Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Інформаційно-аналітичних технологій та менеджменту
(повна назва)
Кафедра Інформатики
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти перший (бакалаврський)

РОЗРОБКА ЗАСТОСУНКУ ДЛЯ МОНІТОРИНГУ ДІЙ УЧАСНИКІВ КОНФЕРЕНЦІЇ ZOOM
(тема)

Виконав:
студент 4 курсу, групи ІТІНФ-19-1

Єременко І.О.
(прізвище, ініціали)

Спеціальності 122 Комп'ютерні науки
(код і повна назва спеціальності)

Тип програми освітньо-професійна

Освітня програма Інформатика
(повна назва освітньої програми)

Керівник доц. Яковлева О.В.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри _____ Кобилін О.А.
(підпис) (прізвище, ініціали)

2023 р.

Харківський національний університет радіоелектроніки

Факультет Інформаційно-аналітичних технологій та менеджменту
(повна назва)

Кафедра Інформатики
(повна назва)

Рівень вищої освіти перший (бакалаврський)

Спеціальність 122 Комп'ютерні науки
(код і повна назва)

Тип програми освітньо-професійна

Освітня програма Інформатика
(повна назва освітньої програми)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
«____»_____ 2023 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Єременку Івану Олексійовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Розробка застосунку для моніторингу дій учасників конференції Zoom

затверджена наказом університету від 15 травня 2023 року № 474 Ст

2. Термін подання студентом роботи до екзаменаційної комісії 29 травня 2023 р.

3. Вихідні дані до роботи аналіз інтернет джерел, методи детектування, розпізнавання облич, аналізу емоцій, нейронна мережа SSD, моделі VGG-Face, Facenet, Facenet512, OpenFace, DeepFace, ArcFace, Zoom Video SDK, C++, WPF, C#, SQL Server, Python, OpenCV.

4. Перелік питань, що потрібно опрацювати в роботі
1. Отримання кадрів з відеопотоків конференції Zoom Video SDK.
2. Огляд методів детектування, розпізнавання облич, аналізу емоцій.
3. Розробка архітектури та реалізація застосунку для моніторингу дій учасників конференції Zoom.

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (п.5 включається до завдання за рішенням випускової кафедри) Актуальність, постановка задачі, особливості створення застосунків для конференції Zoom, детектування облич, верифікація облич, розпізнавання емоцій, специфікація вимог до застосунку, розробка застосунку, ілюстрація роботи застосунку.

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

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата
Консультант з дотримання діючих стандартів та норм	Доцент Творошенко І.С.		

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Отримання завдання на кваліфікаційну роботу	10.04.2023	
2	Аналіз завдання, підбір літератури	11.04.23-17.04.23	
3	Аналіз літератури з досліджуваної проблеми	18.04.23-20.04.23	
4	Огляд ПО та інструментарію для розробки	21.04.23-25.04.23	
5	Моделювання застосунку	25.04.23-27.04.23	
6	Програмна реалізація	28.04-25.05.23	
7	Оформлення пояснювальної записки	26.05.23-30.05.23	
8	Перевірка на плагіат	31.05.23	
9	Рецензування	1.06.23	
10	Підготовка презентації та доповіді	1.06.23-04.06.23	
11	Занесення роботи в електронний архів	07.06.23	
12	Попередній захист кваліфікаційної роботи	07.06.23	

Дата видачі завдання 10 квітня 2023 р.

Студент _____
(підпис)

Керівник роботи _____ _____доц. Яковлева О.В._____
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ/ABSTRACT

Пояснювальна записка до кваліфікаційної роботи: 60 с., 11 табл., 28 рис., 40 джерел.

ZOOM VIDEO SDK, ДЕТЕКТУВАННЯ ОБЛИЧЧЯ, OPENCV, АНАЛІЗ ЕМОЦІЙ, ВЕРИФІКАЦІЯ ПЕРСОНИ, DEEPFACE, ГЛИБОКІ НЕЙРОННІ МЕРЕЖІ.

Об’єктом роботи є питання моніторингу дій та стану учасників онлайн конференцій.

Метою роботи є розробка застосунку для моніторингу дій учасників конференції Zoom.

Було використано методи глибокого навчання для дослідження методів детекції обличчя, аутентифікації особи, аналізу емоцій, спираючись на відеопотоки, отримані за допомогою Zoom SDK.

У результаті роботи здійснена програмна реалізація системи для моніторингу дій учасників відеоконференії.

ZOOM VIDEO SDK, FACE DETECTION, OPENCV, EMOTION ANALYSIS, PERSON VERIFICATION, DEEPFACE, DEEP NEURAL NETWORKS.

The object of the work is the issue of monitoring the actions and state of online video conference participants.

The purpose of this work is to develop a service for monitoring the actions of video conference participants, based on data provided by the Zoom SDK.

Deep learning methods were used to investigate methods of face detection, person authentication, and emotion analysis based on video streams obtained through the Zoom SDK.

As a result of the work, a software implementation of the system for monitoring the actions of video conference participants has been carried out.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

ШІ – штучний інтелект

SVM – Support Vector Machine (метод опорних векторів)

CNN – Convolutional Neural Network (згорткова нейронна мережа)

YOLO – You Only Look Once (дивишся тільки раз)

CPU – Central Process Unit (центральний процесор)

GPU – Graphics Process Unit (графічний процесор)

SSD – Single Shot MultiBox Detector (однокадрове багатоблочне виявлення)

MTCNN – Multi-task Cascaded Convolutional Network (багатозадачна каскадна згорткова мережа)

ВСТУП

COVID-19 суттєво вплинув на розвиток та використання відеоконференцій в освітній сфері. За період пандемії COVID-19, коли в Україні та багатьох інших країнах світу були запроваджені карантинні обмеження, відеоконференції стали найефективнішим інструментом для організації дистанційного навчання. Відеоконференції дозволяють вчителям та студентам продовжувати навчання, не залежно від місця перебування, та взаємодіяти між собою, ділитися матеріалами та знаннями.

Одним з найбільших викликів в дистанційному навчанні є відсутність прямого контакту між викладачем та студентами. Це може призводити до випадків шахрайства під час іспитів, коли інша особо буде складати іспит за студента. Навіть при ввімкнутих камерах при великій кількості студентів викладачу не зручно слідкувати за всіма на протязі всього іспиту. Також під час дистанційного навчання у студентів швидше розсіюється увага та знижується мотивація.

Наразі не існує автоматизованих систем, які б допомогали викладачам слідкувати за присутністю необхідних студентів в кадрі під час екзаменів, а також оцінювали їх емоції протягом занять.

Актуальність роботи полягає в зменшенні вірогідності написання іспиту іншим студентом, без потреби викладача витрачати час на перевірки на протязі усього екзамену за допомого технології верифікації особи по зображенню. А також у допомозі вчителям підтримувати відповідний рівень зосередженості та комунікації за допомогою технології аналізу емоцій.

В ході роботи буде розроблений сервіс моніторингу дій студентів під час відеоконференції за допомогою обробки кадрів, які надходять з камер студентів в режимі реального часу.

1 АКТУАЛЬНІСТЬ ПИТАННЯ МОНІТОРИНГУ ДІЙ УЧАСНИКІВ ОНЛАЙН КОНФЕРЕНЦІЙ

1.1 Сучасний стан питання онлайн навчання

Питання онлайн навчання є надзвичайно актуальним у сучасному світі та має значний вплив на освітні процеси та розвиток людства в цілому. За останні кілька років онлайн навчання значно розширило свої можливості, завдяки швидкому розвитку технологій та збільшенню доступу до інтернету. В даний час онлайн курси стали доступними для мільйонів людей у всьому світі, що дає можливість здобувати освіту в будь-який час та в будь-якому місці.

Пандемія COVID-19 та війна в Україні призвели до переходу більшості навчальних закладів на дистанційну форму навчання. Онлайн навчання стало єдиним доступним варіантом навчання для багатьох студентів, що забезпечило продовження освіти та досліджень у важких умовах війни та карантину. Це стало великим викликом для системи освіти та вимагало відповідних змін у підходах до навчання та організації навчального процесу.

Таким чином, питання онлайн навчання є надзвичайно актуальним і вимагає не тільки забезпечення всіх студентів інтернетом та обладнанням, а також продовження розвитку технологічних та методичних підходів для проведення занять та контролю знань.

1.2 Огляд у порівняльному аспекті сервісів для проведення онлайн конференцій

Microsoft Teams – це застосунок для відеоконференцій, який може бути відмінним вибором для великих компаній. Цей інструмент входить до складу

Microsoft 365. Зовнішні учасники, які користуються Microsoft 365 можуть долучатися до зустрічей без необхідності завантаження Teams. Також користувачі можуть легко обмінюватись електронними листами та додактами до них, використовуючи цей застосунок. Багато програм Microsoft, включачи Outlook та Office 365, пов’язані з застосунком. Teams мають додаткові функції, такі як розмивання фону, демонстрація екрану, запис дзвінків, можливість підняття руки, покращена якість шумозаглушення, корисні боти та додатки до чату, пошук файлів, резервне копіювання. З обмежень: велике споживання пам’яті, не завжди приходять повідомлення, обмежена кількість каналів.

Google Meet – інструмент відеоконференцій, який був розроблений спеціально для задоволення потреб бізнесу будь-якого розміру в проведенні відеозустрічей. Для використання цього програмного забезпечення потрібний обліковий запис Google. Застосунок сумісний з різноманітними продуктами Google, наприклад Google Chat. Доступні функції: демонстрація екрану, запис дзвінка, повноекранний перегляд, включення субтитрів, налаштування розташування елементів та інше. Користувачі можуть приєднуватися за допомогою різних методів: спільні електронні листи, посилання або запрошення у календарі. Програма підтримує проведення великих зустрічей з максимально кількістю учасників до 250 осіб на кожний дзвінок, а також пряму трансляцію для до 100000 глядачів в межах домену. З додактових плюсів: не потрібно завантажувати програмне забезпечення, зустрічі можна записувати безпосередньо на Google Drive. З мінусів: немає можливості передавати мультимедійні документи, споживає значну кількість апаратних ресурсів.

Zoom – програмне забезпечення для відеоконференцій, яке було розроблене для допомоги корпораціям, але зараз є однією з найбільш популярних програм відеодзвінків впринципі. Програма дозволяє проводити великі зустрічі до 1000 учасників з 49 учасниками на екрані. Zoom дозволяє користувачам легко записувати зустрічі та надавати доступ до них тим, хто не був присутній. Під час участі в зустрічах можна також обмінюватися файлами.

Існували значні проблеми з безпекою, однак з ростом популярності продукту, компанія доклала великих зусиль, щоб вирішити ці питання. Додатковими плюсами є підтримка інтеграції з Google calendar, Facebook, DropBox і багато інших програм від сторонніх розробників. Також Zoom пропонує такі елементи залучення, як підняття руки, опитування, показ екрана, невербальний відгук, великий діапазон можливостей контролю над відео. Проблемою є загроза вторгнення незнайомих на відеоконференцію, а також дуже велика кількість підписок та доповнень, що займає час на те, щоб в них розібратися. Порівняння сервісів наведено в таблиці 1.1.

Таблиця 1.1 – Порівняння сервісів для проведення відеоконференцій

№	Параметри	Teams	Google Meet	Zoom
1	**2**	**3**	**4**	**5**
1	Максимальна кількість учасників	Безкоштовна версія: 100 Найдорожчка підписка: 300	Безкоштовна версія: 100 Найдорожчка підписка: 250	Безкоштовна версія: 100 Найдорожчка підписка: 1000
2	Обмеження по часу конференцій	Безкоштовна версія: 60 хвилин Найдорожчка підписка: 30 годин	Безкоштовна версія: 60 хвилин Найдорожчка підписка: 24 години	Безкоштовна версія: 40 хвилин Найдорожчка підписка: 30 годин
3	Можливість запису	Тільки за наявності підписки	Тільки за наявності підписки	Доступна для всіх планів
4	Локальне зберігання записів	Відсутнє	Відсутнє	Присутнє для PC
5	Якість відео	Для всіх планів: 1080p	Для всіх планів: 1080p	Безкоштовний план: 720p Платний: 1080p
6	Хмарне сховище для записів	При найдорожому плані 1 тб на організацію та 10 гб на окрему ліцензію	Необмежене сховище для найдорожчого плану	Необмежене сховище для найдорожчого плану

Продовження таблиці 1.1

1	2	3	4	5
7	Можливість розділення сесії на підгрупи	Доступна при наявності підписки	Доступна при наявності підписки	Доступна при наявності підписки
8	Кількість сесій	Необмежена для всіх планів	Необмежена для всіх планів	Необмежена для всіх планів
9	Віртуальний фон	Доступний для всіх планів	Доступний для всіх планів	Доступний для всіх планів
10	Можливість показувати екран	Доступна для всіх планів	Доступна для всіх планів	Доступна для всіх планів
11	Найдорожча підписка на користувача за рік	150 доларів	216 доларів	200 доларів

Кожен з перелічених сервісів відеоконференцій має свої переваги та недоліки, саме Zoom було обрано через найбільшу максимальну кількість учасників, які можуть брати участь у конференції.

1.3 Особливості створення застосунків для конференції Zoom

Інтегрувати Zoom конференції у свій застосунок можна за допомогою таких рішень: Meeting та Video SDK. Meeting SDK є комплектом інструментів для розробників, який використовує безпосередньо інтерфейс Zoom і дозволяє вбудувати досвід зустрічей Zoom в такі платформи: Android, iOS, macOS, Web, Windows. Надається мало можливостей модифікувати інтерфейс користувача, але це приносить перевагу в тому, що витрачається менше часу на інтеграцію відеоконференцій у проєкт та звикання кінцевого користувача до інтерфейсу, адже майже кожен знайомий з Zoom.

Video SDK не надає інтерфейс користувача, замість цього дозволяє створювати будь-який інтерфейс користувача, в залежності від того, як саме вам треба використовувати відео [1]. Також при використанні Video SDK неможливо приєднатися до звичайних Zoom конференцій, так як для Video SDK створюються окремі конференції, які проходять через інші сервери Zoom. Також окрім інтерфейсу повна свобода надається при роботі з відео/аудіо потоками учасників мітингу.

Обидва SDK надають можливість отримувати відео дані кожного учасника окремими кадрами у форматі YUV420 [2]. Аудіо дані є можливість отримувати від кожного учасника окремо, або аудіо всієї зустрічі, тобто те, що чують учасники. Різниця полягає в тому, що у Video SDK всі учасники конференції можуть отримувати відео та аудіо, так як разом з Video SDK не надається відповідний клієнт Zoom, який відповідає за рендерінг відео учасників. Натомість Meeting SDK використовує клієнт Zoom, тому можливість записувати відео та аудіо отримують тільки ти учасники, яким надав відповідне право власник зустрічі, або які отримали необхідні права за допомогою протоколу авторизації OAuth.

Отримувати необроблені дані за допомогою Meeting SDK є можливим тільки для macOS та Windows. Необроблені дані для Video SDK є можливим отримувати на всі платформи, які підтримують Video SDK з обмеженням для Web: не більше ніж для 25 учасників одночасно. Video SDK надає функціонал не тільки для отримання необроблених даних, а також для передачі їх таким стрімінговим платформам як Facebook Live, YouTube Live і т.д.

Окрім цього у Video SDK є можливість налаштувати параметри якості відео між роздільною здатністю та частотою кадрів при обмеженій пропускній здатності мережі. При задовільній пропускній здатності користувач буде отримувати відео найкращої якості .

Таким чином, Video SDK доцільніше вибирати для досліджень в області моніторингу дій учасників через більш багатий функціонал по роботі з відео потоками, або для комерційних рішень, які потребують свого інтерфейсу та

додаткового функціоналу для відеоконференцій. В той час як Meeting SDK краще підійде маленьким комерційним рішенням, яким швидко треба інтегрувати Zoom конференцій у свій проєкт на великій кількості платформ.

1.4 Сучасний прогрес в вирішенні задач комп'ютерного зору

Комп'ютерний зір – сфера штучного інтелекту, яка дозволяє комп'ютерам отримувати корисну інформацію з цифрових зображень, відео та інших візуальних даних і виконувати або пропонувати дії у відповідь на цю інформацію. Якщо штучний інтелект дає комп'ютерам здатність мислити, то комп'ютерний зір дає їм можливість бачити, спостерігати та розуміти.

Поштовхом для розвитку комп'ютерного зору стали експерименти у 1950-1960-их роках, коли нейрофізіологи показували кішці низку зображень, намагаючись співвіднести реакцію в її мозку. Озброївшись отриманими результатами, вчені зосередилися на відтворенні людських неврологічних структур у цифровій формі і незабаром у 1974 році була розроблена система розпізнавання символів, яка могла розпізнавати текст, надрукований будь-яким шрифтом або гарнітурою [3].

Досягнення в дослідженнях і розробках глибокого навчання і згорткових глибинних нейронних мереж з 2012 року призвели до значного прогресу в області комп'ютерного зору [4]. Завдяки вдосконаленому апаратному забезпеченню для запуску алгоритмів і доступності великої кількості даних для навчання, методи на основі глибокого навчання вважаються найпоширенішими та найефективнішими алгоритмами для обробки багатьох завдань комп'ютерного зору. Ключова відмінність між традиційними методами, які використовують створені вручну функції і методами на основі глибокого навчання полягає в тому, що останні здатні вивчати особливості вхідних зображень наскрізним способом без необхідності вилучення ознак (рис. 1.1).

