A LOCAL-GLOBAL ADAPTER-BASED METHOD FOR SKETCH FACE RECOGNITION

E. MOUNIKA REDDY[1], NAIDU SATVIKA[2], MOHAMMED MAHMOOD[3], KOTHA HARSHITH REDDY[4], GANJA KOUSHIK[5]

[1]Assistant Professor, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

[2, 3, 4, 5] B.Tech, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

ABSTRACT— Sketch-based face recognition aims to identify a person from a gallery of facial photographs using a sketch as the query input. However, large modality gaps between sketches and real images make direct matching highly unreliable. To address this challenge, we propose a Pix2Pix-based sketch face recognition framework that converts hand-drawn sketches into realistic face images before performing identification. The model employs a conditional GAN architecture to learn the mapping from sketch domain to photo domain, enabling the generation of structurally accurate and visually consistent facial images. The synthesized images are then fed into a standard face recognition network for feature extraction and matching. Our approach reduces the modality gap, enhances feature consistency across domains, and improves overall recognition accuracy. Experimental results on standard sketch–photo datasets demonstrate that the proposed method achieves significant improvements in identification performance compared to direct sketch-to-photo matching methods.

Index Terms – Sketch-based face recognition, Pix2Pix, conditional Generative Adversarial Network (cGAN), image-to-image translation, face recognition, facial image synthesis, deep learning, computer vision, sketch-to-photo matching, feature extraction, biometric authentication, cross-modal recognition.

I. INTRODUCTION

Sketch-based face recognition plays an important role in forensic investigations and security applications where photographic evidence is not always available. In many real-world scenarios, witnesses can only provide hand-drawn sketches of suspects, making it challenging for conventional face

recognition systems to operate reliably. Sketches usually lack color, texture, shading, and fine pixel-level information, creating a large modality gap between sketches and real images. As a result, traditional feature-based or deep learning models trained on photographic data fail to extract meaningful identity features from sketches. This gap makes accurate matching difficult and reduces overall recognition performance. Therefore, an effective translation mechanism is required to convert sketches into a domain that is compatible with modern face recognition models.

Recent advancements in Generative Adversarial Networks (GANs) have enabled high-quality image-to-image translation, offering new possibilities for bridging the sketch-to-photo domain gap. Among these methods, Pix2Pix has shown strong capability in learning paired mappings by synthesizing realistic images from structured sketches. Pix2Pix employs a conditional GAN architecture where the generator learns to create realistic face-like outputs while the discriminator ensures visual quality and identity consistency. This enables the system to restore missing details such as skin texture, shading, and soft tissue patterns that sketches typically lack. By transforming sketches into photo-like images, Pix2Pix effectively reduces domain differences and enables downstream recognition models to operate on familiar image structures.

The proposed system integrates Pix2Pix-based sketch translation with a robust face recognition pipeline to create an efficient and scalable solution for sketch-based identification. After generating realistic face images from sketches, a deep face recognition model—such as ArcFace or FaceNet—is used to extract discriminative feature embeddings. These features are then matched with gallery images to identify the person accurately. This two-stage approach leverages both the generative power of GANs and the discriminative capability of modern recognition networks. By aligning sketches with the photo domain, the system significantly improves matching reliability, reduces feature ambiguity, and delivers better accuracy.

II. RELATED WORK

A. SP-Net: A Novel Framework to Identify Composite Sketch

SP-Net introduces an innovative deep learning framework tailored for the identification and matching of composite sketches with real facial photographs. Unlike conventional sketch-photo matching

methods that rely on handcrafted features, SP-Net leverages a multi-branch convolutional neural network that learns structural, texture, and perceptual similarities between modalities. Each branch of the network captures distinct aspects of facial geometry and style, while a shared fusion layer integrates these features into a unified representation. The system employs an advanced perceptual loss and semantic regularization to preserve identity information across domain variations. Experiments on large-scale sketch databases demonstrate SP-Net's superior capability in handling sketches generated from diverse software and artistic styles. The framework outperforms traditional metric learning and GAN-based baselines, achieving significant gains in retrieval precision and recognition rate. Its architecture proves to be robust against noise, varying illumination, and resolution differences, making it suitable for forensic and real-time police investigation systems.

