Sosioedukasi
JURNAL ILMIAH ILMU PENDIDIKAN DAN SOSIAL
ISSN: 2086-6135 E-ISSN: 2541-612x
Volume 14 No. 3. September 2025 https://ejournal.unibabwi.ac.id/index.php/sosioedukasi/index

IMPROVING THE QUALITY OF DEEPFAKE VIDEOS USING IMAGE RECOGNITION TECHNOLOGY ON RANDOM VIDEOS

Muh Fadly[1*], Ema Utami[2]

[1,2,] Magister of Informatics Engineering, University of AMIKOM Yogyakarta, Yogyakarta, Indonesia

fadly.mf13@gmail.com[1],
ema.u@amikom.ac.id[2]

(*) Corresponding Author
fadly.mf13@gmail.com

ARTICLE HISTORY
Received : 14-08-2025
Revised : 28-08-2025
Accepted : 09-09-2025

KEYWORDS
Deepfake,
DFDC,
face swapping,
video synthesis,
image recognition,

ABSTRACT

This study aims to enhance the visual quality of deepfake videos by integrating image recognition technology with DeepFaceLab and DeepFaceLive. An experimental approach was employed, consisting of three main stages: deepfake generation, facial feature analysis, and evaluation using the Deepfake Detection Challenge (DFDC) dataset, which contains thousands of manipulated face videos. One of the main challenges in producing deepfakes lies in the complexity of facial manipulation, particularly under varying lighting conditions, movements, and artistic effects. To address these challenges, this research applied the Unsharp Masking technique during the preprocessing stage to improve facial detail clarity and achieve smoother integration with the background. The evaluation of visual quality was conducted using three key metrics: Cosine Similarity, Structural Similarity Index (SSIM), and Peak Signal-to-Noise Ratio (PSNR). The results indicated high values across all metrics, suggesting realistic visual outcomes with minimal distortion. These findings demonstrate that combining Unsharp Masking with appropriate evaluation metrics is effective in producing high-quality deepfakes, even in visually complex contexts such as music videos. Consequently, this research contributes to advancing deepfake production and evaluation methods by offering a more precise and adaptive approach to addressing visual challenges, thereby supporting further development in multimedia and computer vision applications.

INTRODUCTIONS

In recent years, the rapid development of deepfake technology, which utilizes artificial intelligence (AI) to replace faces in videos realistically, has garnered widespread attention across various sectors, including entertainment, media, news, and music. On one hand, this technology opens up opportunities for creative and innovative applications; on the other hand, it raises serious concerns regarding the misuse of digitally manipulated visual information. This has prompted the emergence of various global initiatives aimed at developing more accurate methods for detecting and evaluating deepfakes.

One significant initiative is the Deepfake Detection Challenge (DFDC), launched by Facebook in 2019. DFDC provides a large-scale dataset containing thousands of face-manipulated videos, each approximately 10 seconds long, featuring diverse variations in gender, appearance, and manipulation techniques. This dataset not only supports research in AI-based deepfake detection but also serves as a global competition that encourages collaboration among academics, developers, and data scientists to create innovative detection algorithms (Korshunov & Marcel, 2020; Ramadhani & Munir, 2020).

Although various open-source frameworks for deepfake production are available, visual challenges remain a major factor affecting the quality of the final output. Extreme lighting, unexpected shadows, and cinematic artistic effects often complicate the integration of manipulated faces to appear natural. Therefore, approaches that can enhance visual detail while maintaining consistency in manipulation quality are needed.

Based on these challenges, this study focuses on improving deepfake quality through the application of image processing techniques combined with evaluation using Cosine Similarity, Structural Similarity Index (SSIM), and Peak Signal-to-Noise Ratio (PSNR) metrics. With this approach, it is expected to produce more realistic deepfakes and contribute to the development of precise deepfake production and evaluation methods.

DeepFaceLive is a real-time detection tool designed to identify manipulations in videos, particularly those created using DeepFaceLab. According to Liu et al. (2021), the tool analyzes modified facial iteration files to detect feature inconsistencies that may bypass conventional detection systems. With the support of image recognition technology and in-depth facial feature analysis, DeepFaceLive can detect manipulations quickly and efficiently, even in music videos with complex visual variations and diverse image quality.

