International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-3, February 2019

Design Framework for Facial Gender Recognition Using MCNN

Sangita Choudhary, Manisha Agarwal, Manisha Jailia

Abstract*: *Facial Gender Recognition that allows automatic identification of gender from facial images, plays an important role in various applications. Even though it's a challenging task, it has gained immense popularity recently, especially with the development and popularity of social platforms and social media. The main aim of this paper is to use the proposed framework to classify the facial images based on their gender. The proposed framework uses a modified form deep convolution neural network (CNN), to obtain greater performance and accuracy. This frame can be used even for processing huge quantity of data. Hence by combining both modified deep convolution neural network and KNN-classifier we have created an application that can classify gender accurately. The rate of accuracy can be increased by increasing the number of layers and simultaneously training the images using back propagation. The parallel processing concept can be enhanced using this framework.

Index Terms*: *convolution neural network (CNN), deep convolutional neural network, deep learning.

I. INTRODUCTION

The human face plays an important role in transmitting visual information from one person to another. The diversity of information provided by the face enables us simultaneously to recognize another person, identify his gender and ethnicity, and estimate his age and emotional state. a wide range of scientists have been interested in studying the characteristics of the human face. It has been studied for a decade by psychologists. Furthermore, it has become a widely active research area for computer vision and machine learning communities. Face analysis refers to a set of tasks that could be used for solving different problems related to facial recognition, classification, and detection, that humans perform easily in their daily lives. For many decades, Facial-related analysis has been one the most important researches, in the fields of computer vision. It can be used in various places and applications like biometric authentication, security systems, multimedia management, and advanced Human-computer interaction. Studies in two-dimensional (2D) images-based face recognition gained significant interest in the computer vision community. Nevertheless, they are still limited to variations in illumination conditions, occlusions, and facial expressions. High accuracies have been achieved through several algorithms that were created during the last decade. However, these approaches perform adequately only when the face is frontal and normalized. Recently, facial data was captured in unconstrained environments and algorithms were developed to tackle this problem. Some approaches have used recent machine-learning algorithms to solve this problem. But still, the general problem of recognizing faces under certain unconstrained conditions remains a failure under illumination, and pose conditions. to provide a solution to this, we propose a face classification approach, based on the isolation of face parts. We conduct a series of experiments on the fusion of the texture and depth features. Furthermore, we employ the fusion of face parts decision to determine the overall decision rendered by the individual parts. Then inorder to process data's of various quantities, either small or large effectively we implement convolution neural network architecture. Hence by combining both modified deep convolution neural network and KNN-classifier we have created and implemented a gender classification too, whose accuracy can be increased increasing the number of layers and training the images using back propagation and eventually increase the quality of parallel processing concept. The contents of the remaining paper are organized into 6 sections. Section I is introduction. Section II gives a description about the applications variable in other domains, Section III give a detailed information on the overall work related to the research, Section IV describes the approach we proposed, Section V provides the result of the experiments and compares them with the baseline methods and finally the conclusion is given in the Section VI.

II. RELATED WORK

Yoo, B et al[1] the main aim is to design and construct a deep learning structure that can identify both, age and gender from the data provided. This can be done by using certain conditional age estimation along with gender inferences. By using conditional multitask learning (CMT) model that has gender-conditioned age probability we can build our deep learning model. Also inorder to solve the problem of data security and data scarcity this model develops an semi-supervised labeling strategy. This is first ever attempt made, that combines both age and gender classification within a single architecture with label expansion. Even though there were previous methods to find gender and age, none were in the combined form like this and this is way more effective compared to them.

Manuscript published on 28 February 2019.
* Correspondence Author (s)
Sangita Choudhary*, Research Scholar, Banasthali Vidyapith, Banasthali , Jaipur ,India.
Dr. Manisha Agarwal, Associate professor, Banasthali Vidyapith, Banasthali , Jaipur ,India..
Dr. Manisha Jailia, Associate professor, Banasthali Vidyapith, Banasthali , Jaipur ,India..