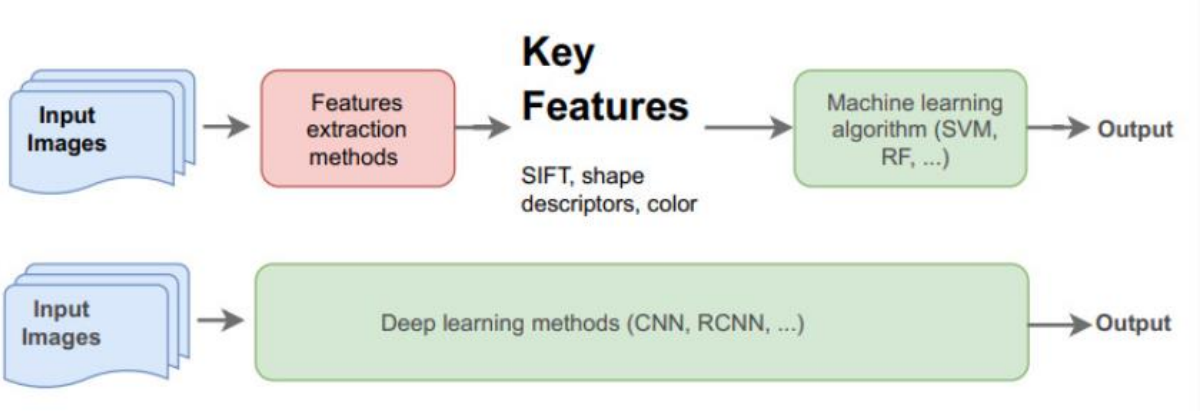


Рисунок 1.1 – Різниця між традиційними методами комп'ютерного зору та методами глибокого навчання [4]

Класифікація об'єктів використовується у багатьох галузях, наприклад під час планових медичних діагностиках серця, мозку, легенів і інших органів комп'ютерний зір класифікує наявність захворювань, що дозволяє відсіяти велику кількість здорових людей при цьому не витрачаючи дорогий час лікарів. Також класифікація зображень є невід'ємною частиною систем автономного водіння. Використовується для виявлення та класифікації об'єктів навколо транспортного засобу, таких як інші автомобілі, пішоходи, дорожні знаки тощо. Використання автоматичних систем, які вміють класифікувати зображення пов'язане з тим, що вони визначають об'єкти на зображенні з дуже високою точністю, наближеною до людської, а іноді з ще кращою точністю, наприклад, під час дослідження передачі навчання з використанням згорткових нейронних мереж для завдань класифікації медичних зображень. А саме використовуючи InceptionV3, натренованою вагами з ImageNet і перенесення навчання на датасет з медичними зображеннями, який містить 108312 зображень оптичної когерентної томографії вдалося досягти accuracy в 96,6%, sensitivity 97,8%, specificity 97,4% [5].

Класифікація часто поєднується з детектуванням, задача якого полягає у визначенні наявності об'єктів на зображені та їх локалізації. Прикладом комбінування класифікації і детектування є система, яка спочатку визначає, чи

є обличчя на зображенні, а потім визначає до якої категорії воно належить, наприклад, до категорії чоловік чи жінка.

Значних успіхів у детектуванні об’єктів було досягнуто саме за останні 20 років. Одним з успіхів вважається виявлення маленьких об’єктів на великих сценах. Так у 2015 році був впроваджений Single Shot Multibox Detector, який використовує одну згорткову нейронну мережу для передбачування розташування областей і їх класів, без застосування другого етапу класифікації [6]. На виході нейронної мережі формується декілька тисяч прогнозів для можливих регіонів розташування об’єктів різної форми на різних масштабах, після цього відбувається вибір декількох найбільш вірогідних областей. Можливими застосуваннями є підрахунок кількості людей в толпі, або тварин на відкритому повітрі, а також виявлення військових цілей на зображеннях супутника

Дуже затребуваним є виявлення та відстеження об’єктів у реальному часі для систем відоспостереження та автономного водіння. Традиційні детектори об’єктів зазвичай розробляються для виявлення об’єктів на зображення, не враховуючи зв’язки між кадрами відео.

Одна з нейронних мереж, яка ефективно працює з відео – YOLO [7]. Використовуючи YOLO на CPU можна обробляти до 30 кадрів на секунду, якщо підключити GPU та використовувати YOLOv5n, то можна опрацьовувати до 230 кадрів на секунду, що є дуже вражаючим результатом. Як видно з назви цього детектору для виявлення об’єкта потрібно лише один раз пройти через нейронну мережу. Поперше зображення розбивається на сітку з великою кількістю комірок.

В кожній клітинці будуть виявлятися об’єкти, які з’являються всередині неї. Наприклад, якщо центр об’єкта з’являється в певній комірці, то ця комірка відповідатиме за виявлення цього об’єкта.

Після цього YOLO передбачує висоту, ширину, центр об’єкта та його клас. Нарешті YOLO встановлює відповідність між реальними та передбачуваними рамками (рис. 1.2).

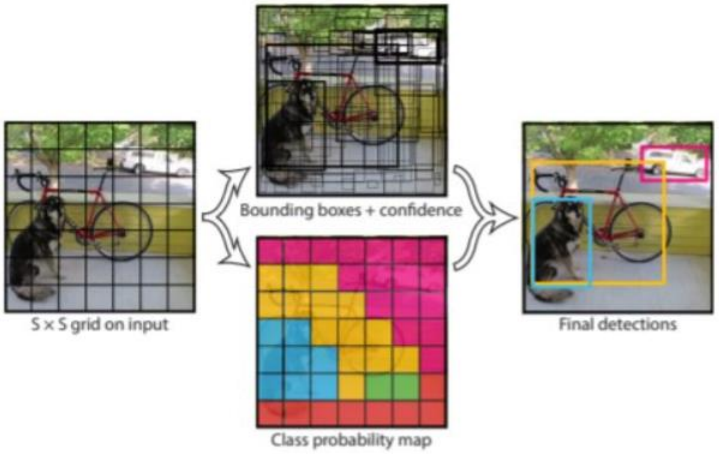


Рисунок 1.2 – Демонстрація трьох технік, які використовує YOLO [7]

Оцінка 3D моделі за 2D зображенням є фундаментальним завданням для комп’ютерного зору, хоч люди і інстинктивно сприймають сцени за одним зображенням, але одночасне врахування геометрії та семантики є дуже складним завданням, що вивчається дисятиліттями. Можливість передбачення тривимірного середовища за допомогою зображення відкриває нові можливості застосування. А саме поліпшується застосування віртуальної реальності, використання в іграх тощо. Відтворення тривимірних об’єктів з тривимірних даних, таких як хмари точок, або зображення глибини, зазвичай є більш точним, ніж відтворення з двовимірних зображень, оскільки тривимірні дані надають більше інформації про форму та геометрію об’єкта. Однак, сенсори для збору цих даних часто є більш коштовними, менш компактними, аніж камери, що використовуються у смартфонах, дронах, транспортних засобах та інших пристроях.

Розвиток глибокого навчання спонукав вчених до вирішення проблем, пов’язаних з розумінням тривимірних об’єктів з двовимірних зображень. Серед них методи, які виконують відтворення тривимірної сцени з декількох зображень або відео, досягли хороших результатів. Однак, якщо доступне тільки одне двовимірне зображення, якість відтворення є незадовільною через розмірність, що втрачається при проекції з тривимірного простору на площину двовимірного зображення. Крім того, більшість існуючих досліджень відтворюють об’єкти на зображенні як одну мережу, не розділяючи окремі

об’єкти в сцені. В результаті, ці відтворені сцени не підходять для використання у віртуальній реальності або іграх, оскільки не дозволяють ідентифікувати та взаємодіяти з окремими об’єктами. Тому відтворення 3D сцени з урахуванням окремих об’єктів розглядається для створення взаємодії з 3D сценою.

Корейським дослідникам з університету Донгук в Сеулі вдалося створити мережу реконструкції 3D сцени, яка враховує об’єкти, оцінює позицію камери, 3D макет, позу об’єкту в 3D та його форму [8]. Вони запропонували використовувати мережу генерації характеристик глибини на етапі уточненої оцінки, щоб вирішити проблему невизначеності глибини в розумінні зображень 2D у 3D. Запропонували використання мультизадачного навчання для мереж реконструкції сіток, щоб отримати більш повні сітки. Дослідження проводилося на реальних наборах даних – SUN RGB-D та Pix3D і було порівняно з іншими сучасними методами. Експериментальні результати на наборі даних SUN RGB-D показали, що цей метод покращив оцінку огороджуючого паралелепіпеда 3D для більшості категорій об’єктів. Якість реконструкції сітки на наборі даних Pix3D показала, що запропонована мережа реконструкції сітки на основі мультизадачного навчання може бути корисною для отримання повної форми об’єкта (рис. 1.3). Одне з обмежень цього дослідження полягає в тому, що поточні мережі реконструкції можуть оцінювати лише форми об’єктів, що входять до вивчених категорій.

1.5 Існуючі бібліотеки та програмні засоби для роботи із зображеннями

Багато задач комп’ютерного зору вже вирішені і їх реалізації зібрані у відкриті бібліотеки, більшістю з яких можуть користуватися навіть програмісти, які не є спеціалістами в області комп’ютерного зору.

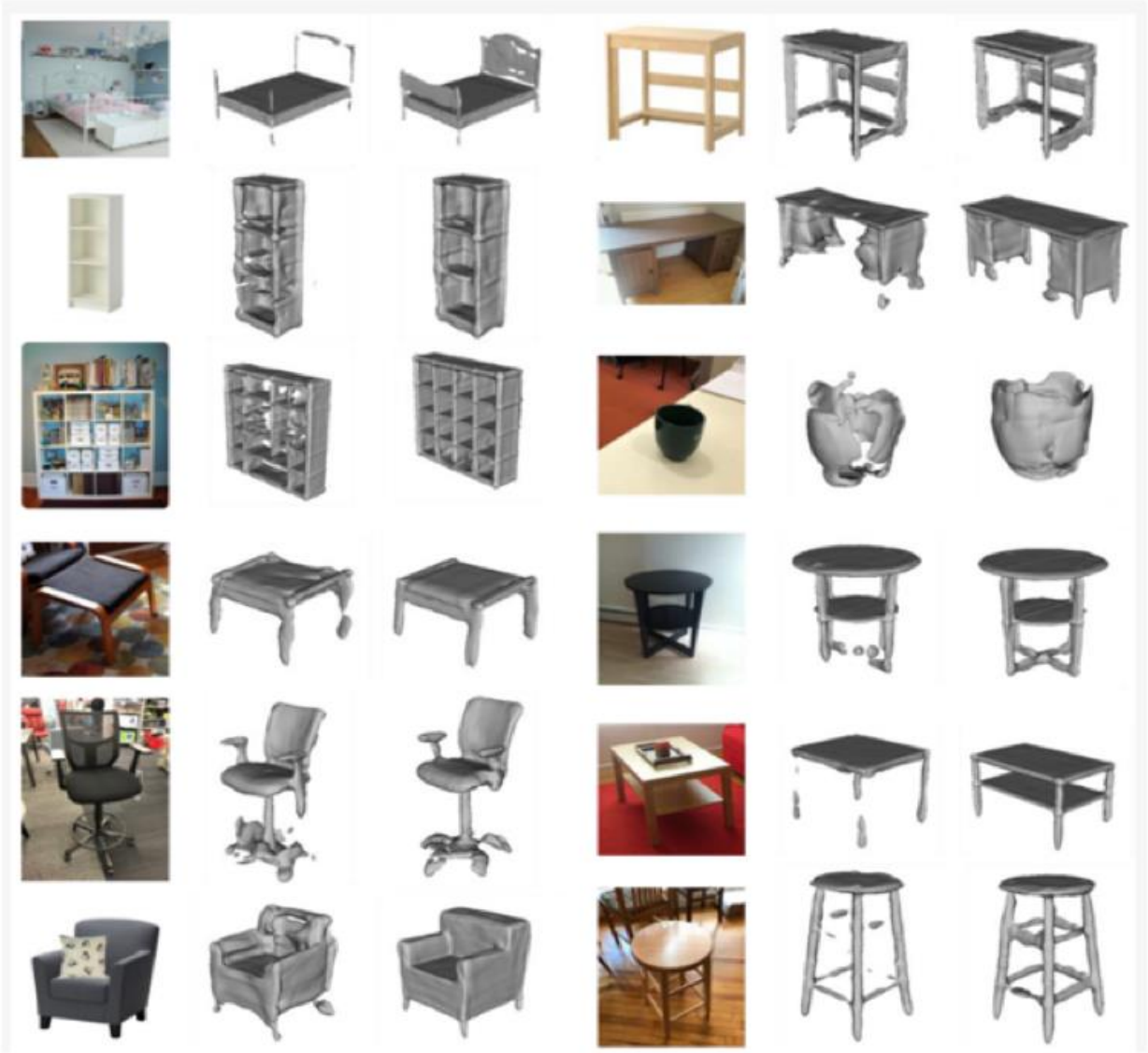


Рисунок 1.3 – Результати реконструкції для набору даних Pix3D:
(а, г) вхідні зображення; (б, г) результат методу Implicit3D; (в, е) результат методу, запропонованого у дослідженні [8]

Однією з найстаріших і найпопулярніших відкритих бібліотек для роботи із зображеннями є OpenCV. Перший альфа-реліз компанія Intel випустила ще у 2000 році [9]. Бібліотека є мультиплатформною, підтримує Windows, Android, Linux, macOS, і може використовуватись різними мовами програмування: C++, Python, Java, тощо. OpenCV може вирішувати такі задачі: обробка зображень, а саме їх фільтрація, геометричні перетворення, обробка кольору, розпізнавання жестів, обличчя, розуміння рухів, детектування об’єктів, сегментація, видалення червоних очей, тощо.

TensorFlow є однією з найпопулярніших платформ машинного навчання з відкритим кодом і з повним набором інструментів, ресурсів і бібліотек [10]. TensorFlow дозволяє користувачам розробляти моделі машинного навчання, пов’язані з комп’ютерним зором для таких завдань, як розпізнавання облич, класифікація зображень, виявлення об’єктів тощо. Має індивідуальні рішення, такі як TensorFlow.js – бібліотека JavaScript для навчання та розгортання моделей у браузері та на Node.js. TensorFlow Lite – полегшена бібліотека для розгортання моделей на мобільних пристроях.

DeepFace – бібліотека, основним напрямком якої є обробка обличчя, а саме: детектування, аналіз емоцій, верифікація, стать, вік. Підтримує такі моделі розпізнавання облич: Facebook DeepFace, VGG-Face, Google FaceNet, OpenFace, DeepID, Dlib, ArcFace. Доступні детектори обличчя: OpenCV, Dlib, SSD, MTCNN, RetinaFace. У 2014 році DeepFace досягнув точності в 97,35% на знаменитому датасеті LFW benchmark, наблизившись до людської продуктивності у 97,53% [11]. Це вдалося завдяки тренуванню 9 шарової моделі на 4-ох мільйонах зображень облич.

Caffe – фреймворк, написаний на C++ дослідниками з університету Берклі. Він підтримує популярні алгоритми глибокого навчання, такі як CNN, RCNN, LSTM. Найбільше підходить для проєктів, пов’язаних із класифікацією та сегментацією зображень. Через високу швидкість Caffe доцільно використовувати для відстеження об’єктів у реальному часі.

1.6 Постановка задачі

Таким чином, питання моніторингу дій учасників онлайн конференцій є актуальною задачею, вирішення якої допомагає організатору верифікувати учасників, оцінити їх зацікавленість та зібрати статистичну інформацію для подальших досліджень взаємозв’язку між поведінкою під час конференції та цільовим результатом конференції, або серії конференцій.

Об’єктом роботи є питання моніторингу дій та стану учасників онлайн конференцій.

Метою роботи є розробка застосунку для моніторингу дій учасників конференції Zoom.

Для досягнення цієї мети необхідно вирішити такі завдання:

– дослідити сучасний стан онлайн навчання та сервісів для проведення онлайн конференцій;

– розглянути існуючі SDK конференції Zoom, проаналізувати можливості доступу до відеопотоку конференцій Zoom, інших подій конференції, роботи на різних платформах;

– розглянути методи розпізнавання облич та емоцій, дослідити навчені нейромережеві моделі для вирішення цих задач;

– розглянути бібліотеки для роботи із зображенням та відео;

– спроєктувати архітектуру застосунку для моніторингу дій учасників конференції, який повинен фіксувати час під’єднання та від’єднання учасників, факт підняття руки, вмикання мікрофону, проводити верифікацію учасників та розпізнавати їх емоції;

– розробити інтерфейс для організатора конференції та інших учасників;

– реалізувати спроєктований застосунок та провести його тестування;

– зробити висновок щодо роботи розробленого застосунку на основі обраного SDK, використаних методів аналізу зображень, бібліотек та фреймворків.

2 МАТЕМАТИЧНІ МОДЕЛІ ТА РОЗРОБКА АЛГОРИТМУ ДЛЯ МОНІТОРИНГУ ДІЙ ТА СТАНУ УЧАСНИКІВ КОНФЕРЕНЦІЇ ZOOM

2.1 Перетворення кольорових моделей

Наразі комп’ютерні системи використовують різноманітні кольорові моделі: RGB, CMYK, HSV, LAB, YUV, тощо для представлення кольору в цифровому форматі. Так Zoom для передачі кадрів відеоконференції використовує YUV формат. Ці кадри ми отримуємо через Video SDK, але більшість моделей глибокого навчання не підтримують вхідні дані у форматі YUV. Через що потрібно конвертувати кадри у RGB формат, який є більш звичним для нейронних мереж, для того щоб надалі була можливість робити їх аналіз.