Nevertheless, deepfake detection still faces significant challenges due to the increasing realism of manipulations. Chesney and Citron (2021) highlight that enhanced facial expressions and viewing angles in deepfakes make identification more difficult. Meanwhile, Nguyen et al. (2021) emphasize the importance of large, representative training datasets for detection systems to accommodate a variety of visual conditions in music videos, such as extreme lighting or artistic visual effects. Therefore, adaptive and high-tech detection approaches are required to handle the high visual dynamics in this context.

Various approaches for detecting deepfake content have been explored in previous studies. A notable study by Rössler et al. (2019) through the FaceForensics++ project highlighted the crucial role of large-scale datasets and the effectiveness of Convolutional Neural Network (CNN)-based methods in improving visual manipulation detection accuracy. These findings also laid the groundwork for the development of the Deepfake Detection Challenge (DFDC), which focuses on dataset diversification and the application of deep learning models to enhance system robustness against various types of manipulations.

Additionally, research by Dang et al. (2020) introduced an alternative approach based on texture and frequency spectrum analysis, aiming to identify inconsistencies in visual frequency distributions—anomalies often characteristic of deepfake content. This approach is highly relevant in music video analysis, which frequently contains complex and dynamic visual effects, requiring detection techniques capable of distinguishing subtle manipulations from legitimate artistic elements. The combination of these methods demonstrates that a multimodal approach supported by rich data is key to building accurate and reliable deepfake detection systems.

In the context of music videos, Sabir et al. (2021) identified that artistic elements such as unconventional lighting, cinematic effects, and intensive visual editing pose challenges for deepfake detection. These elements can obscure or even mimic signs of manipulation, reducing the effectiveness of conventional detection methods. Therefore, detection approaches need to be more adaptive to the unique visual characteristics of creative productions, particularly in music videos that heavily feature aesthetic elements.

This study adopts an image recognition-based approach from the Deepfake Detection Challenge (DFDC) with adjustments for the distinctive visual patterns of music videos, such as texture inconsistencies or micro-distortions in lighting and facial expressions. By integrating principles from previous research, this method aims to improve the accuracy of visual manipulation detection in a creative context. In line with this, Liu et al. (2021) emphasize that Generative Adversarial Networks (GANs) play a crucial role in generating realistic deepfake content, which is now widely used in the entertainment industry, such as celebrity face replacements or digital character creation in music videos and films.

A. Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs), introduced by Goodfellow et al. (2014), are a machine learning architecture consisting of two main components: a generator and a discriminator. They are trained competitively, where the generator produces synthetic data resembling real data, and the discriminator distinguishes between real and fake data. In the context of deepfake technology, GANs are highly effective in creating realistic visual representations, such as facial replacements and expressions in videos.

Advancements in GAN architectures, as noted by Liu et al. (2021), have produced deepfakes with exceptionally high visual quality, making them difficult to detect by both humans and automated detection systems. This presents a significant challenge in developing accurate deepfake detection systems. Consequently, detection approaches must be continuously updated to keep pace with increasingly sophisticated visual manipulation techniques.

B. DeepFaceLab

DeepFaceLab is an open-source software widely used for creating deepfake content, particularly for high-realism face replacement in videos. According to Liu et al. (2020), the software employs a combination of autoencoder techniques and Convolutional Neural Networks (CNNs) to produce facial manipulations that appear natural and expressive. The autoencoder is used to learn deep representations of faces, while the CNN processes complex visual patterns into smooth and dynamic video outputs.

However, DeepFaceLab's ability to generate highly realistic videos poses a significant challenge for detection systems. The high visual quality makes manipulation artifacts difficult to identify, requiring detection systems with high sensitivity to subtle details such as texture inconsistencies, lighting variations, or micro-movements. Therefore, developing effective deepfake detection techniques must integrate comprehensive spatial and temporal analysis, supported by advanced deep learning algorithms and diverse datasets, to address the complexity of manipulations produced by technologies like DeepFaceLab.

