

Особенности использования свёрточных нейронных сетей для распознавания возраста человека по изображению лица

В.К. Колодин
Самарский национальный исследовательский университет им. академика С.П. Королева
Самара, Россия
vkkolodin@gmail.com

Д.А. Савельев
Самарский национальный исследовательский университет им. академика С.П. Королева
Самара, Россия
dmitrey.savelyev@yandex.ru

Аннотация—В работе исследуются влияние различных факторов на результат обучения свёрточных нейронных сетей при решении задачи определения возраста человека. Рассматриваются разные подходы к решению задачи, в том числе различные конфигурации нейронных сетей и различные интерпретации исходной задачи. Было показано, что при использовании функции активации Swish и распределения возрастов $[-1, 1]$, а также архитектуры сети EfficientNetV2_B1 возможно уменьшить ошибку распознавания на 43 %.

Ключевые слова—анализ цифровых изображений, машинное обучение, свёрточные нейронные сети, распознавание возраста.

1. ВВЕДЕНИЕ

В настоящее время решение задачи определения возраста человека, в том числе автоматическое, может быть использовано для ряда приложений, среди которых можно отметить ограничение доступа к Интернет-ресурсам вследствие возрастного порога, для улучшения работы рекомендательных систем, как уточняющий фактор при использовании других биометрических данных, в том числе в системах безопасности [1]. Следует отметить, что задача определения возраста человека по изображению его лица не является тривиальной, поскольку прямой корреляции между точным возрастом и лицом нет. Человек при решении таких задач часто допускает достаточно большие ошибки [2].

Такая важная антропоморфная характеристика как возраст может быть получена рядом методов с различной степенью точности, в том числе с помощью анализа голоса [3] и лица [4-5]. Оценка возраста с помощью анализа лица предполагает поверхностную биометрию, также известно использование методов машинного обучения и нейронных сетей [5].

В данной работе рассматривается решение задачи распознавания возраста человека с помощью свёрточных нейронных сетей. Исследуется влияние различных факторов на результат оценки возраста, таких как предобработка данных, функция активации и использования различных архитектур.

2. ОПРЕДЕЛЕНИЕ ВОЗРАСТА ЧЕЛОВЕКА С ПОМОЩЬЮ СВЁРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ

В данной работе исследования проводились на собственном наборе данных, собранном из различных, предварительно отфильтрованных, наборов данных: IMDB-WIKI [6], UTKFACE [7], APPA-REAL [8]. Всего в наборе данных присутствует 80000 изображений, представленных 80 различными возрастными (от 1 до 80 лет), в равной пропорции.

В контексте решения задачи регрессии (определение возраста как вещественного числа) проводилось исследование с помощью архитектуры MobileNetV2 [9], с классифицирующей частью из трех полносвязных слоев со следующим количеством нейронов: 1024, 512, 1, и функциями активации Swish, кроме последнего слоя. В качестве функции ошибки была взята MAE (Mean Absolute Error – среднее арифметическое абсолютной величины ошибки).

Исследовалось влияние предобработки данных на результат оценки возраста, в качестве экспериментальных данных взяты 4 интерпретации распределения возрастов, а именно в отрезках $[1; 80]$, $[-39,5; 39,5]$, $[0; 1]$, $[-1; 1]$.

Исследования показали, что наименьшая средняя ошибка среди рассмотренных у сетей с распределением возрастов в $[-1; 1]$, которая составила 10,4 года. Тем не менее, следует отметить, что для распределения $[0; 1]$ ошибка у лучшей сети больше данного результата на 0,3 года. Для распределения $[0; 1]$ особенностью является то, что большие ошибки сконцентрированы для возрастов меньше 40 лет, а для распределения $[-1; 1]$ высокие ошибки сконцентрированы для возраста больше 60, что делает сеть с последним распределением возрастов более полезной для прикладных задач. График распределения ошибки для распределения $[-1; 1]$ представлен на Рис. 1.

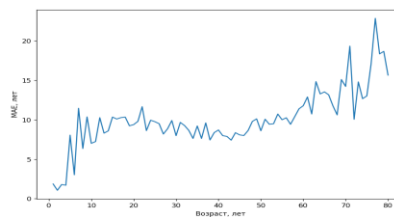


Рис. 1. Распределение погрешности определения возраста по возрастам для сети с распределением возрастов $[-1, 1]$, представленная в распределении возрастов от 1 до 80 лет

В случае с распределениями $[1; 80]$, $[-39,5; 39,5]$ наблюдалась большая ошибка, или вовсе отсутствие сходимости к решению.

Для выявления лучшей функции активации для данной задачи были рассмотрены 4 функции активации: Tanh, Sigmoid, Leaky ReLU, Swish для полносвязных слоев. В качестве основы сети взята сеть с такой же архитектурой, что и рассматривалась ранее, а распределение возрастов лежало в отрезке $[-1; 1]$. Получены следующие ошибки на тестовой выборке в годах: 20,15; 20,82; 12,36; 10,41; для функций активации Tanh, Sigmoid, Leaky ReLU, Swish соответственно. Таким образом, лучший результат достигается при использовании функции Swish, которая является гладкой, в отличие от Leaky ReLU, а также не

ограниченной сверху, что позволяет обойти проблему исчезающего градиента.

