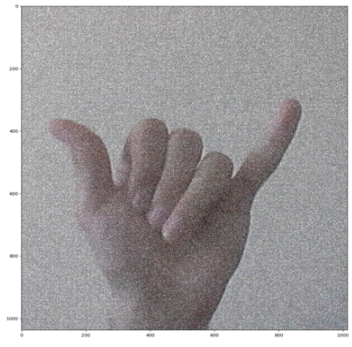
3.3.3. Розширення (аугментація) даних та розподіл

Розширення (аугментація) даних проводиться з метою збільшити розмір набору даних без ручного створення нових зображень. Аугментація здійснюється за допомогою низки методик, які дозволяють збільшити кількість зображень у наборі даних у декілька разів, урізноманітнити їх та, що важливо, зменшити здатність нейромережі до перенавчання особливостей, які присутні у оригінальному зібраному наборі даних, роблячи модель, таким чином, більш стійкою до зміни вхідних даних. У запропонованій технології відбувається аугментація даних у розмірі 3:1 оригінального зображення. Також підходи до зміни зображень можуть комбінуватися, аби внести ще більші спотворення до

оригінального набору даних. Таким чином, можна змінювати середовища, в яких тестується тренована модель.

- Вхід: зібраний набір даних зображень дактилем
- Проводиться ряд операцій над зображеннями, результуючі зображення додаються до оригінального набору даних:
    - гаусівський шум;
    - афінне перетворення;
    - обрізання + зсув;
    - віддзеркалення;
    - спотворення перспективи;
    - розмиття
- Вихід: аугментований набір даних

Рис. 3.6. Аугментація жестів

Оригінальне зображення	
Гаусівський шум	

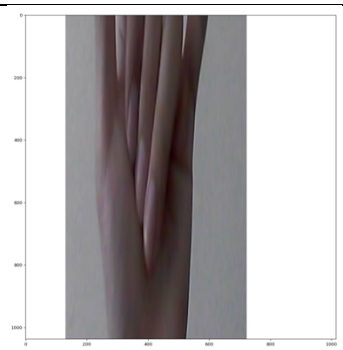
Спотворення перспективи	
Обрізання + зсув	
Віддзеркалення та афінне перетворення	
Розмиття	

Рис. 3.7. Приклади підходів до аугментації даних

3.3.4. Задання набору конфігурацій архітектур

Обрана архітектура MobileNetv2, яка демонструє високу якість та швидкодію на мобільних пристроях та пристроях з обмеженими обчислюваними потужностями, може, тим не менш, на етапі тренування бути зконфігурована гіперпараметрами, які обираються для кожного тренування індивідуально

(learning rate, batch size, кількість епох), та самою архітектурою, тобто кількістю та конфігурацією повторюваних однотипних шарів.

Даний набір конфігурацій та можливих значень гіперпараметрів формує сітку, в межах якої відбувається навчання набору нейромереж та порівняння їх на єдиному тестовому наборі.

Приклад сітки:

$$\{\text{“}learning\_rate\text{”}: [0.001, 0.0001],$$

$$\text{“}batch\_size\text{”}: [8,16,32],$$

$$\text{“}layers\_config\text{”}: [\text{“}config1\text{”},\text{”}config2\text{”},\text{”}config3\text{”}]\}$$

У запропонованій технології було згенеровано 5 конфігурацій архітектури нейронної мережі з різною кількістю шарів та кількістю параметрів, що дозволило знайти збалансовану архітектуру нейромережі з обмеженим розміром та великими показниками на тестовому наборі даних.

У результаті навчання за сіткою та тестування на тестовому наборі даних отримується відповідний набір тренувальної конфігурації та якості (наборі метрик) на тестовому наборі даних

$$\{\text{“}grid\_param\_1\text{”}: \{\text{“}f1\_score\text{”}: N, \text{“}conf\_matrix\text{”}: [\dots]\}$$

- Вхід: аугментований набір даних зображень дактилем, набір конфігурацій нейронних мереж та набір можливих гіперпараметрів
- Проводиться навчання заданої кількості нейронних мереж залежно від можливих варіантів параметрів із сітки:
    - learning rate;
    - batch size;
    - конфігурація нейромережі.
- Вихід: відповідність набору параметрів із сітки та якість нейромережі на тестовому наборі даних:
    - F1-score;
    - Confusion Matrix

Рис. 3.8. Задання набору конфігурацій нейромереж

3.4. Розпізнавання (прогнозування)

3.4.1. Виділення вхідних даних

Відеофрагменти отримують з безперервних відеозаписів, що містять приклади показаних дактилем. Запис проводиться безперервно одним файлом без змін параметрів зйомки (положення камери відносно руки, освітленості тощо).

Отримані відео за допомогою алгоритмів відеокомпресії перетворюються на набір відео для подальшої обробки і анотації.

- Вхід: сегментована множина відеозаписів дактилем на відео
- Для кожного відео створюється набір з 3 відео, що використовуються для сегментації, для отримання mocap-даних та для попереднього перегляду відео з характеристиками, що відповідають вимогам до відео, якщо потрібні характеристики, відмінні від наявних в оригінальному відео.
- Вихід: сегментована множина відеозаписів мімічних проявів на відео

Рис. 3.9. Виділення вхідних даних

3.4.2. Обробка даних

Даний крок полягає в обробці даних, щоб уніфікувати вхідні дані для нейромережі та привести їх до вигляду, який дозволить виконати прогнозування нейромережі, уникаючи певних особливостей вхідного зображення, які можуть виникнути під час його отримання.

Обробка даних скаладється з трьох кроків:

- нормалізація;
- зменшення шуму;
- приведення до єдиного розміру.

3.4.3. Розпізнавання та отримання результатів

Розпізнавання відбувається за допомогою натренованої нейронної мережі з оптимальним набором параметрів та архітектурою. В запропонованій інформаційній технології було обрано оптимальну архітектуру з максимальним можливим значенням метрик якості за мінімального набору параметрів (розміру) нейромережі.

Послідовність зображень передається на вхід нейромережі, яка за допомогою запропонованого механізму тривимірної згортки прогнозує на послідовності кадрів, а не на єдиному зображенні, що дозволяє брати до уваги темпоральний (часовий) контекст поточного зображення. Таким чином було вдосконалено підхід до розпізнавання послідовності зображень із дактилемами української абетки.

Результат розпізнавання аналізується на предмет співпадіння з результатами прогнозу на суміжних (попередніх) кадрах, для відкидання аномальних прогнозів.

- Вхід: сегментована множина відеозаписів дактилем на відео
- Відбувається розподіл вхідних зображень на послідовності, які перетинаються між собою (overlapping)
- Послідовність вхідних зображень обробляється
- Оброблена послідовність подається на вхід нейромережі
- Відбувається прогноз (inference) та отримується результат у вигляді розподілу класу дактилеми, ймовірно зображеної на послідовності зображень, із ймовірностями.
- Обирається клас із найбільшою ймовірністю та несуперечністю попереднім прогнозам
- Вихід: розпізнаний клас дактилеми

Рис. 3.10. Розпізнавання та отримання результатів

3.5. Кросплатформена розробка

Завдяки вибраним інструментам для кросплатформеної реалізації запропонована технологія вирішує проблему виконання свого завдання на декількох платформах без реалізації під кожну платформу окремо.

Програмне забезпечення, що пропонується використовувати при впровадженні інформаційної технології, є кросплатформеним і функціонує без змін незалежно від операційної системи (Windows, Linux, Android, iOS), типу процесора (x86, arm) та типу апаратного забезпечення (мобільного або стаціонарного пристрою).

Оскільки немає конкретних апаратних вимог щодо інформаційної технології для моделювання жестової мови, існують об'єктивні перешкоди для швидкості роботи пристроїв старшого покоління.

Подальша реалізація модулів використовуватиме існуючі кросплатформені технології. Модулі вивчення жестів і розпізнавання жестів, розроблені за допомогою кросплатформених технологій (Python [110], Tensorflow [111]), будуть вбудовані в інформаційну та жестову комунікаційну технологію [12]. У випадку мобільного додатка (iOS, Android) або додатка на пристрої зі стаціонарною операційною системою (Windows, Linux) під час встановлення інформаційна технологія аналізує наявне обладнання і платформу та, залежно від їх ємності, проводить серію коригувань: 1) кількість полігонів ручної моделі змінюється на пріоритет швидкості виконання; 2) під час анімації моделі руки змінюється крок анімації з пріоритетом швидкості. Якщо наявне обладнання не відповідає мінімальним вимогам інформаційних технологій, користувачеві надається рекомендація вибрати режим «онлайн», в якому обчислення не проводиться на апаратному забезпеченні.

3.5.1. Моделювання руки

Модель руки, вбудована у модуль моделювання жестів, має 27 кісток: 8 кісток — у зап'ясті, 3 — у великому пальці (1 п’ясткова кістка та 2 фаланги), 4

п’ясткові кістки та 12 фалангових — в інших пальцях. Кожна кістка з'єднана з іншою за допомогою різних типів суглобів.

Розробка власного кросплатформеного рушія для відтворення руки є нетривіальним завданням, тому в якості основної технології моделювання тривимірної моделі руки та анімації жестів між дактилемами була обрана кросплатформена технологія Unity3D. Unity3D здатна ефективно відтворити реалістичну модель руки, яка складається з понад 70 000 полігонів.

На основі анатомії кисті в рамках Unity3D була розроблена модель руки з 25 ступенями рухливості, чотири з яких розташовані в зап’ястково-п’ясткових суглобах I та V п’ясткових кісток для забезпечення руху долоні. Великий палець має 5 ступенів свободи, середній та вказівний пальці мають 4 ступені свободи (п’ястково-фаланговий суглоб має два ступені рухливості, а дистальний та проксимальний міжфалангові суглоби один). Для збереження жесту було обрано формат YAML [112].

3.5.2. Розпізнавання жестів

Модулі тренування жестів та розпізнавання жестів, розроблені за допомогою кросплатформених інструментів (програмні бібліотеки, засновані на Python, C++), можуть бути вбудовані в інформаційну та жестову комунікаційну кросплатформену технологію. Автоматичне розпізнавання мови жестів може бути досягнуте аналогічно розпізнаванню мови, при цьому знаки обробляються аналогічно фонемам або словам. Умовно розпізнавання мови жестів полягає в отриманні введених відеофрагментів, у виділенні ознак руху, що відображають мовні терміни жестової мови, а потім у застосуванні методів інтелектуального аналізу або підходів машинного навчання до вхідних даних.

Згорткові (конволюційні) нейронні мережі (CNN) [113] показали надійні результати в питаннях класифікації та розпізнаванні зображень, і вони успішно впроваджуються для розпізнавання жестів в останні роки. Зокрема, глибокі CNN були використані в дослідженнях, проведених у галузі розпізнавання жестової мови, з розпізнаванням вводу, що використовує не тільки піксельні зображення.

За допомогою глибинних камер процес значно полегшується шляхом розробки характерних профілів глибини та руху для кожного жесту мови жестів. Численні існуючі дослідження, проведені на різних жестових мовах, показують, що CNN досягають найсучаснішої точності розпізнавання жестів [113], [114].

3.6. Вдосконалена для розпізнавання дактилем архітектура MobileNetv2

Згорткові нейронні мережі мають такі переваги: відсутність необхідності в ручному оформленні особливостей жестів на зображеннях; передбачувана модель здатна розрізнити користувачів та оточуючих, що не беруть участі в тренуванні; стійкість до різних масштабів, умов освітлення й оклюзій. Також обраний підхід має декілька недоліків, які можуть бути подолані порівняно великим набором даних (1000 зображень на кожен жест, виконаних більш ніж 10 людьми різного віку, статі, національності, причому зображення мають бути зроблені в різних умовах та масштабах). Використання кросплатформеної нейронної мережі, такої як Tensorflow, дозволяє реалізувати розпізнавання жестів як кросплатформений модуль запропонованої технології та розгорнути навчену модель розпізнавання на сервері або перенести її на пристрій.