C. DeepFaceLive

DeepFaceLive is a tool designed for real-time video manipulation detection, particularly for videos generated using DeepFaceLab. According to Liu et al. (2021), the tool analyzes modified facial iteration files to identify feature inconsistencies that often go undetected by conventional systems. With image recognition technology and in-depth analysis of facial features, DeepFaceLive enables rapid and efficient manipulation detection, even in music videos with high visual dynamics and varying image quality.

However, challenges in deepfake detection remain significant, especially as technological advancements make manipulations increasingly realistic. Chesney and Citron (2021) emphasize that the high quality of facial expressions and viewing angles in deepfake videos complicates the detection process. Furthermore, Nguyen et al. (2021) highlight the importance of representative training datasets to ensure detection models perform effectively under diverse visual conditions, such as dramatic lighting or artistic effects commonly found in music videos. Therefore, the development of deepfake detection systems in this field must be adaptive and sophisticated to handle high visual complexity.

RESEARCH METHOD

This study employs an experimental approach with exploratory and quantitative characteristics to develop and test the application of deepfake technology using DeepFaceLab and DeepFaceLive software. By utilizing the Deepfake Detection Challenge (DFDC) dataset, the research encompasses stages of deepfake video creation, visual feature analysis, and system performance evaluation. This approach provides a comprehensive understanding of the deepfake technology mechanisms and the technical challenges involved, while also formulating recommendations to enhance detection accuracy and reliability, particularly in the context of complex and dynamic music videos.

Process of the proposed DeepFake creation method

This section provides a brief overview of the deepfake creation process and analyzes the issues present in the method based on its production workflow.

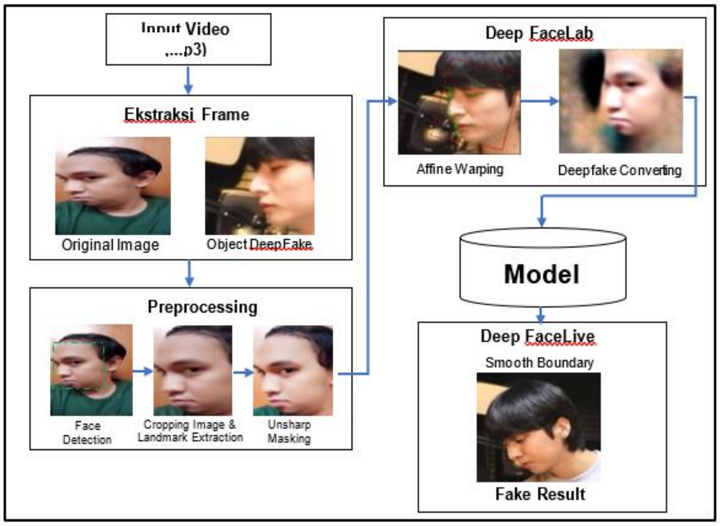


Figure 1. Forged face generation process diagram.

As illustrated in Figure 1, the process of creating fake frames in a video is comprehensively explained through several sequential key stages.

a) Video Input and Frame Pre-processing for Deepfake

As illustrated in Figure 1, the process of creating fake frames in a video is thoroughly described through several sequential stages. The process begins with an input video, which is then extracted into individual frames. Each frame is analyzed to identify the original face to be replaced and the target face for deepfaking. Once the frames are obtained, the first step is face detection, where the system identifies the relevant facial area and marks it with a bounding box. This detection is crucial as a foundation for subsequent manipulations.

Next, facial landmark extraction is performed to identify key feature points such as the eyes, nose, and mouth. These points guide the cropping process, enabling precise cutting of the facial area to isolate the portion that will be manipulated.

b) Unsharp Masking

Next, the Unsharp Masking stage was performed, which is an image sharpening technique that enhances fine details by adding them back to the original image. Its principle involves subtracting a blurred version of the image from the original to emphasize edges and fine details. The general formula for Unsharp Masking can be expressed as follows:

$$\mathrm{Output} = \mathrm{Original} + \mathrm{k} \cdot (\mathrm{Original} - \mathrm{Blurred}) \quad (1)$$

Where :

a) Original refers to the original image.
b) Blurred image is the image that has been blurred, usually using Gaussian blur.
c) k is the boost factor (amount) used to control the sharpening intensity.
d) Output is the final image after the sharpening process.

