FEATURE PYRAMID NETWORK BASED REAL-TIME YOLO FACIAL MASK RECOGNITION

Mrs. M.K. Geedtha ,
Dept of Computer Science and Engineering,
geedthamk2709@gmail.com

Guide
Mr. A.N. Arun B.E.,M.E.,M.B.A.,SCPM
Head of the department
arunsvistcse@gmail.com
Dept of Computer Science and Engineering
Sri Venkateshwara Insitute of Science and technology, Thiruvallur ,Tamilnadu

Abstract— **In recent years, automated recognition of faces and facial features has grown in popularity and significance. Face changes and the presence of masks make this task extremely difficult. A large amount of research has been done to recognise faces under various conditions such as changing pose or illumination, degraded images, and so on. Despite this, recognising people's faces behind masks remains difficult. The primary goal of this work is to make the system detect facial masks and, in particular, to improve the recognition accuracy of various masked faces. The proposed method involves first detecting the facial regions. A Feature Pyramid Network (FPN) learner is used to solve the occluded face detection problem.**

Keywords— *Feature Pyramid Network (FPN), YOLO (You Only Look Once), Support Vector Machine (SVM)*

I. INTRODUCTION

Deep Learning is one of the fascinating fields that allows a machine to train itself using various datasets as input and to provide the proper output during testing by using various learning algorithms. Technology today attempts to convey a great deal of knowledge-based technical breakthroughs. Face recognition is a straightforward method of recognising faces and facial features using technology, particularly hardware like video cameras. By comparing any image or video to a database of recognised faces, face recognition software or an app maps the facial features from the image or video using biometrics. It is one such piece of technology that can identify a person by recognising their face through a mask. It employs an individual's biometric data along with an AI-based pattern recognition technology. It pulls out

A. Supervised machine learning

In order to anticipate future events, algorithms might use labelled examples to apply what they have learnt in the past to fresh data. The learning algorithm creates an inferred function to forecast the values of the outputs starting from the examination of a known training dataset. After sufficient training, the system is capable of providing objectives for any new input. The learning algorithm can also check its output against the desired, correct output to identify mistakes and fix the model as necessary

.B. Unsupervised machine learning

When the information used to train is neither classified nor labelled, algorithms are used. Unsupervised learning investigates how systems can infer a function from unlabeled data to describe a hidden structure. The system does not determine the appropriate output, but it explores the data and can draw inferences from datasets to describe hidden structures in unlabeled data.

Semi-supervised machine learning algorithms fall somewhere between supervised and unsupervised learning because they train on both labelled and unlabelled data - typically a small amount of labelled data and a large amount of unlabelled data. This method

allows systems to significantly improve learning accuracy. Semi-supervised learning is typically used when the acquired labelled data necessitates the use of skilled and relevant resources in order to train.

C. Reinforcement machine learning

Algorithms are a type of learning method that interacts with its surroundings by performing actions and discovering errors or rewards. The most important aspects of reinforcement learning are trial and error search and delayed reward. This method enables machines and software agents to automatically determine the optimal behaviour in a given context in order to maximise their performance. For the agent to learn which action is best, simple reward feedback is required; this is known as the reinforcement signal.

Machine learning allows for the analysis of massive amounts of data. While it generally produces faster and more accurate results in identifying profitable opportunities or dangerous risks, it may also necessitate more time and resources to properly train it. Combining machine learning with AI and cognitive technologies has the potential to improve

D.NATURAL LANGUAGE PROCESSING

Some formal grammars claim that a parse tree represents the syntactic structure of a sentence. Machines can read and comprehend human language thanks to natural language processing (NLP). Natural language user interfaces and the direct acquisition of knowledge from human-written sources, such as newswire texts, might be possible with a sophisticated enough natural language processing system. Information retrieval, text mining, question answering, and machine translation are a few straightforward uses of natural language processing.

