GENDER RECOGNITION AND AGE DETECTION USING HUMAN FACIAL FEATURES

Dr. Zankhana Shah[1], Jayraj Malamdi[2], Bhavya Dave[3], Shrey Bhatt[4], Dhruv Shah[5]

12345Information Technology Department,

Birla Vishvakarma Mahavidyalaya,

Anand, Gujarat, India

[1]zankhana.shah@bvmengineering.ac.in

[2]jayrajmalamdi31@gmail.com

[3]bhavyadave700@gmail.com

[4]sbbhatt010@gamil.com

[5]ds4365778@gmail.com

Abstract: One of the active study fields in recent years is the extraction of information from the human face. Numerous investigations on the identification of the most prevalent face variety, including age and gender, have been done. The approach for automatically determining a person's age and gender from their face is suggested in this paper and is based on convolutional neural networks and support vector machines. The four steps of this method are face detection, pre-processing, feature extraction, and classification. A variety of convolutional neural networks have been trained to detect faces in live feed and also determine their age and gender are part of our system. Our model, was trained using Adience benchmark dataset for age and gender prediction available on Kaggle, which has produced positive results.

Keywords: **Convolution neural network, Support Vector Machine, Face Detection, Multi-Task Cascaded Convolution Neural Networks, Principal Component Analysis, Adience benchmark dataset.**

1. INTRODUCTION

Numerous real-world applications, such as behavior research, online advertising, social understanding, identity verification, video surveillance, human computer interface, and many others, depend on age and gender information. Despite their many applications, it can be difficult to automatically determine a person's age and gender from facial pictures. The project's objective is to create a gender and age predictor that uses deep learning& machine learning concepts on the Adience Dataset to roughly guess the gender and age of the person (face) in live feed using a flask-made web application from live video feed.

The predicted gender may be one of ‘Male’ and ‘Female’, and the predicted age may be one of the following ranges-

(0– 3), (4 – 6), (8 – 13), (15 – 20), (25 – 32), (35 – 43), (45 – 53), (60 –150).

It is already known that predicting age and gender of a person by just seeing its face is a difficult task for even a human. There are numerous problems that we faced during this research work. First is illumination i.e., not adequate lighting in the space at time of input (video/image acquisition). Second are foreign objects (like glasses, masks, hair, etc.) in between the subject face and camera cause’s poor results [1]. Third is Frontal Face i.e., face rotations in and out of planes affects the results, hence, subject should face the camera directly instead of having a sideways gaze.

2. LITERATURE SURVEY

In the field of machine learning and computer vision, numerous studies and research work have been done on human age and gender detection. In this section, a brief overview of the work done by previous studies on age and gender detection is shown.

For Face Detection, we started by using the method proposed by Viola and Jones [2], which provided good performance in real time. However other research [3], demonstrate that when there are greater variances in human faces, this approach deteriorates dramatically in real-world applications. Inspired by CNN performance, several studies [4] proposed numerous CNN-based face detection approaches as a result of this. For face detection, we employed MTCNN [5], because it shows that performance has improved significantly when MTCNN is used, and it also performs better in real-time application.

For Gender Recognition, various techniques, such as facial photographs [6], hand images, and pose/body images [7], can be used to identify gender. By extracting two different sorts of characteristics, gender recognition is accomplished [8, 10]. The first type of feature extraction is geometric-based features, which describes the elements of the shape and position of the face using geometrical principles. It also discovers the locations of facial points. Second type of feature extraction is appearance-based features. For Age Detection, two different types of features are extracted from the facial photos [8, 9, 10]. (i) Wrinkle Features: The f5 properties are estimated. Additionally, as we age, facial wrinkles are increasingly obvious. (ii) Geometric Features: These characteristics are based on ratios of different face feature data (e.g., eyes, nose, mouth, chin, etc.).

Ref. [11] is one of the most popular research papers in age and gender classification using a neural network; it shows that usage of deep trained neural networks can dramatically increase performance. Also, Reference [11] suggested using deep CNN on the Adience dataset, which is effective for recognizing low resolution facial expressions. However, it faces challenges to identify the face image if there is some rotation or tilt.

