IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Relative Age And Gender Position Learning For Face Based Estimation

Lincy C T
Computer Science and Engineering
Mar Ephraem College Of Engineering and Technology
Marthandam, India

Arpitha Renjan
Computer Science and Engineering
Mar Ephraem College Of Engineering and Technology
Marthandam, India

***Abstract*—Automatic gender classification and age detection may be a fundamental task in computer vision which has recently attracted immense attention. Here, a new type of intelligent system which is based on several neural networks connected on two levels of classification with an efficient and simple Convolutional Neural Network (CNN) model architecture by considering gender and age estimation as a multi-label classification problem. The proposed model is trained and then evaluated on the publicly available audience bench mark dataset. The model first performs feature extraction on the input image which can classify eyes, lips, beard, hair etc supporting these featured the model will classify the gender as male or female. Here we have used the Haar Cascade for feature extraction purpose. Age is estimated with the assistance of Caffe Model. We have used various detection methods for analysis like skin colour Segmentation and lip detection and many more. The CNN model is built by using VGG-16 architecture. The CNN model is then trained using epochs, where each epoch contains a certain number of training images. To remove distorted and unwanted images, the loss Gauss function is used. For testing, the input image is given by the user. The model makes the predictions to estimate age, gender and emotion of that input image by comparing with the trained images.**

Keywords*—Convolutional Neural Network, Recurrent Neural Network, Support Vector Machines, Deep EXpectation, Generative Adversarial Network, Conformal Embedded Analysis

I. INTRODUCTION

Estimating the age and gender of a person is important in person to person communication since grammar rules are different for each gender and different vocabularies are used to represent different age groups. In addition, age and gender are considered as soft biometric [1] features and hence they are able to be used to identify an individual with the combination of other soft biometric features such as weight and height. After the emergence of computer vision and machine learning, automated computerized age and gender estimation systems became popular and are used by many applications. Police departments use automated age and gender estimation systems to efficiently identify the subjects and corpses [2]. Human computer interaction and marketing fields use them to identify the customer clusters and hence to post suitable advertisements [3]. Security access control systems use them to provide age and gender based security protocols.

Although age estimation studies such as have confirmed a direct correlation between age and gender, there has been insufficient research into the concept of estimating age based on the subject's gender. We attempt to solve the abovementioned issue by introducing a gender-specific age estimation system that consists of two models. Each model is based on the Visual Geometry Group (VGG16) architecture and trained on the UTKFace dataset. Model A is only trained on images of female subjects, while model B is only trained on images of male subjects. We use the letters "A" and "B" for labelling purposes. In addition, a custom-built gender estimation model is employed to detect the gender of the subject from an input image and produces a label that is then used to load the appropriate age model. We divide the images into four age classes: 0–12, 13–19, 20–59 and 60+ and test our implementation on a testing portion of the UTKFace dataset, which was not included in the training phase. The proposed system is also cross-validated on the FG-NET [10] dataset to confirm whether gender separation affects performance. Our results demonstrate an improvement in the classification accuracy when two separate models are trained compared to a single model. Our contributions are summarized as follows:

1) We propose a novel gender-based age classification system consisting of two age classifiers, where each model is trained on a specific gender group.

2) We introduce a robust custom-built facial gender classifier that produces a gender label responsible for loading the appropriate age model.

3) We propose two modified VGG16 networks to estimate age groups from input images.

Automated age and gender estimation approaches should have excellent prediction performance. Decisions of a machine learning model should be visually explained to identify the discriminative features in age and gender estimation. To achieve these objectives, we have developed a CNN based model architecture by treating gender and age estimation as a multi-label classification task. The proposed CNN model architecture is simple but powerful and is able to estimate the gender and age of an individual in real-time. The proposed approach showed excellent performance in publicly available benchmark datasets when compared to similar approaches.

