Automated Detection of Gender from Face Images

Revathi Ramachandran Nair[1], **Reshma Madhavankutty**[2], **Dr. Shikha Nema**[3]

[1]*UG Student, Dept. of Electronics and Communication, UMIT, SNDT, Mumbai, Maharashtra, India*
[2]*UG Student, Dept. of Electronics and Communication, UMIT, SNDT, Mumbai, Maharashtra, India*
[3]*Head of Dept., Dept. of Electronics and Communication, UMIT, SNDT, Mumbai, Maharashtra, India*

---***---

Abstract - *Humans are capable of determining an individual's gender relatively easily using facial attributes. Although it is challenging for machines to perform the same task, in the past decade incredible strides have been made in automatically making prediction from face image. The project identifies or detects the gender from the given face images. The tools used involve Convolutional Neural Network along with programming language like Python. The project has been motivated by problems like lack of security, frauds, child molestation, robbery, criminal identification.*

Key Words: **CNN, Gender, Machine Learning, Python, Deep Learning**

1. INTRODUCTION

Automatically predicting demographic information such as gender from face images is becoming increasingly significant for law enforcement and intelligence agencies. Humans are capable of determining an individual's gender relatively easily using facial attributes. Although it is challenging for machines to perform the same task, in the past decade incredible strides have been made in automatically predicting gender from the face. The complexity of predicting demographic information depends on the type of demographic category being predicted and the availability of adequate dataset [3]. The project involves identifying or detecting the gender from the given face images. The project uses Deep Learning Technology where Convolutional Neural Network (CNN) acts as a classifier. CNN is used in applications where both classification speed and maximum accuracy is considered important unlike Neural Networks which focuses on classification speed [1].

1.1 Motivation and Problem Statement

The number of crimes has been increasing daily at a much faster rate. It has become a necessity to identify criminals as soon as possible. The traditional way of identification is a slow process while the proposed approach can be used to counter terrorism by identifying the features at a much faster rate. The project can also be used to overcome the frauds that can take place during voting i.e. can be used for voter identification. The old generation has the difficulty to operate computers with ease. This bridge can be lessened by improving Human-Computer Interaction (HCI). The child molestation cases can be tackled at a faster rate by comparing school surveillance camera images to know child molesters and the same can be used for verifying the court records thereby minimizing victim trauma. Similarly, it can also be used for surveillance at banks and residential areas. The technologies used in the project are Machine Learning - supervised, Image Processing - Digital images of the face region, Deep Learning - Convolutional Neural Network and Deep Learning - Tensor Flow. Supervised learning is a machine learning algorithm wherein the input is mapped to the output with the help of training data consisting of input output pairs. TensorFlow, an open source library, is used for mathematical computation, dataflow programming and various machine learning applications. TensorFlow computations are expressed as stateful dataflow graphs. These arrays are referred to as tensors. Convolutional Neural Network (CNN) as one of the most prevalent algorithm has gained a high reputation in image features extraction [2].

2. LITERATURE REVIEW

E. Makinen et al. [1] presented a system which classifies the detected and aligned face images based on the gender. The paper concluded that manual alignment method provides better classification rates as compared to that of automatic alignment method. One of the findings was that different input image sizes did not affect the classification accuracy rates. S. U. Rehman et al. [2] presented a new architecture for face image classification named unsupervised CNN. A CNN is required where a single CNN handles multitask (i.e. Facial detection and emotional classification) by merging CNN with other modules/algorithms. N. Jain et al. [4] presented a hybrid deep CNN and RNN (Recurrent Neural Network) model. The model was proposed to improve the overall result of face detection. The proposed model is assessed based on MMI Facial Expression and JAFFE dataset. G. Levi et al. [5] proposed a convolutional network architecture which classifies age and gender with a small amount of data. They have trained the model on Adience Benchmark. S. Turabzadeh et al. [6] proposed a system where a real-time automatic facial expression system was designed, implemented and tested on an embedded device that can be a first step for a specific facial expression recognition chip for a social robot. The system was built and simulated in MATLAB and then was built on an embedded system. N. Srinivas et al. [3] explores the hardship of performing automatic prediction of age, gender and ethnicity on the East Asian Population using a Convolutional Neural Network (CNN). Predictions based on a refined categorization of the human population (Chinese, Japanese, Korean, etc.) are known as fine grained ethnicity. According to earlier results, the prediction of the fine-grained ethnicity of an individual is the most challenging task, followed by age and lastly gender. A. Dehghan et al. [7] presented a paper consisting of an automated recognition system for age, gender and emotion which was trained using the deep neural network. A. Krizhevsky et al. [8] participated in ImageNet LSVRC-2010 contest and proposed a paper in which 1.2 million images were segregated into 1000 different categories by training a large, deep convolutional neural network. The results obtained using the technique proposed in the paper shows that outstanding results can be achieved with the help of supervised learning. Most of the datasets mentioned in the papers above are of no use. This is because some are paid data sets and some are datasets for age which is not required for the project. Also some datasets contain annotations or landmarks on face images which is not useful for face recognition. In some of the papers mentioned above, RNN is used but this algorithm is not applicable for the project because the input for RNN is text or speech whereas the input for the project is image. Hence, CNN is used as an algorithm for the project. Also, in some papers unsupervised CNN is used. But for the project, supervised learning is a more appropriate approach and hence supervised CNN is used. In the project, the dataset used for gender is UTKFace.

