2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI) | 978-1-7281-9675-6/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICDABI51230.2020.9325696

Gender Classification and Age Prediction using CNN and ResNet in Real-Time

Asad Mustafa
Department of Computing
Letterkenny Institute of Technology
Letterkenny, Ireland
L00151146@student.lyit.ie

Kevin Meehan
Department of Computing
Letterkenny Institute of Technology
Letterkenny, Ireland
ORCiD : 0000-0001-5447-5878

***Abstract*—The face is the most dominant part of the human body, we can get a lot of information from facial features such as detecting the face of person, gender classification and even age prediction. In current times, Computer Vision (CV) has been used to train machines to comprehend and envision the real world. In this research, a novel artifact has been presented to detect face, classify genders, and predict age from the human facial images all in real-time using a live stream from a camera source. Convolutional Neural Networks (CNN) have been used for training purposes along with the CV library, Keras. To power this novel study, each model has been trained separately and finely tuned before merging them into the final system. The use of the careful modern architecture of CNN and current regularisation methods have been properly evaluated and implemented.**

The accuracy of the developed model has been calculated manually achieving an overall accuracy of 85%. All the testing has been performed in real-time. After extensive testing and evaluation, a state of the art novel system has been developed with a combination of simple pre-processing steps. This system can be broadly deployed for security purposes, as in airports and police checkpoints, and also to restrict the access of alcohol from vending machines to under-aged people.

Keywords— Convolutional Neural Networks (CNN), ResNet, Classification.

I. Introduction

In our daily lives, a significant part of communication messages that are shared among individuals are nonverbal, which involves facial expressions and face gestures. It has always been very easy for humans to analyse these expressions, but achieving the same output with the help of a machine is quite challenging. To interpret these messages effectively has the utmost of importance in every field of life. There has been a tremendous growing interest in the research area of predicting human age by looking at one's face and classifying the gender. Gender classification plays an important part as various social interactions depend upon correct gender perception.

Gender classification emerged as an interesting research topic in the field of Computer Vision and Machine Learning. Although it is much challenging, it will contribute to enhancing the Human-Computer Interaction (HCI), as well as improving facial recognition applications, increasing efficiency, and accuracy of facial recognition. With the ever-increasing use of social media, automatic gender classification of human facial images has a tremendously growing amount of applications. It will also escalate the process of gathering demographic data for commercial purposes, which will, in turn, assist in a lot of further studies.

Face Detection is a significant Computer Vision problem for identifying faces in images as well as in a real-time video [1]. Facial and emotion recognition has been deployed in various sectors, these sectors include surveillance, clinical and psychological diagnosis, hiring purposes, and also in market research.

To date, Gender Classification may be classified as a binary classification problem, either male or female. This may not be politically correct but a lot of datasets are set up to assess this as a binary problem. The system developed for gender classification needs to be swift, reliable, and time-efficient. There are a lot of databases available that are utilised for gender classification. The early attempts started from classifying the gender from textual data, names of peoples, and audio. This process can also be performed in real-time using a live feed from a webcam.

The gender classification and age prediction system can be implemented in following sectors:

1) Security Control, as it can control minors from purchasing alcohol and other prohibited objects under the underage acts.

2) Human-Computer Interaction, as the system can show only the content based on the age of the person.

3) Law Enforcement, as the automatic age and gender detection, can determine potential suspects very efficiently based on age and gender.

There are many other applications of age and gender detection systems such as criminal and terrorist investigations, social security, border control, passport control, medical records, and others. This can be deployed to alcohol vending machines as well to refrain minors from buying these prohibited products. Such a system has not been deployed until now and it proves to be a major contribution to the society from the field of Computer Vision.

II. Deep Learning Techniques for Computer Vision

In recent times, deep learning techniques proved to be a big success in the Computer Vision discipline. Deep learning enables multi-layered computing models to determine and interpret data with multiple abstraction levels and imitating how the information is perceived and translated by the brain.

978-1-7281-9675-6/20/$31.00 ©2020 IEEE

So, it implicitly captures large scale data structures. Deep learning has outperformed many of the previously existing techniques. Deep learning has enabled various techniques in computer vision to increase accuracy and efficiency, which includes object detection [2], action recognition, human emotion recognition, and others. Further, types of deep learning techniques will be discussed with their comprehensive details.

