MOBILE AND FORENSICS
Vol. 8, No. 1, March 2026, pp. 50-64

Forensics of Low-Quality Facial Images from CCTV Using The Generative Adversarial Network (GAN) Method

Muhammad Adil Kustian[a,1,*], Ahmad Luthfi[a,2]

[a] Master's program in Informatics, Universitas Islam Indonesia, Yogyakarta, Indonesia
[1] 21917034@students.uii.ac.id*; [2] 125230405@uii.ac.id
* corresponding author

ARTICLE INFO

Article history
Received : October 25, 2025
Revised. : December 20, 2025
Accepted : March 10, 2026

Keywords
CCTV
Digital Forensics
Generative Adversarial Network
Real-ESRGAN
GFPGAN

ABSTRACT

CCTV facial images often suffer from poor quality due to low resolution, motion blur, and poor illumination, complicating forensic investigation and identification. This situation necessitates more modern image restoration methods. This study proposes a Generative Adversarial Network (GAN)-based pipeline that combines two architectures, Real-ESRGAN for resolution enhancement and GFPGAN for more natural facial feature recovery. Experimental results show significant improvements in perceptual quality with a decrease in NIQE values from 12.56 to 7.81 and BRISQUE from 70.81 to 44.23, with an 82% image recovery success rate. Additional evaluations using texture entropy and gradient histograms demonstrate consistency in facial structure and edge sharpness. This study contributes by demonstrating an integrated two-stage GAN approach as an effective solution for face recovery in low-quality CCTV images, while highlighting the need for standardized forensic protocols and facial identity validation in real-world applications. Thus, this pipeline can serve as a pre-processing stage to improve image readability and has the potential to be used as a tool for forensic investigations.

1. Introduction

Closed-Circuit Television (CCTV) is a surveillance system that functions to monitor and maintain security in various environments such as rooms, commercial areas, and infrastructure [1]. Since the 1990s, the application of Closed-Circuit Television (CCTV) has expanded globally with the main goal of strengthening crime prevention efforts, supporting the effectiveness of law enforcement processes, and reducing the level of public concern about criminal acts [2]. This technology has an important role in the field of surveillance and security, including in supporting forensic investigation processes [3]. However, the limitation of CCTV is that it often produces low-quality images [4]. Low-quality images on CCTV include motion blur, occlusion, low-light conditions, compression artifacts, and low resolution. This degradation is not artificially generated, but occurs naturally from operational constraints in the real world [5], [6].

Low-quality CCTV is one of the main obstacles in forensics, because the resulting images often do not contain sufficient visual detail to support the process of investigating the perpetrator or understanding the chronology in detail [7]. In it is also emphasized that CCTV recordings that is of low quality, may limit its usefulness in the investigation process or individual identification [8]. Especially the face, the face is the most sensitive component of personal identity, because it directly represents the individual [9]. Low quality facial images in CCTV can be caused by occlusion, noise, low resolution, and uneven lighting [10]. Therefore, low-quality facial images can hinder the process of accurately investigating criminals [11].

One approach to addressing this problem is to use a Generative Adversarial Network (GAN). GAN is a subset of deep learning and can generate new data consisting of two types of data network,

https://doi.org/10.12928/mf.v8i1.14805

namely a generator and discriminator that can be used in image restoration in various situations [12], [13]. The two GAN models used are Real-ESRGAN and GFPGAN. Real-ESRGAN is a part of GAN which is a development of SRGAN and ESRGAN. Real-ESRGAN introduces a new technique called blind super- resolution (blind SR) which functions to improve the quality of low-resolution images [14]. Just like Real-ESRGAN, GFPGAN is a sophisticated model that is also used for image degradation removal, image processing and feature extraction, but is more commonly used for facial restoration or recovery [15], [16].

Both models show promising technical potential. However, in the context of forensic image quality restoration [17], highlights that the application of generative models such as GANs carries great potential to assist the digital investigation process, particularly in the recovery of damaged or unclear visual data. However, this technology also poses new challenges regarding the authenticity of the results and the ethical responsibilities in its use in the forensic realm, in addition, GANs can also cause hallucinations in high-resolution details that do not correspond to the actual facial structure [18]. In line with the SWGDE Best Practices guidelines, image enhancement or image restoration in the forensic context is positioned as a technique to improve visual readability, so that the restoration results are used as a forensic analysis tool [19]. Therefore, GAN-based technology needs to be directed carefully as a forensic investigation tool, not as a stand- alone form of legal evidence.

Several previous studies have examined facial image restoration using GAN-based models. In research [20] aimed to develop a Real-World Super-Resolution (RWSR) framework for faces from CCTV footage by constructing realistic LR/HR data pairs through estimation of blur, noise, and compression artifacts. The results showed significant improvements, but the synthetic degradation approach used did not fully represent real-world conditions, resulting in the model still failing on images with motion blur, extreme lighting, or out-of-focus. The study only highlighted the general image improvement using a single model without testing the model combination in a pipeline, while the use of synthetic test data made the model less adaptive to the actual characteristics of CCTV footage. The authors also suggested further development through attention mechanisms and image degradation variations to make the model more robust. Meanwhile, in research [21], the GAN method was used to reduce noise in CCTV images. The results showed that GAN can reduce noise in CCTV images with the ability to visually enhance quality and important details, although it still has limitations in handling complex degradation.

In line with this, research [22] reviewed deep learning methods for Face Restoration (FR) based on degradation types such as denoising, super-resolution, deblurring, and artifact removal. Most models still rely on synthetic test data and have not studied the application to CCTV images that have noise and motion distortion. Although Transformer and GAN-based models provide high visual results, both are still limited by computational complexity. This research emphasizes the importance of human perception-based evaluation metrics, especially for non-reference images such as CCTV footage in forensic and security contexts. From these studies, it was identified that previous research still uses test data. Synthetic models that do not reflect real-world conditions and only use a single GAN model, resulting in less than optimal results. The limitations of this single approach are reinforced by the findings. Next. Research [23] states that Real-ESRGAN is not optimal when used alone, because it tends to produce artifacts and is ineffective when the image conditions experience complex degradation. Meanwhile, states that GFPGAN has proven effective in performing facial restoration by utilizing generative facial priors, but its performance is highly dependent on the quality of the initial input [24]. In facial images that have lost a lot of geometric structure, GFPGAN is not able to restore the face well. In a forensic context, these limitations indicate that a single approach has not been able to perform image restoration consistently in complex degradation conditions, especially for facial objects from CCTV that suffers from a combination of low lighting, motion blur, and limited resolution. Unlike previous studies that only used a single GAN model, this study proposes an approach in the form of a Real-ESRGAN and GFPGAN pipeline tailored for low-quality facial image recovery, which is expected to produce more optimal images and support the identification process and forensic investigation more accurately.

