Enhancing Image Quality in Facial Recognition Systems with GAN-Based Reconstruction Techniques

Beni Wijaya[1*], Arief Suryadi Satyawan[2], Mokh. Mirza Etnisa Haqiqi[3], Helfy Susilawati[4], Khaulyca Arva Artemysia[5], Sani Moch. Sopian[6], M. Ikbal Shamie[7], Firman[8]

[1,4,5,6,7,8]Departement of Electrical Engineering, Faculty of Engineering, Universitas Garut, Garut, West Java, Indonesia
[2]Research Center for Telecommunication, National Research and Innovation Agency (BRIN), DKI Jakarta, Indonesia
[3]Departement of Electrical Engineering, Faculty of Engineering, Universitas Indonesia, Depok, West Java, Indonesia
E-mail: [1*]beniw163@gmail.com, [2]arie021@brin.go.id, [3]mokhammad.mirza@ui.ac.id, [4]helfy.susilawati@uniga.ac.id, [5]khaulyca@gmail.com, [6]sanimochsopian1@gmail.com, [7]muhamadikbalsh1@gmail.com, [8]firmanfardiansyah2592@gmail.com

(Received: 9 Jan 2025, revised: 19 Jan 2025, accepted: 20 Jan 2025)

Abstract

Facial recognition systems are pivotal in modern applications such as security, healthcare, and public services, where accurate identification is crucial. However, environmental factors, transmission errors, or deliberate obfuscations often degrade facial image quality, leading to misidentification and service disruptions. This study employs Generative Adversarial Networks (GANs) to address these challenges by reconstructing corrupted or occluded facial images with high fidelity. The proposed methodology integrates advanced GAN architectures, multi-scale feature extraction, and contextual loss functions to enhance reconstruction quality. Six experimental modifications to the GAN model were implemented, incorporating additional residual blocks, enhanced loss functions combining adversarial, perceptual, and reconstruction losses, and skip connections for improved spatial consistency. Extensive testing was conducted using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) to quantify reconstruction quality, alongside face detection validation using SFace. The final model achieved an average PSNR of 26.93 and an average SSIM of 0.90, with confidence levels exceeding 0.55 in face detection tests, demonstrating its ability to preserve identity and structural integrity under challenging conditions, including occlusion and noise. The results highlight that advanced GAN-based methods effectively restore degraded facial images, ensuring accurate face detection and robust identity preservation. This research provides a significant contribution to facial image processing, offering practical solutions for applications requiring high-quality image reconstruction and reliable facial recognition.

Keywords: Facial Recognition Systems, Image Reconstruction, Generative Adversarial Networks (GANs), PSNR, SSIM.

I. INTRODUCTION

The human face is a crucial aspect of personal identity, communication, and social interaction. It conveys emotions, intentions, and identity-related information that are essential for both personal and professional engagements [1]. In modern society, facial recognition technology plays a pivotal role in various domains, including security systems, healthcare diagnostics, forensics, digital communication, and access to public facilities [1][2][3]. For example, facial recognition is widely used in airports for identity verification, in mobile devices for unlocking and authorization, and in smart city infrastructures for enhancing security and personalized services. Given this reliance on facial imagery, maintaining the quality, integrity, and completeness of facial images is vital for accurate identification and smooth access to these essential services [4].

However, facial images are frequently corrupted, degraded, or intentionally obfuscated due to various factors [5]. These can include environmental conditions (e.g., poor lighting, motion blur, or noise), data compression during transmission, and deliberate occlusions to protect privacy or conceal identity (e.g., masks, objects, or partial censorship) [6]. Such distortions degrade the quality of facial images and impede accurate identification, reducing the effectiveness of systems that rely on facial data [7]. In particular, failures in facial recognition can lead to denied access to critical services

DOI: 10.34148/teknika.v14i1.1180

such as healthcare facilities, transportation hubs, financial transactions, and secure building entry systems, posing both practical and security-related challenges [8][9]. For individuals who depend on these systems for daily activities, any disruption caused by image corruption can significantly hinder their ability to access services, thereby emphasizing the urgency of developing robust reconstruction solutions.

Existing approaches to facial image reconstruction often face significant limitations. Traditional image processing techniques struggle to restore facial details with high accuracy, especially when large portions of the image are corrupted or occluded [10]. Even advanced deep learning models, while effective in certain scenarios, frequently produce reconstructions with noticeable artifacts, loss of fine-grained features, or inconsistencies in global structure. For example, many methods fail to account for the semantic and contextual relationships between facial features, leading to outputs that are unrealistic or unfit for reliable facial recognition [11]. Furthermore, current GAN-based methods often prioritize visual realism over identity preservation, which is critical for applications involving security and identity verification. These limitations highlight the need for more robust and specialized reconstruction techniques that can address both visual quality and identity consistency in challenging scenarios.

Generative Adversarial Networks (GANs) have emerged as a revolutionary tool in the realm of image processing, particularly for tasks involving image inpainting and reconstruction [12][13]. By leveraging the adversarial framework of a generator and discriminator, GANs can produce highly realistic and coherent images that fill in missing details or correct distortions [14]. In the context of facial image reconstruction, GANs are particularly promising due to their ability to model intricate facial structures, textures, and identity-specific features [15][16]. This capability allows GANs to restore images in ways that preserve the subject's unique identity while maintaining global coherence, even when substantial portions of the image are corrupted or missing.

Recent advancements in GAN architectures have significantly enhanced their performance [17][18]. Techniques such as attention mechanisms allow the model to focus on critical regions of the face, ensuring that important features like eyes, nose, and mouth are accurately reconstructed [19]. Multi-scale feature extraction helps in capturing both fine-grained details and broader contextual patterns, improving the overall quality and realism of the reconstructed image [20]. Additionally, the use of contextual losses ensures that the output maintains semantic consistency with the original image, reducing artifacts and producing smoother results.

The urgency for developing advanced GAN-based techniques for facial image reconstruction is further heightened by the increasing integration of facial recognition in public infrastructure [21][22]. Airports, government buildings, healthcare institutions, and financial services are adopting facial recognition technologies to streamline access control, improve security, and enhance user experience. In such settings, a corrupted or incomplete facial image can lead to misidentification, delays, or complete denial of access. For example, in healthcare, failure to verify a patient's identity can result in incorrect treatment or delayed care. In airports, an inability to match a traveler's face to their identity records can disrupt travel plans and compromise security protocols. Therefore, improving the robustness and accuracy of facial image reconstruction systems is crucial to ensuring reliable and seamless access to these services.

