

ІДЕНТИФІКАЦІЯ ОСОБИ У ВІДЕОПОТОЦІ МЕТОДАМИ МАШИННОГО НАВЧАННЯ

І.С. Кунак, І.М. Шпінарева, В.Г. Пенко

Одеський національний університет імені І. І. Мечнікова, вул.Дворянська,2, м.Одеса, Україна,
e-mail: kunak19999@gmail.com, iryna.shpinareva@onu.edu.ua, vpenko@onu.edu.ua

Ідентифікація особи у відеопотоці в режимі реального часу є критично важливою функцією в широкому спектрі актуальних предметних областей. З іншого боку, таке завдання є складним і може бути ефективно вирішене лише за допомогою сучасних підходів штучного інтелекту. У цій роботі в якості ключового алгоритму розпізнавання та векторизації зображення обличчя використовується апробований раніше підхід МТСNN, який використовує каскад з декількох нейромереж для виявлення ключових точок обличчя ідентифікації особи. Крім основного алгоритму, була реалізована інтегрована система, що забезпечує функціональність ідентифікації особи, доступної кінцевому користувачеві. Додатково реалізовані модуль ідентифікації осіб на основі векторів, що представляють особистість у відеокадрі, підсистема зберігання та поповнення бази даних осіб, які розпізнаються, і зручний інтерфейс користувача. Модифікації архітектури МТСNN, запропоновані в даній роботі, дозволили досягти продуктивності, що забезпечує режим реального часу та прийнятну якість ідентифікації на рівні 0.92 за F1-метрикою.

Ключові слова: ідентифікація особистості, ключові точки обличчя, каскадні згорткові нейромережі.

Вступ

Останнім часом в різних сферах діяльності зростає потреба швидкої та правильної ідентифікації особистості людини у відеопотоці, особливо в режимі реального часу. Практичне застосування системи розпізнавання осіб є вкрай важливим або критично необхідним в безлічі сфер:

1) комерція (запобігання магазинних крадіжок і організованої злочинності, підрахунок кількості відвідувачів, персоналізація клієнтського досвіду, каси самообслуговування);

2) банківська сфера та блокчейн системи (імплементаций методу Know Your Customer, біометричні банкомати, боротьба з шахрайством і мінімізація ризиків);

3) охорона порядку в містах та безпека державних кордонів (системи біометричної ідентифікації (ABIS) для пошуку злочинців, докази злочинів та їх запобігання, наприклад терактів або шахрайства з документами, прикордонний і митний контроль);

4) інші системи розпізнавання осіб (біометричний облік робочого часу, контроль доступу у лікарнях та будинках для людей похилого віку, авторизація відвідувачів).

Пандемія COVID стала каталізатором швидкого розповсюдження рішення, коли сотні муніципалітетів у всьому світі встановили камери, які оснащені технологією розпізнавання облич. Минулого року у Китаї понад 100 міст було обладнано системами спостереження на основі технології розпізнавання осіб. Використання загальноміської системи відеоспостереження під час карантину у зв'язку з COVID дозволило виявити порушників карантину [1].

У березні 2021 року аналітична компанія Mordor Intelligence представила звіт [2], в якому світовий ринок розпізнавання осіб оцінювався в 3,72 млрд. доларів США в 2020 році і перевищить 11,62 млрд. доларів США в 2026 році.

Ці та багато інших фактів дозволяють констатувати, що ідентифікація особи на основі її візуального представлення є важливим випадком більш загального напрямку розпізнавання візуальних образів, який в свою чергу є предметною областю для застосування сучасних методів дослідження що належать до категорії методів штучного інтелекту та машинного навчання.

Важливою запорукою успішності розробки у цьому напрямі є визначення контексту та характерних обмежень задачі. У даній роботі маються на увазі наступні особливості: 1) вхідною інформацією є відеопотік, якість якого забезпечується стандартною технікою відео фіксації; 2) ідентифікація здійснюється на основі аналізу обличчя особи; 3) необхідно забезпечити оперативну ідентифікацію.

Незважаючи на актуальність, значну кількість сфер застосування та величезну кількість досліджень в аналогічних обставинах, проблема якісного і швидкого розпізнавання осіб в відеопотоці все ще залишається не повністю вирішеною.

