



СИСТЕМА ВИЯВЛЕННЯ ОБЛИЧЧЯ НА ЗОБРАЖЕННІ З ВИКОРИСТАННЯМ ГЛИБИННОЇ ЗГОРТКОВОЇ НЕЙРОННОЇ МЕРЕЖІ

Здійснено огляд попередніх досліджень та актуальність систем з пошуку обличчя. Виявлено недоліки і труднощі, що впливають на їх ефективність. Запропоновано для реалізації систему для пошуку обличчя на зображенні на підставі алгоритмів розпізнавання обличчя. Пошук обличчя є першим кроком у різних інших застосуваннях: відстеження, аналіз, розпізнавання осіб. Окрім використання у поєднанні з технологіями, описаними вище, можна використовувати пошук обличчя, щоб: підрахувати кількість людей, які є на екрані або дивляться на цифровий дисплей; визначити, які ділянки зображення потрібно розмивати, щоб забезпечити конфіденційність. Використано глибинні згорткові нейронні мережі для реалізації, оскільки вони ефективніші для цієї задачі. Науковим результатом під час розроблення поданої системи є удосконалення алгоритмів розпізнавання обличчя на зображенні: пришвидшення, підвищення точності розпізнавання. Реалізовано першу версію запропонованої системи. Розроблена система може бути використана у різних сферах людської діяльності та в багатьох задачах аналітики. Для цього використано TensorFlow – спеціальний фреймворк для розробки, навчання власних глибинних нейронних мереж. Вирішено використати набір даних LFW – People (Face Recognition) для навчання моделі. Використано Python бібліотеку LabelImg, щоб попередньо промаркувати навчальні дані, які нейронна мережа повинна знаходити на зображенні. Зроблено карту міток для навчання моделі. Запущено 25000 ітерацій для навчання моделі. Отримано рівень втрат 0,069 (похибка у 6,9 %) на останній ітерації. Модель показує однозначний результат для більшості тестових даних ($\geq 99\%$), але рівень точності у випадках, коли на зображенні містилися головні убори, є дещо нижчим (97 %). Модель поки що не тестувалася на даних із великим кутом повороту обличчя, тому важливо перевірити її стійкість для цього фактору і за змогою знизити його вплив на точність моделі.

Ключові слова: комп'ютерний зір; алгоритми розпізнавання; аналіз зорових образів; машинне навчання; штучний інтелект.

Вступ / Introduction

Великого прогресу було досягнуто в області виявлення і розпізнавання обличчя для цілей безпеки, ідентифікації та відвідуваності, але все ще є проблеми, що перешкоджають прогресу в цій сфері. Виявлення обличчя на зображенні є більш ніж простим завданням для людського зору, однак у разі спроби побудови автоматичної системи пошуку обличчя доводиться зіткнутися з такими труднощами: велика варіація зовнішнього вигляду; зміна кута повороту обличчя тягне за собою зміну зображення; можлива присутність індивідуальних особливостей; зміна виразу обличчя; частина обличчя може бути невидима; умови зйомки; низьке розширення зображень.

Усе це потребує застосування різних алгоритмів попереднього оброблення, що ускладнює процес розпізна-

вання, робить його громіздкішим, збільшує обсяг і час обчислювальних процесів. Тому на сьогодні досі існує потреба в розробленні методів і алгоритмів, які вирішують перераховані вище проблеми.

Пошук обличчя є першим кроком у різних інших застосуваннях: відстеження, аналіз, розпізнавання осіб. Окрім цього, можна використовувати пошук обличчя, щоб: підрахувати кількість людей, які є на екрані або дивляться на цифровий дисплей; визначити, які ділянки зображення потрібно розмивати, щоб забезпечити конфіденційність.

Об'єкт дослідження – пошук обличчя на зображенні.

Предмет дослідження – алгоритми виділення та розпізнавання обличчя, що дасть змогу підрахувати кількість людей у приміщенні.

