© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 13 Issue: 03 | NCASIT-2022 | April 18th 2022
Paper is available at http://www.jrps.in | Email : info@jrps.in
Refereed & Peer Reviewed

# Real-time Age, Gender and Emotion Detection using Caffe Models

Prathamesh Kirpal, Nihal Kuthe, Prof. M.M. Gudadhe, Snehal Gajbhiye, Achal Tumsare, Anoop Chahande
Dept. of Information Technology, Priyadarshini College of Engineering
Nagpur, India

**ABSTRACT**

Age and gender classification has become applicable to an extending measure of applications, particularly resulting in the ascent of social platforms and social media. Regardless, execution of existing strategies on real-world images is still fundamentally missing, especially when considering the immense bounce in execution starting late reported for the related task of face acknowledgment. In this paper we exhibit that by learning representations through the use of significant Convolutional Neural Network (CNN) and Extreme Learning Machine (ELM). CNN is used to extract the features from the input images while ELM defines the intermediate results. We experiment our architecture on the recent Audience benchmark for age and gender estimation and demonstrate it to radically outflank current state-of-the-art methods. Experimental results show that our architecture outperforms other studies by exhibiting significant performance improvement in terms of accuracy and efficiency.

**Introduction**

A face contains much information, such as age, gender, and emotions . Age and gender have features of personal identity, which play important roles in social life. And the recognition of emotions can help us understand the psychological state of the participants [4].The estimations of age, gender and emotion can make the feedback more comprehensive [4].

We found a system to estimate the age, gender, and emotions of the face by using convolutional neural networks and residual network structures. This system can be applied in retail industries, by helping merchants to build their own databases to analyze shopping preferences and shopping experiences of users with different ages and genders, and then adjust purchase lists or operating models in a timely manner [4].

In many cases, these systems are able to outperform humans. However, this still remains a difficult problem and existing commercial systems fall short when dealing with real world scenarios [2]. In this work, we present an end-to-end system capable of estimating facial attributes including age, gender and emotion with low error rates [2]. In order to support our claims, we tested our system on several benchmarks and achieved results better than the previous state-of-the-art [2].

**METHODOLOGY**

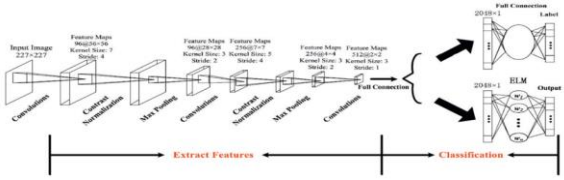


**Fig.The architecture of Age and Gender detection using CNN+ELM Model [3].**

Fig. shows the architecture of our CNN–ELM [3]. It can be seen from the figure that our network includes two stages, feature extraction

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 13 Issue: 03 | NCASIT-2022 | April 18th 2022
Paper is available at http://www.jrps.in | Email : info@jrps.in
Refereed & Peer Reviewed

and classification [3]. The stage of feature extraction contains the convolutional layer, contrast normalization layer, and max pooling layer [3]. The first convolutional layer consists of 96 filters, and the size of its feature map is 56 ×56 while its kernel size is 7 and the stride of the sliding window is 4 [3]. A single convolution layer is implemented after the two stages, and a full connection layer converts the feature maps into 1-D vectors which is beneficial to the classification [3]. Finally, the ELM structure is combined with the designed CNN model, and this architecture is used to classify the age and gender tasks [3].

**Related**

Several works have been done so far for real-time face detection, facial emotion recognition, and gender-age classification [1]. For this project, we reviewed the current literature on convolutional face detection and age, gender and classification and facial emotion recognition [1]. We found that convolutional face detection and age gender & emotion classification is still evolving as a technology, despite outranking other face detection and gender classification methods For the free availability of datasets and pre-trained networks, it is possible to make a functional implementation of a deep neural network without permission to specialist hardware. Pretrained networks can also be used as a starting point for training new networks, decreasing costly training time such as vgg16 and inception v3 [1].

**Face Detection**

The Facial appearance detection and recognition system performs in the three learning stages in just one Convolution neural network (CNN) [1]. The proposed function operates in two main phases: first one training and second test. During training, the system receives training data comprising gray-scale images of faces with their respective expression id and eye center locations and learns a set of weights for the network [1]. To ensure that the training performance is not expected by the order of presentation of the examples, a few images are separated as validation and are used to choose the best set of weights out of a set of training performed with samples presented in different orders [1]. During the test, the system receives a gray-scale image of a face along with its eye center locations and outputs the predicted expression by using the neural network weights learned during training [1].

Face Detection is a technology which is used in different applications that detect human faces in digital images [1]. Face detection is also used for the psychological process by which humans locate and attend to faces in a visual scene. We used OpenCV to catch the live image [1]. Here, for detection of the human faces, we used the HaarCascade image processing method [1]. We saw that there was a situation where it didn't detect the human faces in the live images for the lack of contrast [1]. So, we used histogram equalization to improve detection by increasing contrast. Haar-cascade: Face detection using Haar-cascade is based upon the training of a Binary classifier system using the number of positive images that represent the object to be recognized (such as faces of different peoples at the different scene) and even large number of negative images that indicate objects or feature not to be detected (images that are not human faces but can be anything else like a table, chair, wall, etc.) Actual Image Extracted human face [1].