Moreover, face recognition technology supports adaptive control systems in autonomous electric vehicles by enabling real-time detection of driver or passenger emotions and behaviors [23]. This capability allows the vehicle to respond proactively to specific situations, such as adjusting environmental settings for comfort or issuing warnings in case of unsafe actions. Furthermore, the integration of face recognition enhances fleet management systems, enabling precise access control and usage tracking for shared or commercial autonomous vehicles

This paper focuses on leveraging and improving GAN-based techniques to address these challenges by reconstructing partially corrupted or occluded facial images. The primary objective is to achieve visually realistic reconstructions that not only enhance the quality of the images but also preserve identity-specific features necessary for accurate recognition. By addressing issues such as artifact removal, noise reduction, and global coherence, this work aims to contribute to the broader field of computer vision, particularly in applications that depend on facial recognition for identity verification and secure access.

II. RESEARCH METHODOLOGY

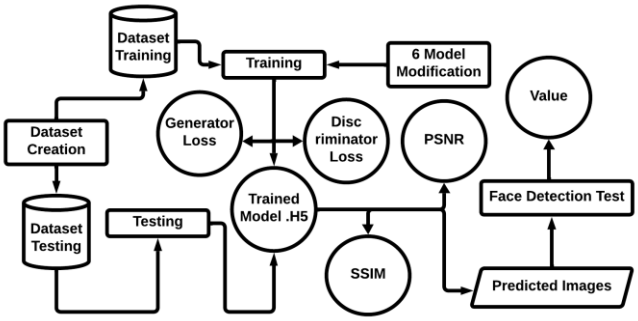

Figure 1. Process Architecture

The proposed method as shown in Figure 1 follows a systematic process for facial image reconstruction using Generative Adversarial Networks (GANs). The workflow begins with dataset creation, which is split into training and testing sets. During the training phase, the GAN model is optimized by minimizing the Generator Loss and Discriminator Loss. After training, the model is saved in .h5 format.

The trained model is then evaluated on the test dataset to produce predicted images. The quality of the reconstructions is assessed using metrics such as PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index Measure).

Additionally, six model modifications are applied to enhance performance throughout the process.

The reconstructed facial images are further evaluated to determine whether the faces are accurately detected and compared to the ground truth. This evaluation is conducted using a pre-trained face recognition model, such as SFace [24], which generates facial embeddings for both the reconstructed and original images.

A. Dataset Preparation

In this study, we utilized the CelebA dataset for training and evaluating the GAN model as shown in Figure 2. The CelebA dataset contains over 200,000 facial images with diverse attributes, making it suitable for tasks involving facial image reconstruction [21].

Figure 2. Paired Dataset

A total of 1,000 images were selected, comprising 500 occluded images paired with 500 ground-truth images for model training. Additionally, a subset of 200 images was reserved for testing. The test set consists of images previously encountered by the model during training to evaluate its reconstruction performance on familiar data. This approach ensures consistency in assessing the model's ability to restore corrupted facial images while maintaining identity-specific features.

B. Model Development Modification Experiment

Generative Adversarial Networks (GANs) are a type of deep learning framework that specializes in generating realistic data, particularly images, by leveraging a competitive training approach [25]. The model comprises two neural networks: the generator, which synthesizes images to mimic real ones, and the discriminator, which evaluates and identifies whether the input is real or generated (Figure 3). Through this adversarial interaction, the generator progressively improves its ability to create authentic-looking outputs. GANs have demonstrated remarkable success in tasks such as image inpainting and restoration due to their capacity to model intricate data distributions and produce coherent, high-resolution results.

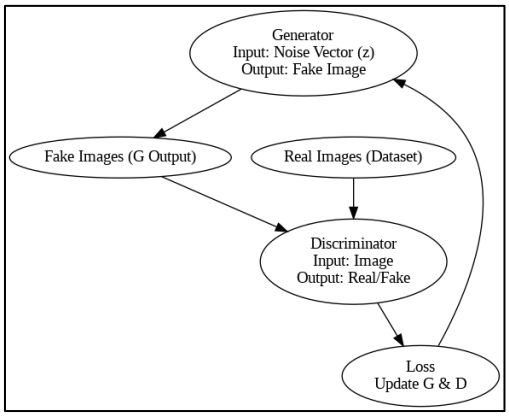

Figure 3. GAN Architecture

To enhance the performance and robustness of the GAN architecture for facial image reconstruction, a series of modifications were implemented across six experiments. These changes aimed to improve image quality, coherence, and identity preservation by progressively refining the generator and discriminator architectures, as well as their loss functions.

In the initial experiment, the generator utilized seven downsampling and seven upsampling layers, complemented by five residual blocks with 64 filters, while the discriminator was enhanced with five residual blocks of 512 filters to improve its differentiation capability. The second experiment increased the generator's depth to eight downsampling and eight upsampling layers, maintaining five residual blocks, while introducing adversarial loss to the discriminator for more effective feedback during training. The third experiment adjusted the generator by reducing upsampling layers to seven, keeping its residual blocks unchanged, with the discriminator architecture and adversarial loss remaining consistent.

The fourth experiment introduced a combination of adversarial loss, VGG16-based perceptual loss, and reconstruction loss in the generator, aiming to balance realism, feature-level consistency, and pixel-level fidelity. In the fifth experiment, the VGG16 perceptual loss was replaced with VGG19 perceptual loss to leverage deeper feature extraction for improved perceptual accuracy.

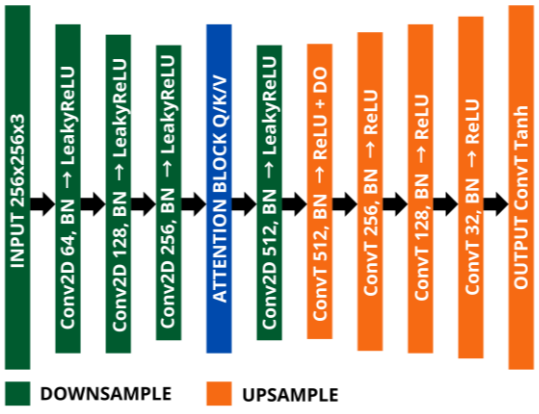

Figure 4. Generator Architecture for 6th Experiment

DOI: 10.34148/teknika.v14i1.1180

Finally, the sixth experiment, as illustrated in Figure 4, evaluates the generator's performance using a U-Net-based architecture that incorporates an integrated attention mechanism. The U-Net framework consists of an encoder-decoder structure designed to capture and reconstruct spatial features effectively. The encoder compresses the input image into a latent representation by progressively reducing spatial dimensions, while the decoder reconstructs the image by upscaling the latent representation back to the original resolution. Skip connections are added between corresponding layers of the encoder and decoder to retain spatial details and prevent the loss of crucial low-level information during feature extraction. The attention block improves the model's ability to perform accurate mappings between masked and original images.