ML algorithms are also used for age and gender detection, such as Support Vector Machine (SVM), Naïve Bayes algorithm, K-means clustering. K-means method is employed in reference [9] to identify age groupings. SVM is used for image classification in reference [12, 13, 14], and it provides good accuracy. PCA is used which is employed in reference [15] to accomplish dimensionality reduction in order to produce better and quicker results.

3. METHODOLOGY

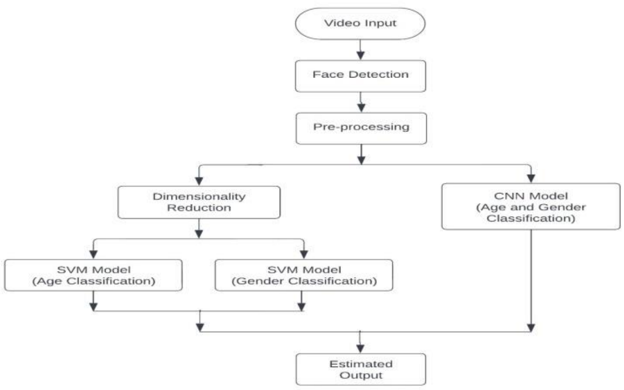

Fig. 1. Proposed Methodology

3.1. Data Acquisition

The Adience benchmark dataset is selected as our dataset for detecting the age and gender in images. The entire Adience collection includes 26,580 256×256 color facial images of 2,284 subjects, with eight age group classes:(0– 2), (4 – 6), (8 – 13), (15 – 20), (25 – 32), (38 – 43), (48 – 53), (60 –100).

To align faces, we employ the in-plane face aligned technique. A typical five-fold, subject-exclusive cross-validation methodology is used to test for age classification, and the results are averaged to get the final age group classification.

Fig.2. Sample Images from Adience Benchmark Dataset

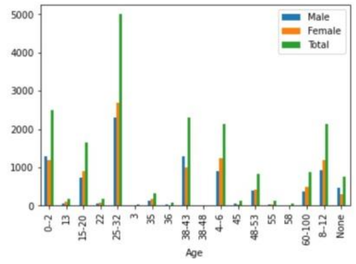

Fig.3. Combined Age and Gender Dataset

3.2. Face Detection

We utilized MTCNN for facial detection (multi-Task Cascaded CNN) [5]. It is capable of Landmark detection, which is the recognition of additional face features like the eyes and mouth. The network uses a cascade structure with the networks: The image is first rescaled to a variety of different sizes (referred to as an image pyramid). The first model (Proposal Network, or P-Net), which proposes candidate facial regions, is followed by the second model (Refine Network, or R-Net), which filters the bounding boxes, and the third model (Output Network, or O-Net), which proposes facial landmarks.

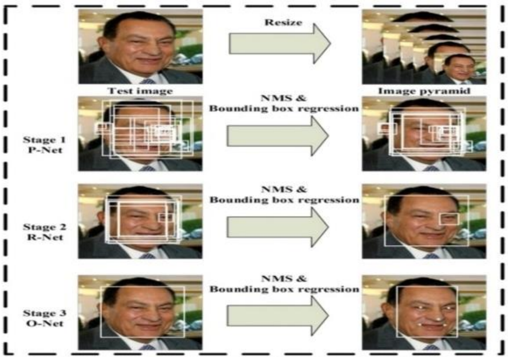

Fig.4. MTCNN process

The above image is taken from the paper [5], which provides a helpful summary of the three stages from top-to-bottom and the output of each stage left-to-right.

As each one of the three models in the cascade (P-Net, R-Net, and O-Net) is trained on three tasks, such as making three different sorts of predictions, the model is known as a multi-task network facial landmark localization, bounding box regression, and face categorization.

3.3. Pre-Processing

3.3.1 Data Augmentation

Data augmentation is the technique of creating additional data points from current data in order to artificially increase the amount of data. By creating fresh and varied examples for training datasets, augmented data is enhancing the efficiency and outcomes of deep learning models.

3.3.2 Data Normalization

The technique of projecting picture data pixels (intensity) to a preset range, such as [0, 1], is also known as data re-scaling**.** Here, we are normalizing our RGB images which are of data type uint8 and are in the range of [0, 255]. The new range will be [0, 1].

