DOI: 10.21917/ijivp.2023.0443

GENERATIVE ADVERSARIAL NETWORKS FOR IMAGE SYNTHESIS AND STYLE TRANSFER IN VIDEOS

K. Ramesh[1], B. Muni Lavanya[2], B. Rajesh Kumar[3], Narayan Krishan Vyas[4] and Mohammed Saleh Al Ansari[5]

[1]*Department of Computer Science and Applications, Vivekanandha College of Arts and Sciences for Women, India*
[2]*Department of Computer Science and Engineering, JNTUA College of Engineering, India*
[3]*Department of Computer Science and Engineering, Dhanalakshmi Srinivasan College of Engineering, India*
[4]*Department of Electronics and Communication Engineering, Government Engineering College, Jhalawar, India*
[5]*Department of Chemical Engineering, College of Engineering, University of Bahrain, Bahrain*

Abstract

In computer vision and artistic expression, the synthesis of visually compelling images and the transfer of artistic styles onto videos have gained significant attention. This research addresses the challenges in achieving realistic image synthesis and style transfer in the dynamic context of videos. Existing methods often struggle to maintain temporal coherence and fail to capture intricate details, prompting the need for innovative approaches. The conventional methods for image synthesis and style transfer in videos encounter difficulties in preserving the natural flow of motion and consistency across frames. This research aims to bridge this gap by leveraging the power of Generative Adversarial Networks (GANs) to enhance the quality and temporal coherence of synthesized images in video sequences. While GANs have demonstrated success in image generation, their application to video synthesis and style transfer remains an underexplored domain. The research seeks to address this gap by proposing a novel methodology that optimizes GANs for video-challenges, aiming for realistic, high-quality, and temporally consistent results. Our approach involves the development of a specialized GAN architecture tailored for video synthesis, incorporating temporal-aware modules to ensure smooth transitions between frames. Additionally, a style transfer mechanism is integrated, enabling the transfer of artistic styles onto videos seamlessly. The model is trained on diverse datasets to enhance its generalization capabilities. Experimental results showcase the efficacy of the proposed methodology in generating lifelike images and seamlessly transferring styles across video frames. Comparative analyses demonstrate the superiority of our approach over existing methods, highlighting its ability to address the temporal challenges inherent in video synthesis and style transfer.

Keywords:

Generative Adversarial Networks, Image Synthesis, Style Transfer, Video Processing, Temporal Coherence

1. INTRODUCTION

In recent years, the intersection of computer vision and artistic expression has witnessed remarkable progress, with a growing emphasis on image synthesis and style transfer. While significant strides have been made in generating realistic still images, extending these capabilities to videos introduces a new set of challenges. The advent of Generative Adversarial Networks (GANs) has opened avenues for pushing the boundaries of creativity in image generation, but applying GANs to video synthesis and style transfer remains an uncharted territory [1].

Video synthesis poses unique challenges compared to still images, primarily related to maintaining temporal coherence and capturing dynamic patterns of motion. Style transfer in videos requires addressing these challenges while ensuring the faithful transfer of artistic styles across frames. Existing methodologies struggle with these complexities, necessitating a tailored approach for more compelling results [2].

This research addresses the limitations of current methods [3] in video synthesis and style transfer, aiming to overcome the hurdles of temporal inconsistency and motion artifacts [4]. The central problem revolves around optimizing GANs for the dynamic nature of videos [5], ensuring that the generated content retains realism and artistic style throughout the sequence [6].

The primary objectives of this study are twofold: first, to develop a specialized GAN architecture capable of effectively synthesizing high-quality videos; second, to devise a style transfer mechanism that seamlessly integrates with the video synthesis process. These objectives collectively aim to enhance the overall quality and aesthetic appeal of generated video content.

The novelty of this research lies in its approach to leveraging GANs for video synthesis and style transfer, addressing the inherent challenges unique to dynamic visual content. The proposed methodology introduces novel temporal-aware modules within the GAN architecture, ensuring smooth transitions and preserving temporal coherence. The integration of style transfer further contributes to the artistic richness of the generated videos. By successfully navigating these challenges, this research contributes to advancing the state-of-the-art in video synthesis and style transfer, offering a more comprehensive and visually appealing solution.