2. Method

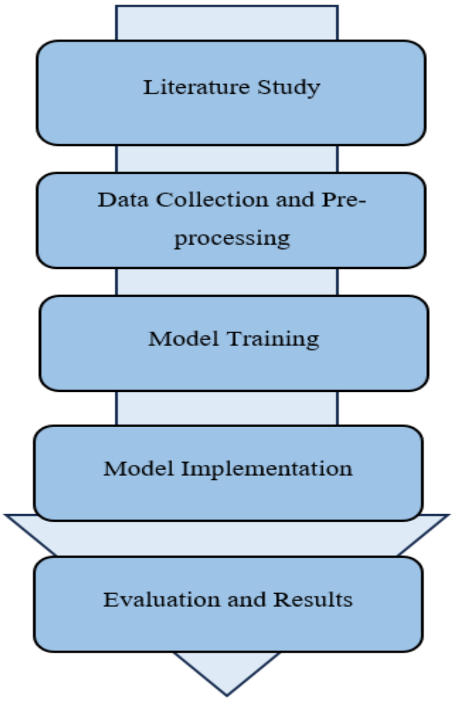


Fig. 1. Research Methodology

This research begins with a literature study as shown in Fig. 1, namely searching for references in the form of books or journals related to the relevant research topic, to identify problems in image forensics, especially the problem of facial image quality from CCTV results. Based on the literature study conducted, images originating from CCTV generally often experience low quality which complicates the forensic investigation and identification process so that it is necessary to improve the technology of facial visual quality so that it can be recognized more clearly. Next, the data collection process is carried out, where the collected data is then divided into two, namely training data and test data.

This research uses a two-stage pipeline based on Generative Adversarial Network (GAN), where Real-ESRGAN is used for super-resolution enhancement, and GFPGAN functions for facial feature restoration to make it look more natural. In general, this research process includes four main stages: (1) fine- tuning Real-ESRGAN using a high-resolution public dataset, (2) applying the pre-trained GFPGAN model, (3) implementing the pipeline on a CCTV test dataset, and (4) evaluating the results using perceptual metrics such as NIQE, BRISQUE, entropy, and gradient histogram.

2.1. Training Data

The training data used in this study is a combination of three datasets: DIV2K, Flickr2K, and OST (Outdoor Scene Training). These three datasets were chosen because they have excellent high resolution, providing a strong foundation for the model to accurately learn texture and structure details. Before using the data, preprocessing was performed, where HR images were preprocessed using an image degradation simulation method. This process produces low-quality (LR) images in model training. Meanwhile, HR images serve as ground truth or target outputs for comparison with model predictions.

Although the training datasets (DIV2K, Flickr2K, OST) are not sourced from CCTV sources, a synthetic degradation simulation process was performed to mimic the conditions of surveillance images. This step aims to minimize the effects of domain shift so that the model remains adaptive to the characteristics of CCTV facial images. This approach is also commonly used in surveillance image restoration research, as real CCTV-based datasets typically lack ground truth required for the supervised training process. Meanwhile, the GFPGAN model uses pre-trained data previously developed by Tencent ARC. This model was previously trained on high-resolution facial datasets like FFHQ and CelebA. By applying this pre-trained data, GFPGAN focuses on facial restoration.

2.2. Testing Data

Testing data in this research the QMUL-SurvFace facial dataset is a collection of low-quality facial images obtained from surveillance video or CCTV due to low resolution, motion, and insufficient lighting. This study used 200 images for testing, considering the limitations of available computing resources.

The selection was made from the QMUL SurvFace dataset to avoid duplicate images and to focus solely on faces, ensuring the test data remains representative and relevant to the model's performance testing objectives. Before testing, the entire test data set, consisting of 200 faces, was then grouped into a single folder to facilitate batch inference and evaluation. The details of the dataset used in the training data and testing data can be seen in Table 1.

Table 1. Dataset Details

No	Name Dataset	The Image Used	Objevtive Use
1	DIV2K	800	Fine-Tuning Real-ESRGAN
2	Flickr2K	2.650	Fine-Tuning Real-ESRGAN
3	OST (Outdoor Scene Training)	+3000	Fine-Tuning Real-ESRGAN
4	CelebA	30.000	Pre-Trained GFPGAN
5	FFHQ	+70.000	Pre-Trained GFPGAN
6	QMUL-SurvFace	200	Data Testing

The next step is to further train the Real-ESRGAN model. This model is built on the RRDBNet (Residual-in-Residual Dense Block Network) architecture with 23 residual blocks. The architecture and training flow of the Real-ESRGAN model can be seen in Fig. 2.