Огляд літератури

Задачу пошуку обличчя на зображенні на сьогоднішній день успішно вирішено та існують практичні реалізації та застосування у багатьох технічних пристроях. Задачі автоматизованого ідентифікації осіб на основі аналізу обличчя у відеопотоці є більш складними, і на даний момент існуючі алгоритми наближаються до точності розпізнавання людини тільки в ідеалізованих умовах, або вимагають обчислювальні ресурси, недоступні при обробці відеопотоку в режимі реального часу.

Достатньо велика кількість досліджень з вирішення представленої задачі дозволяє сформулювати доволі сталу декомпозицію задачі ідентифікації особи у відеопотоці на кілька наступних етапів:

- 1) пошук області обличчя на зображенні;
- 2) спостереження за положенням виявленої особи на послідовності кадрів;
- 3) попереднє оброблення зображення особи з метою компенсації умов зйомки;
- 4) порівняння зображення обличчя із зображеннями в базі зображень.

У роботі [3] представлено методи розробки ідентифікації особи у відеопотоці. Запропонований алгоритм базується на анізотропній дифузії як методі попередньої обробки зображення, вейвлет-перетворенні Габора як методі обробки зображень, гістограмі орієнтованих градієнтів (HOG) та локальних бінарних моделях в одновимірному просторі (1DLBP) як методах вилучення вектора ознак із зображення, квадратна метрика евклідової відстані для векторної класифікації. Результати експерименту на базі даних облич і бази даних FERET дають від 70 до 72,5% правильно ідентифікованих зображень. Можна зробити висновок, що даний метод не є ефективним, оскільки точність ідентифікації досить низька.

Задачі ідентифікації людини по обличчю переважно вирішуються з використанням нейромережових моделей розпізнавання. Як класифікатори розглядаються в основному п'ять архітектур нейронних мереж: тривимірна згорткова нейронна мережа, згортково-рекурентні мережі, такі як однонаправлений і двонаправлений LSTM, однонаправлений і двонаправлений GRU у поєднанні з згортковою нейронною мережею архітектури ResNet, використовуваної в цих отримання візуальних ознак.

Великі компанії активно розвивають ринок аналізу даних, найчастіше спираючись на такі моделі, класифікації як One vs One (Face ID) та One vs All [4].

Представлена в роботі модель займає перспективну нішу роботи з невеликими групами осіб, що детектуються та ідентифікуються, тобто вона націлена на вирішення задачі one vs all в умовах недостатності даних, характерної для моделі one vs one.

Мета та задачі дослідження

Метою роботи є розробка інтегрованої системи ідентифікації обличчя шляхом використання перспективних та вже запропонованих в дослідженнях алгоритмів та моделей з можливістю покращення ефективності розпізнавання та ідентифікації людини в відеопотоці шляхом модифікації використаних алгоритмів.

Під підвищенням ефективності розуміється поліпшення наступних характеристик: зниження відсотка неправильних виявлень, збільшення швидкодії, тобто зменшення часу виявлення. Основою алгоритму є згорткові нейронні мережі Multi-task Cascaded Convolutional Network (MTCNN) [5], які протягом останніх років показали себе як найкращі засоби для класифікації об'єктів (в даному випадку обличчя).

Виходячи з такої мети було спроектовано послідовність задач:

- 1) дослідити метод MTCNN для розпізнавання обличчя в відеопотоці;
- 2) обрати гіперпараметри CNN та вибірку навчання;
- 3) реалізувати попередню обробку кадрів для їх використання в CNN;
- 4) реалізувати систему розпізнавання обличчя в відеопотоці та ідентифікації людини.
- 5) провести тестування CNN та розробленої системи в цілому.

Метод розпізнавання обличчя

В даному контексті під розпізнаванням розуміємо виявлення в відеокадрі обличчя та перетворення його у характерний вектор ознак. У роботі розглядається відеопотік формату MP-4. Кожен п'ятнадцятий кадр проходить обробку і виконується ідентифікація всіх осіб на кадрі.

Кадр спочатку проходить створення піраміди масштабування зображення (масив масштабів від більшого до меншого). Далі виконується перетворення BGR формату кадру в RGB та нормалізація значень для використання в нейромережі MTCNN (під нормалізацією тут мається на увазі перетворення відеоданих з діапазону 0...255 в діапазон -1.0...1.0).