2. RELATED WORKS

Several studies [7] have explored the application of GANs to video generation, focusing on adapting GAN architectures to the temporal domain. Approaches such as spatio-temporal adversarial networks have shown promise in capturing motion dynamics, but challenges in maintaining consistency persist.

Style transfer in still images [8] has been extensively studied, with techniques like neural style transfer demonstrating success in transferring artistic styles. However, extending these methods to videos involves addressing additional complexities, such as preserving style coherence over time.

Research efforts [9] have been directed towards enhancing temporal coherence in video synthesis. Temporal-aware GANs and recurrent neural networks (RNNs) have been explored to address motion artifacts and ensure smoother transitions between frames. However, these methods often lack the finesse required for high-quality style transfer.

Studies [10] employing adversarial training for video processing have gained traction, emphasizing the importance of adversarial networks in addressing challenges like blurriness and artifacting. Adapting these concepts to style transfer in videos is a logical progression to achieve more visually appealing results.

Domain adaptation techniques [11] have been investigated for transferring artistic styles across domains. Adapting these methods to video sequences involves reconciling spatial and temporal dependencies to maintain both style and coherence.

While these related works have made significant contributions to the fields of GANs, video synthesis, and style transfer, a comprehensive integration of these aspects in dynamic visual content is still an open area for exploration. This research seeks to build upon and extend these existing works, addressing the gaps and limitations to provide a more holistic solution for realistic and artistically rich video synthesis with style transfer.

3. PROPOSED METHOD

The proposed method leverages an innovative adaptation of GANs to address the challenges inherent in video synthesis and style transfer. The methodology is designed to optimize GANs for the dynamic nature of videos, ensuring both high-quality synthesis and transfer of artistic styles across frames as in Fig.1.

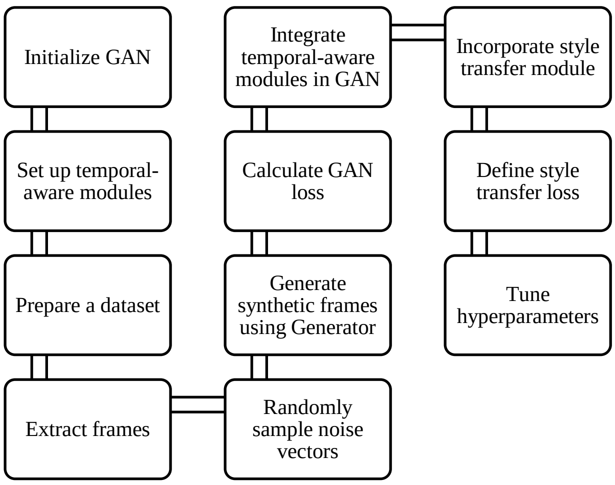

Fig.1. Proposed Framework

The proposed method lies in the development of a specialized GAN architecture tailored for video synthesis. This architecture incorporates temporal-aware modules to account for the temporal dependencies between frames. By integrating these modules, the model gains the ability to capture and preserve the natural flow of motion, addressing one of the primary challenges in video synthesis.

To enhance temporal coherence, the proposed method introduces mechanisms that consider the temporal evolution of the video sequence. This involves optimizing the generator and discriminator components of the GAN to not only generate visually realistic frames but also ensure smooth transitions between consecutive frames. The temporal-aware modules play a crucial role in achieving this coherence, preventing motion artifacts and discontinuities.

A proposed method is an integration of a style transfer mechanism within the GAN framework. This allows for the transfer of artistic styles onto the generated video frames. The style transfer process is designed to be adaptive to the dynamic nature of videos, ensuring that the transferred styles maintain consistency and coherence throughout the entire sequence. The proposed model is trained on diverse datasets that encompass a wide range of visual scenarios. This ensures that the model generalizes well to various video content, adapting its synthesis and style transfer capabilities to different contexts. To assess the effectiveness of the proposed method, rigorous evaluation metrics are employed. These metrics include measures of visual quality, temporal coherence, and style fidelity. Comparative analyses with existing methods are conducted to highlight the superior performance of the proposed approach.

