

Age estimation based on facial images using Deep Learning techniques

Paulo Rijnberg (snr. 2033812), April 2020

1. Introduction

Age estimation plays a significant role in different scenarios, from law enforcements identifying individuals to social interactions. Because the aging process does not only depend on intrinsic factors (e.g. genetic factors) but also on extrinsic factors (e.g. environment, lifestyle), estimating someone's age based only on their face is a challenging problem¹.

Therefore, the literature offers a wide variety of approaches to this problem. Lots of researcher tried approaching this problem by using traditional Machine Learning techniques such as K-Nearest Neighbors² and Support Vector Machines³. Since deep learning showed great potential on images, some have proposed working with deep learning techniques. Sithungu and Van der Haar for example focussed on a lightweight model and proposed a modified LeNet-5 architecture⁴. While Ozbula, Aytar and Ekenel researched the potential of using transfer learning with pre-trained model such as AlexNet and VGG1. ⁵ Lastly, Fariza, Mu'arifin and Arifin also proposed using convolutional layers by following the ResNe(x)t-50 architecture⁶.

Also, most of these proposed models used the imbalanced dataset UTKFace. For this reason, this paper will focus on the following research question: 'To what extent can we apply deep learning techniques on age estimation using a balanced dataset of facial images?' Where we take the paper from Sithungu et al. as a starting point.

2. Experimental procedure

2.1 Dataset

We used the Aligned&Cropped Face images from UTKFace dataset (24,106 images). We binned the data into 7 classes, applied under sampling (the data was highly imbalanced) and removed images with ages over 70 (see Table 1). We split the data into three sets, 75% training, 20% validation and 5% test.

2.2 Experimental setup

We started by implementing the (modified) LeNet-5 architecture just like Sithungu et al. We tried modifying these by adding extra layers, changing the activation and optimizer. After this, we turned to transfer learning VGG16 (Ozbula et al.) and ResNet-50 (Fariza et al.). We tried modifying these by (un)freezing layers, adding extra dense layers and training the model after importing the imagenet weights/from scratch. In every experiment we used a generator, rescaled the data, used batch sizes of 32 and ran 15 epochs.

2.3 Evaluating the performance

Like mentioned previously, we focused on using a balanced dataset, and most other papers used an imbalanced dataset. Thus, our starting point (the modified LeNet-5) that reached an accuracy of 56% in the paper of Sithungu et al., is not representative in our case. Therefore, we have a baseline of 14% (1,183 / 8,281 total balanced samples). In the provided [GitHub page](#), we also applied a confusion matrix.

3. Results

Table 3 shows the most interesting results of the experiment. We display here the model architecture, input shape, compiler and highest accuracy achieved in 15 epochs. [MCA???](#) We also show some layer activations in Figure 1.

Table 2: Models used in the experiment

Model architecture	Input shape	Compiler	Accuracy
1. Original LeNet-5 architecture	32, 32, 1	Optimizer = rmsprop	0.56
2. Modified LeNet-5 architecture proposed by Sithungu et al.	200, 200, 3	Metrics = accuracy Loss = categorical cross entropy	0.56
3. Non-trainable VGG16 layers (weights of 'imagenet') with three trainable dense layers (512, 512, 7)	200, 200, 3		0.49
4. Retrain ResNet-50 layers from scratch with two extra layers (GlobalAvgPool2D, dense 7)	224, 244, 3	Optimizer = adam Metrics = accuracy	0.51
5. Trainable ResNet-50 layers (weights of 'imagenet') with two layers (GlobalAvgPool2D, dense 7)	224, 244, 3	Loss = categorical cross entropy	0.56

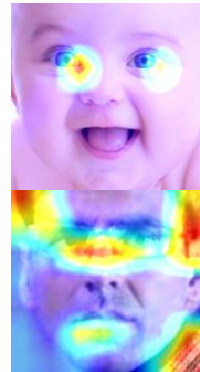


Figure 1: Activations ResNet-50 model (5) on the test set.

