Sketch Recognition Using Siamese Network

Pooja Sadarangani[1], Prachi Sadarangani[1], Vatsal Gupta[1], Mrunalini Pimpale[2]

1. U.G. student, Electronics and Telecommunication Department, DJSCE, Vile-Parle (W), Mumbai –400056.
2. Asst. Professor, Electronics and Telecommunication Department, DJSCE, Vile-Parle (W), Mumbai – 400056.

Abstract: This document proposes the design of an efficient and easy system to identify individuals by comparing outdated hand-drawn or computerized sketches with recent pictures to predict whether they are of the same person. The system makes use of Siamese network to understand and learn the facial appearance patterns us it to match newer inputs. This system will be extremely useful in the criminology industry where identification of criminals is required every day. Often there are times when criminals repeat their crimes but despite having their information stored in the database, the officers are unable to use it because they don't have proper resources and systems to extract that information. The proposed system will expedite this process and will save up on a lot of time. In this paper, we use a holistic approach instead of a facial recognition of hybrid recognition approach because the federal authorities only have the outline of face and not the actual features.

Key Words: Siamese Network, Holistic Recognition, Deep Learning, Convolutional Neural Network, Siamese Network, Data Augmentation.

I INTRODUCTION

A human cerebrum can store and recollect a huge number of faces in an individual's lifetime; however, it is hard for an automated machine to produce similar outcomes. Faces are intricate and multidimensional making extraction of facial features to be very difficult, still it is fundamental for our face recognition frameworks to outperform our mind's abilities. Every individual is distinguished by their face, as well as a variety of physiological biometrics such as unique fingerprints, hand geometry, retina, iris, and ear. A significant body of evidence supports the idea that the human visual system processes a face as an integrated perceptual whole rather than a series of separable facial features. Hence, in this paper we centre predominantly around a holistic approach on sketch recognition for forensic use. This methodology to sketch recognition eliminates many of the challenges posed by facial features and hybrid recognition. Holistic face recognition utilizes global data from faces to perform face recognition on the worldwide information from faces is essentially delineate by a small variety of options, that are directly derived from the constituent information of face images. These small variances unmistakably catch the fluctuation among various individual appearances and in this manner are utilized to interestingly recognize people.

Facial sketches are commonly used by federal authorities to aid in the detection and capture of suspects involved in illegal activity. Sketches used in forensic investigations are either drawn by forensic artists or generated with computer software following the verbal description given by the eyewitness or the victim. The project focuses on the recovery of face sketches for easy identification, detection and recognition of suspects by comparing them to an existing database of sketches. This will produce much accurate outcomes. Since, a hand-drawn or computerized sketch only represents the outline of the face rather than actual face features, a holistic approach for recognition is well suited in such scenarios. What makes forensic sketch matching a holistic task even additional so, is that the undeniable fact that often the bottom truth image within the information of a criminal becomes obsolete because of ageing. For instance, such a case may emerge in which the criminal commits a crime again after 30 years wherein, a feature-based approach would inevitably perform poorly. This demands for a mechanism for processing global face information resembling different facial characteristics to search out an appropriate match.

In today's era of deep learning, neural networks are almost good at every task. However, their drawback is that they depend immensely on large quantities of data to perform well and make accurate predictions. For face recognition purposes, it is difficult to gather an extensive dataset for training and testing purposes and thus this approach isn't the most ideal approach. Nonetheless, this drawback is eliminated via Siamese Networks, which are a class of neural networks that are capable of one-shot learning. They can make good predictions using few images (data) and this unique feature of them has made them useful for face and sketch recognition purposes.

In the traditional neural network approach, multiple classes are predicted and therefore this method fails when new classes are added, or some classes are removed. Siamese networks overcome this problem by understanding and learning similarity function, thereby allowing us to compare two images and check whether they are similar or not.

Hence, we are using Siamese Network to garner information concerning the worldwide linguistics variations between classes, that may additional accurately represent the semantic distance between different categories.

II LITERATURE SURVEY

In reference to paper [1], they proposed a convolutional neural network based on Siamese network for sketch-based image retrieval. The aim was to bring output feature vectors closer together for input sketch-image pairs that were classified as identical and drive them apart otherwise. They achieved this by integrating two convolutional networks with a single loss function. The proposed method outperformed other methods, according to the results of the experiments. This improved the efficiency of sketch-based image retrieval by taking into account the complexity of sketches and addressing the problems they cause.

In reference to paper [2], they proposed a content-based image retrieval (CBIR) approach for law enforcement fields which could be used to align forensic sketches with digital human face. One of the most useful technologies for researchers is the CBIR technique, which can help them distinguish possible evidence from a large number of digital images. It looks for related image collections with characteristics such as case(s) of interest. In this paper, along with proposing a CBIR method for face recognition, the authors have also investigated and outlined the problems face sketch recognition systems encounter.