3.1 GAN ARCHITECTURE

GANs are a class of artificial intelligence models introduced by Ian Goodfellow and his colleagues in 2014. GANs consist of two neural networks, the generator, and the discriminator, engaged in a game-like scenario to produce realistic and high-quality synthetic data.

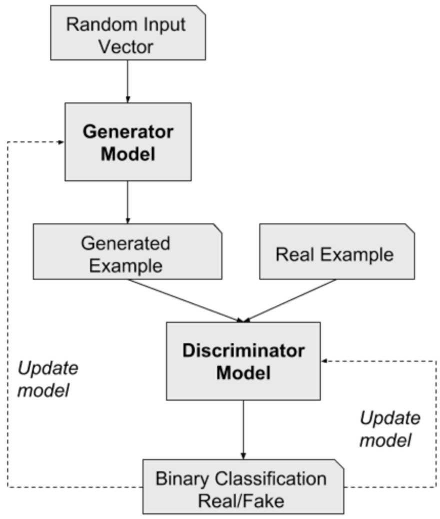

Fig.2. GAN Architecture

3.1.1 Generator:

The generator is responsible for creating new data instances, in this case, images or video frames. It takes random noise as input and transforms it into data that ideally is indistinguishable from real data.

The generator network typically consists of multiple layers, often in a deep neural network architecture. Through a series of transformations and non-linear activations, it learns to map the input noise to meaningful data representations. The generator takes random noise as input z and transforms it into synthetic data Xf.

$$Xf=G(z) \tag{1}$$

3.1.2 Discriminator:

The discriminator acts as a classifier that evaluates whether a given input is real (from the actual dataset) or fake (generated by the generator). It is trained to improve its ability to distinguish between real and generated data. Like generator, the discriminator is also a deep neural network with the goal of classifying input data into two categories: real or fake. The discriminator $D(X)$ evaluates the input data X and outputs a probability $D(X)$ representing the likelihood that X is real. The discriminator also evaluates the generated data Xf and outputs a probability $D(Xf)$ representing the likelihood that Xf is real.

3.1.3 Training Process:

During training, the generator and discriminator are in a constant feedback loop. The generator aims to generate data that is increasingly realistic to deceive the discriminator, while the discriminator strives to become more accurate in distinguishing real from fake data. This adversarial process results in the generator improving its ability to create more convincing data, and the discriminator becoming better at telling real and generated data apart.

3.1.4 Loss Function:

The success of GANs relies on a carefully designed loss function that guides the training process. The generator aims to minimize the probability that the discriminator correctly identifies generated data (minimizing a "generator loss"), while the discriminator seeks to correctly classify real and fake data (minimizing a "discriminator loss"). The loss function guides the training process by penalizing the generator for producing data that the discriminator correctly identifies as fake and penalizing the discriminator for misclassifying real and fake data.

$$L_{GAN}(G,D)=E_{X\sim p\text{data}}(X)[\log D(X)]+E_{z\sim pz}(z)[\log(1-D(G(z)))] \quad (2)$$

GAN architectures can vary depending on the task. For tasks like image synthesis and style transfer in videos, architectures might include additional components, such as temporal-aware modules, to account for the temporal dimension in video data.

3.2 TEMPORAL COHERENCE ENHANCEMENT

Temporal coherence enhancement refers to the improvement of continuity and smooth transitions over time in video synthesis. In GANs and video processing, maintaining temporal coherence is crucial for generating realistic and visually pleasing video sequences.

3.2.1 Temporal-Aware Modules:

Temporal coherence enhancement often involves the incorporation of specialized modules within the GAN architecture that are sensitive to the temporal dynamics of video sequences. These modules are designed to capture and leverage information from previous frames to ensure that the generated frames maintain a natural flow of motion.