Кольоровий формат RGB(red-green-blue) ґрунтується на можливостях колбочок сітківки ока реагувати на різні довжини хвиль. RGB описує як треба змішати три основних кольори: червоний, зелений та синій за допомогою адитивного змішування, щоб отримати бажаний відтінок.

Кольоровий формат YUV також можна вважати подібним до сітківки людського ока, причому яскравість (Y) – описує інтенсивність світла, як і палички в сітківці ока. У темряві, коли конусні клітини не мають достатньої інтенсивності світла для активації та визначення кольорів, палички є основним джерелом інформації.

Однак, коли інтенсивність світла трохи збільшується, тоді додактова інформація з конусних клітин стає доступною. У форматі YUV два додаткових канали – компоненти, які несуть інформацію про колір [12]. Тобто YUV окремо зберігає чорно-білу та кольорову інформацію. Це переважно використовувалося в аналогових телевізійних стандартах, коли інформацію про кольори додавали до існуючого каналу яскравості. Таким чином була

можливість передавати інформацію одразу для чорно-білих та кольорових телевізорів.

YUV конвертується в RGB за допомогою наступного перетворення [13]:

$$\begin{cases} R = Y + 1.140V, \\ G = Y - 0.395U - 0.581V, \\ B = Y + 2.032V. \end{cases} \quad (2.1)$$

2.2 Основні принципи згорткових нейронних мереж

Згорткові нейронні мережі (CNN) – це клас глибоких нейронних мереж, які найчастіше використовуються для аналізу візуальних зображень. Нейронні мережі загалом складаються з набору нейронів, які організовані в шари, кожен з власними вагами та зміщенням [14].

Першим є вхідний шар, що представляє вхідне зображення CNN [15]. Цей шар може складатися з декількох каналів, наприклад, якщо ми надаємо RGB зображення на вхід, то шар буде мати три канали, що відповідають червоному, зеленому та синьому кольорам.

Далі іде згортковий шар, який використовується для вилучення різних ознак з вхідних зображень. У цьому шарі виконується математична операція свертки між вхідним зображенням та фільтром певного розміру.

Переміщуючи фільтр по вхідному зображенню, виконується скалярний добуток між фільтром та частинами вхідного зображення відносно розміру фільтра. Вихід називається картою ознак, яка дає нам інформацію про зображення, таку як кути та краї. Пізніше ця карта ознак передається до інших шарів, щоб навчитися декількох інших ознак вхідного зображення.

У більшості випадків після згорткового шару використовується шар пулінгу. Основною метою цього шару є зменшення розміру карт ознак, отриманих зі згорткового шару, з метою зменшення обчислювальних витрат.

Це досягається шляхом зменшення зв’язків між шарами і незалежної роботи з кожною картою ознак.

Є різні методи пулігну, але ідея полягає в тому, щоб узагальнити ознаки, отримані у минулому шарі. У максимальному пулінгу з кожної картки ознак береться найбільший елемент, в середньому – середнє значення елементів в попередньому шарі, а в сумі – сума всіх елементів. Цей шар допомагає зменшити обчислювальні витрати мережі і загальний час навчання.

Повноз’єднанний шар складається з ваг і зміщень і використовується для з’єднання нейронів між двома різними шарами. Ці шари зазвичай розміщують перед вихідним шаром та утворюють останні кілька шарів архітектури.

Вхідне зображення з попередніх шарів перетворюється на одновимірний вектор. Вектор проходить декілька інших повнозв’язних шарів, де зазвичай відбуваються математичні операції функцій. На цьому етапі починається процес класифікації.

Коли всі ознаки пов’язані з повноз’єднаним шаром, тоді це може призвести перенавчання на навчальному наборі даних. Це відбувається, коли певна модель працює настільки добре, що це має негативний вплив на її продуктивність при використанні на нових даних.

Для подолання цієї проблеми використовується шар відмови, в якому під час навчання деякі нейрони видаляються з нейромережі, що призводить до зменшення розміру моделі.

Нарешті, одними з найважливіших компонент згорткової нейронної мережі є функції активації. Простими словами вони визначають чи повинен бути активований нейрон чи ні.

Існує кілька типів функцій активації: ReLU, Softmax, tanH, Sigmoid. Кожна з цих функцій має специфічне призначення. Так наприклад Sigmoid переважно використовується для бінарної класифікації, а Softmax для багатокласової класифікації.

Базова архітектура згорткових нейронних мереж представлена на рисунку 2.1.

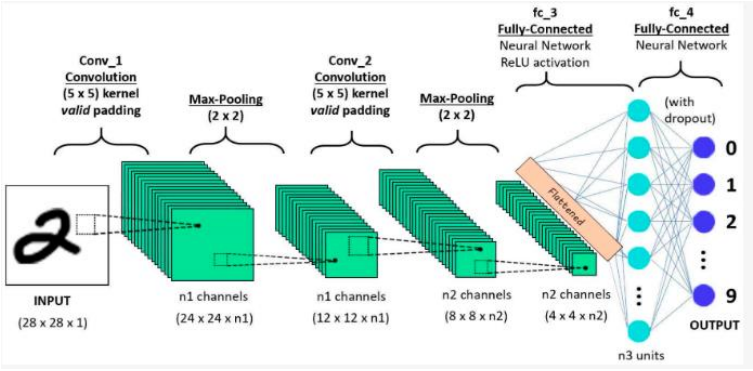


Рисунок 2.1 – Архітектура згорткових нейронних мереж [16]

2.3 Навчання нейронних мереж та датасети

Існує декілька типів навчання нейронних мереж, а саме: з вчителем, без вчителя, з підкріпленням. Під час навчання з вчителем програма навчається на датасеті з розміченими даними. Програма використовує надані бажані входи та виходи для визначення кореляцій та логіки, які вже в подальшому використовуються для прогнозування відповіді. Під час навчання алгоритм порівнює свої результати з наданими правильними відповідями і у разі потреби коригує модель. Навчання відбувається до того моменту, коли алгоритм почне надавати результати в межах заданого діапазону точності. Для загальних задач можна використовувати наявні датасети, такі як MNIST, ImageNet, Titanic Dataset, CIFAR-10, тощо. Для більш специфічних задач краще створювати власний датасет для навчання [17]. Види навчання з вчителем поділяються на класифікацію: наївний Баєсовський класифікатор [18], логістична регресія, SVM, а також регресії: лінійна, нелінійна та лінійна Баєсівська регресії. Прикладом застосування є прогноз продажів та оцінка ризиків.

При навчанні без вчителя правильні відповіді у датасеті відсутні, програма сама вивчає дані і виявляє закономірності. Програма робить висновки та групує їх за подібним принципом до людини, коли вона довгий час природно спостерігає за навколишнім світом. По мірі зростання

оброблених даних «інтуїція та спостереження» програми стають більш досконалими. Не завжди є можливість класифікувати дані вручну для навчання, саме в таких випадках вдаються до навчання без вчителя. Алгоритми навчання без вчителя: пошук правил, кластерізація, зменшення розмірності, самоорганізаційна карта Кохонена [19–21]. Прикладом застосування є створення систем рекомендацій, виявлення аномалій.

Останнім видом є навчання з підкріпленням, для якого визначаються набір правил та станів. Програма виконує різні дії та спостерігає реакції на основі яких вчиться використовувати правила для досягнення бажаного результату. По суті вчитель замінюється середовищем. Q-навчання [22], пошук стратегії Марковського процесу вирішування є поширеними методами навчання з підкріпленням. Використовується в ботах для комп'ютерних ігор, трейдингу, машинах з автономним керуванням, в логістиці при плануванні завдань та складанні графіків, медичній сфері.

2.4 Етапи розпізнавання облич та емоцій

Розпізнавання облич та емоцій поділяється на пайплайн з моделей та алгоритмів. Загалом можна виділити чотири етапи: детектування, трансформація, вилучення ознак та відповідність ознак для верифікації, а для розпізнавання емоцій – класифікація ознак.

В першу чергу на зображені шукається обличчя, вилучається, якщо присутнє, обрізається для передачі на наступтний етап трансформації. Існує декілька алгоритмів для даної задачі, починаючі з більш ранніх Haar cascade, HOG, які були вперше представлені у 2001 та 2005 роках відповідно. Більш сучасними є CNN та MTCNN, які були розроблені у 2016 році.

Наступним етапом є трансформація, а саме вирівнювання обличчя. Google дослідники у 2015 році заявили, що точність їх FaceNet системи розпізнавання облич покращилася з 98,87% до 99,63% при використанні

вирівнювання [23]. Виявити очі можна за допомогою модуля OpenCV (cascade Haar). Після цього треба дізнатися кут нахилу, щоб знати на скільки треба повертати обличчя. Кут нахилу знаходиться за допомогою розрахунку Евклідової відстані та формули косинусу:

$$d = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2}, \tag{2.2}$$

$$\cos x = \frac{b^2+c^2-a^2}{2bc}. \tag{2.3}$$

На етапі вилучення ознак отримуються біологічні ознаки обличчя. Ці ознаки є характерними рисами облич, які відрізняються від людини до людини. Комбінації цих ознак називаються векторами ознак. Жодна людина не може мати ідентичні вектори ознак як у іншої людини, за вийнятком однояйцевих близнюків. Найпопулярнішими моделями для вилучення ознак для подальшої верифікації є VGG-Face, Google FaceNet, DeepID, Facebook DeepFace, OpenFace, ArcFace. Для аналізу емоцій використовуються такі моделі: FaceMesh, VGG-16.

На останньому етапі розпізнавання облич отримані ознаки порівнюються з ознаками з бази даних, в той час для розпізнавання емоцій ці ознаки класифікуються. Використовуються такі методи класифікації та порівняння ознак: Евклідова відстань між векторами, косинусна подібність, SVM, метод *K*-найближчих сусідів, наївний Баєс, CNN, RNN, тощо [24, 25].

2.5 Детектування обличчя за допомогою Single Shot MultiBox Detector

Наразі існує багато методів детектування обличчя. Одними з перших були розроблені каскадний метод Хаара (Haar) та гістограма орієнтованих градієнтів (HOG), які були презентовані у 2001 та 2005 роках відповідно. Haar

та HOG були взяті для порівняння через задовільну точність, а саме від 90% при фронтальному положенні обличчя без відхилень, що є переважаючим положенням обличчя під час відеоконференцій [26].

Більш сучасними є однокадрове багатоблочне виявлення (SSD) та багатозадачна каскадна згорткова нейронна мережа (MTCNN), які були презентовані у 2016 році.

Haar витягує приблизно 180000 ознак з зображень, потім найкращі 6000 відфільтровуються за допомогою алгоритму Adaboost. Останнім етапом є каскадний класифікатор, який складається з серії етапів, де кожен етап є набором слабких учнів.

Результат отримується із середнього прогнозу усіх слабких учнів. Етапи призначені для якнайшвидшого відхилення негативних зразків, оскільки більшість вікон не містять нічого цікавого [27]. Для роботи був взятий класифікатор haarcascade_frontalface2 з бібліотеки OpenCV.

Алгоритм HOG ділить зображення на маленькі клітинки, обчислює орієнтацію та величину градієнта кожної клітинки, потім об’єднує інформацію про градієнти в гістограму орієнтованих градієнтів.

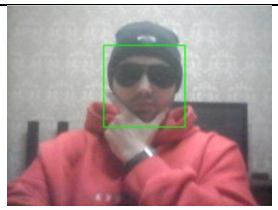
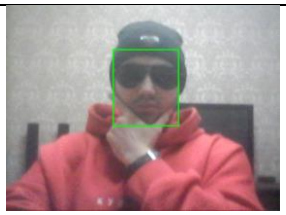
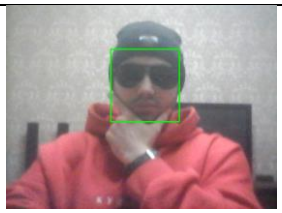
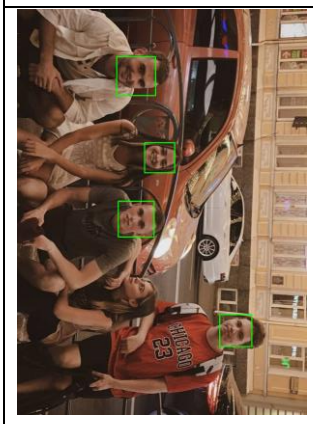
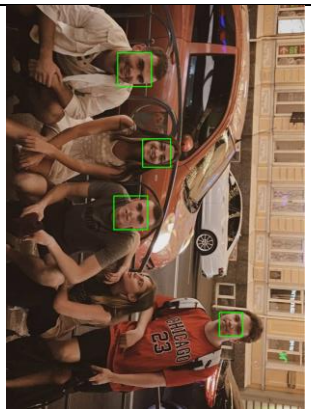
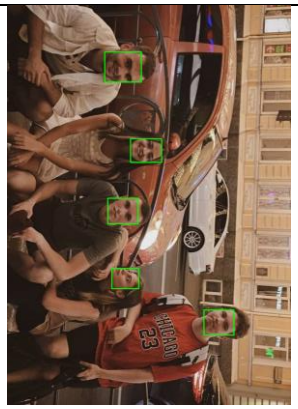
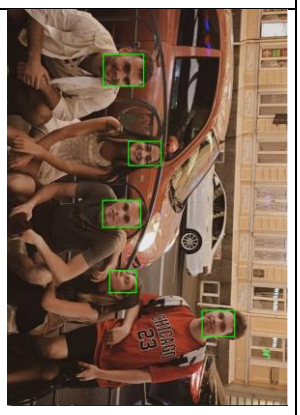
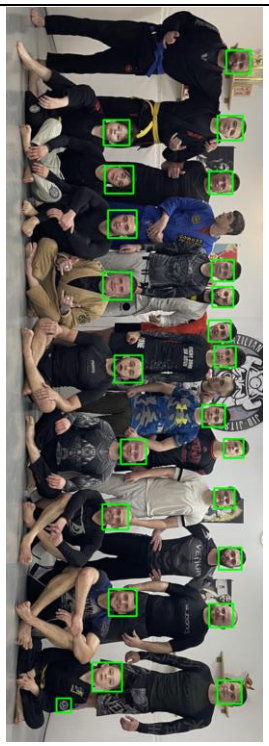
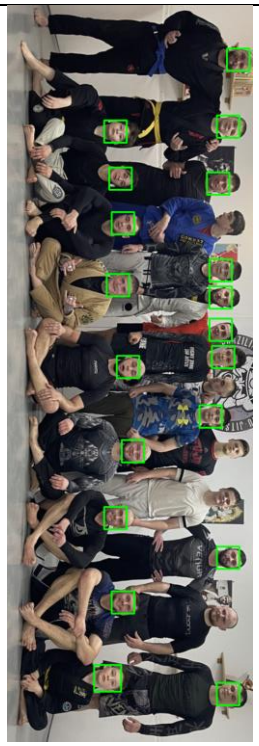
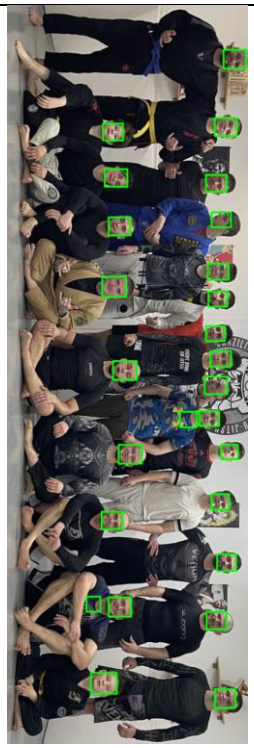
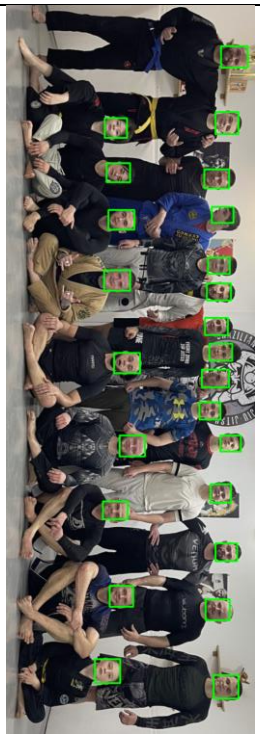
Для роботи використана реалізація HOG з бібліотеки Dlib, в якій отримані ознаки за допомогою HOG потім проганяються через SVM. MTCNN складається з 3 етапів.

Пешим етапом є повністю згорткова мережа(FCN), яка відрізняється від CNN тим, що не має щільного шару. Відповідає за отримання вікон-кандидатів і їх векторів регресії обмежувальної рамки.

Наступним етапом є CNN, яка зменшує кількість кандидатів, виконує калібрування за допомогою регресії обмежувальної рамки та використовує немаксимальне придушення (NMS) для об’єднання кандидатів, що перетинаються. Останній етап подібний до другого, але ця мережа спрямована на більш детальний опис обличчя та знаходження основних 5 орієнтирів обличчя: очі, ніс, краї рота [28].

Приклад порівняння алгоритмів наведенений в таблиці 2.1.

