FACE PHOTO-SKETCH RECOGNITION USING DEEP LEARNING TECHNIQUES - A REVIEW

Poornima E[1], Pragisha K[2]

[1]*MTech student, Dept. of computer science and engineering, LBS Institute of Technology for Women, Trivandrum, Kerala, India*

[2]*Assistant Professor, Dept. of computer science and engineering, LBS Institute of Technology for Women, Trivandrum, Kerala, India*

Abstract - *In recent years, face photo-sketch synthesis has considerably improved due to the development of deep convolutional neural networks. The technique known as FACE photo-sketch synthesis, which aims to produce a face sketch from a face photo and vice-versa has drawn a lot of attention and found useful in both law enforcement and digital entertainment. Facial image sketches are frequently created using applications for social networks and short videos for entertainment purposes. This problem has been addressed in a variety of ways, all of which have proven successful. However, there are still problems that need to be fixed in apps that are used in the real world. Face recognition (FR) makes it nearly impossible to directly utilize a sketch to search for the appropriate photo in the gallery due to the glaring contrasts between the two face domains. Transferring the faces into the same domain and using an off-the-shelf high-performance face verifier, which demands that the produced faces keep their distinguishing characteristics, is one way for identifying the target person. Numerous Face Recognition systems have been developed in recent years, but these systems still need to be improved in a number of aspects, including availability, performance, efficiency, and accuracy. Here, a comparison of various methods being developed to accurately detect the Face is being done.*

Key Words: **Deep learning, Face sketch synthesis, Generative Adversarial network, Convolutional neural network.**

1.INTRODUCTION

A cross-modality recognition technique called photo-sketch image recognition uses a face sketch as input and searches a face photo dataset for a matching photo. A face sketch is frequently the only description available when a person's appearance wasn't captured on film. This technology also helps in the search for the missing person.

Because pictures and sketches have different features, it is not possible to directly apply face recognition methods for face sketch recognition.

The challenge of matching facial images of the same person known as FACE recognition has evolved significantly with the emergence of deep learning. In the features that are recovered through a number of hidden layers, deep convolutional neural networks (DCNNs) contain representative information that is utilized to distinguish a person. However, it can be challenging toidentify a face from its distinguishing traits when the environment, lighting, or facial expression has changed.

Through FACE sketch synthesis, a target sketchface is meant to be generated from an intake sketch face image, or vice versa. It serves a variety of purposes in real life. Law enforcement can identify the offender without a monitor by using the face sketch made from witness descriptions. Deep learning algorithms have recently been applied to the recognition of faces in photo drawings in order to find similar qualities in a common context. Some of the methods convert the images into the same modality by synthesizing fake images from sketches and using GAN networks to extract the deep picture features from multi- channel neural networks like Siamese networks and triplet networks.

To anticipate the face using sketch and color photos, numerous machine learning and deep learning-based techniques or algorithms have been created during the past few years. In spite of this, the current systems still need to be improved in a number of areas, including availability, performance, efficiency, and accuracy. In this study, various cutting-edge methods being developed to accurately detect people's faces arebeing compared. The rest of the paper is organized as follows: Section 2 presents various studies that address a particular issue that is taken into consideration in the study, and Section 3 compares the outcomes of these various techniques. The full list of references is followed by remarks on the conclusion in Section 4.

2. LITERATURE REVIEW

This section provides a summary of the studies that used Deep Learning (DL) and numerous other efficient and creative techniques to identify faces in photo-sketch images. Due to the rising number of criminal cases worldwide, researchers were motivated to create a way to identify faces using either color or sketch images and inform the public. Numerous strategies have been used in the past and are continuously being examined. Below is a brief summary of some of the earlier works that were taken into consideration.

2.1 FEATURE INJECTION AND ICNN FOR FACE RECOGNITION

1. They provide FI as a way of keeping identifiable information in face sketches and photo-synthesis. This is the main study object in the work.

2. A lightweight version of ICNN with two domain-specific decoders for simultaneously creating sketches and images and a shared encoder for obtaining general input and removing domain-specific features were shown.

3. According to evaluation results, the approach provides a practical solution for the use of face sketches and image synthesis.

