

ВИЗНАЧЕННЯ ВІКУ ЛЮДИНИ ЗА ФОТО НА ОСНОВІ НЕЙРОННИХ МЕРЕЖ

Свєгеній Вербенко

ORCID: <https://orcid.org/0009-0001-8438-4990>

Дніпропетровський університет імені Олеся Гончара, Дніпро, Україна

Ольга Мацуґа

ORCID: <https://orcid.org/0000-0001-6444-8566>

Дніпропетровський університет імені Олеся Гончара, Дніпро, Україна

Вступ

У сучасному світі технології штучного інтелекту та машинного навчання знаходять широке застосування в різних галузях. Однією з цікавих і корисних задач, яка активно досліджується та впроваджується в практику, є визначення віку людини на основі аналізу зображень її обличчя. Складність цієї задачі пов'язана з тим, що процес старіння залежить від багатьох факторів, таких як генетика, спосіб життя та навколишнє середовище. З розвитком комп'ютерного зору та глибокого навчання стало можливим досягати високої точності у розв'язанні цієї задачі, що відкриває нові можливості для її застосування у різних сферах.

ЦІЛЬ ТА ЗАДАЧІ

Ціллю даної роботи було порівняти різні архітектури нейронних мереж в задачі визначення віку людини за зображеннями її обличчя. Оскільки вік є неперервною величиною, то задача визначення віку людини зводиться до задачі регресії. Її розв'язання за допомогою нейронних мереж передбачало виконання таких етапів:

- 1) обрання достатньо великого датасету, який би містив обличчя людей різних вікових категорій та рас;
- 2) вибір архітектур нейронних мереж;
- 3) підбір гіперпараметрів та тренування нейронних мереж;
- 4) порівняння якості моделей.

МАТЕРІАЛИ ТА МЕТОДИ

Для розв'язання поставленої задачі було обрано датасет UTKFaces [1]. Він містить близько 24 000 розмічених зображень. Зображення розмічені по статі, расі та віку. На рисунку 1 наведено розподіл зображень за статтю, расою та віком: кругові діаграми

показують розподіл за статтю та расою, а гістограма ілюструє розподіл за віком. Кількість зображень чоловіків і жінок в наборі даних майже однакова. За расою датасет незбалансований, більша частина зображень людей раси White. Розподіл за віком є багатомодальним. Можна відмітити наявність чотирьох мод на рівні 0-4, 24-28, 56-60 та 72-76 років.

З цих характеристик можна зробити висновок, що нейронній мережі буде складніше визначити вік людей, які старші 80 років, через те, що зображень людей такої вікової групи не дуже багато.

