JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Image Recognition Using Machine Learning

Sushma Thatikonda[1],Sai Kaushil Goud Kandunuri[1],Rohit Botu[1],Hari Shanmuksai[1], Mr.B.Rajesh[2]

Students, Department of Computer Science Engineering, GITAM Deemed to be University, Visakhapatnam, India-530045[1]
Assistant Professor, Department of Computer Science Engineering, GITAM Deemed to be University, Visakhapatnam, India[2]

Abstract : Image Recognition is an assortment of algorithms and technologies that attempt to analyze images, understand the obscure representations of features underlying, to apply them todifferent tasks like classifying images into different categories. The human brain is competent in processing images that the eye glimpses in a time period as short as thirteen milliseconds, as found by neuroscientists from MIT. This rapid processing can be done approximately thirty times faster than the blink of an eye. For neural networks to recognize one or more concepts inan image, it is necessary to train them. Models can simulate the task of image identification through a variety of approaches. Face recognition is a technology competent for identifying andverifying a subject through an image, video or any audio-visual element of any face. Algorithmsoften collapse when forced to make predictions on data in cases where inadequate supervised information is available. Deep face recognition has made remarkable progress with the recent development of deep learning techniques and is widely used in many real-world applications. Given an image or video frame as input, an end-to-end deep face recognition system outputs the face feature for recognition, and with the help of a distance metric, the similarity between these features can be determined. The plan is to evaluate new categories based on learned feature.

Keywords: Machine Learning, Deep Neural network, Recurrent Neural Network(RNN), Back propagation.

I. INTRODUCTION

Face identification and recognition are paramount aspects of social interaction and define critical agilities acquired early in human life. Socially meaningful information concerning levels of acquaintance, attractiveness, and emotional standing can emanate from facial recognition, which shapes behavioural patterns. Humans have a remarkable capacity to learn and detect new patterns. [3] Face identification and recognition, as well as recognition of their features, patterns, and other characteristics, have been discovered to be exact procedures requiring brain networks. People appear to grasp new concepts rapidly when provided with stimuli and then spot departures from these notions in later situations. The brain's ability for feedforward processing, or the influx of information from the retina to the visual processing centres in just one direction, is sufficient for the brain to recognise ideas without having to undertake any additional processing. After visual input reaches the retina, the information travels to the human brain, where it is analysed for specifics such as shape, colour, lighting, and orientation. This capacity to recognise pictures seen for a brief moment aids the ability to recognise images glimpsed for a brief moment.

Face recognition is a technique for recognising and authenticating a person based on a photograph, video, or other audio-visual feature of their face. It's a well-researched subject with a diverse set of algorithms and strategies. Face recognition is the most extensively utilised solution in real-world applications among existing human biometrics technology. Deep learning-based techniques have significantly improved several computer vision applications, including face recognition, thanks to the tremendous progress of deep convolutional neural networks. Machine learning has been utilised to achieve high-level performance in a variety of real-time applications, including online search, spam detection, caption generation, and speech and picture recognition. A mathematical duplicate of a real-world approach is a machine learning model. [3] Patterns are discovered by the algorithm.

IMAGE CLASSIFICATION

The task of clustering and arranging pixels into classifications is known as image classification. Image classification is the process of categorising pictures into one of several predetermined categories, which can involve image sensors, vision pre-processing, object identification, object segmentation, feature extraction, and object classification. The techniques to picture categorization are as follows:

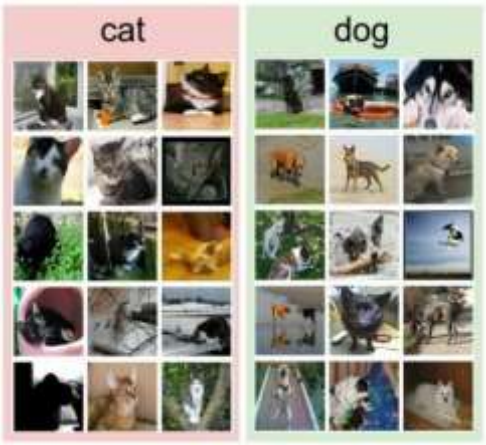


Figure 2.1 - Cat and Dog classes of images

- Supervised and Unsupervised Classification

Supervised image classification is a procedure for determining similar patterns on an image byidentifying the testing sets of known targets and then inferring those signatures to other sites ofunknown targets. It is based on the concept that a user can pick sample pixels representing distinct classes and use training sites as considerations for classifying other pixels in the image.In comparison, unsupervised image classification is the procedure by which each image is pinpointed as an associate or a member of one of the intrinsic categories available in the collection without labelled training samples. The outcomes of this method reflect the analysis of an image without providing sample classes. The user, however, must possess knowledge ofthe photos being classified. It is also possible to create final output analyses and categorised maps using a combination of supervised and unsupervised classification techniques.