Таблиця 2.1 – Порівняння алгоритмів детектування обличчя

Haar	**HOG**	**MTCNN**	**SSD on Caffe**
			
1246 мс; 1/1	179 мс; 0/1	1355 мс; 1/1	39 мс; 1/1
			
1435 мс; 4/5	451 мс; 4/5	1732 мс; 5/5	294 мс; 5/5
			
1770 мс; 22/24 (1 хибний позитив)	578 мс; 19/24	2101 мс; 24/24 (2 хибних позитива)	232 мс; 24/24

Haar очікувано показав найгірший результат по співвідношенню швидкості та точності, він не завжди детектував обличчя при значному відхиленні.

Також давав хибні позитиви і не завжди детектував всі обличчя на зображенні з великою кількістю людей. Цим підтвердив, що є доволі застарілим методом.

HOG та MTCNN мали досить рівні результати, при цьому HOG вигравав по швидкості, але був гіршим у точності, а саме був не стійким до частично закритого обличчя та зображень з великою кількістю людей.

MTCNN періодично давав хибні позитиви. Також HOG не розпізнавав обличчя на зображеннях менше 80×80, тому при потребі обробляти такі зображення треба збільшувати їх масштаб, що збільшить час обробки.

SSD з Caffe моделлю надав найкращі результати з єдиною проблемою в неможливості детектувати маленькі обличчя на зображеннях з великою роздільною здатністю, так як класично SSD стискає зображення до 300×300 і після цього обличчя важко розпізнавати навіть людським зором, тому для роздільної здатності від 2000×2000 треба робити додаткові налаштування по ступеню стискання.

Для задачі детектування облич у відеоконференції в реальному часі підійдуть MTCNN та SSD через їх точність, але враховуючи значну перевагу SSD у швидкості та з невеликою роздільною здатністю зображень у відеоконференціях доцільніше буде обрати саме цей метод.

На рисунку 2.2 можна побачити, що архітектура базується на відомій мережі VGG-16, але не має повністю зв’язних шарів. Замість зв’язних шарів було додано набір вспоміжних згорткових шарів: починаючи з conv6, що дозволяє вилучати ознаки на кількох масштабах та поступово зменшувати розмір входу до кожного наступного шару.

SSD використовує техніку регресії обмежувальної рамки для отримання швидких пропозицій координат обмежувальних рамок незалежно від класу. Ще однією особливістю SSD є використання двох функцій втрат під час

навчання мережі. Перша вимірює впевненість мережі в передбачених об’єктах, друга – відстань між прогнозованими обмежувальними рамками та рамками в навчальному наборі.

Окрім цього, під час навчання використовується техніка видобутку складних негативних зразків, суть якої полягає в зосереджені на найскладніших негативних прикладах, які найлегше сплутати з позитивними. Останньою особливістю навчання SSD є такі методи збільшення навчальних зразків як випадкове кадрування, перевертання та масштабування, які допомагають мережі навчитися розпізнавати об’єкти за різних умов і точок зору.

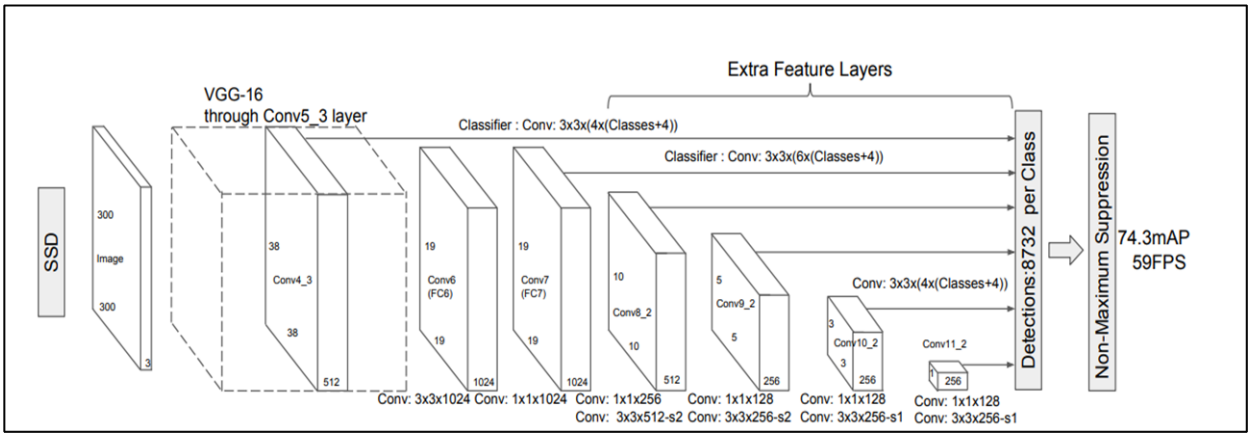


Рисунок 2.2 – Архітектура SSD [29]

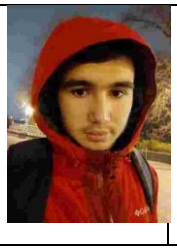
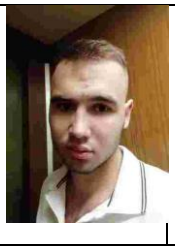
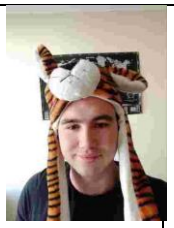
В 2022 році дослідникам з Тяньцзінського університету покращеним SSD методом на датасеті WiderFace вдалося досягти точності 0,834%, при цьому Haar та MTCNN на тих же зображеннях показали результат в 0,137% та 0,6% відповідно [30].

Модель була навчена на WiderFace датасеті (40% зображень цього датасету призначені для навчання, 10% для перевірки та 50% для тестування [31]). При тестуванні швидкості в середньому SSD обробляв 46 кадрів на секунду, в той час MTCNN лише 16.

2.6 Розпізнавання облич за допомогою DeepFace

У той час коли більшість бібліотек розпізнавання облич обслуговують лише одну модель штучного інтелекту, бібліотека DeepFace відразу надає можливість скористатися багатьма передовими моделями розпізнавання облич. Більшість цих моделей базуються на найсучасніших згорткових нейронних мережах і забезпечують найкращі результати в своєму класі. Результати порівняння моделей наведені в таблиці 2.2.

Таблиця 2.2 – Порівняння моделей розпізнавання обличчя

						
	Пройшов перевірку					Еталон
	Час обробки					
VGG-Face	так	так	так	так	ні	
	2,15 с	3,27 с	2,32 с	4,36 с	3,24 с	
Facenet	так	так	так	так	ні	
	1,77 с	1,92 с	1,78 с	2,16 с	2,75 с	
Facenet512	так	так	ні	ні	ні	
	1,81 с	1,9 с	1,89 с	2,22 с	2,71 с	
OpenFace	так	ні	ні	ні	ні	
	2,07 с	2,15 с	1,92 с	2,06 с	2,65 с	
DeepFace	так	ні	так	так	так	
	2,06 с	2,24 с	2,37 с	2,79 с	2,99 с	
ArcFace	так	так	так	так	ні	
	1,85 с	1,86 с	1,89 с	2,17 с	2,6 с	

Заявлена точність моделей розробниками DeepFace на датасетах LFW та YTF наведена в таблиці 2.3 [32].

Таблиця 2.3 – Точність моделей розпізнавання обличчя

Model	**LFW Score**	**YTF Score**
Facenet512	99,65%	-
ArcFace	99,41%	-
Facenet	99,20%	-
VGG-Face	98,78%	97,40%
OpenFace	93,80%	-

Найкращі результати по точності та швидкості показали ArcFace та Facenet.

ArcFace використовує механізм навчання подібності, що дозволяє вирішувати задачу метрики в задачі класифікації, впроваджуючи Angular Margin Loss замість Softmax Loss.

Різниця між обличчями в ArcFace обчислюється за допомогою косинусної подібності.

У типовій задачі класифікації після обчислення ознак, повністю зв’язаний шар бере скалярний добуток ознак та ваг і застосовує Softmax до виходу.

У ArcFace косинусна подібність обчислється шляхом нормалізації [33] ознак та ваг повністю зв’язаного шару та беручи їх скалярний добуток.

Втім, функція втрат розраховується шляхом застосування Softmax до косинусної подібності.

Наступним кроком є застосування оберненої косинусної функції до значень косинусної подібності після взяття скалярного добутку та додавання кутової межі лише для правильних міток (рис. 2.3).

Цим способом запобігається занадто велика залежність вагів повнозв’язного шару від вхідного набору даних (табл. 2.4).

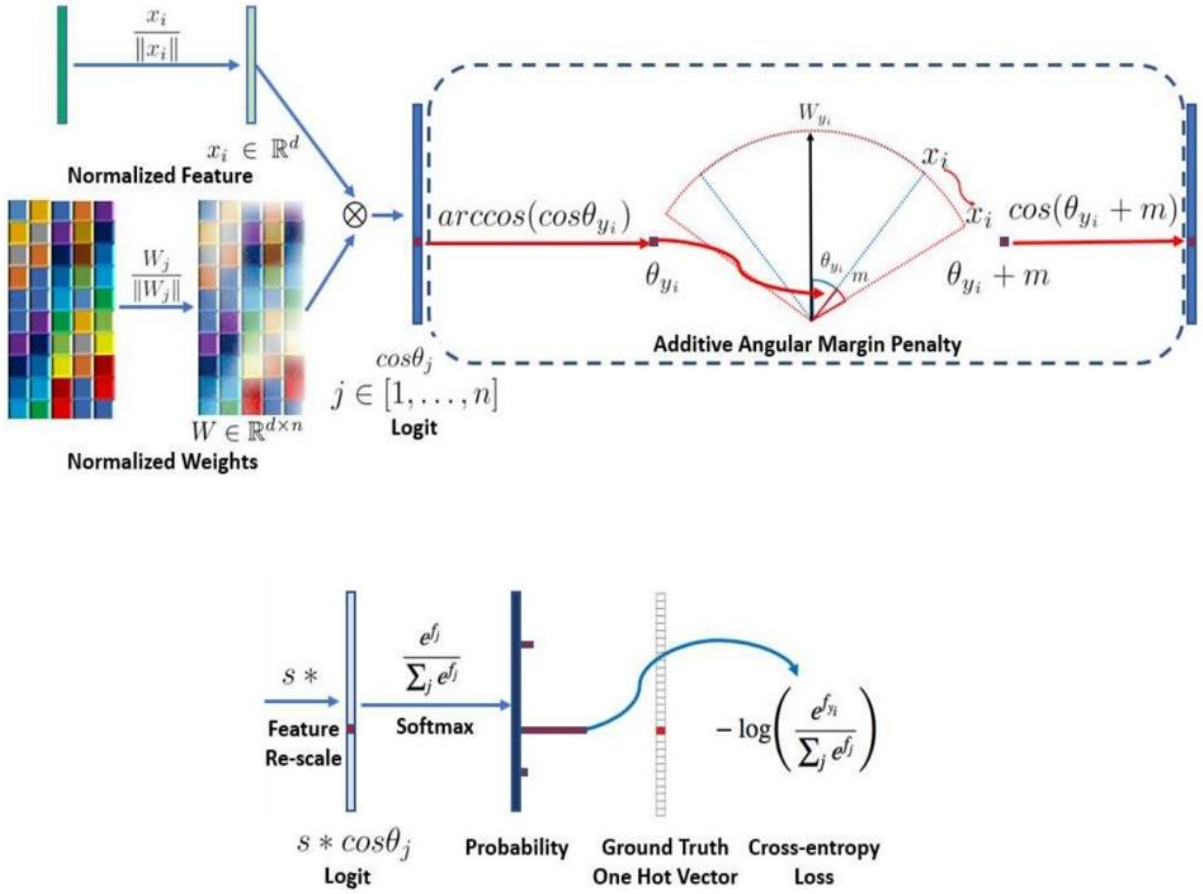


Рисунок 2.3 – Архітектура ArcFace [34]

Таблиця 2.4 – Конфігурація параметрів FaceNet [35]

№	Тип шару	Розмір на вході	Розмір на виході	Розмір ядра (шаг згортання)
1	**2**	**3**	**4**	**5**
1	Conv	224 × 224 × 3	112 × 112 × 64	7 × 7 × 3 (2)
2	Max pool	112 × 112 × 64	56 × 56 × 64	3 × 3 × 64 (2)
3	Inception (2)	56 × 56 × 64	56 × 56 × 192	3 × 3 × 192 (1)
4	Max pool	56 × 56 × 192	28 × 28 × 192	3 × 3 × 192 (2)
5	Inception (3a)	28 × 28 × 192	28 × 28 × 256	1 × 1 × 64 (1), 3 × 3 × 128 (1), 5 × 5 × 32 (1), 1 × 1 × 32(1)

Продовження таблиці 2.4

1	2	3	4	5
6	Inception (3b)	28 × 28 × 256	28 × 28 × 320	1 × 1 × 64 (1), 3 × 3 × 128 (1), 5 × 5 × 64 (1), 1 × 1 × 64(1)
7	Inception (3c)	28 × 28 × 320	14 × 14 × 640	3 × 3 × 256 (2), 5 × 5 × 64 (2)
8	Inception (4a)	14 × 14 × 640	14 × 14 × 640	1 × 1 × 256 (1), 3 × 3 × 192 (1), 5 × 5 × 64 (1), 1 × 1 × 128(1)
9	Inception (4b)	14 × 14 × 640	14 × 14 × 640	1 × 1 × 224 (1), 3 × 3 × 224 (1), 5 × 5 × 64 (1), 1 × 1 × 128(1)
10	Inception (4c)	14 × 14 × 640	14 × 14 × 640	1 × 1 × 192 (1), 3 × 3 × 256 (1), 5 × 5 × 64 (1), 1 × 1 × 128(1)
11	Inception (4d)	14 × 14 × 640	14 × 14 × 640	1 × 1 × 160 (1), 3 × 3 × 288 (1), 5 × 5 × 64 (1), 1 × 1 × 128(1)
12	Inception (4e)	14 × 14 × 640	7 × 7 × 1024	3 × 3 × 256 (2), 5 × 5 × 128 (2)
13	Inception (5a)	7 × 7 × 1024	7 × 7 × 1024	1 × 1 × 384 (1), 3 × 3 × 384 (1), 5 × 5 × 128 (1), 1 × 1 × 128(1)
14	Inception (5b)	7 × 7 × 1024	7 × 7 × 1024	1 × 1 × 384 (1), 3 × 3 × 384 (1), 5 × 5 × 128 (1), 1 × 1 × 128(1)
15	Avg Pool	7 × 7 × 1024	1 × 1 × 1024	7 × 7 × 3 (1)
16	FC	1 × 1 × 1024	1 × 1 × 128	1024 × 128 (1)

2.7 Аналіз емоцій за допомогою DeepFace

DeepFace здатна аналізувати такі атрибути вік, стать, емоції (злість, страх, нейтральність, сум, огида, радість і здивування) та расу (азіатську, білу, середньосхідну, індійську, латиноамериканську та чорну). Нейронна мережа навчена на датасеті FER2013, на якому була отримана точність 57% при аналізі емоцій [36]. Вікова модель обробляє зображення з точністю ±4,65 MAE, гендерна модель отримала 97,44% accuracy, 96,29% precision та 95,05% recall.

FER2013 складаєтья з 30000 RGB зображень з різними виразими обличчя, розміром 48×48 [37]. Вираз огиди має мінімальну кількість зображень – 600, тоді як інші вирази мають майже по 5000 зразків. Через це DeepFace найгірше розпізнає саме емоцію огиди навіть при гарному освітленні та задовільній якості зображення (рис. 2.4).

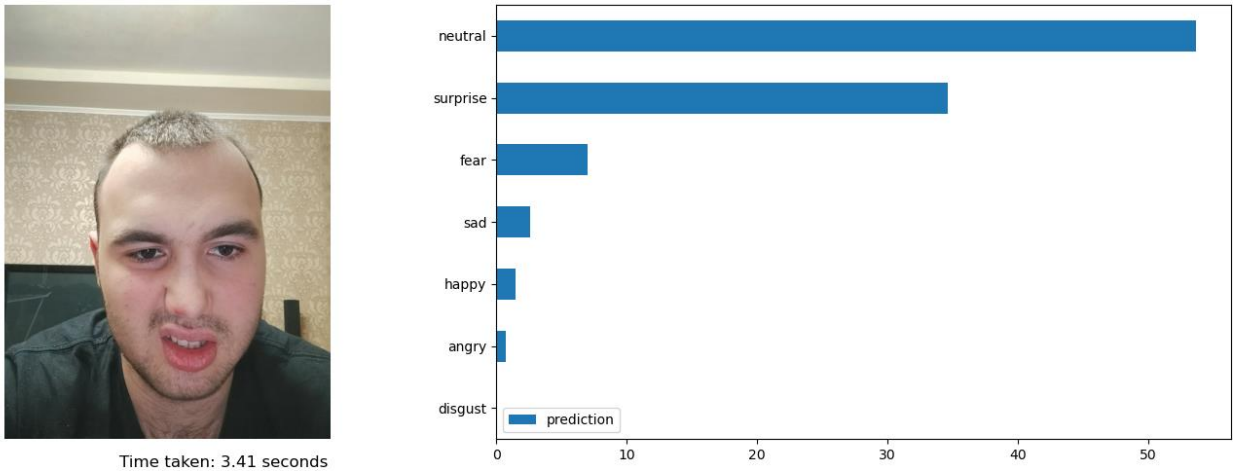


Рисунок 2.4 – Приклад некоректного розпізнавання емоції огиди

Такі емоції як злість, страх, нейтральність, сум, радість розпізнавалися навіть при поганій якості зображення та освітлення, при цьому здивуванні вдалося розпізнати тільки при гарному освітленні та досить виразній міміці (рис. 2.5).