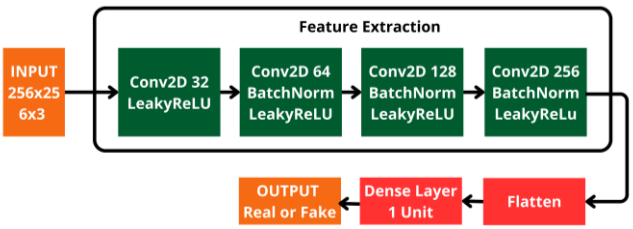

Figure 5. Discriminator Architecture for 6th Experiment

Also in the sixth experiment, the generator's loss function incorporated adversarial loss, perceptual loss based on EfficientNetV2B0, and reconstruction loss, while the discriminator retained its five residual blocks with adversarial loss, as depicted in Figure 5. The discriminator architecture, consisting of sequential convolutional layers followed by BatchNormalization and LeakyReLU activations, effectively distinguishes between real and generated images, ensuring the generator produces high-quality outputs. This design emphasizes the balance between generator and discriminator training for optimal adversarial performance.

These modifications were essential to address limitations in existing GAN frameworks by enhancing the preservation of facial identity, improving the coherence of reconstructed features. By systematically refining both architecture and loss functions, the proposed framework achieves a robust balance between realism and identity fidelity, critical for accurate facial image reconstruction.

C. Training Process

The training process utilized an AMD Ryzen 5 5500U CPU with integrated AMD Radeon Graphics, serving as a trial to evaluate the feasibility of training deep learning models on CPU-based systems. Despite the inherent limitations in processing speed compared to GPU-based systems, the AMD Ryzen 5 5500U provided sufficient computational power for handling image reconstruction tasks, showcasing the versatility and accessibility of deep learning implementations on more economical hardware. This trial underscores the broader applicability of the proposed method, particularly in resource-constrained settings, making it a significant contribution to the field. Each experiment was conducted over 50 epochs, ensuring sufficient iterations for the model to learn and refine its performance while maintaining consistency across all trials to facilitate a fair and comprehensive comparison of results.

The training strategy in the code also involves alternating optimization of the generator and discriminator to achieve balanced adversarial training. The generator uses a multi-component loss function—adversarial, perceptual loss, contextual, and reconstruction losses—to produce realistic and semantically consistent images. The discriminator is trained with adversarial loss to effectively distinguish real and generated images. Both networks are optimized with Adam (learning rate 2e-4, $\beta_1 = 0.5$), and data augmentation techniques, such as random flipping and brightness adjustments, are applied to enhance generalization and robustness. This approach ensures stable training and high-quality outputs.

D. Evaluation Parameter

The performance of the Generative Adversarial Network (GAN) model is evaluated using several critical metrics. Among these are the Generator Loss, which determines the generator's capability to create convincing outputs that successfully fool the discriminator, the Discriminator Loss, which measures how well the discriminator can differentiate between real and generated data. Additionally, the Structural Similarity Index Measure (SSIM) is employed to evaluate perceptual similarity by analyzing aspects like luminance, contrast, and structural features between the generated images and ground truth. Finally, the Peak Signal-to-Noise Ratio (PSNR) is used to assess the quality of reconstructed images by quantifying their pixel-wise similarity to the ground truth.

1. Generator Loss

The generator loss measures the effectiveness of the generator in producing realistic data that can deceive the discriminator [26]. It is typically derived from the adversarial loss, which quantifies how well the generator's output aligns with the target distribution. For GANs, the generator loss can be expressed as shown in Equation 1.

$$L_G = - E_{z \sim pz\,(z)} [\log (D(G(z)))] \quad (1)$$

Where:

- L_G : Generator loss, measuring how well the generator fools the discriminator.
- $z \sim pz(z)$: Noise vector sampled from a prior distribution.
- $G(z)$: Generated image from the generator.
- $D(G(z))$: Discriminator's probability that $G(z)$ is real.

2. Discriminator Loss

The discriminator loss evaluates the discriminator's ability to distinguish between real and generated data. It consists of two terms: the loss for real data and the loss for generated data[27]. Discriminator loss is given by Equation 2.

$$L_D = -E_{x \sim Pdata}(x)[\log (D(x))] - E_{z \sim pz(z)} [\log(1-D(G(z)))] \quad (2)$$

DOI: 10.34148/teknika.v14i1.1180

Where:
- L_D : Discriminator loss, measuring the performance of the discriminator.
- $x \sim P_{data}(x)$: Real data samples from the data distribution.
- $z\ p_z(z)$: Noise vector sampled from a prior distribution.
- $D(x)$: Discriminator's probability that x is real.
- $D(G(z))$: Discriminator's probability that the generated image G(z) is real.

3. Structural Similarity Index Measure (SSIM)

SSIM is a perceptual metric that measures image similarity by considering luminance, contrast, and structure . It is widely used to evaluate the quality of reconstructed or generated images [28]. The SSIM value ranges from -1 to 1 (perfect similarity). SSIM is calculated using the formula as in Equation 3.

$$SSIM\ (x, y) = \frac{(2\mu x\mu y + C1)(2\sigma xy + C2)}{(\mu_x^2 + \mu_y^2 + C1)\,(\sigma_x^2 + \sigma_y^2 + C2)} \quad (3)$$

Where:
- μx and μy are the means of the images x and y.
- $\sigma x2$ and $\sigma y2$ are the variances of the images x and y.
- σxy is the covariance between the images x and y .
- $C1$ and $C2$ are constants used to maintain the stability of the calculation when the denominator approaches zero.