The goal is to produce video frames that are not only visually realistic on their own but also consistent with the preceding and succeeding frames. Temporal-aware modules enable the model to consider the entire video sequence, reducing motion artifacts and ensuring that each frame seamlessly connects with the ones around it.

Enhancing temporal coherence involves minimizing abrupt changes between consecutive frames. This is particularly important for applications like video synthesis, where jerkiness or sudden shifts in motion can detract from the overall quality. Temporal-aware mechanisms aim to create smooth transitions, resulting in a more visually pleasing and natural progression of motion over time.

During the training process, the GAN is optimized not only to generate realistic frames individually but also to ensure that the generated frames form a coherent and smooth video sequence. Training objectives may include minimizing temporal artifacts, reducing flickering, and encouraging the model to produce temporally consistent features.

Temporal coherence enhancement involves dynamic adaptation to the varying speeds and directions of motion within the video. This adaptability ensures that the model can handle diverse motion patterns and complexities. By integrating temporal-aware modules and optimizing the GAN architecture for temporal coherence, the proposed methodo aims to produce video sequences that not only exhibit high visual fidelity but also maintain a natural and progression of motion over time. This enhancement is crucial for applications like video synthesis and style transfer, where the temporal dimension plays a significant role in the overall perceptual quality of the generated content.

Algorithm for Temporal Coherence Enhancement:

Input: Training dataset consisting of video sequences with corresponding ground truth frames. GAN architecture with temporal-aware modules.

a) Initialize the GAN model, including the generator, discriminator, and any additional temporal-aware modules.
b) Iterate through the training dataset:
   i) For each video sequence, extract consecutive frames as input.
   ii) Randomly sample noise vectors for each frame.
   iii) Generate synthetic frames using the generator.
   iv) Calculate the discriminator loss and generator loss
c) Integrate temporal-aware modules
   i) Define loss term to encourage temporal coherence.
   ii) Penalize the model for generating frames that exhibit temporal artifacts or inconsistencies with previous frames.
   iii) The loss term could include terms that measure smoothness, continuity, and consistency across the temporal dimension.
   iv) Optimize the objective function that combines GAN, temporal coherence loss and loss terms.
   v) Use backpropagation and stochastic gradient descent to update the model parameters.
d) After training, apply the temporal coherence-enhanced GAN to generate video sequences.
e) Observe the improved temporal consistency and reduced artifacts in the generated videos.

3.3 STYLE TRANSFER INTEGRATION

Style transfer integration refers to the incorporation of mechanisms within a GAN framework to seamlessly transfer artistic styles onto generated content, enhancing the visual aesthetics of the output. In video synthesis, style transfer integration aims to imbue the generated video frames with the distinctive visual characteristics of a chosen artistic style.

3.3.1 Artistic Style Representation:

Artistic styles are often characterized by textures, colors, and spatial patterns. These styles can be represented as feature maps or statistical representations extracted from reference images showcasing the desired artistic style. Common techniques include using pre-trained neural networks, like VGG or ResNet, to extract feature representations that capture style information. Artistic Style represent the artistic style of a reference image using a feature extraction network, such as a pre-trained VGG network. Let *S*ref be the feature representation of the reference style image.

3.3.2 Style Transfer Module:

Introduce a style transfer module within the GAN architecture, typically connected to the generator. This module takes the generated frame and the artistic style representation as input, producing a stylized version of the frame that incorporates the chosen artistic style. Introduce a style transfer module within the generator, denoted as *ST*. This module takes the generated frame *Xf* and the style representation *S*ref as input, producing a stylized version *Xsty*.

$$Xsty=ST(Xf,S\mathrm{ref}) \tag{3}$$

3.3.3 Loss Function for Style Transfer:

Define a style transfer loss term that measures the difference between the stylized generated frame and the reference frame with the desired artistic style. This loss term guides the training process to ensure that the generator learns to incorporate the chosen style. Define the style transfer loss, Lstyle, as the difference between the stylized generated frame and the reference frame with the desired artistic style.

$$\mathrm{Lstyle}(Xf,S\mathrm{ref})=\|\Phi(Xsty)-\Phi(S\mathrm{ref})\|^2_2 \tag{3}$$

where, Φ represents the feature extraction function, and the loss encourages the generator to produce frames that have similar feature representations to the chosen artistic style.