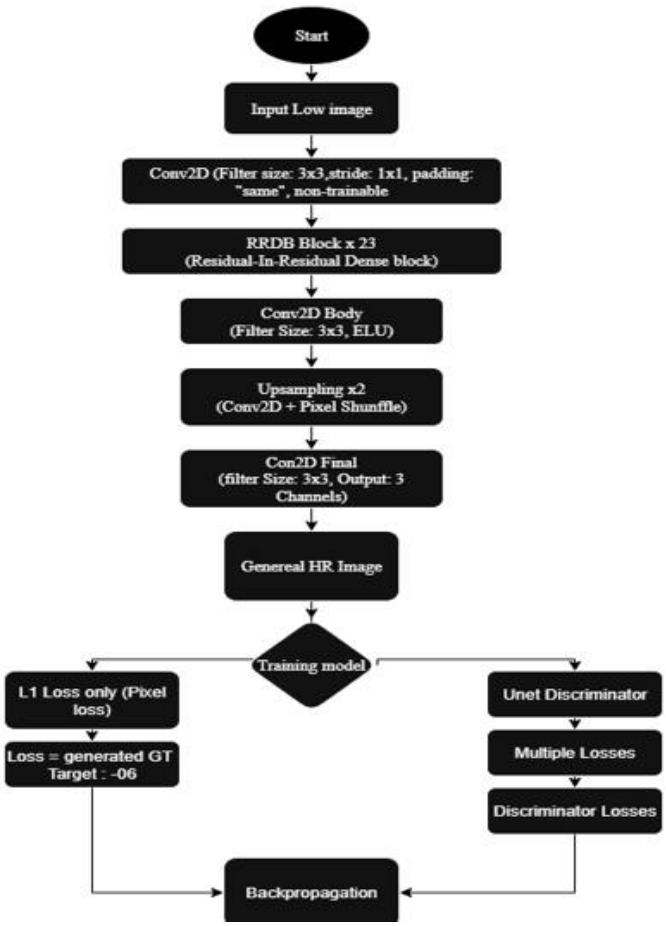


Fig. 2. Architecture And Process Training Real ESRGAN

The training process uses a process called fine-tuning by utilizing the Real-ESRGAN pre-trained model that has been previously trained on a common dataset. Training is carried out using the pytorch framework, adam optimizer, 1e-4 learning rate, loss functions that include pixel loss (L1 loss) to maintain the consistency of pixel restoration plus VGG 19-based perceptual loss and adversial loss to maintain the quality of the resulting image, and using the DIV2K, Flickr2K and OST (Outdoor Scene Training) datasets with a synthetic degradation pipeline to simulate the characteristics of low-quality images. The next stage is the facial restoration process using GFPGAN, but GFPGAN training is not done from scratch. Instead, a pre-trained GFPGAN model provided by (TencentARC) is used which can be accessed at https:// github.com/TencentARC/GFPGAN. This GFPGAN model uses a U-Net encoder-decoder architecture with StyleGAN2 priors, and is equipped with identity preserving loss, perceptual loss, and pixel-wise loss which function to restore facial details naturally and consistently. With this two-stage approach, Real-ESRGAN training and the use of pre-trained GFPGAN are able to effectively restore facial images from low-quality CCTV which can facilitate the process of digital forensic investigation and identification needs. At this stage of the process. Fourth, the model was implemented on the previously mentioned dataset and the Real-ESRGAN model was run which focused on super resolution and GFPGAN for face restoration using PyTorch on google Colab. The architectural diagram of the pipeline can be seen in Fig. 3.

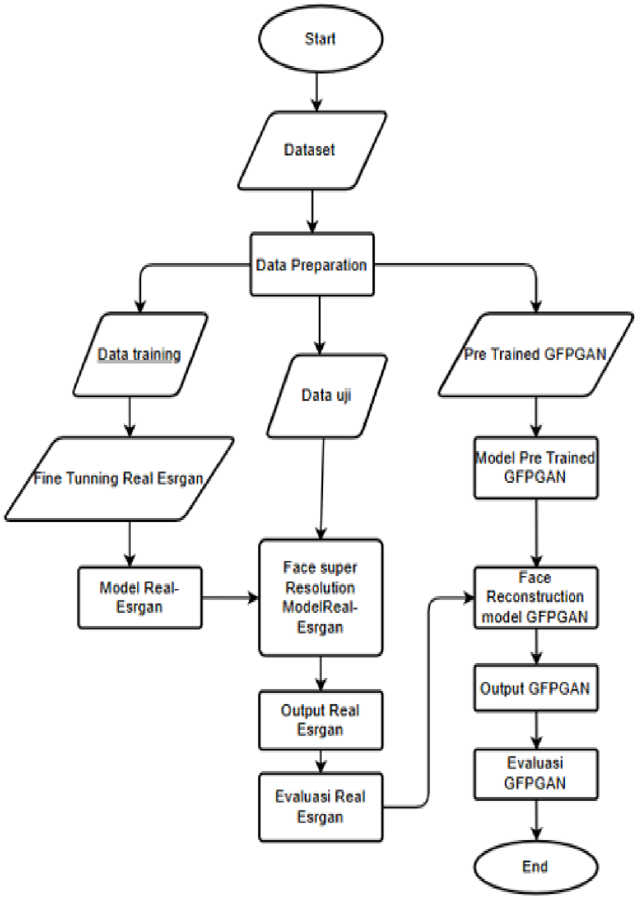


Fig. 3. Pipeline Architecture Diagram

Fig. 3 shows the pipeline architecture, where low-quality facial images from the QMUL-SurvFace dataset are first loaded as input. A fine-tuned Real-ESRGAN model is used to increase the image resolution to obtain high-resolution (HR) images. The results are then processed using a pre- trained GFPGAN model to restore facial structure and details for a more natural appearance. The restored images are then saved and evaluated using the non-reference metrics NIQE and BRISQUE, as well as additional analysis in the form of texture entropy and gradient histogram to assess the consistency of facial structure and edge sharpness.

The fifth stage, namely evaluation and results, aims to evaluate the extent to which the model is able to produce faces with a better level of clarity and structure compared to the image before processing.

1) Forensic Risk and Mitigation
GAN-based models have the potential to produce visual hallucinations, which are the appearance of synthetic features not present in the original image, such as changes in facial contours or skin texture. This risk could impact the reliability of forensic interpretation, especially since this study used a non-reference approach, using original comparison imagTeos. Mitigate this, this study utilizes GFPGAN's built-in regularization mechanism, which maintains the stability of facial structures through facial priors from StyleGAN2, so that facial features remain proportional and do not deviate. In addition, qualitative evaluations based on visual perception and non-reference metrics (NIQE, BRISQUE) were conducted to ensure that the restoration results did not produce artifacts or excessive distortion. This pipeline is limited to its use as an aid in forensic investigations, with the aim of facilitating facial identification from low-quality CCTV images, not as stand-alone legal evidence.

2) Software and Hardware Specifications
This study utilizes specific hardware and software to perform deep learning calculations, allowing for model training, inference, and evaluation in the context of facial image recovery for digital forensic analysis. Table 2 lists the hardware and software used in this study.

Table 2. Specification Device Hard and Soft Used

No	Type	Name
1	Hardware	CPU 4 Core
2	Hardware	Nvidia T100 16 GB x2
3	Hardware	RAM 32 GB
4	Software	Python 3.10
5	Software	Library Pytorch
6	Software	Windows

3. Results and Discussion

Experiments were conducted using a combination of the Real-ESRGAN and GFPGAN pipelines. Quantitative evaluations were performed using two non-referenced metrics: Natural Image Quality Evaluator (NIQE) and Blind/Referenceless Image Spatial Quality (BRISQUE). The formula used to calculate the score from the metrics used is as follows:

Here is the Niqe Formula Used

$$NIQE(I) = \sqrt{(fI - \mu_n)T_{\Sigma}{}_{n}^{-1}(fI - \mu_n)}$$

With:

fI : feature NSS image test

$\mu\, n\, , \sum n$: Mean and covariance feature data dataset natural image high quality.

