Features of using convolutional neural networks for human age estimation from a face image

V. K. Kolodin , D. A. Savelyev

Abstract

In this paper we describe the various factors' influence on the convolutional neural networks' performance on task of determining a person's age. Different approaches to solving the problem are considered, including different neural network architectures and various interpretations of the original problem. The result of studies revealed that by using the Swish activation function and age distribution [-1, 1], as well as the Xception base model, it is possible to reduce the recognition error by 16 %. Moreover, by using above mentioned techniques and joining regression and classification models it is possible to reduce error by 32 % total.

Dataset

The combination of IMDB WIKI, UTKFace and APPA-REAL datasets was used. All images were cropped and resized to 224 x 244 pixels.

Fig. 1. Example of images from dataset

Architecture

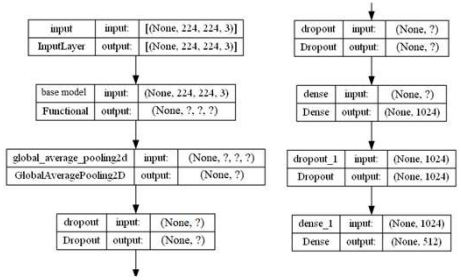

Fig. 2. Architecture for all models. Layers' number of inputs / outputs marked with "?" depends on a base model.

Classification

For classification problem it is necessary to divide all ages into age groups – classes, in this paper 3 different age divisions were examined, and a neural network was trained for each. All divisions are represented in fig. 3, points show boundaries of classes.

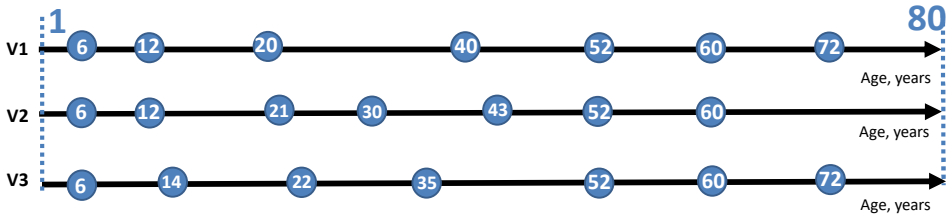

Fig. 3. Representation of age classes in 3 versions of dataset.

Classification results

Table 1. Results of models trained on different versions of dataset

Base architecture	Dataset version	Test accuracy
Xception	1	0.555
Xception	2	0.463
Xception	3	**0.655**

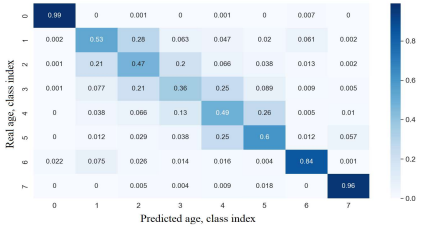

Fig. 4. Confusion matrix for model trained on V3 dataset

Regression

Influence of age mapping, activation function and base model on the performance were considered. Model with Xception base architecture, Swish activation and [-1;1] age mapping preformed better than the others: 7.94 MAE.

Table 2. Results of models trained on different age mapping

Age mapping	Test MAE, years	Error sample variance
[1; 80]	8.75	8.38
[-39.5; 39.5]	8.8	7.18
[0; 1]	8.67	8.62
[-1; 1]	**8.49**	**5.41**
[-0.5; 0.5]	8.63	7.12

Table 3 and 4. Results of models with different base models and trained with different activations

Activation	Test MAE
Tanh	19.75
Leaky ReLU	8.69
Swish	**8.49**

Base model	Test MAE
Mobilenet V2	8.424
Xception	**7.94**
Resnet50	9.25

To improve regression performance, best of trained regression (Table 4, model 2) and classification (Table 1, model 3) models were combined and new network was trained on dataset once more, but only new layers were trained. The performance of the new model was 6.28 years MAE on test dataset.

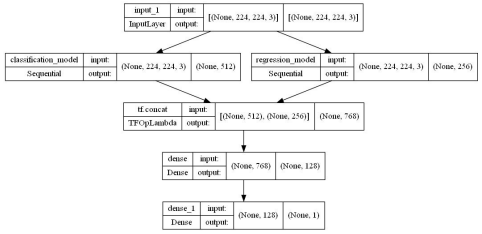

Fig. 5. Architecture of the model, that contains regression and classification models

Conclusion

In this paper, the impact of various factors on neural networks performance on the age recognition problem were studied. It occurred, that for classification problem class division is the most significant factor, whilst for regression problem age mapping, base model and activation function are affecting the result in the same intensity. Moreover, models trained for classification rely on different image features than regression models. Hence the combination of regression and classification models rely on both feature sets, what allows it to perform better.

САМАРСКИЙ УНИВЕРСИТЕТ
SAMARA UNIVERSITY

[2] SAMARA NATIONAL RESEARCH UNIVERSITY
34, Moskovskoye shosse, Samara, Russia (www.ssau.ru)