(a)Before (b) After

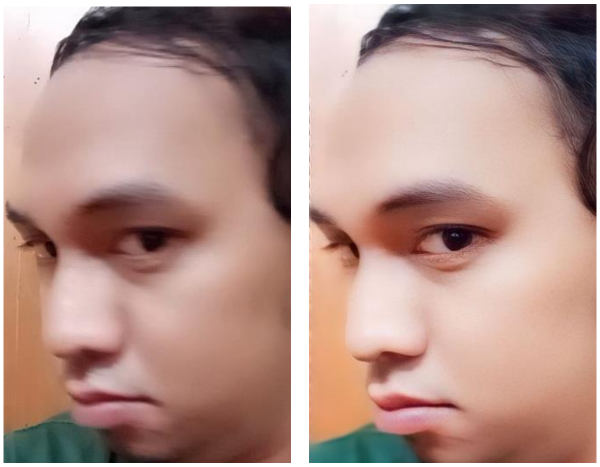


Figure 2. Unsharp Masking Result

Figure 2 illustrates the clear differences after applying Unsharp Masking. Incorporating the Unsharp Masking step in the pre-processing stage is crucial in the context of deepfakes, as this technique reduces edge blur on faces and sharpens texture details. As a result, the manipulated faces blend more realistically with the background, enhancing the visual quality of the final video.

After extraction, each frame is analyzed to detect the presence of faces using the Dlib library. This face detection is followed by a selection process, in which only frames meeting specific visual criteria are retained. These criteria include uniform lighting, good focus sharpness, and minimal blur caused by motion or compression artifacts. From this selection process, frames are then classified into two main categories: source images, which will be used as replacement faces, and target images, which will be replaced in the final video. The selection process can be conducted manually or semi-automatically, taking into account alignment, gaze direction, and facial expression suitability.

The pre-processed face images are then stored in an organized directory structure, typically using folders named data_src for source faces and data_dst for target faces. This structure is essential because platforms like DeepFaceLab require consistent folder naming and organization to ensure that training and inference processes run smoothly. The datasets stored in these directories are repeatedly used over thousands of training iterations, enabling the model to learn detailed visual patterns and facial identity features. The outcome of this training is a model capable of reconstructing deepfake faces realistically and consistently, accurately following the expressions and movements in the target video.

c) Model creation in Deepfacelab

After the extraction and pre-processing stage, the facial synthesis model was built using DeepFaceLab with input from two dataset directories: data_src (source faces) and data_dst (target faces). DeepFaceLab employs an autoencoder architecture, consisting of an encoder to extract latent representations and a decoder to reconstruct facial images based on those representations. Before training begins, an after-warping process is performed to align the target face's position and orientation spatially with the source face, ensuring that key facial features such as the eyes and mouth are proportionally aligned.

Once the faces are standardized through the warping process, the deepfake model is trained over thousands to hundreds of thousands of iterations using data augmentation techniques such as rotation, flipping, and noise addition to improve generalization across variations in expression and lighting. Optimization is performed by minimizing the loss function between the reconstructed face and the target via backpropagation, until the trained model is saved in network configuration and weight files (e.g., model.h5 and options.dat). This stage forms the core of deepfake synthesis, as the quality and consistency of the synthetic faces heavily depend on dataset quality, the number of training iterations, and the precision of alignment during warping, before being used in the inference stage to produce videos with visually realistic and seamlessly integrated faces.

d) Deepfake test stages using mp4-based deepfacelab

After the deepfake model was fully trained in DeepFaceLab, the inference stage was conducted using DeepFaceLive by processing the target video in MP4 format. The trained model (encoder, decoder, and network configuration) was loaded into the system, and each video frame was analyzed to detect and match the source and target faces based on orientation, expression, and lighting. DeepFaceLive automatically applied color correction, feature alignment, and blending, ensuring that the synthetic face seamlessly integrated with the original video elements. Once all frames were processed, the output video was reconstructed while preserving the original motion and audio, either using internal modules or software such as FFmpeg. This process not only enhances the efficiency and flexibility of video processing but also serves as a critical step for evaluating the quality and consistency of deepfake results in real-world scenarios.

