Proceedings of the 8th International Conference on Computing Methodologies and Communication (ICCMC-2025)
DVD Part Number: CFP25K25-DVD; ISBN: 979-8-3315-1210-1

Fusion-Based Super-Resolution for Facial Identification and Sketch-to-Image Translation

Santhana Krishnan. G
Department of Computer Science and Engineering
Vels Institute of Science, Technology and Advanced Studies (VISTAS),
Chennai, India
santhanakrishnang13@gmail.com

Kumar. N
Department of Computer Science and Engineering
Vels Institute of Science, Technology and Advanced Studies (VISTAS),
Chennai, India
kumar.se@vistas.ac.in

Sheela Gowr. P
Department of Computer Science and Engineering
Vels Institute of Science, Technology and Advanced Studies (VISTAS),
Chennai, India
sheela.se@vistas.ac.in

***Abstract*— Advancements in technology have inadvertently enabled easier access to crime-related data, resulting in an increase of criminal activities. Law enforcement agencies encounter substantial difficulties in identifying suspects and preventing further crimes. To tackle these challenges, this research introduces a proposed solution of fusion-based super-resolution approach combined with sketch-to-image translation techniques. The method enhances low-quality facial images and using ESRGAN and Real-ESRGAN. A fusion model integrates multiple enhancement strategies, followed by a Siamese model for face verification system. The experimental results demonstrate improved recognition accuracy and image quality, supporting efficient suspect identification. The proposed model shows significant improvement in areas of criminal sketches. This approach effectively strengthens facial recognition systems in real-world surveillance and forensic scenarios.**

Keywords*— *Crime Detection, Neural network, CycleGAN, ESRGAN, Real-ESRGAN, CCTV Surveillance, Sketch-to-Image Translation.

I. INTRODUCTION

Crime remains one of the most critical and persistent challenges confronting society today. The steady rise in crime rates, fueled by the growing frequency of offenses committed daily which necessitates constant surveillance of criminal activities and the maintenance of detailed databases for effective crime prevention. Diverse forms of crime, including data fraud, theft, and burglary, disrupt public safety and diminish the quality of life and mental well-being of individuals. In light of the rising challenges over crime and public security, the widespread deployment of closed-circuit television (CCTV) systems across public and private domains has become essential to support law enforcement. These System play a crucial role in post-crime analysis, with video footage serving as crucial evidence for investigation and incident assessment. However, the proliferation of CCTV installations has significantly increased the demand for human operators to monitor these video streams. To address this operational challenge, deep learning models have emerged as a promising solution, offering faster and more accurate performance than traditional approaches while enabling real-time data verification and assisting law enforcement agencies in enhancing operational efficiency.

Advancements in Artificial Intelligence (AI) have led to the development of numerous facial analysis applications. Convolutional Neural Networks (CNNs) in particular have driven significant progress in computer vision, natural language processing, and facial feature mapping. Despite these advancements, real-world facial identification remains challenging due to factors such as partial occlusions and adverse lighting conditions, which hinder accurate recognition and feature extraction. Viola and Jones introduced a cascade-based face detection framework that leverages Haar-like features combined with the AdaBoost algorithm to train cascaded classifiers, achieving notable detection efficiency. Nonetheless, several studies have shown that such detectors and classifiers, even when enhanced, encounter limitations in practical scenarios due to overly smoothed and low-quality input images. To overcome these constraints, Chao Dong pioneered Deep learning-driven image super-resolution techniques, which learn end-to-end mappings Between lower-resolution and higher-resolution images to improve image quality. Although effective, these methods are computationally intensive and often produce overly smooth images lacking fine details, rendering them might be less suitable for real-time applications. Additionally, Gargi Tela explored criminal identification through facial recognition and fingerprint analysis by employing CNNs to extract and analyze high-dimensional facial features. Despite various research efforts to apply video-based facial recognition technology, challenges such as lighting variations, pose changes, facial expressions, and environmental obstructions continue to limit practical implementation.

The proposed model uses a Siamese Neural Network for One-Shot Face Verification [14]. For image enhancement it uses ESRGAN [8] and Real-ESRGAN models [9], generating two high-resolution outputs images then These outputs are fused using an Attention-Based Fusion Network, which applies channel and spatial attention mechanisms, edge refinement, and smoothing to enhance feature preservation. The fused image is then passed through a Siamese model [14], which extracts deep feature embeddings and performs one-shot verification for criminal identification.

II. RELATED WORKS

The accuracy of public facial recognition are primarily affected by occlusion such as masks, glasses … which occurs in the faces and the quality of the input image used for detection The following is a review of existing literature in the field of research relevant to this research topic.