II. RELATED WORK

Early approaches relied on hand-crafted soft-bio metric features in age and gender estimation. Hayashi et al., [16] have extracted the line segments of wrinkles in a face at the corner of eyes and then used them in age and gender estimation. Karimi and Tashk [17] have manually obtained the facial features such as positions of nose, mouth, eyes, and top of head and chin. Fukai et al., [18] have extracted multiple facial features such as skin colour, wrinkle, and pigmented spot and then combined them in gender and age estimation. Eidinger et al., [11] have developed a model by using local image features and SVM classifier. Since these approaches rely on handcrafted features and their locations, they are not capable of handling the appearance variances of facial images such as view changes, rotations, and occultations. Most of the recent approaches rely on Convolutional Neural Network (CNN) based model architectures, and they are using raw facial images in age and gender estimation. Liao et al., [9] have divided a facial image into fine patches and then fed each of them to a Neural Network (NN) architecture to obtain classification scores. Then they used the average classification score of these nine patches to recognize the gender and age group of a person. Similar to this, Zhang and Xu [10] have proposed another approach by extracting landmark patches in facial images and then used them to train a NN architecture. Levi and Hassner [12] have developed a simple but powerful CNN model architecture for gender and age estimation and it is able to train the architecture with a fewer number of samples. Recently, Abu Nada et al., [13] developed a CNN model to detect the faces first and then to detect the age and gender in an automated manner.

• Age and gender that are the two key facial attributes, play a foundational role in social interactions, making age and gender estimation from one face image a crucial task in intelligent applications, like access control, human-computer interaction, enforcement, marketing intelligence and visual surveillance.

• The basic aim of this project is to develop an algorithm that estimates age and gender of a person correctly.

• One of the most widely used techniques is Haar cascade.

• In this project we propose a model which can predict the gender of a person with the assistance of Haar Cascade.

• The model trained the classifier with different male and female images as positive and negative images.

• Different facial features are extracted.

• With the assistance of Haar Cascade classifier will determine whether the input image is male or female.

• We made use of Deep Convolution neural network

A. *Skin Colour Segmentation*

For skin colour segmentation, first we contrast the image. Then we perform skin colour segmentation. Then, we have to find the largest connected region. Then we have to check the probability to become a face of the largest connected region. If the largest connected region has the probability to become a face, then it will open a new form with the largest connected region. If the largest connected regions height & width is larger or equal than 50 and the ratio of height/width is between 1 and 2, then it may be face.
For face detection, first we convert binary image from RGB image. For converting binary image, we calculate the average value of RGB for each pixel and if the average value is below than 110, we replace it by black pixel and otherwise we replace it by white pixel. By this method, we get a binary image from RGB image. Then, we try to find the forehead from the binary image. We start scan from the middle of the image, then want to find a continuous white pixels after a continuous black pixel. Then we want to find the maximum width of the white pixel by searching vertical both left and right site. Then, if the new width is smaller half of the previous maximum width, then we break the scan because if we reach the eyebrow then this situation will arise. Then we cut the face from the starting position of the forehead and its high will be 1.5 multiply of its width. Then we will have an image which will contain only eyes, nose and lip. Then we will cut the RGB image according to the binary image.

B. *Eyebrow Corner Detection*

For eyes detection, we convert the RGB face to the binary face. Now, we consider the face width by W. We scan from the W/4 to (W-W/4) to find the middle position of the two eyes. The highest white continuous pixel along the height between the ranges is the middle position of the two eyes. Then we find the starting high or upper position of the two eyebrows by searching vertical. For left eye, we search w/8 to mid and for right eye we search mid to w – w/8. Here w is the width of the image and mid is the middle position of the two eyes. There may be some white pixels between the eyebrow and the eye. To make the eyebrow and eye connected, we place some continuous black pixels vertically from eyebrow to the eye. For left eye, the vertical black pixel-lines are placed in between mid/2 to mid/4 and for right eye the lines are in between mid-+(w-mid)/ 4 to mid- +3*(w-mid)/ 4 and height of the black pixel-lines are from the eyebrow starting height to (h- eyebrow starting position)/4. Here w is the width of the image and mid is the middle position of the two eyes and h is the height of the image. Then we find the lower position of the two eyes by searching black pixel vertically. For left eye, we search from the mid/4 to mid - mid/4 width. And for right eye, we search mid + (w-mid)/ 4 to mid-+3*(w- mid)/ 4 width from image lower end to starting position of the eyebrow. Then we find the right side of the left eye by searching black pixel horizontally from the mid position to the starting position of black pixels in between the upper position and lower position of the left eye. And left side for right eye we search mid to the starting position of black pixels in between the upper position and lower position of right eye. The left side of the left eye is the starting width of the image and the right side of the right eye is the ending width of the image. Then we cut the upper position, lower position, left side and the right side of the two eyes from the RGB image.

C. Lip Corner Detection

For lip detection, we determine the lip box. And we consider that lip must be inside the lip box. So, first we determine the distance between the forehead and eyes. Then we add the distance with the lower height of the eye to determine the upper height of the box which will contain the lip. Now, the starting point of the box will be the ¼ position of the left eye box and ending point will be the ¾ position of the right eye box. And the ending height of the box will be the lower end of the face image. So, this box will contain only lip and may some part of the nose. Then we will cut the RGB image according the box.

