Face Image Synthesis

Parvathy S Kurup
CSE Department
Sree Buddha College of Engineering,
Alappuzha, India
cs.parvathy@sbcemail.in

Pranav P Nair
CSE Department
Sree Buddha College of Engineering,
Alappuzha, India
pranavadoor2002@gmail.com

Sai Kishor
CSE Department
Sree Buddha College of Engineering,
Alappuzha, India
saikishor1331@gmail.com

Aryan S Nair
CSE Department
Sree Buddha College of Engineering,
Alappuzha, India
aryansnair23@gmail.com

Pranav P
CSE Department
Sree Buddha College of Engineering,
Alappuzha, India
pranavpramod422@gmail.com

***Abstract*— Facial image synthesis has made rapid dynamic progress with the fast expansion of deep learning techniques. Reference samples can give complete primary information about texture and content in this job and improve the visual quality of synthetic images. A normalized network with a multi-scale pyramid structure is used in this instance. The dual-channel normalization architecture at the center of the normalization network is capable of obtaining previous knowledge about various semantics from reference samples. There are two conditional normalization branches in the DNC specifically. Through the first branch, the reference image can be spatially adaptively normalized based on the input image's semantic mask. The second branch is used to normalize the adaptive representation of the modified input image on the reference image. By dividing the complete cross-domain mapping into two branches, DNC may highlight the distinct significance of structural and spatial elements. To avoid information redundancy and improve the final performance, the Gated Channel Attention Fusion module is used to differentiate and merge useful information from the two branches. This generated synthetic image is then compared with photos of criminals in the crime database. If a match is found, the image will be displayed along with details about that criminal. Comparing the images is done by using pillow library with image hashing.**

Keywords— ***Double Channel Normalization network, Gated Channel Attention Fusion, Semantic mask, Spatially-adaptive normalization, Adaptive instance normalization.***

I. Introduction

Facial sketch-to-image synthesis is a swiftly developing field that has the potential to change the law enforcement investigations. By automatically generating realistic images from hand-drawn sketches, this technology can greatly assist in identifying and apprehending suspects. Facial sketches are often created based on eyewitness accounts or grainy surveillance footage. By synthesizing realistic images from these sketches, investigators can better understand a suspect's appearance, making identification and arrest easier. Creating realistic images from sketches can help improve a witness's memory and improve their ability to provide an accurate description of a suspect.

This can lead to more detailed sketches, making identification efforts easier. Composite images can provide investigators with additional information to pursue potential leads. For example, the generated images can reveal distinctive featuresthat can narrow down the number of suspects or link the suspect to other incidents. The generated image must accurately represent the facial identity captured in the sketch while maintaining a realistic appearance.

The sketches vary widely in detail, style, and artist interpretation. The synthesis process must be robust to these variations to produce consistent and accurate results. The images created will be realistic, capturing subtle facial features and textures, making them more reliable for identification. The technology will adapt to handle sketches from a variety of sources, including eyewitness drawings, police artist drawings and even caricatures.

The ability to produce images in real time will enable immediate use in investigations, providing investigators with useful information while the incident is still fresh in the minds of witnesses. Facial sketch-to-image synthesis has the potential to transform law enforcement investigations, providing a powerful tool for identifying suspects and generating leads. As technology advances, its impact on crime prevention and case resolution is expected to increase significantly. The system includes features such as analysis, use of reference images, mapping from one domain to another, conversion from one domain to another.

It also uses a semantic mask for each face sketch to analyze the image and convert it into a binary image. Then, the content information of the binary image and the reference image are transmitted through the DNC network to create a composite image [1]. This image is then used to compare against the criminal database, and if a match is found, it displays that specific match and details about that suspect.

A. Generative Models:

GAN (Generative Adversarial Network): GAN is a general class of models consisting of a generator and a discriminator. The generator generates synthetic images, while the discriminator evaluates their authenticity. The generator and discriminator are trained iteratively, with the ultimate goal of creating images that are indistinguishable from the real thing.

B. Data Augmentation:

Techniques such as data augmentation play an important role in training robust facial image synthesis models. By introducing variations into the training data, the model can better generalize to different facial expressions and poses.

C. *Conditional GANs***:**

Conditional GANs extend the basic GAN architecture by introducing conditional information. In the context of face image synthesis, this means providing additional input, such as a facial attribute or specific features, to guide the generation process.

D. Applications:

Face image synthesis has a variety of applications, including video game development, virtual reality, and creating training datasets for facial recognition systems. However, it is essential to use these technologies responsibly to avoid abuse. As technology continues to advance, the ability to synthesize facial images will likely see further improvements, addressing challenges such as improving realism, reducing bias, and increasing interpretability. Researchers and practitioners in this field must remain vigilant about ethical considerations and potential social impacts.