- Per-pixel and Sub-pixel Classification

Per-pixel methods classify pixels into distinctive classes based exclusively on the spectral andancillary knowledge within that particular pixel. It is widely used in many areas, especially those that require classification based on pixel by pixel, straightforward calculations of environmental indices to advanced specialist systems to assign urban land covers. In contrast, sub-pixel classification methods aim to quantify the assortment of surfaces statistically to improve overall classification accuracy. This classification deals with accomplishing feature classification by splitting the pixel into more pixels based on spectral unmixing by identifyingthe surplus classes employing fuzzy logic.

- Parametric and Non-Parametric Classification

Non-parametric or soft classifiers explicitly assess the class's conditional probabilities and thenperform classification with the estimated probabilities as reference. These strategies take into account the fact that the real world is diverse. Assigning a fraction of the inland cover type found inside each pixel, for example. On the other hand, parametric or hard classifiers instantly target the classification decision boundary with no task of assembling the probabilityestimation. These algorithms achieve classification using discrete categories and includesupervised and unsupervised classifications.

- Object-based image analysis

Object-based image analysis, OBIA, starts with segmenting an image by clustering the pixels into groups of more than one pixel. The method renders objects with different geometries and orientations. In OBIA classification, various aspects like shape, colour, texture, spectral and geographical context can be used to classify objects. The two most common segmentation algorithms include multi-resolution segmentation in recognition and segment mean shift in ArcGIS.

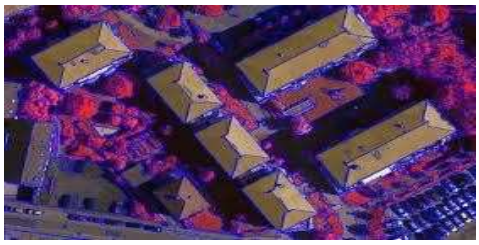


Figure - OBIA of Urban Land Cover

FACE RECOGNITION

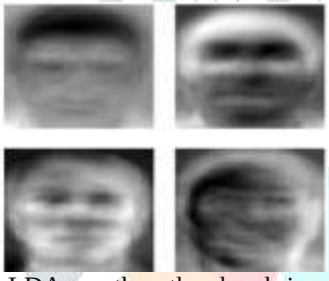
Computer algorithms are used in facial recognition to recognize particular, unique, and distinguishing information about a person's face. It's a new technology capable of recognizing and matching a human face from an image or a video frame against a group of faces. Face recognition technology identifies and measures facial characteristics from a given image and is commonly used to verify users through identity verification services. Facial recognition systems were constructed primarily using portraits of human faces until the 1990s. Governments and commercial enterprises all across the world are now using these systems to attain state-of-the-art performance. Their effectiveness varies. Despite the fact that alternative identification methods are more accurate, facial recognition has always been a good study focus, because to its non-intrusive nature and the fact that it is the most used means of personal identification. Face recognition algorithms extract landmarks and characteristics from a picture to detect

facial qualities. For example, an algorithm may examine attributes such as the size, shape, relative position, form of the eyes, ears, expression, cheekbones, nose, and jaw and use them to match other images .

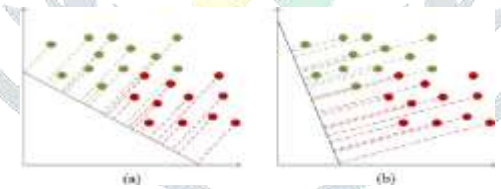
In computers, facial recognition is a difficult pattern recognition issue. Even while this behaviour comes naturally to people, it is difficult to replicate in robots. Face detection, face alignment, and face representation are the three fundamental parts of an end-to-end deep face recognition system. The impact of all of these parameters to face recognition ability is a noteworthy finding. Inferiority in any one of the factors will cause the final performance to fail.