4. Peak Signal-to-Noise Ratio (PSNR)

PSNR measures the quality of the generated or reconstructed image compared to the original ground truth image. It is expressed in decibels (dB) and is calculated based on the Mean Squared Error (MSE) between two images [29]. Higher PSNR values indicate better image quality. The formula is given by Equation 4.

$$PSNR = 10.\mathrm{Log}_{10}\left(\frac{MAX^2}{MSE}\right) \quad (4)$$

Where:
- MAX is the maximum pixel intensity value
- MSE is the Mean Squared Error between the original image and the generated image., and MSE is given by Equation 5.

$$\mathrm{MSE} = \frac{1}{m\,.n} \sum_{i=1}^{m} \sum_{j=1}^{n} [x\ (i,j) - y(i,j)]^2 \quad (5)$$

Where:
- $x(I,j)$ and $y(i,j)$ represent the pixel values of the original and generated images at position (i,j) respectively.
- m and n denote the dimensions of the image (height and width), with mn being the total number of pixels

5. Cosine Similarity (for Face Recognition Confidence)

Cosine similarity measures the cosine of the angle between two vectors (embeddings) [24]. Higher values indicate greater similarity is given by Equation 6.

$$Confidance\ Level = \cos(\theta) = \frac{A.B}{|A||B|} \quad (6)$$

Where:
- A and B Feature embeddings of the reconstructed and ground truth images, respectively.
- |A| Magnitude (norm) of vector A
- Values range from −1 to 1

6. Confidence from Detection Model (for Face Detection)

In face detection, confidence often represents the model's probability output for detecting a face [24]. This is derived from a softmax activation or similar probability estimation, given by Equation 7.

$$Confidence\ Level = P(Face) \quad (7)$$

Where P(Face) the probability assigned by the model to the presence of a face.

III. RESULT AND DISCUSSION

A. Training Evaluation

The evaluation of a Generative Adversarial Network (GAN) model typically involves analyzing the loss functions of both the generator and the discriminator. These losses are crucial for understanding the model's performance during training and its ability to generate high-quality, realistic images.

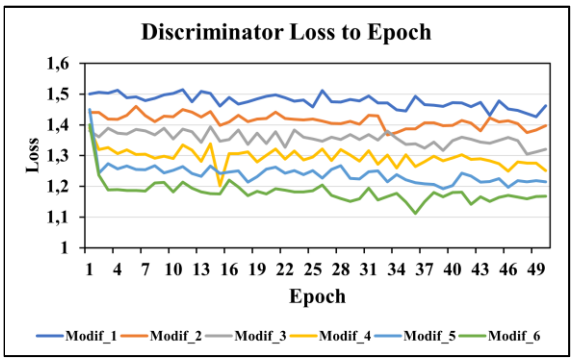

Figure 6. Discriminator Loss to Epoch for Six Modification

The Figure 6 shows the discriminator loss across 50 epochs for six different model modifications (Modif_1 to Modif_6). The y-axis represents the discriminator loss, while the x-axis indicates the epochs. Initially, all models experience a significant reduction in loss, followed by stabilization and minor fluctuations. Among them, Modif_6 exhibits the lowest loss, suggesting better discriminator performance and potentially more robust training stability. In contrast, Modif_1 maintains the highest loss values, which could indicate challenges in learning or an imbalance between the generator and discriminator. Modif_3, Modif_4, and Modif_5 show moderate performance with relatively consistent trends, while Modif_2 exhibits intermediate results with slightly higher fluctuations. This analysis suggests that

DOI: 10.34148/teknika.v14i1.1180

architectural or parameter differences in Modif_6 likely enhanced its capability to differentiate between real and generated samples effectively, leading to a more stable and lower loss trajectory.

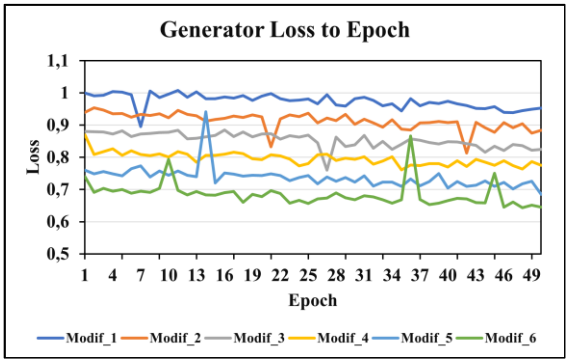

Figure 7. Generator Loss to Epoch for Six Model Modification

The Figure 7 represents the generator loss across 50 epochs for six model modifications (Modif_1 to Modif_6). Modif_1 consistently exhibits the highest loss values, indicating relatively weaker generator performance in producing realistic outputs. Conversely, Modif_6 achieves the lowest loss values, reflecting better training dynamics and stronger generator performance. Modif_3, Modif_4, and Modif_5 have moderate and comparable losses, with occasional spikes in loss, particularly noticeable around epochs 10, 30, and 40. These spikes may suggest moments of instability in adversarial training. Modif_2 demonstrates intermediate loss values but with fewer pronounced fluctuations. The results imply that Modif_6 outperforms the other models, likely due to enhancements in architecture or training strategies, while Modif_1 struggles to match the adversarial learning process effectively.

The training loss patterns of both the discriminator and generator, shown in Figures 3 and 4, offer insight into the training stability and convergence of each model.Modif_6 and Modif_5 exhibit the most stable loss curves, particularly in their generator loss. Both models show consistent and gradual reductions in loss over time, indicating effective learning and better convergence.

B. Qualitative Visual Test Result

During the evaluation phase, the trained model was tested using previously unseen images to assess its performance. Below is the visualization of the results from the six modifications. Additionally, some sample images are provided to illustrate the reconstructed outputs compared to the ground truth, highlighting the performance differences across the modifications

The Figure 8 showcases results from modifications 1 and 2 of the Generative Adversarial Network (GAN) model applied to reconstruct occluded face images. The first column represents the occluded input images, while the second and third columns display the reconstructed outputs from modifications 1 and 2, respectively The fourth column illustrates the ground truth images.

