Contest

Guess The Age: Age Estimation From Facial Images with Deep Convolutional Neural Networks

The Guess The Age (GTA) contest is an international competition among methods based on modern Deep Convolutional Neural Networks (DCNNs) for age estimation from facial images. In order to allow the participants to train effective models, we provide a dataset, the Mivia Age Dataset, including 575.073 images annotated with age labels; it is among the biggest publicly available datasets of faces in the world with age annotations. The performance of the proposed methods will be evaluated in terms of accuracy and regularity on a test set of more than 150.000 images, different from the ones available in the training set.

Age estimation from face images is nowadays a relevant problem in several real applications, such as digital signage, social robotics and business intelligence. In the era of deep learning, many DCNNs for age estimation have been proposed, so effective to achieve performance comparable to those of humans.

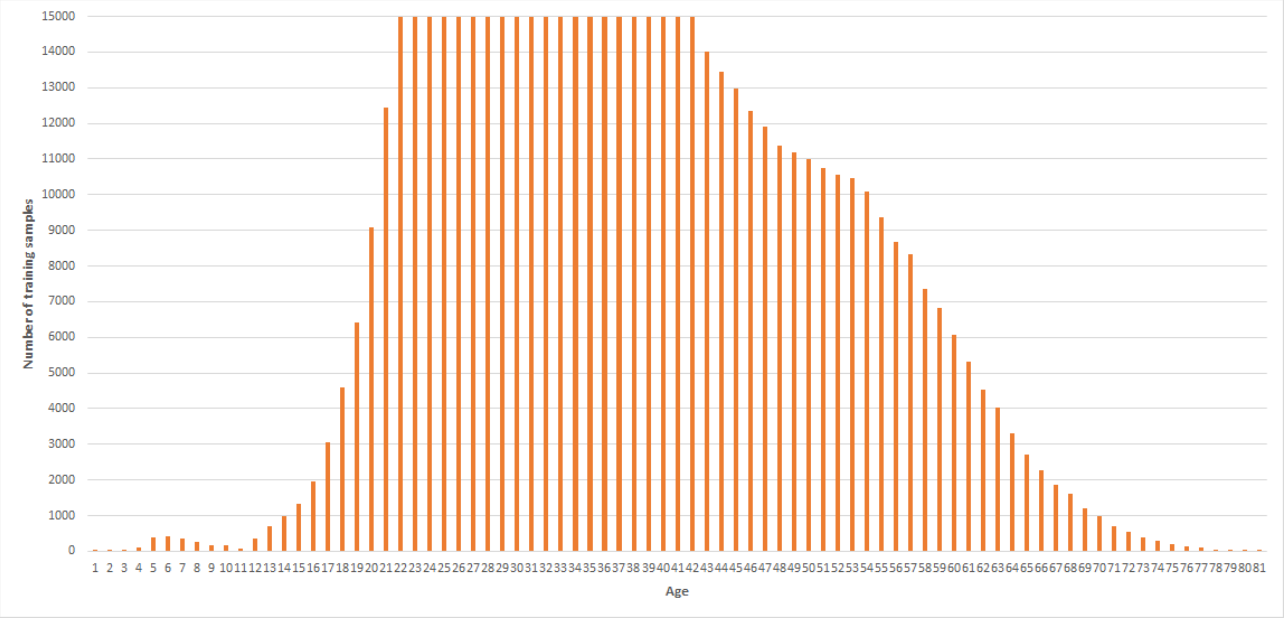
To this concern, it is worth pointing out that the most promising methods use complex ensembles of DCNNs, making the obtained classifier not usable in real applications, as they require prohibitive computational resources not always available; in addition, their training procedure is made complex by the plurality of neural networks and typically require huge training sets not simply collectable.

Within this framework, we propose the usage of a new huge dataset, the MIVIA Age Dataset, and we restrict the competition to methods based on a single neural network or small ensembles composed at most of 3 classifiers.

Description of the dataset to be used

The MIVIA Age Dataset is composed of 575.073 images of more than 9.000 identities, got at different ages. They have been extracted from the VGGFace2 dataset and annotated with age by means of a knowledge distillation technique, making the dataset very heterogeneous in terms of face size, illumination conditions, facial pose, gender and ethnicity. Each image of the dataset contains a single face, already cropped. We make available to the participants an archive with all the training samples and a CSV file with the age annotations.

The distribution of the dataset samples over the age is depicted in the following figure:



Evaluation protocol

The methods proposed by the participants will be evaluated in terms of accuracy and regularity.

We will adopt a modified version of the Mean Absolute Error (MAE) for evaluating the accuracy of the methods submitted to the contest. Suppose that the test set consists of K samples. The age prediction of a method for the i-th sample of the test set is p_i , while the real age label is r_i. Therefore, the absolute age estimation error on the i-th sample is:

$$e_i = |p_i - r_i|$$

The MAE is the average age estimation error over all the K samples of the test set:

$$MAE = \frac{\sum_{i=1}^{K} e_i}{K}$$

However, we will compute the following versions of the MAE:

1. MAE^1 is computed over the samples whose age annotation is in the range 1-10
2. MAE^2 is computed over the samples whose age annotation is in the range 11-20
3. MAE^3 is computed over the samples whose age annotation is in the range 21-30
4. MAE^4 is computed over the samples whose age annotation is in the range 31-40
5. MAE^5 is computed over the samples whose age annotation is in the range 41-50
6. MAE^6 is computed over the samples whose age annotation is in the range 51-60
7. MAE^7 is computed over the samples whose age annotation is in the range 61-70
8. MAE^8 is computed over the samples whose age annotation is in the range 70+

The $mMAE$, our measure of accuracy that takes more into account the regularity of the error, is computed as follows:

$$\frac{\sum_{j=1}^{8} MAE^j}{8}$$

The lower is the $mMAE$ achieved by a method, the higher is its average accuracy over the age groups.

The standard deviation σ is the performance index we will use for evaluating the regularity of the methods over the different age groups:

$$\sigma = \sqrt{\frac{\sum_{j=1}^{8}(MAE^j - MAE)^2}{8}}$$

The lower is the standard deviation achieved by a method, the higher is its regularity.

The final score is the Age Accuracy and Regularity (AAR) index, computed as follows:

$$AAR = \max(0;\ 5 - mMAE) + \max(0;\ 5 - \sigma)$$

The AAR index can assume values between 0 and 10. Methods which achieve $mMAE \geq 5$ and $\sigma \geq 5$ will obtain $AAR = 0$. A perfect method ($mMAE = 0$ and $\sigma = 0$) will achieve $AAR = 10$. Methods which achieve intermediate values of MAE and σ obtain intermediate values of AAR.

The method which achieves the highest AAR will be the winner of the GTA contest.

Rules

1. The first deadline for the submission of the methods is **4th January, 2023**. The second deadline is **8th February 2023**. The submission is intended to be applied to the entire group. It must be done with an email in which the participants share (directly or with external links) the trained model, the code and the report.
2. The participants will receive the training set and the age annotations through an email sent to the first member of the team.
3. The participants can use only the received samples for training and validation, without the possibility to take samples from other datasets. Additional samples of ages less represented in the dataset can be obtained with data augmentation techniques, which are encouraged both for extending the training set and for increasing its representativeness.
4. The participants can use as DCNNs single neural networks or ensembles composed at most of 3 classifiers.
5. The participants must submit their trained model and their code by carefully following the detailed instructions reported below.
6. The participants must produce a PDF report of the proposed method, by following a template that will be sent together with the material described in 2.

Instructions

Follow the instructions reported on the website of the GTA 2021 competition: https://gta2021.unisa.it/ .