- K. Kranthi Kumar et al [1], proposed a method for real-time criminal identification system utilizing MTCNN for face recognition and a Siamese Neural Network for one-shot learning-based verification [14]. it achieved an 86% accuracy but has limitations like poor performance in low-light conditions, high computational costs, and a small

dataset for training and testing. There is no specific methodology to enhance the low-resolution input images.

- The Xiaoguang Li et al [2], introduced a approach to enhance low-quality face images. The authors proposed a Discriminative Self-Attention Cycle-Consistent Generative Adversarial Network (CycleGAN) [7] which uses unpaired samples to simultaneously train degradation and reconstruction of the networks, A self-attention mechanism is designed to restore intricate facial details. It outperforms traditional methods in scenarios like processing Real-world facial images of low quality but it suffers from Identity Preservation Issues and Generalization Challenges.
- Chengkun Song et al [3], proposed a methodology where the authors employ the Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) [8] to upscale low-resolution images and restore finer facial details. This approach demonstrates a significant improvement in the recognition of individuals from low-quality images but the ESRGAN are prone to generating artifacts during the super-resolution process and these artifacts can distort facial features and adversely affecting the accuracy.
- Helena Dewi Hapsari et al [4], addresses challenges in forensic investigations caused by poor-quality facial images. The study involved preparing low-resolution facial images, enhancing these images using the ESRGAN [8] model, and evaluating the results with metrics such as PSNR and SSIM. The findings indicate that ESRGAN significantly enhances the visual quality of the facial images but has higher Sensitivity to Extreme Noise.
- Xintao Wang et al [5], presents an extension of The Real-Enhanced Super-Resolution Generative Adversarial Network (Real-ESRGAN) [9] is utilized to tackle the challenges of super-resolution of images in real-world and artifacts problem. The Real-ESRGAN incorporated an advanced degradation modeling process which is used to more accurately replicate complex real-world image distortions and the comparisons demonstrates that the Real-ESRGAN achieves superior visual performance over prior models. The authors also offers efficient implementations for generating synthetic training pairs dynamically during training. The disadvantages of this model is Dependence on Synthetic Data and struggles with Artifact Handling in images with unseen artifacts.
- Xingfang Yuan [6] introduced a CycleGAN-based technique [7] aimed at transforming forensic sketches into realistic facial images. This approach trains on unpaired sketches and photographs, providing enhanced flexibility, and the model quickly converges within 500 epochs, producing high-fidelity images with natural skin tones and lighting; however, the model suffers from mode collapse when training is extended beyond 1000+ epochs. A key limitation of this method is its difficulty in handling extreme data variations.

The previous approaches uses deep learning models which helps to identify the inherent correlation between facial landmark localization and bounding box regression; however, this proposed model helps to achieve the detection process more efficiently and Image enhancement with the fusion network used in this model helps to achieve more accuracy than the existing algorithms. it helps to achieve performance boost in poor quality images that are usually obtained from CCTV and also works more efficiently with criminal sketches which many algorithms usually fall short in real-life. The process is leveraged by one-shot learning, thus the model can verify identities with minimal data, making it more robust for real-time applications like biometric authentication and surveillance.

III. PROPOSED METHODOLOGY

The criminal face identification framework operates by capturing frames from live CCTV footage and subjecting them to pre-processing steps. Following this, facial regions are extracted from the processed images and subsequently verified against an existing criminal database to retrieve detailed information about identified individuals. The efficiency and accuracy of this pipeline are further improved through the integration of the proposed algorithm, as illustrated in (Fig.1).

A. Registration of new criminal

The new criminal details such as name, age, crime committed … are collected from the user using the GUI application and these collected data are stored in database in CSV format for efficient retrieval and the GUI application also has features to store the log details of the user for security purpose. It also uses AES Encryption in EAX mode to ensure data security during transmission.

B. Image pre-processing using Fusion Model

The pipeline is started with detecting faces using RetinaFace model [11] which makes the bounding box on the faces and crop them to store it, then the captured faces are transferred to the pre-processing phase of the image. In the image pre-processing stage, the input images are resized to 250×250 and converted into tensors to ensure uniformity across the data. To enhance the facial details the images are passed through pipeline of ESRGAN, Real-ESRGAN, and the fusion model. In ESRGAN [8] a RRDBNet-based generator is used to enhance the images and are applied with CLAHE algorithm for contrast enhancement and Gaussian blur for denoising the image, this ESRGAN model[8] helps to reconstruct the fine facial details like hair and skin textures but it has some drawbacks of introducing artificial skin-color and artifacts.