Following is Formula Brisque Which used

$$f(x; \alpha, \sigma^2) = \frac{\alpha}{2\beta T\,(1/\alpha)} \exp\,(-(\frac{|x|}{\beta})^{\alpha} \qquad (2)$$

With:

α : shape parameters

β : $\sigma\sqrt{\frac{r(1/\alpha)}{r(3/\alpha)}}$

T (.) : main function

Facial images from the QMUL-SurvFace dataset are used as input in a study to restore facial images low-quality faces. The model's output is a facial image with clearer details. The test data used in this study can be seen in Fig. 4.

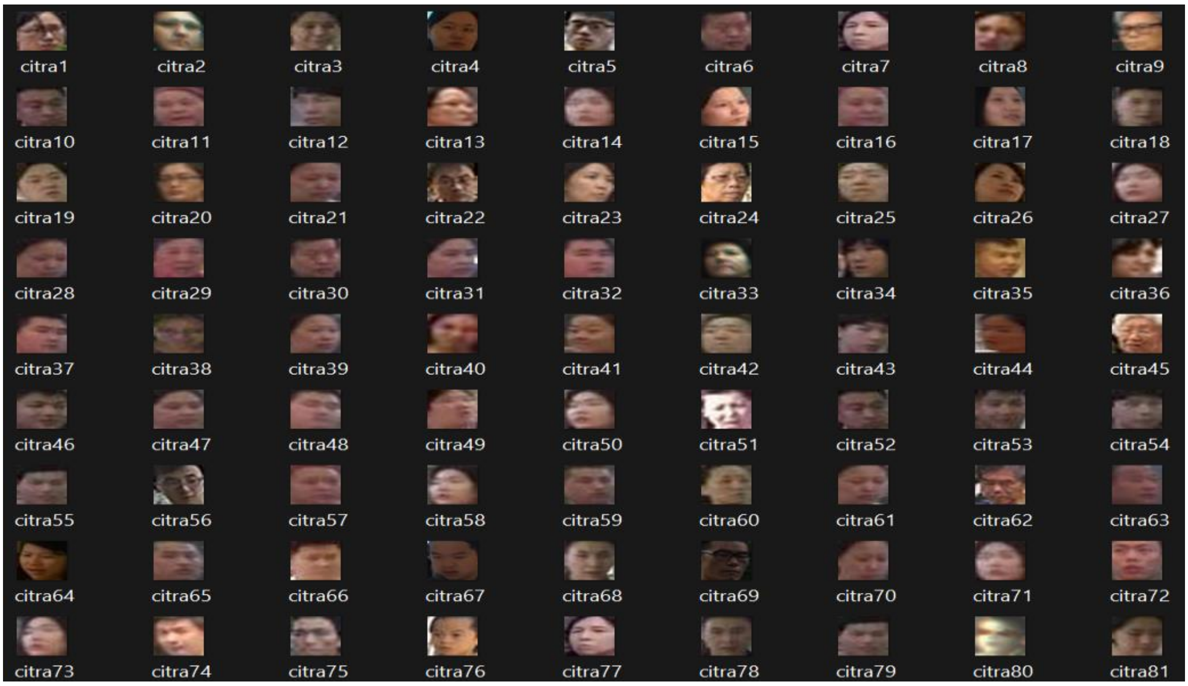


Fig. 4. Test Data Used

The images in Fig. 4 were used as test data to measure the performance of the Real-ESRGAN and GFPGAN models. Twenty-one data sets were processed to test the model performance, and the results obtained from these data sets can be seen in Table 3.

Table 3. Result Real-ESRGAN and GFPGAN

Input (original image)				
Metrics	**Average**	**Minimum**	**Maximum**	**Stand Deviation**
Niqe	12.56	10.11	14.24	0.64
Brisque	70.40	64.48	80.05	2.93
Real-ESRGAN				
Niqe	11.46	5.42	17.67	2.39
Brisque	58.56	6.10	82.50	12.08
GFPGAN				
Niqe	7.81	4.78	16.00	2.19
Brisque	44.23	27.44	81.35	10.86

Table 3 shows that the input image has an average NIQE value of 12.56 and a BRISQUE value of 70.31, indicating low image quality. After processing using Real-ESRGAN, the average NIQE value decreased to 11.46 with a range of 5.42 – 17.67 and BRISQUE to 58.56 with a range of 6.10 – 82.50. Furthermore, the results of Real-ESRGAN reprocessed using GFPGAN showed a further decrease in values, namely NIQE of 7.81 (range 4.78–16.00) and BRISQUE of 44.23 (range 27.44–81.35). The decrease in NIQE and BRISQUE values at each stage indicates an increase in the perceptual quality of the image after layered processing with Real-ESRGAN and GFPGAN. However, the increase is not uniform, as seen from the increased standard deviation and the maximum value of BRISQUE GFPGAN (81.35) which is higher than the original image (80.04).

Evaluation focused on intra-model improvements (Original ÿ Real-ESRGAN ÿ GFPGAN) without comparison with other baseline methods, because the QMUL-SurvFace dataset has low-quality surveillance image characteristics that are significantly different from general high-resolution datasets, so that cross-method comparisons will not provide scientifically equivalent results.

The visualization results of the process that has been carried out can be seen in Fig. 5.