The middle convolutional layers of a trained CNN's feature maps retain the multilayer texture information from the original image[1]. The various feature maps produced by a well-trained CNN can be viewed as the approximations of the eigenvectors clustered in various regions that can be classified using the same attributes in the same feature space. The suggested method is put into practise using GANs[13]. The generator is the proposed ICNN, which includes a shared encoder, two identical decoders, and can simultaneously generate sketches and images. Due to the wide modality difference between a set of sketches and images, domain-specialized information may cause the synthesis's performance to degrade. The shared encoder is designed to obtain common latent properties of paired faces and lessen domain interference, allowing the two specialized decoders to concentrate more on their goal generation. Instead of using deconvolution to increase the size of feature maps, as Odena et al. [14] did, they use interpolation and a convolutional layer, which can reduce artifacts for generated images. To further reduce the model, the encoder and decoder designs are identical. They just used a small convolutional filter with size-3, padding-1, and stride-1 to up and down-sample thechannel dimension, no further techniques were used. Both the encoding and decoding stages use FI for revision after sampling. Thus, a method for simultaneously producing high-quality sketches and photos while preserving one's identity is presented in this proposed work. It consists of two essential parts.

The first is the hypothetical FI, which adds auxiliary traits to the synthesis procedures so that the resulting faces can have more identification data. The other is the theoretical ICNN, which has a simple architecture composed of two domain-specific decoders and a common encoder. It used interpolation and minute convolutional filters, respectively, to change the size and channel dimension of feature maps[1]. The goal of the common encoder was to reduce the particular interference between the two face domains so that the domain-specific decoders could effectively complete the synthesis. Evaluations showed that this approach outperformed cutting-edge methods without regard to identity preservation or visual quality, demonstrating that it provides a practical alternative for the employment of face sketching and photosynthesis in the real world. Consequently, 73.78% accuracy was attained.

2.2 IDENTIFICATION OF A FACE USING KNOWLEDGE DISTILLATION

This work's primary contributions are:

1. This suggested KD model introduces the face photo-sketch synthesis job and performs well even with limited training data.

2. A network architecture is created that is efficient and investigates three forms of knowledge transfer loss in order to effectively distil and transfer knowledge.

3. They investigate a KD+ model by combining GANs with KD in order to further improve the perceived quality of synthetic images. 4. This model outperforms state-of-the-art methods, according to extensive testing and a user study[2].

They provide a KD framework in this study for the purpose of making face photo-sketches. Based on KD, they develop two models: KD and KD+. The KD approach can create effective student networks on little amounts of training data by transferring information from a more thorough and accurate instructor network trained on sufficient training data. In addition to absorbing knowledge from the instructor network, the student networks can exchange their own knowledge to enhance learning. By combining GANs with KD, the KD+ model can generate images with higher perceptual quality. They compare the proposed models to the most recent state-of-the-art methods on two open databases. Results from qualitative and quantitative evaluations demonstrate that the proposed technique yields considerable advantages. Consequently, 88.77% accuracy was achieved [2].

2.3 REGENERATION OF HETEROGENEOUS FACE USING RELATIONAL MODULE

The following are the main contributions to this work:

• The graph-structured module RGM, which describes face components as node vectors and relational information edges, is available to bridge the fundamental domain gap.

It is also re-calibrated by accounting for global node correlation via NAU.

• A C-softmax is suggested that effectively projects traits from diverse domains into a common latent space by conditionally exploiting the inter-class margin.

• The suggested module can overcome the restrictions of the HFR database by hooking into a universal feature extractor. It empirically demonstrates improved performance for five HFR databases and three different backbones. The Relational Graph Module (RGM) is able to extract the representative relational information of each identity by embedding each face component into a node vector and modeling the interactions between them. A structured approach based on extracting relations was employed in this graph-structured module to address the discrepancy problem between HFR domains. Furthermore, by linking to and optimizing a pre-trained face extractor, the RGM overcame the problem of an inadequate HFR database[4].

Additionally, to focus on globally informative nodes among propagating node vectors, node-wise recalibration was done using the Node Attention Unit (NAU). Furthermore, the creative C-softmax loss helped with the adaptive learning of common projection space by using a larger margin as the class similarity increased. They investigated performance improvements utilizing the RGM module on a variety of pre-trained backbones for the NIR-to-VIS and Sketch-to-VIS tasks. Each recommended tactic also showed how its function had an impact on performance in tests with ablation. The display of relational information in VIS, NIR, and sketch pictures, which reveals representative domain-invariant features, further demonstrates that relationships within the face are comparable in each individual. On the CASIA NIR-VIS 2.0, IIIT-D Sketch, BUAA- VisNir, Oulu-CASIA NIR-VIS, and TUFTS, this suggested strategy fared better than cutting-edge methods. It is also used to connect to an universal feature extractor in order to get over the HFR databases limitations. As a result, this method obtained an accuracy of 95.65%.