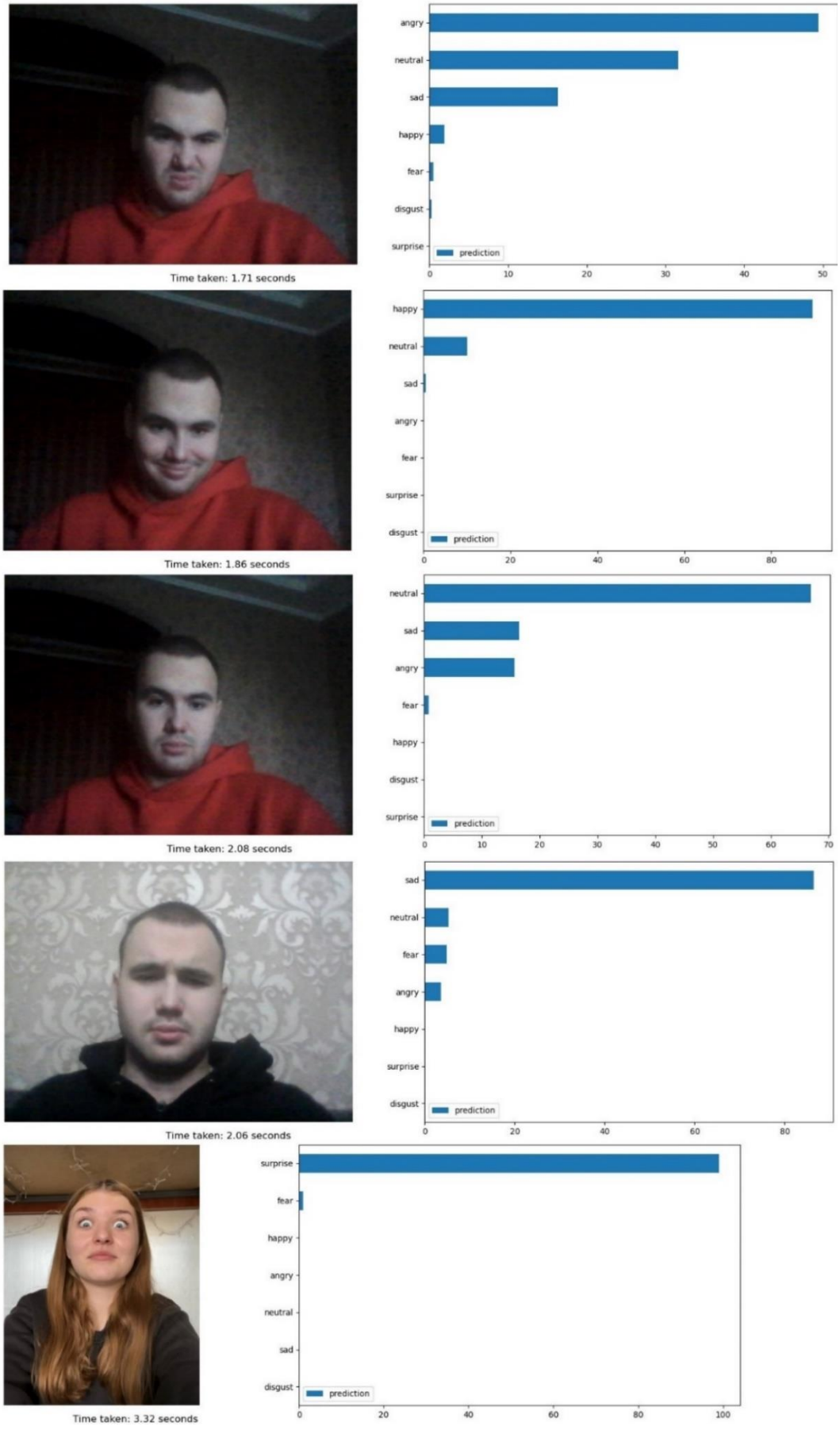


Рисунок 2.5 – Приклад коректного розпізнавання емоцій

3 РОЗРОБКА ЗАСТОСУНКУ

3.1 Специфікація вимог до застосунку

Zoom Video SDK повинен виконувати такі функції:

– давати можливість користувачам створювати та приєднуватись до відеоконференцій, виходити з них та завершувати;

– авторизуватися при при’єднанні до конференцій;

– передавати і відображати аудіо, відеопотоки (з камери та екрану), текстові повідомлення, реакції у вигляді emoji користувачів, припиняти та відновлювати передачу відповідних потоків;

– записувати кадри з відоепотоків користувачів, які їх передають для подальшої обробки, обробляти та записувати такі події, як підняття руки зDa допомогою emoji, вмикання та вимикання мікрофону для подальшого збору статистики.

Також потрібно реалізувати сервіс, який буде виконувати такі функції:

– генерувати JWT токен, який користувачі будуть вводити на етапі авторизації;

– працювати паралельно з Zoom Video SDK та при надходженні кадрів учасників надсилати їх сервісу аналізу даних;

– отримувати проаналізовані результати від сервісу аналізу даних, на основі цих даних приймати рішення чи були порушені правила конференції;

– записувати отримані дані в базу даних;

– при порушенні правил відеоконференції – сповіщувати про це та виводити іншу корисну інформацію, як поточний статус верифікації учасників, їх домінуючу емоцію;

– показувати статистику емоцій учасників за різні періоди мітингів, не пройдена верифікація, зайва кількість людей в кадрі, кількість підняття рук, час проведений з ввімкнутим мікрофоном, тощо.

Система аналізу в свою чергу повинна вміти приймати кадри та робити аналіз:

– детектувати обличчя;

– верифікувати обличчя;

– виділити домінуючу емоцію обличчя, зрозуміти вік, стать, расу особи присутньої в кадрі.

При цьому система аналізу повинна надавати не тільки задовільні результати, а і працювати з доволі високою швидкістю для тих випадків, коли застосунок буде налаштований на запис кадрів кожну секунду, або при надто великій кількості учасників конференції.

3.2 Специфікація правил проведення відеоконфернції

Учасник відеоконфернціїї повинен дотримуватися таких правил:

– в кадрі повинна знаходитися тільки одна особа, в разі відсутності або присутності двох і більше осіб в кадрі – власник конференції буде повідомлений;

– в нікнейм особа повинна вказати пошту, яка надавалася власнику конференції разом з еталонним зображенням цієї особи; еталонне зображення повинно бути задовільної якості з гарним освітленням, на зображенні повинна бути присутня тільки дана особа, 30% і більше повинно займати лице; якщо в кадрі буде особа, яка не відповідає еталонному зображенню – власник конференції буде повідомлений;

– камера учасника повинна передавати зображення задовільної якості, робоче місце повинно бути гарно освітлене, обличчя повернуте в бік камери, якщо система не зможе детектувати або верифікувати обличчя через недостатню якість, погане освітлення або положення обличчя – власник конференції буде повідомлений.

3.3 Розробка бази даних для зберігання інформації щодо моніторингу

Для розробки застосунку був обраний Microsoft SQL Server як система управління реляційною базою даних. Структура бази даних наведена на рисунку 3.1.

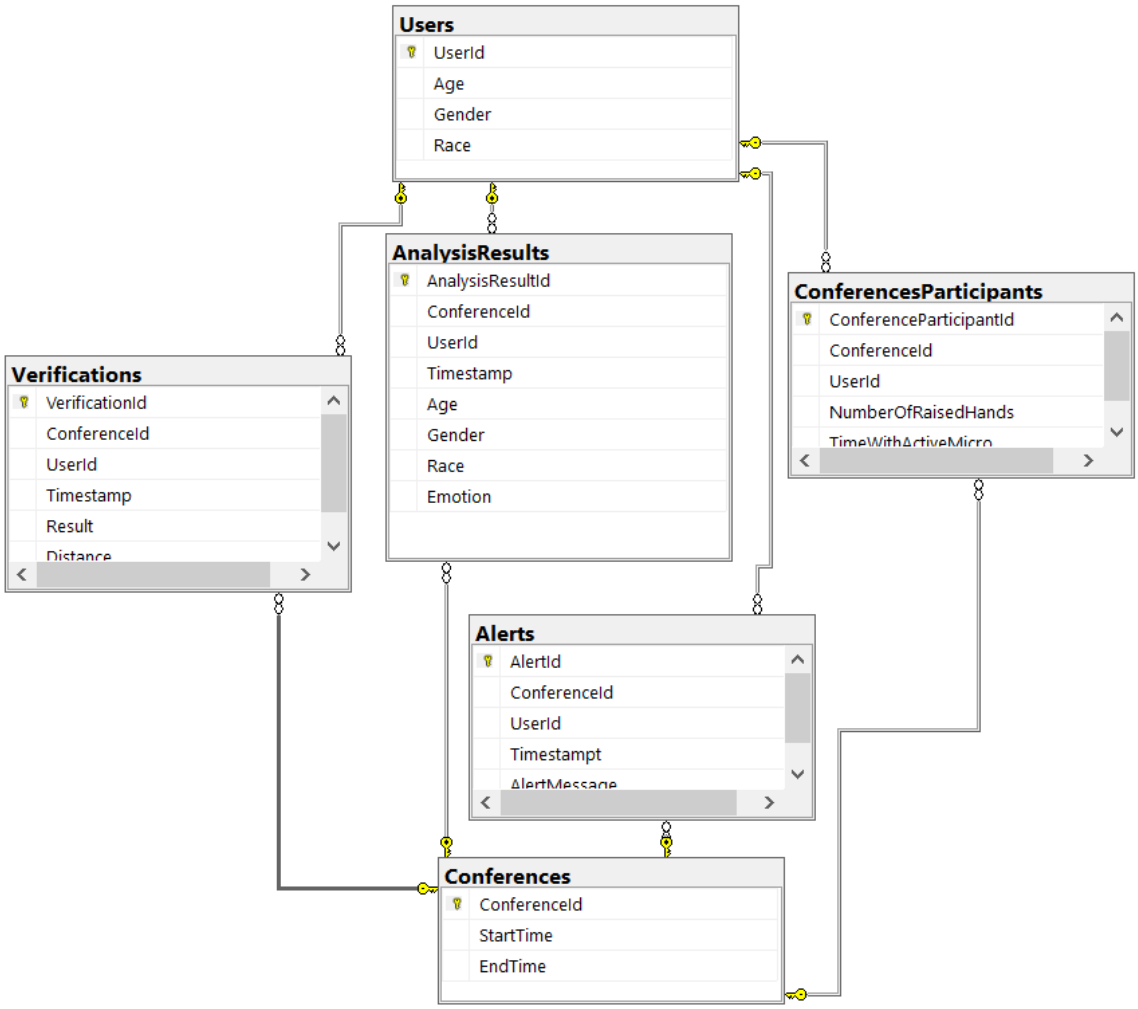


Рисунок 3.1 – Структура бази даних

Використання реляційної бази даних обумовлено її швидкістю, особливо при потребі отримувати велику кількість записів, заздалегідь їх обробивши (фільтрація, сортування, тощо), що буде застосовуватись для відображення статистики.

При показі статистики відеоконференцій за пів року в базі даних буде дуже велика кількість записів, наприклад, якщо система налаштована

аналізувати емоції кожні 5 секунд, то якщо в конференції тривалістю 90 хв. будуть приймати участь 30 людей – в базі даних з’явиться 32400 записів в одній тільки таблиці емоцій. При показі статистики будуть запити по типу: показати 10 людей з найбільшим рівнем суму за останні 5 конференцій – саме тут SQL стане в нагоді, з його здатністю швидко обробляти подібні запити. Атрибути бази даних та їх типи даних описані в таблицях 3.1 – 3.6.

Таблиця 3.1 – Таблиця Conferences

Поле	**Тип даних**
ConferenceId	INT
StartTime	DATETIME
EndTime	DATETIME

Таблиця 3.2 – Таблиця Users

Поле	**Тип даних**
UserId	NVARCHAR(50)
Age	INT
Gender	NVARCHAR(10)
Race	NVARCHAR(50)

Таблиця 3.3 – Таблиця ConferencesParticipants

Поле	**Тип даних**
ConferenceParticipantId	INT
ConferenceId	INT
UserId	NVARCHAR(50)
NumberOfRaisedHands	INT
TimeWithActiveMicro	TIME

Таблиця 3.4 – Таблиця Verifications

Поле	**Тип даних**
VerificationId	INT
ConferenceId	INT
UserId	NVARCHAR(50)
Timestamp	DATETIME
Result	BIT
Distance	FLOAT

Таблиця 3.5 – Таблиця AnalysisResults

Поле	**Тип даних**
AnalysisResultId	INT
ConferenceId	INT
UserId	NVARCHAR(50)
Timestamp	DATETIME
Age	INT
Gender	NVARCHAR(10)
Race	NVARCHAR(50)
Emotion	NVARCHAR(20)

Таблиця 3.6 – Таблиця Alerts

Поле	**Тип даних**
AlertId	INT
ConferenceId	INT
UserId	NVARCHAR(50)
Timestampt	DATETIME
AlertMessage	NVARCHAR(300)

3.4 Розробка архітектури

Архітектура застосунку представлена на рисунку 3.2, розроблялася з урахуванням необхідності обробляти та аналізувати кадри в реальному часі, заміняти реалізації окремих компонент та масштабувати застосунок за допомогою інверсії залежностей.

Рисунок 3.2 – Архітектура застосунку

Застосунок складається з наступних компонент:

– Zoom Video SDK – це повністю налаштовуваний Zoom Client реалізований за допомогою C++, який відповідає за створення відеоконференцій, під’єднання та від’єднання учасників, передачу відео та аудіо потоків учасників їх рендеринг та показ, чат та інший базовий функціонал Zoom. Окрім цього він також записує отримані відео дані до Image Storage та події підняття руки, вмикання, вимикання мікрофону до Database, тощо;

– Image Storage – сховище, в якому зберігаються отримані в Zoom Video SDK кадри учасників конференції до того моменту, коли Session Observer витягне їх звідти і передасть на подальшу обробку та аналіз;

– Database – база даних, реалізована за допомогою Microsoft SQL Server, в яку Zoom Video SDK та Session Observer зберігають результати детекції, верифікації, аналізу емоцій, Zoom подій, тощо. При потребі показати

ці результати вони витягуються за допомогою Session Observer задля демонстрації ним статистики;

– Session Observer – сервіс, який працює паралельно з Zoom Video SDK, при надходженні нових кадрів в Image Storage – асинхронно відправляє їх на аналіз до Image Analyzer за допомогою брокера повідомлень RabbitMQ, який будує чергу повідомлень та гарантує їх отримання очікуваними споживачами. Після повернення RabbitMQ результату – записує його в базу даних, перевіряє чи не були порушені правила проведення конференції, в разі порушень – повідомляє. При потребі показати статистику – вилучає результати з бази даних, фільтрує, сортує та візуалізує їх. Показ поточних результатів, статистики, повідомлення про порушення правил конференції реалізовано за допомогою C# та Windows Presentation Foundation. Також перед початком конференції Session Observer відповідає за створення JSON Web Token, який знадобиться користувачам для авторизації та приєднання до відеоконференції;

– Image Analyzer – сервіс, який містить в собі навчені моделі для детектування, верифікації та аналізу емоцій учасників конференції. Кадри надходять за допомогою RabbitMQ, верифікації та аналіз відбуваються паралельно (перед цим проводиться детектування і для верифікації і для аналізу), що дає можливість не чекати закінчення обробки іншою моделлю, результати надсилаються назад в RabbitMQ. Паралельне виконання та ізоляція моделей дозволяє збільшувати кількість їх кількість задля покращення пропускної здібності. Сервіс реалізовано за допомогою Python, заздалегідь натреновані моделі взяті з бібліотек OpenCV та DeepFace.

UML діаграма послідовності дій користувача та сервісів застосунку показана на рисунку 3.3. На діаграмі був зображений процес взаємодії користувача з застосунком, починаючи з генерації JWT токена та приєднання до відеоконференції, закінчуючи отриманням результатів аналізу кадрів та завершенням зустрічі.