У запропонованій технології для експерименту був зібраний набір даних з літерами української мови дактилів. Кожен жест складається з 1000 зображень, 50 різних людей показували жести, зафіксовано жести у виконанні 70% чоловічих та 30% жіночих рук. Були використані різні умови освітлення (з розподілом: 20% зображень в умовах поганого освітлення, 30% в умовах посереднього світла і 50% при хорошому освітленні). Близько 10% зображень були спотворені шумом і розмиттям.

В якості основи для архітектури CNN було використано архітектуру MobileNetv2 [115] (рис. 3.11). Вона має низку переваг, таких як оптимальний компроміс між точністю та продуктивністю, особливо на мобільних пристроях. Модель MobileNet заснована на глибоко відокремлених згортках (конволюціях), що є формою факторизованих згорток, яка перетворює стандартну згортку на

глибоку згортку і 1×1 згортку, що називається точковою згорткою. Для MobileNets глибока згортка застосовує один фільтр до кожного вхідного каналу. Точкова згортка застосовує згортку 1×1 для об'єднання виходів глибокої згортки. Стандартна згортка фільтрує і поєднує входи в новий набір виходів за один крок. Глибоко відокремлювана згортка розділяє це на два шари, окремий шар для фільтрації та окремий шар для комбінування. Ця факторизація призводить до різкого зменшення кількості обчислень та розміру моделі.

Процес навчання мережі MobileNet для виявлення жестів займає 200 000 ітерацій, що становить приблизно 25 епох.

Table 1. MobileNet Body Architecture

Type / Stride	Filter Shape	Input Size
Conv / s2	3 × 3 × 3 × 32	224 × 224 × 3
Conv dw / s1	3 × 3 × 32 dw	112 × 112 × 32
Conv / s1	1 × 1 × 32 × 64	112 × 112 × 32
Conv dw / s2	3 × 3 × 64 dw	112 × 112 × 64
Conv / s1	1 × 1 × 64 × 128	56 × 56 × 64
Conv dw / s1	3 × 3 × 128 dw	56 × 56 × 128
Conv / s1	1 × 1 × 128 × 128	56 × 56 × 128
Conv dw / s2	3 × 3 × 128 dw	56 × 56 × 128
Conv / s1	1 × 1 × 128 × 256	28 × 28 × 128
Conv dw / s1	3 × 3 × 256 dw	28 × 28 × 256
Conv / s1	1 × 1 × 256 × 256	28 × 28 × 256
Conv dw / s2	3 × 3 × 256 dw	28 × 28 × 256
Conv / s1	1 × 1 × 256 × 512	14 × 14 × 256
5× Conv dw / s1	3 × 3 × 512 dw	14 × 14 × 512
5× Conv / s1	1 × 1 × 512 × 512	14 × 14 × 512
Conv dw / s2	3 × 3 × 512 dw	14 × 14 × 512
Conv / s1	1 × 1 × 512 × 1024	7 × 7 × 512
Conv dw / s2	3 × 3 × 1024 dw	7 × 7 × 1024
Conv / s1	1 × 1 × 1024 × 1024	7 × 7 × 1024
Avg Pool / s1	Pool 7 × 7	7 × 7 × 1024
FC / s1	1024 × 1000	1 × 1 × 1024
Softmax / s1	Classifier	1 × 1 × 1000

Рис. 3.11. Архітектура MobileNet

3.7. Висновки до Розділу 3

Інформаційна технологія узагальнює результати, отримані в Розділі 2, зокрема, розроблену інфологічну модель, описує процес збору та обробки даних, що поступає у набір даних для подальшого тренування моделі розпізнавання

жестів (дактилем). Дані, отримані в результаті, являють собою тензор, утворений із перетворень зображення шляхом нормалізації та стандартизації окремих кадрів відеоряду.

Експериментальна просторова анімована високополігональна модель руки людини, отримана в результаті побудови параметричної моделі з різноманітними ступенями свободи, які відображають реалістичні обмеження людського скелету, показала застосовність математичної формалізації для моделювання жестів української дактильної абетки. Параметри скелету руки зручно та компактно описуються для всієї множини жестів (дактилем) української абетки.

Побудована глибинна нейронна згорткова мережа на базі архітектури MobileNetv2, модифікована тривимірними згортками, показала вдосконалення результатів розпізнавання жестів із вхідного зображення та продемонструвала переваги у розпізнаванні з відеопотоку. Використані кросплатформені засоби для розробки запропонованої технології показали можливість використання даної технології на широкому спектрі платформ та пристроїв.

В даному розділі також розглянуто основні рекомендації щодо розробки програмного забезпечення для моделювання та розпізнавання дактилем української абетки, формалізовані у вигляді інформаційної технології, алгоритмічних реалізацій компонентів технології та концептуальної схеми, яка може використовуватися як прототип програмних модулів, і запропонована процедура підготовки та експериментальної перевірки даних, що може бути використана в експериментальних дослідженнях технології як для валідації даних, так і для перевірки застосовності алгоритмів до даного типу даних.

З метою покращення якості отриманих в запропонованій інформаційній технології результатів та підвищення якості розпізнавання жестів запропоновано провести експериментальні дослідження різних алгоритмів розпізнавання об’єктів на зображеннях, знайти шляхи покращення роботи цих алгоритмів, порівняти різні реалізації алгоритмів класифікації жестів, залежність якості їх роботи від обсягу вхідних даних, обрати з них ті, що дають найкращий результат, оцінити ефективність ознак на основі запропонованої моделі.

Розділ 4. Експериментальна програмна реалізація та результати тренувань і тестувань моделей розпізнавання дактилем

В четвертому розділі розглянуто структуру модулів і класів експериментальної програмної реалізації. Наведені приклади елементів даних та окремих складових програмної реалізації, зокрема окремих функцій, елементів керування та баз даних. Описані експерименти, що проводилися за допомогою програмної реалізації, а також результати цих експериментів під час випробування на різних наборах даних в різних умовах. Наочно показано результати окремих експериментів у вигляді графіків і таблиць.

4.1. Користувацький інтерфейс

Інтерфейс користувача виконано за допомогою бібліотеки Unity3D.

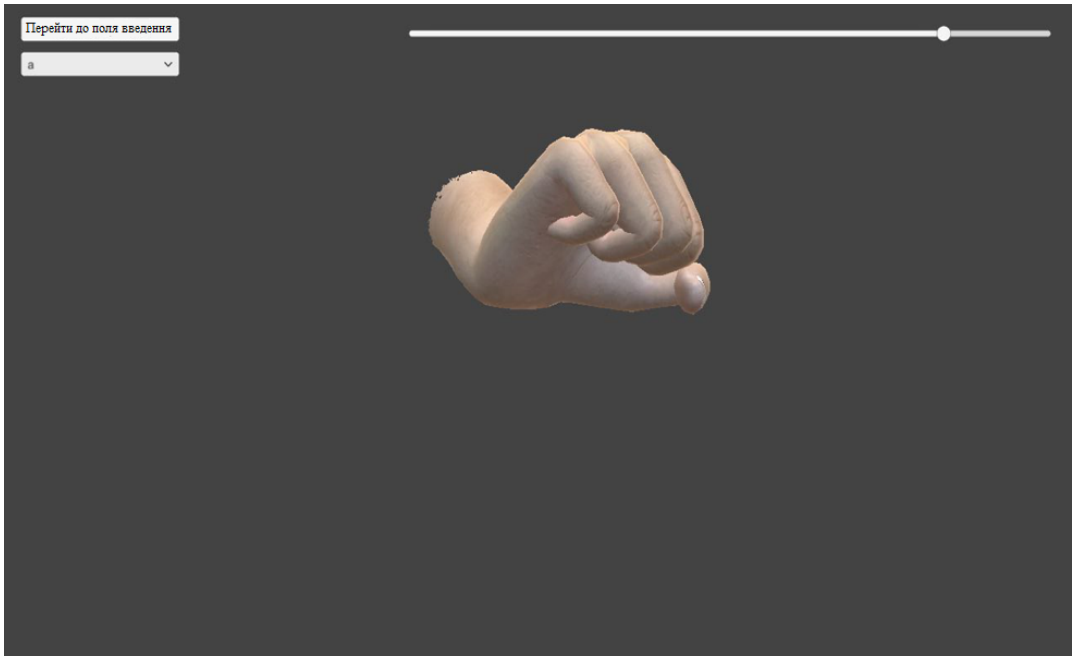


Рис. 4.1. Інтерфейс користувача моделювання жестів запропонованої технології

Рис. 4.2. Інтерфейс користувача модуля розпізнавання жестів

4.2. Структура модулів і класів програмної реалізації

4.2.1. Структура модулів програмної реалізації

Програмна реалізація містить два основні модулі (розпізнавання та моделювання) та базу даних PostgreSQL, яка використовується модулем моделювання.

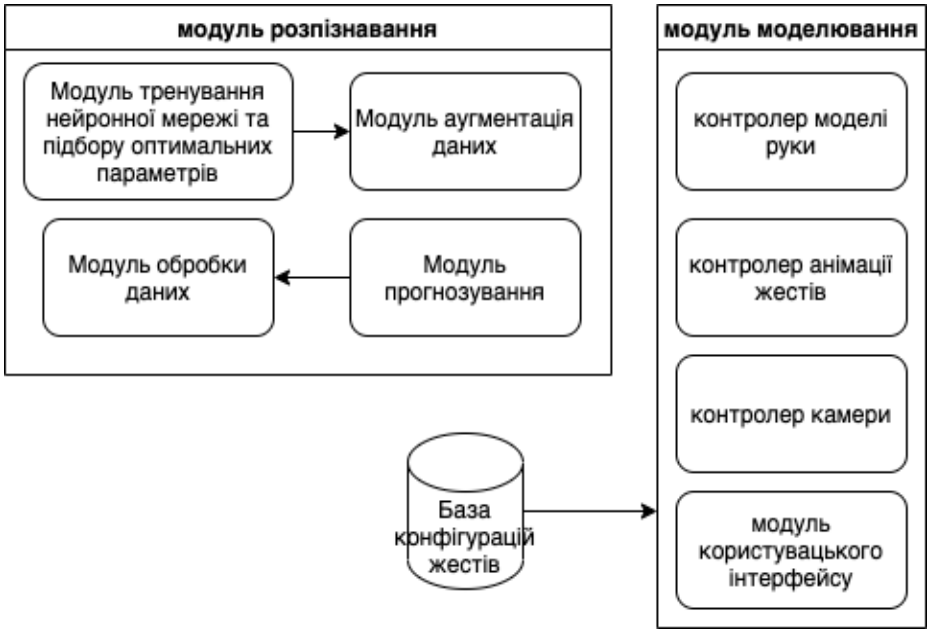


Рис. 4.3. Схема модулів програмної реалізації

4.2.2. Структура класів програмної реалізації

Далі розглянуто більш детально класи програмної реалізації, які реалізують основні методи, функції або процедури, що стосуються модулів програмної реалізації.

Модуль моделювання програмної реалізації написаний на мові C#, яка підтримується Unity3d, та складається з:

- головного класу користувацького інтерфейсу Player;
- класу роботи з базами даних DatabaseClass;
- класу зміни користувацького інтерфейсу залежно від платформи та орієнтації (горизонтальна/вертикальна) LayoutSwitcher;
- класу, який відповідає за керування відтворення заданої послідовності дактилем, KaraokeHandler;
- класу, який відповідає за відтворення анімації, AnimationCreator;
- класу, який відповідає за розширення та зміну анімації, AnimationExtender;
- класу, який відповідає за роботу тривимірної моделі руки із скелетом з заданими обмеженнями, HandController;
- класу, який відповідає за сцену та роботу камери, CameraController;
- бібліотек для роботи з базами даних PostgreSQL.