3.3.3 Dimensionality Reduction

Since Scikit-learn classifiers only accept training features in the form of 2D arrays, we are unable to directly feed RGB photos into any scikit-learn classifier. To solve this problem, we must flatten or restructure the 3D RGB images into a 1D array. However, a 256x256x3 RGB image can be flattened to create an array of the shape 1x196608 that has several properties.

We have performed dimensionality reduction technique by applying PCA (Principal Component Analysis)[17] in SVM model [12, 13]. Here in our project, we have taken the first 500 components which cover up to 90% information when summed up the individual component's variance. This means we have successfully reduced 196608 features to just 500 features by losing only 10% of the information from the dataset.

3.4 Model Building

After the preprocessing is done, the dataset is split for train and test purposes. Test size is 20% of the train size. We have created two models, first is using SVM algorithm and second is by using CNN model.

SVM Model

The SVM algorithm's objective is to establish the best line or decision boundary that can divide n-dimensional space into classes, allowing us to quickly classify fresh data points in the future. A hyper plane is the name given to this optimal decision boundary.

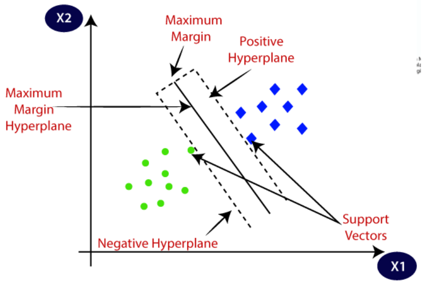

Fig.5. SVM

Now that the dataset has been generated, it's time to train many classifiers before selecting the best one. Hyperparameter tuning will be used during classifier training.

Hyperparameter tuning is done using the Scikit-learn package referred to as GridSearchCV. We use GridSearchCV to automate the tuning of hyperparameters because doing it manually could take a lot of time and resources.

In our project we have created two different models, one for age and other for gender. Because SVM doesn't support multiclass classification natively. It supports binary classification.

CNN Model

A Convolutional Neural Network (ConvNet/CNN) [11] is a Deep Learning method that can take in an input image, give various elements and objects in the image importance (learnable weights and biases), and be able to distinguish between them. The structure of a ConvNet is similar to the connectivity pattern of neurons in the human brain and was modeled after how the visual cortex is organized.

Through the use of pertinent filters, a ConvNet can effectively capture the spatial and temporal dependencies in a picture. Because there are fewer parameters used and the weights can be reused, the architecture does a better job of fitting the picture dataset.

CNN is specifically designed to process input images. Their architecture is then more specific: it is composed of two main blocks:

The first block uses convolution filtering to do template matching. A number of convolution kernels in the first layer filter the image, producing "feature maps" that are subsequently normalized (using an activation function) and/or shrunk.

The second block occurs at the end of all classification neural networks and is not a hallmark of a CNN. To produce a new vector at the output, the input vector values are converted (using a number of linear combinations and activation functions).

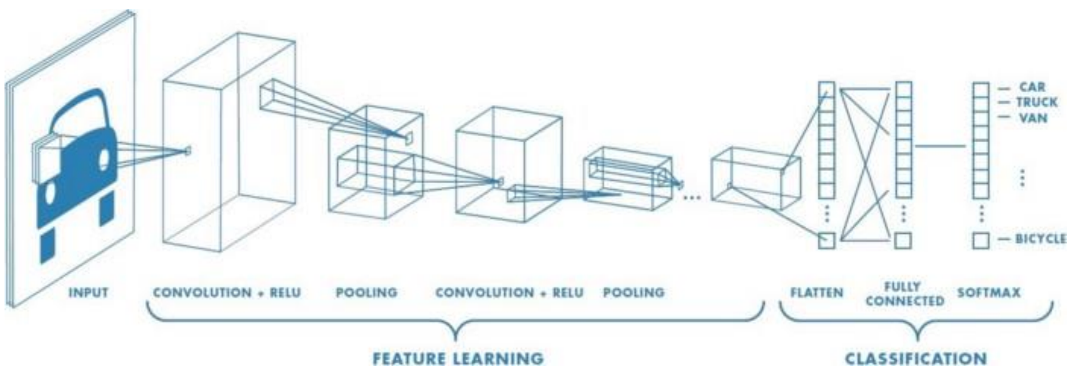

Fig.6. CNN

CNN layers

There are three types of layers that make up the CNN which are the convolutional layers, pooling layers, and fully-connected (FC) layers.