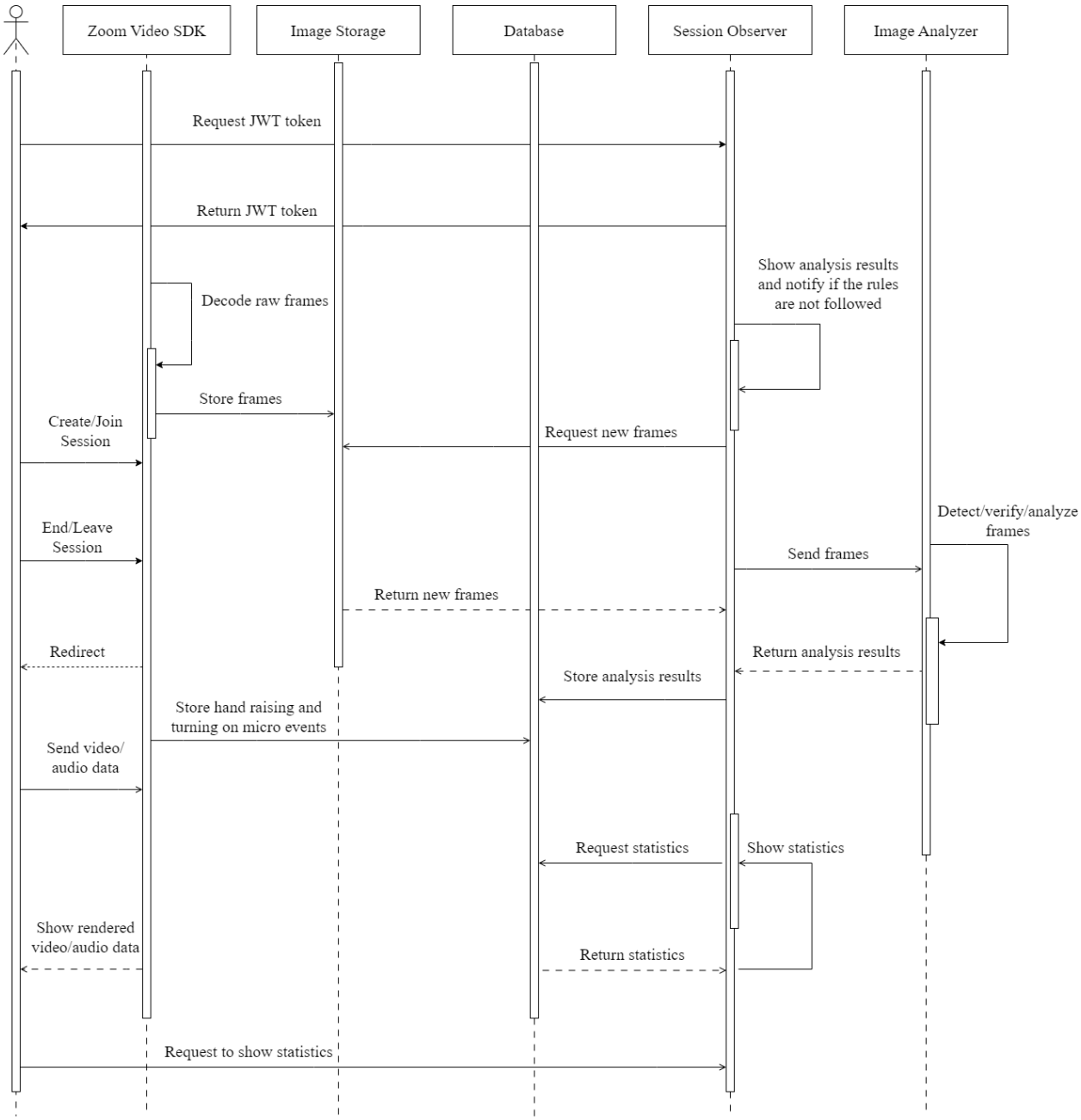


Рисунок 3.3 – Діаграма послідовності взаємодії сервісів застосунку

3.5 Ілюстрація роботи застосунку

Перше, що робить користувач, який хоче створити відеоконференцію – запускає Session Observer для генерації JWT токена та надає його людям, які будуть приєднуватись до конференції. JWT токен потрібен для аутентифікації під час приєднання до відеоконфернції. Початкова сторінка Session Observer відображена на рисунку 3.4.

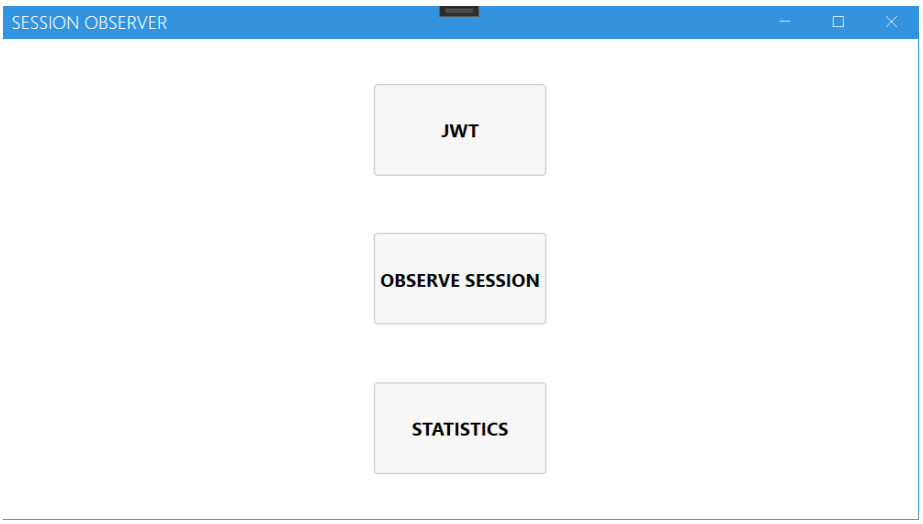


Рисунок 3.4 – Початкова сторінка Session Observer

Після запуску Session Observer майбутній власник конференції переходить на сторінку JWT, вводить key, secret та session name (рис. 3.5). Key та secret отримуються під час створення Zoom Video SDK акаунту. Акаунт треба створювати не всім користувачам, які будуть на конференції, а тільки її власнику.

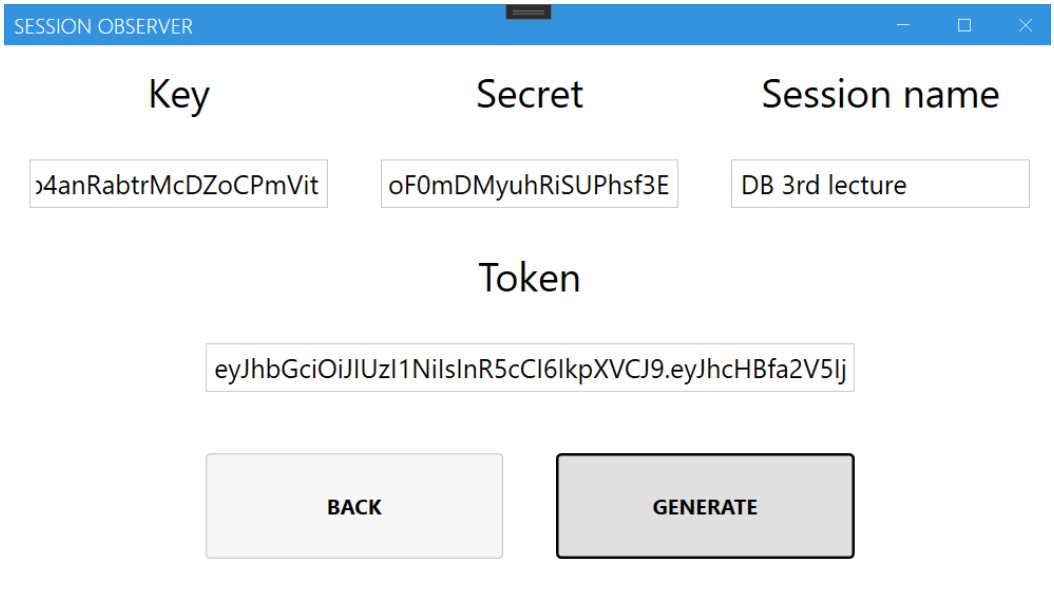


Рисунок 3.5 – Сторінка створення токена

Максимальний термін дії цього токену – 48 годин, після чого його треба наново генерувати. Токен генерується за допомогою HS256 алгоритму хешування.

Потім користувач запускає Zoom Video SDK (рис. 3.6), переходить в налаштування та вводить токен (рис. 3.7).

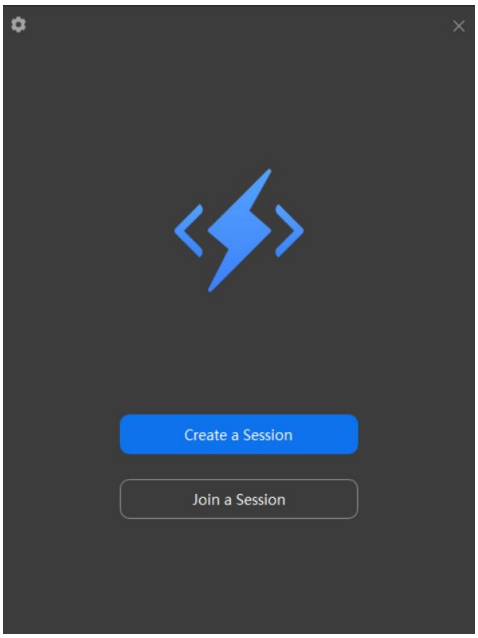


Рисунок 3.6 – Початкова сторінка Zoom Video SDK

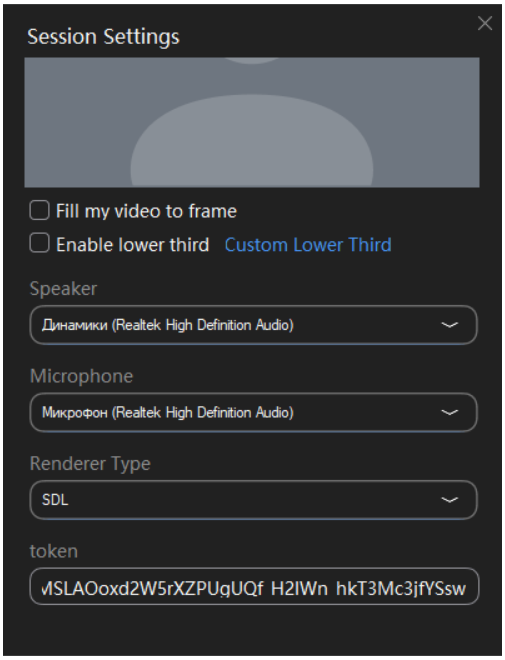


Рисунок 3.7 – Налаштування Zoom Video SDK

Далі майбутньому власнику конференції її треба створити (рис. 3.8), а учасникам – приєднатися (рис. 3.9).

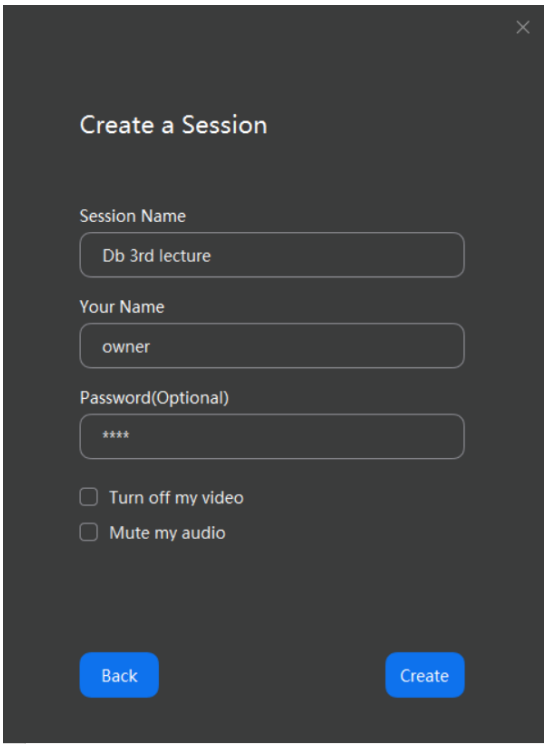


Рисунок 3.8 – Сторінка створення відеоконференції

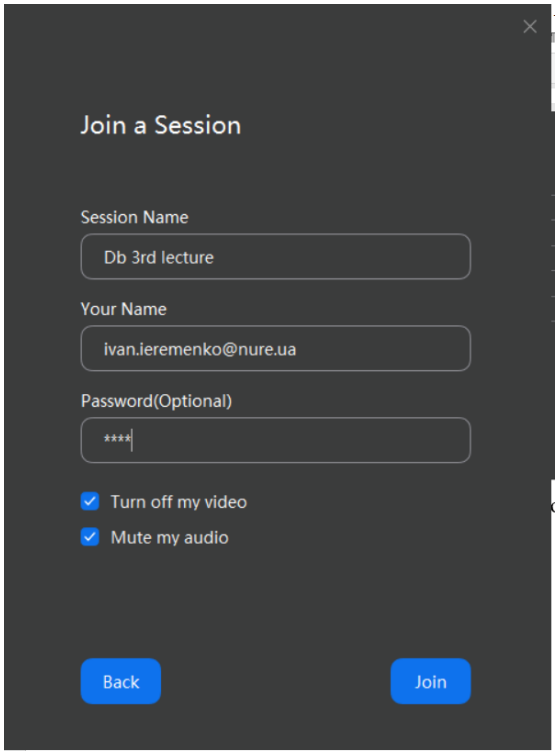


Рисунок 3.9 – Сторінка приєднання до відеоконференції

Після приєднання до відеоконференції користувачі бачать інтерфейс з більшістю елементів, які є в звичайному застосунку Zoom (рис. 3.10).

Рисунок 3.10 – Інтерфейс відеоконференції

Для того, щоб Session Observer почав аналізувати надходячі кадри – треба перейти на сторінку Observe Session. На рисунку 3.11 можна побачити візуалізовані результати верифікації, детектування та аналізу облич учасників конференції. Таблиця Participants status відповідає за відображення ім’я учасника, останнього результату верифікації, аналізу домінуючої емоції, статі, віку, раси, кількості піднімання руки, часу, проведеного з працюючим мікрофоном. Також на цій сторінці можна налаштувати інтервали, по яким буде відбуватися детектування, верифікація та аналіз (рис. 3.12). Налаштування реалізовані у вигляді модального вікна із блокуванням доступу до інтерфейсу решти інтерфейсу Session Observer. Налаштування зберігаються при закритті модального вікна щоб запобігти зайвим змінам. Також заблокована можливість зробити інтервал детектування більшим за інтервали верифікації та аналізу. Через те, що перед верифікацією та аналізом обличчя повинно бути детектовано.