CONFUSION MATRIX RESNET

4. Discussion and conclusion

Based on the accuracy, confusion matrix and layers activations we can conclude that the best performing model was model 5 (Trainable ResNet-50 layers with two GlobalAvgPool2D and a dense layer). Even though the original and modified LeNet-5 model achieved the same accuracy, we believe that this is not representable. After visualizing the activations of these models on multiple different images from the test set, we didn't see any activations. This could imply that the model suffers from high variance.

We also noticed that the best performing model takes noticeable less amount of time to make a prediction, so we would propose users of the model to consider this. This obvious, because there is a significantly difference between the number of layers the ResNet-50 has compared to e.g. the original LeNet-5.

Like shown in section 2.1 Dataset, we applied under sampling leaving us with only 1,183 samples per class. We could argue that these are not enough images to fully (re)train a model (e.g. model 4). We suggest that further work could focus on using a bigger dataset and/or apply image augmentation to the minority class(es).

¹ Han, H., Otto, C., & Jain, A. K. (2013). Age estimation from face images: Human vs. machine performance. *2013 International Conference on Biometrics (ICB)*. <https://doi.org/10.1109/icb.2013.6613022>

² Guzey, A., & Nahiye, V. V. (2008). Automatic age classification with LBP. *2008 23rd International Symposium on Computer and Information Sciences*. <https://doi.org/10.1109/isicis.2008.4717936>

³ Guo, G., Guowang, M., Fu, Y., & Huang, T. S. (2009). Human age estimation using bio-inspired features. *2009 IEEE Conference on Computer Vision and Pattern Recognition*. <https://doi.org/10.1109/cvpr.2009.5206681>

⁴ Sithungu, S., & Van der Haar, D. (2019). Real-Time Age Detection Using a Convolutional Neural Network. *Business Information Systems*, 245-256. https://doi.org/10.1007/978-3-030-20482-2_20

⁵ Ozbula, G., Aytar, Y., & Ekenel, H. K. (2016). How Transferable Are CNN-Based Features for Age and Gender Classification? *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*. <https://doi.org/10.1109/biosig.2016.7736925>

⁶ Fariza, A., Mu'arifin, & Arifin, A. Z. (2019). Age Estimation System Using Deep Residual Network Classification Method. *2019 International Electronics Symposium (IES)*. <https://doi.org/10.1109/iesecsym.2019.8901521>

Age estimation based on facial images using Deep Learning techniques

Paulo Rijnberg (snr. 2033812), April 2020

Age estimation based on facial images using Deep Learning techniques

Paulo Rijnberg (snr. 2033812), April 2020

Table 3 displays which models are used in this experiment:

Table 3: Models used in the experiment

Model architecture	Compiler
LeNet-5 architecture	Optimizer = rmsprop, metrics = accuracy, loss = categorical_crossentropy
LeNet-5 architecture (with batch normalization)	Optimizer = rmsprop, metrics = accuracy, loss = categorical_crossentropy
Modified LeNet-5 architecture proposed from paper Sithungu and Van der Haar (2019)	Optimizer = rmsprop, metrics = accuracy, loss = categorical_crossentropy
VGG16 (weights = 'imagenet', include_top = False, layers set to non-trainable) and adding trainable three dense layers (512, 512, 7).	Optimizer = rmsprop, metrics = accuracy, loss = categorical_crossentropy
ResNet-50 (weights = 'imagenet', include_top = False, layers set to non-trainable) and adding GlobalAveragePooling2D and output dense layer with 7 nodes.	Optimizer = adam, metrics = accuracy, loss = categorical_crossentropy
ResNet-50 (weights = 'None', include_top = False, layers set to trainable) and adding GlobalAveragePooling2D and output dense layer with 7 nodes.	Optimizer = adam, metrics = accuracy, loss = categorical_crossentropy

3. Results

4. Discussion and conclusion