II. LITERATURE REVIEW

Confront outline union is the method of creating confront outlines from face photographs. It features a wide extend of applications, such as advanced excitement, law authorization, and facial picture altering. In recent years, generative adversarial networks have shown promising results in face sketch synthesis. It proposes a novel GAN architecture for face sketch synthesis. The proposed mode of MHGAN comprises of three modules in here. They are a Region module, a Mask module, and a fusion module [2]. The region module starts to learn the nitty gritty highlights of diverse neighborhood districts of the confront by GAN.

The mask module produces a coarse facial structure of outline and employments the facial highlight extractor to improve the high-level picture and learn the inactive space's highlight. The fusion module produces the ultimate outline by combining fine neighborhood locales and coarse facial structure. The creators conducted broad tests on the CUFS and CUFSF standard datasets and photographs on the web.

The comes about appear that the proposed MHGAN beats the state-of-the craftsmanship strategies in terms of both subjective and objective measurements [3]. The proposed MHGAN could be a promising unused approach to confront portray union. It is able to produce tall-quality confront portrays with fine nearby subtle elements and coarse facial structures. The proposed strategy is based on GANs, which have appeared promising comes about in a assortment of image-to-image interpretation issues. The proposed strategy is additionally computationally proficient, making it reasonable for real-time applications.

The paper "Face Recognition via Multi-level 3D- GAN Colorization" proposes a novel approach to face sketch colorization using a multi-level 3D generative adversarial network. The proposed method utilizes a 3D-GAN architecture to generate colorized images from face sketches, preserving the details and features of the original sketch while adding realistic textures and color information [4]. The creators assessed the proposed strategy on the Pakistani Lawmakers Face-sketch Dataset (PPFD) and compared it to other state-of- the-art confront outline colorization strategies. The comes about illustrate that the proposed strategy beats existing strategies in terms of precision, auxiliary similitude file degree (SSIM), signal-to-noise proportion (SNR), and top signal-to-noise proportion (PSNR).

The proposed multi- level 3D-GAN approach presents a promising heading for confront outline colorization. Its hierarchical structure and utilization of 3D convolutions enable effective colorization while preserving facial features and identity. The method's performance on the PPFD dataset demonstrates its applicability to real-world scenarios. Face detection is the most important task in computer vision with a wide range of applications, including surveillance, human - computer interaction, and facial recognition. Traditional face detection methods, such as the Viola-Jones framework, rely on handcrafted features, such as Haar like features, which are limited in their ability to capture the complex variations in human faces.

The paper "Face Detection Based on Receptive Field Enhanced Multi-Task Cascaded Convolutional Neural Networks" proposes a novel approach to face detection using a multi-task cascaded convolutional neural network architecture. The proposed RFE- MTCNN enhances the receptive field of the network, allowing it to capture larger facial regions and improve its ability to detect small faces. Additionally, the network employs multi-task learning, jointly training the network for face detection, facial landmark localization, and facial attribute estimation.

The RFE-MTCNN architecture consists of three stages: Stage 1: Fast Face Proposal Generation: This stage generates candidate face regions using a lightweight CNN. Stage 2: Fine-Grained Face Classification: This stage performs fine-grained classification of the candidate face regions, eliminating false positives. Stage 3: Face Landmark Localization and Attribute Estimation: This stage locates facial landmarks and estimates facial attributes, further refining the face detection results.

Face photo-sketch synthesis is the process of generating realistic face sketches from photographs. It has a wide range of applications, such as digital entertainment, law enforcement, and facial image editing. Traditional face photo- sketch synthesis methods rely on handcrafted features and shallow neural networks, which are limited in their ability to capture the complex details and variations of human faces. The paper "Double normalization channel Pyramid Network for Face Photo-Sketch Synthesis" proposes a novel approach to face photo-sketch synthesis using a double normalization channel pyramid network (DNCP-Net) architecture.

The proposed DNCP-Net utilizes a hierarchical multiscale representation to capture subtle facial features and textures, and incorporates double normalization channel (DNC) layers to enhance the network's ability to adapt to different input images and sketch styles. The DNCP-Net architecture consists of four main components: Encoding Pyramid: This component extracts multi-scale feature representations from the input photo using a series of convolutional layers and down sampling operations. Sketch- Style Encoder: This component encodes the input sketch into a feature representation using a separate convolutional network. DNC-Net Synthesis: This component synthesizes the sketch from the encoded photo and sketch features using a series of DNC layers and up sampling operations. Output Reconstruction: This component reconstructs the final sketch from the synthesized sketch features using a series of convolutional layers.

