Automatic Age Range Estimation on Mobile Devices

Gabriel Bertocco*, Fernanda Andaló, Ricardo Torres, Jacques Wainer, Anderson Rocha

Abstract
Automatic age estimation research has grown over the last years due to its importance in many applications, such as security control and Biometrics, and many machine-learning techniques have been applied in order to perform the task. This study explores age range estimation using state-of-art techniques based on Convolutional Neural Network and transfer learning (fine tuning). Our experiments show high accuracy for some age ranges in challenging datasets and hint at several improvements to research in the area.

Key words: age range estimation, deep learning, convolutional neural network

Introduction
Automatic age estimation from face images is one of the most challenging problems in facial analysis, mainly because aging is determined not only by intrinsic factors (genes), but also extrinsic factors (lifestyle, health condition, etc.). Nevertheless, in various contexts, it is of great importance to automatically detect the age of a person. In security control, access to specific content can be allowed or denied based on age; in Biometrics, an age estimator can be used to detect age features from the user to improve results of traditional biometrics systems. The main goal of this research is to perform age range estimation using machine learning and deep learning techniques given a face image of a user taken by a mobile device. This leads to a better user experience and give greater reliability to user interaction.

Results and Discussion
In this research, we leverage a Convolutional Neural Network (CNN) to perform age range estimation and classify a face image in a specific age range. CNNs are the state-of-art methods to perform image classification. We are considering 5 age ranges: lower than 10, 11 to 18, 19 to 40, 41 to 60, and greater than 60 years. The CNN structure comes from the VGG-Face\(^1\) (Figure 1b), which was proposed originally for face recognition, properly adapted for our problem of 5 classes.

We used a transfer learning technique (fine tune) to retrain the VGG network with the specified classes, using popular datasets: FG-Net\(^2\) (Figure 1a) and FACES\(^3\). We used an intra-dataset protocol to fine tune VGG, i.e., the training, validation, and test sets are formed by images from the two datasets, but keeping all images of same identity in just one of the three sets, to ensure fairness and avoid training contamination. The sets are balanced according to the frequency of the classes in the datasets.

In order to augment data, we employ two approaches: random crop and mirror. The goal is to create more samples from the original ones, thus improving the training and the generalization of the model. Table 1 shows the results for the performed experiments.

Table 1. Confusion matrix.

<table>
<thead>
<tr>
<th>Ground Truth</th>
<th>&lt; 10</th>
<th>10-18</th>
<th>19-40</th>
<th>41-60</th>
<th>&gt; 60</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 10</td>
<td>95%</td>
<td>2%</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10-18</td>
<td>5%</td>
<td>78%</td>
<td>6%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19-40</td>
<td>0</td>
<td>20%</td>
<td>44%</td>
<td>3%</td>
<td>1%</td>
</tr>
<tr>
<td>41-60</td>
<td>0</td>
<td>0</td>
<td>48%</td>
<td>69%</td>
<td>1%</td>
</tr>
<tr>
<td>&gt; 60</td>
<td>0</td>
<td>0</td>
<td>2%</td>
<td>28%</td>
<td>98%</td>
</tr>
</tbody>
</table>

Conclusions
The best proposed classifier thus far has yield a reasonable result for young and senior individuals – with the CNN being capable of properly singling out discriminative features for those classes. However, research is still needed to deal with the confusion in the remaining cohorts – accuracy for classes 2 (19-40), and class 3 (41-60) tends to be low, especially when we consider images at the boundary of these classes, and we do not have yet enough samples to achieve better results. The next steps will be devoted to explore new methods and techniques to achieve better results for classes 2 and 3, and to collect more data to train a model with better generalization capability.

Acknowledgement
We are thankful to Motorola and Recod Lab for their support.