Zhang, K., Tan et al[2] proposed strategies exploit the inherent correlation between face identity, smile, gender and other face attributes to relieve the problem of over-fitting on the small training set aimprove the performance of the classification. They also propose the tasks-aware face cropping scheme to extract attribute-specific regions. The effectiveness of our proposed methods is easily demonstrated by the experimental results on the ChaLearn 16 FotW dataset which was specially designed for gender and smile classification.

Haseena, S. et al[3] in this section the main focus is to increase the accuracy of the deep convolution neural network architecture by increasing the number of layers in the system. Then with the help of dlib package the image is preprocessed. The package detects all the 68 facial landmarks, and finally produces a frontal face image. Finally using the dataset obtained from Labelled Faces in the Wild (LFW) we can test the network. Liu, X. [4] the Classical LeNet-5 model is used along with convolutional neural network to recognize and classify images and handwritten digits. Even though it has a simple and basic architecture, due to development of structural optimization, the convolutional neural network has expanded it field of application. Deep Belief Network, Convolutional Deep Belief Network, Convolutional Deep Belief Network and Fully Convolutional Network are some of the models that have originated with time and conditions. Recently the technique of transfer learning is also used with convolutional neural network for further development.

Gupta, N., et al[5] here the similar feature between two images in a specific attribute is calculated. Instead of considering the output of the final layer alone, here we calculate the weighted combination of all the features vectors from all the hidden and non hidden layers of CNN. Each hidden layer consists of attribute specific weight, which when minimized we can calculate the triplet loss criteria on labeled images. Hence the scoring mechanism adapts itself to each and every attributes and domains. It also exploits the composite face recognition model which is trained on a large number images.

Singh, M et al[6] this research by using the iris images classifies both ethnicity and gender. It uses the proposed Deep Class-Encoder, with robust feature extraction capabilities. It is used for deeply learning and discriminate the techniques of the supervised models. This research is divided into two steps: initially an supervised model called Deep Class Encoder is presented. The model is based on auto encoder. The supervision is done by learning weights that has its hidden representation mapped into the class label. The proposed model by using the Alternating Direction Method of Multipliers, reduces the training time to a greater extent.

Hyun, C. et al[7] have showed the possibility to improve the performance and recognize facial expressions. This was done by developing a multi-task learning system for identifying diverse facial features. In addition this , this research also concentrates on recognizing and learning six different facial features like, identity, expression, gender, race, age, and pose. It is done by initially learning and comparing the features with one another. It involves a multi-task learning of different combinations of the features. Based on these computed experiments, we can find a perfect suitable method to recognize these varying facial attributes or features.

Zibrek, K., et al[8] the main aim of his research is to use a real time virtualistic character to record both the body movements and facial emotions with it. With the data collected from these experiments he aims to study different gender features for different emotions. Even though the above method can be used to identify the different genders from body and facial features, it cannot distinguish how the gender judgments can be different for each and every single emotional expressions by using the data collected from both face and body of the virtual person. A Man and a Woman model are used respectively to record the motions of male and female body. And they also have explored how the basic visual appearance of the character can affect the gender quality, especially since previous work shows that the choice of the model affects the perception of motion.

Jeon, J., et al[9] CNN having received the best performance on Kaggle facial expression recognition challenge, is used here for facial expression recognition task. The SoftMax layer is replaced with Linear-SVM with the help of CNN model in the classification step. His model provided an accuracy of 69.77% which was the best. Bergstra used Null model, which consists of linear readout, pooling, convolution, normalization and Principal Component Analysis (PCA). His model mainly concentrated on optimizing the hyperparameters.