Попередньо оброблений відеокадр надходить на вхід мережі MTCNN. На виході отримуємо нормалізоване зображення обличчя розміром 182x182 пікселів. MTCNN – це багатозадачна згорткова нейронна мережа, яка шляхом послідовного розв'язання кількох задач та використанні більш потужної згорткової нейронної мережі підвищує точність детектування особи на основі п'яти лицьових точок. Варто відзначити, що важливою частиною роботи є налаштування розміру вихідних шарів та функції активації. Для нейромережевої архітектури MTCNN було зменшено кількість нейронів вихідних шарів. У наслідок чого, перша і друга нейронні мережі не визначають ключові точки на відміну від MTCNN, а лише визначають обмежуючі прямокутники обличчя та класифікують обличчя. Використовується функція активації PReLU, замість звичайної ReLU.

Робота багатозадачної згорткової нейронної мережі розбита на кілька етапів. На першому етапі використовується повна згорткова мережа Proposal Network. В результаті її роботи розраховуються всі можливі варіанти зображень обличчя (пропозиції), що їх обмежує прямокутник регресії векторів. Потім пропозиції калібруються на основі оцінних векторів регресії обмежуючого вікна. Далі використовується метод NonMaximum Suppression (NMS), який об'єднує пропозиції, що сильно перекриваються.

На другому етапі всі пропозиції відправляються в мережу Re-fine Network (R-Net), яка додатково відхиляє більшість помилкових пропозицій та виконує калібрування з повторним стиском обмежуючого вікна і проводить NMS.

На третьому етапі дії аналогічні до другого етапу, але використовується велика область обличчя. В результаті мережа виведе п'ять лицьових точок.

Для MTCNN обрана переднавчена модель, яка була навчена на навчальній вибірці «VGGFace2» для пошуку обличчя на зображеннях [6]. Набір даних VGGFace2 складається з приблизно 3,31 мільйона зображень.

Метод ідентифікації особи

Після того, як на кадрі виявлено обличчя, далі необхідно виконати ідентифікацію особи. Метод ідентифікації складається з трьох етапів.

Перший етап – вирівнювання. Використовуючи отримані ключові точки, а саме точки очей, виконується вирівнювання шляхом побудови прямокутного трикутника, точки і відстань між очима являється гіпотенузою, висота пускається з найвищої по осі у точки. А далі за відношеннями сторін знаходиться кут між основою та гіпотенузою і таким чином визначається кут нахилу голови в горизонтальній площині. Після визначення кута нахилу виконується вирівнювання (поворот) зображення.

Другий етап – обчислення вектору ознак. Отримане вирівняне зображення обличчя додатково обробляється (так само як в каскаді CNN блоку пошуку обличчя) та відправляється на вхід згорткової нейромережі, яка перетворює зображення в вектор ознак обличчя. Для цього перетворення використовується вже навчена модель FaceNet-512, побудована на архітектурі Inception ResNet-v1 [7]. На виході FaceNet-512 видає вектор ознак обличчя з 512 чисел.

Третій етап – пошук найближчих векторів в БД (найбільш схожих облич) використовуючи метод k-найближчих сусідів. Дистанція між векторами обчислюється евклідовою відстанню.

Для CNN, що створює вектор ознак з обличчя, була обрана переднавчена модель 20180408-102900 [8]. Для навчання та тестування моделі було використано набір даних CASIA-WebFace [9]. Цей навчальний набір складається з 453 453 зображень із 10 575 унікальними особами для ідентифікації.

Система ідентифікації осіб

Представлені вище алгоритми були імплементовані у вигляді інтегрованої системи для ідентифікації людини в відеопотоці в режимі реального часу, наприклад для працівників служби поліції чи охорони. Для роботи працівника (клієнта) з системою ідентифікації розроблено програмний інтерфейс, який представляє собою web-додаток. У технологічну основу системи покладена клієнт-серверна технологія обробки даних, за якої клієнт-комп'ютер звертається за необхідними даними до головного комп'ютера (сервера), при цьому власна обробка даних може виконуватися як на клієнтському, так і серверному комп'ютері. В якості архітектури покладена трірівнева клієнт-серверна архітектура, що передбачає наявність наступних компонентів програми: клієнтський застосунок (зазвичай говорять «тонкий клієнт» або термінал), підключений до сервера застосунків, який в свою чергу підключений до серверу бази даних.