A. Convolutional Neural Networks (CNN):

Convolutional Network Network (CNN) was first proposed in 1962, and it consists of the following layers.

- Convolutional Layer: As it is known that CNN utilises various kernels, so the convolutional operation of the layer increased the learning time of the developed model [3].
- Pooling Layers: It reduces the spatial dimensions of the input volume for the next convolutional layer. It only affects the length and height, and not the depth [4].
- Fully connected Layers: Several connected neural network layers have been used to perform any high-level reasoning [3].

There are some difficulties that might also arise with CNN, such as overfitting, which is due to the CNN training of a large number of parameters. The solution to it is the pre-training of parameters, which accelerates the learning process of the model as well as improves the generalising capability of the model. In short, CNN has outperformed the usual and traditional machine learning algorithms.

B. Deep Boltzmann Machines(DBM) and Deep Belief Networks (DBN)

Deep Boltzmann and Deep Belief Networks are deep learning models from the "Boltzmann family" [3]. The model used in them is the Restricted Boltzmann Machine (RBM) which is a generative stochastic neural network. There is a single difference between DBM and DBN. DBM have undirected connections between all layers of the model, whereas DBN has undirected connections between top two layers, the remainder are all directed.

III. RELATED WORK

Gender Classification may be classified as a binary classification problem, either male or female. A survey has been performed by Ng *et al.* about recognising the gender using Computer Vision techniques, as gender is a vital demographic attribute of humans [5]. According to them, gender classification depends on various facial features involving hair, neck region, lips, eyes, and others. They performed a comparison of multiple techniques and datasets involved in gender classification using human facial images.

The main aim of gender classification is to determine whether it is a male or a female, given a facial image, and also to predict the age of that person. As far as techniques are concerned, studies have proved that neural networks have outperformed other algorithms remarkably, including SVM in gender classification. Haseena *et al.* proposed a deep Convolutional Neural Network (CNN) to handle a large dataset for gender classification with improved accuracy [6]. The architecture of the system consists of the steps outlined.

- Data pre-processing, detect the face image using facial landmarks using the dlib library of Computer Vision
- Extracting the features from the input image using a deep convolutional neural network. For this system, they use ten convolutional layers and four max-pooling layers with an input of 5 x 112 x 112. Figure below represents the detailed structure of each separate single in the architecture of the proposed system.
- Classifying the image as male or female after they have been trained using the neural network. KNN classifier has been used in this architecture for classification.

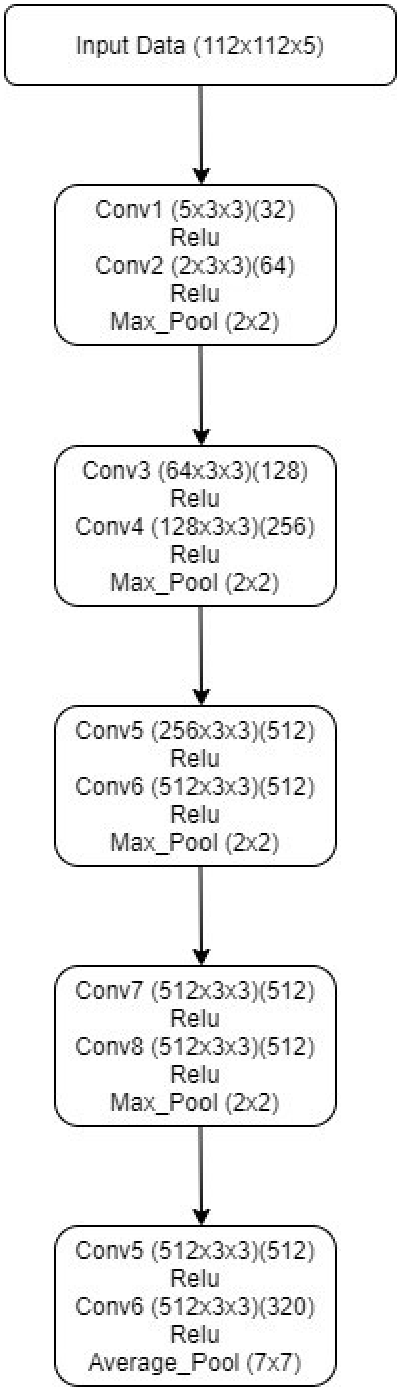

Fig 1. Detailed Layers of CNN [6]

The same architecture can be further enhanced by increasing the number of convolutional layers.