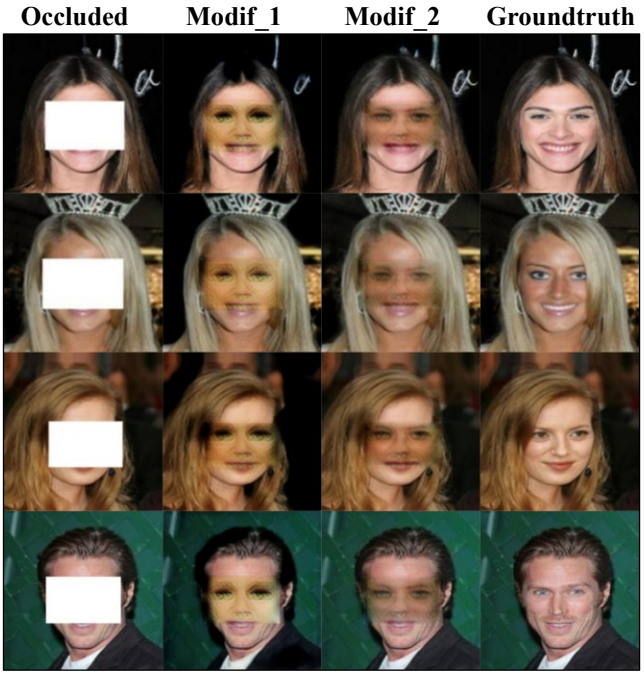

Figure 8. Result for Modification 1 and 2

Compared to modification 1, modification 2 demonstrates improved reconstruction quality, with smoother textures and better alignment of facial features, particularly in regions affected by occlusion. However, subtle artifacts are still visible, indicating potential areas for further enhancement.

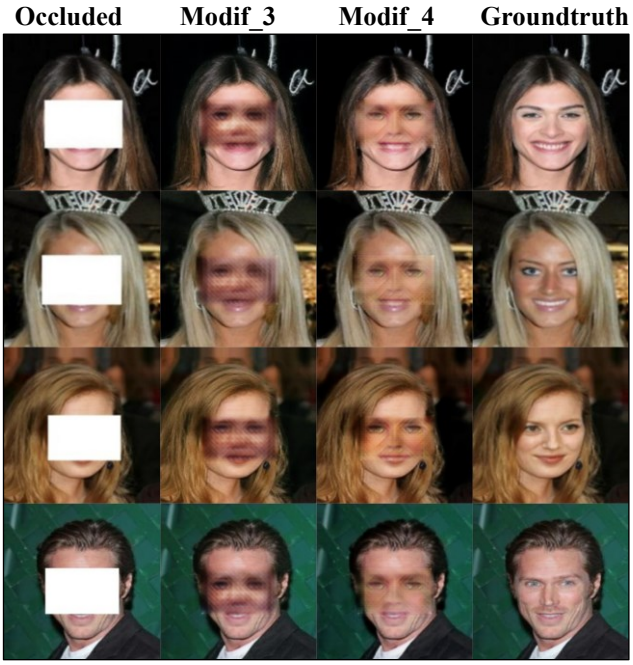

Figure 9. Result for Modification 3 and 4

DOI: 10.34148/teknika.v14i1.1180

In Figure 9. Compares results from modifications 3 and 4 for occluded face reconstruction using GANs. The second and third columns display outputs from modification 3 and modification 4, respectively. Modification 4 demonstrates improved feature alignment and reduced artifacts compared to modification 3, resulting in reconstructions that are closer to the ground truth shown in the fourth column

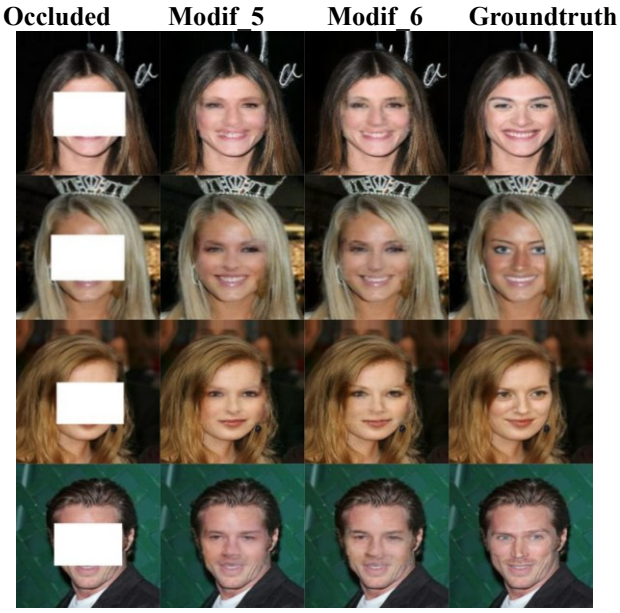

Figure 10. Result for Modification 5 and 6

The visual output of each GAN model, presented in Figure 8, 9 and 10 offers a direct comparison of how well each model reconstructs occluded images. Modif_6: This modification produces the clearest and most accurate reconstructions, closely resembling the ground truth images. Fine details and textures are well-preserved, making it the best performing modification in terms of visual quality. The ability of Modif_6 to reconstruct the occluded areas with minimal blurring or artifacts demonstrates its robustness in handling complex features.

Modif_5: Similar to Modif_6, this modification also performs very well, though with slightly more visible artifacts in some cases. The reconstructions are sharp and visually convincing, particularly in areas with strong textures and high-frequency details. Modif_4: This model offers decent reconstructions, but there is a noticeable drop in visual quality compared to Modif_5 and Modif_6. Some of the occluded areas appear blurry or lose finer details, suggesting that Modif_4 has less capacity to fully restore the masked regions with high fidelity.

Modif_3: The results here are moderately good but still fall behind Modif_4. The images generated by Modif_3 exhibit some blurring and a slight loss of structure in certain regions. While the model performs well in restoring general shapes, it struggles with fine textures and edges. Modif_2: The quality of image reconstruction by Modif_2 is slightly better than Modif_1but remains below the other modifications. The model tends to produce less sharp images, with visible noise in the occluded areas. This reflects a moderate ability to reconstruct the images but with noticeable quality loss in detail.

Modif_1: This modification generates the poorest quality reconstructions. The restored areas are visibly distorted or unclear, and the generated images lack sharpness, making them the least accurate among all modifications. The GAN struggles to recover important visual information, suggesting suboptimal model configuration or training stability issues.