С учетом проведенных исследований далее была рассмотрена более сложная архитектура EfficientNetV2_B1 [10], с помощью которой удалось добиться уменьшения ошибки распознавания на 43 %, относительно худшей сети, то есть ошибка составляет 8,86 лет.

Также была проведена серия исследований по решению рассматриваемой задачи как задачи классификации. Существует два подхода к классификации возраста – определение точного возраста и определение возрастной группы. Классификация точного возраста подразумевает разбиение на 80 классов, в то время как классификация возраста по группам – разбиение на возрастные группы, например, (0,6); (7,12) и так далее.

Для варианта с точным определением были обучены сети, аналогичные сетям для регрессии, но в качестве последнего слоя взят полносвязный слой из 80 нейронов с функцией активации Sparse Categorical Crossentropy, а в качестве функции потерь – Softmax. Средняя ошибка в пересчете на года – 11,18 лет.

Для варианта с определением принадлежности к группе возрастов были обучены нейронные сети с архитектурами Xception [11] и MobilenetV2 с аналогичными варианту с точным определением возраста полносвязными слоями. Более того, для некоторых сетей добавлялись полносвязные слои с размером в 1024 или 512 нейронов. Обучение и тестирование проводилось на различных разбиениях возрастов на 8 групп.

Было установлено, что разбиение на группы оказывает ключевое влияние на результат. При смене количества полносвязных слоев или смене архитектуры с MobilenetV2 на Xception (то есть увеличения количества обучаемых параметров более чем в 7 раз), точность в пределах одного разбиения у разных сетей не менялась больше, чем на 5 %. Таким образом, наихудший результат для различных разбиений – 44,3 % точности, наилучший – 65,5 %.

Наилучший результат достигается при использовании следующего разбиения: (0,6); (6,14); (14,22); (22,35); (35,52); (52,60); (60,72); (72, 80), и при использовании архитектуры Xception с тремя полносвязными слоями. Несмотря на то, что результат не слишком высок, ошибки сети сконцентрированы вблизи реального класса, что видно из диаграммы ошибок классов на Рис. 2. Большинство сетей для групповой классификации возрастов не имеют связи между классами [6], в то время как рассматриваемая сеть ей обладает, что сводит практически все ошибки к неточности, но не к принципиально неверному результату.



Рис. 2. Диаграмма ошибок классов для сети, обученной на данных с разбиением (0,6); (6,14); (14,22); (22,35); (35,52); (52,60); (60,72); (72, 80), архитектурой Xception и тремя полносвязными слоями

3. ЗАКЛЮЧЕНИЕ

В данной работе проведены исследования влияния функции активации, архитектуры сети, распределение возрастов, а также разбиение их на группы на результат распознавания возраста по изображению. Установлено, что разбиение возрастов больше остальных рассмотренных факторов влияет на результат классификации. Подбор лучших параметров позволил добиться улучшения результатов, при обучении удалось уменьшить ошибку распознавания на 43 % при использовании архитектуры EfficientNetV2_B1, функции активации Swish и распределения возрастов в [-1; 1].

ЛИТЕРАТУРА

- [1] Alkaabi, S. Deep Convolutional Neural Networks for Forensic Age Estimation: A Review / S. Alkaabi, S. Yussof, H. Al-Khateeb, G. Ahmadi-Assalemi, G. Epiphaniou // Cyber Defence in the Age of AI, Smart Societies and Augmented Humanity. – 2020. – P. 375-395.
- [2] Han, H. Age estimation from face images: Human vs. machine performance / H. Han, C. Otto, A.K. Jain // 2013 international conference on biometrics (ICB). – 2013. – P. 1-8.
- [3] Kwasny, D. Gender and age estimation methods based on speech using deep neural networks / D. Kwasny, D. Hemmerling // Sensors. – 2021. – Vol. 21(14). – P. 4785.
- [4] Benkaddour, M.K. CNN based features extraction for age estimation and gender classification / M.K. Benkaddour // Informatica. – 2021. – Vol.45(5). – P. 697-703.
- [5] Castellano, G. Real-time age estimation from facial images using yolo and efficientnet / G. Castellano, B. De Carolis, N. Marvulli, M. Sciancalero, G. Vessio // International Conference on Computer Analysis of Images and Patterns. – 2021. – P. 275-284.
- [6] Rothe, R. Deep expectation of real and apparent age from a single image without facial landmarks / R. Rothe, R. Timofte, L. Van Gool // International Journal of Computer Vision. – 2018. – Vol. 126(2). – P. 144-157.
- [7] Zhang, Z. Age progression/regression by conditional adversarial autoencoder / Z. Zhang, Y. Song, H. Qi // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2017. – P. 5810-5818.
- [8] Clapés, A. From apparent to real age: gender, age, ethnic, makeup, and expression bias analysis in real age estimation / A. Clapés, G. Anbarjafari, O. Bilici, D. Temirova, E. Avots, S. Escalera // Proceedings of the IEEE conference on computer vision and pattern recognition workshops. – 2018. – C. 2373-2382.
- [9] Sandler, M. Mobilenetv2: Inverted residuals and linear bottlenecks / M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L. Chen // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2018. – P. 4510-4520.
- [10] Tan, M. Efficientnetv2: Smaller models and faster training / M. Tan, Q. Le // International Conference on Machine Learning. – 2021. – P. 10096-10106.
- [11] Chollet, F. Xception: Deep learning with depthwise separable convolutions / F. Chollet // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2017. – P. 1251-1258.