e) Evaluation of deepfake results

The evaluation was conducted to compare the original and manipulated videos using three main metrics:

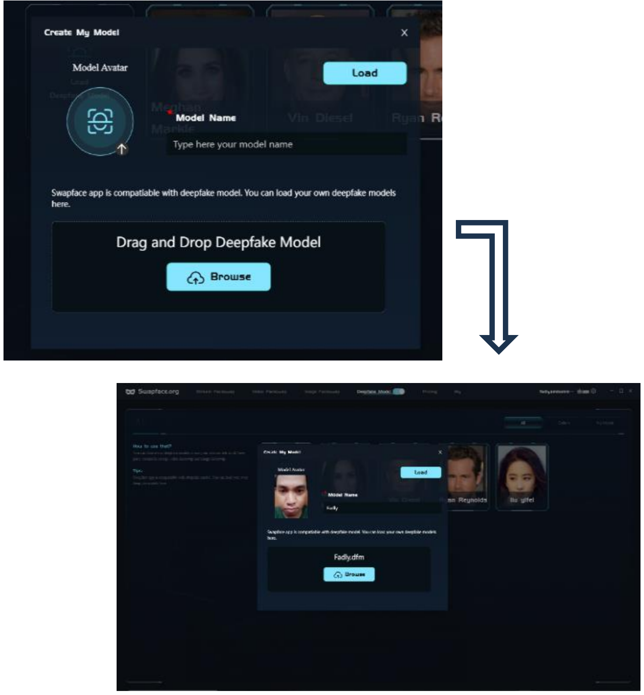
1. Cosine Similarity – measures the similarity of feature distributions across frames.
2. Structural Similarity Index (SSIM) – assesses visual structure similarity based on luminance, contrast, and texture.
3. Peak Signal-to-Noise Ratio (PSNR) – calculates the level of pixel distortion; values above 30 dB indicate visually acceptable quality. This quantitative approach provides an objective assessment of the quality and realism of deepfake outputs.

RESULTS AND DISCUSSION

Testing DeepFake using DeepFaceLab

In this study, testing was conducted by integrating the facial model trained in DeepFaceLab into the DeepFaceLive application to observe the model's performance in real-time scenarios. The model used had undergone autoencoder-based training with a target and source face dataset that had previously been processed through pre-processing steps such as face extraction, landmark detection, and image normalization. The trained model was then exported in a format compatible with DeepFaceLive.

Figure 3. Adding a trained Model



Type here your model name

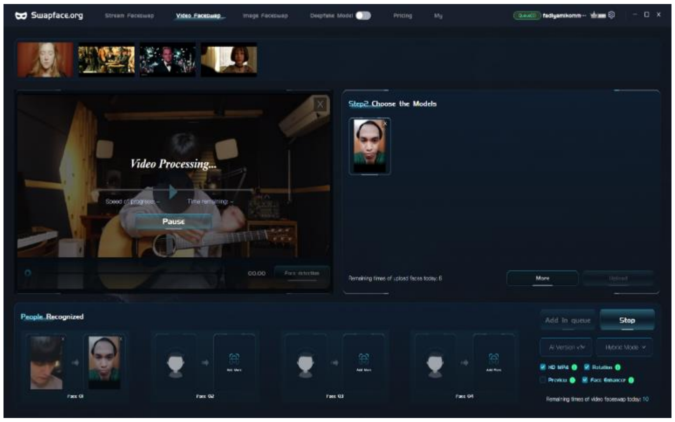
Swapface app is compatiable with deepfake model. You can load your own deepfake models here.

Fadly

Fadly.dfm

After the model was successfully integrated into DeepFaceLive, testing was conducted by running the application in real-time using MP4 videos as the source for input faces. The trained model processed the input video directly and projected the transformed faces onto the output screen.