C. Quantitative Test Result

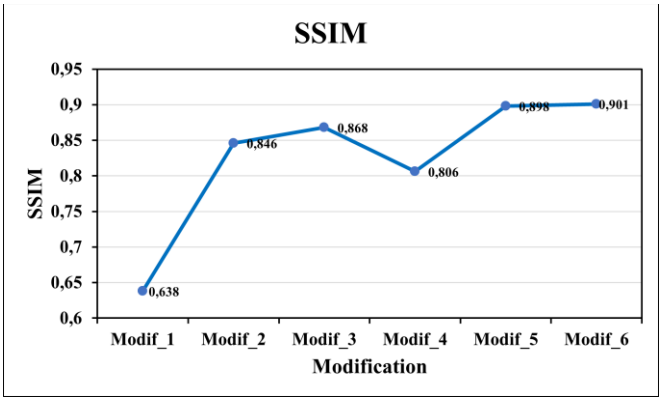

Figure 11. Average SSIM Graph for Each Modification

Modif_6 achieves the highest average SSIM of 0.901, indicating excellent preservation of image structure and high similarity to the ground truth, as shown in Figure 11. Modif_5 closely follows with an average SSIM of 0.898, reaffirming its strong structural reconstruction capabilities. Modif_4 also performs well, achieving an average SSIM of 0.806, though it shows a slight decrease in its ability to maintain structural integrity, particularly in more complex regions.

Modif_3 and Modif_2 achieve average SSIM scores of 0.86 and 0.84, respectively, demonstrating a moderate ability to preserve image structure but with noticeable degradation in more intricate areas. Modif_1 records the lowest average SSIM of 0.63, reflecting poor structural preservation, as the model struggles to reconstruct occluded areas with sufficient detail.

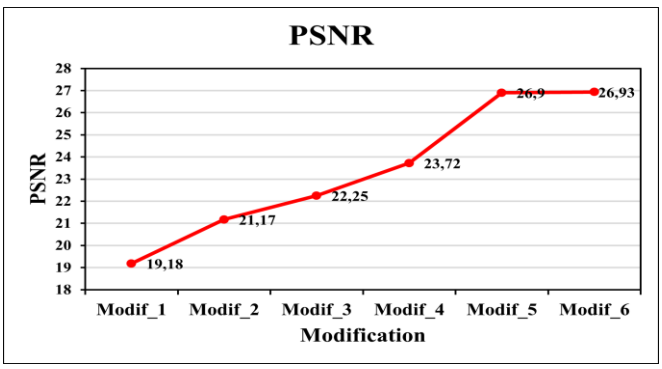

Figure 12. Average PSNR for Each Modification

Modif_6 achieves the highest average PSNR value (26.93), meaning it has the best noise reduction and highest fidelity in reconstructing the original image. Modif_5 is close

behind with an average PSNR of 26.9, which confirms its excellent performance in generating high-quality images with minimal noise shown in Figure 12.

Modif_4 scores 23.72, still performing well but with more noticeable noise and lower reconstruction accuracy compared to the top two. Modif_3 and Modif_2 have lower averages PSNR values of 22.25 and 21.17, respectively, reflecting their moderate ability to reduce noise but also their limitations in fully capturing image details.

Modif_1 exhibits the lowest average PSNR (19.18), highlighting its inability to accurately reconstruct the original image with high fidelity, as substantial noise and distortions remain in the output.

Compared to previous studies on facial image reconstruction, which utilized a VAE-based generator and curriculum learning with local and global discriminators, the proposed method demonstrates clear advancements. Earlier methods achieved an average SSIM of 0.651 and a average PSNR of 21.067, reflecting moderate structural preservation and noise reduction [30]. In contrast, this study employs a U-Net-based generator with an attention mechanism and a combined loss function incorporating perceptual loss (EfficientNetV2), achieving significantly higher average SSIM (0.901) and average PSNR (26.93). These results highlight the improved structural accuracy and noise reduction capabilities of the proposed approach.

D. Face Detection Test with Sface

At this stage, the reconstructed facial images are tested using a pre-trained face recognition model, such as SFace [24]. This testing evaluates whether the reconstructed faces can be accurately detected and compared to the ground truth, ensuring the preservation of essential facial features. By computing embedding similarities between the reconstructed and original faces, this method provides a quantitative measure of how well the reconstruction retains identity and structural integrity.

To illustrate the results more effectively, several representative samples of the reconstructed images are presented, highlighting the model's performance under various conditions. This evaluation serves as a foundational step for assessing the effectiveness of the proposed reconstruction method and its applicability in real-world tasks, such as facial recognition or identity verification

In Figure 13, the reconstruction test using SFace demonstrates the progression of facial reconstruction quality across different modifications. Confidence levels, manually assigned for this visualization, range from poor similarity in Modif_1 (Conf: -0.14) to significantly improved similarity in Modif_6 (Conf: 0.65). Bounding boxes indicate detected facial features, validating the presence of recognizable structures in the reconstructed images. The ground truth serves as a reference for evaluating the overall reconstruction performance.

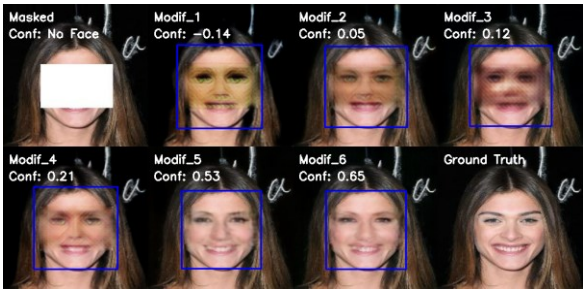

Figure 13. Reconstruction Test With SFace, Showing Bounding Boxes and Confidence Levels for Each Modification.

In addition to the primary evaluation, testing was also conducted under different challenges, such as occlusion and noise. The inclusion of these challenges allows for a more comprehensive evaluation of the model's ability to maintain facial identity and structural coherence under adverse conditions, providing valuable insights into its practical applicability and limitations.

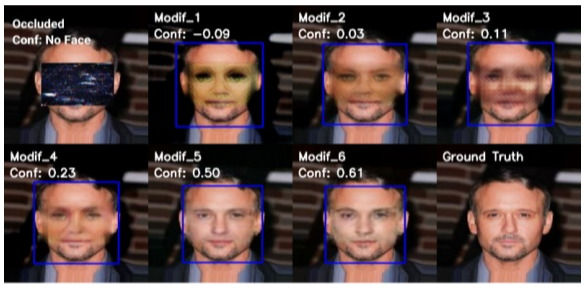

Figure 14. Sface Test Under Occlusion Condition

Reconstruction test results under challenging conditions, including occlusion shown in Figure 14. The figure illustrates the gradual improvement in facial reconstruction quality across different modifications (Modif_1 to Modif_6). Confidence levels, displayed alongside bounding boxes, indicate the similarity between the reconstructed images and the ground truth. Starting from Modif_1 (Conf: -0.09), the confidence levels progressively increase, reaching Modif_6 (Conf: 0.61), which demonstrates the best reconstruction performance. The ground truth image serves as a reference, while the "Occluded" image shows the initial heavily masked input with no detectable face.