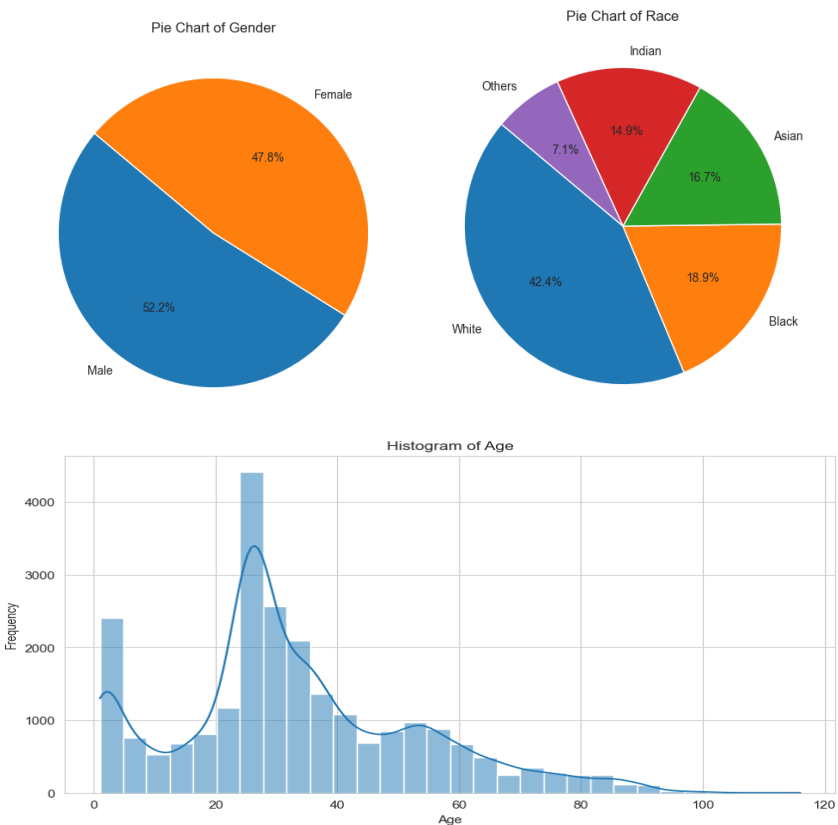


Рисунок 1 – Статистичні характеристики датасету UTKFaces

Було відібрано 4 архітектури для порівняння: AlexNet, VGG-19, ResNet-50, Inception-v4. Це перші архітектури згорткових нейронних мереж, які показали хороші результати в класифікації зображень на ImageNet (великому датасеті зображень).

AlexNet [2] (Рис. 2) – це одна з найперших архітектур, що вплинула на розвиток глибокого навчання. Вона містить вісім шарів, включаючи п'ять згорткових і три повнозв'язаних шари, і використовує ReLU в якості функції активації, дроп-аут та макс-пулінг для покращення продуктивності класифікації зображень.

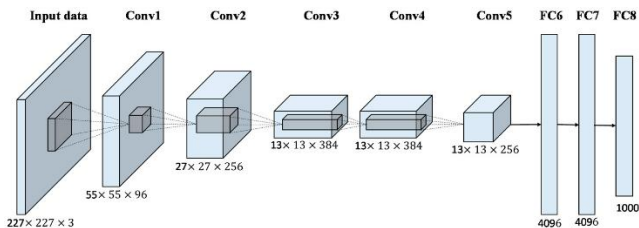


Рисунок 2 – Архітектура AlexNet

VGG-19 [3] (Рис. 3) – це архітектура згорткової нейронної мережі, яка досягла високих результатів у класифікації зображень на конкурсі ImageNet 2014 завдяки своїй простій і глибокій архітектурі. У порівнянні з AlexNet, VGG-19 значно глибша, має 19 шарів. Особливістю VGG-19 є використання малих (3x3) фільтрів, що дозволяє краще виділяти складні ознаки зображень.

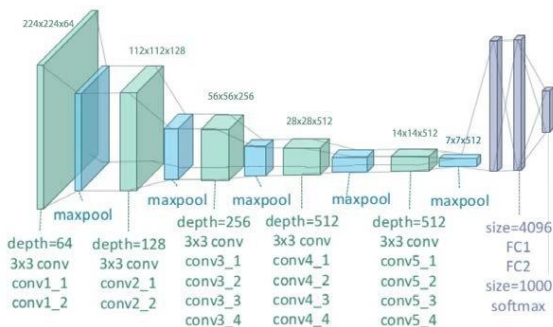


Рисунок 3 – Архітектура VGG-19

ResNet-50 [4] (Рис. 4) використовує глибоку архітектуру з 50 шарами. Ця модель може навчатися на великих наборах даних та досягати непоганих результатів. Ключовим нововведенням цієї архітектури є використання резидуальних зв'язків (residual connections), які дозволяють навчати глибші моделі без затухання градієнта та сильного перенавчання на менших об'ємах даних.