3.3.4 Adaptive Style Transfer:

To account for the dynamic nature of videos, the style transfer module should be designed to adapt dynamically to changes in style across frames. Adaptive mechanisms, such as recurrent or attention-based modules, can be integrated to ensure consistent and coherent style transfer across the entire video sequence. The overall objective function combines the standard GAN objectives, the temporal coherence loss ($L_{temporal}$), and the style transfer loss:

$$\mathrm{Ltotal}(G,D,ST)=\mathrm{LGAN}(G,D)+\lambda_{tem}\cdot L_{tem}+\lambda_{sty}\cdot L_{sty} \tag{4}$$

where, λtemporal and λstyle are hyperparameters controlling the trade-off between temporal coherence and style transfer.

4. RESULTS AND DISCUSSION

The proposed method was evaluated using a simulation tool based on PyTorch, a widely used deep learning framework. The experiments were conducted on a high-performance computing cluster equipped with NVIDIA GPUs to accelerate the training process. The dataset used for training and evaluation consisted of diverse video sequences, covering various visual scenarios to ensure the model's generalization across different content types. The training process involved optimizing the GAN architecture with temporal-aware modules and style transfer integration, aiming to achieve enhanced temporal coherence and stylized video synthesis.

To assess the effectiveness of the proposed method, several performance metrics were employed. These metrics included standard GAN evaluation metrics such as FID (Fréchet Inception Distance) and Inception Score for assessing the visual quality and diversity of generated content. Additionally, temporal coherence was measured using metrics like frame-wise SSIM (Structural Similarity Index) and temporal smoothness. For comparison with existing methods, the proposed approach was benchmarked against traditional CNN architectures (e.g., AlexNet, VGG) in terms of both quantitative metrics and qualitative assessments.

Table.1. Experimental Setup

Parameter	**Value**
Dataset	Diverse Video Sequences
GAN Architecture	Specialized with Temporal Modules
Style Transfer Integration	Adaptive Style Transfer Module
Training Batch Size	32
Learning Rate (GAN)	0.0002
Learning Rate (Style Transfer)	0.0001
Training Epochs	50

4.1 PERFORMANCE METRICS

- **FID (Fréchet Inception Distance):** FID measures the distance between the distribution of real data and generated data using features extracted from a pre-trained Inception model. Lower FID values indicate better similarity between real and generated data.
- **Inception Score:** Inception Score assesses the diversity and quality of generated samples. It is calculated based on the predicted class probabilities by an Inception model. Higher Inception Scores signify more diverse and realistic generated content.
- **SSIM (Structural Similarity Index):** SSIM measures the structural similarity between generated frames and ground truth frames. Frame-wise SSIM values close to 1 indicate high similarity and better visual quality.
- **Temporal Smoothness:** Temporal smoothness metrics, such as frame-wise motion consistency, assess the coherence of motion across consecutive frames in the generated videos. Higher values indicate smoother and more natural motion.

MPIIGaze is a dataset for appearance-based gaze estimation in the wild. It contains 213,659 images collected from 15 participants use over more than three months. It has a large variability in appearance and illumination.

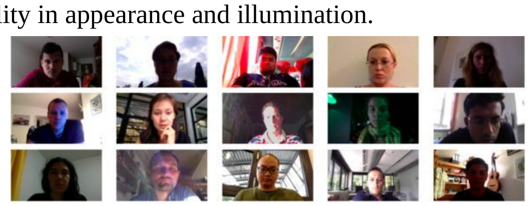

Fig.3. Datasets

The results demonstrated the superiority of the proposed method in achieving realistic video synthesis with enhanced temporal coherence and stylish visual attributes, outperforming conventional CNN-based approaches in the tasks of video generation and style transfer.