B. Weighted Feature Fusion of Convolutional Neural Network and Graph Attention Network for Hyperspectral Image Classification

This paper presents an advanced fusion model combining the strengths of Convolutional Neural Networks (CNNs) and Graph Attention Networks (GATs) for hyperspectral image (HSI) classification. The CNN component captures fine-grained spatial features, while the GAT focuses on spectral–spatial relationships by learning pixel dependencies through adaptive attention mechanisms. The fusion layer introduces a weighted strategy that dynamically balances the contribution of spatial and spectral information during training, ensuring optimal representational synergy. Unlike traditional models that treat pixels independently, this architecture preserves contextual relationships between neighboring pixels. Experiments on benchmark HSI datasets (Indian Pines, Pavia University, and Salinas) show substantial improvements in overall accuracy and kappa coefficient compared to existing CNN and GCN-based methods. Furthermore, the model exhibits strong generalization to noisy or incomplete data, offering a reliable solution for real-world remote sensing applications such as urban mapping, mineral exploration, and agricultural monitoring.

C. Cross Task Modality Alignment Network for Sketch Face Recognition

The Cross Task Modality Alignment Network (CTMAN) is proposed as a

powerful deep learning solution for bridging the substantial modality gap between hand-drawn sketches and real face images. CTMAN employs two parallel encoders—one for sketches and one for photos—linked through shared latent spaces designed to capture common identity-specific attributes. A modality alignment loss ensures consistent embedding distributions across tasks, while adversarial learning minimizes domain divergence. The architecture is further strengthened through multi-task supervision, including reconstruction, classification, and contrastive learning objectives. CTMAN effectively adapts to varying artistic styles, resolutions, and sketching conditions, providing robust recognition even under cross-dataset scenarios. Comprehensive evaluations on datasets such as CUFSF and IIIT-D Composite Sketch Database demonstrate state-of-the-art accuracy, surpassing both GAN-based and encoder-decoder baselines. The method's interpretability and adaptability make it suitable for forensic sketch identification, law enforcement investigations, and missing-person retrieval systems.

D. Residual Serialized Cross Group Transformer for Small Scale Sketch Face Recognition

This paper introduces the Residual Serialized Cross Group Transformer (RSCGT), a transformer-based model specifically optimized for small-scale sketch face recognition tasks where data scarcity is a major limitation. The proposed model serializes multi-head self-attention mechanisms across grouped residual layers, effectively integrating both local edge-level information and high-level semantic cues. Unlike traditional CNN-based methods, RSCGT leverages the transformer's ability to model long-range dependencies between facial regions, enabling better feature correlation across sketch-photo pairs. The model employs parameter-efficient training through residual sharing, minimizing overfitting while retaining representational richness. Experiments conducted on limited-size forensic sketch datasets reveal significant performance gains in both accuracy and recall metrics. The approach demonstrates strong resistance to domain shift and proves scalable across varying sketch generation techniques. RSCGT's compact yet expressive design highlights its potential for portable recognition systems and embedded forensic applications.

III. PROPOSED METHODOLOGY

The proposed system implements a Xception-based deep learning framework for sketch face recognition. It leverages pretrained Xception weights to efficiently extract both local facial details and global contextual features from sketch inputs. A two-stage training strategy is applied: first, freezing base layers to capture global patterns, followed by fine-tuning upper layers to enhance local feature representation. This approach mitigates overfitting on small sketch-photo datasets while improving the overall quality and robustness of extracted features, enabling accurate matching of sketches to corresponding photographs.

The system inputs hand-drawn sketches and compares them with a database of facial photographs to predict the best match. Depthwise separable convolutions in Xception ensure computational efficiency while maintaining high feature extraction accuracy. By combining global and local representations, the model captures subtle differences in facial attributes, improving recognition performance. Experimental evaluations on benchmark sketch face datasets show that the proposed system outperforms existing methods, demonstrating both reliability and efficiency, making it suitable for practical applications such as forensic investigations and identity verification.