Модель руки знаходиться у форматі .FBX у відповідній папці з моделями (Assets/Models), файли анімацій у форматі .anim знаходяться у відповідній папці з анімаціями (Assets/Animations). Текстури знаходяться у Assets/Textures у форматах .png, .hdr.

Модуль розпізнавання складається з декількох відокремлених частин, які не взаємодіють напряму. Зокрема можна виділити 2 частини: тренування та прогнозування.

- Частина тренування поєднує модуль для тренування нейронної мережі та підбору оптимальної архітектури разом із модулем аугментації даних.

- Частина прогнозування поєднує модуль для прогнозування та модуль із обробки даних (який також, в свою чергу, використовується модулем аугментації даних).

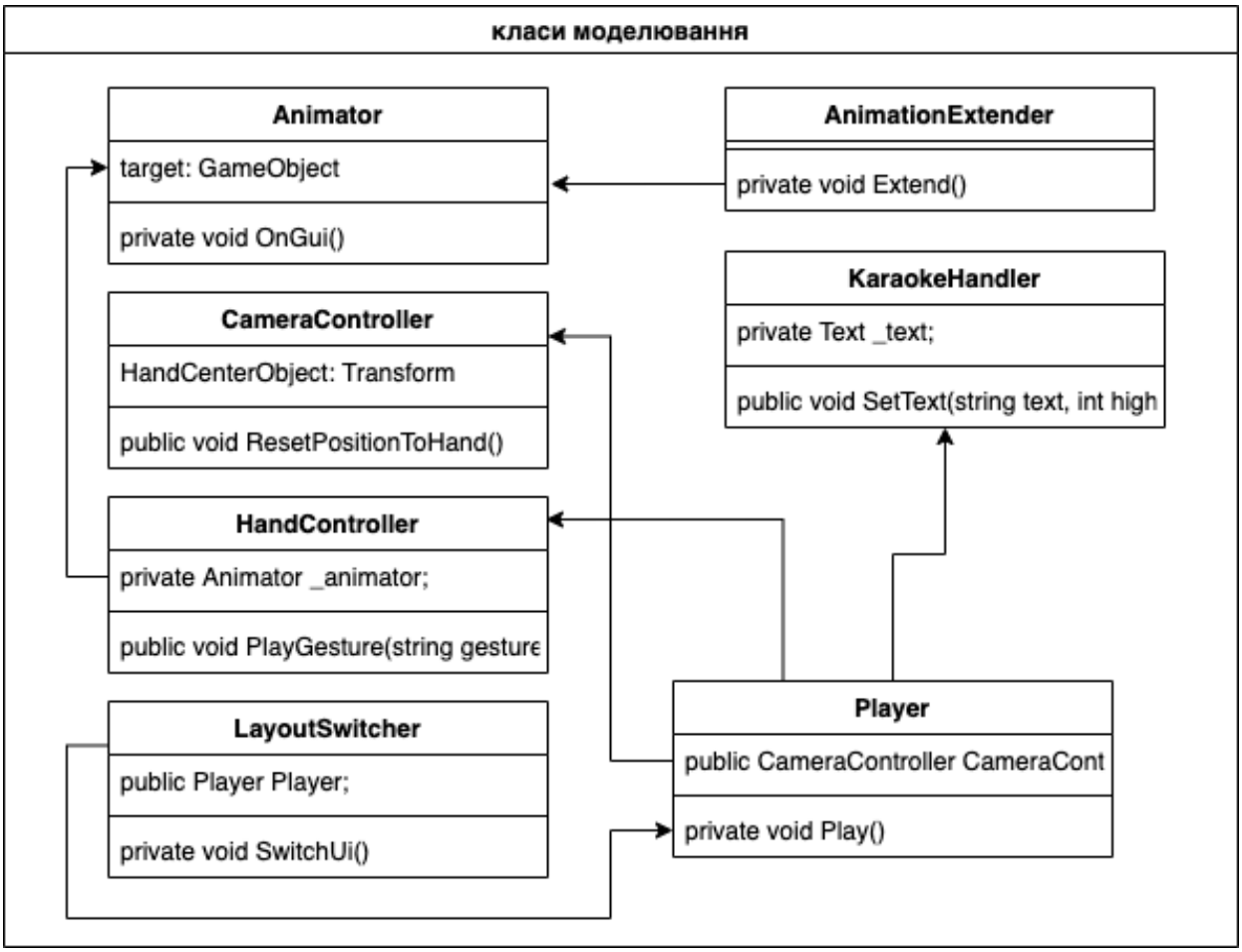


Рис. 4.4. Структура модулів, бібліотек і класів програмної реалізації модуля моделювання

Модуль розпізнавання програмної реалізації написаний мовою Python, яка є кросплатформеною, за допомогою бібліотеки для навчання глибоких нейромереж Tensorflow, яка також є кросплатформеною, та дозволяє тренувати моделі, які можна розгортати (serving / deploy) на серверах, мобільних пристроях та інших платформах.

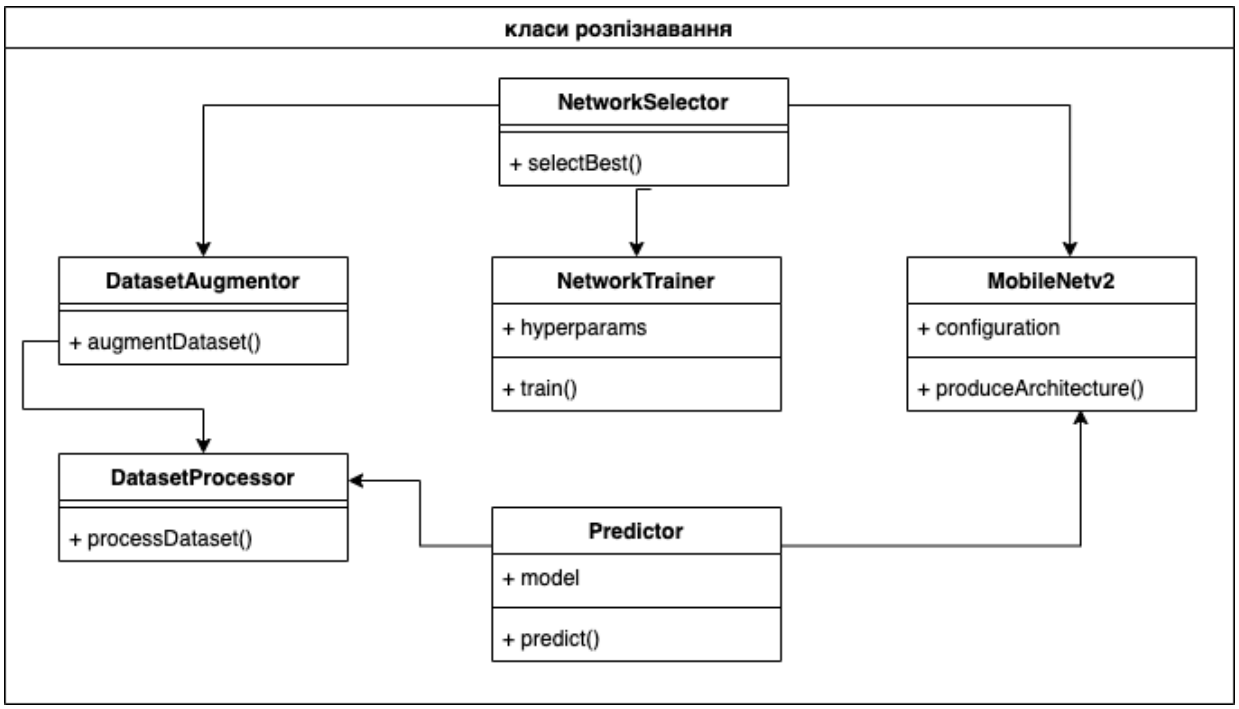


Рис. 4.5. Структура модулів, бібліотек і класів програмної реалізації модуля розпізнавання

4.2.3. Структура бази даних

У запропонованій програмій реалізації використовується база даних PostgreSQL, яка зберігає сукупність конфігурацій жестів української дактильної абетки. Кожна конфігурація має формат YAML та зберігається у базі даних в єдиній таблиці у вигляді рядка

$$\%YAML: 1.0\ rotation: [\ -9.5845758914947510e-02, 8.3791027449819921e-09,\ -9.9539619684219360e-01, -7.5690525770187378e-01,\ -6.4944797754287720e-01, 7.2881683707237244e-02,\ -6.4645802974700928e-01, 7.6040601730346680e-01, 6.2246840447187424e-02\ ]$$

$$hand\_joints:\ finger1joint1: [\ 0., 0., 0.\ ]\ finger1joint2: [\ 0., 0., 0.$$

ID	GestureName	Language	Config
1	A	Ukr	“%YAML:1.0 rotation: […]

Рис. 4.6. Приклад таблиці, яка зберігає конфігурацію дактилем для модуля моделювання

4.3. Набір даних

Для навчального процесу моделі розпізнавання жестів української дактильної абетки Convolutional Neural Network, заснованої на архітектурі MobileNet, необхідно зібрати відповідний набір даних через відсутність набору для української мови жестів у вільному доступі.

Було розроблено спеціальне програмне забезпечення (рис. 4.7) для запису коротких відеопослідовностей жестів української дактильної абетки, показаних різними людьми. Оскільки програмне забезпечення для запису — це не основна частина запропонованої технології, а скоріше допоміжний інструмент, воно було розроблене лише в операційній системі сімейства Windows, використовуючи мову програмування C# та .NET фреймворк.

Процедура запису виглядає так:

1. Людина сидить перед веб-камерою, підключеною до програмного забезпечення для запису.
2. Людині потрібно перемістити руку в потрібну область програмного забезпечення для запису.
3. Людина показує конкретний жест з української дактильної абетки.
4. Оператор починає запис.
5. Людина, що показує жест, починає плавно переміщувати руку по різних осях.
6. Після запису відео відповідної довжини оператор припиняє запис.
7. Процес продовжується записом наступного жесту.

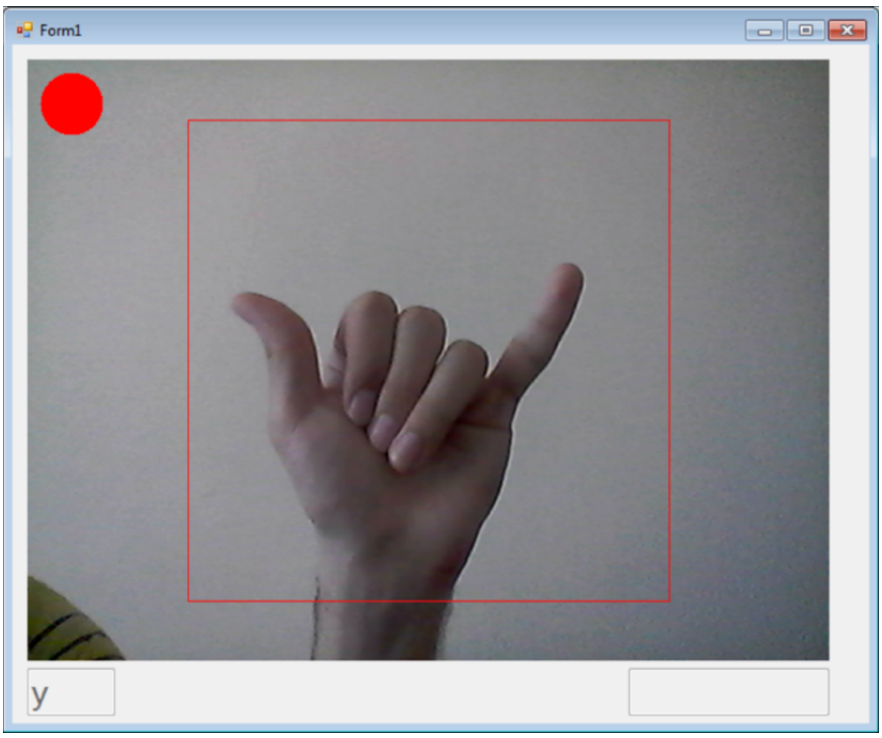


Рис. 4.7. Інтерфейс програмного забезпечення для запису набору даних дактилів

Оскільки підготовка згорткової нейронної мережі сильно залежить від великого та різноманітного набору даних, для досягнення достатньо високого рівня метрики точності було зібрано набір даних українських літер мови жестів з різними характеристиками. Кожен жест складається з 1000 зображень, 50 різних людей показували жести, з розподілом: 70% чоловічої та 30% жіночої руки.