PCA with eigenfaces, elastic bunch graph matching using the Fisherface method, linear discriminant analysis, the HMM, and neural inspired dynamic link matching are all popular recognition techniques. With the principle component analysis, research on face recognition and verification to find a face in an image with other items gained pace in the early 1990s. Eigenface is a PCA face identification approach developed by Matthew Turk and Alex Pentland. To develop a linear model, Turk and Pentland combined the conceptual method of the Karhunen–Loève theorem with factor analysis. Eigenfaces are calculated as a weighted composite of numerous Eigenfaces based on orthogonal and global characteristics in human faces. Before starting with the study, the data is pre-processed using the PCA approach. It can eliminate duplicate information, noise, and key characteristics from picture data in several dimensions, resulting in faster processing and lower costs.

Figure - Eigen Faces from AT&T Labs, Cambridge



Linear Discriminate Analysis, or LDA, on the other hand, is utilised for face recognition datasets with labels. It is most commonly used as a dimensionality reduction step in the pre-processing stage for machine learning and pattern categorization applications. To guarantee that the data distribution may be as wide as possible, PCA requires the data variance after dimensionality reduction to be as large as possible. LDA, on the other hand, advises the variance within the same category of data sets after projection to be as little as feasible, hence increasing the significance of the variation across groups. [8] The label information should be used to segregate distinct types of data and supervise the dimensionality reduction in the LDA technique.



[2]

Figure – Distinguishing (a) PCA and (b) LDA

II. LITERATURE SURVEY

In Lixiang et al. investigate current security threats from two perspectives: the preparation process and the testing / inferring process. Threats to machine learning security were identified.Different categories have been created a brief summary of then there were relevant efforts on machine learning security issued.[2] Adversaries can reduce a team's performance. In terms ofaccuracy, classification or regression models are better. There was a diagram of poisoning attacks as well as categories of poisoning assaults. A case of poisoning for several situations, acomparison was made. Against machine learning, there are a number of tactics that may be used. They different defensive strategies were presented and contrasted. Then there are the current study trends in safety hazards and protective strategies. The methodologies for machinelearning were offered. D. V. Sang et al. exploit recent advances in deep learning to create deep CNNs that can automatically read semantic data in faces without the requirement for hand-crafted feature descriptors. A few different types of CNN architectures have been suggested for recognizing facial expressions.

Kirby and Sirovich tackled the face recognition issue in 2008 using principal component analysis, a classic linear algebra approach. This was a watershed moment since it demonstratedthat just a few hundred values were needed to appropriately code a properly aligned and normalised facial image. Turk and Pentland found in 2008 that the residual error might beutilised to find anomalies when employing Eigen faces algorithms. faces in photos; a breakthrough that allowed for the development of reliable real-time automatic face recognitionsystems. The public's attention was first drawn to the technology as a result of the media's reaction to a test run during the Super Bowl in January 2008. Despite the fact that the strategy was unconventional, Despite being confined by environmental considerations, it sparked greatinterest in further expansion Face recognition software that is automated.

III. SYSTEM METHODOLOGY

3.1 Model

Our typical model is a siamese convolutional neural network with L layers and Nl units for each twin, where h1,l represents the hidden vector in layer l for the first twin and h2,l represents the same for the second twin. In the initial L 2 layers, we employ only rectified linear (ReLU) units. the remaining levels have sigmoidal units A series of convolutional layers makes up the model. Each use a single channel with a variety of filters. number of convolutional layers and a fixed stride of 1 To improve performance, filters should be a multiple of 16. A ReLU activation function is used in the network. Optionally, maxpooling with a filter size of 2 and a stride of 2 is applied to the output feature maps. As a result, the kth filter is used. The map in each layer looks like this:

$$a_{1,m}^{(k)} = \text{max-pool}(\max(0, \mathbf{W}_{l-1,l}^{(k)} * \mathbf{h}_{1,(l-1)} + \mathbf{b}_l), 2)$$
$$a_{2,m}^{(k)} = \text{max-pool}(\max(0, \mathbf{W}_{l-1,l}^{(k)} * \mathbf{h}_{2,(l-1)} + \mathbf{b}_l), 2)$$

Wl−1,l is the three-dimensional tensor that represents the feature maps for layer l, and we've taken? should be a legitimate convolutional operation that returns only those output units that are the consequence of complete overlap between each convolutional filter and the input feature maps.

3.2 Data Pre-Processing

Data pre-processing is essential for grooming the data and making it suitable for the neural network. It normalizes the data, makes it consistent and suitable for the task, and also aids in enhancing the accuracy and efficiency of a machine learning model. For this model, the imageshave been resized and scaled to range between zero and one.