- **Convolutional Layer** - The first layer utilised to extract the different features from the input photos is this one.

- **Pooling Layer** - A Pooling Layer often comes after a Convolutional Layer. This layer's main goal is to lower the convolved feature map's size in order to save on computational expenses.

- **Fully Connected Layer** - The Fully Connected (FC) layer, which connects the neurons between two layers, is made up of the weights and biases as well as the neurons. This process flattens the input image from the preceding layers and feeds it to the FC layer.

4. RESULT ANALYSIS

The testing accuracy and validation accuracy of our model are checked, model loss is noted, and the confusion matrix is plotted. The misclassified images count of each type is also determined.

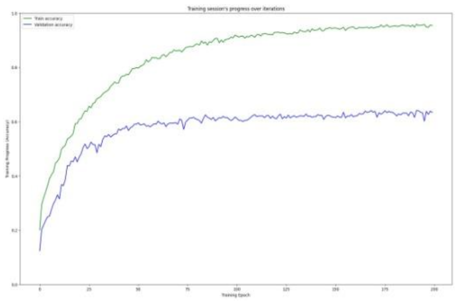

Fig.7. Training Accuracy v/s Validation Accuracy for age

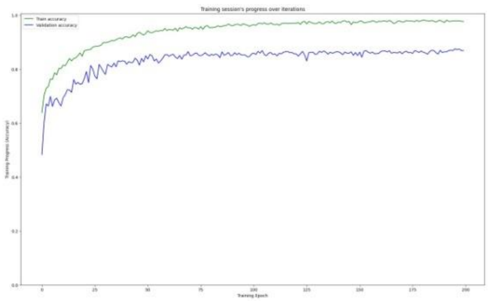

Fig.8. Training Accuracy v/s Validation Accuracy for Gender

CONFUSION MATRIX:

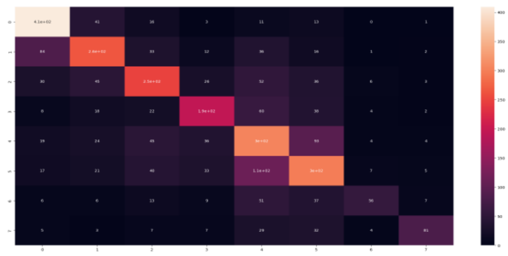

Fig.9. Confusion Matrix of SVM age

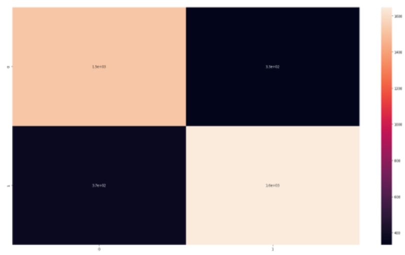

Fig.10. Confusion Matrix of SVM gender

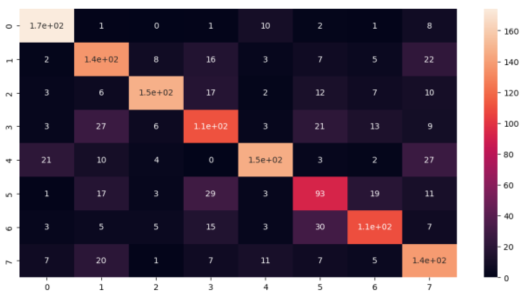

Fig.11. Confusion Matrix of CNN age

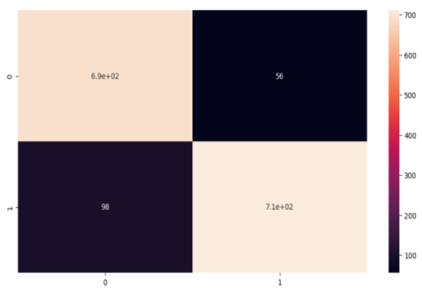

Fig.12. Confusion Matrix of CNN gender

5. CONCLUSION

The Result obtained in our project is:
Accuracy for CNN model for Gender is 90.10% and for Age it is 67.80%. whereas, accuracy for SVM model for Gender is 81.74% and for Age it is 60%.
Thus, two important conclusions can be made from our results:

- First, CNN can be used to provide improved age and gender classification results, even considering the much smaller size of contemporary unconstrained image sets labeled for age and gender.
- Second, the simplicity of our model implies that more elaborate systems using more training data may well be capable of substantially improving results beyond those reported here.