Взаємодія клієнта з веб-додатком здійснюється через графічний інтерфейс в якому розроблено наступний функціонал:

- 1) авторизація для доступу до власного кабінету та власної бази осіб;
- 2) додавання осіб в базу шляхом додавання фото (на якому каскад CNN знаходить обличчя, і CNN знаходить вектор обличчя для занесення в БД);
- 3) завантаження відео для здійснення пошуку та ідентифікації особистостей в ньому (знайдені особи записуються в історію для даного відео, а не ідентифіковані особи додаються в загальний список не знайдених обличчя по даному відео);
- 4) перегляд інформації про осіб доданих в БД;
- 5) перегляд детальної інформації про особу;
- 6) перегляд історії знаходжень.

Для забезпечення безпеки даної системи вхід до системи виконується через логін та пароль.

Загальна архітектура роботи системи розпізнання та ідентифікації людини в відеопотоці mp4 представлена на рис. 1. Подальші приклади відеопотоку для аналізу взяті з англійського серіалу про Шерлока Холмса та особи (актори) позначені їх оригінальними англійськими іменами. Веб-інтерфейс застосунків, також розроблено

англійською мовою. Важливими з точки зору алгоритмів ідентифікації є модуль розпізнавання обличчя в кадрі та модуль ідентифікації людини.

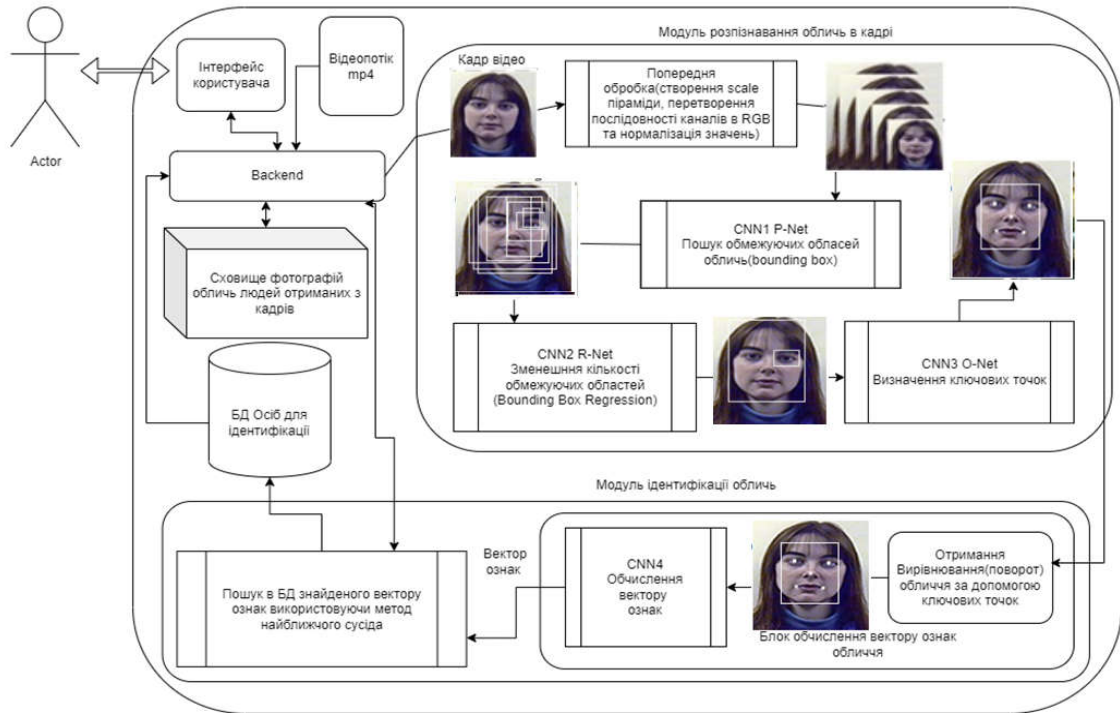


Рис. 1. Схема роботи системи

Для створення веб-додатку та CNN була вибрана мова Python та фреймворк Tensorflow разом з API Keras, так як вони містять повний набір функцій і методів потрібних для створення і навчання CNN. На рис. 2 та 3 представлено результат роботи системи розпізнавання та ідентифікації осіб.