Були використані різні умови освітлення (з розподілом: 20% зображень в умовах поганого освітлення, 30% в умовах посереднього світла і 50% при хорошому освітленні). Близько 10% зображень були спотворені шумом і розмиттям. Загалом в якості навчального набору даних було зібрано 50 000 оригінальних зображень. Після застосування додаткових методів збільшення даних (таких як обертання, випадкове обрізання, дзеркальне відображення тощо) кінцевий набір даних становив близько 150 000 зображень (рис. 4.8). Для тестування було відібрано 10% набору даних, що склало кінцевий набір із 135 000 зображень та кінцевий тестовий набір — 15 000 зображень (рис. 4.9).

Рис. 4.8. Зразок зібраного набору даних

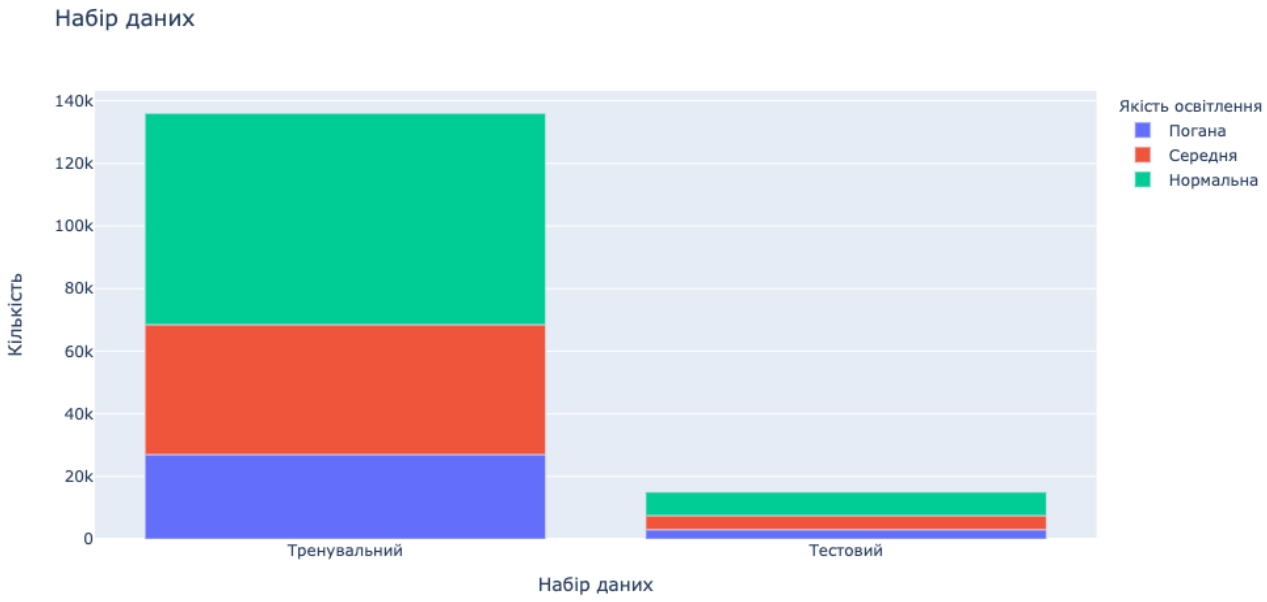


Рис. 4.9. Розподіл зібраного набору даних за якістю освітлення

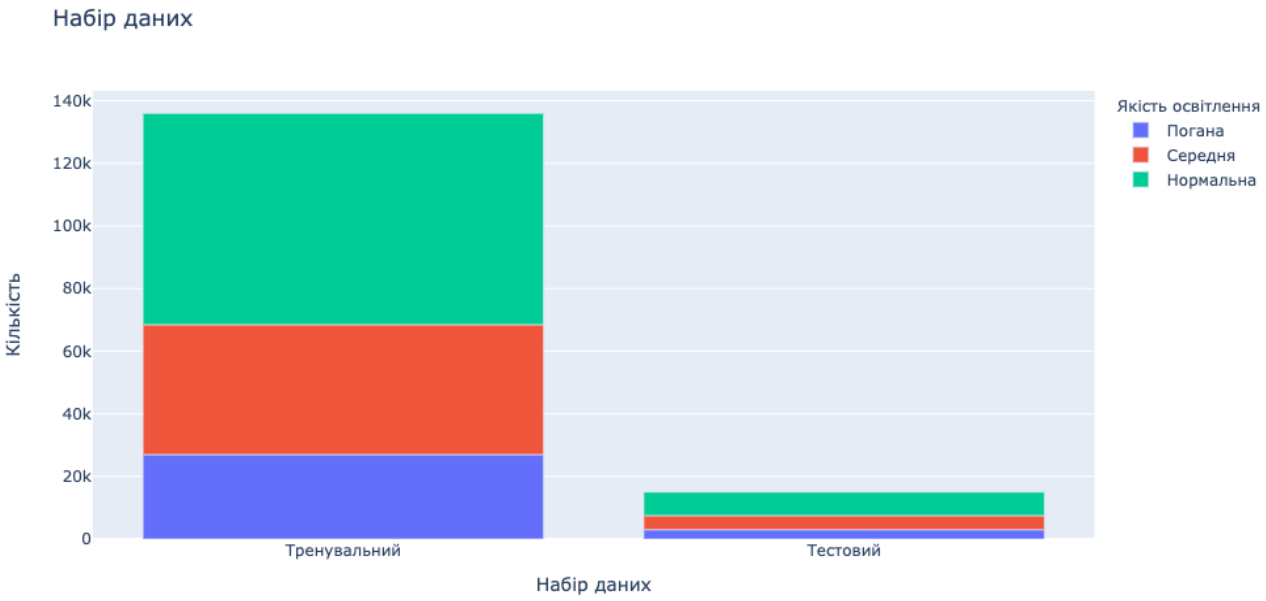


Рис. 4.10. Розподіл зібраного набору даних за якістю зображення

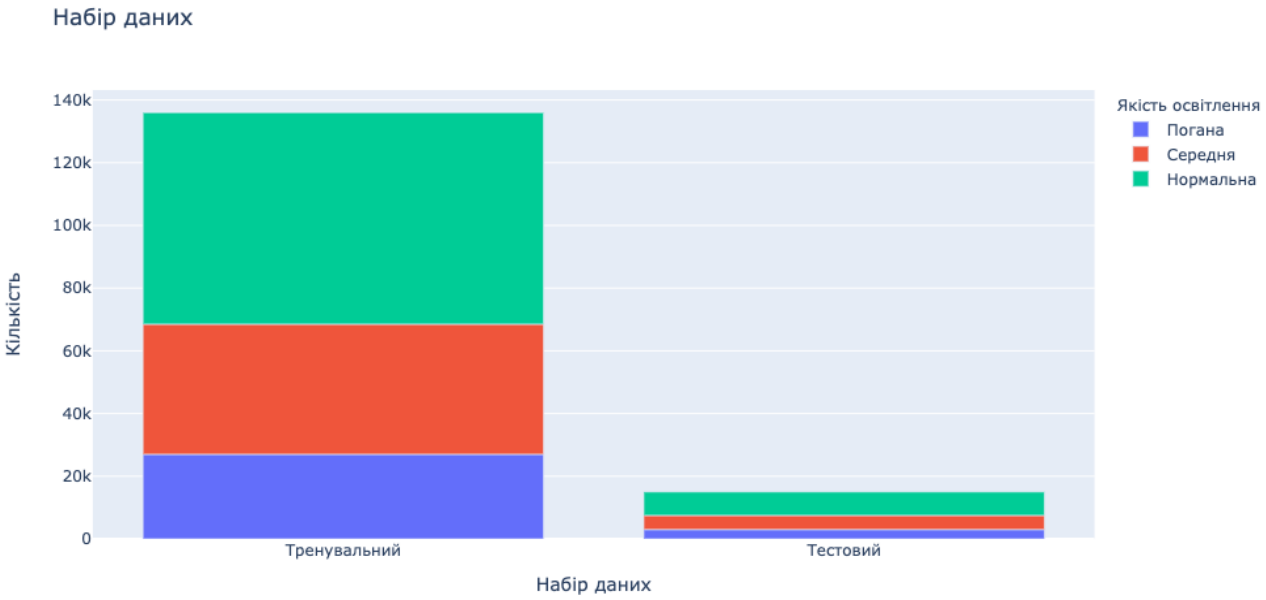


Рис. 4.11. Розподіл зібраного набору даних за статтю

4.4. Експериментальні тренування та досліди навченої моделі

Під час навчального процесу Convolutional Neural Network на основі архітектури MobileNet було створено кілька модифікацій архітектури, щоб знайти найкращий компроміс між кількістю шарів та точністю.

З часом точність навченої моделі припинила зростати, що показано на рисунках 4.18, 4.19, тому архітектура №4 була визнана оптимальною з точки зору найменшої архітектури з найкращою точністю (середній макробал f1).

При кожному навчанні застосовувались стандартні прийоми боротьби з перенавчанням нейронної мережі.

4.5. Вибір архітектури моделі

У цьому розділі показано 5 різних архітектур (таблиця 4.1) та їх матриці метрики та помилок. Архітектура 5 припинила демонструвати зростання балів f1, хоча і має більш складні показники. В якості остаточного варіанту запропонованої технології було обрано архітектуру 4 як найкращий компроміс за розміром та продуктивністю.