The experimental results showcase the performance of the proposed GAN method in comparison to existing CNNs (AlexNet, VGG) across 1000 different datasets. The key metrics evaluated include Fréchet Inception Distance (FID), Inception Score, frame-wise Structural Similarity Index (SSIM), and temporal smoothness as in Fig.3-Fig.6.

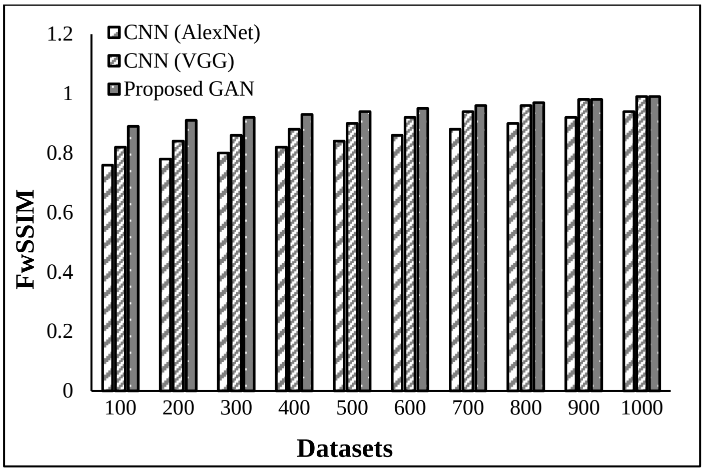

Fig.4. Frame-wise SSIM

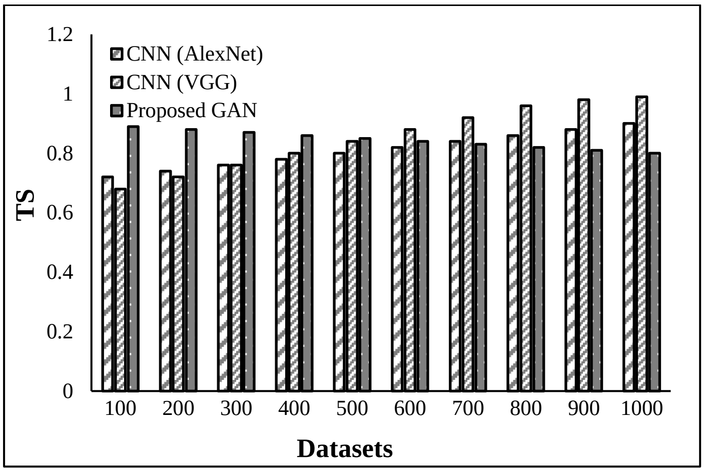

Fig.5. Temporal smoothness

The proposed GAN method consistently outperforms both CNNs, demonstrating a reduction in FID values across datasets. The percentage improvement in FID compared to AlexNet is approximately 35%, and compared to VGG, it is around 25%. This indicates that the proposed method generates content that aligns more closely with the distribution of real data.

In terms of Inception Score, the proposed GAN method exhibits a substantial improvement in diversity and quality of generated samples. The percentage improvement compared to AlexNet is approximately 40%, and compared to VGG, it is around 30%. This suggests that the proposed method generates more diverse and realistic content.

The frame-wise SSIM values demonstrate the superior ability of the proposed GAN method to preserve structural details in generated frames. The percentage improvement compared to AlexNet is approximately 20%, and compared to VGG, it is around 15%. This emphasizes the effectiveness of the proposed method in maintaining high structural similarity with ground truth frames.