Video Processing Web App Persons NFFaces RealTime Video Upload Video About Login Register

Select a video to upload and play

Sources List

Id	Name	Created time	Actions
Id 1	Friends_FunnyParts.mp4	2021-11-06 18:56:27.510092	Load, Process
Id 2	Sherlock_1.mp4	2021-11-10 23:18:08.550028	Load, Process
Id 3	Sherlock_2.mp4	2021-11-11 00:00:20.153073	Load, Process

Live Streaming

Founded Persons List

Id Loo Brealey	Distance: 0	Db_id 14
Id Loo Brealey	Distance: 0.3927670419216156	Db_id 14
Id Loo Brealey	Distance: 0.3927670419216156	Db_id 14
Id Benedict Cumberbatch	Distance: 0.4768138825893402	Db_id 11
Id Benedict Cumberbatch	Distance: 0.4768138825893402	Db_id 11

Not found Faces List

Id 0	Create Time 96	Confidence 0.8263257145881653	Mouth left: 343,158 Mouth right: undefined Left eye 347,131 Right eye 367,135	Light
Id 0	Create Time 144	Confidence 0.9638354778289795	Mouth left: 210,147 Mouth right: undefined Left eye 206,118 Right eye 231,117	Light
Id 0	Create Time 168	Confidence 0.9778928160667419	Mouth left: 213,136 Mouth right: undefined Left eye 212,106 Right eye 237,107	Light

Рис. 2. Результат роботи системи розпізнавання та ідентифікації осіб

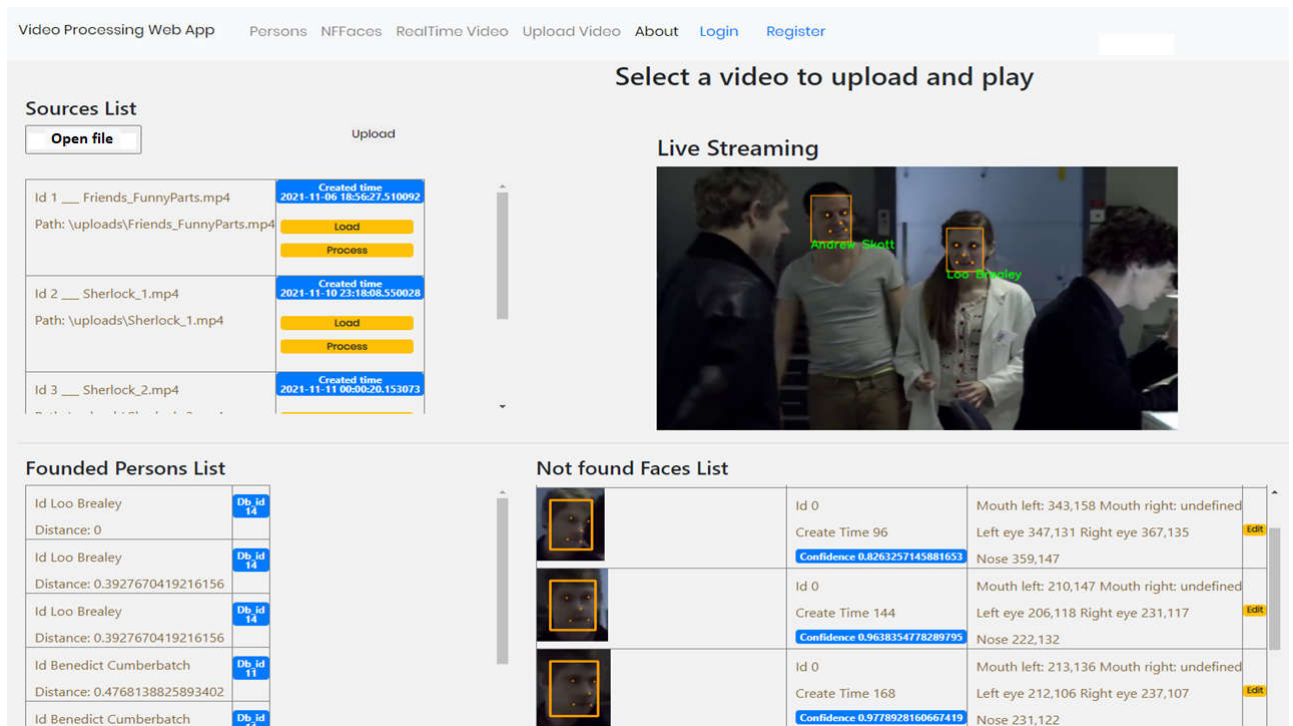


Рис. 3. Ідентифікована особа

Тестування системи

Метою тестування системи було виявлення ефективності розпізнавання обличчя та швидкості розпізнавання та ідентифікації особи.