Video-based gender classification is comparatively more challenging than for still images due to variations in appearance and other factors. Chen *et al.* came up with the idea of classifying gender using live video feed, which works on CNN architecture [7]. They proposed a Multi Branch Voting Convolutional Neural Network (MBV-CNN) with three branches. Adaptive Brightness Enhancement (ABE) has been applied to each branch to eliminate the illumination issue. Majority Voting Scheme has been implemented for the final classification, which was integrated into the architecture. The system was tested on several large datasets achieving different accuracies for each one with the best results of 95.4% GCVL dataset. The results prove that the architecture successfully makes significant improvements as compared to previous models for gender classification.

Venugopal *et al.* proposed an architecture used to determine the gender of children between the ages of 6 and 8 using the local binary and non-binary classifier [8]. Local Binary Pattern (LBP) is a type of texture descriptor, where the results are indicated by a binary digit. Local Directional Pattern (LDP) is an advanced version of LBP which uses some better kernels. SVM has been used for classification purposes based on all the extracted features from LBP and LDP. The accuracy metrics of the architecture can be measured by the following equation:

$$Accuracy = \frac{Correct\ Predictions}{Total\ Number\ of\ Predictions}$$

Three descriptors are applied to the SVM system; Linear, Polynomial, and RBF, with the highest accuracy of Linear SVM i.e., 96.66%.

Liu *et al.* proposed a somewhat advanced Deep CNN network for age and gender classification in real-time [9]. They train the GoogleNet [10], which is a 22 layered neural network, with four high tech GPU. Then they load the trained neural networks into highly specialised NVIDIA Jetson TX_1 which will take input image from the webcam and gives the classification output. The framework architecture is outlined in Figure.

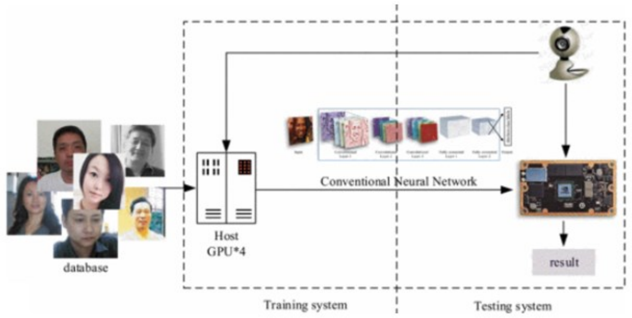

Fig 2. GoogleNet based architecture framework [7]

After adequately testing the system on live data, the GoogleNet based age and gender classifier proved itself accurate a total of 98% in the realistic complex scenarios.
Talking about the analysis system as a whole, Dehghgan *et al.* proposed an emotion, age, and gender classifier (DAGER) using CNN and Computer Vision algorithms [11]. The accuracies achieved were generally very good relating to age and emotion classification, whereas gender classification attained an accuracy of 91%, which is remarkable. The emotion recognition also outperformed the emotion face API of Microsoft by a margin of 15% [11].

Another study in the field of neural networks is age, race, emotion, and gender classification from the human facial images, which was proposed by Gudi [12]. The dataset was FER-2013, and the architecture consists of 3 convolutional layers, one fully connected layer, and some small layers in between. The system was able to achieve an accuracy of 67%. The main breakthrough in this study is that it comes up with a valuable analysis of the effect of adjusting various variables, including dropout, network size, and pooling.

From the detailed analysis of several techniques for Face and Emotion analysis along with gender classification, it can be concluded that Convolutional Neural Networks (CNN's) are the best to solve this widespread issue. CNN proves to be the best neural network architectures designed to improve big data handling. As the computing power is increasing exponentially on a daily basis, CNN provides a learning architecture that depicts the "brain-like" structure, which is capable of performing wonders in the field of Computer Vision.

IV. Methodology And Implementation

With the aid of research carried out in the previous sections, the detailed design of the system architecture will be discussed in this section, along with how it will be implemented.

Python programming language will be used throughout the research implementation along with various computer vision and machine learning packages and libraries. The research will essentially be a high-level convolutional neural network API, which is written in Python as it also can run Tensorflow. Python, in itself, is a high-level programming language. Python has easy readability, easy codability, object orientedness, and open-sourced. It has been widely used in Big Data, Machine Learning, and Computer Vision due to the availability of a vast amount of packages in it [13]. Moreover, Python has been selected for this experimentation as it is free to use, compatible with Windows Operating System, and also has all the necessary libraries used for Face identification, emotion detection, and gender classification.