IV. SYSTEM METHODOLOGY

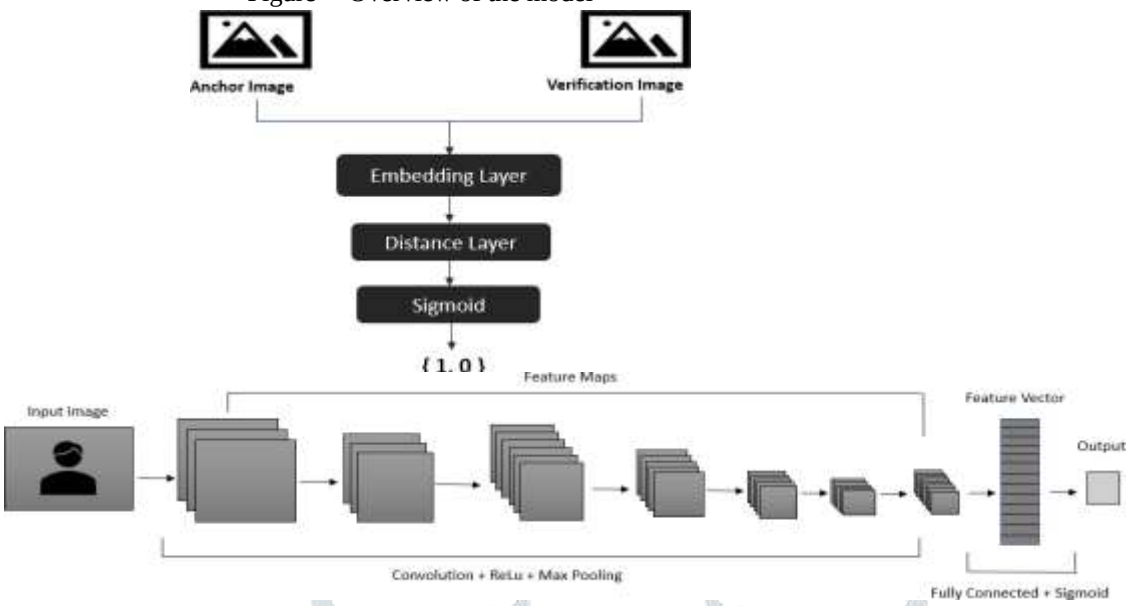
Building a deep learning model requires exploring various tasks and architectures, understanding their contribution to the performance of the model and knowing how to assemblemultiple such components along the way to implement a reliable model. Data is evidently the building block of any algorithm's rendition. A variety of tasks ranging from making simple tocomplex decisions are all data-driven. The first step to kickstart the process is data collection. For this project, we will be working with three types of image sets: an anchor image representing the image being passed to the network for verification, a negative image with the non-matching input, and a positive image marking the matching input. The anchor and positiveimages were collected from the computer's web camera input by using the functionalities of the cv2 package offerings. The negative images are scraped from the internet using a simple web scraper using python's selenium library. The collected images are organized into directories. The dataset can be constructed by configuring the labels to be 1 if anchor and positive images are presented to the model and 0 if anchor and negative images are presented to the model. Data augmentation techniques can also be applied on existing data to create a diverse range of data without having to gather new data. On the other hand, similar results canbe accomplished by using collections of datasets that are available, LFW and FFHQ datasets, or datasets proffered by TensorFlow or Kaggle.

The most important step in building more efficient learning models is data pre-processing. The goal of pre-processing is to

increase the image's quality so that we can better analyze it. To obtain extraordinary results, every data scientist should spend 80% of their time pre-processing data and 20% of their time doing analysis, according to the 80/20 rule. We can suppress unwanted distortions and boost some necessary aspects for the application we're working on via pre-processing. Those features might vary for different applications.The images collected in the prior step are of different formats, sizes, and resolutions. Neural networks take inputs of the same size; therefore, all the images need to be resized to a fixed sizebefore delivering them to the CNN. Hence, it is necessary to maintain consistency before proceeding to work with the data. Therefore, each image must be resized and scaled. The smaller the size of an image the faster the processing is; however, we will be using a 100x100 image with three channels. Apart from same input sizes, since our model heavily relies on distance calculation, the images must be scaled in order to normalize the features. This ensuresthat each feature contributes to the final metric equally.