Мета роботи – розробити систему для пошуку обличчя на зображенні за допомогою алгоритмів його

Інформація про авторів:

Яловега Олеся Володимирівна, магістрант, кафедра програмного забезпечення. Email: olesia.yaloveha.mnpz.2020@lpnu.ua;
<https://orcid.org/0000-0002-4621-9214>

Мельник Роман Андрійович, д-р техн. наук, професор, кафедра програмного забезпечення. Email: ramelnik@polynet.lviv.ua;
<https://orcid.org/0000-0002-4329-6740>

Цитування за ДСТУ: Яловега О. В., Мельник Р. А. Система виявлення обличчя на зображенні з використанням глибинної згорткової нейронної мережі. Науковий вісник НЛТУ України. 2022, т. 32, № 2. С. 55–60.

Citation APA: Yaloveha, O. V., & Melnyk, R. A. (2022). Face detection system on image using deep convolutional neural network. Scientific Bulletin of UNFU, 32(2), 55–60. <https://doi.org/10.36930/40320209>

розпізнавання.

Для досягнення зазначеної мети визначено такі *основні завдання дослідження*: здійснити аналіз літературних джерел; визначити матеріал і методи дослідження; обрати алгоритм розпізнавання обличчя; описати архітектуру запропонованої системи; зібрати матеріали для навчання моделі; здійснити навчання та тестування моделі.

Аналіз останніх досліджень та публікацій. У дослідній роботі [3] подано новий метод, що використовує алгоритм LBP у поєднанні з передовими методами оброблення зображень для вирішення проблем, що перешкоджають точності розпізнавання обличчя. Пропонований метод здебільшого дає точність 99,0 %. Однак дослідження не порушує проблему оклюзії і маскування осіб під час розпізнавання. Автори статті [1] пропонують метод вирівнювання зображення обличчя з використанням як еталона створеної за допомогою FaceGen Modeller тривимірної моделі особи; автори розробляють новий дескриптор для розпізнавання обличчя на підставі фільтрів Габора. Цей дескриптор має кращу продуктивність розпізнавання, досягає точності 97,29 % на наборі даних LFW.

У статті [13] запропоновано метод глибокого навчання системи розпізнавання обличчя з використанням хмарних обчислень. Він забезпечує вищу точність (99,06 %), вищий ступінь відкликання (99,07 %) і вищу специфічність (99,10 %), ніж інші алгоритми, з якими було здійснено порівняння. У статті [2] пропонують новий мультимодальний підхід до вилучення локальних ознак на обличчі, заснований на об'єднанні переваг різноспрямованих перетворень. Основним внеском цього підходу є застосування перетворення Curvelet як швидкого і потужного методу для подання ознак обличчя, а потім оброблення коефіцієнтів Curvelet в різних частотних діапазонах з використанням двох різних підходів фрактального виміру для ефективного відображення текстури обличчя.

У статті [15] досліджують багатозадачне навчання для розпізнавання обличчя. Автори пропонують згортовку нейронну мережу, в якій класифікація обличчя є основним завданням, а оцінка пози, освітлення і виразу – побічними завданнями. Обширні експерименти з набором даних Multi-PIE показують ефективність запропонованих підходів. У статті [8] автори подали нову модель HCLM, яка об'єднує локальне і цілісне визначення орієнтирів обличчя завдяки інтеграції таких трьох методів: оцінка пози голови, визначення розріджено-цілісних орієнтирів і виявлення щільних локальних орієнтирів. Запропонований метод демонструє найсучасніші характеристики і надійний, особливо за наявності великих варіацій пози голови. Точність класифікації для наборів даних 300-Vt і AFLW становить 95,2 і 91,5 % відповідно.

Автори статті [12] показують, як можна використовувати розширення даних для конкретної предметної області для генерації цінних додаткових даних. Привабливі потенційні методи збільшення, які тут не використовуються, – це синтез зображень однієї і тієї ж особи з різним віком або маніпуляції з волоссям на голові, обличчі. У статті [10] автори запропонували новий метод розпізнавання людських облич за виразом обличчя. Функції злиття містять більше інформації, ніж зобра-

ження людських облич, тому такий спосіб можна використовувати для ідентифікації та підвищення швидкості розпізнавання. Метод, запропонований у цій статті, дає 100 % точності розпізнавання навпроти фронтального освітлення, а результат на підставі PCA становить 97 %.