The proposed DNC layers play a crucial role in the network's ability to adapt to different input images and sketch styles. The DNC layers incorporate a conditioning signal derived from the sketch features, allowing the network to learn and apply style-specific normalization parameters. This enables the network to generate sketches that accurately reflect the style and details of the input sketch while maintaining the overall facial structure and identity from the input photo.

III. METHODOLGY

1 Dataset Collection:

This dataset comprises 100,000 facial images and their corresponding hand-drawn sketches, carefully curated from public databases and research collaborations. The images capture a wide range of facial features, expressions, and ethnicities, including smiles, frowns, closed eyes, and individuals from various backgrounds. This diversity ensures the network trained on this data can generalize well to unseen images with similar characteristics, promoting its robustness and real- world applicability.

2. Data Pre-processing:

The collected images are preprocessed to ensure consistency and compatibility with the network architecture. This includes resizing the image to standard resolution, normalizing pixel intensity values, and removing any background noise or artifacts.

3. Semantic Mask Generation:

For each face image, a corresponding semantic mask is generated. A semantic mask is a binary image that shows the location of facial features, like eyes, mouth, nose, and eyebrows. These masks are created using image segmentation techniques or manually annotated by experts.

4. Data Augmentation:

To further improve the training data and avoid overfitting, data augmentation techniques were applied. This involves random transformations like flipping, rotating, cropping, and adding noise to the image.

5. Data Partitioning:

The dataset is divided into three subsets: training, validation, and testing sets. The training set is used to train the network parameters, the validation set is used to monitor training progress and tune hyper parameters, and the test set is used to evaluate the final performance of the network on invisible data.

6. Data Format Conversion:

The images and semantic masks were converted into a format compatible with the network architecture. This might involve converting the images into tensors or other data structures that can be efficiently processed by the network.

7. Comparing:

- Image Hashing: Each image is converted into a unique "fingerprint" called a hash using the image hash library. This hash captures the image's essential features in a compact representation.
- Similarity Measurement: The Hamming distance between the reference image's hash and each image's hash is calculated. This distance measures how different the hashes are, with a lower distance indicating greater similarity.
- Identifying Similar Images: Images within a user-defined threshold of the reference image's hash distance are considered similar. The code returns a list of filenames of these similar images.

IV. PROPOSED SYSTEM

The Dual Channel Normalization (DNC) module is a component used to synthesize facial sketch images. It consists of two branches, the spatially adaptive non-normalized branch and the adaptive representation-normalized branch.

The SPADE (Spatial Adaptive Denormalization) module is used to synthesize optical images based on semantic segmentation masks. The scaling value (g) and bias value (b) in the DNC module are learned from the input semantic segmentation mask. The normalized activation value is calculated using the learned scale and bias values as well as the input activation value.

The DNC module plays an important role in adapting the normalization process based on the input semantic segmentation mask, allowing for more precise and contextual analysis. conscious synthesis of photographs and facial sketches.

Adaptive instance normalization is an instance normalization layer that normalizes each feature channel using a different set of parameters.

The DNCP architecture uses two branches, one for each normalization technique. The Spade branch normalizes the input tensor using Spade normalization, and the Adaptive Instance branch normalizes the input tensor using Adaptive Instance normalization. The output of the two branches is then concatenated and passed through a final ReLU activation function.

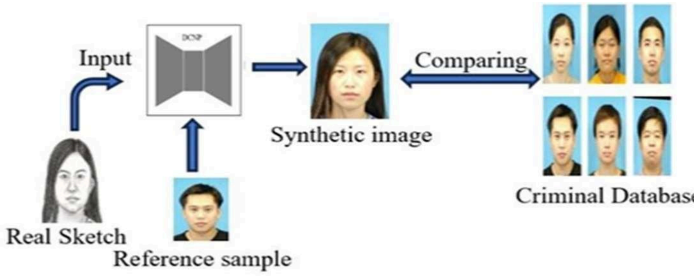

Fig.1 Illustration of the Proposed System

The synthetic image is then compared with criminal database to find a match. Method used to compare is pillow algorithm with image hashing.

V. MODULES

1 Encoder:

The encoder is responsible for extracting features from the input image and semantic mask. It consists of a series of convolutional layers that gradually sample input images, capturing high- level semantic information about facial features and their spatial relationships.

2 Decoder:

The decoder takes the features extracted from the encoder and generates the corresponding sketch. It consists of a series of convolutional layers that gradually sample feature maps, reconstruct sketch details, and preserve the overall structure of the face.