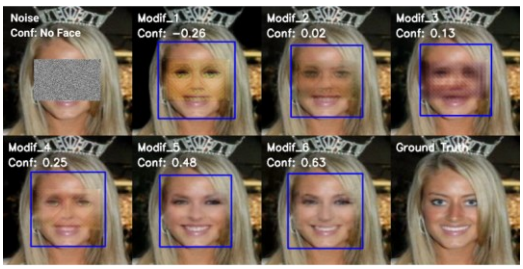

Figure 15. Sface Test Under Noise Condition

Reconstruction test results under noise distortion, as shown in Figure 15. The figure demonstrates the model's

DOI: 10.34148/teknika.v14i1.1180

ability to reconstruct facial features when the input image is corrupted by noise. Confidence levels show a gradual improvement from Modif_1 (Conf: -0.26) to Modif_6 (Conf: 0.63), reflecting an increase in similarity to the ground truth. The "Noise" input, where no face is detected, serves as the starting point for reconstruction. Modif_6 exhibits the best performance, closely approximating the ground truth, while earlier modifications (Modif_1 to Modif_3) struggle to capture essential facial details. The ground truth image provides a benchmark for evaluating reconstruction accuracy.

The testing results indicate that the proposed reconstruction method performs well overall. Notably, Modif_6 consistently achieves the highest performance across various conditions, with an average confidence level exceeding 0.55. This demonstrates the model's robustness and effectiveness in preserving facial identity and structural details under both standard and challenging scenarios.

IV. CONCLUSION

The performance and robustness of the GAN modifications for facial image reconstruction demonstrate a clear progression in quality and effectiveness across the six experiments. The results underline the impact of architectural and loss function refinements on the visual quality, structural preservation, and quantitative fidelity of the reconstructed images.

Modif_6 emerges as the most effective modification, achieving the highest average PSNR (26.93) and average SSIM (0.90), and producing visually accurate reconstructions with minimal artifacts. The incorporation of skip connections and the EfficientNetV2B0-based perceptual loss proved instrumental in enhancing spatial consistency and feature extraction. Modif_5, with its VGG19 perceptual loss, performed nearly as well, with slightly lower fidelity but retaining excellent structural preservation and feature detail.

While Modif_4 showed competitive results, the drop in average PSNR (23.72) and average SSIM (0.80) indicates that the addition of perceptual loss from VGG16, while beneficial, did not match the depth and feature extraction capability of VGG19 or EfficientNetV2B0. The earlier modifications, Modif_3, Modif_2, and Modif_1, exhibited a gradual decline in reconstruction quality, with Modif_1 scoring the lowest across all metrics. The limitations in architectural design and loss functions in these earlier models resulted in less effective learning, leading to higher noise, blurring, and structural inconsistencies in reconstructed images.

The testing conducted with SFace further validates the model's performance. Modif_6 consistently achieved confidence levels exceeding 0.55 across different conditions, such as occlusion and noise, demonstrating its ability to retain facial identity and structural integrity in challenging scenarios. These results emphasize the applicability of the proposed method for tasks requiring robust face detection and recognition.

The training loss trends provide further insights into the observed performance differences. Modif_6 and Modif_5 demonstrated stable and smooth convergence in their loss curves, correlating with their superior reconstruction outcomes. In contrast, the earlier modifications, particularly Modif_1, suffered from significant instability during training, which likely hindered the models' ability to accurately capture the data distribution and resulted in suboptimal reconstructions.

Overall, the progressive enhancements in GAN architecture and training methodologies underscore the importance of balanced architectural depth, effective loss function design, and stable training dynamics in achieving robust facial image reconstruction. These findings highlight the potential of advanced GAN models, such as Modif_6, for real-world applications in facial image restoration and beyond.

While the results demonstrate significant progress, several avenues for further exploration remain. These include applying the method to video frame reconstruction for dynamic scenarios, integrating the approach with multi-modal systems (e.g., combining image and audio data), and extending the method to handle extreme distortions or larger occlusions. Addressing these areas can further enhance the robustness and versatility of the proposed reconstruction framework.

ACKNOWLEDGMENT

This research was supported by funding from the Indonesia Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan) through the National Research and Innovation Agency's Research and Innovation for Advanced Indonesia (RIIM) program (BRIN). The collaboration between the Faculty of Engineering, Universitas Garut, and BRIN has been invaluable in facilitating this research, and we would like to express our sincere gratitude for their continued support and cooperation.

REFERENCES

[1] M. Sajjad *et al.*, "A comprehensive survey on deep facial expression recognition: challenges, applications, and future guidelines," *Alexandria Eng. J.*, vol. 68, pp. 817–840, 2023, doi: https://doi.org/10.1016/j.aej.2023.01.017.

[2] N. Jain, J. Hawari, P. Jha, P. C. Nair, and N. Sampath, "Interpretable Deep Learning for Facial Feature Detection: A Comprehensive Study on Face and Eyes Recognition with LIME Explanations," in *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, 2024, pp. 1–7. doi: 10.1109/I2CT61223.2024.10544194.

[3] L. Wang, "Protection of personal information in the application environment of face recognition technology," *Front. Soc. Sci. Technol.*, vol. 5, no. 15, pp. 63–72, 2023, doi: 10.25236/fsst.2023.051512.

[4] A. N and K. Anusudha, "Real time face recognition system based on YOLO and InsightFace," *Multimed. Tools Appl.*, vol. 83, no. 11, pp. 31893–31910, 2024, doi: 10.1007/s11042-023-16831-7.

[5] R. Munir, "Pengolahan Citra Digital Dengan Pendekatan

Algoritmik," 2011. [Online]. Available: https://api.semanticscholar.org/CorpusID:196067817

[6] H. Ali, M. Hariharan, A. H. Adom, S. K. Zaaba, M. Elshaikh, and S. Yaacob, "Facial emotion recognition under noisy environment using empirical mode decomposition," in *2016 3rd International Conference on Electronic Design (ICED)*, 2016, pp. 476–480. doi: 10.1109/ICED.2016.7804691.