Матеріал та методи дослідження. Для навчання моделі вирішено використати LFW – People (Face Recognition) [Ошибка! Источник ссылки не найден.]. Набір даних містить понад 13 000 зображень облич. Кожне обличчя позначено ім'ям людини, яка зображена. 1680 зображених людей мають дві або більше різних фотографій. Кожне зображення зосереджено на одному обличчі. Кожен піксель кожного каналу (колір у RGB) кодується float в діапазоні 0,0-1,0. Оригінальні зображення мають розмір 250×250 пікселів.

Обрані навчальні дані потрібно попередньо промаркувати, щоб виділити об'єкти, які нейронна мережа повинна знаходити на зображенні. Для створення анотацій використаємо Python бібліотеку Labellmg. Анотації зберігаються як файли XML у форматі PASCAL VOC. На рис. 1 подано приклад використання бібліотеки Labellmg:



Рис. 1. Приклад використання бібліотеки Labellmg / Example of using the Labellmg library

Після цього у папці зі зображеннями має бути створений набір нових *.xml файлів, по одному для кожного зображення. Після поділу всі навчальні зображення разом з відповідними *.xml файлами потрібно помістити в окремі папки train – для навчання та test – для тестування.

Результати дослідження / Research results and their discussion

Запропонована система. Технології для реалізації: мова програмування Python, середовище розробки PyCharm, бібліотека для навчання нейронних мереж Tensorflow. Апаратні вимоги: процесор Intel® Core™ i3-4000M 2.4 ГГц, ОЗП 2 Gb і вище, 100 мб вільного місця на жорсткому диску. Програмні вимоги: потрібно встановити python версії 3.8 або вище; операційна система Windows 10 або Linux Ubuntu 20.04. Для реалізації вирішено використати глибинні нейронні мережі, оскільки вони ефективніші для цієї задачі. Діаграму прецедентів для системи пошуку облич з використанням нейронної мережі зображено на рис. 2.

Алгоритм розпізнавання обличчя. За виявлення об'єктного класу (обличчя) завдання полягає в тому, щоб знайти розташування та розміри всіх об'єктів на зображенні, які належать до цього класу. Нейронна мережа – це алгоритм, призначений для імітації людсько-

го мозку для виконання різноманітних задач. Глибинне навчання – це галузь машинного навчання на підставі нейронних мереж. Воно може виявити особливості, необхідні для класифікації автоматично в процесі навчання без етапів вилучення ознак [10].

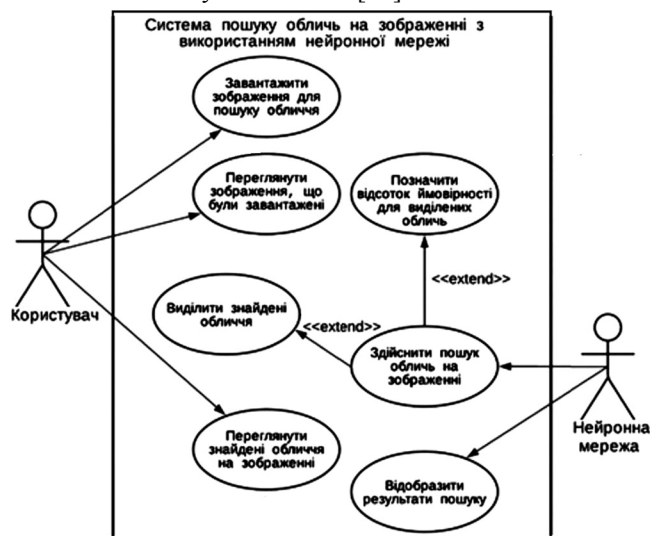


Рис. 2. Діаграма прецедентів / Use case diagram

Для пошуку обличчя на зображенні реалізовано глибинну згорткову мережу. Згортковий шар застосовує фільтр згортки до зображення, щоб виявити його особливості: згортка бере набір ваг і множить їх на вхідні дані з нейронної мережі; під час процесу множення ядро (2D-масиви ваг) або фільтр (3D-структури) проходить над зображенням кілька разів. Фільтр застосовується справа наліво і зверху вниз. Кожен фільтр множить ваги з різними вхідними значеннями. Загальні вхідні дані підсумовуються, забезпечуючи унікальне значення для кожної позиції фільтра.