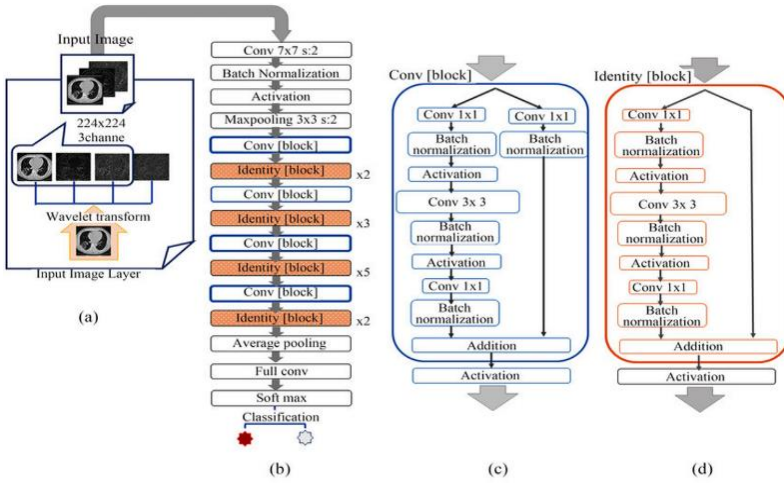


Рисунок 4 – Архітектура ResNet-50

Inception-v4 [5] (Рис. 5) була представлена дослідниками з Google у 2016 році і являє собою удосконалену версію моделі Inception-v1 (також відомої як GoogLeNet). Inception-v4 відрізняється від першої версії більш глибокою архітектурою, оптимізованими інцепційними блоками для підвищення ефективності обчислень, наявністю резидуальних зв'язків для боротьби з проблемою затухання градієнта, а також покращеними нормалізаційними шарами.

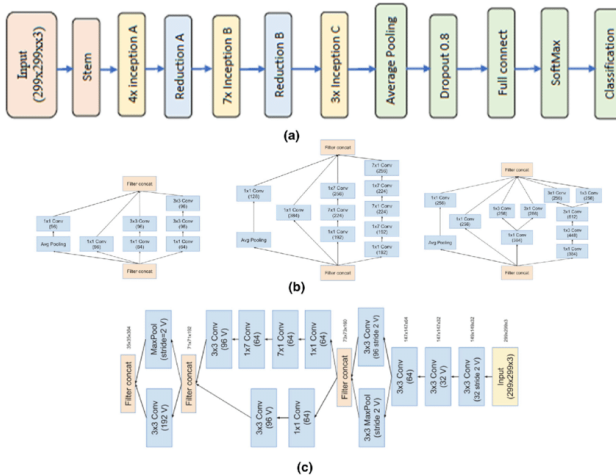


Рисунок 5 – Архітектура Inception-v4

РЕЗУЛЬТАТИ

У ході виконання роботи було навчено 4 зазначені вище нейронні мережі на наборі даних UTKFaces. У всіх мереж останній шар був замінений на повнозв'язний шар з одним нейроном і лінійною функцією активації. Під час навчання використовувалася в якості функції втрат величина MSE (середня квадратична помилка), а як метрика якості – MAE (середня абсолютна помилка).

Дані перед навчанням були розділені на тренувальну та тестову вибірки у пропорції 90% та 10%. Перед поданням на навчання зображення попередньо нормалізувалися (це важливий крок, щоб пікселі варіювалися в діапазоні від 0 до 1, тому що нейронна мережа погано вчиться на даних, які мають різний масштаб) та зменшувалися до розмірів, які вимагає кожна нейронна мережа.

Для нейронних мереж AlexNet та VGG-19 був використаний метод оптимізації SGD зі швидкістю навчання 0.2, для ResNet-50 – Adam з початковою швидкістю навчання 0.02, для Inception-v4 – Adadelta з початковою швидкістю навчання 0.2. Вказані методи та їх параметр були обрані як найкращі в результаті обчислювальних експериментів.

Кожна мережа навчалась різну кількість епох, яка була потрібна для збіжності ітераційного процесу.

Моделі тренувались на системі з наступними характеристиками: відеокарта RTX 3060 TI, процесор Intel i5-12400F, оперативна пам'ять 48ГБ RAM.

Порівняльний аналіз моделей наведений у таблиці 1. В таблиці вказані значення MAE та MSE, розраховані за тестовою вибіркою на епосі, яка передувала тій, на якій було виявлено перенавчання.

Таблиця 1 – Порівняльний аналіз моделей

Характеристика	VGG-19	ResNet-50	Inception-v4	AlexNet
Приблизна кількість параметрів	114 млн	21 млн	30 млн	58 млн
Середній час на одну епоху	390 секунд	140 секунд	340 секунд	96 секунд
Швидкість навчання	0.2	0.02	0.2	0.2
Метод оптимізації	SGD	Adam	Adadelta	SGD
MAE	2,7	3,5	3,87	12,8
MSE	14,5	16,4	101,6	331,3
Час на обробку одного зображення	1.4 секунди	0.44 секунди	0.97 секунд	0.7 секунд

Графіки, які ілюструють зміну MAE протягом навчання, наведені на рисунку 6. Найкращі результати показали моделі VGG-19, ResNet-

50 та Inception-v4. В них найменше перенавчання та найкращий показник якості. Модель AlexNet має високий рівень перенавчання.