In reference to paper [3] Deep Neural Networks (DNNs) have as of late beat traditional object recognition on numerous large-scale datasets, for example, ImageNet. For the model trained on ImageNet, the information source is overwhelmed by photographs and a wide range of representations are generally marked as 'animation' instead of their own sorts (e.g., 'feline'). Due to this reason the traditional approach fails at recognizing objects. The vast majority of sketch recognition strategies still intensely depend on the hand-make feature extraction procedures, in this way it is fascinating to know whether DNNs can kill the necessities of explicit element designing here, and how the engineering is intended for sketch recognition. Based on the Deep Neural Network model for classification of sketches which has outperformed state-of-the-art results in the TU- Berlin sketch benchmark, they have proposed a sketch image retrieval system.

In reference to paper [4], intends to increment the matching rate utilizing the Siamese convolution network architecture. In order to decrease the modality gap, the system extracts useful features from each image pair. In addition, data augmentation is utilized to dodge overfitting. They have investigated the exhibition of three loss functions and analyzed the closeness between each image pair. The results showed that their system is sufficient for a composite sketch dataset. Likewise, it diminishes the impact of overfitting by utilizing data augmentation and changing the organization structure Criminal suspect images cannot be taken immediately at the crime scene, so this technology has a great potential in criminalogy. The police creates a hand-drawn face sketch picture or a composite sketch based on eyewitness descriptions and matching it to datasets using software. However, computers fail to differentiate between photos and sketches as their visual information is different. Due to this, computers weren’t ideal for the application and thus the authors have proposed an approach using deep learning to achieve the goals.

In paper [5], they introduced a method sketch-specific data augmentation (SSDA) approach, which automatically utilises the quantity and quality of the sketches. They've introduced a Bezier pivot dependent deformation (BPD) technique to enhance training datasets from a quantity standpoint. They pose a mean stroke reconstruction (MSR) method to developing a collection of new kinds of sketches with smaller intra-class variances in order to improve the quality. These solutions are unregulated by multi-source data or sketch temporal cues. Furthermore, they have shown that some recent models trained on universal categories of image data are superior to most extensive architectures built specifically for sketch recognition. SSDA has significant benefits over conventional approaches because it can be combined with any convolutional neural network. On the TU-Berlin dataset, their detailed experimental analyses show that the proposed method produces state-of-the-art results (84.27 percent), surpassing human output by a significant 11.17 percent increase. They also introduce Sketchy-R, a new benchmark for potential research in sketch recognition.

In reference to paper [6], the author focused on a particular facial geometric function that could be used to measure RATIOS of similarity between the reference photograph database and the forensic sketches. The framework of the system is based on computer vision such as Two-Dimensional Discrete Cosine Transform (2D-DCT) and the Self-Organizing Map (SOM) Neural Network simulated in MATLAB. The cropped facial features of the image database, such as the frontal face, left eye, right eye, nose, and lips, are compressed using the 2D-DCT image compression technique. The image pixels are reshaped after compression to prepare the image classes as an input for the neural network. For the training of image data, the SOM (Self Organizing Maps) neural network algorithm is designed. During the training and learning in simulink for various numbers of epochs, the unsupervised weight is allocated to identify and inputted face sketch. The design of facial recognition systems is the basis for this project. MATLAB and Simulink are used to run the software source code and simulation.

Paper [7] provides Using touch screen applications, a method for providing access to visual information based on user-drawn partially coloured sketches. The key problem for sketch-based image retrieval systems is to solve the difficulty involved with drawings cause by lack of colours, textures, shading, and drawing flaws. To address these issues, they suggest fine-tuning a CNN in a sketch-based image retrieval system using the augmented dataset to extract features from partially coloured hand-drawn query specification sketches. Natural photographs, edge maps, hand-drawn drawings, de-colored and de-texturized images make up the enhanced dataset, which allows CNN to accurately predict visual content delivered to it in a number of ways. Deep features extracted from CNN allow image retrieval using both sketches and full colour images as queries. They also looked at whether partial colouring or shading in sketches could help with retrieval. The proposed method was tested on two large datasets for sketch recognition and it outperformed several existing methods.

III BACKGROUND

Deep Learning - Deep Learning is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called artificial neural networks. A computer model learns to perform classification tasks directly from pictures, text, or sound in deep learning. Deep learning models can achieve state-of-the-art precision, even surpassing human output in some cases. Models are trained using a wide collection of labelled data and multilayer neural network architectures. Deep learning models are trained by using large sets of labeled data and neural network architectures that learn features directly from the data without the need for manual feature extraction.