У запропонованій нейронній мережі є три згорткових шари. Перший – це згортка 1×1 , вона розширює кількість каналів даних; далі йде шар згортки 3×3 по глибині, він фільтрує вхідні дані; третій шар це згортка 1×1 , він поєднує відфільтровані значення для створення нових функцій. Схема описаної нейронної мережі представлено на рис. 3.

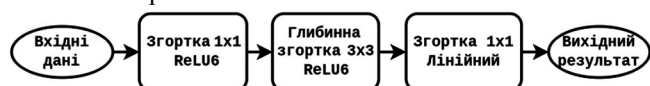


Рис. 3. Архітектура запропонованої нейронної мережі / The architecture of the proposed neural network

Функцією активації, що застосовується для першого та другого шарів, є ReLU6. Формула функції активації ReLU6 така:

$$y = \min(\max(0, x), 6). \quad (1)$$

Графік функції активації ReLU6 подано на рис. 4.

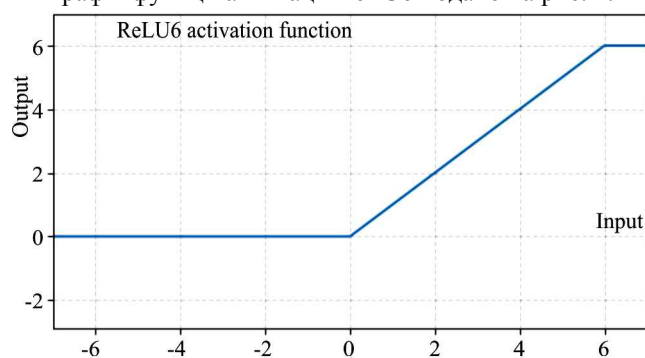


Рис. 4. Функція активації ReLU6 / ReLU6 activation function

Навчання моделі. TensorFlow потребує карти міток, яка відображає кожну з використаних міток та ціле цифрове значення. Ця карта міток використовується як у процесі навчання, так і в процесі виявлення. Після цього потрібно перетворити наші анотації у так званий TFRecord формат. Для цього написано простий сценарій, який перебирає всі *.xml файли і генерує *.record файл для кожного з двох. TensorFlow використовує файли protobuf для налаштування процесу навчання та оцінки. Каркас файлу конфігурації для навчання моделі такий:

```
model { (...) }
train_config { (...) }
train_input_reader { (...) }
eval_config { (...) }
eval_input_reader { (...) }.
```

Для навчання моделі було запущено 25000 ітерацій, що зайняло у часовому проміжку 5 год 59 хв. Нижче подано п'ять основних показників навчання: втрати класифікації, втрати локалізації, втрати регуляризації, загальні втрати, темп навчання.

Показники втрат направляють модель до правильної траєкторії навчання. Якщо на якомусь із кроків навчання модель помилилася, вона отримує більший рівень втрат. На кожному кроці загальна втрата є сумою втрат класифікації, локалізації та регуляризації. Мінімізація загальних втрат вимагає мінімізації всіх трьох цих втрат.

На TensorBoard, спеціальному інструменті для відстеження процесу навчання, що надається разом із TensorFlow, представлено графіки усіх втрат на кожному 100 кроці із 25000 для навчання запропонованої моделі нейронної мережі (рис. 6).

Як видно із рисунків, труднощі у навчанні моделі виникали в основному на рівні втрат класифікації. На графіку на рис. 5, а помітні стрибки зростання класифікаційних втрат приблизно кожні 500 циклів. На такий результат міг вплинути той фактор, що на багатьох зображеннях із обраних навчальних даних були присутні одразу декілька об'єктів класифікації, через що модель не завжди розпізнавала у процесі навчання усі об'єкти із необхідних. Для втрат же локалізації і регуляризації не помітно значних стрибків значень. Це означає, що робота із анотування та маркування обраних навчальних даних була зроблена правильно. Внаслідок цього на 25000 кроці отримано такий рівень втрат (рис. 5).

```
Terminal: Local x + v
INFO:tensorflow:Step 25000 per-step time 0.805s
I0319 21:31:39.349698 140713895163712 model_lib_v2.py:705]
Step 25000 per-step time 0.805s
INFO:tensorflow:{'Loss/classification_loss': 0.0065687275,
'Loss/localization_loss': 0.0031937466,
'Loss/regularization_loss': 0.059165865,
'Loss/total_loss': 0.06892834,
'learning_rate': 0.04128206}
I0319 21:31:39.349874 140713895163712 model_lib_v2.py:708]
{'Loss/classification_loss': 0.0065687275,
'Loss/localization_loss': 0.0031937466,
'Loss/regularization_loss': 0.059165865,
'Loss/total_loss': 0.06892834,
'learning_rate': 0.04128206}
```