Architecture 1	Architecture 2	Architecture 3	Architecture 4	Architecture 5
Conv / s2	Conv / s2	Conv / s2	Conv / s2	Conv / s2
Conv dw / s1	Conv dw / s1	Conv dw / s1	Conv dw / s1	Conv dw / s1
Conv / s1	Conv / s1	Conv / s1	Conv / s1	Conv / s1
Conv dw / s2	Conv dw / s2	Conv dw / s2	Conv dw / s2	Conv dw / s2
Conv / s1	Conv / s1	Conv / s1	Conv / s1	Conv / s1
Conv dw / s1	Conv dw / s1	Conv dw / s1	Conv dw / s1	Conv dw / s1
Conv / s1	Conv / s1	Conv / s1	Conv / s1	Conv / s1
Conv dw / s2	Conv dw / s2	Conv dw / s2	Conv dw / s2	Conv dw / s2
Conv / s1	Conv / s1	Conv / s1	Conv / s1	Conv / s1
Conv dw / s1	Conv dw / s1	Conv dw / s1	Conv dw / s1	Conv dw / s1
Conv / s1	Conv / s1	Conv / s1	Conv / s1	Conv / s1
Conv dw / s1	2 x Conv dw / s1	3 x Conv dw / s1	Conv dw / s2	Conv dw / s2
Conv / s1	2 x Conv / s1	3 x Conv / s1	Conv / s1	Conv / s1
Conv dw / s2	Conv dw / s2	Conv dw / s2	4 x Conv dw / s1	5 x Conv dw / s1
Conv / s1	Conv / s1	Conv / s1	4 x Conv / s1	5 x Conv / s1
Avg Pool / s1	Avg Pool / s1	Avg Pool / s1	Conv dw / s2	Conv dw / s2
FC / s1	FC / s1	FC / s1	Conv / s1	Conv / s1
Softmax / s1	Softmax / s1	Softmax / s1	Avg Pool / s1	Conv dw / s2
			FC / s1	Conv / s1
			Softmax / s1	Avg Pool / s1
				FC / s1
				Softmax / s1