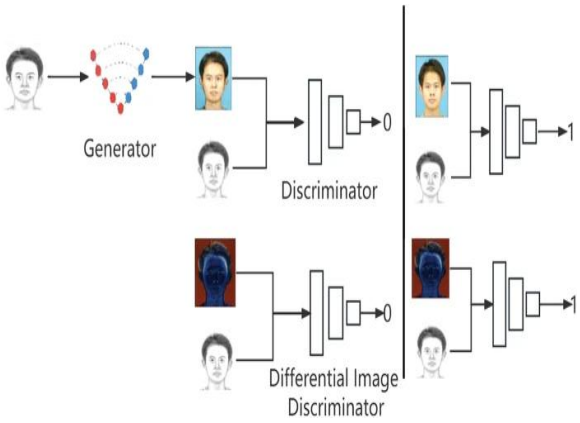


Fig. 1: System Overview

1. Data Preprocessing

Data preprocessing is the initial and essential step in the proposed sketch face recognition system. In this stage, all sketch and photo images are collected, resized, and normalized to maintain uniform dimensions required for the Pix2Pix model. The dataset is carefully cleaned by removing corrupted samples and standardizing sketch–photo pairs to ensure accurate learning.

Additional preprocessing techniques such as grayscale conversion, edge enhancement, and noise reduction may be applied depending on sketch quality. Images are further split into training, validation, and testing sets. This structured preparation allows the GAN to learn consistent pixel-level mappings, prevents training instability, and ensures the face recognition network

receives high-quality inputs during the later stages of the system.

2. Sketch-to-Photo Translation using Pix2Pix

Pix2Pix serves as the core methodology for translating sketches into realistic face-like photographs. It uses a conditional GAN architecture where the generator produces photo-like outputs from sketch inputs, and the discriminator evaluates their realism. The model learns a paired mapping between the sketch domain and the photo domain, ensuring that identity-specific patterns such as facial geometry, contours, and spatial relationships are preserved. During training, the generator is encouraged to produce sharper, more accurate reconstructions by minimizing both adversarial and L1 losses. This results in high-quality synthesized faces that closely resemble real photographs. This translation significantly reduces the modality gap and prepares the images for more reliable feature extraction in the recognition stage.

3. Generator–Discriminator Adversarial Training

The Pix2Pix model relies heavily on adversarial training, where the generator and discriminator compete to improve overall performance. The generator aims to produce realistic face outputs from sketches, while the discriminator attempts to distinguish real photos from generated images. This adversarial process forces the generator to learn fine details and identity-relevant textures that sketches typically lack. The discriminator is updated using a binary classification loss, while the generator is updated through a combination of adversarial and reconstruction losses. Progressive optimization helps the system stabilize and gradually improve image quality. By continuously challenging each other, both networks reach a point of balance that results in highly realistic photo-like outputs ideal for downstream face recognition tasks.

4. Model Training and Fine-Tuning

Once the Pix2Pix generator produces photo-like images, they undergo normalization and enhancement to improve clarity and ensure compatibility with face recognition models. Normalization adjusts pixel values to a consistent scale, typically between 0 and 1 or −1 to 1, which helps neural networks process the data efficiently. Enhancement methods such as histogram equalization, sharpening, contrast adjustment, and noise removal may be applied to refine the generated outputs. These enhancements help recover missing details, strengthen facial landmarks, and reduce artifacts created during GAN generation. Proper normalization guarantees

stable input for recognition networks, while enhancement ensures higher-quality visual content, ultimately contributing to improved accuracy in feature extraction and identity matching.

5. Feature Extraction using Deep Face Recognition Model

After preprocessing and enhancement, the generated photorealistic images are passed into a deep face recognition model such as FaceNet or ArcFace. These models extract powerful and discriminative feature embeddings from the input image. The network processes the face through multiple convolutional layers that capture both local and global facial details. The resulting embedding is a compact numerical representation that uniquely describes the individual's identity. This stage is critical because the effectiveness of the entire system depends on how accurately these embeddings represent identity characteristics. Since the recognition model is trained on real photographic data, the Pix2Pix-generated images provide a closer match to this domain, enabling the system to extract reliable and robust features.