Рис. 5. Результат навчання моделі / The result of model learning

Рівень втрат, що дорівнює 1, рівносильний похибці в 100 %. Доцільно досягти загального рівня втрат принаймні 0,2 (в ідеалі 0,1 і нижче). Як видно із зображен-

ня, загальний рівень втрат становить 0,069 (тобто похибка моделі становить 6,9 %), чого цілком достатньо, але цей результат можна покращити, якщо запустити нові ітерації навчання або збільшити кількість навчальних даних.

П'ятий же параметр, темп навчання, відповідає за ефективність навчання на кожному кроці. Він визначає те, якою мірою нова отримана інформація замінює попередню інформацію, тобто образно представляє швид-

кість, з якою модель машинного навчання навчається. На рис. 7 подано графік темпу навчання розробленої моделі, взятий із інструменту для відстеження навчання моделі TensorBoard.

Як помітно із графіка, на першій 1000 кроків спостерігається різке зростання темпу навчання, однак приблизно з 4000 кроку модель поступово сповільнює свій темп навчання. Для того, щоб знову пришвидшити темп, варто додати для моделі нові дані для навчання.

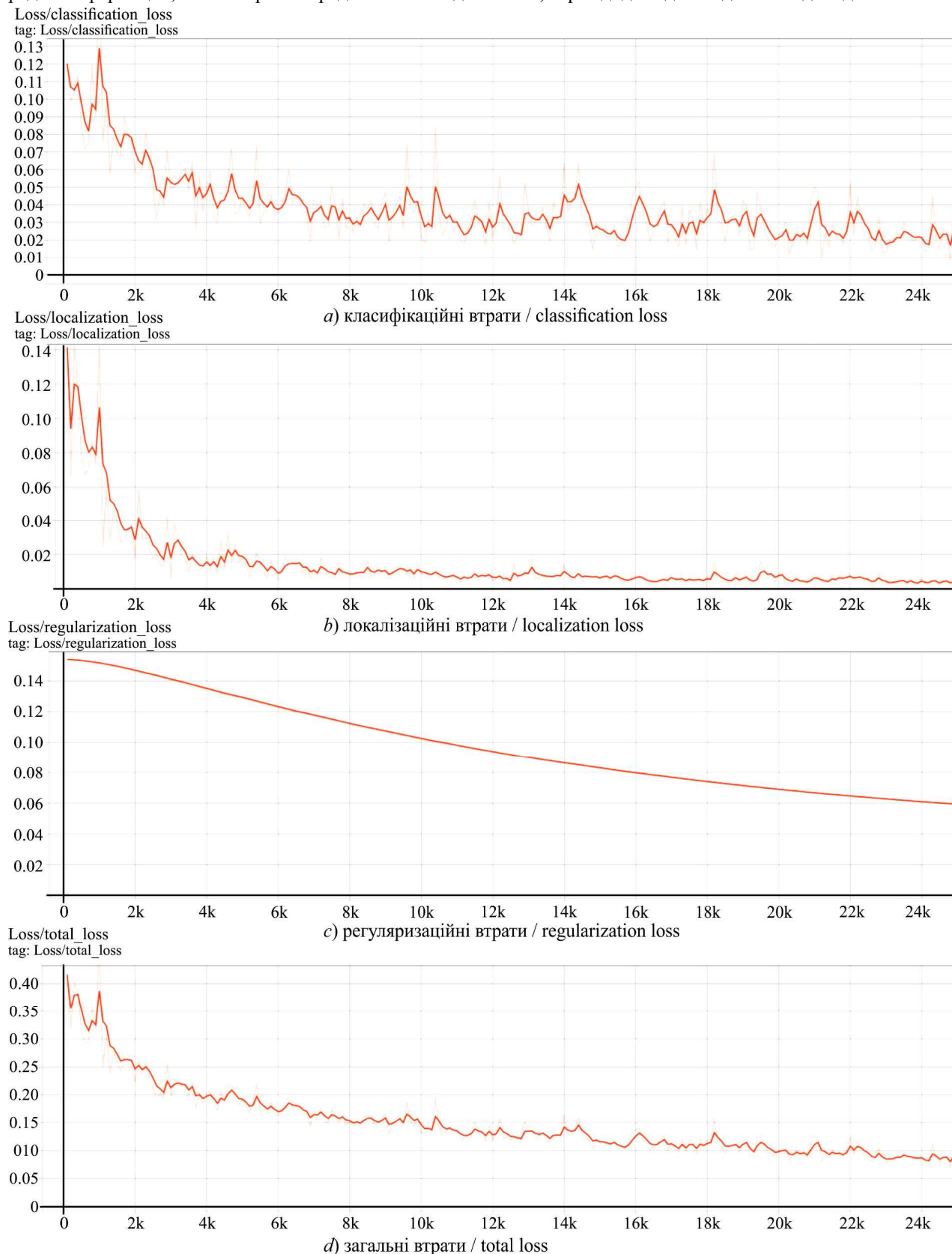


Рис. 6. Рівень втрат для нейронної мережі / Loss level for the neural network [4]

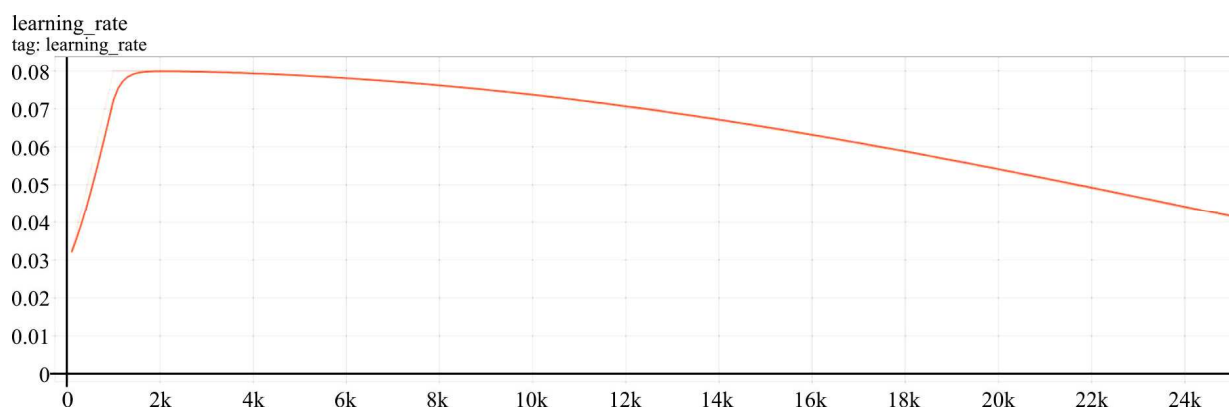


Рис. 7. Темп навчання моделі нейронної мережі / The learning rate of the neural network model [4]

Тестування моделі. На рис. 8. показано приклад роботи моделі на нових даних.

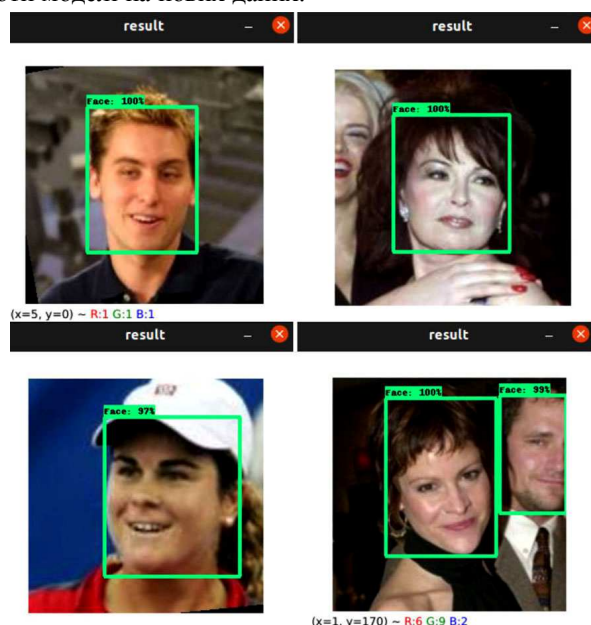


Рис. 8. Результат роботи першої версії моделі / The result of the first version of the model

Обговорення результатів дослідження. Отримані в цій роботі значення точності виявлення обличчя на зображення збігаються із результатами досліджень в інших роботах [1, 3, 10, 13] а також перевищують результати роботи [8].

