**Age and Gender Classification using Convolutional Neural Networks**

Age and gender play fundamental roles in social interactions. Languages reserve different salutations and grammar rules for men or women, and very often different vocabularies are used when addressing elders compared to young people. Despite the basic roles these attributes play in our day-to-day lives, the ability to automatically estimate them accurately and reliably from face images is still far from meeting the needs of commercial applications. This is particularly perplexing when considering recent claims to super-human capabilities in the related task of face recognition (e.g.).

Past approaches to estimating or classifying these attributes from face images have relied on differences in facial feature dimensions or “tailored” face descriptors. Most have employed classification schemes designed particularly for age or gender estimation tasks, including and others. Few of these past methods were designed to handle the many challenges of unconstrained imaging conditions [10]. Moreover, the machine learning methods employed by these systems did not fully challenges of age and gender estimation from real-world, unconstrained images. Most notably, extreme blur (low-resolution), occlusions, out-of-plane pose variations, expressions and more.

In this paper we attempt to close the gap between automatic face recognition capabilities and those of age and gender estimation methods. To this end, we follow the successful example laid down by recent face recognition systems:

Face recognition techniques described in the last few years have shown that tremendous progress can be made by the use of deep convolutional neural networks (CNN). We demonstrate similar gains with a simple network architecture, designed by considering the rather limited availability of accurate age and gender labels in existing face data sets.

**Age and Gender Classification**

Age classification. The problem of automatically extracting age related attributes from facial images has received increasing attention in recent years and many methods have been put fourth. A detailed survey of such methods can be found in and, more recently. We note that despite our focus here on age group classification rather than precise age estimation (i.e., age regression), the survey below includes methods designed for either task.

Early methods for age estimation are based on calculating ratios between different measurements of facial features. Once facial features (e.g. eyes, nose, mouth, chin, etc.) are localized and their sizes and distances measured, ratios between them are calculated and used for classifying the face into different age categories according to hand-crafted rules. More recently, uses a similar approach to model age progression in subjects under 18 years old. As those methods require accurate localization of facial features, a challenging problem by itself, they are unsuitable for in-the-wild images which one may expect to find on social platforms. On a different line of work are methods that represent the aging process as a subspace or a manifold. A drawback of those methods is that they require input images to be near-frontal and well-aligned. These methods therefore present experimental results only on constrained datasets of near-frontal images. Again, consequently, such methods are ill-suited for unconstrained images. Different from those described above are methods that use local features for representing face images.

Gaussian Mixture Models (GMM) were used to represent the distribution of facial patches. In [54] GMM were used again for representing the distribution of local facial measurements, but robust descriptors were used instead of pixel patches. Finally, instead of GMM, Hidden-Marko Model, super-vectors were used in for representing face patch distributions.

An alternative to the local image intensity patches is robust image descriptors: Gabor image descriptors were used in along with a Fuzzy-LDA classifier which considers a face image as belonging to more than one age class. In a combination of Biologically Inspired Features (BIF) and various manifold-learning methods were used for age estimation. Gabor and local binary patterns (LBP) features were used in along with a hierarchical age classifier composed of Support Vector Machines (SVM) to classify the input image to an age-class followed by a support vector regression to estimate a precise age. Finally, proposed improved versions of relevant component analysis and locally preserving projections. Those methods are used for distance learning and dimensionality reduction, respectively, with Active Appearance Models as an image feature. All these methods have proven effective on small and/or constrained benchmarks for age estimation. To our knowledge, the best performing methods were demonstrated on the Group Photos benchmark. In state-of-the-art performance on this benchmark was presented by employing LBP descriptor variations and a dropout-SVM classifier. We show our proposed method to outperform the results they report on the more challenging Adience benchmark, designed for the same task.

**Gender classification.**

Here we quickly survey relevant methods. One of the early methods for gender classification used a neural network trained on a small set of near-frontal face images. In the combined 3D structure of the head (obtained using a laser scanner) and image intensities were used for classifying gender. SVM classifiers were used by, applied directly to image intensities.

Rather than using SVM, used AdaBoost for the same purpose, here again, applied to image intensities. More recently, used the Weber’s Local texture Descriptor for gender recognition, demonstrating near perfect performance on the FERET benchmark. In, intensity, shape and texture features were used with mutual information, again obtaining near-perfect results on the FERET benchmark.

**Deep convolutional neural networks**

One of the first applications of convolutional neural networks (CNN) is perhaps the LeNet-5 network described by for optical character recognition. Compared to modern deep CNN, their network was relatively modest due to the limited computational resources of the time and the algorithmic challenges of training bigger networks. Though much potential laid in deeper CNN architectures (networks with more neuron layers), only recently have they became prevalent, following the dramatic increase in

both the computational power (due to Graphical Processing Units), the amount of training data readily available on the Internet, and the development of more effective methods for training such complex models. One recent and notable examples is the use of deep CNN for image classification

on the challenging Imagenet benchmark. Deep CNN have additionally been successfully applied to applications including human pose estimation, face parsing, facial keypoint detection, speech recognition and action classification. To our knowledge, this is the first report of their application to the tasks of age and gender classification from unconstrained photos.

**A CNN for age and gender estimation**

Gathering a large, labeled image training set for age and gender estimation from social image repositories requires either access to personal information on the subjects appearing in the images (their birth date and gender), which is often private, or is tedious and time-consuming to manually label. Data-sets for age and gender estimation from real-world social images are therefore relatively limited in size and presently no match in size with the much larger image classification data-sets. Overfitting is common problem when machine learning based methods are used on such small image collections.

This problem is exacerbated when considering deep convolutional neural networks due to their huge numbers of model parameters. Care must therefore be taken in order to avoid overfitting under such circumstances.