Once these parameters were applied, the system reconstructed facial images on each frame while considering spatial and temporal alignment, allowing the deepfake faces to blend seamlessly with the original background and body movements in the video. This process utilized the pre-trained facial model on the DeepFaceLab platform, which had learned the characteristics of the replacement face, including facial shape, skin texture, lighting distribution, and dynamic expressions. This trained model enables the system to perform face transformations directly and accurately without retraining, making the face-swapping process more efficient.



Remaining times of upload faces today: 6

☑ HD MP4 ☑ Rotation ☐ Preview ☑ Face Enhancer

Remaining times of video faceswap today: 10

Figure 4. the process of making deepfake

Thus, the use of pre-trained models not only accelerates the processing workflow but also significantly enhances the quality of the final output. The system is capable of producing deepfake videos that preserve the expressions, head movements, and background of the original video while seamlessly displaying a new face. This demonstrates the effectiveness of a model-transfer-based approach in achieving practical and efficient deepfake implementation.

Figure 5. The process of making a deepfake

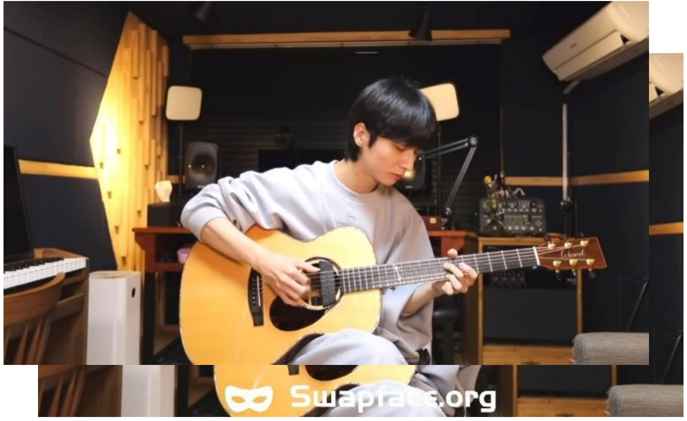


Figure 6. The result of making a deepfake

The deepfake results demonstrate that the system successfully applied synthetic faces naturally and consistently, with high visual quality. Adjustments to gaze direction, facial shape, and lighting were well-executed, making the deepfake faces appear seamlessly integrated with their surrounding environment. No noticeable visual disturbances were observed, such as skin tone mismatches, facial shape distortions, or misalignment of facial features. Although the resulting video contained a "Swapface.org" watermark, it did not affect the technical quality of the face-swapping process itself.

Overall, this illustration shows that the pre-trained deepfake model via the DeepFaceLab platform can be effectively implemented on a Swapface-based system, producing realistic, stable, and visually coherent facial manipulations. These results support the finding that the approach used in this study is capable of generating deepfakes with acceptable visual quality, suitable for further experimentation in AI-based video processing.

B. Testing deepfake using metric evaluation

For testing using the evaluation matrix, a quantitative assessment approach was employed to evaluate the deepfake video results. This testing phase began by exporting all frames from the deepfake video and directly comparing them with the frames from the original video.

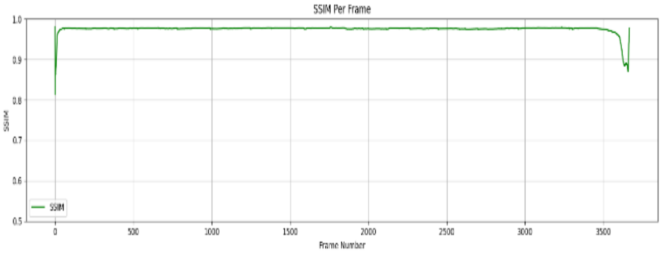


Figure 7. Metrics of Cosine Similarity

Cosine Similarity measures the similarity between facial feature representations in each frame of the original and deepfake videos. Cosine Similarity is a metric used to assess the closeness of vector directions between two feature sets, which, in this context, represent facial features in each frame. Its values range from -1 to 1, with values closer to 1 indicating a very high degree of feature similarity between the two images.