In many contemporary methods, syntactic representations of text are built using word co-occurrence frequencies. Popular and scalable, yet stupid, are keyword detection algorithms for search A search for "dog" might only turn up documents with the word "dog" in italics, leaving out documents with the word "poodle." The occurrence of in lexical affinity methods

II. FEATURE PYRAMID NETWORK

A feature extractor known as a "feature pyramid network" produces appropriately scaled feature maps at various levels in a fully convolutional manner from a single-scale image of any size. The underlying convolutional designs have no bearing on this procedure. In order to be employed in applications like object detection, it serves as a general method for creating feature pyramids inside deep convolutional networks. The YOLO model is used to extract the face traits. An input image is divided into grids using the YOLO (You Only Look Once) object detector, and each grid cell predicts a single object. Finally, a Support Vector Machine is used to carry out the classification operation (SVM). The SVM constructs, in a higher dimensional space, to complete the classification objective.

III. DATASET CONSTRUCTION

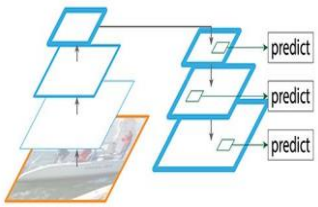
A collection of data is known as a data set (or dataset).

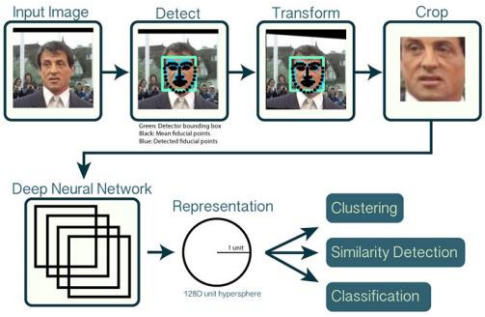
Images make up the dataset we used for this project. These pictures are kept in folders. Images of people's faces wearing and without wearing masks are kept in the designated folder and given names. These annotated image folders serve as our dataset for both the training and testing phases of our work. There are numerous ways to acquire the input image. Either directly using the camera or utilising the database folder where the images are saved, we can snap an image. the process of obtaining an image for processing from a source, typically one that is hardware-based. Datasets: A dataset is a collection of occurrences, and when using machine learning techniques

A. EMBEDDING

The given image is examined using embedding, which then produces numerical vectors that each recognised face in the image is represented by. the 64,128,256,512 different sized vectors. Face matching, face verification, and face recognition applications can all make use of this embedding. a deep learning-based technique for representing a face's identity.

A face image is used as the input, and a vector of 128 numbers representing the key facial traits is produced. This vector is known as an embedding in machine learning. Because this vector contains all of the essential data from an image, embedding is employed. The face of a person is basically compressed into a vector of 128 values. Ideally, similar faces also have similar embeddings.





B. *TRAINING*

Giving an ML algorithm—the learning algorithm—training data to use as a learning resource is the process of training an ML model. The model artefact produced during training is referred recognised as a "ML model." The ML model can be used to make predictions on fresh data for which you are unsure about the aim.

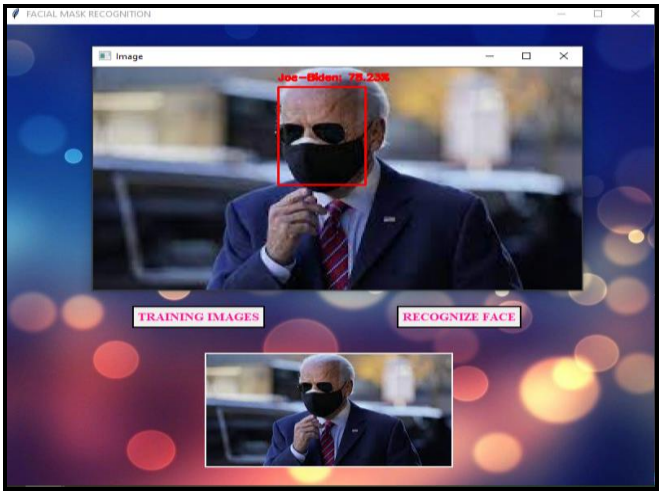
The neural network learns to produce comparable vectors for faces that resemble one another as it is being trained. For instance, if I have many pictures of my face taken over time, some of the traits will undoubtedly alter, but not significantly. Therefore, in this instance, the vectors associated with the faces are comparable, or to put it another way, they are

C. *FACE RECOGNITION*

Face recognition is performed in the comparison module. When the camera detects a face, it compares the corresponding values of the currently visible face to values stored in the file. If the values match, the face is identified and the name associated with it is displayed.