D. Detection using Deep Learning

The photo is first taken from the webcam stream live by the cv2 module. At present we have a technique that can instantiate a model with five different stages which are namely; width, height, depth, classes, and final activation. The measurement of the input image which is in pixels will be taken in the form of width and height. The shallowness which definably represents the number of channels, which will be 3 to handle our different image. Classes will be the number of categories we have in our dataset. In this scenario there will be 8 different situations for our types of vehicles and colour likelihood combined. In spite of the fact that the end activation would be by default set to "soft-max" if there is any modification then the value to "sigmoid" will immediately enable Keras to perform multi-label classification for our result. The prototype itself contains of many different stages which will be followed by alike pattern. The convolution layers which are in 2D tend to be activated and it double the filter size on every moment to allow for more and more abstraction. Dropout is used to randomly disconnect nodes leading to the next layer that will work to prevent over-fitting our model. The CNN model is built by using VGG-16 architecture. The CNN model is then trained using epochs, where each epoch contains a certain number of training images. To remove distorted and unwanted images, the loss Gauss function is used. For testing, the input image is given by the user. The model makes the predictions to estimate age, gender and emotion of that input image by comparing with the trained images.

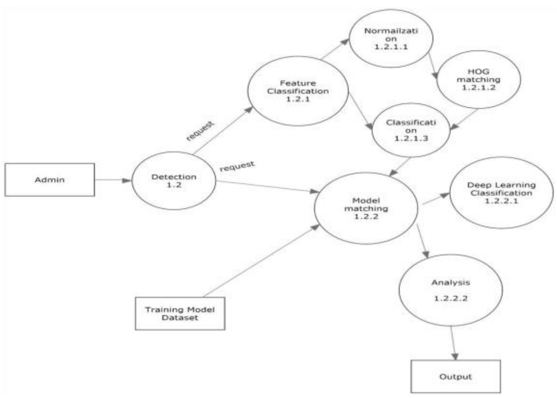

Fig 1: Data Flow Diagram for Detection

III. CNN Algorithm

1) *Machine learning algorithm* has to work with three different data (R-G-B values) to extract features of the images and classify them into their appropriate categories. (Grey level conversion)
2) *The convolution layer* consists of one or more Kernels with different weights that are used to extract features from the input image.
3) *Based on the weights of the Kernel calculating* features for different pixels based on their surrounding/neighbouring pixel values (Stride).
4) *When we use a stride value of 1 (Non-Strided)* operation we need 9 iterations to cover the entire image.
5) *The result is a feature map* that basically detects features from the images rather than looking into every single pixel value.
6) *Collect Image features*, such as edges and interest points from pre-processing stages
7) *When an image is passed through a convolution layer,* it and tries and identify the features by analysing the change in neighbouring pixel intensities.
8) *We can always add more than one convolution layer* when building the neural network, where the first Convolution Layer is responsible for capturing gradients whereas the second layer captures the edges.
9) *Rectified linear unit* is a process of applying an activation function to increase the non-linear of the network without affecting the receptive fields of convolution layers.
10) *The pooling layer* applies a non-linear down-sampling on the convolved feature maps.
11) *Max Pooling* ,that returns the maximum value from the portion of the image covered by the Pooling Kernel and the Average Pooling that averages the values covered by a Pooling Kernel.
12) *Once the training is done* the output converted to an object that can used to perform the classification.
13) *For detection*, matching can perform with the help of the Step12 object.

IV. RESULTS AND DISCUSSION

The method is utilized during the course of our experiments for both age and gender classification. Our approach is inspired by the research advancement in the computer vision fields, such as image classification [48,69,70], object detection [71], age estimation [23,24], and gender classification [23] fuelled by deep learning. At the very beginning of our age and gender estimation process, we ensure a class-wise balance of the image samples for training. To do this task, we down-sample those classes up to a specified threshold where the number of images is huge, then up-sample other classes by importing images from another dataset built on the same setup and perform data augmentation where necessary. Additionally, we manually filter out those images that seem wrongly annotated using human visual perception. In the later stages, before feature extraction, we detect the face from raw faces and prepare the images for network input performing some pre-processing tasks like rescaling and resizing. We use the same CNN structure for the feature extraction and the Soft-max layer for classification output regardless of the estimation task, but a further formulation is performed in case of age estimation.
We calculate the expected value over the Soft-max probabilities for age regression.