REFERENCES

[1] Miss Vaishnavi Mali and Dr. Babasaheb G. Patil, "Human gender, classification using machine learning",International Journal of Engineering Research and Technology (IJERT),2020,doi: 10.17577/IJERTV8IS120228

[2] P. Viola and M. J. Jones, "Robust real-time face detection",International Journal of Computer Vision,2004,pp.137-154, doi: 10.1023/B:VISI.0000013087.49260.fb

[3] B. Yang, J. Yan, Z. Lei, and S. Z. Li, "Aggregate channel features for multi-view face detection", IEEE International Joint Conference on Biometrics, 2014, pp. 1-8., doi:10.48550/arXiv.1407.4023

[4] Danai Triantafyllidou and Anastasios Tefas, "Face detection based on deep convolutional neural networks exploiting incremental facial part learning", in 2016 23rd International Conference on Pattern Recognition (ICPR),2016, doi:10.1109/ICPR.2016.7900186

[5] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, Senior Member, IEEE, and Yu Qiao, Senior Member, IEEE, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks", 2016, IEEE Signal Processing Letters,pp. 1499 – 1503, doi: 10.1109/LSP.2016.2603342

[6] TahminaAkter Sumi ,Mohammad Shahadat Hossain,Islam, Raihan UI Islam and Karl Andersson, "Human Gender Detection from Facial Images Using Convolution Neural Network",in Applied Intelligence and Informatics: First International Conference (Nottingham, UK),2021, pp. 188-203, doi: 10.1007/978-3-030-82269-9_15

[7] Liangliang Cao and MertDikmen, "Gender recognition from body", in Proceedings of the 16th International Conference on Multimedia 2008(Vancouver, British Columbia, Canada, October 26-31, 2008),2008,pp. 725–728, doi: 10.1145/1459359.1459470

[8] Prajakta A. Mélange and Dr. G. S. Sable, "Age Group Estimation and Gender Recognition Using FaceFeatures",The International Journal of Engineering and Science (IJES),2018, pp. 01-07,doi:10.9790/1813-0707010107

[9] Rahul Kumar, Akhilesh Kumar Srivastava and Deepak Kumar Agarwal, "Estimation Of Age Group based on Facial Features", 2018, International Journal of Engineering Research & Technology (IJERT),pp. 115-121, doi: 10.17577/IJERTV7IS070012

[10] Ramesha K., K B Raja ,Venugopal K R and Lalit M Patnaik, "Feature Extraction based Face Recognition, Gender and Age Classification", 2010, International Journal of Advanced Trends in Computer Science and Engineering, 2010, pp.14-23,

[11]Gil Levi and Tal Hassner, "Age and Gender Classification using Convolutional Neural Networks", The Open University of Israel, 2015,doi:10.1109/CVPRW.2015.7301352

[12] Dr.c.k.gomathy, a.lokesh, ch.harshavardhanreddy and a.sai kiran, "AGE AND GENDER DETECTION", International Journal of Scientific Research in Engineering and Management(IJSREM),2021

[13] Miss Vaishnavi mali and Dr.Babasaheb G. Patil, "Human gender, classificationusing machine learning",2020, International Journal of Engineering Research and Technology (IJERT), 2019, doi: 10.17577/IJERTV8IS120228

[14]Sonu Dhall and Poonam Sethi, "Geometric and appearance feature analysis for facial expression recognition",International Journal of Advanced Engineering Technology, 2014, pp. 01-11

[15]Sidharth Mishra,UttamSarkar,SubhashTaraphder and Sanjoy Datta, "Principal Component Analysis",International Journal of Livestock Research,2017, doi:10.5455/ijlr.20170415115235

Figures

[16] Figure 5: https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm

[17] Figure 6: https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148