III. PROPOSED METHODOLOGY

For several decades, this facial age and gender recognition have been studied and researched upon continuously to be used in several softwares. The main applications that use these methods involve human-computer interaction, soft biometrics, and audience measurement systems. Recently, deep convolutional neural networks (CNN) have developed and gained the ability to recognize these facial attributes. But, we still are not able to estimate the accurate facial and physical age and appearance of a person might be very different from the estimated recorded age. Also the facial characteristics will vary with gender. The convolutional neural network with the help of its three characteristic like, observation, sharing weights, and locally connecting and pooling the sampling can be used to reduce the quantity of weights, decrease the complexity and increase the robustness to zoom, provides rotation and translation. The CNN is an artificial neural network consisting of many convolutional and pooling layer. It supplies an end to end model which learns the features to extract and classify the image by making use of the stochastic gradient descent algorithm. Features of each layer are extracted from a local region of the last layer by sharing weights. The convolutional neural network is the best application to learn and express image features. The Classical LeNet-5 model is used along with convolutional neural network to recognize and classify images and handwritten digits. Even though it has a simple and basic architecture, due to development of structural optimization, the convolutional neural network has expanded it field of application.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Deep Belief Network, Convolutional Deep Belief Network, Convolutional Deep Belief Network and Fully Convolutional Network are some of the models that have originated with time and conditions. Recently the technique of transfer learning is also used with convolutional neural network for further development.

CLASSIFICATION FRAMEWORK

The main aim of gender classification system is to classify the given image into Female and Male. The CNN architecture is used to here to classify large quantity of data, with increased speed and accuracy. The proposed gender classification architecture (see fig.1) consists of three main steps.

- Pre-processing module, using 68 facial landmarks, detects the frontal face image.
- Feature Extraction module using the convolution neural network architecture extracts the features from the image.
- Classification module with the help of extracted features classifies the images of face into female and male.

MCNN BASED RECOGNITION

The aim of this process is to provide a better way to learn the new attributes and similarities by using as little training samples as possible. Here it does not concentrate on face matching algorithms or performance of the basic models.

Our approach helps in two ways by:

(a) Reducing both overhead of collection and manual annotation of samples for model training;

(b) Learn some specific attributes along with the inherent properties of the parent model.

This method provides a greater advantage of maintaining the basic properties of the parent model and its specific attributes. For example, if we use face similarity model as a parent model inorder to learn the attributes like gender, then the new learned gender model not only places the same gender images in the top, but also place the other similar looking images in the top too. The major contribution of the proposed work is:

(a) A similarity learning approach has been provided by the novel attribute;

(b) Instead of creating individual models for every single attribute using a large quantity of samples, we can learn attributes by extracting and learning hidden layers present in the parent model and using minimal training samples.

(c) Creation of new complex dataset for the social media. This helps us to identify and extract data from the hidden layers of the parent model to learn new attribute specific features (weighted hidden layers). These extracted data and attribute features are used to construct a perfect attribute scoring system for image ranking and several other tasks.

a) CNN model:

Our CNN model consists of carpooling layers and two fully-connected layers which joined together forms the convolution layer. It was later implemented by using public C++ deep learning library Caffe [9].

b) Training phase

During this phase, the faulty images, especially the images of all the black pixels were removed in the first step. Then more data's were augmented and trained into our CNN models. The 42px×42px randomly cropped images were used to train the inputted dataset. As a result this method generates 8 times more data for training and it also induces the spatial variance. The images are shuffled randomly before the training. The model uses stochastic gradient descent method along with back-propagation to provide training with minimum loss function. To avoid over fitting of the fully connected layers we apply technique called dropout.

c) Testing phase

Here in this phase we use an averaging method to reduce the outliers (see Figure 5). The images cropped on four corners along with their mirrored images are used for data testing and calculate the average of the probability to produce a final result output. The classification error is comparatively reduced in this method.