Face Identification involves three steps. From a given image, detect which part is the face, then train our classifier for that data images dataset, and then lastly, to predict the face. For this, OpenCV will be used, which is an open-source library of Python for computer vision. For face detection, a Haar cascade frontal face default classifier will be used, which is a pre-trained model available free online [14]. There will be default Haar classifier used for emotion recognition and gender and age prediction. The first step is to detect the face in the image, which will be done by using some test images. All the images used for the training purposes are available online and open-sourced. The training images consists of a few of my static photos so that it will recognise me in real-time identification. Also, the training set consists of public domain figures such as celebrities and politicians in which gender and skin color will be classified as well.

As discussed earlier, the whole experimentation will be implemented in Keras running upon Tensorflow in the backend. The entire Convolutional Neural Network will be built upon these along with the OpenCV library of Computer Vision. Keras has the ability to check if the current epoch of the model has performed better than the previously saved epoch [15]. In such a case, the best model weights will be saved into a file that will allow loading the weights directly without any retraining if the model has to be used in some other situation. As compared to other similar libraries, Keras has modularity, extensibility, and Python nativeness. Keras has also proved itself an ideal library among all the Python Deep Learning libraries due to its quick experimentation, which is required in this research. Keras can also be used to perform image augmentation, which is to be done in this experimentation for training purposes.

As far as CNN is concerned, from the comparison stated in the previous section, CNN outperformed many machine learning algorithms both in speed and accuracy.

Gender Identification is to classify gender as male or female. It can be taken as a binary model as there will only be two possibilities, either male or female. This will also be implemented in the same project as Face identification and emotion detection. OpenCV and Keras are the main libraries to be used for Gender Classification. For this purpose, there is a different training file used, which is available open-source on GitHub [16]. This hdf5 file contains the training model results, which has been trained on more than 10,000 images. Due to the availability of this pre-trained model, there will be no need for any other training for gender classification. The model will be tested manually in evaluation section of this study. In the live feed from the webcam, the model will analyse the image from the webcam and give output by telling the gender and predicting the age of the person in the webcam.

In this particular part, we are using a model that has been trained on more than 10,000 images of various people with ages between 2 to 80. The face detection has been done again by the frontal face Haar Classifier which will detect the face from the image. The model has been pre-trained, and the trained files are saved in HDF5 format with its weights. These models have been chosen due to the scarcity of resources and time constraints for training purposes. The weights file is of around 200 MB, which will be used in the final API of the experimentation along with other trained models.

Keras, Tensorflow, and OpenCV libraries are used for this model training. The ResNet model has been used in the implementation of gender and age predictor. ResNet has the ability to train thousands of layers of the neural network with the achievement of compelling performances [17]. The model would give 'M' in case of male and 'F' in the case of the female along with the predicted age.

All the models will be trained individually before merging them into a single system. They will also be tested separately to measure the accuracy and evaluate the model. As there is no single dataset and model being used, so it would not be feasible to evaluate the accuracy and evaluate the system as one.

V. Results

This model is capable of predicting ages between 2 to 80, and classify genders as Male or Female by analyzing human facial features in real-time. The model has been trained on ten thousand grayscale human facial images. In this research, the trained model and weights are being used. Due to a lack of resources and time constraints, the pre-trained model has been acquired, as discussed previously.

There is no proper dataset available for this model training and test, so the model has been tested manually on real-time human faces and images. First, the model has been tested individually to assess the accuracy in real-time over human images manually. After that, it has been integrated with the final model to achieve the required output in the end. As the model is predicting age in real-time, so it is subject to change with every frame captured by the webcam.

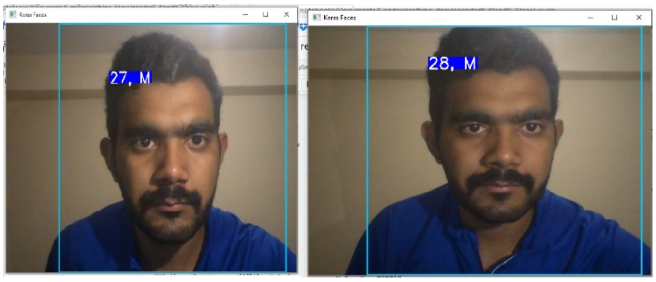

Fig 3. Gender and Age Prediction within two frames

Figure 3 shows the prediction results of the developed model. The model has been tested in real-time with the researcher in front of the webcam. As the predictions are changing with every frame, the figure shows two instances of the predictions taken at the same time. The accuracy of the model can be assessed manually, as it shows nearly accurate results. The model classifies gender accurately. The real age of the subject in Figure 3 is 26 years, and the model is predicting it almost correctly, with accurate gender classification m.