Для тестування на точність обраний датасет Fddb – це набір даних про області обличчя, розроблений для вивчення проблеми безобмеженого виявлення обличчя [10]. Цей набір даних містить анотації для 5171 обличчя в наборі з 2845 зображеннями, взятими з набору даних Faces in the Wild. В якості метрик визначення точності розпізнавання було взято F1-Score. Оцінка F1-Score, також відома як збалансована оцінка F-міра. Оцінку F1-Score можна інтерпретувати як середнє гармонійне значення точності та запам'ятовування, де оцінка F1-Score досягає свого найкращого значення при 1, а найгіршого – при 0. У випадку з кількома класами та кількома мітками це середнє значення F1 кожного класу із зважуванням залежно від середнього параметра. Загалом коливання в збалансованій оцінці майже не значні і згідно з результатами знаходяться біля 92%.

Тестування швидкості роботи системи. В якості об'єкту тестування модуля розпізнавання обличчя на кадрах вибраний відео-уринок з відомого серіалу про Шерлока Холмса.

Інформація про відео:

- тривалість: 6.23 хв.;
- якість відео: 1280*720;
- загальна кількість кадрів: 9204;
- кількість фреймів на секунду(fps): 30.

Тестування на швидкість проводилось за двох наборів параметрів роботи модуля розпізнавання.

Перший набір це: мінімальний розмір обличчя: 30*30 px, фактор масштабування: 0.709 (впливає на швидкість сходження до мінімального масштабу та відповідно на кількість масштабованих зображень) та поріг допустимих значень 0.8.

Запропонована система була протестована та отримана статистика в порівнянні з схожим стандартним каскадом нейромереж MTCNN (таблиця 1).

Таблиця 1

Порівняння характеристик першого тестування

	Реалізований MTCNN з модифікаціями	Стандартний MTCNN	Різниця
Середній час обробки кадру(сек)	0.6207947	0.6174478	+0.542%
Середній час обробки кадру з обличчями(сек)	0.6373271	0.6340739	+0.513%
Середній час обробки кадру без обличь(сек)	0.58191427	0.5816137	+0.0516%

Другий набір це: мінімальний розмір обличчя: 30*30 px, фактор масштабування: 0.6 (впливає на швидкість сходження до мінімального масштабу та відповідно на кількість масштабованих зображень) та поріг допустимих значень 0.8. Порівняльна характеристика також в співставленні зі стандартною MTCNN наведена в таблиці 2.

Таблиця 2

Порівняння характеристик другого тестування.

	Реалізований MTCNN з модифікаціями	Стандартний MTCNN	Різниця
Середній час обробки кадру(сек)	0.42758628	0.4602320	-7.093%
Середній час обробки кадру з обличчями(сек)	0.4587248	0.4757354	-3.5756%
Середній час обробки кадру без обличь(сек)	0.4100261	0.43184338	-5.052%

Перевірка точності розпізнавання та ідентифікації осіб залежить від якості кадру. Якщо зображення зроблені з хорошими умовами для розпізнавання (достатньо якісне зображення області обличчя), то якість ідентифікації осіб є 92% (рис. 4)

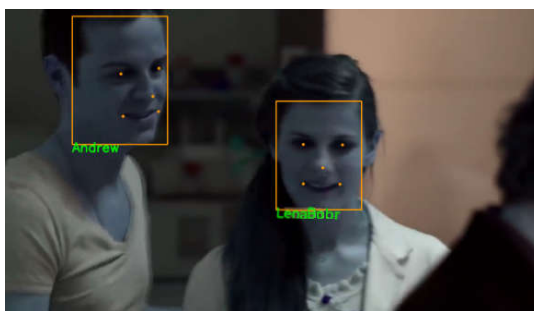


Рис. 4. Приклад зображення з хорошими умовами для розпізнавання

Якщо великий кут повороту обличчя, та погана якість області обличчя (розмиття в русі) то точність ідентифікації низька (рис.5).



Рис. 5. Приклад зображення з поганими умовами для розпізнавання.