3. SYSTEM DESIGN

In the project, the dataset images are input into the algorithm to identify gender. The UTKFace dataset is used for gender classification. Here, the CNN is used as a classifier/ algorithm. Convolutional networks were inspired by biological processes. CNNs require less amount of pre-processing in comparison with other image classifiers. Their applications include image and video recognition as well as natural language processing. CNNs are often used in image recognition systems. In one of the research papers, the error rate was found to be very low. The learning process was reported to be relatively fast when CNN was used. The error rate decreases drastically when CNN is used for facial recognition. The CNN works similar to that of humans. In CNN, multiple hidden layers are present between the input layer and the output layer. A large amount of data is usually required in CNN to avoid overfitting.

As shown in Fig. 1., the input given is in the form of a face image to the pre-processing unit. The pre-processing unit analyzes the image features based on the algorithms. Data preprocessing for machine learning is a technique that is used to convert the raw data into a clean dataset i.e whenever the data is gathered from different sources, it is collected in raw format and raw format is not appropriate for the analysis. After pre-processing the data, the model is trained using this clean dataset. An unknown image is then inputted to predict the gender of the unknown image.

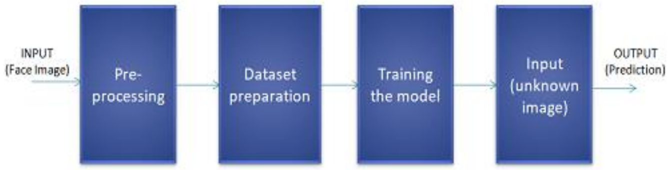


Fig -1: Basic Block Diagram

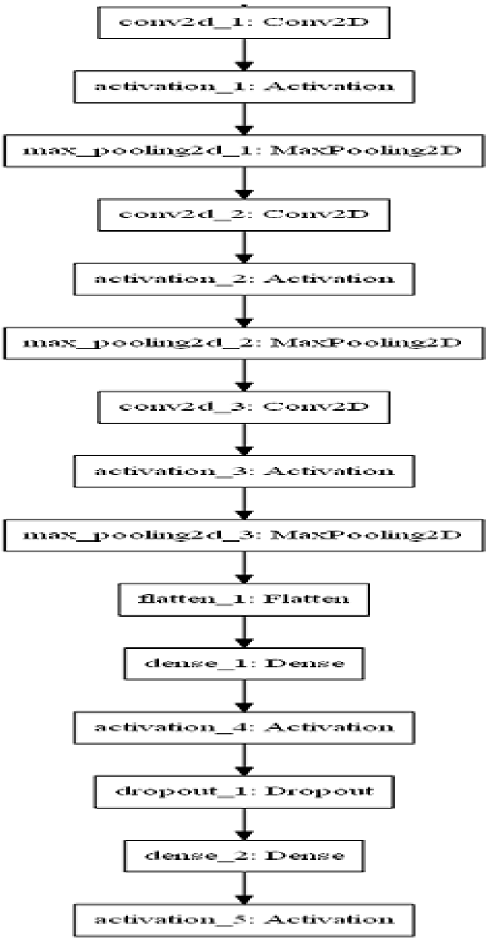


Fig -2: Model for training the dataset

The output will have the information regarding the gender of the unknown input face image. As shown in Fig. 2., the model consists of five stack of layers. The layers include Convolutional layer, Activation layer, Max Pooling layer, Flatten layer, Dense layer and Dropout layer. The first three stacks consist of Convolutional layer, Activation layer and Max Pooling layer. The activation layer used is Rectified Linear Unit (ReLu). The fourth stack consists of Flatten layer, Dense layer and Activation layer. After the first three layers, the output is in 3D and needs to be converted to 1D form. The Flatten layer is used to convert the 3D output to 1D format. The Dense layer is also known as Fully Connected layer. It is used to convert the matrix into a list format and all the nodes are connected to each other. The Activation layer used is Rectified Linear Unit (ReLu). The fifth stack consists of Dropout layer, Dense layer and Activation layer. The