3 Double normalization channel Modules:

The DNC module is an important innovation in network architecture. They are integrated into encoder-decoder structures at multiple scales, allowing efficient integration of spatial and semantic information from reference images.

4 Multi-Scale Pyramid:

The DNC modules are organized in a multi-scale pyramid, enabling the network to capture both global and local features of the face.

5 Residual Connections:

The remaining connections are combined across the network to facilitate information flow and avoid the vanishing gradient problem. These connections allow the network to learn deeper representations and improve the overall stability of the training process.

6 Output Layer:

The final layer of the decoder generates the generated sketch. This layer typically includes a tanh activation function to ensure that the output pixel values are within the appropriate range for the grayscale sketch.

7 Parser:

The parser used here is for actual sketching. The sketch is actually passed to the parser to create the sketch mask. The parser used in Double Normalized Channel Pyramid (DNCP) for facial image sketch synthesis is a convolutional neural network responsible for extracting and interpreting structural information from the sketch input. It plays an important role in the synthesis process by guiding the DNCP network to generate realistic sketches that accurately reflect facial structure and recognition from the input image.

VI. RESULT AND CONCLUSION

In summary, Face Image Synthesis has successfully developed a dual-channel normalized pyramid network for facial image sketch synthesis. The DNCP framework uses multi-scale hierarchical representation, dual-channel normalization (DNC) layers, and sketch-style encoders to effectively capture facial features, adapt to different input styles, and generate sketches of realistic faces.

Experimental results on benchmark datasets demonstrate that DNCP-Net outperforms state-of- the-art facial image synthesis methods on both subjective and objective metrics. The generated sketches demonstrate high fidelity, preserve facial structure and identity, and accurately reflect the style of the input sketch.

Project results show that DNCP is a promising method for facial image sketch synthesis and has potential for many real-world applications, such as digital entertainment, law enforcement and video, facial image editing. Future work may involve exploring different network architectures, loss functions, and applications as well as addressing potential limitations such as generalizability, computational complexity, other trade-offs.

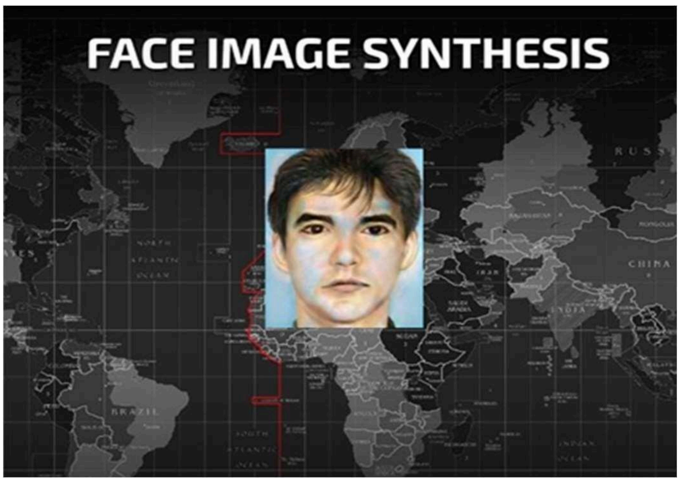

Fig.2 Illustration of Synthetic Image Generated

The main function of the Spade branch is to learn different normalization parameters for each feature channel in the input tensor. This adaptive approach ensures that each channel is appropriately normalized, based on its unique characteristics and distribution.

On the other hand, the Adaptive Instance branch focuses on instance normalization, which normalizes each feature channel independently using a separate set of parameters. It consists of a single convolutional layer followed by ReLU activation.

The output of the convolutional layer is multiplied by a set of parameters learned by the network, and this normalized output undergoes another ReLU activation. The main function of the Adaptive Instance branch is the ability to adapt to different input distributions.

It normalizes each feature channel based on its statistical properties, making the network more robust against variations in input data. The combination of Spade and Adaptive Instance branches in the DNCP architecture provides a comprehensive standardization strategy that leverages the strengths of both techniques. The Spade branch's adaptive normalization improves the network's ability to handle feature-specific variations, while the Adaptive Instance branch's instance normalization contributes to the network's robustness against change in input distribution. The Spade and Adaptive Instance branches work together.

