Expert Systems With Applications 141 (2020) 112942

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Short communication

Robust regression with deep CNNs for facial age estimation: An empirical study

F. Dornaika[a,b,*], SE. Bekhouche[c,d], I. Arganda-Carreras[a,b,e]

[a] University of the Basque Country UPV/EHU, San Sebastian, Spain
[b] IKERBASQUE, Basque Foundation for Science, Bilbao, Spain
[c] Center for Machine Vision and Signal Analysis (CMVS), University of Oulu, Finland
[d] Department of electrical engineering, University of Djelfa, Algeria
[e] Donostia International Physics Center (DIPC), San Sebastian, Spain

ARTICLE INFO

Article history:
Received 30 April 2019
Revised 8 August 2019
Accepted 6 September 2019
Available online 9 September 2019

Keywords:
Age estimation from facial images
Deep convolutional neural networks
Robust loss functions
Fine tuning
Regression

ABSTRACT

Recent works have shown that deep Convolutional Neural Networks (CNNs) can be very effective for image-based age estimation. However, the proposed approaches significantly vary, and there are still some open problems. Almost all deep regression networks for age estimation have exploited the Mean Square Error loss only. These deep networks have not considered the influence of aberrant and outlier observations on the final model. In this letter, we introduce the use of robust loss functions in order to learn deep regression networks for age estimation. More precisely, we explore the use of two robust regression functions: (i) the ℓ_1 norm error, and (ii) the adaptive loss function that retains the advantages of the ℓ_1 and ℓ_2 norms. Experimental results obtained on four public databases demonstrate that learning a deep CNN with robust losses can improve the age estimation.

1. Introduction

Age estimation from face images received a lot of attention in recent times (Angulu, Tapamo, & Adewumi, 2018). This problem gave rise to a significant amount of research since it has a broad spectrum of practical applications like age-oriented commercial advertisement, police investigation, security, soft biometrics, and age specific Human-Computer Interaction. In the literature, there are two types of age estimation problems. The first one addresses real age estimation, which is to predict the biological age of a subject using his or her face image (Fernandez, Huerta, & Prati, 2014). The second one addresses age group estimation. In this type, the objective is to predict the age group in which the person's age falls (Levi & Hassner, 2015). Broadly speaking, there are two categories of approaches for solving the age estimation from images depending on the type of image representation: (i) hand-crafted methods (e.g., Wang, Yau, & Wang, 2009; Lu, Liong, & Zhou, 2015; Nguyen, Cho, & Park, 2014; Sai, Wang, & Teoh, 2015) and (ii) deep learning methods (e.g., Dehghan, Ortiz, Shu, & Masood, 2017; Dornaika, Arganda-Carreras, & Belver., 2019; Han, Jain, Wang, Shan, & Chen, 2018; Huerta, Fernndez, Segura, Hernando, & Prati, 2015; Ranjan et al., 2015). In the first category, the approaches use a generic-purpose image descriptor and then apply a classification or regression on the obtained shallow image descriptors. In the second category, an end-to-end solution is provided by learning a set of linear and nonlinear functions directly from the raw images. Due to their superior performance in several computer vision tasks, deep learning methods have been recently used and proposed for facial image analysis (e.g., Liu, Lu, Feng, & Zhou, 2017; Rothe, Timofte, & Gool, 2018; Taheri & Toygar, 2019.

In Rothe et al. (2018), the authors introduced a deep learning method for age prediction. They generated a large database (IMDB-WIKI database) which is among the largest public age databases. Their solution adopted the use of age groups but the final estimation provides a real number since the expected age can be estimated from the obtained probability distribution. In Shen et al. (2018), the authors introduced the use of Deep Regression Forests (DRFs) and used it as an end-to-end solution for age prediction. These DRFs receive as input the output of a given deep CNN or any other image representation. The authors devise an algorithm that learns the CNN, the data clusters at the split nodes and the data representation at the leaf nodes. In Taheri and Toygar (2019), the authors estimated ages by exploiting Directed Acyclic Graph Convolutional Neural Networks (DAG-CNNs). Their method uses multi-stage outputs from different layers of the CNN.

* Corresponding author at: University of the Basque Country UPV/EHU, San Sebastian, Spain.

E-mail addresses: fadi.dornaika@ehu.es (F. Dornaika), ignacio.arganda@ehu.eus (I. Arganda-Carreras).

https://doi.org/10.1016/j.eswa.2019.112942
0957-4174/

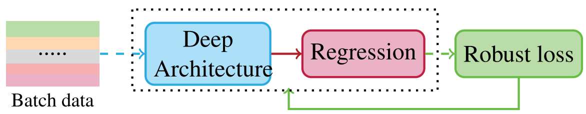

Fig. 1. The flowchart of the proposed facial age estimation method. During the offline stage, the whole network is trained using a given robust loss function.

