Proceedings of the International Conference on Sustainable Expert Systems (ICSES-2024)
IEEE Xplore Part Number: CFP24VS6-ART; ISBN: 979-8-3315-4036-4

Artistic Fusion: AI Powered Artistry for Story Boarding

Bhuvaneshwaran A
Department of AI&DS
Kumaraguru College of Technology
Coimbatore, India
bhuvanesh0312@gmail.com

Kavya P
Department of AI&DS
Kumaraguru College of Technology
Coimbatore, India
kavyaprabahar@gmail.com

Dhaksana N
Department of AI&DS
Kumaraguru College of Technology
Coimbatore, India
dhaksananataraj@gmail.com

Amuthaguka D
Department of Computer Applications
Kalasalingam Academy of Research and Education
Krishnankoil, India
d.amuthaguka@klu.ac.in

Eraianbu R
Department of AI&DS
Kumaraguru College of Technology
Coimbatore, India
eraianbu873@gmail.com

Agilan S
Department of AI&DS
Kumaraguru College of Technology
Coimbatore, India
agilan.s.190705@gmail.com

2024 4th International Conference on Sustainable Expert Systems (ICSES) | 979-8-3315-4036-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICSES63445.2024.10763187

***Abstract*— Storyboarding is a crucial pre-production step in filmmaking, facilitating the visualization and planning of scenes. However, traditional methods have obstacles such as manual labour, artistic skill requirements, and limitations in dynamic scene representation. This paper presents an innovative AI-powered storyboard generator leveraging Stable Diffusion, implemented in PyTorch. The research addresses obstacles in traditional methods by automating the storyboard creation process, enhancing visual realism, and promoting collaborative filmmaking workflows. The AI-powered system integrates advanced generative techniques to simulate complex visual scenarios, from lighting variations to dynamic camera movements, ensuring a more accurate preview of the final cinematic experience. By utilizing text prompts, the model interprets narrative elements and mood, aligning generated visuals with creative intent. The integration of Stable Diffusion model allows for controlled noise introduction, enhancing realism and immersion in storyboarded scenes. Further, an average improved accuracy of 82% is witnessed when compared to existing methodologies like Dreambooth and LoRA. Key features include a user-friendly interface for intuitive operation, compatibility with diverse media types, and iterative refinement capabilities to fine-tune storyboard quality. Evaluation metrics such as Structural Similarity Index Measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR) demonstrate the system's efficacy in producing high-quality visuals that meet filmmakers' expectations.**

Keywords: Storyboard generator, Stable Diffusion, PyTorch, Structural Similarity Index Measure (SSIM), Visual realism, Peak Signal-to-Noise Ratio (PSNR).

I. INTRODUCTION

The film industry relies heavily on storyboarding to plan and visualize scenes before they're filmed. However, traditional storyboarding faces challenges like time-consuming manual creation and limited dynamic representation, leading to inefficiencies and delays. The objective is develop an AI-powered storyboard generator using Stable Diffusion and advanced generative techniques to overcome traditional storyboarding challenges by automating the creation process, enhancing visual realism, and supporting collaborative filmmaking workflows. This study aims to addresses these issues with an AI-powered storyboard generator using Stable Diffusion Models. By leveraging PyTorch, it automates the process, overcoming manual labour and artistic skill requirements. Additionally, it introduces dynamic scene representations, promoting collaboration and seamlessly integrating into existing workflows.

The study aims to develop a robust AI-powered Storyboard Generation System that streamlines the labour-intensive process while enhancing visual realism. It is crucial to ensure compatibility with established workflows and adaptability to diverse media types, ensuring widespread utility in the filmmaking landscape. The user-friendly interface nature fosters accessibility, facilitating widespread adoption and transforming the storyboard creation process.

Storyboarding faces practical issues such as being time-consuming, requiring artistic skill, and struggling with limited detail and flexibility. These challenges are addressed by current technologies like digital storyboarding tools and software that allow for easy revisions, integration with other production tools, and interactive features, enhancing efficiency and communication.