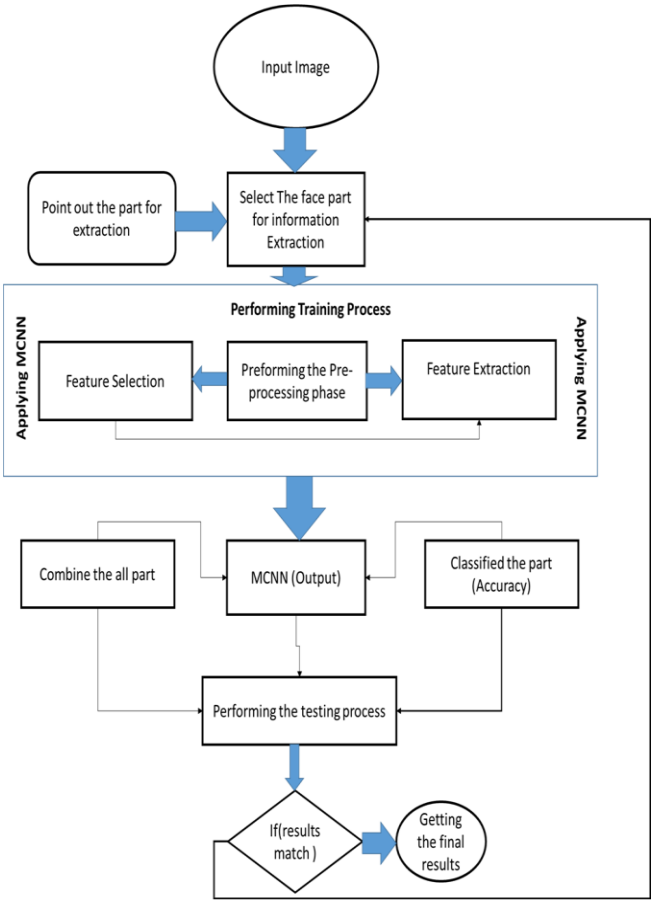

Figure 1: proposed Framework

IV. RESULTS ANALYSIS

In order to study the gender and emotion features on the virtually characters and determine the effect on perceiving these images from these characters, an experiment is conducted. In our initial findings our study on emotion and effects provided nullified gender information but not overridden ones. For example, when the male emotion of anger was enacted by a female actor the result was neutral instead of male or female. Also the female emotions like sad and fear, when expressed by male actors, it was recognized as ambiguous. The fear and happy emotions expressed by both male and female actors when emphasised, produced a gender that were identified correctly.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Trials were also carried out to find the effect on appearance found that the appearance of the actors didn't emphasise the gender in a concordant pair (e.g. male motion applied to the Man model), neither did it changed the gender recognition in terms of in-congruous pairs. But still our findings did include that the accuracy depended on the model especially when they convey it. The sad and the fear emotions were more easily identified on the Man model when compared to the female model. This result shows that the face of the model and his acting will affect the perception of emotion directly. Like if the eyebrows of the model is lower or looks neutral it would directly affect the perception of the emotion. So while selecting an model it should be made sure that all the features fit them exactly and the emotions expressed by them are very accurate without mixing their personal emotions. The difference could also be due the result of motion mapping, skinning, or retargeting effects. So the experiment should be conducted on a variety of models in the future to make sure the emotions and variations are due to models or gender difference. In our research, the partakers were asked to value the motion only and ignore the appearance. But still we doubt that the model has unknowingly altered the emotions with his personal unique emotions. But unfortunately this not to be the case and hence proposed that in-congruous pairs should be circumvented as much as possible, to avoid ambiguity or unnatural results. One of the disadvantages in our study is that we did not emphasise on the controlling of emotions and strength of our emotions during our creation of the stimuli. Since we felt that focussing on the intensity of the emotion would destroy its natural occurrence.(e.g., Anger when expressed in low intensity is as bad as sadness exhibited with high intensity). But still we made the contributor to perform it in their point of view to explore the practicability and hoping higher intensity will provide strong gender indications. Our results prove that intensity has no hand in perceiving the gender. It is based on emotions alone. Further studies will investigate the correlation between emotion recognition, intensity and gender perception. Our model initially detects a human face and then its expression by using two different steps: face detection & tracking and CNN based recognition. Provides the evaluated comparative accuracy of the proposed algorithm and traditional algorithm. For calculating the accuracy, Recall, Precision, FMeasure to compute the performance which are defined in the following formula is used.