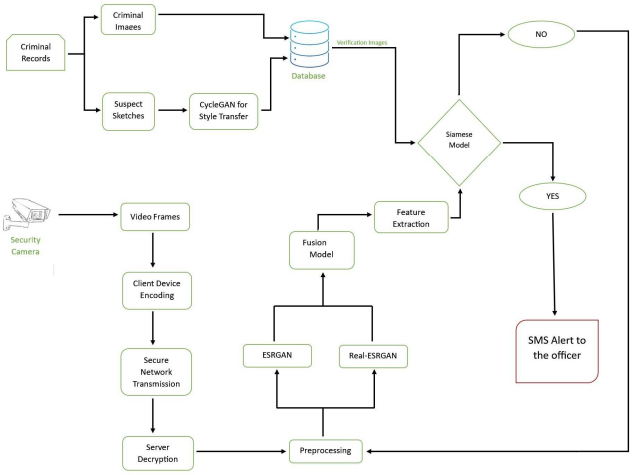

Figure 1: System Flow Diagram

The Real-ESRGAN model[9] also uses a RRDBNet based architecture which helps to retain facial details ,skin-color and enhance the overall image sharpness even on a high frequency noise image but it has the disadvantage of making the facial textures smoother. The proposed Fusion model manages to take advantages on both images enhancing methods and combing them to generate a single image. To begin with the outputs images of both models are altered with PIL-based image enhancer to fine-tune image sharpness. The resulting images are fused together based on a Attention-Based Fusion Model.

The proposed Fusion model uses Channel Attention with Squeeze-and-Excitation[12] module where the image tensor is passed for Global Average, followed by Two fully connected layers (fc1 and fc2) are used to compute channel weights, then computed channel weights are applied to the input using a sigmoid activation and this helps to refine features importance across colour channels and to learn which details need to be preserved from both inputs. The sigmoid activation results are passed to a lightweight convolutional refinement layer of Spatial Attention.

The Spatial Attention has two Convolutional layers (spatial_conv1, spatial_conv2) which are applied to the output of the channel attention, followed by Squeeze-and-Excitation [12] applied to compute spatial weights, at the end of this module spatial weights are performed to the result of the fully connected layer using a sigmoid activation function. This method is used to extract finer details without compromising the balancing sharpness across different image regions. The output tensor is passed to a convolutional layer which is used to refine edge artifacts and Gaussian blur is applied to smooth the subject outline and background(Fig.2)

$$Fused = \alpha \cdot Real-ESRGAN + (1-\alpha) \cdot ESRGAN \quad (1)$$

In short, the Fusion model dynamically assigns fusion weights similar to equation (1) where α is based on pixel-wise importance calculated by the Channel attention and spatial attention. Thus, resulting in an enhanced image formed by stacking two super-resolved images along the channel and spatial dimensions. The comparison showing the effectiveness of the enhancement model is in (Fig.3).

C. Siamese Network for Criminal Identification

The architecture of the Siamese network [14] comprises two identical Convolutional Neural Networks (CNNs) designed to learn facial similarity. The initial stage of the network, known as ConvNet-64, consists of convolutional layers activated by ReLU non-linearity, followed by batch normalization layers. This stage is primarily responsible for capturing low-level spatial features of the face. The resulting feature maps are then down sampled using 2 × 2 max pooling layers. These outputs are subsequently passed into EmbeddingNet-128, a deeper CNN module that refines the feature embeddings by emphasizing high-frequency facial attributes such as eyes and other fine details. The features extracted from this network are initially flattened and then passed through a fully connected layer which results in a 128-dimensional feature vector.

The final output vectors generated by the CNNs are then compared using the Euclidean distance metric. If the