2.4 DOMAIN ALIGNMENT EMBEDDING NETWORK FOR FACE RECOGNITION

This study proposes a deep metric learning method termed a domain alignment embedding network for sketch face recognition. The limited sample problem is addressed by a method known as the training episode technique, which is based on few-shot learning methods [6]. A small number of training episodes are randomly selected from the training set in order to mimic few-shot jobs, and after that, domain-specific query sets and support sets are further built to incorporate domain knowledge. Based on the recommended training episode technique, a domain alignment embedding loss isproposed to guide the feature embedding network while it learns discriminative features. The domain alignment embedding loss is created by combining the sketch domain embedding loss and the photo domain embedding loss[10]. According to the sketch domain embedding loss, photo features in the support set and query set that belongto the same class will be close to one another while those that do not will be far apart. The picture domain embedding loss predicts that photo features in the query set and sketch features in the support set should be close to one another if they belong to the same class, but they should be far apart if they belong to separate classes. The recognition of sketch faces was thus shown here using a domain alignment embedding network (DAEN).

To represent new tasks, a small number of training episodes were randomly selected from the training set, and domain-specific query sets and support sets were made to include domain expertise. A domain alignment embedding loss has been constructed for each training episode utilizing the domain-related query set and support set in order to learn discriminative features. Several test scenarios were conducted using the datasets from the PRIP-VSGC and UoMSGFSv2 projects. The effectiveness of the suggested training episode design and domain embedding loss has been demonstrated by ablation research. This method significantly surpasses existing sketch face recognition techniques, according to comparisons with a number of intra-model and inter-model methodologies. Using this model 74% accuracy was achieved.

2.5 CNN IS USED FOR FORENSIC FACE PHOTO SKETCH RECOGNITION

Scale-invariant feature transform (SIFT) and multi-scale local binary pattern (MLBP) are two innovative algorithms that make use of hand-crafted features [7]– [6]. Given that these qualities were not intended for inter-modality face identification, it would be preferable to utilize descriptors that are more suited for the task of identifying faces from photo sketches [8]. Deep learning is a technique that may be used to learn pertinent descriptors and has shown effective in a variety of disciplines, including traditional face recognition[9]-[10]. The recognition of faces in photo sketches hasn't seen significant usage of deep learning, though[3].

One of the main reasons for this is that there aren't many publicly available photo-sketch pairs, despite the fact that numerous instances are needed to train deep networks successfully and avoid issues like over-fitting and local minima. Additionally, a deep network finds it difficult to learn robust characteristics because there is frequently only one sketch per topic. This work's contributions are:

1. To overcome the "one drawing per subject" problem, a three-dimensional (3-D) Morphable model is used to alter facial traits and automatically synthesize a new large collection of pictures.

2. 2. A state-of-the-art deep network (trained on face photographs) is optimized for the task of recognizing faces in the photo- drawings utilizing the synthetic images by using transfer learning.

3. Synthetic sketches sometimes resemble the corresponding photo more closely than the actual sketch because forensic sketches might contain a variety of inaccuracies. Actually, by comparing numerous sketches of each subject with gallery photographs, performance is improved.

4. It is shown that the suggested architecture performs even better when coupled with a top algorithm on both visible and forensic sketches.

The 3-D Morphable model used by the face photo-sketch identification system presented in this study to automatically create fresh photos eliminates the problem of having only one drawing per subject. This enables a deep convolutional neural network to surpass current techniques by learning the connections between pictures and sketches. It has also been shown that merging multiple sketches during testing enhances performance for real-world draws because variations in facial features may result in sketches that are more similar to the matched photo than the original sketch. This is one of the few pieces of art that uses deep learning to identify topics from several sketches and hand-drawn facial sketches.

The ideas were also discovered to function effectively for genuine forensic sketching. It was further demonstrated that the proposed approach performs better when coupled with different methodology for both observed and forensic sketching. Future work will involve employing a more complex Morphable model that allows for more change in the facial features and applying the recommended approaches to other HFR challenges. As a result, this approach achieved an accuracy of 80.7%.