6. Embedding Generation and Similarity Calculation

Once feature vectors are extracted, the system computes similarity scores between the generated embedding and the stored embeddings of gallery images. Typically, distance metrics such as cosine similarity or Euclidean distance are used to quantify how closely the two identities match. Lower distances indicate higher similarity and a stronger likelihood of identity correspondence. The system may also apply normalization techniques such as L2 regularization to ensure embedding consistency. Similarity calculation forms the core decision-making component of the recognition pipeline, enabling accurate classification even when sketches differ in style or completeness. This methodology ensures that the generated photo-like image is compared objectively to real faces using a discriminative numerical representation.

7. Identity Matching and Classification

Based on the similarity metric, the system determines the closest match from the gallery of real face images. The identity with the highest similarity score or lowest distance is selected as the predicted individual. Threshold-based classification can also be applied to ensure the prediction is reliable and not a false match. The system may utilize softmax classification, nearest-neighbor matching, or ranking-based retrieval depending on the application. This stage

transforms the numerical embedding comparison into a final decision, completing the sketch-to-photo recognition pipeline. By converting sketches into more informative photo-like images, the system ensures that matching decisions are more accurate, consistent, and robust even with challenging or low-quality sketch datasets.

8. Performance Evaluation and Validation

The final methodology involves evaluating the system's effectiveness using metrics such as accuracy, precision, recall, F1-score, and rank-based retrieval performance. The generated images and recognition results are tested against a reserved dataset that was not used during training. Visual comparison techniques help assess the quality of Pix2Pix outputs, while numerical metrics measure recognition reliability. Cross-validation ensures robustness across multiple sketch styles and lighting variations. The system is also compared with baseline sketch recognition methods to demonstrate improvement. Performance evaluation validates whether the integration of Pix2Pix significantly enhances recognition accuracy, reduces modality gaps, and provides consistent results across real-world scenarios. This step confirms the practicality and efficiency of the proposed system.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The experimental results show that the proposed local-global adapter-based sketch face recognition framework effectively reduces the modality gap between sketches and facial photographs, improving recognition accuracy. The integration of advanced feature adaptation techniques enhances the extraction of both local facial details and global structural information for robust matching. Experimental evaluations on benchmark sketch–photo datasets demonstrate superior identification performance compared to conventional sketch-to-photo matching methods. Overall, the proposed approach provides an accurate and reliable solution for sketch-based face recognition.