У трьох і п'яти випадків модель видає однозначний результат – 100 %, для випадку, де велика частина обличчя прихована, результат майже однозначний – 99 %, проте для одного зображення результат нижчий – 97 % (див. рис. 8). Причиною зниження точності відповіді нейронної мережі є наявність головного убору на зображенні. Загалом можна зробити висновок, що модель працює добре, але наявність предметів, що частково перекривають обличчя, знижує точність роботи. Аналогічними факторами, що впливатимуть на рівень точності, є наявність бороди, окулярів, тому потрібне подальше навчання моделі. Саме це буде наступним напрямом дослідження. Модель поки що не тестувалася на даних із великим кутом повороту обличчя, важливо перевірити її стійкість для цього фактору і за змогою знизити його вплив на точність моделі.

Отже, внаслідок проведеного дослідження отримано такі *основні результати*: розроблено систему для по-

шуку обличчя на зображенні на підставі алгоритмів його розпізнавання.

Наукова новизна отриманих результатів дослідження – удосконалення алгоритмів розпізнавання обличчя на зображенні: пришвидшення, підвищення точності розпізнавання.

Практична значущість результатів дослідження – систему можна використовувати у різних сферах людської діяльності та в багатьох задачах аналітики, наприклад для того, щоб підрахувати кількість осіб на зображенні.

Висновки / Conclusions

Розроблено систему для пошуку обличчя на зображенні на підставі алгоритмів розпізнавання обличчя.

Здійснено аналіз літературних джерел, визначено актуальність теми, виявлено недоліки і труднощі, що впливають на ефективність роботи систем з пошуку обличчя. Для реалізації вирішено використати глибокі згорткові нейронні мережі, оскільки вони ефективніші для цієї задачі.

Визначено матеріали і методи дослідження. Обрано набір даних LFW – People (Face Recognition) для навчання моделі нейронної мережі. Здійснено попередню підготовку навчальних даних – їх маркування за допомогою бібліотеки LabelImg.

Описано архітектуру глибокої згорткової нейронної мережі для запропонованої системи. Для розроблення моделі нейронної мережі використано бібліотеку глибокого машинного навчання Tensorflow. Запущено 25000 ітерацій навчання, на останньому кроці отримано рівень втрат 0,069. Найбільший вплив на рівень втрат мали класифікаційні втрати, помітно також поступове зниження темпу навчання моделі.

Здійснено тестування моделі на нових даних. Отримано однозначний результат для більшості тестових даних ($\geq 99\%$). Помітно зниження рівня точності для зображення, на якому є головний убір (97 %).

References

1. Aksasse, B., Ouanan, H., & Ouanan, M. (2017). Novel approach to pose invariant face recognition. *Procedia Computer Science*, 110, 434–439. <https://doi.org/10.1016/j.procs.2017.06.108>
2. Al-Waisy, A. S., Qahwaji, R., Ipson, S., & Al-Fahdawi, S. (2017). A multimodal deep learning framework using local feature representations for face recognition. *Machine Vision and Applications*, 29(1), 35–54. <https://doi.org/10.1007/s00138-017-0870-2>
3. Bah, S. M., & Ming, F. (2020). An improved face recognition algorithm and its application in attendance management system. *Ar-ray*, 5. <https://doi.org/10.1016/j.array.2019.100014>