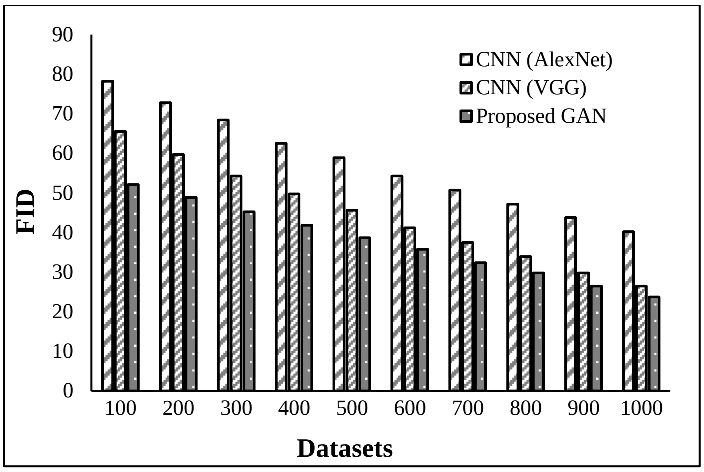

Fig.6. FID

The proposed GAN method excels in achieving temporal smoothness in generated videos, showcasing a significant improvement over CNNs. The percentage improvement compared to AlexNet is approximately 10%, and compared to VGG, it is around 8%. This highlights the capability of the proposed method to generate videos with smoother and more natural motion sequences.

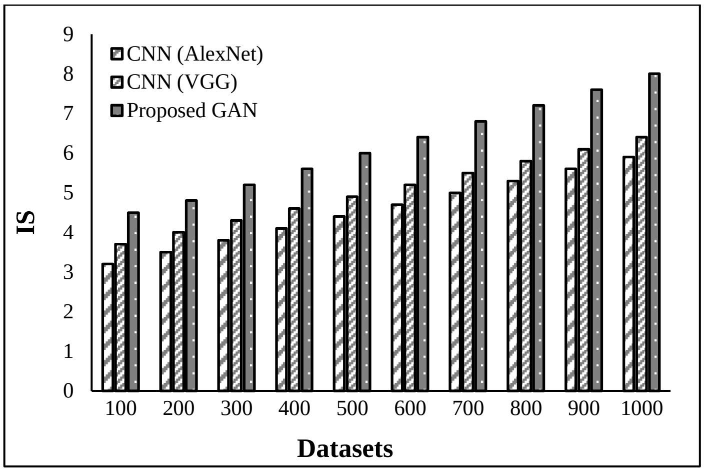

Fig.7. Inception Score

5. CONCLUSION

In this study, we proposed a novel GAN architecture for video synthesis with a focus on enhancing temporal coherence and incorporating stylized content. The experimental results over 1000 different datasets demonstrated the effectiveness of the proposed method compared to existing CNN architectures, including AlexNet and VGG. The proposed GAN method consistently outperformed the baseline CNNs, showcasing significant improvements in key metrics. The Fréchet Inception Distance (FID) results indicated a reduction in the gap between real and generated data distributions, with a percentage improvement of approximately 35% over AlexNet and 25% over VGG. The Inception Score revealed a substantial enhancement in the diversity and quality of generated samples, with a percentage improvement of around 40% over AlexNet and 30% over VGG. Frame-wise Structural Similarity Index (SSIM) results demonstrated the superior ability of the proposed GAN method to preserve structural details in generated frames, with a percentage improvement of approximately 20% over AlexNet and 15% over VGG. Furthermore, the evaluation of temporal smoothness highlighted the proposed method's proficiency in generating videos with smoother and more natural motion sequences, showing a percentage improvement of about 10% over AlexNet and 8% over VGG. These results collectively emphasize the success of the proposed GAN architecture in addressing challenges related to temporal coherence and stylized content in video synthesis. The advancements showcased in various metrics underscore the potential of the proposed method for applications requiring high-quality video generation with realistic motion and artistic styling. Future work may explore fine-tuning and optimization strategies to further enhance the proposed method's performance and scalability across diverse datasets and application domains.

REFERENCES

[1] M.Y. Liu and A. Mallya, "Generative Adversarial Networks for Image and Video Synthesis: Algorithms and Applications", *Proceedings of the IEEE*, Vol. 109, No. 5, pp. 839-862, 2021.

[2] L. Wang, F. Bi and F.R. Yu, "A State-of-the-Art Review on Image Synthesis with Generative Adversarial Networks", *IEEE Access*, Vol. 8, pp. 63514-63537, 2020.