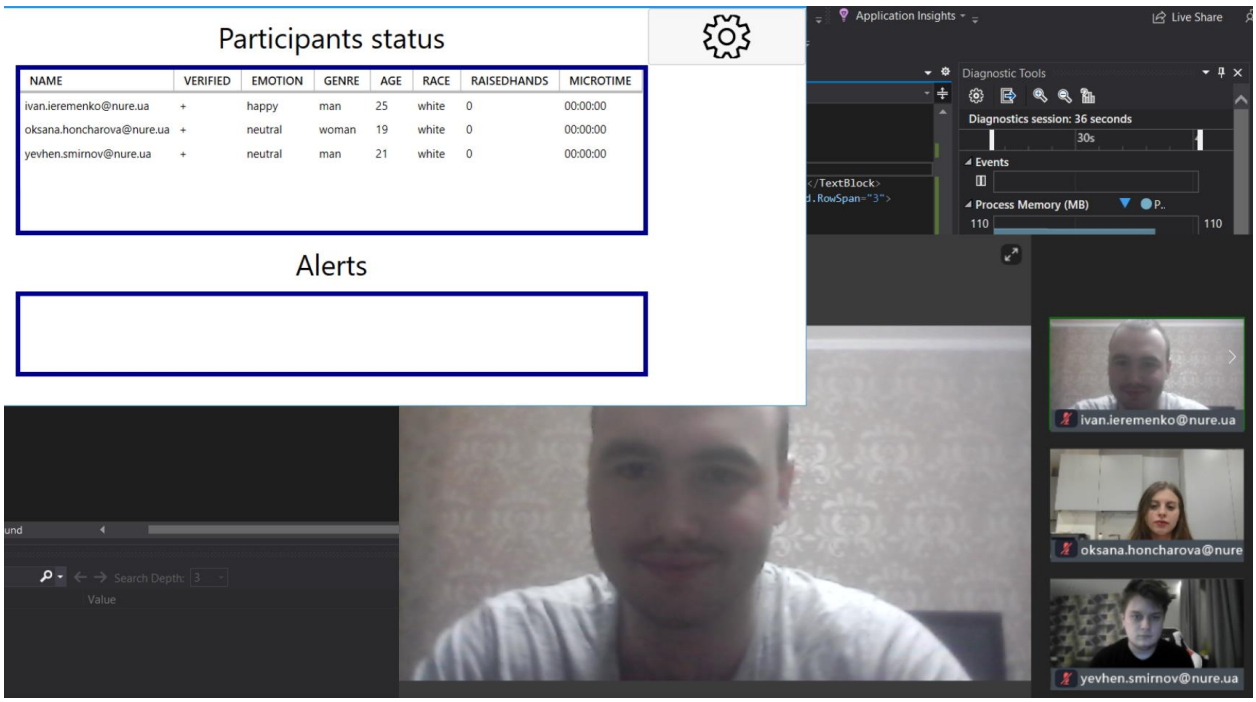


Рисунок 3.11 – Демонстрація роботи сторінки Observe Session

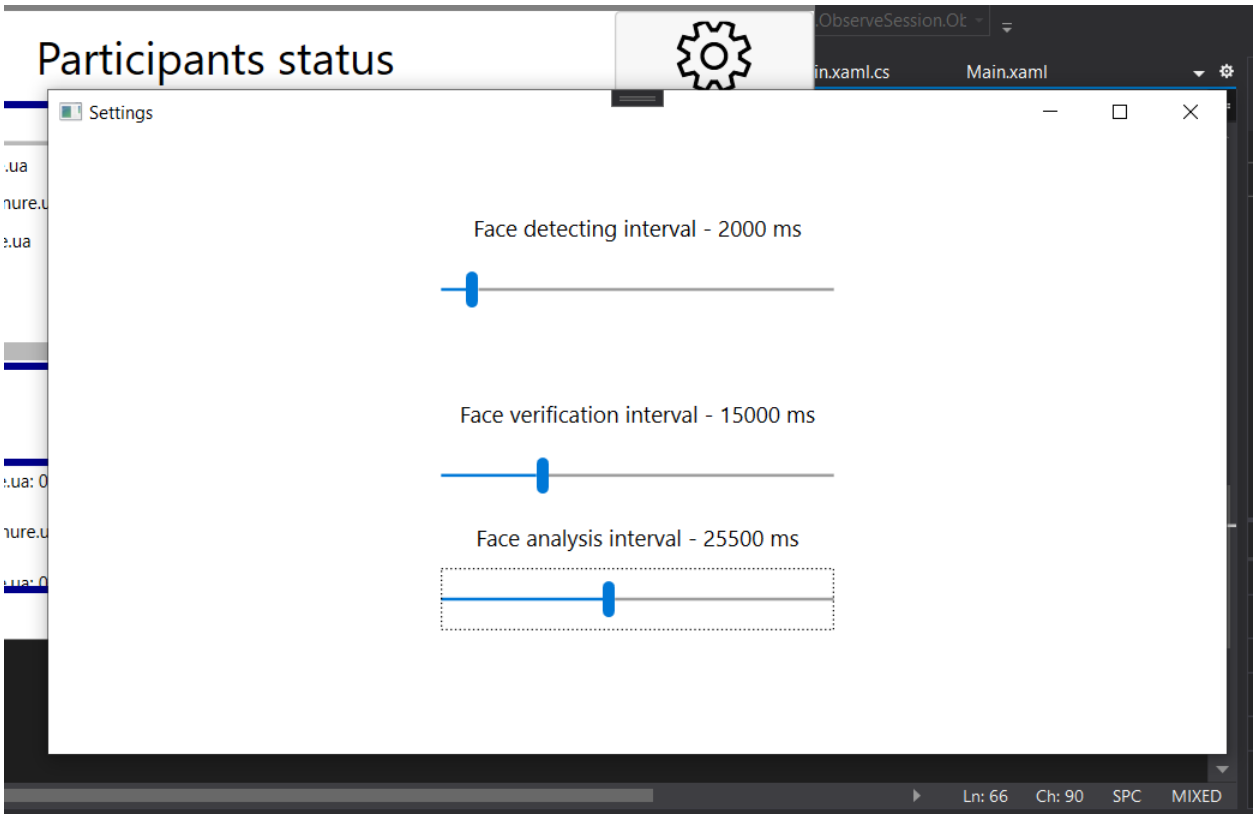


Рисунок 3.12 – Налаштування інтервалів детектування, верифікації, аналізу

Список Alerts відповідає за оповіщення власника конференції про порушення правил конференції, а саме: кількість осіб в кадрі менше (рис. 3.13) або більше (рис. 3.14) за одну людину, провалена верифікація.

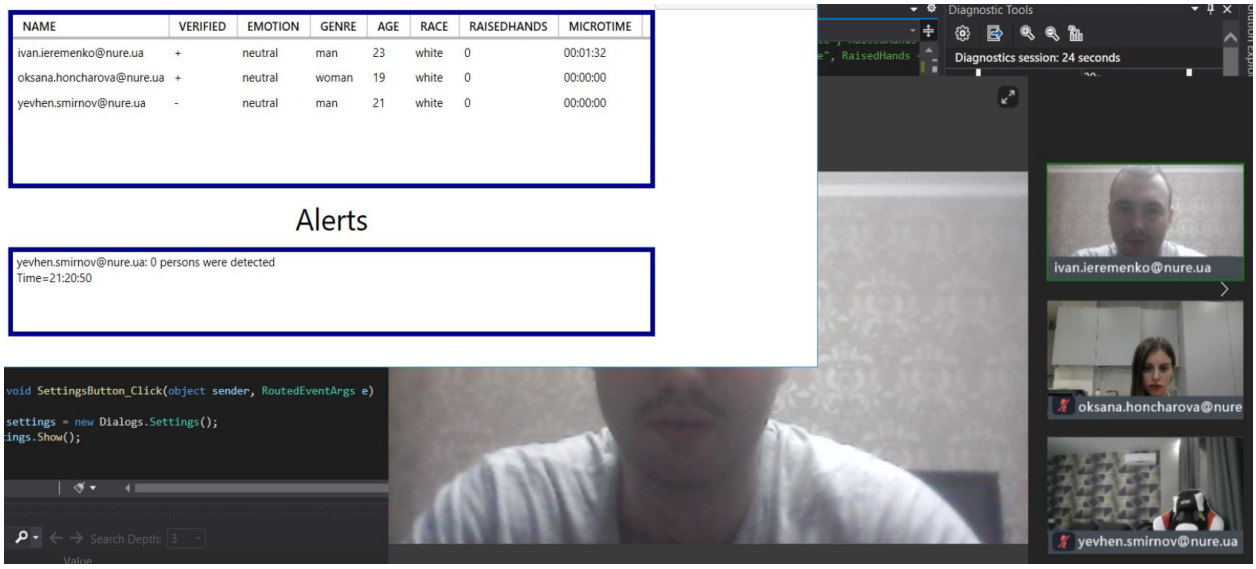


Рисунок 3.13 – Оповіщення при відсутності учасника конференції в кадрі та нарахування часу при увімкненому мікрофоні

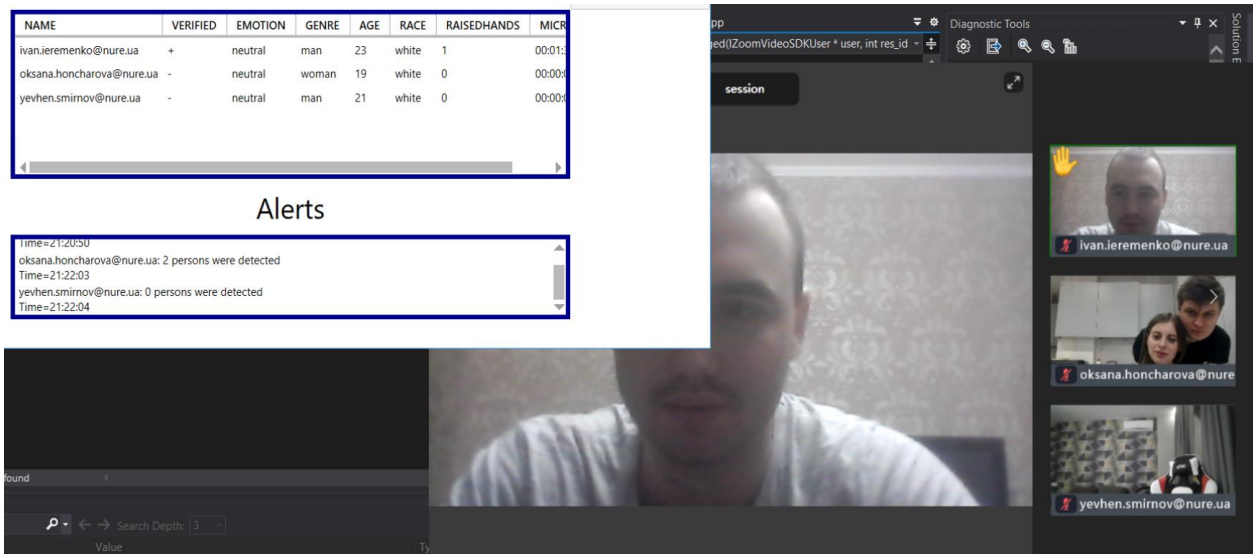


Рисунок 3.14 – Оповіщення при надмірній кількості учасників в кадрі та нарахування кількості підняття рук

Приклад неуспішної верифікації двох учасників відеоконференції наведений на рисунку 3.15.

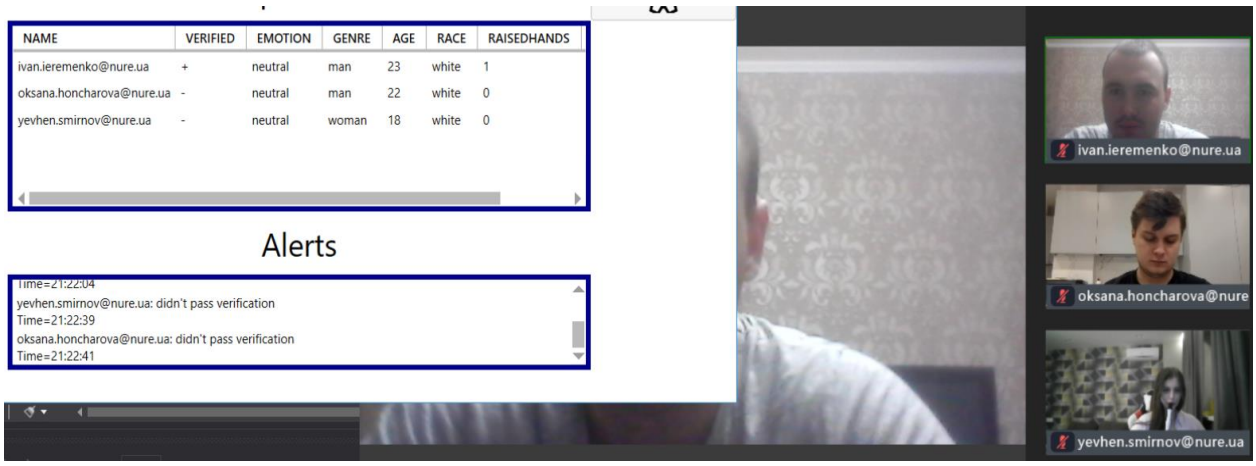


Рисунок 3.15 – Приклад неуспішної верифікації двома учасниками відеоконференції

Приклад розпізнавання емоцій гніву, суму та радості наведений на рисунку 3.16.

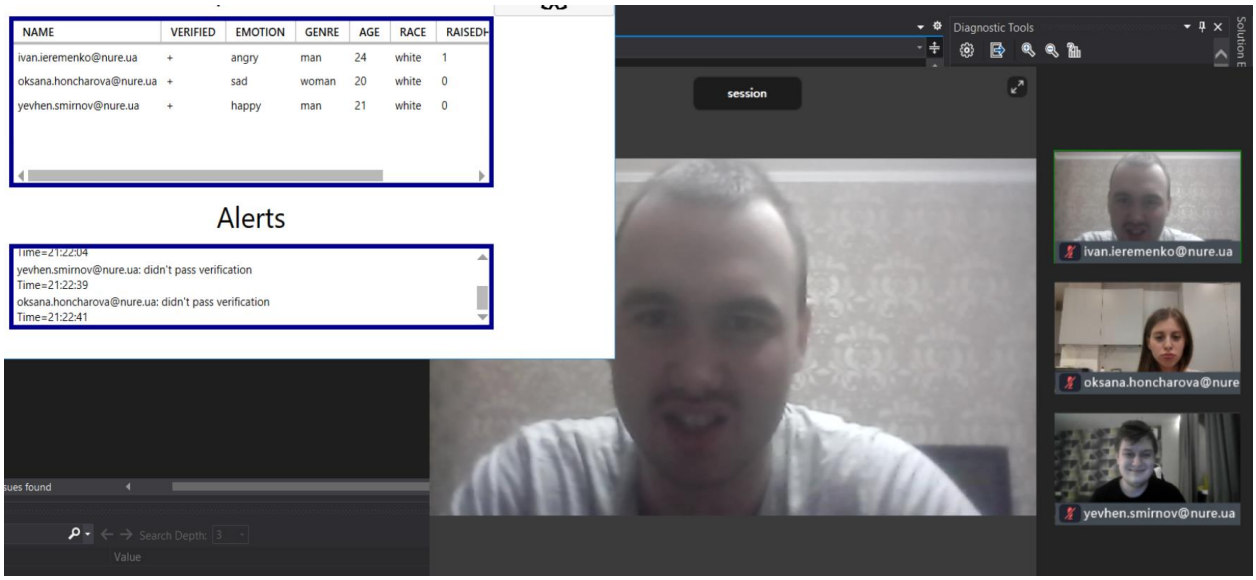


Рисунок 3.16 – Приклад розпізнавання емоцій гніву, суму та радості

По завершенню відеоконференції власник може подивитися статистику на сторінці Statistics. На першій вкладці можна переглянути емоції окремих учасників у відсотковому співвідношенні, кількість успішно пройдених верифікацій, середня відстань під час верифікації, час з увімкненим мікрофоном та кількість підняття рук за деякий проміжок часу (рис. 3.17).

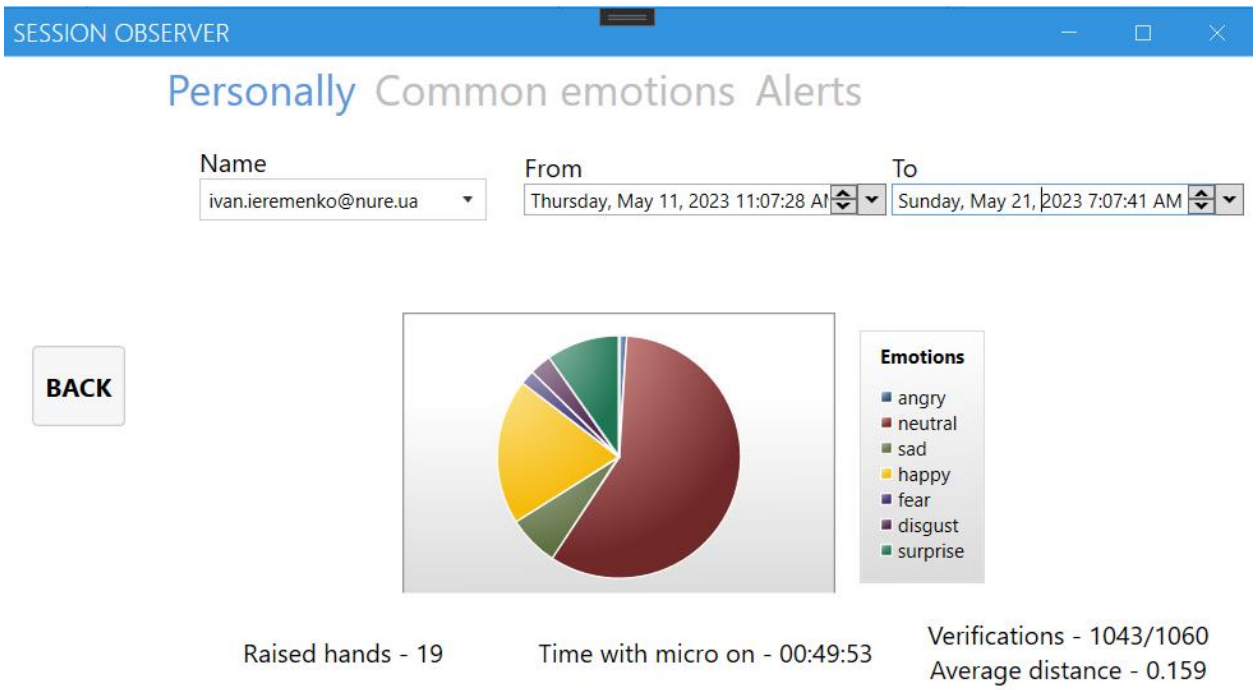


Рисунок 3.17 – Вкладка Personally на сторінці Statistics

Окрім цього можна подивитися графік емоцій всіх учасників за певний проміжок часу у вкладці Common emotions (рис. 3.18).

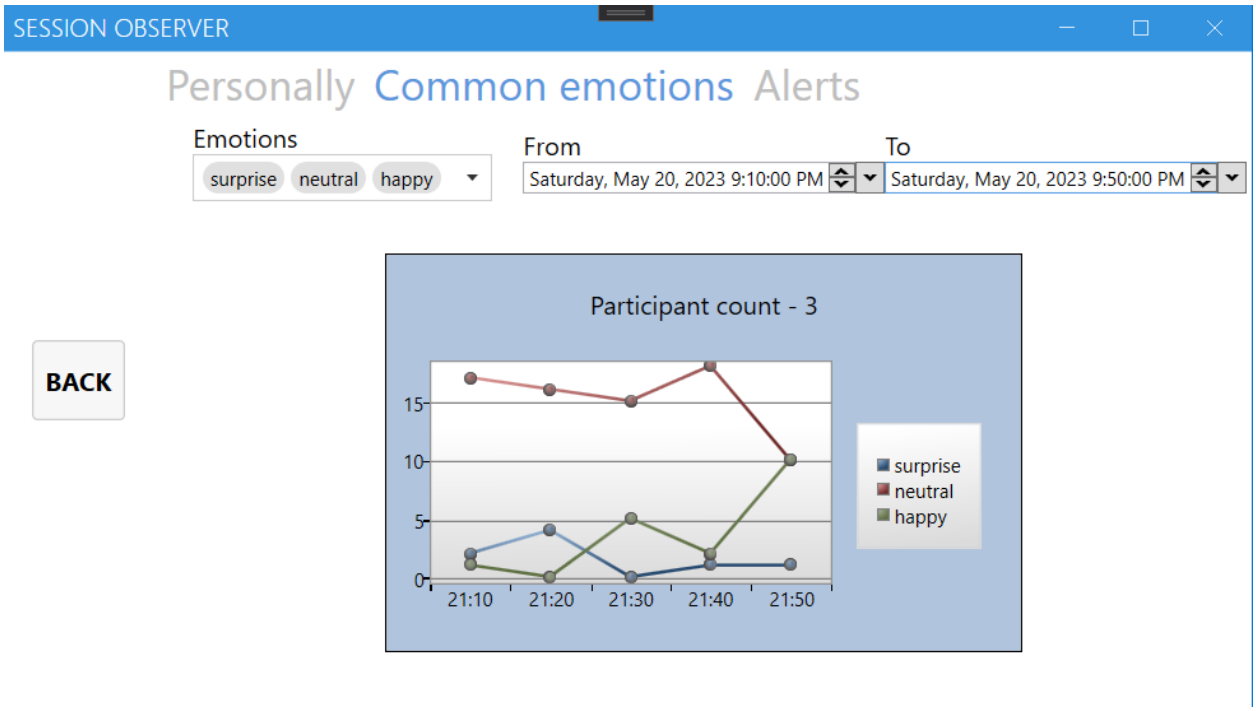


Рисунок 3.18 – Вкладка Common emotions на сторінці Statistics

При побудові застосунку були використані стилі та деякі компоненти (наприклад MultiCombobox) з фреймворку MahApps.Metro. Цей фреймворк дозволяє створювати інтерфейс для WPF, що пришвидшує розробку і є доцільним для використання при побудові невеликих проєктів (рис. 3.19). Для побудови графіків була використана колекція WPF елементів керування Extended WPF Toolkit [38].