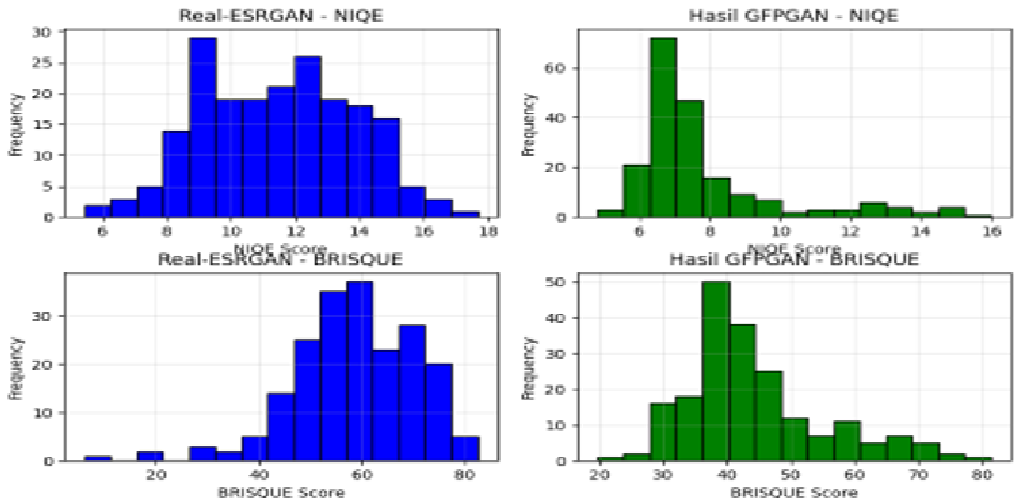


Fig. 5. Result Visualization from Real-ESRGAN And GFPGAN

Fig. 5 shows a visualization of the distribution of NIQE and BRISQUE metrics at each processing stage, namely Original, Real-ESRGAN, and GFPGAN. It can be seen that the original stage has relatively high NIQE and BRISQUE values with a narrow distribution, indicating that the image quality is still low. In the Real-ESRGAN stage (blue), the distribution of NIQE and BRISQUE values becomes wider, indicating that the image quality improvement varies between frames. Meanwhile, the GFPGAN stage (green) shows a more centralized distribution of values and is in the lower score range, indicating that GFPGAN produces images with more consistent and stable recovery compared to the original image.

Based on the test results on 200 low-quality facial image data, the recovery process was carried out using the Real- ESRGAN and GFPGAN pipelines. The amount of data that was successfully recovered can be seen in Table 4.

Table 4. Number of Data Successfully Recovered

Category	Amount Image	Percentage
Total test data	200	100%
Successfully recovered	164	82%
Not Successful	36	18%

Table 4 shows the number and percentage of images that were successfully and unsuccessfully recovered. The success criteria were determined based on the metric values between the input images and the pipeline results, as well as visual evaluation during the testing phase. Of the 200 test images, 164 (82%) were successfully recovered, while 36 (18%) were not.

These results indicate that the method used has a fairly high success rate in recovering low-quality facial images, although there are still a small number of images that cannot be optimally recovered. To facilitate visualization and interpretation of the results, 200 test images were grouped into 10 categories with each category containing 20 images based on the test sequence, as shown in Fig. 6.

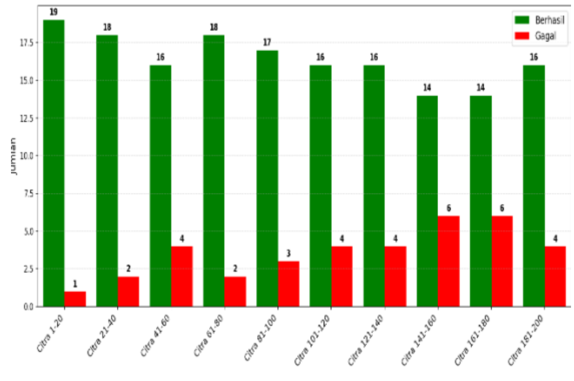


Fig. 6. Distribution of the Number of Successful and Failed Images

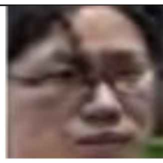
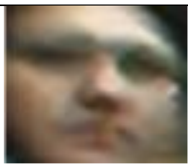
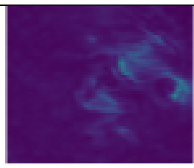
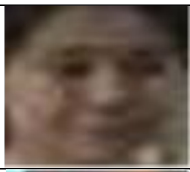
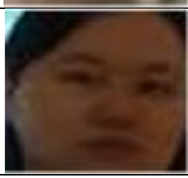
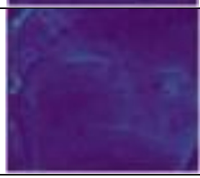
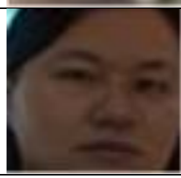
Based on Fig. 6, it can be seen that the success rate is quite consistent across most groups, with the number of successful images ranging from 14-18 images per group. However, there are two groups that show a higher failure rate, namely the 141-160 Image group and the 161-160 Image group 180 with 6 failed images each. Overall, this distribution indicates that the applied image recovery method has stable performance with an average success rate of 82%.

Some images that cannot be recovered properly are caused by degradation that is outside the distribution of the training data (out-of-distribution complicated degradations). This condition occurs when the level of damage to the original image is very severe (severe degradation), so that the model is unable to reconstruct facial details optimally [25].

This research specifically focuses on restoring and improving facial image quality (image restoration), not on the face recognition stage. Therefore, identification accuracy testing was not conducted, as it was beyond the scope of this study. The evaluation focused on improving perceptual quality and visual structure as a foundation for further research in facial forensics.

This study also presents several image samples from the QMUL-SurvFace dataset that have been preprocessed, allowing us to see the quality of facial images before and after processing through the pipeline. The pipeline visualization results show the images before and after processing, along with visualizations. The heatmap to show the distribution of image improvement can be seen in Table 5.

Table 5. Visualization Results of the Original Image,Real-ESRGAN and GFPGAN Pipelines

Name	Original Image	Real-ESRGAN	Heatmap Real-ESRGAN	GFPGAN	Heatmap GFPGAN
Image 1					
Image 2					
Image 3					
Image 4					

The heatmap visualization in Table 5 shows that areas of high color intensity are primarily focused on the eyes, nose, and mouth. These three regions are key biometric landmarks that form the basis of modern forensic facial recognition systems. These regions contribute most to the facial recovery process, making the interpretation of the results more forensically meaningful. Thus, the GFPGAN model can highlight key structural features that can facilitate the identification process. However, it should be noted that GAN-based models have the potential to produce hallucinations of facial features, such as changes in contour shape or skin texture that are not present in the original image. This phenomenon needs to be considered because it can threaten visual authenticity and forensic admissibility. The results of the evaluation of the images representing the test data can be seen in Table 6.

Table 6. Results NIQE and BRISQUE

Name	Metrics	Image Original	Real-ESRGAN	GFPGAN
Citra 1	Niqe	13.09	9.31	6.49
	Brisque	73.85	46.70	28.51

Citra 2	Niqe	12.71	11.71	6.71
	Brisque	76.37	65.13	38.47
Citra 3	Niqe	13.68	11.02	6.61
	Brisque	69.50	60.73	41.12
Citra 4	Niqe	10.33	8.02	6.54
	Brisque	66.59	37.71	36.82

Based on Table 6, the calculation results of the NIQE and BRISQUE metrics for each image can be seen. In image 1, the NIQE value was initially 13.09 in the original image, decreased to 9.31 after Real-ESRGAN processing and again decreased to 6.49 in the GFPGAN. The same thing is seen in the metrics BRISQUE which is down from the original image which is 73.85 to 46.70 after processing with Real-ESRGAN to 28.50 on GFPGAN. The same thing is also seen in images 2, 3, and 4 where the NIQE and BRSIQUE values decrease after processing with Real-ESRGAN and decrease further after processing with GFPGAN. Thus, the results of the NIQE and BRISQUE metrics calculations show consistent improvements at each processing stage.