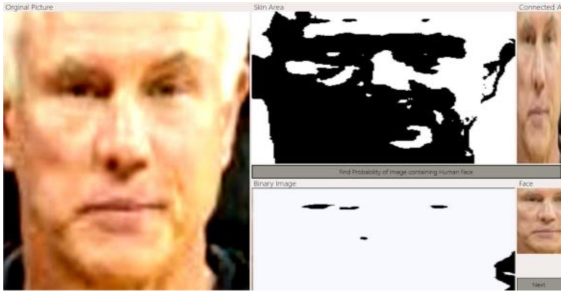

Fig 2: Analysing skin and conversion to Binary Image

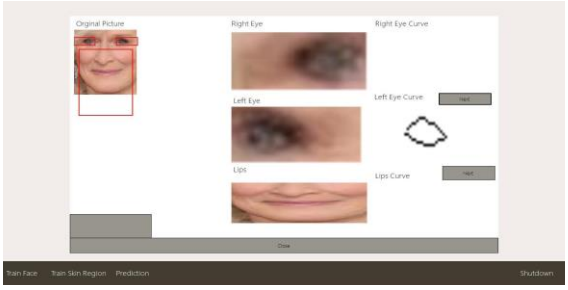

Fig 3: Analysing each skin area

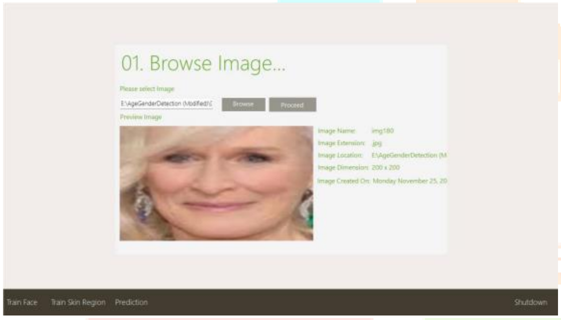

Fig 4: Training Of Skin Region

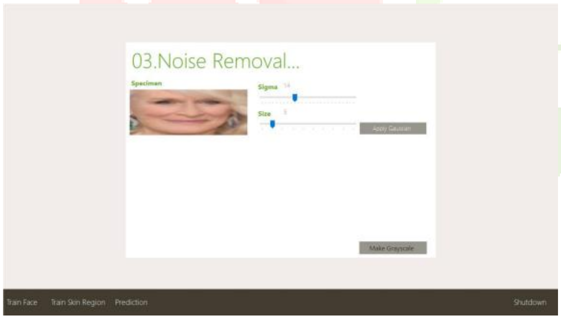

Fig 5: Making into grayscale

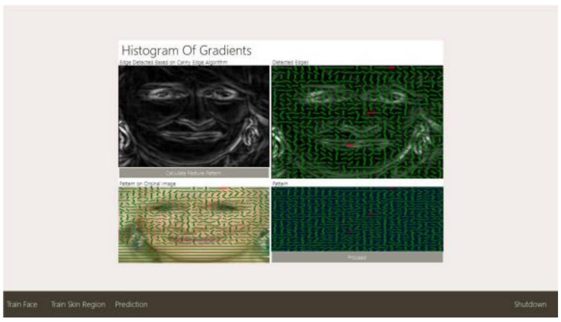

Fig 6: HOG visuals

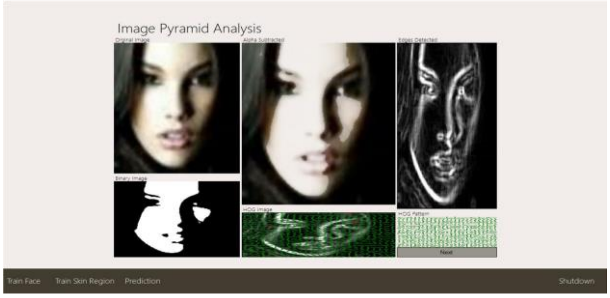

Fig 7: Predicting Age and Gender using HOG Visuals

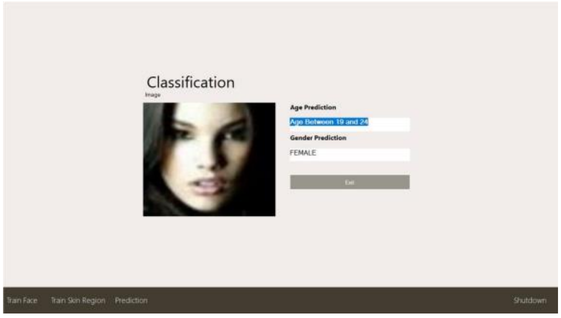

Fig 8: Prediction of Age and Gender

V. CONCLUSION

There are many previous methods that have addressed the problems of age and gender classification but recently much of this work has focused on constrained images taken in lab settings. These settings do not adequately reflect appearance variations which common to the real-world images in social websites and online repositories. Internet images are not simple but more challenging and also they are abundant. The image collections which are easily available they provide modern machine learning based systems with effectively endless training data, However this data is not always suitably labelled for supervised learning. While considering the example from the related problem of face recognition we explore how well deep CNN perform on these tasks using data.

REFERENCES

[1] S. Kumar et al., "Face Spoofing, Age, Gender and Facial Expression Recognition Using Advance Neural Network Architecture-Based Biometric System," Sensors,vol.22,no 14, p. 5160, 2022.
[2] T. Grubl and H. S. Lallie, "Applying Artificial Intelligence for Age Estimation in Digital Forensic Investigations," arXiv preprint arXiv:03045, 2022.
[3] K. S. Htet and M. M. Sein, "Effective Marketing Analysis on Gender and Age Classification with Hyperparameter Tuning," in IEEE 2nd Global Conference on Life Sciences and Technologies (LifeTech), pp. 247-248, 2020.
[4] S. K. Gupta and N. Nain, "Single attribute and multi attribute facial gender and age estimation," Multimedia Tools Applications, pp. 1-23, 2022.
[5] D. Kwasny and D. Hemmerling, "Gender and age estimation methods based on speech using deep neural networks," Sensors, vol. 21, no. 14, p. 4785, 2021.
[6] T. U. Islam, L. K. Awasthi, and U. Garg, "Gender and age estimation from gait: A review," in International Conference on Innovative Computing and Communications, pp. 947-962, 2021.