Dropout layer is used to remove the duplicate images present in the dataset to avoid system to undergo overfitting. The Activation layer used is Sigmoid. The sigmoid classification is best for binary classification problems such as gender. Generally, the network is trained using a larger and domain related dataset. After the convergence of the network parameters, an extra training step is done to optimize the network weights using in-domain data. This allows the system to apply convolutional neural networks on small training sets.

4. IMPLEMENTATION

To start with the project, the first step that needs to be done is data collection. Datasets play an important in deep learning as it is used to train the system to get the required output. Some datasets are available publicly while some are not. The UTKFace is a publicly available dataset which can be used for gender classification. This dataset has images of people of different age, gender and ethnicity. In the project, the dataset images are input into the algorithm to identify the gender. The UTKFace dataset is used to train the model to perform gender classification. The input given is in the form of a face image and the image features are analyzed based on the algorithm. An unknown image is then inputted to predict the gender of the same. An output will be generated which will contain the gender prediction of the unknown face image. The UTKFace dataset is categorized into 'Train' and 'Validation', each of which contains 'Male' and 'Female'. The project is trained using 8,000 images in each class and validated using 1,000 images in each class. The model is trained using ConvNet (Convolutional Neural Network) consisting of 5 layers. The CNN consists of many hidden layers such as Convolutional layer, ReLu layer, Max Pooling layer, Fully Connected layer, etc. Using these layers, the input face image is converted into weights and saved in '.h5' format. These weights are then used to predict an unknown image. The average accuracy achieved in the project is 90%.

The training program includes data augmentation. Data augmentation means increasing the number of images in the dataset because plentiful high-quality information is the key to significant machine learning models. Foremost, training examples needs to be augmented via a variety of random transformations, so that the model would never see twice the exact same picture and this helps prevention of overfitting thereby generalizing model in a better way.

Fig -3: Output:Augmentation

In the project, Keras is used to work on Tensorflow. Keras is an open source neural network library. It is user-friendly and provides many features like activation function, layers, optimizers, etc. Keras support both CNN and RNN. Using Java Virtual Machine (JVM), deep models can be created on iOS and Android. This can be achieved by using an appropriate class in keras. Keras permits the model to perform random transformations and normalization operations on batches of image data. The attributes used are rotation range, width shift and height shift, rescale, shear range, zoom range, horizontal rip and fill mode. By using these attributes, the system can automatically rotate pictures, translate pictures, rescale pictures, zoom into pictures, apply shearing transformations, rip images horizontally, fill newly created pixels, etc. Convnet is the right tool for image classification. Data augmentation is a way to fight overfitting, however, it is not enough since the augmented samples are still highly correlated. The main focus for fighting overfitting must be the entropic capability i.e. the abundant information that the model is allowed to store. A model which can store a lot of information has the potential to be more accurate by leveraging more features in comparison to a model that can only store a few features. But the former is also more at risk as it will start storing irrelevant features whereas a model that can only store a few features will have to focus on the most significant features found within the data. The one which stores fewer features are more likely to be truly relevant and are easy to generalize. The setup for the project is as follows:

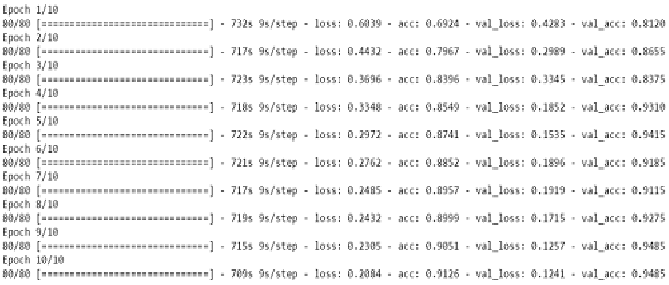
- 16000 training examples (8000 per class)
- 2000 validation examples (1000 per class)

Training Dataset: The training dataset is a set of examples employed to train the model i.e. to fit the parameters. Most of the approaches used for training the samples tend to overfit if the dataset is not increased and used in variety.