Табл. 4.1. Показники навчуваності різних архітектур

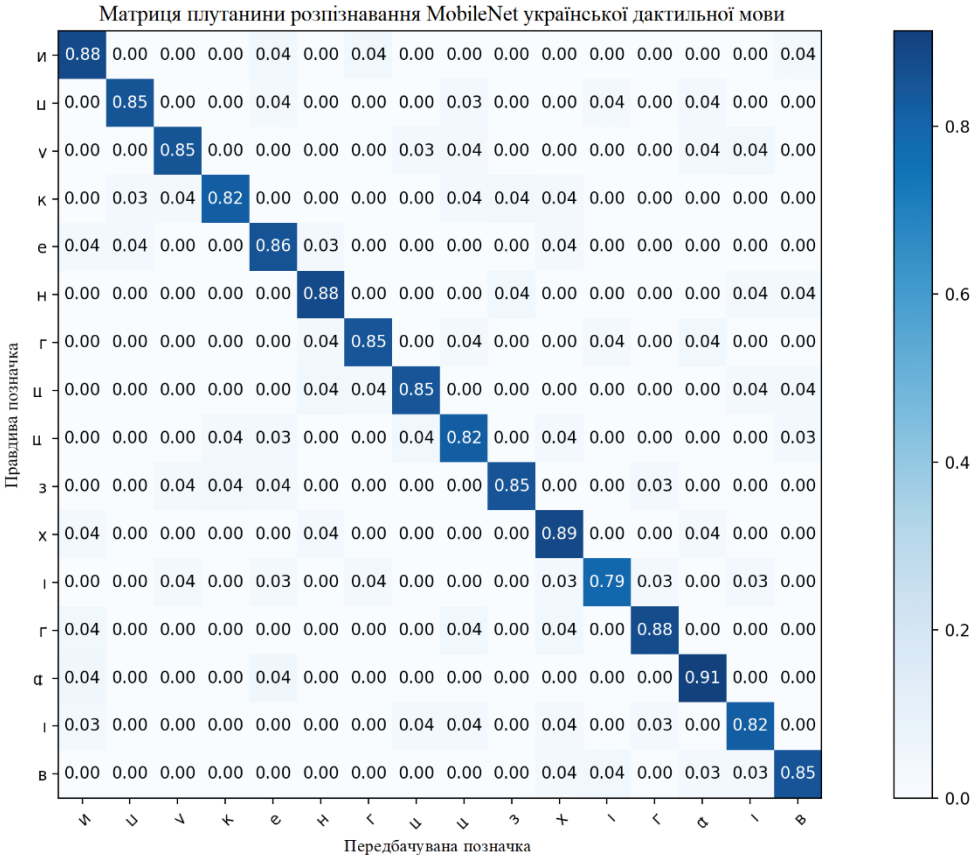


Рис. 4.12. Результати матриці помилок архітектури 1 на тестовому наборі даних

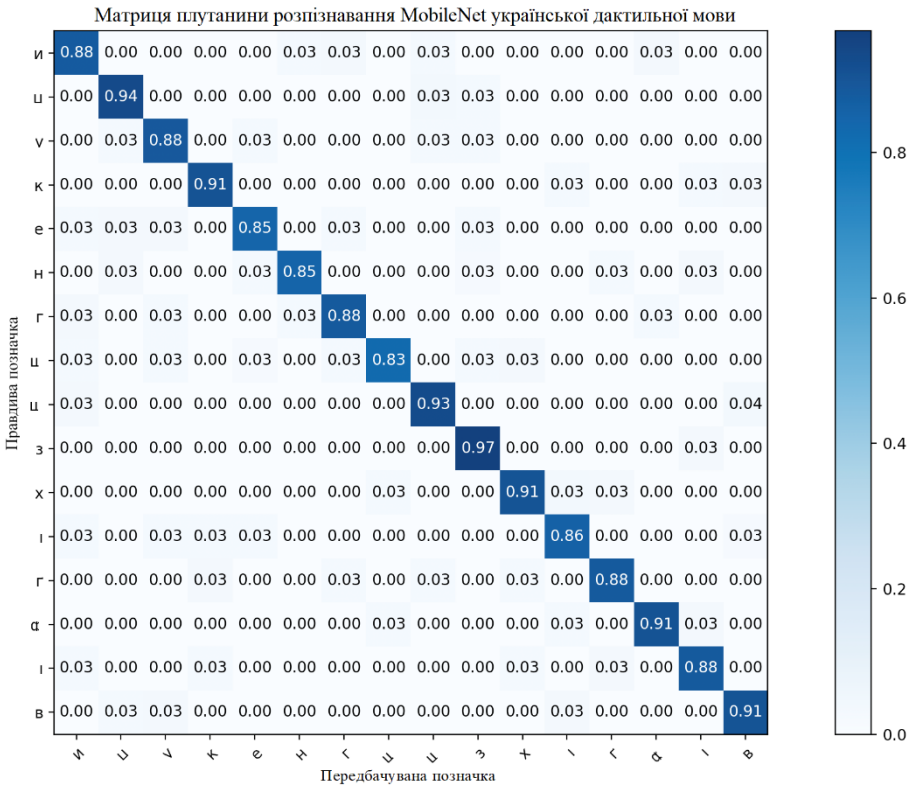


Рис. 4.13. Результати матриці помилок архітектури 2 на тестовому наборі даних

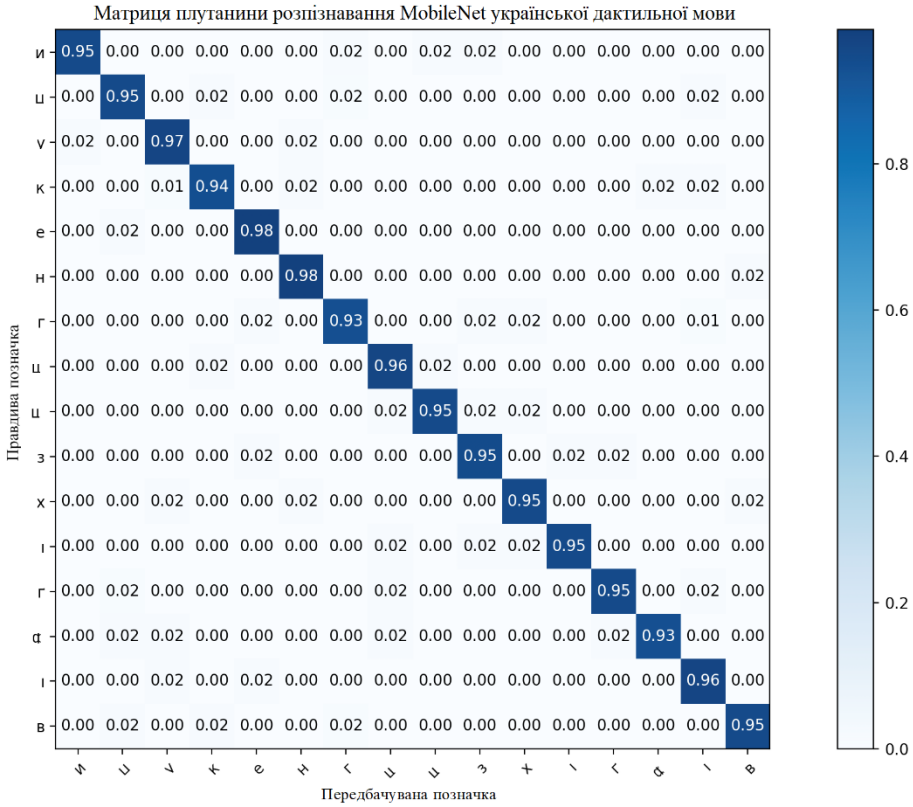


Рис. 4.14. Результати матриці помилок архітектури 3 на тестовому наборі даних

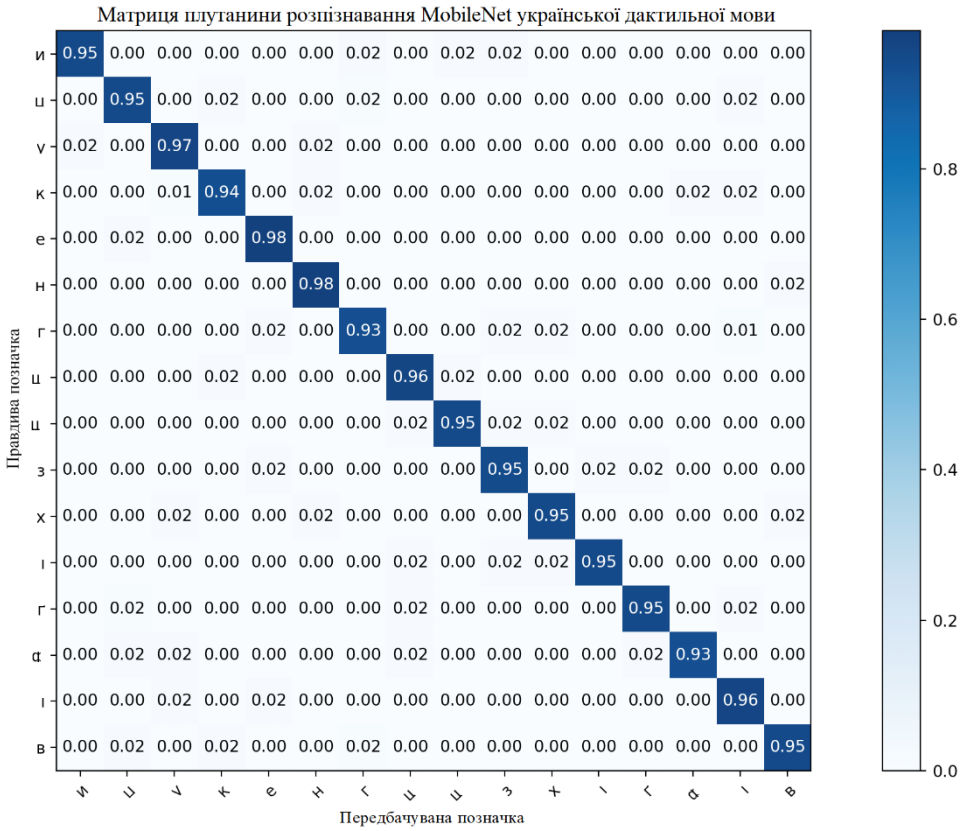


Рис. 4.15. Результати матриці помилок архітектури 4 на тестовому наборі даних

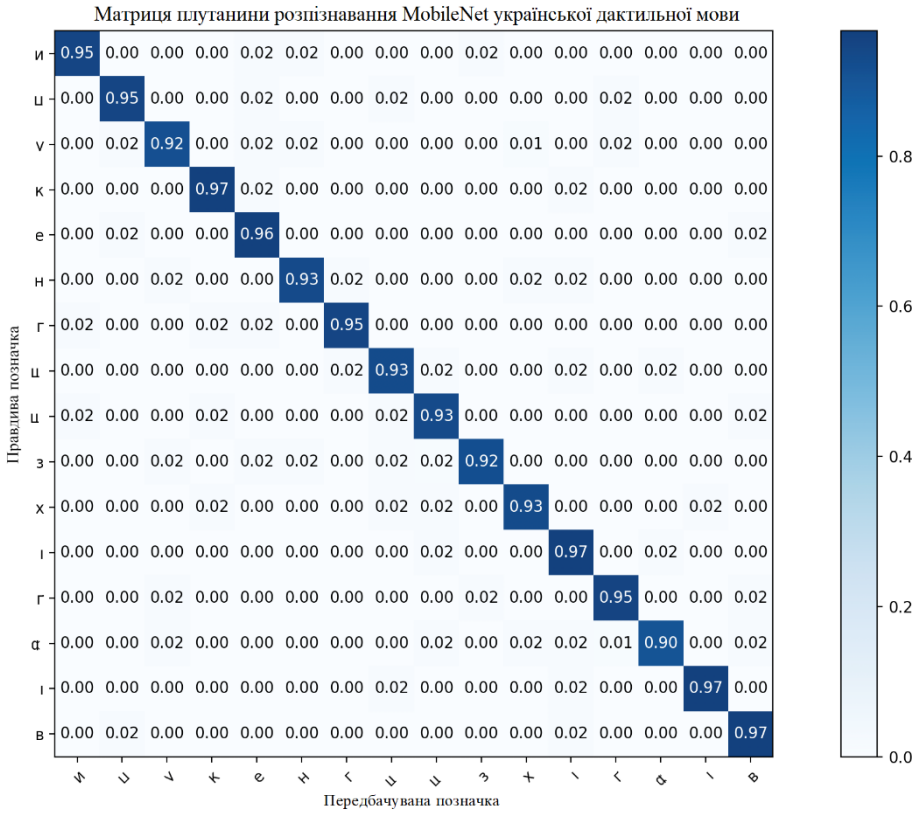


Рис. 4.16. Результати матриці помилок архітектури 5 на тестовому наборі даних

architecture	macro avg f1
1	0.853890064
2	0.891860278
3	0.917950451
4	0.953757086
5	0.94369286

Табл. 4.2. Середній макробал f1 для порівняння підготовлених архітектур

Рис. 4.17. Діаграми, що показують прогрес навчання моделі MobileNet

Approach	macro avg f1
SVM-based classifier	0.75491
MobileNet CNN arch v4	0.95375
ImageNet CNN	0.92431
Contour analysis based	0.85421

Табл. 4.3. Порівняння навченої моделі з сучасними підходами

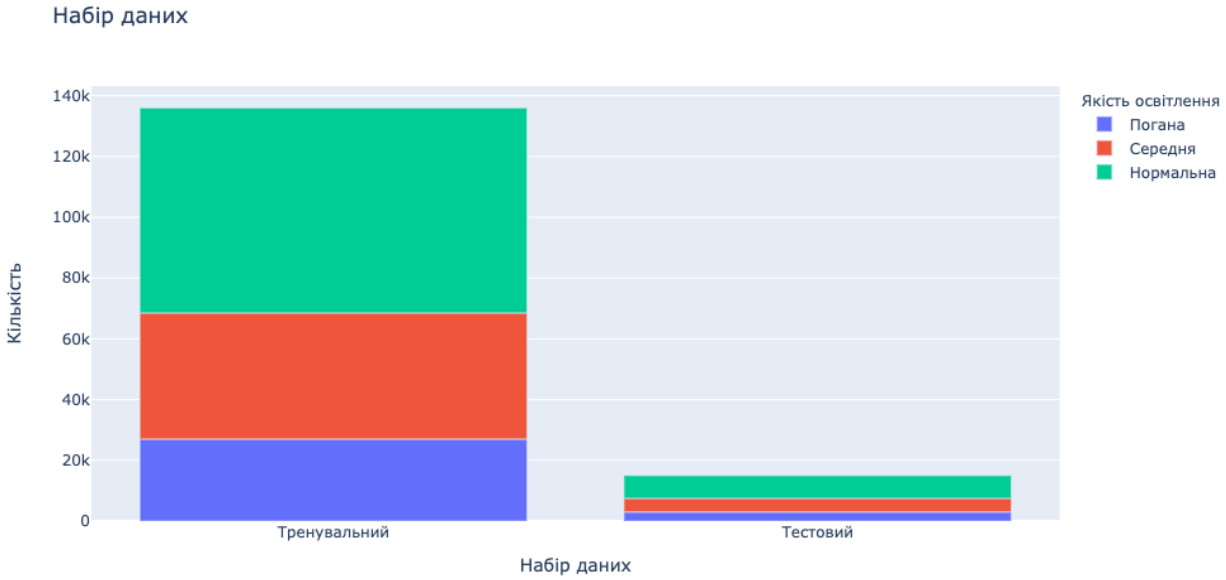


Рис. 4.18. Діаграма, що показує прогрес навчання моделі з певною архітектурою залежно від кількості ітерацій

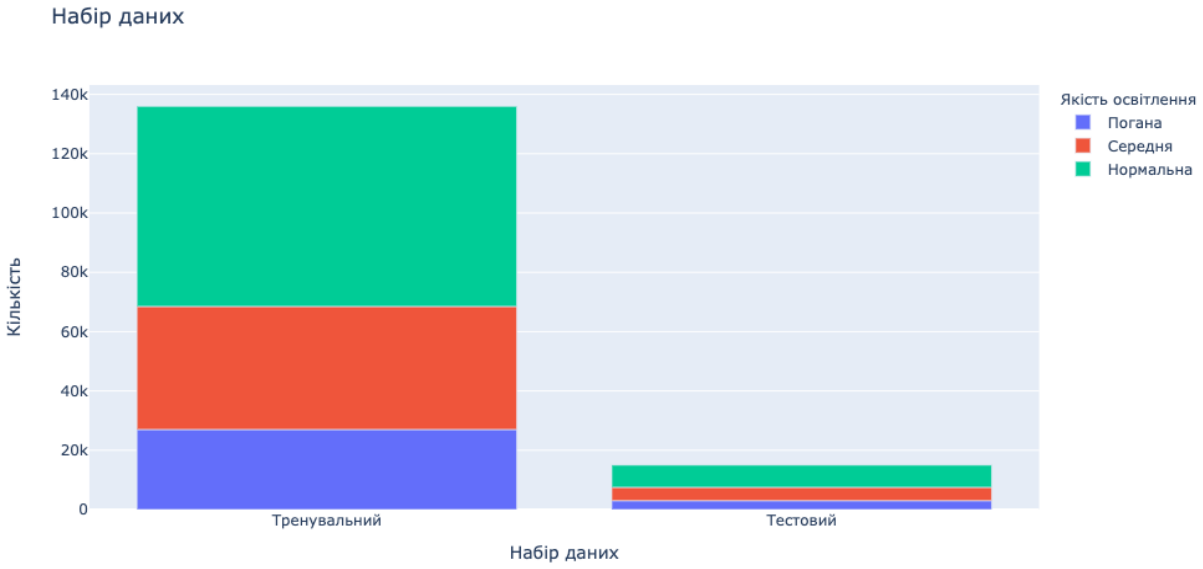


Рис. 4.19. Діаграма, що показує якість навченої моделі залежно від складності архітектури та наявності в ній тривимірної згортки

4.6. Висновки до Розділу 4

Отримана програмна реалізація технології є першою версією розробки повноцінної технології моделювання та розпізнавання жестів і потребує вдосконалень у подальших дослідженнях.

Випробувані алгоритми інтелектуальної обробки даних показали, що розпізнавання жестів у послідовності зображень (відео) вимагає спеціальної архітектури моделей та процедури підготовки даних для покращення результатів.

Отримані результати випробувань експериментальної технології також дали змогу стверджувати, що математична модель скелету руки може бути застосована для реалістичного моделювання жестів.

Отримані результати дослідів із різними архітектурами моделей та різними наборами даних продемонстрували підвищення якості розпізнавання з використанням вдосконаленої архітектури MobileNetv2 з тривимірною згорткою. Процес підбору архітектури продемонстрував оптимальне співвідношення між складністю моделі та її ефективністю в розпізнаванні на заданому наборі даних. Досягнуто якості моделі у 0.97 макробалів f1 на заданому тестовому наборі даних.

В рамках запропонованої реалізації було вперше зібрано набір даних розміром у 50 тисяч зображень із усіма дактилемами української дактильної абетки, продемонстрованими 50 різними людьми, та аугментовано до 150 тисяч зображень.

Висновки по роботі

У даному дисертаційному дослідженні були розглянуті питання, присвячені вирішенню проблем моделювання і розпізнавання мімічної складової проявів емоцій. В результаті дослідження було запропоновано шляхи вирішення поставленої задачі дисертаційного дослідження, що полягали у розробці математичних моделей, методів, алгоритмів та їх реалізацій, зокрема:

1. Створено технологію, що складається з двох основних модулів: моделювання жестів та розпізнавання жестів, які використовують базу даних із специфікаціями жестів, що зберігаються у форматі YAML у базі даних PostgreSQL.

2. Створено технологію, яка вперше реалізує моделювання жестів та розпізнавання жестів для українських жестів дактильної абетки за допомогою інструментів кросплатформеної розробки. Моделювання жестів було реалізовано за допомогою фреймворку Unity3D, який є кросплатформеним та демонструє задовільну продуктивність на різних платформах (мобільних, веб- та настільних), демонструючи реалістичну тривимірну модель руки. Кількість полігонів та анімаційний крок переходів жесту можна регулювати задля виконання роботи.

3. Було вперше зібрано набір даних із понад 50 000 зображень, використовуючи різні умови та руки 50 людей різного віку та статі. Набір даних доповнено за допомогою методик аугментації, а кінцевий набір даних складається з 150 000 зображень. Модуль розпізнавання жестів був реалізований за допомогою фреймворку Tensorflow, що забезпечує можливість функціонування моделі на різних платформах без будь-яких модифікацій бази даних та моделі або модулю для тренування моделі.

4. Було вдосконалено підхід до розпізнавання дактилем. В якості моделі для розпізнавання жестів була обрана архітектура MobileNetv2 як модель з найкращим компромісом між розмірами та точністю, особливо на низькопродуктивних платформах (таких як мобільні та веб). Модель була підготовлена на зібраному наборі даних української дактильної мови. Завдяки

вдосконаленню за допомогою технології тривимірних згорток модель показала ультрасучасний рівень якості.

5. Проведено експериментальні випробування, на основі яких була обрана оптимальна архітектура моделі з метою збереження найкращого рівня швидкодії з найменшим можливим розміром моделі.

6. Були показані порівняльні результати різних архітектур із тривимірними згортками та без них, отримані в ході експерименту. Продуктивність моделі CNN порівнювалася з іншими підходами та продемонструвала аналогічні або вищі значення, які потенційно збільшуються із збільшенням розміру набору даних.

Запропоновану технологію комунікації жестами можна додатково доповнити іншими жестами, мовами й іншими кросплатформеними модулями.

Результати дисертаційного дослідження мають як теоретичне, так і практичне спрямування. Результати окремих елементів даного дослідження були використані в наукових дослідженнях, в наукових темах і звітах.

Значення результатів дослідження полягає у застосовності їх в різних галузях. Найбільш очевидним є використання з залученням технологій розпізнавання і моделювання дактилем при навчанні жестової мови людей, які мають в оточенні особу з вадами слуху, для розробки інформаційних технологій для інклюзивної освіти для шкільних і дошкільних закладів із вивченням жестової мови, для самонавчання батьків дітей із вадами слуху або соціальних працівників (медицина, національна поліція та інші служби), що контактують із людьми з вадами слуху.