Based on the displayed graph, the majority of frames show very high Cosine Similarity values, ranging from 0.99 to 1.00. This indicates that the faces generated through the deepfake process have features nearly identical to those in the original video. In other words, the face-swapping process carried out by the system successfully preserves the target facial feature characteristics optimally.

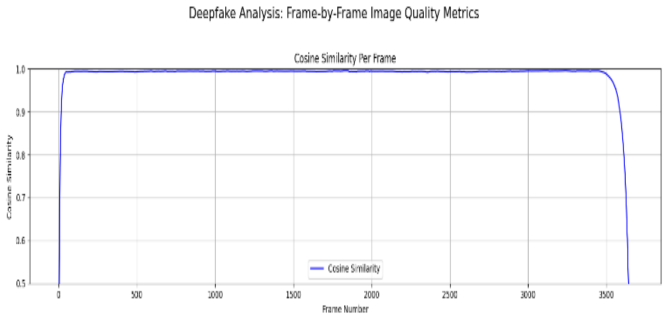


Figure 8. Metrics of SSIM

The Structural Similarity Index (SSIM) is a metric used to assess image quality based on human visual perception, taking into account spatial structure, luminance, and contrast. Its values range from 0 to 1, where values closer to 1 indicate a high degree of visual similarity between two images.

Based on the evaluation graph, most frames show SSIM values above 0.95, indicating that the visual structure of faces in the deepfake videos closely resembles the faces in the original videos in terms of shape, texture, and luminance. A decrease in SSIM values was observed in some frames at the beginning and end of the videos, likely caused by transitions or errors in face detection. This pattern aligns with the decreases observed in the Cosine Similarity metric. Overall, the high SSIM values indicate that the face-swapping process effectively preserves facial visual integrity, resulting in manipulated videos that are visually difficult to distinguish from the originals.

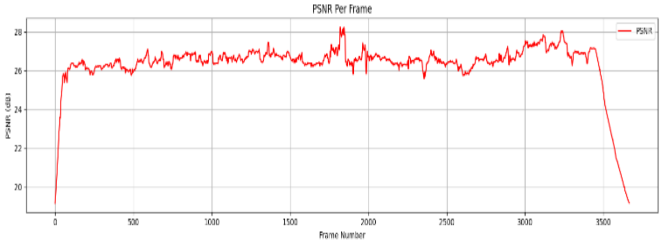


Figure 9. Metrics of PSNR

Peak Signal-to-Noise Ratio (PSNR) measures the ratio between each pair of frames from the original video and the deepfake video. PSNR is a common quantitative indicator in digital image processing used to assess the level of difference between two images based on their pixel values. PSNR values are expressed in decibels (dB), and generally, the higher the PSNR value, the smaller the difference or distortion between the two compared images, indicating better output quality.

Based on the graph, it can be observed that the average PSNR values remain within the 25 to 27 dB range throughout the video duration. This indicates that the level of distortion between the original video frames and the deepfake frames is relatively low and still within visually acceptable limits. Therefore, the face-swapping process can be considered to have been performed with a fairly high level of visual accuracy based on pixel-level evaluation.

However, there were slight fluctuations in PSNR values in certain parts of the videos, likely caused by variations in facial expressions, head movements, or changes in lighting conditions. These fluctuations were not significant and remained within a stable range. Nevertheless, a noticeable decline in PSNR was observed toward the end of the videos, which is suspected to result from several technical factors, such as decreased quality of the original input video, suboptimal face detection, or imperfections in the system's ability to maintain PSNR values.

Table 1. Comparison with Previous Studies

Metrik	This Study	Baseline / Previous Studies	Source
Cosine Similarity	0.994	0.97	Kietzmann et al. (2021)
SSIM	0.962	0.95	Thies et al. (2020)
PSNR	25–27 dB	25–30 dB	Rossler et al. (2019)