Convolutional neural networks have outperformed traditional neural networks in large-scale computer vision applications, notably image recognition. [27] Our model is built using a Convolution Neural Network. It operates with L convolution layers, each having Nl units, where h1,l is a representation of the hidden vector in the first network's layer l. In contrast, $h_{2,l}$is similar but for the second network of the twin. We employed a sigmoid unit for the final layerand rectified linear units for the remaining layers to serve as an activation function. ReLu is simpler but effective, faster and feasible to work with as it activates only one neuron at a time. The sigmoid unit produces an output that ranges from 0 to 1. Since we aim to output the probability of an instance belonging to a class, this activation function is a proper fit.

Figure – Overview of the model



The model depicts the usage of a sequence of convolutional layers that utilises a single channelwith filters of different sizes to extract useful features and work with a stride of 1 fixed. The decision choice regarding the number of convolutional filters directly affects the level of optimisation of the model. Hence, multiples of 16 are considered. Even if convolutional layers are computationally not practical in terms of cost, local connection can minimise the amount of parameters employed in the model, which gives some sort of built-in regularisation. Each feature map is convolved against the input features to recognise and identify patterns as groups of pixels in the convolution process, which has a direct filtering interpretation. As a result, the outputs of each convolutional layer correspond to significant spatial properties in the original input space, and simple changes are more resilient. The network also applies the ReLU activation function to the output feature maps, which is optionally followed by max-pooling with a feasible filter size and a fixed stride of 2.

Figure – Model Architecture

Only those output units arising from the entire overlap between each filter and the input feature maps are returned by the 3D tensor representing the feature maps for layer l. The last layer's output is flattened into a single vector to reflect only the most important characteristics. The sigmoid unit receives the following fully connected layer, as well as the layer that computes the distance metric between each picture input.

Training neural networks require a collection of smaller steps to coordinate and execute to achieve a collective, well-performing model. It starts with creating a data pipeline which is further split into training and testing partitions. The data is processed in batches to ensure fasterand optimised network performance in combination with caching and other measures to preventbottlenecks in the neural network.

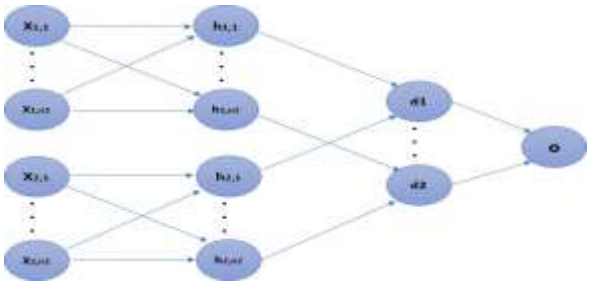


Figure – Two hidden layer network structure for binary classification

Binary Cross Entropy represents the negative average of the log of corrected predicted probabilities. Binary cross entropy compares each anticipated probability to the predicted probability of the actual class output, which can be 0 or 1, then calculates the score that is used to penalise the probabilities based on the expected value's distance metric, i.e., close or far from the actual value. Due to the connected weights, this goal is paired with a typical backpropagation technique, where the gradient is additive across the twin networks. We used an 8-person batch with a learning rate of NJ and a momentum of uj. With a learning rate of 1e-4, Adam's optimiser was employed. Adaptive moment estimate (Adam) is a method of calculating current gradients by leveraging prior gradients. This strategy makes advantage of the concept of momentum.

Further, all the network weights were initialised to be equal and normalised to render an unbiased contribution toward distance calculation. After configuring the weights, learning rate,loss function and train step for each batch of data, we begin training the network against the training data. The results depicted for the initial epochs improve as the training progresses. During training, the model makes a prediction, calculates loss, derives gradients, and back propagates the loss in terms of updating the weights. The loss function depreciates, and accuracy increases as the training proceeds forward.

Once the training process has been completed, use the testing partition of data to verify whetherthe model is producing the expected results for the given inputs or not. Check the performanceagainst a few batches of data or the entire dataset to conclude. Evaluation metrics, for instance,precision and recall can be used to visualise and validate the model's performance numerically. Precision represents the proportion of correct positive identifications according to the test datalabels, while recall shows the proportion of actual positives images that were classifiedcorrectly. The verification function is devised to make predictions for a single shot of input against a few verification images whose probabilities, in comparison with the detectionthreshold, determine the categorization of an instance. The completed model is saved and loaded whenever the verification function needs to be executed. The model is capable and self-sufficient to be able to generalize unknown instances. Hence, no extensive training is requiredeach time there is a new sample.