In order to widen the testing scope of the model, tests have been run using the facial pictures of celebrities available online and open-sourced. As far as gender is concerned, the model classifies between male and female very accurately. For testing purposes, the model has been shown with pictures of 100 celebrities and other people available online, and it achieved an accuracy of 90%, which has been calculated and noted manually in each case. In the case of age prediction, there are variations seen in the results. The model is shown the image of the same person through webcam, and it predicted different ages with a minor difference, as shown in Figure 3.

Figure 4 shows the testing example run on two celebrities. On the left is Donald Trump, and on the right side is Maisie Williams. The model has classified the genders accurately as male and female. It predicted the age of Donald Trump as 55, which is 70 in reality at the time the picture was taken. As far as the female celebrity is concerned, her age was 20, and the model predicted it to be 15. These variations are due to the appearance of the face in real-time and other external effects such as face makeup and lighting.

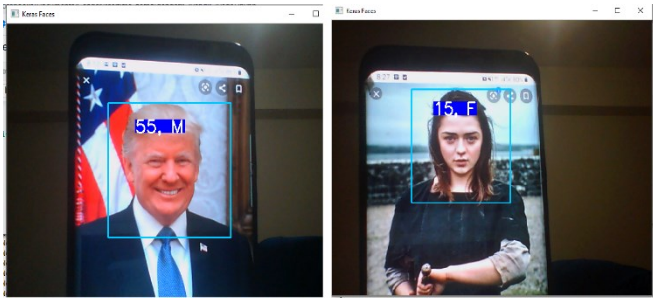

Fig 4. Testing model over Celebrities pictures in real-time

There are few variations in the age predictions due to external factors. The accuracy can be enhanced by more rigorous training of the model. There is another constraint that the model is unable to detect age and gender from the side pose of the view in real-time. This is due to the fact that the proposed CNN model is constrained only to detect frontal face view. There are other Haar cascade models that need to be explored for the detection of faces from the side pose.

VI. Concluding the Paper

Expeditious research efforts have been carried out over the last two decades in the field of age and gender prediction. This field of Computer Vision is growing exponentially. In this research, an efficient method has been implemented to recognise the human face, along with gender and age prediction, all in real-time. The main aim of this research is to grant a system that has the ability to give the single output in which it recognises the human in the live video, classify the gender and predicting age. This experimentation offers human-like abilities to the model using Deep Convolutional Neural Networks (CNN). The main aim of this study is to explore the use of saliency, which is equivalent to the human visual system to classify the gender and predict age.

All the techniques and methods already implemented in this field are discussed with a thorough literature review of various Machine Learning and Deep Learning techniques. Furthermore, the importance and need for Computer Vision in this era highlights the exponential growth of CNN in the field of Human facial feature detection and classification. The dataset for all the three modules of this research is different, and models have been trained separately. This gives the unique and diverse results as compared to the previously existing models.

Facial images have proven their importance in recent decades, and it is primarily due to their promising real-world applications in several emerging fields. The proposed system is capable of performing the classification of classifying the gender as either male or female, and predicts the age from 2 to 80.

The accuracy of the model is calculated separately to provide a better comparison and interpretation of the study. The proposed architecture has been built systematically in order to increase accuracy and reduce the number of parameters.

A CNN visualises the input images as a specific form of Deep Learning, and it will also help in explaining and demonstrating the capabilities of the model learned through different Facial Emotion Recognition (FER) datasets. The main similar feature to all these disruptive models is Artificial Intelligence, and more precisely, it is known as Deep Learning in which the model is capable of learning from data. This research also compares a number of different techniques and algorithms that have already been used in this field of Computer Vision.

Gender classification and age prediction has been tested manually, and they have produced some remarkable results. Gender classification is considered a binary problem in this research, so it has proved itself very efficient with the use of Keras and Resnet and achieves an overall accuracy of approximately 85%. Age prediction varies with a lot of external factors, such as lighting effects, facial expressions, skin tones, and with gender.

