Age and Gender Classification

183.587 Computer Vision System Programming WS16

Christoph Körner (0726266), christoph.koerner@tuwien.ac.at
Patrick Wahrmann (1327120), patrick.wahrmann@student.tuwien.ac.at

Problem Definition

Our goal in this lecture is to develop a classification system for the task of age (in classes of around 10 years each) and gender estimation on human facial images. For this purpose we are going to use Deep Learning (Keras). The final input images are going to be cropped and registered face images from *"The Profiler"*, a project of the CVL for which we deliver an estimation of age and gender of the depicted person.

Datasets

For training and testing we are planning to use the IMDB-WIKI dataset of Rothe et al. [1]. This dataset consists of two parts (images from IMDb and Wikipedia) and stands out because of its very big size (both parts together: 523,051 faces). The following Figure 1 shows 4 sample images per dataset. In addition, quadratic crops of the faces are also available (IMDb: 7GB, Wiki: 1GB).

Figure 1: Sample images from the IMDb/Wiki dataset [1]

The age labels can be computed from the meta attributes *date of birth* of the actor/person and *photo taken*, the year when the picture was taken; the gender label is available as enum. The following Figure 2 shows the distributions of the age values in both datasets (with a different y-scale).

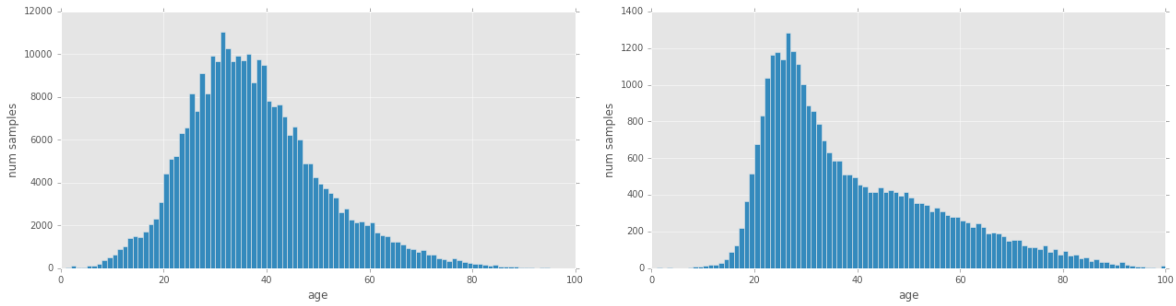

Figure 2: Age distributions from the IMDb (left) and Wiki (right) dataset

Depending on the training time of the Deep Learning model we will might decide on using only a subset of all available images for training.

Solution Approach

In this section we briefly describe the steps that we outlined in order to solve the age and gender classification.

<u>A) Data Preprocessing</u>

The dataset was acquired automatically and hence contains noise (1x1 images, images without faces), outliers (images with multiple faces, images with tiny faces), etc. that need to be filtered. We also observe, that non-quadratic images are very likely to contain outliers (see the images in the following Figure 2). Hence, we decide to use only the square images.

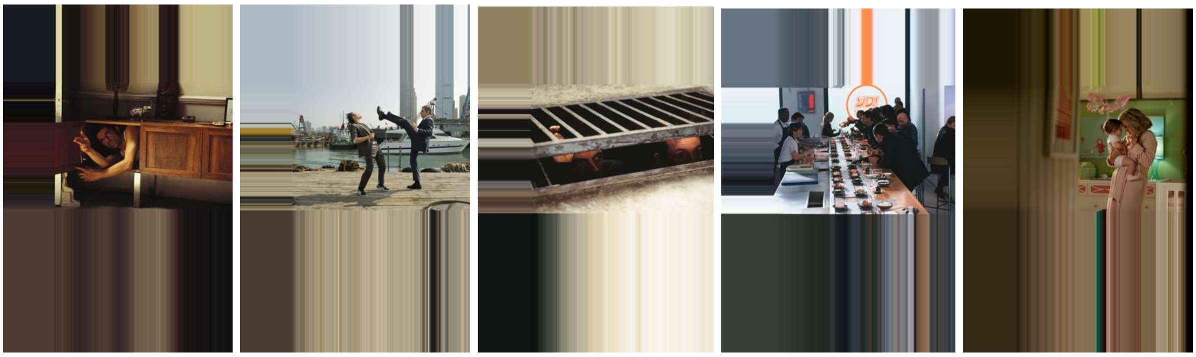

Figure 3: Faulty non-square images from the IMDb dataset

We also discard all images with ages less than 1 and greater than 100.