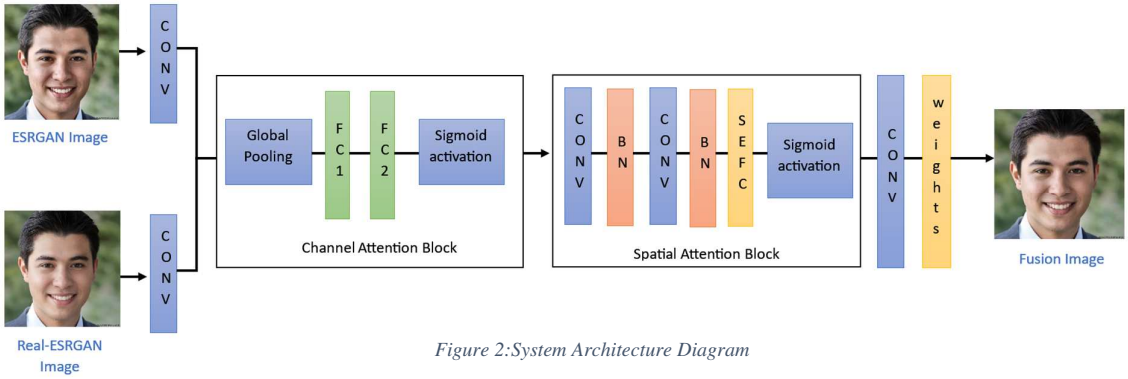

Figure 2:System Architecture Diagram

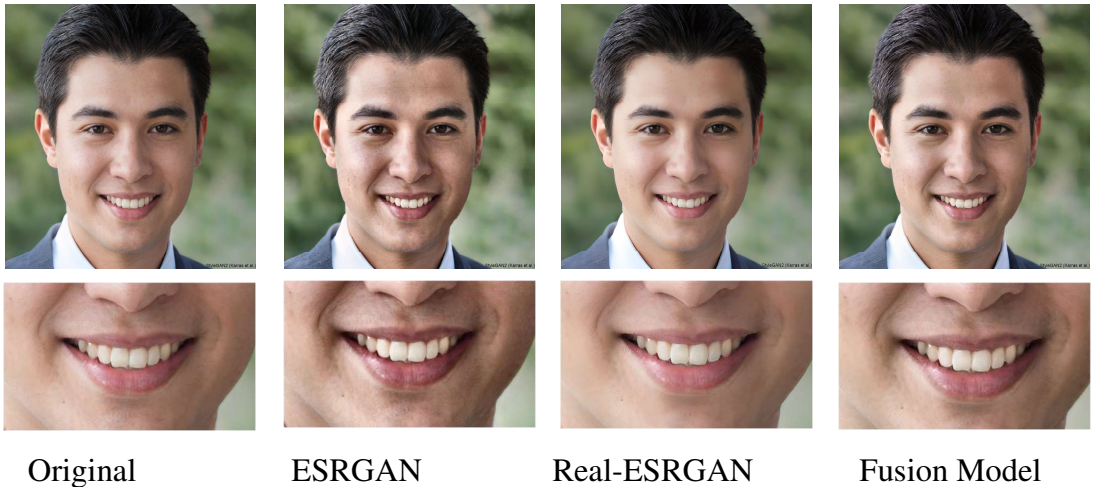

Figure 3:Visual Comparison of Fusion model with traditional models

computed distance between two vectors drops below a predefined threshold, the system classifies the pair as a match; otherwise, they are classified as a mismatch. Upon successful identification, the details of the matched criminal, along with the current location, date, and time, are retrieved from the database and immediately communicated to law enforcement officers via SMS using Python's Twilio library. In the event of a mismatch, the framework continues to iterate over other detected faces within the same CCTV frame, performing the same verification procedure against all criminal records stored in the database. The adoption of the Siamese network [14] enables practical implementation of One-Shot Learning, wherein the model is trained not to classify individual identities but to assess the similarity between two facial images. This allows the system to distinguish between individuals even with minimal reference images.

The dataset used to train the Siamese model [14] is curated from two distinct datasets. Positive image samples are obtained from the VGGFACE2 dataset [13], which has been modified to include 540 individuals, with an average of 170 images per person. For negative samples, images are sourced from the LFW dataset [15], comprising 5,749 images of various individuals. For training purposes, 80% of the combined dataset is utilized.

D. CycleGAN for criminal sketches

Many existing algorithms encounter limitations when dealing with criminal sketches, which are often the only available visual cues during the initial phase of an investigation. To address this challenge, the proposed pipeline incorporates CycleGAN [7] to facilitate sketch-to-image translation. The CycleGAN architecture [7] is chosen due to its capability to learn mappings without requiring paired training data, making it particularly effective for forensic scenarios where such pairs may not exist. Prior to processing, all images are resized and normalized to 250×250 dimensions to maintain consistency in input size for the model. The CycleGAN model [7] is trained using the publicly available "Person Face Sketches" dataset from Kaggle, which contains 20,655 labeled samples, each comprising a sketch and its corresponding real-life facial image. For model evaluation and testing, 80% of the dataset is sampled and reserved prior to the training phase.