In order to strengthen the evaluation results of the metrics used, additional analysis was carried out using texture entropy and gradient histogram on image 1, which can be seen in Fig. 7. Texture entropy analysis useful for measuring the level of complexity and texture in an image while the gradient histogram serves to evaluate the distribution of edge strength and facial contour.

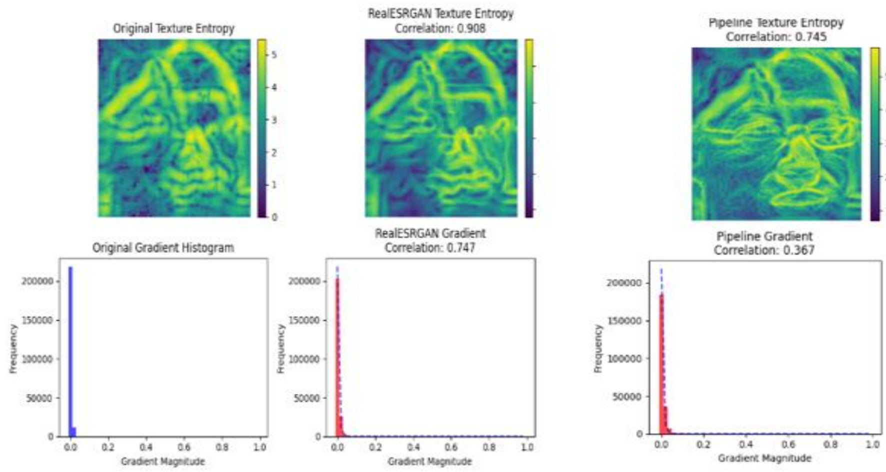


Fig. 7. Result Texture Entropy and Gradient Histogram on Image 1

Based on Fig. 7, it can be seen that the Real-ESRGAN correlation texture entropy is 0.908 and gradient 0.747, after the GFPGAN stage the value decreases where the texture entropy becomes 0.745 and the gradient is 0.367.

As for the texture entropy and gradient histogram in image 2, it can be seen in Fig. 8.

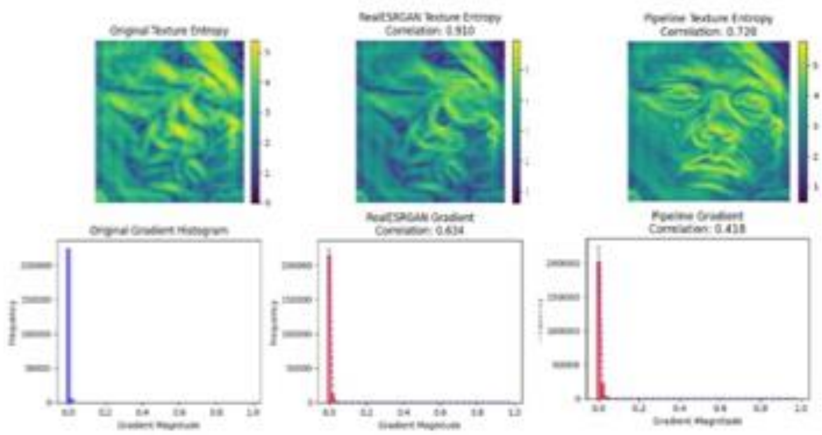


Fig. 8. Result Texture Entropy and Gradient Histogram on Image 2

Based on Fig. 8, it shows that the original image has a more dispersed distribution of texture entropy and gradient. After Real- ESRGAN processing, the entropy correlation is 0.910 and the gradient is 0.634. At the GFPGAN stage, the entropy correlation decreased to 0.720 and the gradient to 0.418. As for the texture entropy and gradient histogram in image 3, it can be seen in Fig. 9.

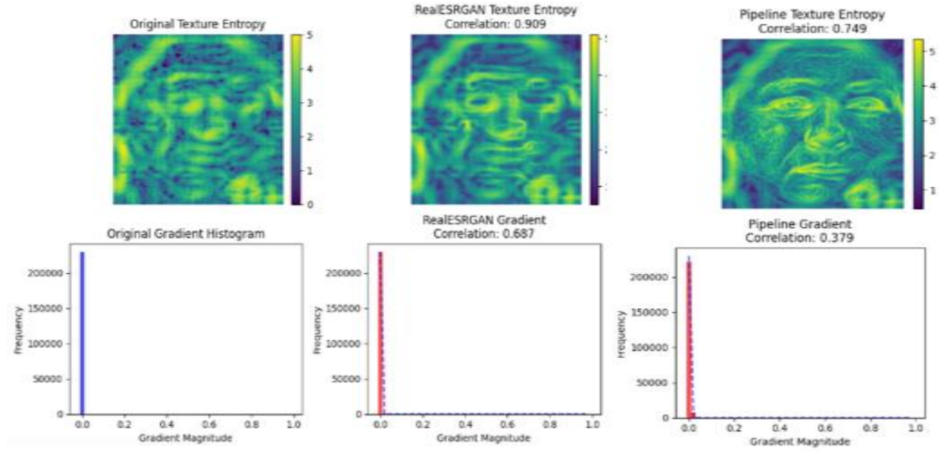


Fig. 9. Result Texture Entropy and Gradient Histogram on Image 3

Based on Fig. 9, it shows that the original image has a diffuse distribution of texture entropy and gradient. After processing with Real-ESRGAN, the correlation entropy is 0.909 and the gradient is 0.687, and in the GFPGAN stage, the correlation entropy decreases to 0.749 and the gradient is 0.379. As for texture entropy and gradient. The histogram in image 4 can be seen in Fig. 10.