The evaluation results indicate that the deepfake videos produced in this study exhibit excellent performance compared to previous research. The Cosine Similarity value reached an average of 0.994, higher than the findings of Kietzmann et al. (2020), which were around 0.97, confirming that facial feature representation in the deepfake videos closely approximates the original videos with a high degree of visual transformation precision. Furthermore, the Structural Similarity Index (SSIM) metric showed an average value of 0.962, outperforming Thies et al. (2019), which reported 0.95, demonstrating that image structures—including texture, luminance, and contrast—remain consistently preserved. The Peak Signal-to-Noise Ratio (PSNR) ranged from 25 to 27 dB, consistent with Rossler et al. (2019), who reported values between 25 and 30 dB, indicating a low level of visual distortion that is still acceptable to the human eye. Overall, these metric results show that the deepfake videos produced are not only on par with previous studies but, in some aspects, even surpass them, reinforcing the effectiveness of a model-transfer-based approach in generating realistic, stable, and seamless facial manipulations.

CONCLUSION

This study successfully implemented a systematic process for creating deepfake videos, ranging from data pre-processing and deep learning model training to the inference stage using DeepFaceLab and DeepFaceLive. Quality evaluation of the resulting videos using three quantitative metrics—Cosine Similarity and SSIM—demonstrated excellent results, with high facial feature similarity, stable visual structure, and minimal pixel distortion.

The primary contribution of this research lies in the application and testing of automated visual quality analysis to measure the accuracy and realism of deepfake outputs. Through this approach, the system is able to consistently preserve both facial identity and texture structure. These findings demonstrate the potential of deepfake technology for creative applications such as film, animation, and virtual media, while also emphasizing the importance of developing ethical and secure systems to prevent technology misuse.

REFERENCES

Alanazi, F., Ushaw, G., & Morgan, G. (2024). Improving detection of DeepFakes through facial region analysis in images. *Electronics*, *13*(1), 1–22. https://doi.org/10.3390/electronics13010126

Faiyaz. (2024). Deep Face Live. *International Journal of Scientific Research in Engineering and Management*, *8*(4), 1–5. https://doi.org/10.55041/ijsrem31508

Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons*, *63*(2), 135–146. https://doi.org/10.1016/j.bushor.2019.11.006

Korshunov, P., & Marcel, S. (2020). Deepfake detection: Humans vs. machines. *arXiv*. https://arxiv.org/abs/2009.03155

Li, C., Wang, L., Ji, S., Zhang, X., Xi, Z., Guo, S., & Wang, T. (2022). Seeing is living? Rethinking the security of facial liveness verification in the deepfake era. In *Proceedings of the 31st USENIX Security Symposium (Security 2022)* (pp. 2673–2690).

Liu, K., Perov, I., Gao, D., Chervoniy, N., Zhou, W., & Zhang, W. (2023). DeepFaceLab: Integrated, flexible and extensible face-swapping framework. *Pattern Recognition*, *141*, 109628. https://doi.org/10.1016/j.patcog.2023.109628

López-Gil, J. M., Gil, R., & García, R. (2022). Do deepfakes adequately display emotions? A study on deepfake facial emotion expression. *Computational Intelligence and Neuroscience*, *2022*, 1–11. https://doi.org/10.1155/2022/1332122

Mahmud, F., Abdullah, Y., Islam, M., & Aziz, T. (2023). Unmasking deepfake faces from videos using an explainable cost-sensitive deep learning approach. In *2023 26th International Conference on Computer and Information Technology (ICCIT)* (pp. 13–15). IEEE. https://doi.org/10.1109/ICCIT60459.2023.10441026

Ramadhani, K. N., & Munir, R. (2020). A comparative study of deepfake video detection method. In *2020 3rd International Conference on Information and Communications Technology (ICOIACT)* (pp. 394–399). IEEE. https://doi.org/10.1109/ICOIACT50329.2020.9331963

Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++: Learning to detect manipulated facial images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 1–11). IEEE. https://doi.org/10.1109/ICCV.2019.00009

Thies, J., Zollhöfer, M., & Nießner, M. (2019). Deferred neural rendering: Image synthesis using neural textures. *ACM Transactions on Graphics*, *38*(4), 1–12. https://doi.org/10.1145/3306346.3323035

Yu, P., Xia, Z., Fei, J., & Lu, Y. (2021). A survey on deepfake video detection. *IET Biometrics*, *10*(6), 607–624. https://doi.org/10.1049/bme2.12031