The proposed CycleGAN [7] model consist of two generator functions. Generator G, converts a sketch (X) to real image (Y') and Generator F, Converts a real image (Y) to sketch (X'). The generator function uses a CNN which has 3 initial convolutional layers for feature extraction and Multiple residual blocks to learn structural transformations. ReLU activation function is applied after convolutional layers for introducing non-linearity and Tanh activation is used to generate smooth images. Similarly, two discriminators are used in the process to discriminate between real and generated images. The VGG Perceptual Loss ensures that the generated images (Y' and X') maintain perceptual similarity to real images which compares deep feature representations, rather than pixel-wise differences. Hinge Loss is Applied to both the discriminators to improve stability in adversarial training, this helps to prevent mode collapse and ensures diverse realistic image generation. The model is optimized using the Adam algorithm during training which helps to stabilize the gradient updates and accelerates convergence process. The Laplacian loss is used for better edge detail preservation in the output image. During the inference, criminal sketch is passed through generator G, which synthesizes a face image with realistic facial details.

To further enhance the realism of the generated facial image, the output images of CycleGAN [7] is passed through GFPGAN (Generative Facial Prior GAN) [10] which is a face restoration model. It is made efficient using a pre-trained StyleGAN-based prior and this helps to enhance facial details while maintaining naturalness. It works by using a multi-stage U-Net architecture, where low-quality facial inputs are first mapped to a latent space, then those are refined using high-quality priors learned from real faces during the training. This model ensures that fine details such as skin texture, eye sharpness, and facial symmetry are preserved while reducing unwanted artifacts. GFPGAN [10] balances perceptual quality and fidelity which makes the synthesized face images more suitable for real-time matching applications. In the final stage, the GFPGAN [10] processed images(Fig.4) are stored in database to match against the criminals in real-time. GANs, like CycleGAN and GFPGAN, play a vital role in converting sketches to photorealistic images. CycleGAN learns the structural mappings using adversarial training, enables it to function without the paired data.

IV. RESULTS AND DISCUSSION

A. Results of Fusion Model

For evaluating the effectiveness of the proposed Fusion model in comparison with the baseline Siamese model [14], 20% of the negative images from the LFW dataset [15] and positive images from the VGGFACE2 dataset [13] were utilized. These datasets were separated prior to the training process to ensure unbiased evaluation.

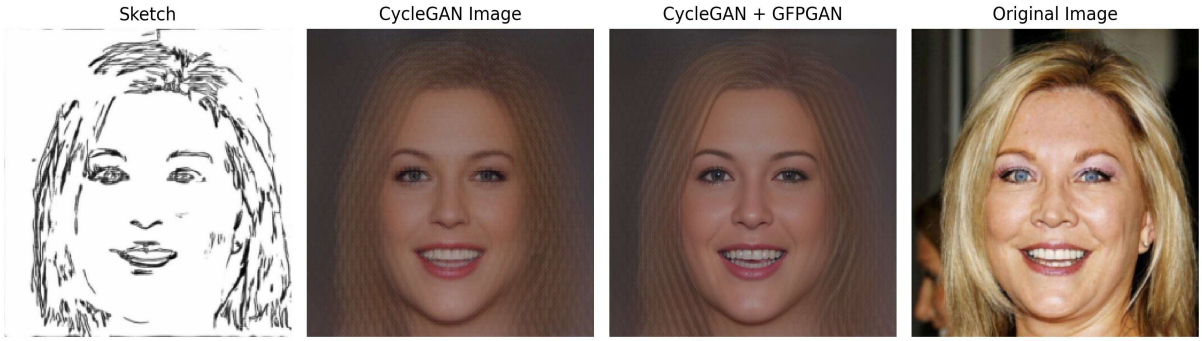

Figure 4: Results of Sketch to realistic image style transfer and comparison of it against the original image

The test setup involves verifying an anchor image against 30 positive and 30 negative images. This process was repeated for all 540 anchor images, and performance metrics were computed accordingly. Furthermore, bar graphs were plotted to compare the results of the Fusion model with the outputs generated by ESRGAN [8] and Real-ESRGAN [9]. The Siamese model addresses variations such as occlusion and lighting using various images with different lightings being involved in training process. Regularization via dropout and batch normalization was applied to avoid overfitting. However, the performance slightly drops under extreme occlusion or aging conditions. Latency in live systems averages 1.8 seconds per verification cycle on system equipped with RTX 3060. Reliability was ensured in this model by evaluating the system across two independent datasets, verifying consistent performance using standardized metrics including PSNR, SSIM, and LPIPS.