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 13 Issue: 03 | NCASIT-2022 | April 18th 2022
Paper is available at http://www.jrps.in | Email : info@jrps.in
Refereed & Peer Reviewed

**Age Classification:**

Recently there have been some efforts in collecting data with corresponding age labels [2]. Among those, the dataset proposed by Rothe et al. in is the largest dataset that contains 523, 051 images and is available for research purposes [2]. However, the dataset is not carefully annotated and contains many mistakes. Additionally the distribution of the data across different ages is highly unbalanced [2]. This led to the authors using only half of the data for training in the original paper [2]. To better address this problem we collected a large dataset of ~ 6,00,000 images with corresponding age labels [2]. In contrast to previous works, our dataset has a more balanced distribution across different ages [2]. For example we have over 120, 000 people in our dataset with labeled ages over 70 or younger than 20 years of age [2]. We used a team of human annotators to further clean our dataset through a semi-supervised procedure [2].

**Gender Classification**

In Gender Classifications there are two important point first is age and second is gender, play a very different role in social interactions, making age and gender estimation from a single human face image an important job in intelligent applications, such as human-computer interaction, access control, marketing intelligence, law enforcement, visual surveillance, etc [1]. A preprocessing method which can collect facial and other physical characteristics from the image, a neural network which can classify the gender from the ensemble, an algorithm which can integrate the part-based information and ensemble, based on the database that connects the peculiarity of these physical features for females and males should work [1].

**Emotion Classification**

Emotion Classification Images can both express and affect people's emotions [1]. It is interesting and essential to understanding what emotions are conveyed and how they are implied by the visual content of images. Inspired by the recent success of deep convolutional neural networks (CNN) in visual recognition, they explore simple, yet effective deep learning-based methods for image emotion analysis [1].

We extract attributes using the fine-tuned CNN at the different addresses at multiple levels to capture both the global and local information. The features at different locations are aggregated using the Fisher Vector for each level and concatenated to form a compact representation.

**Basic Operations and training process on CNN**

Primary Operations on CNN Convolutional Neural Network (CNN or ConvNet) is a class of deep artificial neural networks that have successfully been functional in analyzing visual imagery [1]. Simple ConvNet for Emotion & Gender classification could have the architecture [INPUT - CONV – ReLU - POOL - FC] [1]. There are four main operations in the ConvNet [1]. Figure shows the basic CNN architecture for classification where the first portion describes the feature extraction part and the next portion describes the classification part [1].

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 13 Issue: 03 | NCASIT-2022 | April 18th 2022
Paper is available at http://www.jrps.in | Email : info@jrps.in
Refereed & Peer Reviewed

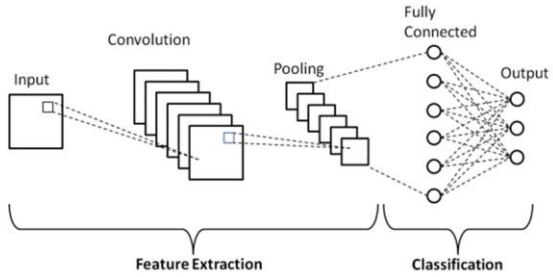


Fig. : Basic CNN Architecture [1].

A: Steps in the training process of CNN

Step 1: Initialize all filters and attributes/weights with random values

Step 2: The method takes a training image as input, goes through the forward propagation step (convolution, ReLU, and pooling operations with forwarding propagation in the completely connected layer and searches the output probabilities for each class [1].

Step 3: Calculate the total error at the output stage Total error = ∑1/2 (Target Probability – Outcome Probability)2

Step 4: Use Back-propagation to evaluate the gradients of the error concerning all weights in the method and use gradient descent to update all filter (values )/weights and parameter values to minimize the output error. Step5: again steps 2-4 with all images in the training set [1].

## Conclusion

We have proposed and tested general building designs for creating real-time CNNs. Our proposed architectures have been systematically built to reduce the number of parameters. We began by eliminating the fully connected layers and by reducing the number of parameters in the remaining convolutional layers via depth-wise separable convolutions. We have shown that our proposed models can be stacked for multi-class classifications while maintaining real-time inferences. Specifically, our vision system can perform face detection, gender classification, and emotion classification in a single integrated module. We have achieved human-level performance in our classification tasks using a single CNN that leverages modern architecture constructs. Our architecture reduces the number of parameters 80× while obtaining favorable results.

Finally, we developed a visualization of the learned features in CNN using the guided back propagation visualization. This visualization technique can show us the high-level features learned by our models and discuss their interpretability.

## References

1. A Convolutional Neural Network for Real-time Face Detection and Emotion & Gender Classification-[1]
2. DAGER: Deep Age, Gender and Emotion Recognition Using Convolutional Neural Networks-[2]
3. The architecture of Age and Gender detection using CNN+ELM Model- [3]
4. Face Recognition with Age, Gender and Emotion Estimations- [4]