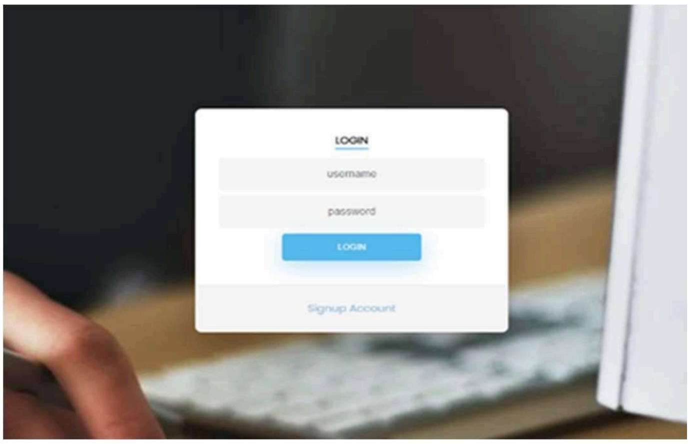

Fig.3 Login Page

Fig.3 shows the Login Page of the Project. It also has Sign Up tab for registering new users. There is an Admin user who can upload the generated images from the sketch. Admin can also enter the details of these criminals.

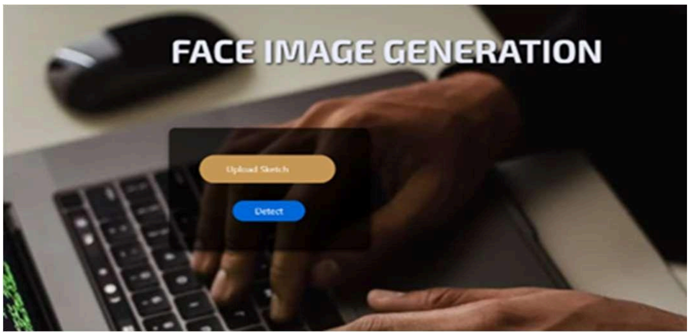

Fig.4 Home Page

Fig.4 shows Home Page where the sketch is to be uploaded. After the sketch is uploaded it checks the system status and takes step to generate the synthetic image.

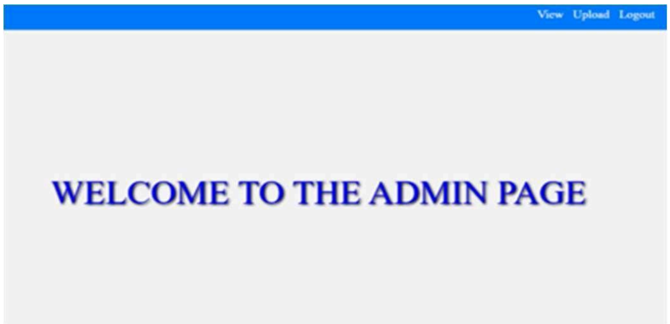

Fig.5 Admin Page

Fig.5 shows the admin page. Admin can upload the generated images from the sketch to the database from here. The generate images are then compared with images from this database. The admin can also view the criminals stored in the database here.

REFERENCES

[1] M. Zhu, Z. Wu, N. Wang, H. Yang, and X. Gao, "Dual Conditional Normalization Pyramid Network for Face Photo-Sketch Synthesis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 9, pp. 5200–5211, Sep. 2023.

[2] Z. Khan *et al.*, "Face Recognition via Multi-Level 3D-GAN Colorization," *IEEE Access*, vol. 10, pp. 133078–133094, 2022.

[3] "Face Detection Based on Receptive Field Enhanced Multi-Task Cascaded Convolutional Neural Networks | IEEE Journals & Magazine | IEEE Xplore." Accessed: Mar. 08, 2024.

[4] "MHGAN: Multi-Hierarchies Generative Adversarial Network for High-Quality Face Sketch Synthesis | IEEE Journals & Magazine | IEEE Xplore." Accessed: Mar. 08, 2024.

[5] Image-to-Image Translation with Conditional Adversarial Networks (Pix2Pix) by Isola et al. (2017)

[6] Unsupervised Image-to-Image Translation Using Cycle-Consistent Adversarial Networks (CycleGAN) by Zhu et al. (2017)

[7] DualGAN: Unsupervised Dual Learning for Image-to-Image Translation (DualGAN) by Yi et al. (2017)

[8] MDAL: Face Sketch Synthesis by Multidomain Adversarial Learning (MDAL) by Sun et al. (2018)

[9] FAG-GAN: Facial Attributes Guided Deep Sketch-to-Photo Synthesis (FAG-GAN) by Guo et al. (2019)

[10] Geo-GAN: Unsupervised Facial Geometry Learning for Sketch to Photo Synthesis (Geo-GAN) by Liu et al. (2019)

[11] MSG-SARL: Multi-Scale Gradients Self-Attention Residual Learning for Face Photo-Sketch Transformation (MSG-SARL) by Li et al. (2020)