Proceedings of the International Conference on Sustainable Computing and Data Communication Systems (ICSCDS-2022)
IEEE Xplore Part Number: CFP22AZ5-ART; ISBN: 978-1-6654-7884-7

Perception of Age and Gender Detection by using Hierarchical Deep Learning Architecture through Vision

1st M. Praveena,
Department of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation, Guntur, AP, India
Praveena.mandapati@kluniversity.in

2nd K. Geetha Srinija,
Department of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation, Guntur, AP, India
geethasrinijakndds@gmail.com

3rd A. Meghana,
Department of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation, Guntur, AP, India
meghanamaggie115@gmail.com

2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS) | 978-1-6654-7884-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICSCDS53736.2022.9760865

***Abstract*--When it comes to software systems that analyze face photos, gender detection is a crucial feature. This research provides a quantum machine learning-based approach for automated gender categorization from face photos utilizing a hybrid classical-quantum neural network, which is based on quantum machine learning. Using the knowledge of a pre-trained off-the-shelf Deep Neural Network (DNN) in conjunction with the transfer learning of a quantum variational circuit, an accurate binary classifier has been produced in this paper. An example of a hybrid network is the binary classifier, which is composed of a convolutional base of the Deep Neural Network (DNN) that has been compressed with a dressed quantum circuit and a dressed quantum circuit. In the classification of a publicly accessible collection of face photographs, better results than that have been previously reported have been obtained.**

Keywords: Deep Neural Network (DNN), Convolutional Neural Network (CNN), OpenCV, blob.

I. INTRODUCTION

The field of computer vision has existed since the late 1960s. Academics are still working on some of the oldest problems in computer vision, such as picture classification and object identification, which have been attempted to solve for decades. We've arrived to a stage where, with the help of neural networks and deep learning, computers can begin to really perceive and recognize objects with high accuracy, sometimes even exceeding humans in some scenarios. The OpenCV Deep Neural Network (DNN) module is also a great place to start learning about neural networks and deep learning in the context of computer vision. Furthermore, beginners may easily get started because to its highly optimized CPU performance, even if they do not have a very powerful GPU-enabled PC.

Automatic gender prediction from face images has sparked a lot of attention in recent years because to its wide range of applications in a number of facial analysis difficulties. Despite the fact that current models are becoming more accurate, they still fall short of the level of precision necessary for usage in real-world applications due to significant disparities in face images (such as variations in lighting, size, and occlusion).

The objective of this study is to construct a lightweight command-line-based application that utilizes Python-based modules to automatically find faces in a static picture and estimate the gender of the identified

humans using a deep learning-based gender recognition model.

OpenCV is an internet-based, free and open-source toolbox for computer vision, machine learning, and image processing. OpenCV supports a broad range of programming languages, including Python, C++, and Java, and is used for a number of image and video analysis applications, including face detection and identification, picture editing, optical character recognition, and more.

OpenCV is a free and open-source computer vision programme. OpenCV has a number of benefits, the most noteworthy of which are as follows:

A. OpenCV is a computer vision library that may be downloaded for free from the internet.

B. OpenCV is very fast because to its C/C++ development.

The majority of operating systems, including Windows, Linux, and macOS, are compatible with OpenCV.

We've all heard about OpenCV, which is generally considered as one of the best computer vision libraries on the market. It also has the capacity to do deep learning inference and a variety of other tasks. The ability to import models from a number of frameworks is particularly notable, since it enables us to execute a range of deep learning procedures. With version 3.3, the OpenCV library added the ability to handle models from a variety of frameworks. Many newcomers to the field are aware of OpenCV's strong capabilities, while many others are not. As a consequence, children miss out on a lot of fun and important learning opportunities.

The only kind of deep learning inference enabled by the OpenCV Deep Neural Network (DNN) module is deep neural networks on photos and videos. It is impossible to fine-tune or train with it. Nonetheless, for anybody new to the world of deep-learning-based computer vision and experimentation, the OpenCV Deep Neural Network (DNN) module is a great place to start.