We will combine the age values into ordinal similar-sized age classes such that all classes contain approximately the same number of samples. The following figure shows an aggregation of the age label to the classes (0,15), (16,20), (21,25), (26,30), (31,35), (36,40), (41,45), (46,50), (51,55) and (56,100) for the combined and cleaned IMDb/Wiki dataset. However, we observe that the classes for teenagers will always contain less samples than the other classes.

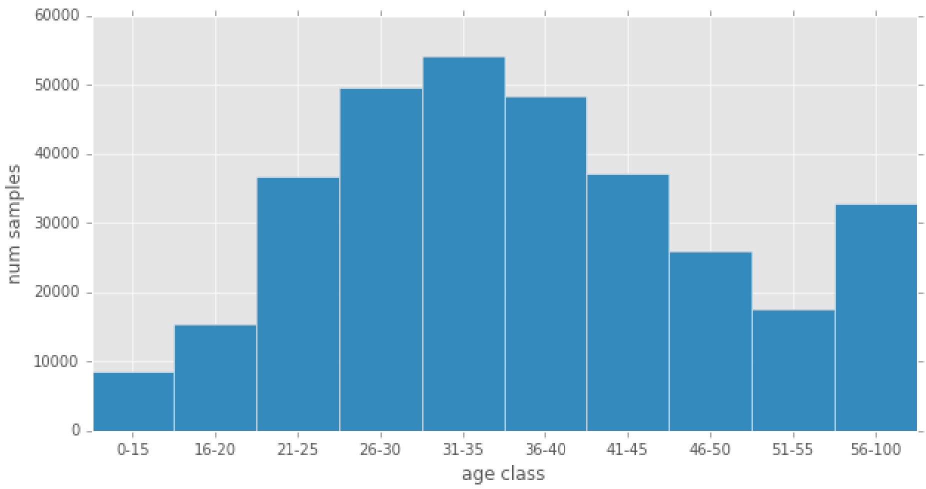

Figure 4: Age classes for the combined and cleaned IMDb/Wiki dataset

B) Dataset Splits

Both datasets will be cleaned and combined to a single face dataset. Then, we will split it into train (90% of the dataset), validation (10% of the training set) and test data (10% of the dataset). The images are resized to a fixed dimension (e.g. 224x224px for the GoogLeNet model). We will use only b/w images for training.

C) Loss Functions

For computing the loss of the gender estimate we will use the standard cross-entropy loss function which is commonly used for binary classification.

For computing the loss of the age estimate we will use a loss function that takes into account the ordinal property of the classes: hence, we will use Mean Squared Error (MSE) or Cohen's Weighted Kappa [2].

D) Deep Neural Network Model

We will use the GoogLeNet [3] architecture to train both the age and gender classification models. According to [4] this model achieves the highest image classification accuracy (top 1 accuracy on ImageNet) with the lowest number of parameters. The model (with pretrained model weights) is available for free with an unrestricted license for commercial use.

Evaluation

The age and gender estimation algorithms will both be evaluated on the test set with reported accuracy, precision, recall and F1-score values. These scores will be compared to state-of-the-art age and gender estimation algorithms and to human classification accuracy (if available).

We will as well evaluate the performance of our algorithm to the pretrained model provided in [1], a state-of-the art deep neural network trained on the same IMDb/Wiki dataset for both age and gender classification.

References

[1] https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/

[2] B. Arie, “Comparison of classification accuracy using Cohen's Weighted Kappa”, *Expert Systems with Applications*, vol. 34/2, pp. 825-832, 2008.

[3] C. Szegedy et al., “Going Deeper with Convolutions”, CoRR, 2014.

[4] A. Canziani, A. Paszke and E. Culurciello, “An Analysis of Deep Neural Network Models for Practical Applications”, CoRR, 2016.