1) PSNR (Peak Signal-to-Noise Ratio)

PSNR, essentially compares the best possible signal strength of an image to the level of unwanted noise that distorts it. Higher PSNR values signify superior image reconstruction quality. It is computed as the logarithmic decibel the proportion of the maximum achievable signal strength intensity to the mean squared error (MSE) between the enhanced and original images. The proposed model demonstrates a 20.63% improvement over ESRGAN and records a 4.95% lower PSNR than Real-ESRGAN (Fig.5).

2) SSIM (Structural Similarity Index Measure)

SSIM assesses the similarity between two images by analyzing their structural patterns and luminance levels, with values that span from 0 to 1. It evaluates brightness, contrast, and structural information to determine perceptual similarity. The Fusion model shows a 3.23% improvement over Real-ESRGAN and achieves an SSIM score equivalent to that of ESRGAN (Fig.6).

3) LPIPS (Learned Perceptual Image Patch Similarity)

LPIPS quantifies perceptual variations between two images by comparing deep feature representations derived from a previously trained neural network. Rather than comparing individual pixels, it assesses how similar two images are in high-level visual perception. Lower LPIPS values indicate higher perceptual similarity. The Fusion model shows a significant improvement of 14.86% over ESRGAN and 49.74% over Real-ESRGAN (Fig.7).

4) AUC (Area Under Curve) and EER (Equal Error Rate)

AUC indicates the model's effectiveness in differentiating between positive and negative image pairs during face verification. It graphs the True Positive Rate (TPR) against the False Positive Rate (FPR) across varying classification thresholds. EER denotes the point where the point where the False Acceptance Rate (FAR) and False Rejection Rate (FRR) meet, with lower EER indicating superior verification performance.