Література

1 Електронний ресурс Українського товариства глухих. Режим доступу: http://utog.org/

2 Електронний ресурс відкрити архіву мов. Режим доступу: http://www.language-archives.org/item/oai:ethnologue.com:ukl

3 Електронний ресурс ВОЗ. Режим доступу: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

4 Електронний ресурс Національного Інституту глухоти та інших вад комунікацій США. Режим доступу: https://www.nidcd.nih.gov/health/statistics/quick-statistics-hearing

5 Електронний ресурс Київстар. Режим доступу: https://kyivstar.ua/ru/mm/news-and-promotions/kolichestvo-4g-smartfonov-v-seti-kievstar-vyroslo-vdvoe

6 Електронний ресурс The Linux Information Project із визначенням платформи. Режим доступу: http://www.linfo.org/platform.html

7 Електронний ресурс Amazon. Режим доступу :https://aws.amazon.com/what-is-cloud-computing/

8 Електронний ресурс The Linux Information Project із визначенням кросплатформеності. Режим доступу: http://www.linfo.org/cross-platform.html

9 Kondratiuk S. Gesture recognition using cross platform software and convolutional neural networks // Штучний інтелект. – т. 85-86, – 2019 – С.107-113.

10 С.С. Кондратюк. Платформонезалежне програмне забезпечення для розробки систем жестової комунікації: моделювання дактильної мови / С.С. Кондратюк, Ю.В. Крак // Штучний інтелект.– т.73 в.3 – 2016. С.36-47.

11 Kondratiuk S. Cross-platform software for the development of sign communication system: Dactyl language modeling. / Kondratiuk S., Krak I. //Proceedings of the 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT – 2017. – p. 167-170.

12 Kondratiuk S. Dactyl Alphabet Modeling and Recognition Using Cross Platform Software. / Kondratiuk S., Krak I. // Proceedings of the 2018 IEEE 2nd International Conference on Data Stream Mining and Processing, DSMP 2018. 21-25 Aug. 2018. Lviv. – 2018. – P. 420 – 423.

13 Електронний ресурс Michigan State University's ASL. Режим доступу: http://commtechlab.msu.edu/sites/aslweb/browser.htm

14 Graschenko L. A., Fisun A. P. i dr. Teoreticheskie i prakticheskie osnovyi cheloveko-kompyuternogo vzaimodeystviya: bazovyie ponyatiya cheloveko-kompyuternyih sistem v informatike i informatsionnoy bezopasnosti: Monografiya / Red. A. P. Fisun. — 2004. — 169 s.

15 Електронний ресурс Samsung TV Gesture book. Режим доступу: www.samsung.com/ph/smarttv/common/guide_book_3p_si/waving.html

16 Електронний ресурс Apple Touchless Gesture System for iDevices. Режим доступу: http://www.patentlyapple.com/patently-apple/2014/12/apple-invents-a-highly-advanced-air-gesturing-system-for-future-idevices-and-beyond.html

17 Rafiqul Zaman Khan. Comparative study of hand gesture recognition system / Rafiqul Zaman Khan, Noor Adnan Ibraheem, Natarajan Meghanathan, et al. // SIPM, FCST, ITCA, WSE, ACSIT, CS & IT 06, – 2012, – pp. 203–213.

18 Michael Neff Gesture Modeling and Animation by Imitation / Michael Neff, Michael Kipp, Irene Albrecht, Hans-Peter Seidel // – 2006.

19 Електронний ресурс Dynamic Controller Toolkit, Ari Shapiro, Derek Chu, Brian Allen, Petros Faloutsos, 2005. Режим доступу: www.arishapiro.com/Sandbox07_DynamicToolkit.pdf

20 Yu. G. Kryvonos. Modelyuvannya rukhiv virtual'noho personazha dlya prostorovoho vidtvorennya zhestovoyi movy / Yu. G. Kryvonos, Yu. V. Krak, O. V. Barmak // 2010 ISSN 1560-9189 Reyestratsiya, zberihannya i obrobka danykh, T. 12, # 2 – 2010

21 S. Sueda, “Musculotendon simulation for hand animation,” / S. Sueda, A. Kaufman, and D. K. Pai // in ACM Transactions on Graphics (TOG), vol. 27, no. 3. ACM, – 2008, p. 83.

22 T. Rhee, “Human hand modeling from surface anatomy,” / T. Rhee, U. Neumann, and J. P. Lewis // in Proceedings of the 2006 symposium on Interactive 3D graphics and games. ACM, – 2006, – pp. 27–34.

23 R. Vaillant, “Implicit skinning: Realtime skin deformation with contact modeling,” / R. Vaillant, L. Barthe, G. Guennebaud, M.-P. Cani, D. Rohmer, B. Wyvill, O. Gourmel, and M. Paulin // ACM Transactions on Graphics (TOG), vol. 32, no. 4, – 2013, – p. 125

24 T. MacLeod, T. P. Rioux, M. Yokota, P. Li, B. D. Corner, and X. Xu, “Individualized human cad models: Anthropmetric morphing and body tissue layering,” – 2014.

25 F. Bogo, “Detailed full-body reconstructions of moving people from monocular rgb-d sequences,” / F. Bogo, M. J. Black, M. Loper, and J. Romero // in Proceedings of the IEEE International Conference on Computer Vision, – 2015, – pp. 2300–2308.

26 Hoevenaren, T. J. Maal, E. Krikken, A. de Haan, S. Berge, and ´ D. Ulrich, “Development of a three-dimensional hand model using 3d stereophotogrammetry: Evaluation of landmark reproducibility,” Journal of Plastic – 2015

27 Електронний ресурс Api overview - leap motion. Режим доступу: https://developer. leapmotion.com/

28 Електронний ресурс Leap motion: Vr best practices guidelines. Режим доступу: https://developer.leapmotion.com/vr-best-practices

29 Y. Bulatov, “Hand recognition using geometric classifiers,” / S. Jambawalikar, P. Kumar, and S. Sethia // in Biometric Authentication. Springer, –2004, – pp. 753–759.

30 H.-L. Yu and B. Strauch, Atlas of hand anatomy and clinical implications. Mosby Incorporated, – 2004.

31 Sing language tutoring tool, Oya Aran, // eNTERFACE’06 – 2006

32 3D hand pose retrieval from a single 2D image Haiying Guan Chin-Seng Chua Yeong-Khing Ho, – 2001

33 H. Rijpkema. Computer animation of knowledge-based human grasping. / H. Rijpkema and M. Girard // Computer Graphics, t. 25(4) –1993 – p.:339-348

34 R. Cipolla and A. Pentland (Editors), Computer Visiorz forHuman-Machine Interaction. Cambridge: Cambridge University Press, – 1998.

35 V. Pavlovic, ‘‘Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review,” / V. Pavlovic, R. Sharma, T. S. Huang // IEEE PAMI, – vol. 19, No. 7, –1997, pp. 677-695.

36 T. Heap and D. Hogg, “Wormholes in shape space: trackingthrough discontinuous changes in shape” –1998

37 C. Chang, “Model-Based Analysis of Hand Gestures From Single Images Without Using Marked Gloves Or Attaching Marks on Hands”, / C. Chang, W. Tsai // ACCV2000, – 2000, – pp. 89-93.

38 C.S. Chua, “Model-based Finger Posture Estimation”, / C.S. Chua, H. Y. Guan and Y. K. Ho // ACCV2000, –2000, – pp. 43-48.

39 J. Kuch “Vision-Based Hand Modeling and Tracking for Virtual Teleconferencing and Telecollaboration”, / J. Kuch and T. S. Huang // ICCV95, – 1995, – pp.666-671.

40 J. Lee “Model-based Analysis of Hand Posture”, / J. Lee, T. Kunii // IEEE Computer Graphics and Applications, Sept., – 1995 – pp. 77- 86.

41 J. Rheg, “Model-Based Tracking of SelfOccluding Articulated Objects”, / J. Rheg, T. Kanade // Proc. of IEEE ICCV95, – 1995 , – pp.63-80,

42 N. Shimada, et al., “Hand Gesture Estimation and Model Refinement Using Monocular Camera-Ambiguity Limitatio by Inequalty Constraints,’’ Proc. of the 3rd Cons On Face and Gesture Recognition, – 1998, – pp. 268-273.

43 Y. Wu, “Capturing Human Hand Motion: A Divide-and-Conquer Approach”, / Y. Wu, T. S. Huang //Proc. of IEEE ICCV99, vol. 1, – 1999,– pp. 606-611.

44 Електронний ресурс gamecareerguide. Режим доступу: http://www.gamecareerguide.com/features/529/what_is_a_game_.php

45 Електронний ресурс Unreal engine. Режим доступу: https://www.unrealengine.com/

46 Електронний ресурс GameMaker Studio. Режим доступу: https://www.yoyogames.com/gamemaker

47 Електронний ресурс Unity3D. Режим доступу: https://unity3d.com/unity

48 Francis Quek, Toward a Vision-Based Hand Gesture Interface, – 1994

49 T.K. Kim. Gesture recognition under small sample size. / T.K. Kim and R. Cipolla //In 8th Asian Conference on Computer Vision (ACCV), volume 4843 of Lecture Notes in Computer Science, – 2007, – pages 335-344.

50 Krueger W.M., Gionfriddo T., Hinrichsen K. Videoplace an artificial reality; Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; San Francisco, CA, USA. 10–13 April 1985; pp. 35–40

51 V.D. Shet. Multi-cue exemplar-based nonparametric model for gesture recognition. / V.D. Shet, V.S.N. Prasad, A.M. Elgammal, Y. Yacoob, and L.S. Davis // In Indian Conference on Computer Vision, Graphics & Image Processing (ICVGIP), – 2004, pages 656-662.

52 C. Kanan. Robust classication of objects, faces, and owers using natural image statistics. / C. Kanan and G. Cottrell //In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), – 2010, pages 2472-2479.

53 F. Parvini. An algorithmic approach for static and dynamic gesture recognition utilising mechanical and biomechanical characteristics. / F. Parvini and C. Shahabi // International Journal of Bioinformatics Research and Applications, 3(1): – 2007, –p. 423.

54 V.I. Pavlovic. Visual interpretation of hand gestures for human-computer interaction: a review./ V.I. Pavlovic, R. Sharma, and T.S. Huang // IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7): –1997, – pp.677- 695

55 B. Swapna. Hand gesture recognition system for numbers using thresholding. / B. Swapna, F. Pravin, and V.D. Rajiv // In 1st International Conference on Computational Intelligence and Information Technology (CIIT), volume 250 of Communications in Computer and Information Science, – 2011, –pp. 782-786.

56 W.T. Freeman. Orientation histograms for hand gesture recognition. / W.T. Freeman and M. Roth // In IEEE International Workshop on Automatic Face- and Gesture- Recognition, –1995, – p. 296-301.

57 K. Symeonidis. Hand gesture recognition using neural networks. Master's thesis, School of Electronic and Electrical Engineering, Centre for Vision, Speech and Signal Processing, Surrey University, –2000.

58 J. Triesch and C. von der Malsburg. Robust classication of hand postures against complex backgrounds. In 2nd International Conference on Automatic Face and Gesture Recognition (AFGR), –1996, – p. 170-175.

59 Licsar and T. Sziranyi. Hand-gesture based lm restoration. In 2nd International Workshop on Pattern Recognition in Information Systems (PRIS), –2002, –p. 95-103.

60 C.C. Chang. New approach for static gesture recognition. / C.C. Chang, J.J. Chen, W.K. Tai, and C.C. Han //Journal of Information Science and Engineering, 22(5): – 2006, – pp. 1047-1057.

61 M.A. Amin and H. Yan. Sign language nger alphabet recognition from GaborPCA representation of hand gestures. In International Conference on Machine Learning and Cybernetics (ICMLC), volume 4, – 2007, p. 2218-2223.