One of the most appealing features of the OpenCV Deep Neural Network (DNN) module is that it is optimized for Intel processors. We can acquire high frames per second while doing inference on real-time videos for object recognition and picture segmentation applications. When utilizing the Deep Neural Network (DNN) module with a model that has been pre-trained using a certain framework, we typically receive more frames per second. Consider the inference time for picture classification using a variety of frameworks.

II. RELATED WORK

Deep learning is a technique that is commonly utilized in picture, sound, and text analysis applications. Face recognition and detection, as well as age and gender detection, are the primary study interests in this discipline. Success rates for vanilla machine learning algorithms created for identification of face traits such as age and gender detection maintained between 75% and 80% for the majority of cases. Studies done in the same area have shown that when deep neural networks are applied, the success rate exceeds 90 percent in most cases. Classical classifiers were utilized for face, gender, and age identification in the majority of the experiments that were reviewed. In identification and detection systems, the classifier has a negative impact on both the performance and the success of the system. This study is founded on the hypothesis that "With the evolving technologies and methodologies, can a face, age, or sex be identified without the need of any classification algorithm?" in order to remove all of the drawbacks mentioned before. The suggested approach is based on a facial recognition project provided on GitHub by Geitgey [2], which serves as its starting point. As part of his research, Geitgey demonstrated how to recognize a person from a single photograph by using the face embedding model of the dlib library (http://dlib.net/). In this research, the approach was refined to be more precise and was used to the

identification of genders. Most significantly, the research demonstrated that gender recognition can be achieved using 128-D average face vectors without the need of a classification algorithm, which was the most significant breakthrough in this new technique.

Many experiments have been undertaken in the literature on the face, age, and gender identification by employing facial landmarks in conjunction with deep learning algorithms, with the results being published in scientific journals. Cha et al. [3] used a multi-task Deep Convolutional Neural Network (DCNN) technique to recognize faces in a variety of stances. They used facial landmarks to detect faces in a variety of poses. In their investigation, they employed the FDDB dataset [4], and it was discovered that the approach they presented outperformed the other state-of-the art methods by a factor of three percent. Face point identification was achieved using a 3-level Deep Convolutional Neural Network (DCNN), which cascades three layers of convolutional networks, as developed by Sun et al. [5]. Based on this drawback, a novel Tasks-Constrained Deep Convolutional Network (TCDCN) for facial point recognition has been proposed [6] that minimizes model complexity while maintaining accuracy. Eidinger et al. [7] have developed a method for predicting age and gender-based on unfiltered faces. Within the scope of the research, they created their own dataset for the prediction of the age and gender of participants. For data classification, they devised a dropout-SVM approach inspired by the deep belief network's dropout learning methodology, which they called "dropout SVM." By using a "frontalization" technique to the face discovered in unconstrained pictures, Hassner et al. [8] were able to correct the front aspect of the face. They included crucial face feature points into the design of their research infrastructure. The success rate for face recognition and gender prediction algorithms has risen as a consequence of the new picture acquired. Levi et al. [9] have developed a basic Convolutional Neural Network (CNN) that may be used to estimate age and gender-based on a small dataset and is easy to implement. Ranjan et al. [10] developed a deep multi-task learning system called Hyperface that can do face identification, landmark localization, posture estimation, and gender recognition all at the same time using Convolutional Neural Networks (CNN). In experiments, it was discovered that the suggested technique is capable of capturing both global and local information on faces and that it outperforms a large number of competing algorithms in each of the four tasks tested. Rothe et al. [11] developed a deep learning model that can estimate age and gender from a single photograph by using machine learning techniques. Within the purpose of the investigation, they made use of the IMDB-WIKI dataset. The pictures in the dataset used for training in earlier research in this area were not all from a single image, but the most essential aspect that separates this work from the others is the use of a single image for training. Some of the convolutional layers in the VGG-16 architecture have been altered to improve their performance. The Local Deep Neural Network (Local-DNN) presented by Mansanet et al. [12] was designed specifically for gender recognition. It is recommended that a Local Deep Neural Network(local-DNN) model be developed, which is based on deep learning architecture and local facial characteristics. Several layers of Feed-Forward Networks, as well as tiny overlapping areas in the visual fields, are used to aid in the model's learning. Another research [13] used Convolutional Neural Network (CNN) architecture to accomplish face-based gender estimation, which was successful. A deep neural network (DNN) model that can predict race and gender, as well as age prediction, was proposed by Xinga et al. [14]. Deep multi-task learning architecture was used in this model. With the help of gender dictionary learning, Moeini et al. [15] were able to execute gender identification based on the aspects of facial position and expression. Qawaqneh and colleagues [10] have developed a deep neural network model that can categorize people based on their age and gender. In addition, they presented a new cost function. The research