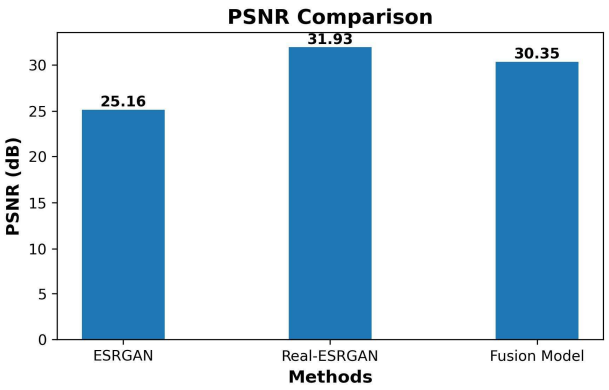

Figure 5: Bar Graph for Fusion model using PSNR Score

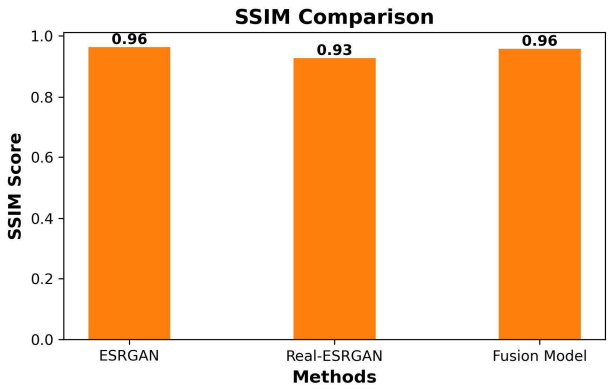

Figure 6: Bar Graph for Fusion model using SSIM Score

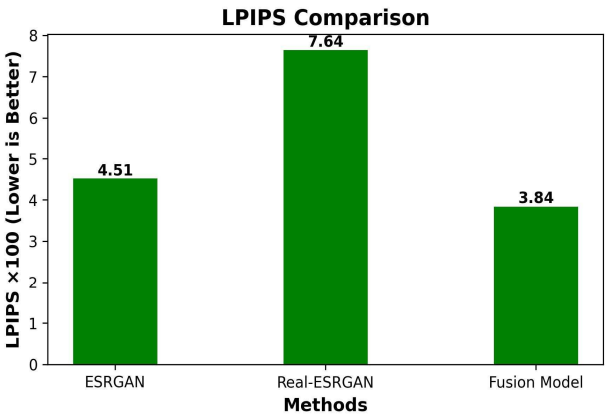

Figure 7:Bar graph for Fusion model using LPIPS Score

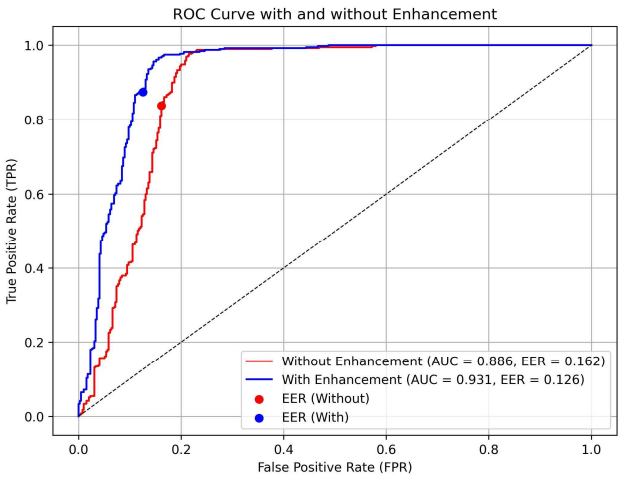

Figure 8: ROC Curve

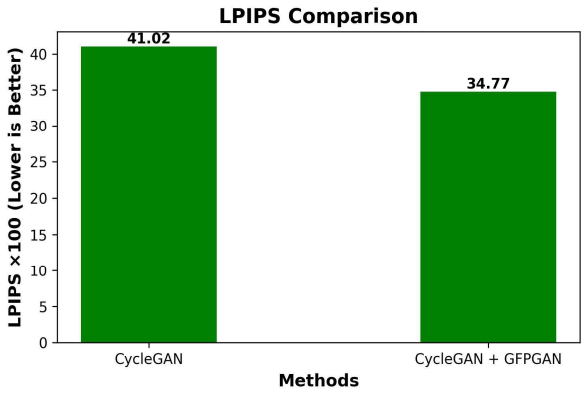

Figure 9: Bar Graph for CycleGAN using LPIPS Score

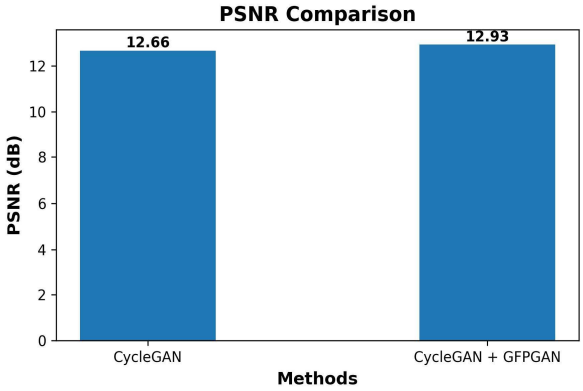

Figure 10: Bar Graph for CycleGAN using PSNR Score

ROC curve plotted for the Siamese model with Fusion model enhancement shows an increase in AUC by 5.08% and reduction in EER by 22.22% (Fig.8).

The results reveal that while the proposed Fusion model records a marginally lower PSNR than Real-ESRGAN, this is attributed to its emphasis on perceptual realism over strict pixel-wise accuracy. The slight PSNR drop is an acceptable trade-off for achieving visually natural and consistent image enhancements. These findings suggest that the Fusion model offers appreciable improvements over traditional methods, particularly in scenarios involving low-frequency images such as CCTV footage. However, the model exhibits limitations under extreme low-light conditions and when significant facial occlusions are present. All models were evaluated using unseen test splits which were split prior, ensuring unbiased performance and No cross-validation were applied. In (Table:1) Positive values indicate that how much percentage the Fusion Model performs better compared to each baseline methods and whereas negative values indicate that slight underperformance of the Fusion Model in the corresponding metric.

Measures\Methods	ESRGAN	Real-ESRGAN
PSNR	20.63%	-4.95%
SSIM	0.00%	3.23%
LPIPS	14.86%	49.74%

Table 1:Percentage improvement of the proposed Fusion Model over baseline methods (ESRGAN and Real-ESRGAN) across PSNR, SSIM, and LPIPS metrics.

B. Results of CycleGAN and GFPGAN

The authenticity of the generated realistic images was additionally assessed by comparing the outputs of CycleGAN [7] against the ground truth images. Additionally, the results of CycleGAN enhanced with GFPGAN [10] were assessed using LPIPS and PSNR metrics where the original image is used as a base image and then it is compared against CycleGAN and CycleGAN with GFPGAN. Similarly, the evaluation was carried out across 25 batches, and the results were averaged to ensure consistency and reliability in the performance metrics. The results are visualized using bar graphs, which indicate a 15.24% improvement in LPIPS (Fig. 9) and a 2.13% improvement in PSNR (Fig. 10) when the CycleGAN outputs are post-processed using GFPGAN [10].