4. Grafiki vtrat ta tempu navchannia rozroblenoi modeli glibinnoi zgorikovoi neironnoi merezhi. Retrieved from: https://www.drive.google.com/drive/folders/13Klb_rBFC-tckJ7S4LBrHQp-3ANB9Rhu?usp=sharing. [In Ukrainian].
5. Grytsiuk, M. Yu., & Hrytsiuk, Yu. I. (2018). Nature and Sustainable Development of Tourism in the Carpathian Region Ukraine. *Scientific Bulletin of UNFU*, 28(2), 99–110. <https://doi.org/10.15421/40280219>
6. Havrysh, V. I., & Hrytsiuk, Yu. I. (2021). Analysis of temperature regimes in heat-sensitive layered elements of digital devices caused by internal heating. *Scientific Bulletin of UNFU*, 31(5), 108–112. <https://doi.org/10.36930/40310518>
7. Hrytsiuk, Yu., Grytsiuk, P., Dyak, T., & Hrynyk, H. (2019). Software Development Risk Modeling. *IEEE 2019 14th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT 2019)*, (Vol. 2, pp. 134–137), 17–20 September, Lviv, Ukraine. Lviv: Lviv Polytechnic National University, 206 p. <https://doi.org/10.1109/stc-csit.2019.8929778>
8. Kim, K., Baltrušaitis, T., Zadeh, A., Morency, L.-P., & Medioni, G. (2016). Holistically Constrained Local Model: Going Beyond Frontal Poses for Facial Landmark Detection. *Proceedings of the British Machine Vision Conference 2016*. <https://doi.org/10.5244/c.30.95>
9. LFW – People (Face Recognition). Retrieved from: <https://www.kaggle.com/atulanandjha/lfwpeople>
10. Li, L., Mu, X., Li, S., & Peng, H. (2020). A Review of Face Recognition Technology. *IEEE Access*, 8, 139110–139120. <https://doi.org/10.1109/access.2020.3011028>
11. Liu, J., Fang, C., & Wu, C. (2016). A Fusion Face Recognition Approach Based on 7-Layer Deep Learning Neural Network. *Journal of Electrical and Computer Engineering*, 1–7. <https://doi.org/10.1155/2016/8637260>
12. Masi, I., Tran, A. T., Leksut, J. T., Hassner, T., Medioni, G. (2016). Do we really need to collect millions of faces for effective face recognition? *Proceedings of the European conference on computer vision 2016*.
13. Salama Abdelminaam, D., Almansori, A. M., Taha, M., & Badr, E. (2020). A deep facial recognition system using computational intelligent algorithms. *PLOS ONE*, 15(12). <https://doi.org/10.1371/journal.pone.0242269>
14. TensorBoard. Retrieved from: <https://www.tensorflow.org/tensorboard>
15. Yin, X., & Liu, X. (2018). Multi-Task Convolutional Neural Network for Pose-Invariant Face Recognition. *IEEE Transactions on Image Processing*, 27(2), 964–975. <https://doi.org/10.1109/tip.2017.2765830>

O. V. Yaloveha, R. A. Melnyk

Lviv Polytechnic National University, Lviv, Ukraine

FACE DETECTION SYSTEM ON IMAGE USING DEEP CONVOLUTIONAL NEURAL NETWORK

A review of previous research and the relevance of facial search systems has been made. Some disadvantages and difficulties affecting their effectiveness have been identified. A system searching for a face in an image based on face recognition algorithms is proposed for implementation. Face search is the first step in various other applications such as tracking, analysis, and face recognition. In addition to using in combination with the technologies described above, it is possible to use face search to count the number of people on the screen or those looking at the digital display, and also determine which areas of the image need to be blurred to ensure privacy. Deep convolutional neural networks were used for implementation, as they are more efficient for this task. The scientific result in the development of the presented system is the improvement of algorithms for facial recognition in the image like acceleration, increase in the accuracy of recognition. The first version of the proposed system has been implemented. The developed system can be used in various fields of human activity and many analytical tasks. Therefore, Tensorflow is used, i.e. a special framework for the development, and training of own deep neural networks. We decided to use the data set LFW – People (Face Recognition) to train the model. The Labelling Python library is used to pre-mark the training data that the neural network should find in the image. A map of labels for model learning has been made. 25,000 iterations were launched to teach the model. The level of losses of 0.069 (error of 6.9 %) on the last iteration was obtained. The model shows an unambiguous result for most of the test data ($\geq 99\%$), but the level of accuracy in cases where the image contained hats is slightly lower (97 %). The model has not yet been tested on data with a large angle of rotation of the face, so it is important to test its stability for this factor and if possible reduce its impact on the accuracy of the model.

Keywords: computer vision; recognition algorithms; analysis of visual images; machine learning; artificial intelligence.