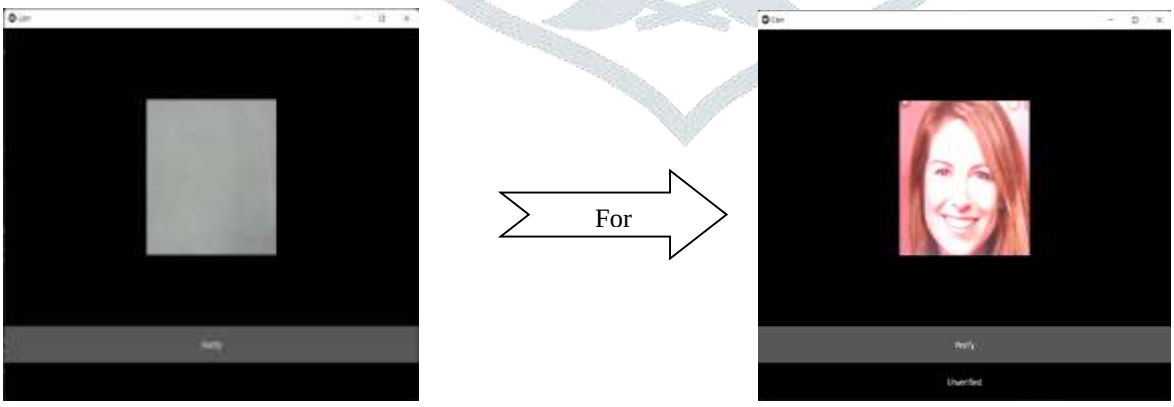
Figure – System Methodology

V. Results and discussion

The real circumstances of the situation encountered is more the degradation of the face angle, that is, the side situation, and face detection and recognition are more directly applied to life. when the face is angularly displaced. The accuracy of face identification and recognition reduces as the degree of face angle change rises. When the angle deviation is less than 20 degrees, the angular offset is considerably reduced, and the identification accuracy is greater than 30%, allowing the face to be detected and identified. However, there is a scenario when the mistake is too big. Face detection and identification accuracy drops to 20% or less when the side angle change is too great, reaching 40 degrees or more.

The circumstance in which the face is obstructed by an obstacle masking the actual appearanceof the detector exists in the market, thus face detection and recognition in the event of face occlusion has enormous research value. This method first learns through face identification and extracts features using the final hidden layer outputs, and subsequently generalizes them for face verification. This network was trained on the custom-made dataset and has demonstrated precision of 1.0 and recall score of 0.978 with improved speed, and with little amount of data. Face detection is also possible when the detector's eyes are obstructed. The five-point placement can be clearly accomplished and the picture and backdrop can be split more properly when the nose and mouth are blocked.

VI. Final Output:



VII. Conclusion and feature scope

The presented model and the strategy for performing one-shot image classification demonstrated high levels of accuracy, faster speed, and optimised performance in terms of costand time taken for training. Beginning with learning deep convolutional neural networks for verification purposes, the new results are outlined by comparing the performance of our network against the non-optimised network architecture. According to the metrics and verification results, the model outperformed the existing baselines by a notable margin and demonstrated good performance. With this metric learning approach, human-level accuracy ispossible with the substantial performance of these networks. A limited number of images is sufficient to generalise any new instance along with known samples of data the model is trainedusing. The model requires no rigorous training to classify unknown examples as the distance metric is capable of identifying the differences between the input image and the existing fixedset verification images.

With the development in technology and the advancements in the area of deep learning, face recognition technology has yet more milestones to cross in practical application. Imagerecognition is excelling in diverse tasks for a wide range of sectors like health care, security, voting, etc. Juniper Research states that facial recognition hardware will be the fastest-growingmeans of mobile biometric hardware, to over 800 million in 2024 from 96 million in 2019. There are millions of doors yet to be opened in domains where this technology may excel. Forinstance, authentication in areas like payment, orders, online examinations, interviews, etc. This one-shot, cost and speed optimised model definitely has room for growth in technological aspects, hardware integration.

VIII. References

[1] Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L. (June 2014). "DeepFace: Closing the Gap to Human-Level Performance in Face Verification". 2014 IEEE Conference on Computer Vision and Pattern Recognition: 1701–1708. doi:10.1109/CVPR.2014.220. ISBN 978-1-4799-5118-5. S2CID 2814088.

[2] Li, Lixiang et al. “A Review of Face Recognition Technology.” IEEE Access 8 (2020): 139110-139120.

[3] Office, U. S. Government Accountability. "Facial Recognition Technology: Federal Law Enforcement Agencies Should Have Better Awareness of Systems Used By Employees". www.gao.gov. Retrieved September 5, 2021.