62 D.Y. Huang. Vision-based hand gesture recognition using PCA+Gabor lters and SVM. / D.Y. Huang, W.C. Hu, and S.H. Chang. // In 5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), – 2009, p. 14.

63 S. Gupta. Static hand gesture recognition using local Gabor lter. / S. Gupta, J. Jaafar, and W.F.W. Ahmad // Procedia Engineering, 41: – 2012, –pp. 827-832.

64 H. Meng. Hidden Markov models based dynamic hand gesture recognition with incremental learning method. / H. Meng, F. Shen, and J. Zhao // In International Joint Conference on Neural Networks (IJCNN), – 2014, – pp. 3108-3115

65 C. Hu. Visual gesture recognition for humanmachine interface of robot teleoperation. / C. Hu, M.Q. Meng, P.X. Liu, and X. Wang // In IEEE/RSJ

International Conference on Intelligent Robots and Systems (IROS), volume 2, – 2003, –pp. 1560-1565, 2003.

66 L. Yun and Z. Peng. An automatic hand gesture recognition system based on Viola-Jones method and SVMs. In 2nd International Workshop on Computer Science and Engineering (WCSE), volume 2, – 2009, – pp. 72-76.

67 C. Bekir. Hand gesture recognition. Master's thesis, Dokuz Eylul University, – 2012.

68 Z. Ren. Robust part-based hand gesture recognition using Kinect sensor. / Z. Ren, J. Yuan, J. Meng, and Z. Zhang // IEEE Transactions on Multimedia, 15(5): – 2013, –pp. 1110-1120.

69 R.K. McConnell. Method of and apparatus for pattern recognition, –1986.

70 80R.Z. Khan and N.A. Ibraheem. Hand gesture recognition : a literature. International Journal of Articial Intelligence & Applications, 3(4): –2012, – pp. 161-174.

71 R.-L. Vieriu. On HMM static hand gesture recognition. / R.-L. Vieriu, B. Goras, and L. Goras // In 10th International Symposium on Signals, Circuits and Systems (ISSCS), – 2011, – p. 14.

72 S. Oprisescu. Automatic static hand gesture recognition using ToF cameras. / S. Oprisescu, C. Rasche, and Su Bochao // In 20th European Signal Processing Conference (EUSIPCO), –2021, – pp. 2748-2751.

73 V. Athitsos and S. Sclaro. Boosting nearest neighbor classiers for multiclass recognition. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), – 2005, –p. 45.

74 G.J. Awcock and T. Awcock. Applied Image Processing. McGraw-Hill, – 1995.

75 R.C. Gonzalez and R.E. Woods. Digital Image Processing. Prentice Hall, 2nd edition, – 2002.

76 Pitas. Digital Image Processing Algorithms and Applications. WileyInterscience, – 2000.

77 S.E. Umbaugh. Computer Vision and Image Processing: A Practical Approach Using Cviptools. Prentice Hall, – 1997.

78 J. Huang. Color-spatial image indexing and applications. PhD thesis, Cornell University, – 1998.

79 90Y. Wu and T. S. Huang. Gesture-Based Communication in Human-Computer Interaction, volume 1739 of LNCS, chapter Vision-based gesture recognition: a review, – 1999, – pp. 103-115.

80 Sotiris B Kotsiantis, I Zaharakis, and P Pintelas. Supervised machine learning: A review of classification techniques, – 2007.

81 Marti A. Hearst. Support vector machines. / Marti A. Hearst, Susan T Dumais, Edgar Osuna, John Platt, and Bernhard Scholkopf // IEEE Intelligent Systems and their Applications, 13(4) –1998, –pp. 18–28.

82 Leo Breiman. Random forests. Machine learning, 45(1) –2001, – pp. 5–32.

83 Liwei Liu. Hand posture recognition using finger geometric feature. / Liwei Liu, Junliang Xing, Haizhou Ai, and Xiang Ruan // In Pattern Recognition (ICPR), 2012 21st International Conference on, –2012, – pp. 565–568.

84 Chenyang Zhang and Yingli Tian. Edge enhanced depth motion map for dynamic hand gesture recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, –2013, –pp. 500–505.

85 Bhumika Gupta. K-nearest correlated neighbor classification for indian sign language gesture recognition using feature fusion. / Bhumika Gupta, Pushkar Shukla, and Ankush Mittal // In Computer Communication and Informatics (ICCCI), 2016 International Conference on, –2016, –pp. 1–5.

86 Alaa Tharwat. Sift-based arabic sign language recognition system. / Alaa Tharwat, Tarek Gaber, Aboul Ella Hassanien, MK Shahin, and Basma Refaat // In Afro-european conference for industrial advancement, –2015, – pp. 359–370.

87 J Rekha. Shape, texture and local movement hand gesture features for indian sign language recognition. / J Rekha, J Bhattacharya, and S Majumder // In Trendz in Information Sciences and Computing (TISC), 2011 3rd International Conference on, –2011, – pp. 30–35.

88 Neha Baranwal and GC Nandi. An efficient gesture based humanoid learning using wavelet descriptor and mfcc techniques. International Journal of Machine Learning and Cybernetics, 8(4) –2017, –pp. 1369–1388.

89 Maxime Devanne. 3-d human action recognition by shape analysis of motion trajectories on riemannian manifold. / Maxime Devanne, Hazem Wannous, Stefano Berretti, Pietro Pala, Mohamed Daoudi, and Alberto Del Bimbo // IEEE transactions on cybernetics, – 2014 – pp. 1340-1352.

90 Natalia Neverova. Multi-scale deep learning for gesture detection and localization. / Natalia Neverova, Christian Wolf, Graham W Taylor, and Florian Nebout // In Workshop at the European conference on computer vision, – 2014, – pp. 474–490.

91 Lawrence Rabiner and B Juang. An introduction to hidden markov models. ieee assp magazine, 3(1) –1986, –pp. 4–16.

92 Julie A. Jacko. Human–Computer Interaction Handbook (3rd Edition). – 2012

93 20Електронний ресурс MP4RA. Режим доступу: http://mp4ra.org/index.html#/

94 30M. W. Bern and P. E. Plassmann, Mesh generation. Pennsylvania State University, Department of Computer Science and Engineering, College of Engineering – 1997.

95 Електронний ресурс Anatomy of a mesh. Режим доступу: https://docs.unity3d.com/

96 P. W. Brand and A. Hollister, Clinical mechanics of the hand. Mosby Incorporated, – 1999.

97 James Steven Supancic III. Depth based hand pose estimation: methods, data, and challenges. arxiv preprint. / James Steven Supancic III, Gregory Rogez, Yi Yang, Jamie Shotton, and Deva Ramanan // arXiv preprint arXiv:1504.06378, – 2015.

98 Stephen J Schmugge. Objective evaluation of approaches of skin detection using roc analysis. / Stephen J Schmugge, Sriram Jayaram, Min C Shin, and Leonid V Tsap // Computer Vision and Image Understanding, 108(1): – 2007, pp. 41–51.

99 Hironori Takimoto. A robust gesture recognition using depth data. / Hironori Takimoto, Jaemin Lee, and Akihiro Kanagawa // International Journal of Machine Learning and Computing, 3(2): –2013, –p. 245.

100 Poonam Suryanarayan. Dynamic hand pose recognition using depth data. / Poonam Suryanarayan, Anbumani Subramanian, and Dinesh Mandalapu // In Pattern Recognition (ICPR), 2010 20th International Conference on, –2010, –pp. 3105– 3108.

101 Alexey Kurakin. A real time system for dynamic hand gesture recognition with a depth sensor. / Alexey Kurakin, Zhengyou Zhang, and Zicheng Liu // In Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European, – 2012, – pp. 1975–1979.

102 Eva Kollorz. Gesture recognition with a time-of-flight camera. / Eva Kollorz, Jochen Penne, Joachim Hornegger, and Alexander Barke // International Journal of Intelligent Systems Technologies and Applications, 5(3- 4): –2008, – pp. 334–343.

103 Jamie Shotton. Realtime human pose recognition in parts from single depth images. / Jamie Shotton, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, and Richard Moore // Communications of the ACM, 56(1): –2013, – pp. 116–124.

104 Jonathan Tompson. Real-time continuous pose recovery of human hands using convolutional networks. / Jonathan Tompson, Murphy Stein, Yann Lecun, and Ken Perlin // ACM Transactions on Graphics (ToG), –2014

105 S. Ioffe. Batch normalization: Accelerating deep network training by reducing internal covariate shift. / S. Ioffe and C. Szegedy // arXiv preprint arXiv:1502.03167, – 2015

106 K. He, Deep residual learning for image recognition. / X. Zhang, S. Ren, and J. Sun // In Proceedings of the IEEE conference on computer vision and pattern recognition, – 2016, – pp. 770–778.

107 Електронне посилання на набір даних Open Images. Режим доступу: https://github.com/openimages.

108 Sinno Jialin Pan. A survey on transfer learning. / Sinno Jialin Pan and Qiang Yang // IEEE Transactions on knowledge and data engineering, 22(10) –2010, – pp. 1345–1359

109 A.K. Musa. Signature recognition and verication by using complex-moments characteristics. Master's thesis, University of Baghdad, – 1998

110 Електронний ресурс мови програмування Python. Режим доступу: https://www.python.org/

111 Електронний ресурс бібліотеки Tensorflow. Режим доступу: https://www.tensorflow.org/

112 Електронний ресурс YAML. Режим доступу: https://yaml.org/

113 Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, – 2012, – p. 1097–1105,.

114 Andrew G. Howard, Weijun Wang. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications

115 Mark Sandler, MobileNetv2: Inverted Residuals and Linear Bottlenecks. / Andrew Howard, Menglong Zhu, Andrey Zhmoginov, Liang-Chieh Chen // arXiv:1801.04381v4 – 2019

116 Y. Jia. Caffe: Convolutional architecture for fast feature embedding. / Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell // arXiv preprint arXiv:1408.5093, – 2014.

Додатки

Додаток 1. Акт впровадження результатів дослідження

Науково-технічна фірма
"ІНФОСЕРВІС"

м.Хмельницький,
вул. Зарічанська, 3/2 офіс № 404
ЄДРПОУ 14169783
☎ (0382) 79-40-57

"14" січня 2021 р. № 1

ДОВІДКА

про впровадження результатів дисертаційної роботи «Моделювання та розпізнавання жестів української дактильної абетки за допомогою кросплатформених технологій» Кондратюка Сергія Сергійовича

Результати дисертаційної роботи «Моделювання та розпізнавання жестів української дактильної абетки за допомогою кросплатформених технологій» Кондратюка Сергія Сергійовича:

- реалістична високополігональна параметрична тривимірна модель руки зі скелетом зі ступенями свободи, який описує природні обмеження у рухах усіх частин руки
- методи та алгоритми розпізнавання жестів української дактильної абетки за допомогою згорткових нейронних мережей із тривимірними згортками
- кросплатформена технологія моделювання та розпізнавання дактилем української дактильної абетки за зображеннями із камери та без додаткового обладнання чи маркерів

знайшли застосування в ТОВ «Науково-технічна фірма «Інфосервіс».

Зазначені результати використовуються при розробці програмного забезпечення для: відтворення жестів за допомогою просторової моделі руки людини; розпізнавання жестів української дактильної абетки з відеопотоку з використанням моделей глибоких нейромереж; аналізу просторових рухів руки людини. Методи та алгоритми, запропоновані у роботі, дозволяються покращити ефективність програмного забезпечення, що розробляється у ТОВ «Науково-технічна фірма «Інфосервіс».

Директор ТОВ "Науково-технічна фірма "Інфосервіс" Павлишин В.В.