team employed both voice data and facial photos in their findings. Philip et al. [1] have been using both VGG19 and VGGface models, which were previously trained Convolutional Neural Network (CNN) based deep neural networks to do their analysis. They've been looking towards transfer learning for model trainings for some time. They have been altering the model parameters in order to improve the overall success of the system. It has been reported that their Convolutional Neural Network (CNN) based models have achieved 98 percent success in gender recognition. Dhomne et al. [18] have presented a VGGNet model based on Deep Convolutional Neural Network (D-CNN) that may be used to identify gender based on face photos. Xu et al. [11] have proposed the Hierarchical Multi-task Network (HMTNet), a deep neural network that can distinguish between a person's sex, race, and face attractiveness based on a portrait photograph of the individual. Document, word, or picture embedding is the representation of a document, word, or image in a two-dimensional or three-dimensional space. That is to say, documents, phrases, or images (things, people, faces, and so on) are represented vectorially in two-dimensional space. Face embeddings are a mathematical representation of faces that are represented as numerical vectors. A variety of procedures are used in the creation of face embeddings. Deep neural networks are one of these technologies. For face embedding extraction, deep neural networks have been used in two key research published recently in the literature: Dlib [14] and the Openface face recognition library [15]. dlib is built in C++ and offers a Python API for interfacing with it. Among other things, Openface makes use of the dlib package for fundamental operations such as face detection, and it makes use of a deep neural network model created in the Torch environment for face embedding extraction. In these two major investigations, the individual recognizing the face employs a 128-D face vector representation to do so. ResNet is the deep neural network that lies at the heart of the Dlib facial recognition model's deep neural network. When it comes to picture recognition, the ResNet (Residual Networks) employed is a 34-layer network created in 2016 by He, Zhang, Ren, and Sun [15] for image recognition. FaceNet is a deep neural network model developed by Schroff et al. [13] from the Google team that is used to extract face embedding vectors. Faces are represented by 128-dimensional vectors in this model. In order to extract face embeddings using FaceNet, at least three photographs of each individual are necessary. Because FaceNet makes use of a triple-based loss function, which is similar to the one employed in the LMNN model [12]. The FaceNet model is composed of layers such as input, Convolutional Neural Network (CNN) for face identification, L2 distance to separate face vectors, generate the face embedding, and the triple loss function, in which the error values are represented by the error values. Dlib is based on the ResNet-34 architecture in its most basic form. In contrast to the RestNet-34 design, the number of filters and layers has been decreased. The number of filters on each layer has been reduced by half. Some layers were deleted, and the network was reconfigured to include 29 layers instead of the previous 28. This has resulted in a reduction in the overall cost of computation. With the newly built network, it was possible to get a 128-dimensional face embedding vector. The VGG and face scrub datasets were utilized in the training of the new network.

III. PROPOSED WORK

A convolutional neural network design that is similar to CaffeNet and AlexNet is used. The network is made up of three convolutional layers, two fully connected layers, and a final output layer. Below is a detailed description of the strata.