[7] C. Busch, "Challenges for automated face recognition systems," *Nat. Rev. Electr. Eng.*, vol. 1, no. 11, pp. 748–757, 2024, doi: 10.1038/s44287-024-00094-x.

[8] M. De Marsico, M. Nappi, D. Riccio, and H. Wechsler, "Robust Face Recognition for Uncontrolled Pose and Illumination Changes," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 43, no. 1, pp. 149–163, 2013, doi: 10.1109/TSMCA.2012.2192427.

[9] R. H. Al-Abboodi and A. A. Al-Ani, "Deep Learning Approach for Face Recognition Applied in IoT Environment- Comprehensive Review," in *2024 Seventh International Women in Data Science Conference at Prince Sultan University (WiDS PSU)*, 2024, pp. 191–197. doi: 10.1109/WiDS-PSU61003.2024.00047.

[10] M. Motmaen, M. Mohrekesh, M. Akbari, N. Karimi, and S. Samavi, "Image Inpainting by Hyperbolic Selection of Pixels for Two-Dimensional Bicubic Interpolations," *26th Iran. Conf. Electr. Eng. ICEE 2018*, pp. 665–669, 2018, doi: 10.1109/ICEE.2018.8472408.

[11] D. K. Basha and T. Venkateswarlu, "Image Restoration by Linear Regression for Gaussian Noise Removal from Natural Images," *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 11S2, pp. 126–130, 2019, doi: 10.35940/ijitee.k1020.09811s219.

[12] P. Chai, L. Hou, G. Zhang, Q. Tushar, and Y. Zou, "Generative adversarial networks in construction applications," *Autom. Constr.*, vol. 159, p. 105265, 2024, doi: https://doi.org/10.1016/j.autcon.2024.105265.

[13] S. Islam *et al.*, "Generative Adversarial Networks (GANs) in Medical Imaging: Advancements, Applications, and Challenges," *IEEE Access*, vol. 12, pp. 35728–35753, 2024, doi: 10.1109/ACCESS.2024.3370848.

[14] Y. Lv, J. Duan, and X. Li, "A survey on modeling for behaviors of complex intelligent systems based on generative adversarial networks," *Comput. Sci. Rev.*, vol. 52, p. 100635, 2024, doi: https://doi.org/10.1016/j.cosrev.2024.100635.

[15] A. Zargaran, S. Sousi, S. P. Glynou, H. Mortada, D. Zargaran, and A. Mosahebi, "A systematic review of generative adversarial networks (GANs) in plastic surgery," *J. Plast. Reconstr. Aesthetic Surg.*, vol. 95, pp. 377–385, 2024, doi: https://doi.org/10.1016/j.bjps.2024.04.007.

[16] *ISBM College of Engineering , Pune Multidisciplinary Emerging Trends in Engineering and Technology*. doi: 10.17492/jpi.ISBM.082401.

[17] M. Velayutham, V. Chin, and W. Shen, "Restoration of Historical Illustrations using Generative Adversarial Networks," vol. 8, no. 3, 2024.

[18] R. Elanwar and M. Betke, "Generative adversarial networks for handwriting image generation: a review," *Vis. Comput.*, 2024, doi: 10.1007/s00371-024-03534-9.

[19] B. Kc, S. Sapkota, and A. Adhikari, "Generative Adversarial Networks in Anomaly Detection and Malware Detection: A Comprehensive Survey," *Adv. Artif. Intell. Res.*, vol. 4, no. 1, pp. 18–35, 2024, doi: 10.54569/aair.1442665.

[20] Y. Wu *et al.*, "Mineral prospecting mapping with conditional generative adversarial network augmented data," *Ore Geol. Rev.*, vol. 163, p. 105787, 2023, doi: https://doi.org/10.1016/j.oregeorev.2023.105787.

[21] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2015 Inter, pp. 3730–3738, 2015, doi: 10.1109/ICCV.2015.425.

[22] X. Gao, W. He, and Y. Hu, "Modeling of meandering river deltas based on the conditional generative adversarial network," *J. Pet. Sci. Eng.*, vol. 193, p. 107352, 2020, doi: https://doi.org/10.1016/j.petrol.2020.107352.

[23] A. S. Satyawan, S. Fuady, A. Mitayani, and Y. W. Sari, "HOG Based Pedestrian Detection System for Autonomous Vehicle Operated in Limited Area," in *2021 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET)*, 2021, pp. 147–152. doi: 10.1109/ICRAMET53537.2021.9650473.

[24] F. Boutros, M. Huber, P. Siebke, T. Rieber, and N. Damer, "SFace: Privacy-friendly and Accurate Face Recognition using Synthetic Data," *2022 IEEE Int. Jt. Conf. Biometrics, IJCB 2022*, 2022, doi: 10.1109/IJCB54206.2022.10007961.

[25] M. Krichen, "Generative Adversarial Networks," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2023, pp. 1–7. doi: 10.1109/ICCCNT56998.2023.10306417.

[26] S. Gao, J. Tian, X. Fu, Y. Li, B. Wang, and L. Zhao, "Complete loss distribution model of GaN HEMTs considering the influence of parasitic parameters," *J. Power Electron.*, vol. 24, no. 1, pp. 119–129, 2024, doi: 10.1007/s43236-023-00710-3.

[27] R. Al Mdanat, R. Georgious, S. Saeed, and J. García, "Power Loss Modeling and Impact of Current Measurement on the Switching Characterization in Enhancement-Mode GaN Transistors," *IEEE J. Emerg. Sel. Top. Power Electron.*, vol. 12, no. 6, pp. 5404–5420, 2024, doi: 10.1109/JESTPE.2024.3431270.

[28] J. Nilsson and T. Akenine-Möller, *Understanding SSIM*. 2020. doi: 10.48550/arXiv.2006.13846.

[29] T. B. Taha, "Modified PSNR Metric for Watermarked-Image Assessment," in *2023 20th International Multi-Conference on Systems, Signals & Devices (SSD)*, 2023, pp. 414–418. doi: 10.1109/SSD58187.2023.10411302.

[30] D. Feature and R. Loss, "Facial inpainting," *JUTI J. Ilm. Teknol. Inf.*, vol. 18, no. 2, pp. 171–178, 2020.