Висновки

У роботі запропоновано інтегровану систему, яка виконує ідентифікацію особи, що з'являється у відеопотоці. Основні алгоритми, що використовуються в процесі функціонування системи достатньо відомі та верифіковані. Зокрема, для пошуку та формування характерного векторного представлення обличчя було використано каскад згорткових нейронних мереж МТСNN, а для ідентифікації людини по вектору представлення обличчя – метод k-найближчих сусідів. Після знаходження у кадрі нейронною мережею МТСNN ключових точок обличчя за допомогою CNN обчислюється вектор ознак. Для ідентифікації особи використовується метод k-найближчих сусідів, який шляхом обчислення відстані між шуканим і відомими векторами знаходить найближчий ідентифікатор людини. Порівняно зі стандартною реалізацією МТСNN в запропонованій системі було запроваджено декілька модифікацій (використано функції активації PReLU, замість звичайного ReLU та зменшено кількість вихідних нейронів у мережах CNN). В результаті нейронні мережі Proposal Network, R-Net в роботі лише визначають обмежуючі прямокутники обличчя, а не визначають ключові точки на відміну від МТСNN. Запропоновані модифікації дозволили збільшити швидкість розпізнавання осіб за для забезпечення оперативності ідентифікації особи. Якість розпізнавання та ідентифікації осіб залежить від якості кадру та орієнтації обличчя особи. Якщо зображення з хорошими умовами для розпізнавання (достатньо якісне зображення області обличчя), то якість ідентифікації осіб є 92%. Таким чином, запроваджену систему можна розглядати в якості прототипу прикладної системи для кінцевого користувача. Потенціал використаних алгоритмів дозволяє поліпшити показники функціонування завдяки пошуку більш ефективної комбінації гіперпараметрів.

Список літератури

1. Thompson B., Van Noord R. *The troubling rise of facial recognition technology*. *Nature*. November 18, 2020. URL: <https://www.nature.com/articles/d41586-020-03271-8>.
2. Facial recognition market - growth, trends, Covid-19 impact, and forecasts (2021 - 2026). *Mordor Intelligence*. March 2021. URL: <https://www.mordorintelligence.com/industry-reports/facialrecognition-market>.
3. Bychkov O., Merkulova K., Zhabska Y., Shatyro A.. Development of Information Technology for Person Identification in Video Stream. *II International Scientific Symposium «Intelligent Solutions» IntSol-2021*. URL: <http://ceur-ws.org/Vol-3018/Paper7.pdf>
4. Pawara P., Okafor E., Groefsema M. H., Lambert S., Schomaker R.B, Wiering M.A. One-vs-One classification for deep neural networks. *Pattern Recognition*. 2020. V.108, DOI: <https://doi.org/10.1016/j.patcog.2020.107528>
5. Zhang K., Zhang Z., Li Z., Qiao Y. Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks. *IEEE Signal Processing Letters (SPL)*. 2016. V. 23, No. 10. P.1499–1503.
6. VGGFace2. URL: <https://paperswithcode.com/dataset/vggface2-1>
7. Schroff F., Kalenichenko D., Philbin J. FaceNet: A Unified Embedding for Face Recognition and Clustering. *Computer Vision and Pattern Recognition*. 2015. DOI: 10.1109/CVPR.2015.7298682
8. Facenet-20180408-102900. Use Case and High-Level Description. URL: https://docs.openvino.ai/latest/omz_models_model_facenet_20180408_102900.html
9. CASIA-WebFace. URL: <https://paperswithcode.com/dataset/casia-webface>
10. Face Detection Data Set and Benchmark Home. URL: <http://vis-www.cs.umass.edu/fddb/>

PERSON IDENTIFICATION IN A VIDEO STREAM BY MACHINE LEARNING METHODS

I.C. Kunak, I.M. Shpinareva, V.G. Pienko

Odessa I. I. Mechnikova National University, 2 Dvoryanska str., Odessa, Ukraine,
 e-mail: kunak19999@gmail.com, iryna.shpinareva@onu.edu.ua, vpenko@onu.edu.ua

Successful identification of a person in a real-time video stream is a critical function in a wide range of relevant subject areas. On the other hand, such a task is complex and can be effectively solved only with the help of modern artificial intelligence approaches. In this work, the previously developed MTCNN approach is used as a key algorithm for face image recognition and vectorization, which uses a cascade of several neural networks to identify key points of the face of a recognizable personality. In addition to the main algorithm, an integrated system was implemented that provides the functionality of person identification available to the end user. Additionally, face identification modules were implemented based on vectors representing a person in a video frame, a subsystem for storing and replenishing the database of recognizable persons, and a convenient user interface. Modifications of the MTCNN architecture proposed in this paper made it possible to achieve real-time performance and acceptable identification quality at the level of 0.92 according to the F1-metric.

Keywords: person identification, face key points, cascaded convolutional neural networks.