$$Accuracy = \frac{Number of Correct Detection \times 100}{Total Number of Test Images}$$

$$\text{Recall}_i = {Q_{ii}}/{\sum_j Q_{ij}}$$

$$\text{Precision}_i = {Q_{ii}}/{\sum_j Q_{ji}}$$

$$F - measure_i = \frac{2 \times Recall_i \times Precision_i}{Recall_i + Precision_i}$$

here, Qii = Quantity of accurate detection of class i as class i; Qij = Complete quantity of detection of class i as class i and j; Mji = Quantity of detection of class j detected as class i and j. F-measure used for determining the regression performances which is the harmonic mean of Precision and Recall value. Huge value of F-measure designates advanced accuracy.

Table 1: Percentage of accuracy dummy datasets in numerous situations.

Feature extraction	Accuracy (%)		
	Dummy data1	Dummy data2	Dummy data3
Proposed MCNN	95.3	80.80	98.25
Traditional CNN	94.63	81.80	97.1

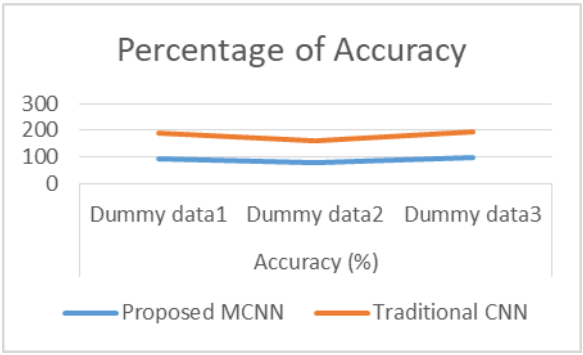

Figure 2: Percentage of accuracy dummy datasets in numerous situations

The HOG feature descriptor along with a linear classifier is used for face detection. This type of detector can be used for both human face and semi-rigid objects. Since localizing a face using face detector consumes more time, we can use the face detector in the first frame and the remaining frames can be detected using correlation tracker[6]. Re-initialisation is done to prevent tracking out of sight objects and compensate tracking errors. (see the left side). After this the facial part is cropped into 42px×42px the ideal dimension to input into the CNN model. The features are then extracted and classified by CNN model. Then we get the results of the real time facial expression (see the right side of Figure 1 and Figure 2).

1) details.

V. CONCLUSION

We have also performed a detailed study and analysis of CNN architecture for the purpose of gender classification which is an optimized one. The accuracy can be obtained by increasing the quantity of layers and image training using back propagation. Instead of convolution, the architecture comprises of combined convolution and sub sampling layers along with cross-correlation that is implemented in the processing layers. In addition, the architecture has a significant speedup in its study and analysis, that enables the processing and classification of a 32x32 pixel input image on the PC platform in less than 0.3ms. Based on the rate of classification and speed in which it is processing of CNN, this is the first method have studied up on the effects of weight flipping on the performance of CNN, to the best of our knowledge. So to produce complete and perfect gender recognition system, this work can be used and implemented for both face detection and analysis tasks by using similar CNN architectures.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

This system is suitable in resource-constrained environments for implementing custom hardware that is targeted for real-time processing. Having sufficient technology to distinguish and recognize genders with accuracy from various faces is one of the major advantages of the proposed method.