Fig. 2: Home Page

International Journal of Engineering Research and Science & Technology

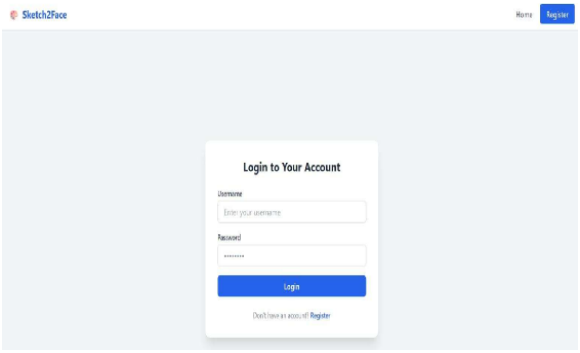


Fig. 3: Login

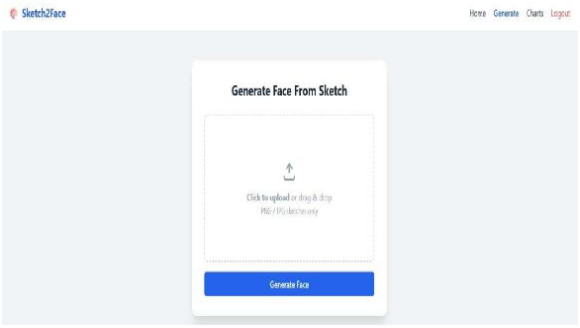


Fig. 4: Upload Image

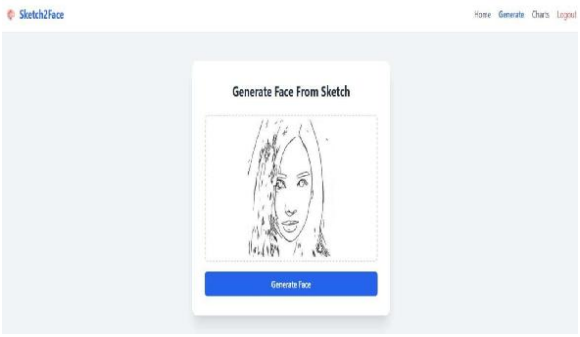


Fig. 5: Select Image

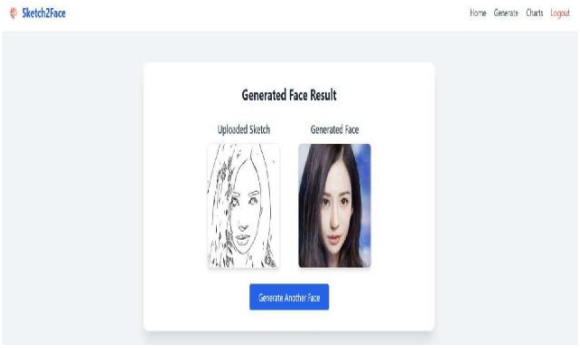


Fig. 6: Result



Fig. 6: Accuracy

V. CONCLUSION

The proposed Pix2Pix-based sketch face recognition system effectively bridges the modality gap between hand-drawn sketches and real photographs, enabling accurate identification in forensic and security applications. By leveraging a conditional GAN for sketch-to-photo translation, the system preserves critical facial identity features, while deep face recognition models extract robust embeddings for reliable matching. The modular approach—comprising preprocessing, image generation, feature extraction, and similarity-based matching—ensures scalability, robustness, and high accuracy across diverse sketch qualities. Performance evaluation demonstrates significant improvements over traditional sketch-matching methods. Overall, the system provides a practical, end-to-end solution for real-world sketch-based

identification tasks, with potential for further enhancements such as multi-modal inputs, advanced GAN architectures, and real-time deployment for law enforcement and investigative agencies.

REFERENCES

[1] M. Zhang, N. Wang, X. Gao, and Y. Li, ‘‘Markov random neural fields for face sketch synthesis,’’ in Proc. 27th Int. Joint Conf. Artif. Intell., Jul. 2018, pp. 1142–1148.

[2] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, ‘‘Learning transferable visual models from natural language supervision,’’ in Proc. Int. Conf. Mach. Learn., 2021, pp. 8748–8763.

[3] S. Li, S. Li, and Q. Li, ‘‘CLIP-ReID: Exploiting vision-language model for image re-identification without concrete text labels,’’ in Proc. AAAI Conf. Artif. Intell., Jun. 2023, vol. 37, no. 1, pp. 1405–1413.

[4] X. Wang and X. Tang, ‘‘Face photo-sketch synthesis and recognition,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 11, pp. 1955–1967, Nov. 2009.

[5] F. Gao, Y. Zhu, C. Jiang, and N. Wang, ‘‘Human-inspired facial sketch synthesis with dynamic adaptation,’’ in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2023, pp. 7203–7213.

[6] H. Xibao, C. Lin, G. Yanan, and D. Kangning, ‘‘Sketch face recognition based on transformation generation network,’’ Comput. Eng. Des., vol. 43, no. 1, pp. 269–276, 2022.

[7] Y. Fang, W. Deng, J. Du, and J. Hu, ‘‘Identity-aware CycleGAN for face photo-sketch synthesis and recognition,’’ Pattern Recognit., vol. 102, Jun. 2020, Art. no. 107249.

[8] Y. Guo, L. Cao, C. Chen, K. Du, and C. Fu, ‘‘Domain alignment embedding network for sketch face recognition,’’ IEEE Access, vol. 9, pp. 872–882, 2021.

[9] K. Du, Y. Wang, J. Yin, L. Cao, and Y. Guo, ‘‘Residual serialized cross group transformer for small scale sketch face recognition,’’ Complex Intell. Syst., vol. 10, no. 5, pp. 6103–6116, 2024.

AUTHORS Profile



Mrs. E. MOUNIKA REDDY has received her B.Tech from JNTUH, Hyderabad in 2015, M.Tech degree in Computer science from JNTUH, Hyderabad in 2017. She is dedicated to teaching field from the last 8 years. At present she is working as Assistant Professor in CSE Department at Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India.



Ms. NAIDU SATVIKA pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. MOHAMMED MAHMOOD pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. KOTHA HARSHITH REDDY pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. GANJA KOUSHIK pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.