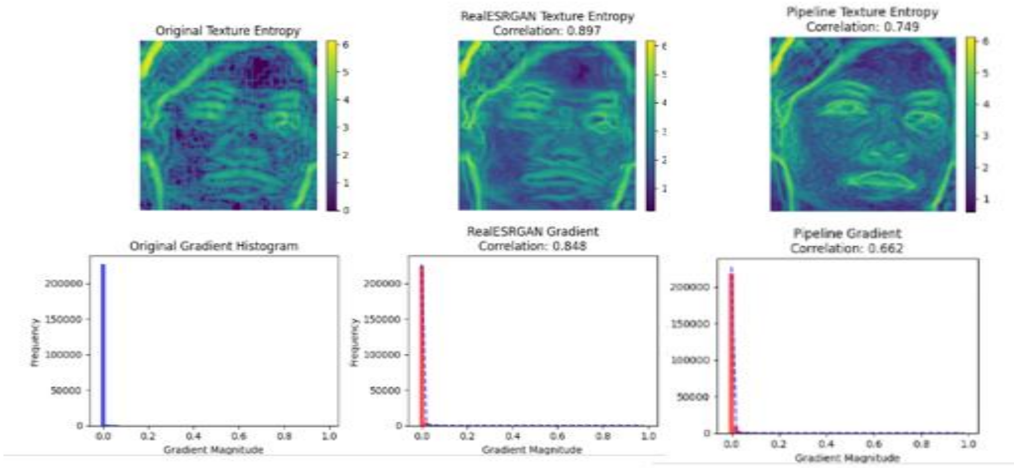


Fig. 10. Result Texture Entropy and Gradient Histogram on Image 4

Based on Fig.10, it shows that the original image has a diffuse distribution of texture entropy and gradient. After processing with Real-ESRGAN, the correlation entropy is 0.897 and the gradient is 0.897, 0.848, and at the GFPGAN stage the correlation entropy decreased to 0.749 and gradient 0.662.

Thus, the results from the four images show a consistent pattern. In the first stage, Real-ESRGAN produces higher entropy and gradient values than GFPGAN. Then, in the second stage, GFPGAN successfully balances the results with entropy and gradient lower. The difference in value between these stages emphasizes the complementary role of the two-stage pipeline, where Real-ESRGAN generates initial details and GFPGAN refines textures to make them more natural.

Overall, although the Real-ESRGAN and GFPGAN pipelines are effective in low-quality face recovery, the acceptability of the results in a forensic context still depends on the consistency of facial identity features and the transparency of the methodology. This research shows that GAN-based recovery can be a valuable tool for forensic investigations, but it must be used with a clear understanding of its limitations and risks.

Further validation using facial recognition systems, or face identification, and standardization of protocols for the use of AI in forensics are essential steps to ensure this technology can be safely and responsibly adopted in forensic practice. Comparisons with other methods were not conducted in this study due to fundamental differences in dataset characteristics. However, comparative research on other datasets with available identity ground truth could be a direction for further research to provide a more comprehensive benchmarking. Therefore, this study focuses on evaluating the effectiveness of the Real-ESRGAN and GFPGAN pipelines in the context of low-quality facial image restoration.

4. Conclusion

This study shows that the combination of Real-ESRGAN and GFPGAN is effective in restoring low-quality facial images from CCTV (NIQE:12.56ÿ7.81, BRISQUE:70.40ÿ44.23) with the success rate of 82% of the images was successfully restored, while 18% were suboptimal due to out-of-distribution complicated degradations. This condition occurs when the level of damage to the original image is very severe, making the model unable to optimally recover facial details. However, this study has several limitations: the evaluation is limited to perceptual metrics, the model was trained on a non-surveillance dataset, which creates a potential domain gap, and the risk of GAN hallucinations that can produce non-authentic facial features. Therefore, the recovery results need to be carefully reviewed for authenticity and admissibility of the evidence, and should be used as an aid to investigations accompanied by original images, not as primary evidence. For further research, it is recommended to test the accuracy of identity recognition (face recognition/ face identification) using recovered images, as well as evaluation with CCTV footage obtained directly in the field, so that the pipeline performance is more representative of real-world conditions. Overall, the Real-ESRGAN pipeline and GFPGAN has the potential to be a pre-processing stage forensics that are useful in improving image readability and as a tool to facilitate the forensic investigation and identification process, while still considering the reliability of identity and the integrity of digital evidence.

References

[1] I. E, C. Jacob, and R. R, “Facial Recognition and CCTV Integration for Enhanced Security Using Deep Learning Techniques,” in *2024 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, Kothamangalam, Kerala, India: IEEE, May 2024, pp. 1–5. doi: 10.1109/RAICS61201.2024.10689986.

[2] M. Tykesson, “Effects of CCTV on Fear of Crime: a Systematic Literature Review,” *Eur. J. Crim. Policy Res.*, Aug. 2025, doi: 10.1007/s10610-025-09633-0.

[3] F. Brookman and H. Jones, “Capturing killers: the construction of CCTV evidence during homicide investigations,” *Polic. Soc.*, vol. 32, no. 2, pp. 125–144, Feb. 2022, doi: 10.1080/10439463.2021.1879075.

[4] K. S. J. Aditya, C. S. Varma, S. P. Teja, P. R. Babu, R. Kuchipudi, and T. S. Murthy, “Low Quality CCTV Image and Video Enhancement using GANs and Audio Enhancement with Equalization Algorithms,” in *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS)*, Chikkaballapur, India: IEEE, Apr. 2024, pp. 1–3. doi: 10.1109/ICKECS61492.2024.10616988.

[5] Y. Han, J. Kang, S. Lee, and T. Kim, “Understanding Perceptual Quality in CCTV Images: A Benchmark Dataset and Entropy-based Insights,” presented at the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2025, pp. 3412–3419. Accessed: Mar. 05, 2026. [Online]. Available: https://openaccess.thecvf.com/content/ICCV2025W/VQualA/html/Han_Understanding_Perceptual_Quality_in_CCTV_Images_A_Benchmark_Dataset_and_ICCVW_2025_paper.html

[6] S. Powale, A. Dhanawade, S. Bagwe, S. Kawale, N. L. Chutke, and S. Chavan, “Person identification in low resolution CCTV footage using deep learning,” in *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, Greater Noida, India: IEEE, Dec. 2020, pp. 236–240. doi: 10.1109/ICACCCN51052.2020.9362764.

[7] S. Fahad *et al.*, "A Modified Singular Value Decomposition (MSVD) Approach for the Enhancement of CCTV Low-Quality Images," *IEEE Access*, vol. 12, pp. 20138–20151, 2024, doi: 10.1109/ACCESS.2024.3349477.