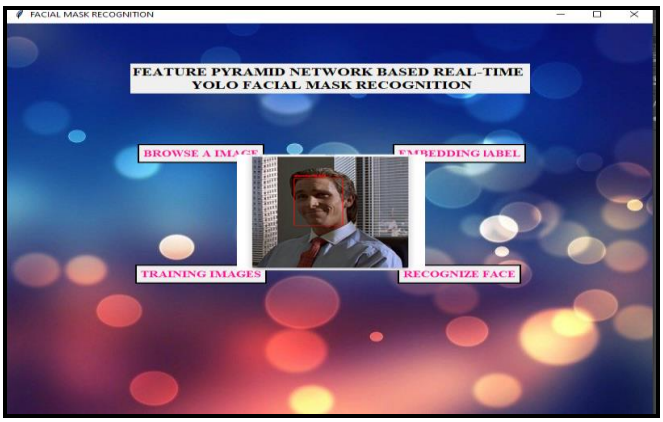
It is a task that humans perform routinely, even in varying lighting and when faces change with age or are obstructed by accessories and facial hair. Nonetheless, it has remained a difficult computer vision problem for decades until recently.

Deep learning methods can use very large datasets of faces to learn rich and compact representations of faces, allowing modern models to perform well at first and then surpass them.



IV. CONCLUSION

In this work We recognise the face if the generated embedding is closer or similar to any other embedding. we passed two images, one of the images is of person A and other of person B. In our example, we did not save the embeddings for person B but we saved the embeddings of person A. Thus, when we compared the two new embeddings with the existing ones, the vector for person A is closer to the other face embeddings of person A whereas the face embeddings of person B are not closer to any other embedding and thus the program cannot recognise him. Similar processes can be used for real time facial recognition. We have implemented the process of transfer learning in our work in order to modify the network so that when we update the dataset a new updated network will be created. Hence, we can create a network as per our data requirement.



V. REFERENCES

[1] Roomi, Mansoor, Beham, M.Parisa, "A Review Of Face Recognition Methods," in International Journal of Pattern Recognition and Artificial Intelligence, 2013.

[2] A. S. Syed navaz, t. Dhevi sri, Pratap mazumder, " Face recognition using principal component analysis and neural network," in International Journal of Computer Networking, Wireless and Mobile Communications (IJCNWMC),2013.

[3] Turk, Matthew A., and Alex P. Pentland. "Face recognition using eigenfaces," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1991.

[4] H. Li, Z. Lin, X. Shen, J. Brandit, and G. Hua, "A convolutional neural network cascade for face detection," in IEEE CVPR, 2015.

[5] Wei Bu, Jiangjian Xiao, Chuanhong Zhou, Minmin Yang, Chengbin Peng, "A Cascade Framework for Masked Face Detection," in IEEE International Conference on CIS & RAM, Ningbo, China, 2017.

[6]. Md. Milon Islam, Md. Kamrul Hasan, Md Masum Billah, and Md. Manik Uddin, "DEVELOPMENT OF SMARTPHONE-BASED STUDENT ATTENDANCE SYSTEM", IEEE 2018.

[7]. S. S. Mahat and S. Mundhe, "PROPOSED FRAMEWORK: COLLEGE ATTENDANCE MANAGEMENT SYSTEM WITH MOBILE PHONE DETECTOR", IEEE, 2018.

[8]. S. Joardar, A. Chatterjee, and A. Rakshit, "A REAL-TIME PALM DORSA SUBCUTANEOUS VEIN PATTERN RECOGNITION SYSTEM USING COLLABORATIVE REPRESENTATION-BASED CLASSIFICATION", IEEE 2018.

[9].Devira Anggi Maharani,Carmadi Machbub,Pranoto Hidaya Rusmin,Lenni Yulianti "Improving the Capability of Real-Time Face Masked Recognition using Cosine Distance", IEEE 2020.

[10]. S.Naveen, R Shihana Fathima, R S Moni ," Face recognition and authentication using LBP and BSIF mask detection and elimination ", IEEE 2016.

[11]. Naser Damer;Jonas Henry Grebe;Cong Chen;Fadi Boutros;Florian Kirchbuchner;Arjan Kuijper," The Effect of Wearing a Mask on Face Recognition Performance: an Exploratory Study" , IEEE 2020.