На останній вкладці можна подивитися перелік недотримань правил відеоконференції конкретним учасником або всіма одразу за певний проміжок часу (рис. 3.20).

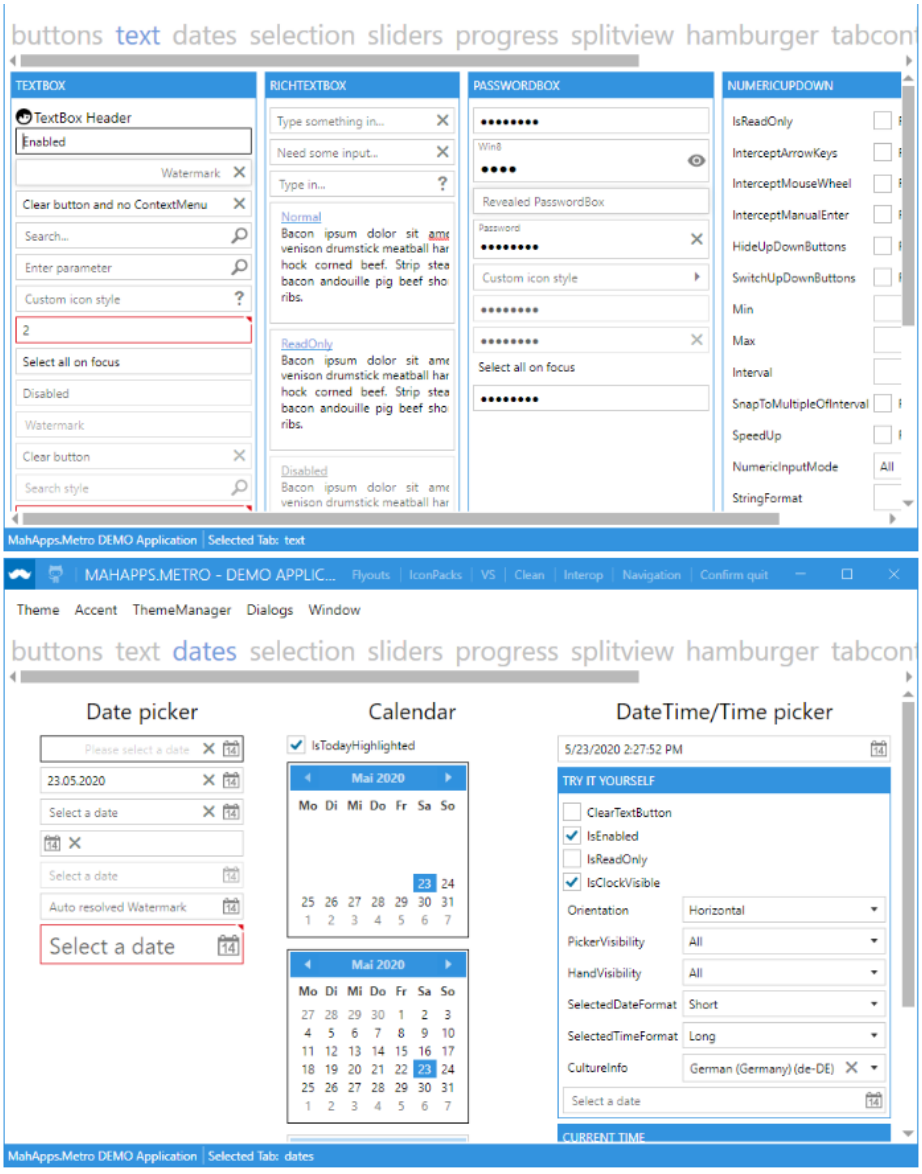


Рисунок 3.19 – Приклад стилей та компонентів MahApps.Metro [39]

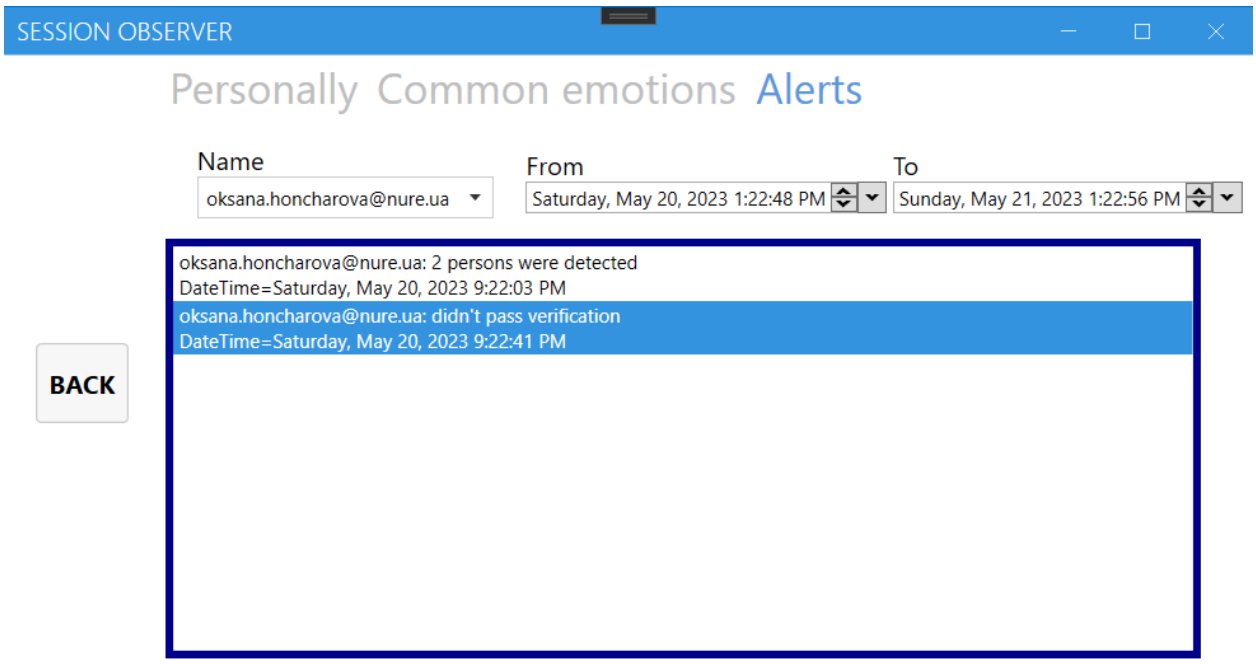


Рисунок 3.20 – Вкладка Alerts на сторінці Statistics

ВИСНОВКИ

У рамках кваліфікаційної роботи був реалізований застосунок для моніторингу дій учасників конференції Zoom.

В ході виконання роботи було досліджено та вирішено теоретичні та практичні питання:

– проаналізовано різні сервіси для відеоконференцій, обрано найбільш підходящу для даної роботи – Zoom Video SDK;

– досліджено та реалізовано перетворення кольорової моделі YUV в RGB;

– опрацьовані основні принципи згорткових нейронних мереж, особливості тренування нейронних мереж та створення датасетів;

– проаналізовано методи детектування, верифікації, обрані найбільш оптимальні з них: SSD, ArcFace;

– ознайомлено з бібліотеками OpenCV, DeepFace та впроваджено детектування, верифікацію та аналіз емоцій у застосунок за допомогою них;

– розроблено вимоги до застосунку, правила проведення відеоконференції, базу даних та архітектуру застосунку;

– реалізовано застосунок та продемонстровано роботу з застосунком.

В результаті тестування розробленого застосунку було показано, що Zoom Video SDK надає відеопотоки з достатньою роздільною здатністю, щоб OpenCV був здатен детектувати обличчя, DeepFace – верифікувати особу та аналізувати емоції.

Використання розробленої системи в навчальному процесі допоможе викладачу як під час проведення заняття, так і надасть змогу зібрати статистику щодо проведених занять на протязі семестра, наприклад, присутність, активність, емоційна складова. Зібрана моніторингова інформація може бути корисною для подальших досліджень взаємодії поведінки студентів під час навчання та їх кінцевим рівнем знань.

Розроблений застосунок працює на платформі Windows. В подальшому бажано б було розробити інтерфейси, які будуть давати можливість приєднуватись користувачам з інших платформ.

Результати роботи апробовано у вигляді 2 тез доповідей під час 14-ої Міжнародної науково-практичної конференції «FREE AND OPEN SOURCE SOFTWARE» [2] та 27-го Міжнародного молодіжного форуму «РАДІОЕЛЕКТРОНІКА І МОЛОДЬ В XXI СТОЛІТТІ» [40].

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Zoom Video SDK. URL: https://developers.zoom.us/docs/video-sdk/ (дата звернення 15.04.23).

2. Єременко, І. О. АНАЛІЗ SDK ПАКЕТІВ ОНЛАЙН КОНФЕРЕНЦІЙ ZOOM ДЛЯ ВИРІШЕННЯ ЗАДАЧІ МОНІТОРИНГУ ДІЙ УЧАСНИКІВ. Матеріали XIV-ої Міжнародної науково-практичної конференції «Free and Open Source Software», Харків, 14-16 лютого 2023 р.–Харків: Харківський національний економічний університет імені Семена Кузнеця, 2023.–110 с., 57.

3. Кобилін, О.А., & Творошенко І.С. (2021). Методи цифрової обробки зображень: навч. посібник. Харків: ХНУРЕ.

4. Elyan, E., Vuttipittayamongkol, P., Johnston, P., Martin, K., McPherson, K., Moreno-García, C. F., ... & Sarker, M. K. (2022). Computer vision and machine learning for medical image analysis: Recent advances, challenges, and way forward. Artificial Intelligence Surgery, 2(1), 24-45.

5. Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., ... & Zhang, K. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. cell, 172(5), 1122-1131.

6. Zou, Z., Chen, K., Shi, Z., Guo, Y., & Ye, J. (2023). Object detection in 20 years: A survey. Proceedings of the IEEE.

7. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

8. Wen, M., & Cho, K. (2023). Object-Aware 3D Scene Reconstruction from Single 2D Images of Indoor Scenes. Mathematics, 11(2), 403.

9. Bradski, G., & Kaehler, A. (2000). OpenCV. Dr. Dobb's journal of software tools, 3(2).

10. Pang, B., Nijkamp, E., & Wu, Y. N. (2020). Deep learning with tensorflow: A review. Journal of Educational and Behavioral Statistics, 45(2), 227-248.

11. Wang, M., & Deng, W. (2021). Deep face recognition: A survey. Neurocomputing, 429, 215-244.

12. Wang, Q., Chen, W., Wu, X., & Li, Z. (2019). Detail-enhanced multi-scale exposure fusion in YUV color space. IEEE Transactions on Circuits and Systems for Video Technology, 30(8), 2418-2429.

13. Koju, R., & Joshi, S. R. (2014). Comparative analysis of color image watermarking technique in RGB, YUV, and YCbCr color channels. Nepal Journal of Science and Technology, 15(2), 133-140.

14. Cherednichenko, O., Vovk, M., Yanholenko, O., & Yakovleva, O. (2020). Towards the technology of employers' requirements collection development. In 67 Integrated Computer Technologies in Mechanical Engineering: Synergetic Engineering (pp. 228-239). Cham: Springer International Publishing.

15. Gorokhovatskyi, O., Peredrii, O., Gorokhovatskyi, V., & Vlasenko, N. (2023). Explanation of CNN image classifiers with hiding parts. In Explainable Deep Learning AI (pp. 125-146). Academic Press.

16. Nadeem, M. W., Goh, H. G., Ali, A., Hussain, M., Khan, M. A., & Ponnusamy, V. A. P. (2020). Bone age assessment empowered with deep learning: a survey, open research challenges and future directions. Diagnostics, 10(10), 781.

17. Cherednichenko, O., Kanishcheva, O., Yakovleva, O., & Arkatov, D. (2020). Collection and Processing of a Medical Corpus in Ukrainian. In COLINS (pp. 272-282).

18. Salmi, N., & Rustam, Z. (2019, June). Naïve Bayes classifier models for predicting the colon cancer. In IOP conference series: materials science and engineering (Vol. 546, No. 5, p. 052068). IOP Publishing.

19. Gorokhovatskyi, V., & Tvoroshenko I. (2020). Image classification based on the Kohonen network and the data space modification.

20. Gorokhovatskyi V., Tvoroshenko I., Kobylin O., and Vlasenko N. (2023) Search for visual objects by request in the form of a cluster representation for the structural image description, Advances in Electrical and Electronic Engineering, 21(1), pp. 19-27.

21. V. Gorokhovatskyi, I. Tvoroshenko (2020). Image Classification Based on the Kohonen Network and the Data Space Modification, in: CEUR Workshop Proceedings: Computer Modeling and Intelligent Systems (CMIS-2020), pp. 1013–1026.

22. Clifton, J., & Laber, E. (2020). Q-learning: Theory and applications. Annual Review of Statistics and Its Application, 7, 279-301.

23. Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 815-823).

24. Kobylin, O., & Lyashenko, V. (2020). Time series clustering based on the k-means algorithm.

25. Daradkeh, V. Gorokhovatskyi, I. Tvoroshenko, S. Gadetska, M. Al-Dhaifallah, (2021). Methods of classification of images on the basis of the values of statistical distributions for the composition of structural description components, IEEE Access 9 92964–92973.

26. Yakovleva, O., Kovtunenko, A., Liubchenko, V., Honcharenko, V., & Kobylin, O. (2023). Face Detection for Video Surveillance-based Security System (COLINS-2023). In CEUR Workshop Proceedings (Vol. 3403). pp. 69-86.

27. Rahmad, C., Asmara, R. A., Putra, D. R. H., Dharma, I., Darmono, H., & Muhiqqin, I. (2020). Comparison of Viola-Jones Haar Cascade classifier and histogram of oriented gradients (HOG) for face detection. In IOP conference series: materials science and engineering (Vol. 732, No. 1, p. 012038). IOP Publishing.

28. Zhang, N., Luo, J., & Gao, W. (2020, September). Research on face detection technology based on MTCNN. In 2020 international conference on computer network, electronic and automation (ICCNEA) (pp. 154-158). IEEE.

29. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). Ssd: Single shot multibox detector. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14 (pp. 21-37). Springer International Publishing.

30. Liu, Y., Liu, R., Wang, S., Yan, D., Peng, B., & Zhang, T. (2022). Video face detection based on improved ssd model and target tracking algorithm. Journal of Web Engineering, 545-568.

31. Yang, S., Luo, P., Loy, C. C., & Tang, X. (2016). Wider face: A face detection benchmark. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5525-5533).

32. Deepface. URL: https://github.com/serengil/deepface (дата звернення 01.05.23).

33. Yakovleva, O., & Nikolaieva K. (2020). Research of descriptor based image normalization and comparative analysis of SURF, SIFT, BRISK, ORB, KAZE, AKAZE descriptors. Advanced Information Systems, 4(4), 89-101.

34. Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4690-4699).

35. Xu, X., Du, M., Guo, H., Chang, J., & Zhao, X. (2020). Lightweight facenet based on mobilenet. International Journal of Intelligence Science, 11(1), 16.

36. Sampaio, E. V., Lévêque, L., da Silva, M. P., & Le Callet, P. (2022, June). Are facial expression recognition algorithms reliable in the context of interactive media? A new metric to analyse their performance. In EmotionIMX: Considering Emotions in Multimedia Experience (ACM IMX 2022 Workshop).

37. FER2013 (Facial Expression Recognition 2013 Dataset). URL: https://paperswithcode.com/dataset/fer2013 (дата звернення 03.05.23).

38. Extended WPF Toolkit. URL: https://github.com/xceedsoftware/wpftoolkit (дата звернення 09.05.23).

39. MahApps.Metro. URL: https://github.com/MahApps/MahApps.Metro (дата звернення 09.05.23).

40. Єременко І. О., Яковлева О.В. (2023), РОЗРОБКА СЕРВІСУ ДЛЯ МОНІТОРИНГУ ДІЙ УЧАСНИКІВ КОНФЕРЕНЦІЇ ZOOM. 27-ий міжнародний молодіжний форум «РАДІОЕЛЕКТРОНІКА І МОЛОДЬ У XXI СТОЛІТТІ», С. 48-49.