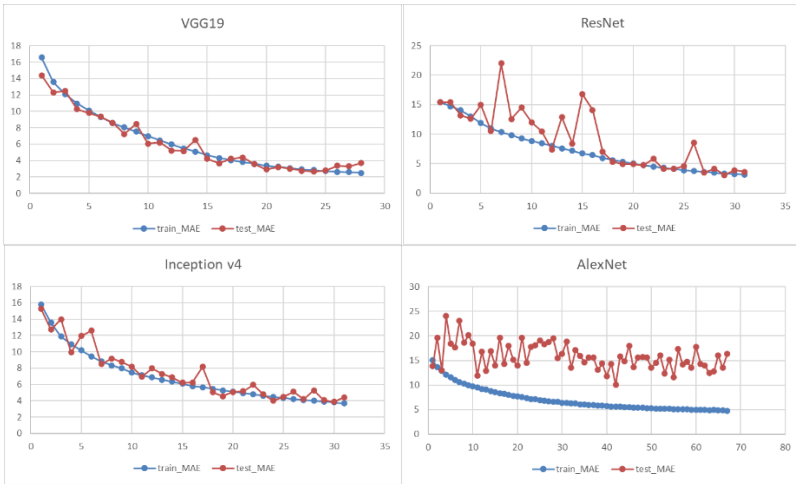


Рисунок 6 – Графіки зміни метрики MAE в ході навчання моделей

На рисунку 7 в якості прикладу наведені результати визначення віку людини за фото на основі навченої моделі VGG-19.

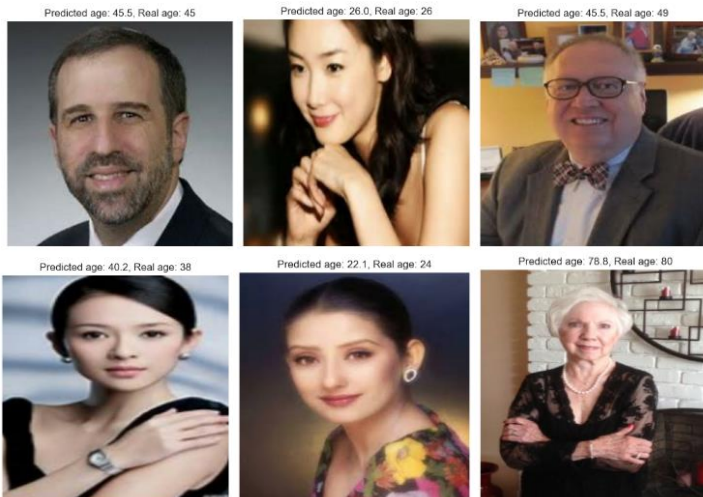


Рисунок 7 – Результати роботи навченої моделі VGG-19

ВИСНОВКИ

В процесі роботи було навчено 4 нейронні мережі (AlexNet, VGG-19, ResNet-50 та Inception-v4) для передбачення віку людини за фото. Для моделей VGG-19 і ResNet-50 вдалося досягти значення метрики MAE у 2,7 та 3,5 відповідно. Для Inception-v4 величина MAE становила 3,87. При цьому модель ResNet-50 найшвидше обробляє зображення. Результати навчання AlexNet виявилися незадовільними, модель була сильно перенавченою.

Якщо є потреба визначати вік людини за фото в режимі реального часу, то згідно з одержаними результатами, можна рекомендувати для цього модель ResNet-50, оскільки вона обробляє зображення втричі швидше, ніж VGG-19. Якщо необхідна висока точність визначення віку, то краще використати VGG-19. Загалом ці дві моделі показали непогані результати для розв'язання поставленої задачі.

ПОСИЛАННЯ

1. *UTKFace*. Kaggle. <https://www.kaggle.com/datasets/jangedoo/utk-face-new/data>
2. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, 60(6), 84–90. <https://doi.org/10.1145/3065386>
3. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations (ICLR 2015)*, 1–14. <https://doi.org/10.48550/arXiv.1409.1556>
4. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
5. Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, 31(1), 4278–4284. <https://doi.org/10.1609/aaai.v31i1.11231>