The output of this research is human gender classification and age prediction combined together to give a single output in real-time from the live feed of the webcam. The research has been successful with the achievement of the required output. Although there is a plethora of research already completed on these topics individually, there is no similar study in which all these applications of Computer Vision are combined together.

VII. References

[1] J. Brownlee, “How to Perform Face Detection with Deep Learning,” *Machine Learning Mastery*, Jun. 02, 2019. https://machinelearningmastery.com/how-to-perform-face-detection-with-classical-and-deep-learning-methods-in-python-with-keras/ (accessed Jun. 09, 2020).

[2] W. Ouyang *et al.*, “DeepID-Net: Object Detection with Deformable Part Based Convolutional Neural Networks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 7, pp. 1320–1334, Jul. 2017, doi: 10.1109/TPAMI.2016.2587642.

[3] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, “Deep Learning for Computer Vision: A Brief Review,” *Computational Intelligence and Neuroscience*, Feb. 01, 2018. https://www.hindawi.com/journals/cin/2018/7068349/ (accessed Jul. 14, 2020).

[4] Y.-L. Boureau, J. Ponce, and Y. LeCun, “A Theoretical Analysis of Feature Pooling in Visual Recognition,” p. 8, 2010.

[5] C. B. Ng, Y. H. Tay, and B.-M. Goi, “Recognizing Human Gender in Computer Vision: A Survey,” in *PRICAI 2012: Trends in Artificial Intelligence*, Berlin, Heidelberg, 2012, pp. 335–346, doi: 10.1007/978-3-642-32695-0_31.

[6] S. Haseena, S. Bharathi, I. Padmapriya, and R. Lekhaa, “Deep Learning Based Approach for Gender Classification,” in *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Mar. 2018, pp. 1396–1399, doi: 10.1109/ICECA.2018.8474919.

[7] J. Chen, S. Liu, and Z. Chen, “Gender classification in live videos,” in *2017 IEEE International Conference on Image Processing (ICIP)*, Sep. 2017, pp. 1602–1606, doi: 10.1109/ICIP.2017.8296552.

[8] A. Venugopal, Y. O. V, and R. N. T, “A SVM based Gender Classification from Children Facial Images using Local Binary and Non-Binary Descriptors,” in

2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), Mar. 2020, pp. 631–634, doi: 10.1109/ICCMC48092.2020.ICCMC-000117.

[9] X. Liu, J. Li, C. Hu, and J.-S. Pan, "Deep convolutional neural networks-based age and gender classification with facial images," in *2017 First International Conference on Electronics Instrumentation Information Systems (EIIS)*, Jun. 2017, pp. 1–4, doi: 10.1109/EIIS.2017.8298719.

[10] C. Szegedy *et al.*, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[11] A. Dehghan, E. G. Ortiz, G. Shu, and Z. Masood, "(1) (PDF) DAGER: Deep Age, Gender and Emotion Recognition Using Convolutional Neural Network," *ResearchGate*, Feb. 2017. https://www.researchgate.net/publication/313713649_DAGER_Deep_Age_Gender_and_Emotion_Recognition_Using_Convolutional_Neural_Network (accessed Jul. 13, 2020).

[12] A. Gudi, "Recognizing Semantic Features in Faces using Deep Learning," *ArXiv*, 2015.

[13] K. Kumar, "WHY PYTHON ROCKS FOR RESEARCH....???," vol. 06, no. 09, p. 3, 2019.

[14] V. Pisarevsky, "opencv/opencv," *GitHub*, 2013. https://github.com/opencv/opencv (accessed Jul. 21, 2020).

[15] I. Lasri, A. R. Solh, and M. E. Belkacemi, "Facial Emotion Recognition of Students using Convolutional Neural Network," in *2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS)*, Oct. 2019, pp. 1–6, doi: 10.1109/ICDS47004.2019.8942386.

[16] K. C. Naik, *krishnaik06/Gender-Recognition-and-Age-Estimator*. 2020.

[17] V. Fung, "An Overview of ResNet and its Variants," *Medium*, Jul. 17, 2017. https://towardsdatascience.com/an-overview-of-resnet-and-its-variants-5281e2f56035 (accessed Aug. 17, 2020).