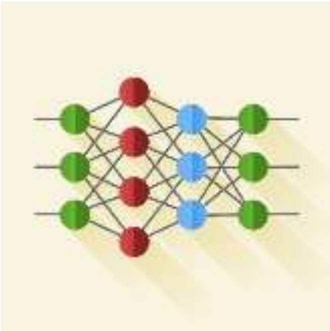

Fig 1. Deep Learning

Siamese Network - The Siamese Neural Network is a subset of neural network architectures that have two or more similar subnetworks. 'Identical' here implies that they have the same configuration with the same parameters and weights. The parameter update is replicated in all sub-networks. This is utilized to determine the similarity of inputs by comparing their vector attributes, such that these networks are used in many applications.

Architecture:

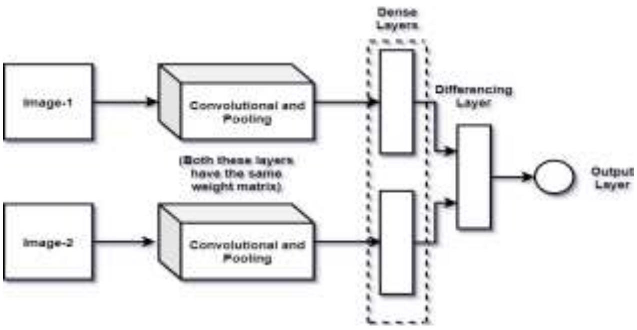

Fig 2. Siamese Network

For CNN, there is a series of layers - convolution layer and pooling layer. These layers are followed by dense layer and outcome layer presumably with a SoftMax function. Convolution layer is answerable for feature extraction of the image while SoftMax layer is liable for the range of likelihood for each class. The neuron with the most probability will then be chosen as a class of the image. In case of Siamese networks, the convoluting and pooling layer working remains similar but there is no SoftMax layer present. Thus, we stop at the dense layer. Now, since we have 2 inputs, there will be two dense layers. We will find the difference between these two layers and yield the output as a single neuron with sigmoid activation function (0 to 1). Accordingly, the training data to this network must be organized so that there is a rundown comprising of two images and a variable either 0 or 1. Since training of Siamese networks involves pairwise classification at the output, it classifies whether the output is of the same class or not. Therefore, the training can use different loss functions. We will be using cross entropy loss in this project. Cross entropy loss can be calculated as:

$$L = -y \log p + (1 - y) \log(1 - p)$$

Where,

L = The loss function

y = The Class Label (0 OR 1)

p = Prediction.

To train the network to differentiate between similar objects and dissimilar ones, we provide it with a positive and a negative example at a time and calculate the sum of losses.

Data Augmentation - In data analysis, data augmentation refers to techniques for increasing the amount of data by inserting slightly changed versions of existing data or creating new artificial data from existing data. When training a machine learning model, it serves as a regularizer and helps to minimise overfitting.

IV WORKING

The dataset that we will be using for this project is CUHK Face Sketch database (CUFS) created for research on face sketch synthesis and face sketch recognition. For this project, we will be using CUHK database which consists of 25 faces. However, the existing sketch datasets used to fine-tune the pre-trained CNN models are much smaller than their real image counterparts and this gives rise to overfitting due to which the recognition accuracy in test sets and new data is low. Moreover, the network fails to map sketches that are skewed, zoom or distorted. This is undesirable as it is impossible to draw the exact same sketch each time and subtle variances in sketches are bound to occur. Hence, to tackle this conundrum, we will be incorporation data augmentation in our project.

Data augmentation techniques such as cropping, padding, and horizontal flipping are commonly used to train large neural networks. We will be performing the following augmentations to each sketch:

1. Random zoom between 130% to 150%

2. Random skew of random magnitude

3. Random distortions within a 4x4 window of varying magnitude

4. Random crops containing between 80-90% of the image

5. Random crops and skews

6. Random zoom and skews

Therefore, we will have 6 augmented images for each sketch. So that total number of sketches that we will have after data augmentation will be: 25+ 6*25=175.

For Data Augmentation, we will be using FaceApp. Augmentation is done on one image from the CUHK database and it is:

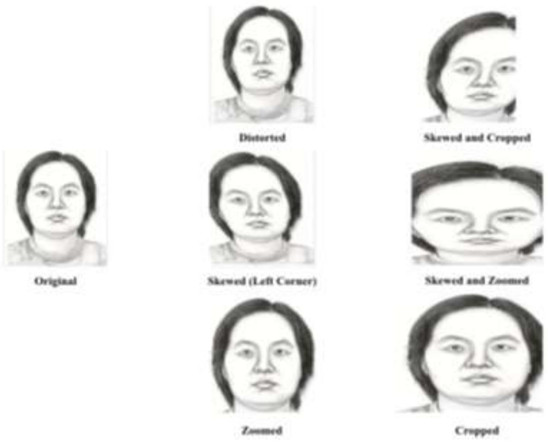

Fig 3. Data Augmentation

The Siamese network is used to extract features from two images using two identical neural networks as two channels and correlate the resemblance after training using a loss function. The Siamese network's input form is an image dual, which comprises of two images and a mark. Each image must be composed of a pair of images that are paired with other images. The network eliminates the issue of overfitting by providing an easy method of comparing similarity. To minimise the gap between features, the proposed neural network design uses two identical convolutional networks as channels that accept different images as samples and share probabilities to extract the features from face photos and sketches, respectively. After the data has been prepared in this way, we mix it and set aside 20% of it as a testing dataset before starting the training. We use the binary cross entropy loss function and Adam as our optimization method so our marks are just ‘1' and ‘0' for positive and negative matches, accordingly.

The learning rate is 0.0006 after 1024 epochs of training. The batch size and regulation will be varied as the hyperparameters and only the best result for each batch will be saved.

V FUTURE SCOPE

1] The automated extraction of images of suspects from criminal records, which can help police narrow down possible suspects rapidly, is a valuable application of face recognition. This is a primary application which uses the sketches of the accused and match them with the photos in the federal database of the offenders. There are a few limitations to the system that this project has eliminated. For instance, facial features change with age and if the offender recommits the crime after 30 years in such case the system tries to find a match with the old picture in the database it might now show the expected results. We overcame this limitation using Siamese network by training the model with a learning rate of 0.0006 with Adam Optimizer for about 1000 epochs. This makes it a suitable system for the given application.

2] Applications in fields like police departments, video monitoring, banking, and security device access authentication are in high demand. In recent times, automatic face recognition has gotten a lot of press. due to its ease of use and advantages over other measures. The The benefits of facial identification over other approaches, such as fingerprint recognition, are largely due to the fact that it does not need the cooperation of those being tested. Face recognition systems are often more user-friendly and cost-effective, so recognition outcomes can be reversed in ambiguous situations by people with little to no experience. With the digitization of the world, everything is switching online. In this scenario, it is of utmost importance that a suitable face identification method is incorporated, particular in sectors like online banking (For instance in digital banking customer onboarding process). Our system with further enhancements and modifications will be suitable in such platforms in the future.

VI CONCLUSIONS

This paper presents a groundbreaking face sketch recognition algorithm. The use of Siamese network and Convoluted Neural Network is proved to be an effective approach for face recognition and comparing the face and the sketch. The use of data augmentation on CUHK Face Sketch database (CUFS) enabled the use of modified or distorted images. The training of the network with learning rate of 0.0006 with Adam Optimizer for about 1000 epochs overcame the limitation of identifying aged sketches of the same person. These features in the algorithm makes it efficient to be applicable for practical purposes and gives a step above other algorithms available.

VII REFERENCES

[1] Yonggang Qi; Yi-Zhe Song; Honggang Zhang; Jun Liu, 2016, “Sketch based image retrieval via Siamese Convolutional Neural Network”

[2] Salah Eddine Lahlali; Abdelalim Sadiq; Samir Mbarki, 2016, “Face Sketch Recognition System: A content based image retrieval approach”

[3] Yongxin Yang & Timothy M. Hospedales, 2015, “Deep Neural Networks for Sketch Recognition”

[4] LIANG FAN, HAN LIU, YUXUAN HOU, 2019, “An improved Siamese Network for face Sketch Recognition”

[5] Ying Zheng, Hongxun Yao, Xiaoshuai Sun1, Shengping Zhang2, Sicheng, Zhao3, Fatih Porikli, 2019, "Sketch-Specific Data Augmentation for Freehand Sketch Recognition"

[6] VINEET SRIVASTAVA, 2020, "Forensic Face Sketch Recognition Using Computer Vision"

[7] Jamil Ahmad, Khan Muhammad, Sung Wook Baik, 2017, "Data augmentation-assisted deep learning of hand drawn partially coloured sketches for visual search"

[8] Abbas, Z.-A. and Duchaine, B. 2008. The role of holistic processing in judgments of facial attractiveness. Perception, 37: 1187–1196. doi:10.1068/p5984.

[9] Karamizadeh, Sasan & Abdullah, Shahidan & Zamani, Mazdak. (2013). An Overview of Holistic Face Recognition. International Journal of Research in Computer and Communication Technology. 2. 738-741.

[10] Liang Fan, Han Liu, Shenzhen University. July, 2009. Yuxuan Hou An Improved Siamese Network for Face Sketch Recognition. doi: 10.1109/ICMLC48188.2019.8949231

[11] Sumit Chopra, Raia Hadsell, Yann LeCun. Learning a Similarity Metric Discriminatively, with Application to Face Verification

[12] Siamese capsule networks with global and local features for text classification. Yujia Wu, Jing Li, Jia Wu, Jun Chang, Neurocomputing, January 2020.