REFERENCES

1. B., Kwak, Y., Kim, Y., Choi, C., & Kim, J. (2018). Deep Facial Age Estimation Using Conditional Multitask Learning With Weak Label Expansion. IEEE Signal Processing Letters, 25(6), 808–812. doi:10.1109/lsp.2018.2822241.
2. Zhang, K., Tan, L., Li, Z., & Qiao, Y. (2016). Gender and Smile Classification Using Deep Convolutional Neural Networks. 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). doi:10.1109/cvprw.2016.97
3. Haseena, S., Bharathi, S., Padmapriya, I., & Lekhaa, R. (2018). Deep Learning Based Approach for Gender Classification. 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA). doi:10.1109/iceca.2018.8474919.
4. Liu, X., Li, J., Hu, C., & Pan, J.-S. (2017). Deep convolutional neural networks-based age and gender classification with facial images. 2017 First International Conference on Electronics Instrumentation & Information Systems (EIIS). doi:10.1109/eiis.2017.8298719.
5. Gupta, N., Gupta, A., Joshi, V., Subramaniam, L. V., & Mehta, S. (2017). Deep Attribute Driven Image Similarity Learning Using Limited Data. 2017 IEEE International Symposium on Multimedia (ISM). doi:10.1109/ism.2017.28.
6. Singh, M., Nagpal, S., Vatsa, M., Singh, R., Noore, A., & Majumdar, A. (2017). Gender and ethnicity classification of Iris images using deep class-encoder. 2017 IEEE International Joint Conference on Biometrics (IJCB). doi:10.1109/btas.2017.8272755
7. Hyun, C., & Park, H. (2017). Recognition of Facial Attributes Using Multi-Task Learning of Deep Networks. Proceedings of the 9th International Conference on Machine Learning and Computing - ICMLC 2017. doi:10.1145/3055635.3056618.
8. Zibrek, K., Hoyet, L., Ruhland, K., & McDonnell, R. (2013). Evaluating the effect of emotion on gender recognition in virtual humans. Proceedings of the ACM Symposium on Applied Perception - SAP '13. doi:10.1145/2492494.2492510
9. Jeon, J., Park, J.-C., Jo, Y., Nam, C., Bae, K.-H., Hwang, Y., & Kim, D.-S. (2016). A Real-time Facial Expression Recognizer using Deep Neural Network. Proceedings of the 10th International Conference on Ubiquitous Information Management and Communication - IMCOM '16. doi:10.1145/2857546.2857642.
10. Azzakhnini, S., Ballihi, L., & Aboutajdine, D. (2018). Combining Facial Parts For Learning Gender, Ethnicity, and Emotional State Based on RGB-D Information. ACM Transactions on Multimedia Computing, Communications, and Applications, 14(1s), 1–14. doi:10.1145/3152125.
11. Vinay, Gupta, S., & Mehra, A. (2014). Gender specific emotion recognition through speech signals. 2014 International Conference on Signal Processing and Integrated Networks (SPIN). doi:10.1109/spin.2014.6777050.
12. Zvarevashe, K., & Olugbara, O. O. (2018). Gender Voice Recognition Using Random Forest Recursive Feature Elimination with Gradient Boosting Machines. 2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD). doi:10.1109/icabcd.2018.8465466
13. Wang, Z.-Q., & Tashev, I. (2017). Learning utterance-level representations for speech emotion and age/gender recognition using deep neural networks. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). doi:10.1109/icassp.2017.7953138.
14. Azimi, M., & Pacut, A. (2018). The effect of gender-specific facial expressions on face recognition system's reliability. 2018 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR). doi:10.1109/aqtr.2018.8402705.
15. Yoon, W.-J., & Park, K.-S. (2011). Building robust emotion recognition system on heterogeneous speech databases. IEEE Transactions on Consumer Electronics, 57(2), 747–750. doi:10.1109/tce.2011.5955217.
16. Narain, B., Shah, P., & Nayak, M. (2017). Impact of emotions to analyze gender through speech. 2017 4th International Conference on Signal Processing, Computing and Control (ISPCC). doi:10.1109/ispcc.2017.8269645.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)