Validation Dataset: A validation dataset is also called the 'development dataset' or 'dev set' and is used to fit the hyper parameters of the classifier. It is necessary to have a validation dataset along with training and test dataset because it helps avoid over fitting. The ultimate goal is to choose a network performing the best on unseen data hence we use validation dataset which is independent of the training dataset.

Test Dataset: The test dataset is not dependent on the training or validation dataset. If a model is fitting both the training dataset as well as test dataset then it can be said that minimum overfitting has taken place. The test dataset is the dataset which is only used to test the performance of the classifier or model. The test dataset is employed to check the performance characteristics like the accuracy, loss, sensitivity, etc.

Here the class, .flowfromdirectory() is used to generate groups of image data (and their labels) directly from jpgs in the respective folders and these generators can then be used to train the model. An epoch is one complete representation of the dataset to be learned by the learning machine. One epoch is when the entire dataset is passed both forward and backward through the neural network only once. This approach gives a validation accuracy of 0.81-0.94 after 10 epochs.



```
Epoch 1/10
80/80 [==============================] - 732s 9s/step - loss: 0.6039 - acc: 0.6924 - val_loss: 0.4283 - val_acc: 0.8120
Epoch 2/10
80/80 [==============================] - 717s 9s/step - loss: 0.4432 - acc: 0.7967 - val_loss: 0.2989 - val_acc: 0.8655
Epoch 3/10
80/80 [==============================] - 723s 9s/step - loss: 0.3696 - acc: 0.8396 - val_loss: 0.3345 - val_acc: 0.8375
Epoch 4/10
80/80 [==============================] - 718s 9s/step - loss: 0.3348 - acc: 0.8549 - val_loss: 0.1852 - val_acc: 0.9310
Epoch 5/10
80/80 [==============================] - 722s 9s/step - loss: 0.2972 - acc: 0.8741 - val_loss: 0.1535 - val_acc: 0.9415
Epoch 6/10
80/80 [==============================] - 721s 9s/step - loss: 0.2762 - acc: 0.8852 - val_loss: 0.1896 - val_acc: 0.9185
Epoch 7/10
80/80 [==============================] - 717s 9s/step - loss: 0.2485 - acc: 0.8957 - val_loss: 0.1919 - val_acc: 0.9115
Epoch 8/10
80/80 [==============================] - 719s 9s/step - loss: 0.2432 - acc: 0.8999 - val_loss: 0.1715 - val_acc: 0.9275
Epoch 9/10
80/80 [==============================] - 715s 9s/step - loss: 0.2305 - acc: 0.9051 - val_loss: 0.1257 - val_acc: 0.9485
Epoch 10/10
80/80 [==============================] - 709s 9s/step - loss: 0.2084 - acc: 0.9126 - val_loss: 0.1241 - val_acc: 0.9485
```