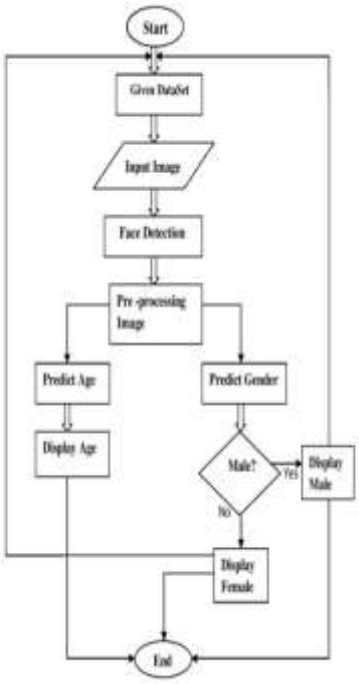
- With 96 nodes and a kernel size of 7, the first convolutional layer is the most complex.
- The second convolutional layer, Conv2, has 256 nodes and a kernel size of 5.

- Conv3: The third convolution layer has 384 nodes and a three-kernel size.

There are 512 nodes in each of the two entirely connected layers.

Adience dataset is used to train the model.

A. Flowchart



B. Gender Prediction

They've structured Gender Prediction as a classification problem. The output layer of the gender prediction network is softmax, with two nodes representing the categories "Male" and "Female".

Hierarchical Deep Learning Neural Network (HiDeNN) is used in Image Recognition and to classify them.

C. Age Prediction

Because the conclusion will be a real number, age prediction should preferably be approached as a regression problem. It is, nevertheless, difficult to accurately forecast age using regression. Even humans are incapable of estimating a person's age just by looking at them. We do know if they are in their 20s or 30s, however. As a consequence, it's best to approach this assignment as a classification issue in which we try to figure out what age group the individual belongs to. Children between the ages of 0 and 2 are put into one category, while those between the ages of 4 and 6 are put into another. The Adience dataset is divided into eight age groups [(0–2), (4–6), (8–12), (15–20), (25–32), (38–43), (48–53), and (60–100)]. As a result, the final softmax layer of the age prediction network comprises eight nodes reflecting the aforementioned age ranges.

It should be mentioned that estimating age from a single image is a difficult problem to solve since age perception is influenced by a range of factors, and people of the same age might seem substantially different in various parts of the world. Furthermore, many people make a determined attempt to hide their genuine ages.

D. Faces are detected

The Deep Neural Network (DNN) Face Detector will be used for face detection. The model is just 2.7MB in size and runs quite quickly, particularly on the CPU. The face detection is done with the help of the getFaceBox function.

E. Identify the Gender

The gender network will be loaded into memory, and the detected face will be sent through the network to check that it is male. For the two classes in the system, the forward pass yields probabilities or degrees of confidence. To arrive at a final gender prediction, the highest of the two results is used.

F. Detect and display the user's age

To acquire the output from the ageing network, we use the forward pass. We may take the maximum number of outputs from all the networks to obtain the age group that is expected since the network architecture is like that of the Gender Network.

G. Display the output

The imshow function will be used to display the network's output on the pictures that were used as inputs to the network.

IV. RESULT ANALYSIS

As showed above, the network is capable of accurately predicting both gender and age to a high degree of precision. The part of the opposite gender has been played by a large number of performers in films.

A. Outputs

We were pleased with the performance of the gender prediction network; however, the age prediction network fell short of our expectations. We looked through the article for an explanation and discovered the confusion matrix for the age prediction model, which is Fig. 1. and Fig. 2.

Fig. 1. Output image with male representation

Fig. 2. Output image with female representation

V. CONCLUSION

While the challenge of distinguishing age and gender is not very difficult in and of itself, it is more difficult than a great many other computer vision tasks. The knowledge required to construct these types of frameworks serves as the basic rationale for this issue spot. While general article discovery errands can regularly approach many thousands or even hundreds of thousands of images for preparation, datasets with age and gender namesa are significantly more modest, typically numbering in the hundreds or even thousands or, in the best-case scenario, a few thousand at most. Python was used to gather pictures, and the model did not perform well in terms of accuracy rate; thus, additional development in the model method is necessary.