3. RESULTS AND DISCUSSIONS

In order to evaluate the algorithms under discussion on actual images, the PRIP-HDC dataset [6], which contains hand-drawn forensic sketches of 47 individuals created from eyewitness evidence in genuine investigations, is used. The mug shots became available when the suspects depicted in the sketches were located. All people were solely used for testing using the same models that were trained on the visible hand-drawn sketches because of the small size of the dataset.

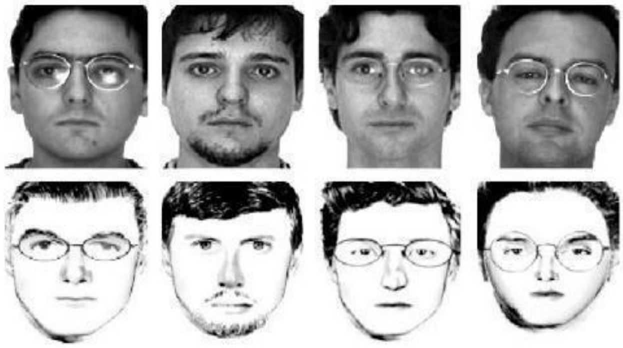

Fig -1: Example images from the PRIP-VSGC database[7].

The ranks at which each subject is retrieved by algorithms are used for analysis right away because typical performance assessments may produce inaccurate results. There are other outcomes in the Supplementary Material as well.

Dataset	Rank 1	Rank 5	Rank 10	Rank 20	Rank 30
CUHK	23.5	40.5	71.3	83.4	91.1
FEI	26.8	39.6	63.3	77.0	85.5
CASPEAL	21.8	28.8	41.5	50.8	61.6
FERET	20.1	27.0	33.6	42.2	54.1
MGDB	39.8	73.33	85.1	87.4	89.2
SCface	22.8	38.7	51.7	69.9	82.8
AR	31.6	55.6	74.2	81.2	93.0

Table-1: Identification accuracy (%) on privatesketch datasets in different ranks[11].

3.1. PERFORMANCE METRICS

Accuracy (AC), recall (RC), and precision (PR) are a few performance metrics that are used to evaluate how effective the suggested models are. The interpretations of these measurements are as follows:

1. **True positive (TP)**: Accurately identified a true color or accurate sketch as such.

2. **True Negative (TN)**: Accurately identified a color or sketch as being false.

3. **False positives (FP):** are images that aren't in color or are sketches but are mistakenly identified as such.

4. **False Negatives (FN):** are coloured or hand-drawn images that have been wrongly categorized as neither coloured nor hand-drawn.

For a deeper examination of system performance, these data can be expressed as a confusion matrix, as seen in Fig 2.

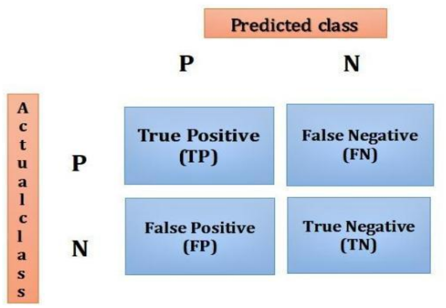

Fig -2: Confusion Matrix sample.

The following are the mathematical formulas used to calculate the model performance analysis measures accuracy, precision, and recall:

$$\text{Accuracy(AC)} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Precision (PC)} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall (RC)} = \frac{P}{P+N} \quad (3)$$

The results obtained for each of the Face detection algorithms taken into consideration for the study are summarized in Table.2. The accuracy percentage element included in performance analysis is the only feature taken into account for an overall analysis of the approaches to make comparisons between them easier.

The compiled statistics make it clear that the method used in the model [4] achieved the highest accuracy percentage, 95.65%, while the model [2] earned the second-highest accuracy, 88.77%. Nearly every other state-of-the-art model taken into consideration was found to have acceptable accuracy levels ranging from 70 to 95 percent. The method outlined in the model [12] demonstrated the least accuracy, at 73.7%.

4. CONCLUSIONS

In this study, a comparison of various state-of-the-art methods being developed to accurately detect a person's face is being done. Despite the fact that many methods and systems based on machine learning and deep learning have been developed over the past few years to predict the face photo sketch using mainly the datasets CUFS,CUFSF, and PRIP, the majority of the current face detection methods only accept either colour or sketch images as input.