Qualitative analysis reveals that the combination of CycleGAN and GFPGAN produces facial images with finer details and structurally accurate features. Nevertheless, the model faces challenges in replicating clean hair details due to CycleGAN's dependence on the global structural transformations, which limits its ability to reconstruct high-frequency components like hair textures.

V. CONCLUSION

Conventional CCTV footage is typically used for investigation only after a crime occurs. The proposed methodology enables real-time recognition and tracking of previously identified criminals, helping to prevent potential offenses. It demonstrates improved accuracy in handling low-quality images and criminal records that include only sketches. The model is suitable for deployment in public or private spaces prone to criminal activities, contributing to proactive surveillance.

The proposed system can be further improved by organizing the primary criminal dataset into smaller subsets based on various attributes such as age, gender, height, and other relevant features. This classification would enable the model to operate more efficiently and precisely. Additionally, the CycleGAN-based style transfer and criminal detection pipeline can be generalized by incorporating diverse facial datasets from different geographic regions, improving the model's adaptability to varying facial structures, skin tones, and hairstyles. The CycleGAN [7] model can also be enhanced by training on sketches of different types and artistic styles, facilitating more versatile and accurate image synthesis.

VI. ACKNOWLEDGMENT

This study makes use of publicly accessible datasets, such as VGGFace2, LFW, and the Kaggle Face Sketch dataset, all of which are licensed for academic use. No private or personally identifiable data were collected or involved in this research. All experiments and data processing were conducted in full adherence to ethical research guidelines. Furthermore, the system's communication layer, which handles alert delivery via the Twilio API, employs AES encryption to maintain the confidentiality and integrity of the transmitted data. We extend our gratitude to the developers and contributors of these datasets and tools, whose valuable work has greatly supported the progress of this research.

VII. REFERENCES

[1] Kumar, K.K., Kasiviswanadham, Y., Indira, D.V.S.N.V., Palesetti, P.P., and Bhargavi, Ch.V, "A Deep Learning Approach to Criminal Face Identification via a Multi-Task Cascade Neural Network (MTCNN)," Materials Today Proceedings, Volume 80, Section 3 (2023).

[2] Xiaoguang Li , Ning Dong , Jianglu Huang , Li Zhuo , Jiafeng Li , "A discriminative self-attention cycle GAN for face super-resolutionand recognition", IET Image Processing, Volume 15, Issue 11, in the year 2021.

[3] Chengkun Song , Zhujiang He , Yinghuai Yu , Zhenni Zhang , "Low Resolution Face Recognition System Based on ESRGAN" , IEEE proceedings of the 2021 (ICAIS).

[4] H. D. Hapsari, A. D. Wicaksana, H. F. Faylasuf, A. Yuaziva, R. M. Adzani, E. P. Giri, and G. P. Mindara, "Enhancing Low-Resolution Facial Images for Forensic IdentificationUsing ESRGAN" , International Journal of Multilingual Education and Applied LinguisticsVolume. 1, Nomor. 4, Tahun 2024.

[5] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-ESRGAN: Training Real-World Blind Super-Resolution With Pure Synthetic Data" , Proceedings of the Proceedings of the IEEE/CVF (ICCV) Workshops, 2021, pages 1905–1914.

[6] Xingfang Yuan , "Translation from sketch to realistic photo based on CycleGAN" , contributions from the Fourth International Conference on Signal Processing and Machine Learning, 2024.

[7] Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros , "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks" , IEEE International Conference on Computer Vision (ICCV), 2017.

[8] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, Chen Change Loy, "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks" , European Conference on Computer Vision Workshops (ECCVW), 2018.

[9] Xintao Wang, Liangbin Xie, Chao Dong, Ying Shan , "Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data", IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 2021.

[10] Xintao Wang, Yu Li, Honglun Zhang, Ying Shan , "Towards Real-World Blind Face Restoration with Generative Facial Prior" , IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

[11] Jiankang Deng, Jia Guo, Evangelos Ververas, Irene Kotsia, Stefanos Zafeiriou, "RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild", IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[12] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, Enhua Wu, "Squeeze-and-Excitation Networks", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[13] Qiong Cao, Li Shen, Weidi Xie, Omkar M. Parkhi, Andrew Zisserman, "VGGFace2: A Dataset for Recognising Faces across Pose and Age", International Conference on Automatic Face and Gesture Recognition (FG), 2018.

[14] Gregory Koch, Richard Zemel, Ruslan Salakhutdinov , "Siamese Neural Networks for One-Shot Image Recognition", International Conference on Machine Learning (ICML) Deep Learning Workshop, 2015.

[15] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller, Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments", University of Massachusetts, Amherst, Technical Report, 2007.