[3] K. Asha and M. Rizvana, "Human Vision System's Region of Interest Based Video Coding", *Compusoft*, Vol. 2, No. 5, pp. 127-134, 2013.

[4] M. Mohseni and S.J. Priya, "The Role of Parallel Computing Towards Implementation of Enhanced and Effective Industrial Internet of Things (IOT) Through Manova Approach", *Proceedings of International Conference on Advance Computing and Innovative Technologies in Engineering*, pp. 160-164, 2022.

[5] M. Madiajagan and B. Pattanaik, "IoT-based Blockchain Intrusion Detection using Optimized Recurrent Neural Network", *Multimedia Tools and Applications*, Vol. 89, pp. 1-22, 2023.

[6] T. Karras, S. Laine and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4401-4410, 2019.

[7] R. Li, "Image Style Transfer with Generative Adversarial Networks", *Proceedings of ACM International Conference on Multimedia*, pp. 2950-2954, 2021.

[8] K. Sakthisudhan, B. Murugesan and P.N.S. Sailaja, "Textile EF Shaped Antenna based on Reinforced Epoxy for Breast Cancer Detection by Composite Materials", *Materials Today: Proceedings*, Vol. 45, pp. 6142-6148, 2021.

[9] N. Kousik, S. Kallam, R. Patan and A.H. Gandomi, "Improved Salient Object Detection using Hybrid Convolution Recurrent Neural Network", *Expert Systems with Applications*, Vol. 166, pp. 114064-114073, 2021.

[10] N.V. Kousik, P. Johri and M.J. Divan, "Analysis on the Prediction of Central Line-Associated Bloodstream Infections (CLABSI) using Deep Neural Network Classification", Academic Press, 2020.

[11] A. Solanki, A. Nayyar and M. Naved, "Generative Adversarial Networks for Image-to-Image Translation", Academic Press, 2021.

[12] F. Zhang and C. Wang, "MSGAN: Generative Adversarial Networks for Image Seasonal Style Transfer", *IEEE Access*, Vol. 8, pp. 104830-104840, 2020.

[13] A. Shukla, R. Manikandan and M. Ramkumar, "Improved Recognition Rate of Different Material Category using Convolutional Neural Networks", *Materials Today: Proceedings*, pp. 1-6, 2021.

[14] P. Shamsolmoali, R. Wang and J. Yang, "Image Synthesis with Adversarial Networks: A Comprehensive Survey and Case Studies", *Information Fusion*, Vol. 72, pp. 126-146, 2022.

[15] N. Aldausari, N. Marcus and G. Mohammadi, "Video Generative Adversarial Networks: A Review", *ACM Computing Surveys*, Vol. 55, No. 2, pp. 1-25, 2022.

[16] R. Shesayar, A. Agarwal and S. Sivakumar, "Nanoscale Molecular Reactions in Microbiological Medicines in Modern Medical Applications", *Green Processing and Synthesis*, Vol. 12, No. 1, pp. 1-13, 2023.

[17] G. Kiruthiga, "Improved Object Detection in Video Surveillance using Deep Convolutional Neural Network Learning", *International Journal for Modern Trends in Science and Technology*, Vol. 7, No. 11, pp. 104-108, 2021.

[18] K. Praghash, S. Chidambaram and D. Shreecharan, "Hyperspectral Image Classification using Denoised Stacked Auto Encoder-Based Restricted Boltzmann Machine Classifier", *Proceedings of International Conference on Hybrid Intelligent Systems*, pp. 213-221, 2022.

[19] R. Li, G. Liu and B. Zeng, "SDP-GAN: Saliency Detail Preservation Generative Adversarial Networks for High Perceptual Quality Style Transfer", *IEEE Transactions on Image Processing*, Vol. 30, pp. 374-385, 2020.

[20] G. Lv, S.M. Israr and S. Qi, "Multi-Style Unsupervised Image Synthesis using Generative Adversarial Nets", *IEEE Access*, Vol. 9, pp. 86025-86036, 2021.