MODEL	JOURNAL	DATASETS	METRICS
Relational Graph Module(RGM)	Relational Deep Feature Learningfor Heterogeneous Face Recognition	Oulu-CASIA NIR- VIS, TUFTS NIR &VIS	Accuracy-95.65
Knowledge Distillation(KD)	Knowledge Distillationfor FacePhoto–Sketch Synthesis	CUFS, CUFSF	Accuracy-88.77
Convolutional Neural Network(CNN)	Forensic Face Photo-SketchRecognition Using a Deep Learning-Based Architecture	PRIP-HDC	Accuracy-80.7
Domain Alignment Embedding Network(DAEN)	Domain Alignment EmbeddingNetwork for Sketch Face Recognition	UoM-SGFSv2, PRIP	Accuracy–74.0
Feature Injection(FI)	Toward Identity Preserving Face Synthesis Between Sketches and Photos Using Deep Feature Injection	CUFS, CUFSF	Accuracy-73.7

Table-2: Summary of different models

The current systems still need to be improved in a variety of areas, including availability, performance, efficiency, and accuracy. In this study, various methods were examined and contrasted with one another. Different strategies for performance improvement could be investigated as far as future improvements to the current aspects of the methodologies for this study that were taken into consideration. Additionally, the features that are taken into account for detection can be increased.

Future work will involve employing a more complex Morphable model that allows for more change in the facial features and applying the recommended approaches to other HFR challenges. By modifying the system to accept both color and sketch images as input, it will be possible to recognize the related face by converting the color image to the sketch image using an image synthesis method.

REFERENCES

[1] Y. Lin, K. Fu, S. Ling, J. Wang and P. Cheng,"Toward Identity Preserving Face Synthesis Between Sketches and Photos Using Deep Feature Injection," in *IEEE Transactions on Industrial Informatics*, vol. 18, no. 1, pp. 327-336,Jan.2022,doi: 10.1109/TII.2021.3074989.

[2] "Knowledge Distillation for Face Photo–Sketch Synthesis," in IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 2,pp.893906,Feb.2022,doi:10.1109/TNNLS.2020.303 0 536.

[3] C. Galea and R. A. Farrugia, "Forensic Face Photo-Sketch Recognition Using a Deep Learning-Based Architecture," in IEEE Signal Processing Letters, vol. 24, no. 11, pp. 1586-1590,Nov. 2017, doi: 10.1109/LSP.2017.2749266.

[4] Cho, MyeongAh, et al. "Relational deep feature learning for heterogeneous face recognition." *IEEE Transactions on Information Forensics and Security* 16 (2020): 376-388.

[5] Y. Guo, L. Cao, C. Chen, K. Du and C. Fu, "Domain Alignment Embedding Network for Sketch Face Recognition," in *IEEE Access*, vol. 9, pp.872882,2021,doi:10.1109/ACCESS.2020.3047108.

[6] W. Zhang, X. Wang, and X. Tang, "Coupled information-theoretic encoding for face photo sketch recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 513–520.

[7] C. Galea and R. A. Farrugia, "A large-scale software-generated face composite sketch database,"in *Proc. Int. Conf. Biometrics SpecialInterest Group*, Sep. 2016, pp.1–5.

[8] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. Brit. Mach. Vis.Conf.*, 2015.

[9] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: closing the gap to human- level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun.2014, pp. 1701– 1708.

[10] Y. Guo, L. Cao, C. Chen, K. Du and C. Fu, "Domain Alignment Embedding Network for Sketch Face Recognition," in *IEEE Access*, vol. 9, pp.872882,2021,doi:10.1109/ACCESS.2020.3047108.

[11] H. Cheraghi and H. J. Lee, "SP-Net: A Novel Framework to Identify Composite Sketch," in *IEEE Access*, vol. 7, pp.131749-131757, 2019,doi:10.1109/ACCESS.2019.2921382.

[12] U. Osahor and N. M. Nasrabadi, "Text- Guided Sketch-to-Photo Image Synthesis," in *IEEE Access*, vol. 10, pp. 9827898289,2022,doi:10.1109/ACCESS.2022.3206771.

[13] I. J. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[14] A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and checkerboard artifacts,"*Distill*, vol. 1, no. 10, 2016, Art. no. e3.