[8] A. Vinay, A. Lokesh, V. R. Kamath, K. N. B. Murty, and S. Natarajan, "Enhancement of Degraded CCTV Footage for Forensic Analysis," in *International Conference on Innovative Computing and Communications*, vol. 1165, D. Gupta, A. Khanna, S. Bhattacharyya, A. E. Hassanien, S. Anand, and A. Jaiswal, Eds., in Advances in Intelligent Systems and Computing, vol. 1165. , Singapore: Springer Singapore, 2021, pp. 617–636. doi: 10.1007/978-981-15-5113-0_50.

[9] X. Piao, Z. Piao, S. J. Yoo, and Y. H. Gu, "Robust sensitive-information de-identification framework based on relative-position estimation of objects in closed-circuit television videos," *Alex. Eng. J.*, vol. 89, pp. 172–183, Feb. 2024, doi: 10.1016/j.aej.2024.01.046.

[10] M. Farooq, M. N. Dailey, A. Mahmood, J. Moonrinta, and M. Ekpanyapong, "Human face super-resolution on poor quality surveillance video footage," *Neural Comput. Appl.*, vol. 33, no. 20, pp. 13505–13523, Oct. 2021, doi: 10.1007/s00521-021-05973-0.

[11] M. Leone, "From Fingers to Faces: Visual Semiotics and Digital Forensics," *Int. J. Semiot. Law - Rev. Int. Sémiot. Jurid.*, vol. 34, no. 2, pp. 579–599, Apr. 2021, doi: 10.1007/s11196-020-09766-x.

[12] S. Aldhaheri and A. Alhuzali, "SGAN-IDS: Self-Attention-Based Generative Adversarial Network against Intrusion Detection Systems," *Sensors*, vol. 23, no. 18, p. 7796, Sep. 2023, doi: 10.3390/s23187796.

[13] H. Yang, P. Shi, D. Zhong, D. Pan, and Z. Ying, "Blind Image Quality Assessment of Natural Distorted Image Based on Generative Adversarial Networks," *IEEE Access*, vol. 7, pp. 179290–179303, 2019, doi: 10.1109/ACCESS.2019.2957235.

[14] Z. Zhu, Y. Lei, Y. Qin, C. Zhu, and Y. Zhu, "IRE: Improved Image Super-Resolution Based on Real-ESRGAN," *IEEE Access*, vol. 11, pp. 45334–45348, 2023, doi: 10.1109/ACCESS.2023.3256086.

[15] X. Zhang and J. Feng, "A Novel Blind Restoration Method for Miner Face Images Based on Improved GFP-GAN Model," *IEEE Access*, vol. 12, pp. 104676–104687, 2024, doi: 10.1109/ACCESS.2024.3435351.

[16] H. Yadav and D. S. Sisodia, "Image Restoration with Conditional Noise Diffusion: Restoring Facial Clarity and Details," *Procedia Comput. Sci.*, vol. 258, pp. 1159–1168, 2025, doi: 10.1016/j.procs.2025.04.350.

[17] O. Emehin, I. Emeteveke, O. J. Adeyeye, and I. Akanbi, "Generative AI in Forensic Data Analysis: Opportunities and Ethical Implications for Cloud-Based Investigations," *Int. J. Res. Publ. Rev.*, vol. 5, no. 10, pp. 2941–2957, Oct. 2024, doi: 10.55248/gengpi.5.1024.2904.

[18] C.-C. Hsu, C.-W. Lin, W.-T. Su, and G. Cheung, "SiGAN: Siamese Generative Adversarial Network for Identity-Preserving Face Hallucination," *IEEE Trans. Image Process.*, vol. 28, no. 12, pp. 6225–6236, Dec. 2019, doi: 10.1109/TIP.2019.2924554.

[19] P. Burnard, "Prelude: Provocations in Creativity Research," in *International Handbook of Research in Arts Education*, vol. 16, L. Bresler, Ed., in Springer International Handbook of Research in Arts Education, vol. 16. , Dordrecht: Springer Netherlands, 2007, pp. 1175–1179. doi: 10.1007/978-1-4020-3052-9_80.

[20] A. Aakerberg, K. Nasrollahi, and T. B. Moeslund, "Real-world super-resolution of face-images from surveillance cameras," *IET Image Process.*, vol. 16, no. 2, pp. 442–452, Feb. 2022, doi: 10.1049/ipr2.12359.

[21] P. B S *et al.*, "Removing/Minimising Gaussian noise from real time CCTV footages using GAN," in *2024 First International Conference on Innovations in Communications, Electrical and Computer*

Engineering (ICICEC), Davangere, India: IEEE, Oct. 2024, pp. 1–8. doi: 10.1109/ICICEC62498.2024.10808611.

[22] K. Zhang *et al.*, "Deep Image Deblurring: A Survey," *Int. J. Comput. Vis.*, vol. 130, no. 9, pp. 2103–2130, Sep. 2022, doi: 10.1007/s11263-022-01633-5.

[23] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data," in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Montreal, BC, Canada: IEEE, Oct. 2021, pp. 1905–1914. doi: 10.1109/ICCVW54120.2021.00217.

[24] X. Wang, Y. Li, H. Zhang, and Y. Shan, "Towards Real-World Blind Face Restoration With Generative Facial Prior," presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 9168–9178. Accessed: Mar. 05, 2026. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2021/html/Wang_Towards_Real-World_Blind_Face_Restoration_With_Generative_Facial_Prior_CVPR_2021_paper.html

[25] C. Hu *et al.*, "RestormGAN: Restormer with generative facial prior towards real-world blind face restoration," *Comput. Electr. Eng.*, vol. 123, p. 110095, Apr. 2025, doi: 10.1016/j.compeleceng.2025.110095.

AUTHORS BIBLIOGRAPHY

Muhammad Adil Kustian was born in Sungai Penuh in 1997, he obtained his diploma in 2019 and his bachelor's degree in 2021. He is currently pursuing a master's degree at the Islamic University of Indonesia, Yogyakarta, majoring in informatics and concentrating on digital forensics.

Ahmad Luthfi is an assistant professor and researcher in the Department of Informatics, Universitas Islam Indonesia (UII), with over 20 years of experience in computer science. He is an experienced lecturer, inspiring tutor, and meticulous examiner with a knack for recognizing and encouraging individual growth. Luthfi earned his doctorate (2021) in Information and Communication Technology from TU Delft, his master's degree (2005) in Computer Science from Gadjah Mada University (UGM), and his bachelor's degree (1999) in Information Systems from UBD (University of Bandung).