Ultimately, I believe that the models' accuracy is acceptable, but that it may be further improved by including additional data, data augmentation, and better network topologies.

If there is sufficient data available, it may be possible to utilize a regression model instead of a classification model for Age Prediction.

ACKNOWLEDGMENT

We would like to show our gratitude to the experts for sharing their pearls of wisdom with us during the course of this research, and we thank 2 reviewers for their insights.

REFERENCES

[1] B. Schuller, S. Steidl, A. Batliner, F. Burkhardt, L. Devillers, C. Mueller, and C. Narayanan, “The Interspeech 2010 Paralinguistic Challenge,” in Proc. Interspeech (2010), 2010, p. no pagination.

[2] T. Bocklet, A. Maier, J. Bauer, F. Burkhardt, and E. Noth, ¨ “Age and Gender Recognition for Telephone Applications Based on GMM Supervectors and Support Vector Machines,” in Proc. ICASSP 2008, vol. 1, 2008, pp. 1605–1608.

[3] G. Dobry, R. Hecht, M. Avigal, and Y. Zigel, “Dimension Reduction Approaches for SVM-Based Speaker Age Estimation,” in Proc. Interspeech 2009, 2009, pp. 2031–2034.

[4] F. Honig, G. Stemmer, C. Hacker, and F. Brugnara, “Re- ¨ vising Perceptual Linear Prediction (PLP),” in Procof the 9th European Conference on Speech Communication and Technology, ISCA, Ed., Bonn, 2005, pp. 2997–3000.

[5] H. Hermansky, “Perceptual linear predictive (PLP) analysis of speech,” Journal of the Acoustic Society of America, vol. 87, no. 4, pp. 1738–1752, 1990.

[6] H. Hermansky and S. Sharma, “TRAPS – classifiers of temporal patterns,” in Proc. International Conference on Spoken Language Processing (ICSLP), Sydney, Australia, 1998.

[7] C. Hacker, Automatic Assessment of Children Speech to Support Language Learning. Berlin: Logos Verlag, 2005.

[8] W. Wahlster, Verbmobil: Foundations of Speech-toSpeech Translation. New York, Berlin: Springer, 2000.

[9] A. Maier, F. Honig, V. Zeissler, A. Batliner, E. K ¨ orner, ¨ N. Yamanaka, P. D. Ackermann, and E. Noth, “A ¨ language-independent feature set for the automatic evaluation of prosody,” in Proc. Interspeech 2009, Brighton, England, 2009, pp. 600–603.

10] S. E. Linville, “The Sound of Senescence,” The Journal of Voice, vol. 10, no. 2, pp. 190–200, 1996.

[11] K. N. Stevens, Acoustic Phonetics. Cambridge,

MA 02141: The MIT Press, 1998.

[12] P. Beyerlein, A. Cassidy, V. Kholhatkar, E. Lasarcyk, E. Noth, B. Potard, S. Shum, Y. C. Song, W. Spiegl, ¨ G. Stemmer, and P. Xu, "Vocal aging explained by vocal tract modelling: 2008 JHU summer workshop final report," Tech. Rep., 2008.

[13] J. A. Nelder and R. Mead, "A Simplex Method for Function Minimization," The Computer Journal, vol. 7, no. 4, pp. 308–313, 1965.

[14] D. Olsson and L. Nelson, "The Nelder-Mead simplex procedure for function minimization," Technometrics, vol. 17, no. 1, pp. 45–51, 1975.

[15] N. Brummer, ¨ FoCal Multi-class: Toolkit for Evaluation, Fusion and Calibration of Multiclass Recognition Scores. available online: http://sites.google.com/site/nikobrummer/focalmulticlass, 2007.