Fig -4: Accuracy and Loss in Each Epoch

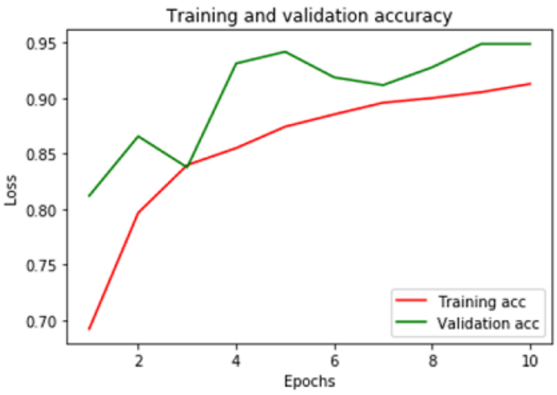


Fig -5: Training and Validation Accuracy Graph

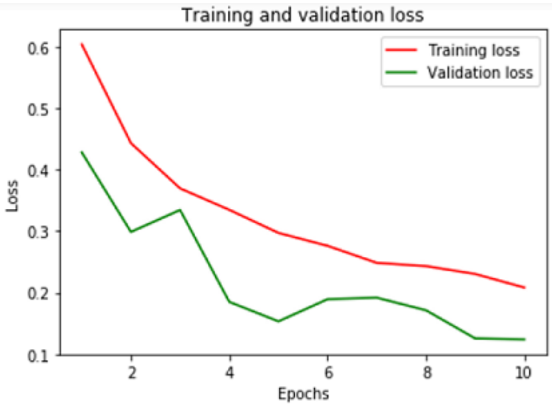


Fig -6: Training and Validation Loss Graph

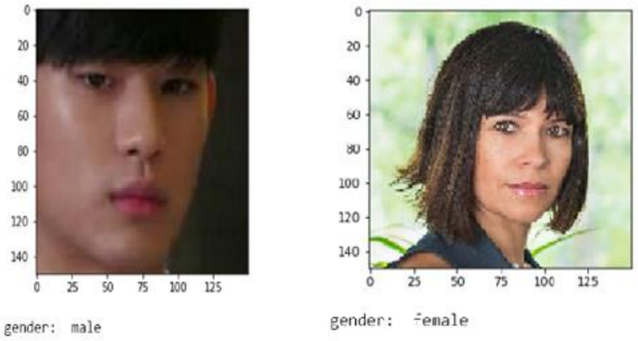


Fig -7: Output Prediction

5. CONCLUSION

Convolutional Neural Network, a supervised machine learning algorithm gives accurate and better results as compared to other algorithms. For gender classification, the model is trained on the pre-processed data and hence is able to determine the gender of the face image. The categories used for gender classification are: male and female. This approach gives an average validation accuracy of 90% after 10 epochs for gender. The variance of validation accuracy is not that high because only less validation samples are used. More the number of samples more will be the accuracy of the model. The accuracy of the system can be increased by increasing the images in the dataset, changing the small ConvNet to VGG16 architecture, using bottleneck features etc.

6. FUTURE WORKS

Upon changing the dataset, the same model can be trained to predict emotion, age, ethnicity, etc. The gender classification can be used to predict gender in uncontrolled real time scenarios such as railway stations, banks, bus stops, airports, etc. For example, depending upon the number of male and female passengers on the railway station, restrooms can be constructed to ease the travelling.

ACKNOWLEDGEMENT

Foremost, we would like to express our sincere gratitude to our guide, Mr. Abhishek Gangwar, Principal Technical Officer, CDAC, for his valuable guidance and continuous support during the project; his patience, motivation, enthusiasm, and immense knowledge. His direction and mentoring helped us to work successfully on the project topic.

We express our sincere gratitude to Dr. Sanjay Pawar, Principal (Usha Mittal Institute of Technology) for his valuable encouragement and insightful comments.

We would also like to thank all the teaching and non-teaching staff for their valuable support. Lastly, we thank our parents and friends for their unconditional and continuous support.

REFERENCES

[1] E. Makinen, and R. Raisamo, “Evaluation of Gender Classification Methods with Automatically Detected and Aligned Faces,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 3, pp. 541547, 2008.
[2] S. U. Rehman, S. Tu, Y. Huang, and Z. Yang, Face recognition: A Novel Un-supervised Convolutional Neural Network Method, IEEE International Conference of Online Analysis and Computing Science (ICOACS), 2016.
[3] N. Srinivas, H. Atwal, D. C. Rose, G. Mahalingam, K. Ricanek, and D. S. Bolme, Age, Gender, and Fine-Grained Ethnicity Prediction Using Convolutional Neural Networks for the East Asian Face Dataset, 12th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2017), 2017.
[4] N. Jain, S. Kumar, A. Kumar, P. Shamsolmoali, and M. Zareapoor, Hybrid Deep Neural Networks for Face Emotion recognition, Pattern Recognition Letters, 2018.
[5] G. Levi, and T. Hassner, ”Age and Gender Classification Using Convolutional Neural Networks,” IEEE Workshop on Analysis and Modeling of Faces and Gestures (AMFG), IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Boston, 2015.
[6] S. Turabzadeh, H. Meng, R. M. Swash, M. Pleva, and J. Juhar, Realtime Emotional State Detection From Facial Expression On Embedded Devices, Seventh International Conference on Innovative Computing Technology (INTECH), 2017.
[7] A. Dehghan, E. G. Ortiz, G. Shu, and S. Z. Masood, Dager: Deep Age, Gender and Emotion Recognition Using Convolutional Neural Network, arXiv preprint arXiv:1702.04280, 2017.
[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet classication with deep convolutional neural networks, Communications of the ACM, vol. 60, no. 6, pp. 8490, 2017.