

AgeEstimAI

Attempt to beat the state of the art
in age estimation based on face images

Project for the course Applied Deep Learning 2023 at the TU Vienna.

Project category: Beat the stars

Introduction

Age estimation (AE) has been a subject of research for many years. Over time, various features have been investigated for their relevance in age prediction, ranging from activity and blood data to medical imaging outputs. Historically, the predominant methods involved manual feature extraction combined with traditional machine learning techniques. However, newer approaches often shifted to deep learning. One prominent application of deep learning in this domain is the use of facial images for age estimation.

In age estimation research, a distinction is made between chronological age and biological age. The former, chronological age, refers to the number of years since birth — the conventional understanding of age. In contrast, biological age refers to the notion that an individual's true age might be different from their chronological age due to many factors influencing the aging process throughout their life. In this project, I focus on predicting chronological age, primarily because it is easier to measure.

Datasets

There are multiple publicly available datasets of faces labeled with their chronological age that I intend to use in a combined way.

FG-NET:

This dataset is used for age estimation and face recognition across ages. It is composed of a total of 1,002 images with 82 people aged 0 to 69.

MORPH 2:

MORPH is a facial age estimation dataset, which contains 55,134 facial images of 13,617 subjects ranging from 16 to 77 years old.

IMDB-WIKI:

This is a dataset containing images of celebrities that was scraped from the internet. It contains more than 500 thousand face images with age and gender labels.

Research and state-of-the-art

Deep learning methodologies applied to age estimation from face images have seen continual advancements over the past years. In my review of the literature, I found a few noteworthy contributions that have set benchmarks on this topic.

Rothe et al. (2015) won the ChaLearn LAP 2015 challenge. A major key to their success was creating the IMDB-Wiki dataset. They used convolutional neural networks of VGG-16 architecture with a softmax expected value refinement and achieved a mean average error (MAE) of 3.2 years.

Bobrov et al. (2018), leveraged high-resolution images of only eye corners and achieved a MAE of 2.3 years with this approach.

Othmani et al. (2020) gives a comprehensive overview of different frameworks. Their study offers insights into the performance of several models across multiple datasets. They also propose a model that outperformed existing benchmarks. Their approach uses the Xception architecture which is based on depthwise separable convolution layers. In combined with a pre-training using the CASIA-WEB Face dataset, which enables task specific transfer learning. They yield the best results with a MEA of 2.01 on the MORPH 2 dataset. To my knowledge, this is the current state-of-the-art performance.

To summarize, generally CNN architectures are used, that result in a single output value or that transform the last layer of the model in a way that yields a single value for the estimated age. The parameter optimized is typically done using the mean squared error function.

Possible ways to beat the state-of-the-art

Building on these findings, several avenues can be explored to enhance the current state-of-the-art in age estimation. Here's a closer look at potential improvements.

1. **Data augmentation techniques:** Exploring various data augmentation methods could enhance the model's robustness, potentially leading to better performance.
2. **Advanced CNN architectures:** Given that the current reference point is from 2020, there's a possibility that newer CNN architectures have emerged since then. Utilizing these could yield better results.
3. **Fine-Tuning parameter optimization:** Putting more effort into optimizing parameters during the fine-tuning phase might further improve predictions.

4. **Image cropping over downsizing:** For large datasets that pose challenges during training, image cropping could be a preferable alternative to downsizing. Bobrov et al.'s work suggests the potential of this method.
5. **Diversified pre-training datasets:** Othmani et al. (2020) highlighted the significant influence of pre-training dataset selection on performance. By combining multiple datasets for pre-training, the extend of transfer learning could be increased.

Workplan

1. Dataset preprocessing (8 hours): Collect and preprocess the datasets. Try to combine them in a meaningful way.
2. Designing and building the network (35 hours): Researching recent literature to identify advanced CNN architectures. Implement and compare the performance of these architectures on the age estimation task.
3. Fine-Tuning (25 hours): Experiment with different optimization algorithms, learning rates, and regularization techniques as well as different datasets and tasks for pre-training the model.
4. Application (15 hours): Build Flutter app that lets the user upload an image and get back an age estimation score.
5. Report (3 hours): Write documentation and report.
6. Presentation (5 hours): Prepare the presentation.

In total, I estimate the workload of this project with 91 hours.

References

Bobrov E, Georgievskaya A, Kiselev K, Sevastopolsky A, Zhavoronkov A, Gurov S, Rudakov K, Del Pilar Bonilla Tobar M, Jaspers S, Clemann S. PhotoAgeClock: deep learning algorithms for development of non-invasive visual biomarkers of aging. *Aging (Albany NY)*. 2018 Nov 9;10(11):3249-3259. doi: 10.18632/aging.101629. PMID: 30414596; PMCID: PMC6286834.

(<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6286834/>)

Alice Othmani, Abdul Rahman Taleb, Hazem Abdelkawy, Abdenour Hadid, Age estimation from faces using deep learning: A comparative analysis, *Computer Vision and Image Understanding*, Volume 196, 2020, 102961, ISSN 1077-3142,

(<https://www.sciencedirect.com/science/article/pii/S1077314220300424>)

Rasmus Rothe and Radu Timofte and Luc Van Gool, Deep expectation of real and apparent age from a single image without facial landmarks, International Journal of Computer Vision, Volume 126 (<https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/>)