[7] N. Dwivedi and D. K. Singh, "Review of deep learning techniques for gender classification in images," in Harmony

Search and Nature Inspired Optimization Algorithms: Springer, 2019, pp. 1089-1099.

8] M. G. Rhodes, "Age estimation of faces: A review," Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory Cognition, vol. 23,no. 1, pp. 1-12, 2009.

[9] Z. Liao, S. Petridis, and M. Pantic, "Local deep neural networks for age and gender classification," arXiv preprint arXiv:08497, 2017.

[10] Y. Zhang and T. Xu, "Landmark-guided local deep neural networks for age and gender classification," Journal of Sensors, vol. 2018, 2018.

[11] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," IEEE Transactions on information forensics security, vol. 9, no. 12, pp. 2170- 2179, 2014.

[12] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 34-42, 2015.

[13] A. M. Abu Nada, E. Alajrami, A. A. Al-Saqqa, and S. S. Abu-Naser, "Age and Gender Prediction and Validation Through Single User Images Using CNN," International Journal of Academic Engineering Research, vol. 4, no. 8, pp. 21-24, 2020.

[14] Z. Khan and Y. Fu, "One label, one billion faces: Usage and consistency of racial categories in computer vision," in Proceedings of the 2021 acm conference on fairness, accountability, and transparency, pp. 587-597, 2021.

[15] R. Angulu, J. R. Tapamo, and A. O. Adewumi, "Age estimation via face images: a survey," EURASIP Journal on Image Video Processing, vol. 2018, no. 1, pp. 1-35, 2018.

[16] J.-I. Hayashi, H. Koshimizu, and S. Hata, "Age and gender estimation based on facial image analysis," in International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, pp. 863-869, 2003.

[17] V. Karimi and A. Tashk, "Age and gender estimation by using hybrid facial features," in 20th Telecommunications Forum (TELFOR), pp. 1725-1728, 2012.

[18] H. Fukai, H. Takimoto, Y. Mitsukura, and M. Fukumi, "Age and gender estimation by using facial image," in 11th IEEE International Workshop on Advanced Motion Control (AMC), pp. 179-184, 2010.

[19] C. Shan, "Learning local features for age estimation on real-life faces," in Proceedings of the 1st ACM international workshop on Multimodal pervasive video analysis, pp. 23-28, 2010.

[20] J.-C. Chen, A. Kumar, R. Ranjan, V. M. Patel, A. Alavi, and R. Chellappa, "A cascaded convolutional neural network for age estimation of unconstrained faces," in IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), pp. 1-8, 2016.