Their introduced architecture used multi-scale features and fused the scores associated with multiple classifiers. In Lou, Alnajar, Alvarez, Hu, and Gevers (2018), the authors proposed an expression-invariant age estimation method by simultaneously learning the expression and age. They learn the relationship between the expression and age by deploying a graphical model adopting a latent layer. In Liu, Lu, Feng, and Zhou (2018b), the authors introduced an ordinal deep learning scheme. They jointly learn face representation and age predictor.

1.1. Paper contribution

The present submission is not about proposing novel robust loss functions. It focuses instead on the practical aspects of using such robust loss functions for the particular problem of age estimation using deep CNNs. To the best of our knowledge, almost all deep regression networks for age estimation have adopted the Mean Square Error loss. This means that the loss function used for training the CNN is set to the ℓ_2 norm of the age error. However, the ℓ_2 norm is a non-robust estimator that can lead to poor generalization in cases where aberrant data are present in the training set. The existing deep regression networks have not considered the influence of aberrant and outlier observations on the final model. In the domain of age prediction from face images, an aberrant image is an image that has a large prediction error with respect to the majority of training images. Since the regression model is unknown, one cannot decide whether or not a face image is aberrant. Therefore, in the training phase, a robust loss function should reduce the effect of these unexpected errors on the whole CNN model that performs the age regression.

In this paper, we propose the use of robust loss function in order to derive deep regression for age estimation. By retraining a given deep CNN architecture with a robust regression function, we are able to improve the accuracy of age estimation as well as the convergence properties of the training process. More precisely, we explore the use of two robust loss functions: (i) the ℓ_1 norm error, and (ii) the adaptive loss function. The adaptive loss function retains the advantages of the ℓ_1 and ℓ_2 norms. It can handle the Gaussian distribution of data with small losses and the Laplacian distribution of outlier data.

2. Approach: deep CNNs with robust loss

In our work, we start from a given deep network architecture and proceed as follows. If the network architecture provides discrete classification (age groups), its corresponding layers are replaced by a linear regression function that estimates the age. Furthermore, the resulting architecture is retrained using the robust loss functions described below. A graphical illustration of the proposed approach is illustrated in Fig. 1. For a given network and a given loss function, the retraining process adopts the Stochastic Gradient Descent method in order to minimize the loss function over the training images. In the literature of estimation one can find several robust estimators. In the sequel, we will briefly describe the classic MSE loss as well as two robust loss functions that will be used in the deep CNN training.

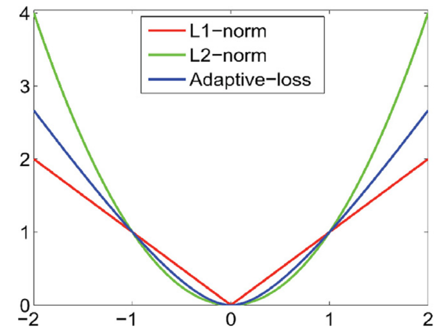

Fig. 2. Three loss functions: MSE, MAE, and the adaptive loss.

2.1. Mean square error (MSE)

The mean square error is given by the square of the residual error over the training set. For N training images, the MSE loss is defined by:

$$L_{MSE} = \frac{1}{N}\sum_{i=1}^{N}(Y_i - T_i)^2 \tag{1}$$

where T_i is the ground-truth age associated with the ith image, and Y_i is the predicted age (the output of the deep CNN). Its first derivative with respect to the prediction Y_i is

$$\frac{\partial L_{MSE}}{\partial Y_i} = \frac{2}{N}\,(Y_i - T_i) \tag{2}$$

2.2. Mean absolute error (MAE)

The Mean Absolute Error (MAE) is the ℓ_1 norm of the error measure between two continuous random variables. For N predictions Y_i and their corresponding targets T_i, the MAE loss is given by:

$$L_{MAE} = \frac{1}{N}\sum_{i=1}^{N}|Y_i - T_i| \tag{3}$$

where N is the number of the training images. The first derivative of the MAE loss with respect to the prediction Y_i is given by:

$$\frac{\partial L_{MAE}}{\partial Y_i} = \frac{1}{N}\,sign(Y_i - T_i) \tag{4}$$

2.3. Adaptive loss

The adaptive loss function (Ding, 2013) takes advantage of both the ℓ_2 and ℓ_1 norms. It is given by:

$$L_{Ada} = \frac{1}{N}\sum_{i=1}^{N}\frac{(1+\sigma)\,(Y_i - T_i)^2}{|Y_i - T_i| + \sigma} \tag{5}$$

where σ is a positive parameter that controls the shape of the loss function. When σ approaches zero the adaptive loss function becomes equivalent to the MAE loss function. If σ approaches infinity then the adaptive loss function becomes equivalent to the MSE loss function.

The derivative of the adaptive loss with respect to the prediction Y_i is given by:

$$\frac{\partial L_{Ada}}{\partial Y_i} = \frac{1}{N}\,\frac{(1+\sigma)sign(Y_i - T_i)(Y_i - T_i)^2 + 2\,\sigma\,(Y_i - T_i)}{(|Y_i - T_i| + \sigma)^2} \tag{6}$$

Fig. 2 illustrates the shape of the above three loss functions.