By integrating stable diffusion model and adjusting hyperparameters strategically, this system expedites visualizations and fosters collaboration. It bridges the gap between creative vision and realization on screen, setting the stage for a more streamlined, collaborative, and visually compelling future in the film industry. Through successful implementation, this study aims to make inefficiencies and delays relics of the past, paving the way for innovation and efficiency in cinematic storytelling.

II. LITERATURE SURVEY

The research paper titled "Generative Adversarial Networks (GANs) for Medical Image Analysis" by Salome Kazeminia et al. explores various GAN approaches for medical applications, highlighting their capability to enhance medical image synthesis, segmentation, reconstruction, detection, de-noising, and classification. The paper discusses how GANs, including Conditional GANs (cGANs) for image-to-image translation, Variational Autoencoders (VAEs) combined with GANs for improved image quality, CycleGAN for unpaired image translation, and StyleGAN for high-resolution image generation, effectively simulate realistic images even with limited labeled data. Despite these advantages, GANs face several drawbacks such as sensitivity to hyperparameters, which can result in unstable convergence, and further it generates similar samples, which reduces their ability to generalize well to diverse scenarios [1].

The research paper titled "SUGAN: A Stable U-Net Based Generative Adversarial Network" by Shijie Cheng et al focuses in increasing the stability and quality of GAN generated images. The approach makes use of SUGAN (Stable U-Net Based GAN) to maintain the trade-off between the quality of generated images and training stability. A gradient normalization module is introduced to the discriminator of U-Net GAN. This module effectively reduces gradient magnitudes, thereby handles the issues of gradient instability and overfitting. To address the issue of blurred edges of the generated images, a modified residual network is used in the generator. This modification enhances its ability to capture image details, leading to higher-definition generated images. The major issues that persist in SUGAN are potential trade-offs between training stability and output diversity, and sensitivity to hyperparameter tuning. Additionally, its enhancements may not generalize well across different datasets or types of GAN architectures [2].

The research paper titled "Neural Storyboard Artist: Visualizing Stories with Coherent Image Sequences" by Shizhe Chen et al. proposes a novel framework called CADM (Coherent Action-Driven Model), which uses neural networks to automate the storyboarding process. Using a deep learning model, it first converts written descriptions into a series of visual scenes. It then creates precise storyboard frames using a sequence-to-sequence architecture to ensure smooth transitions between scenes. Although this approach enhances efficiency by reducing manual labor and artistic skill requirements, it still has limitations. Specifically, it struggles with dynamically simulating complex visual scenarios, such as varied lighting conditions and dynamic camera movements, which are crucial for a more realistic and immersive storyboard [3].

The research paper titled " Storyboarding the Virtuality: Methods and Best Practices to Depict Scenes and Interactive Stories in Virtual and Mixed Reality" by Thomas Brett Talbot, Katherine Elizabeth Thiry, and Michael Jenkins proposes methods to adapt storyboarding techniques for Virtual Reality (VR) and Mixed Reality (MR) by maintaining spatial efficiency while enabling multiple viewpoints in a single timestep. It discusses the use of 3D storyboarding to represent depth and spatial relationships, as well as techniques to combine 2D and point-of-view (POV) depictions to balance dimensionality and sequential action. It also suggests using color-coding to differentiate between real-world and virtual elements, aiding in the visualization of how these elements interact within the storyboard. It faces several disadvantages such as the complexity involved in the visualizing immersive 3D environments which often results in less clear frames. It also has trouble capturing user movements and dynamic interactions. Moreover, putting multiple points of view on a single storyboard page may make it harder to read and organize. The fact that this process takes time is more significant [4].

The research paper titled "Story Diffusion: How to Support UX Storyboarding with Generative-AI" by Zhaohui Liang et al proposes an approach to enhance UX storyboarding through Generative AI by integrating text-to-text and text-to-image models. It allows designers to input textual descriptions and then it is converted into visual storyboard frames, furthermore, automating and streamlining the creation process. The combination of LLMs (Large Language Models) and text-to-image generation models enables quick translation of ideas into visuals. It has several challenges such as, it might have trouble creating abstract or complicated sceneries, which could result in biases or inconsistent graphics. Furthermore, the absence of interactive feedback mechanisms in the system may hinder the process of fine-tuning and personalizing created graphics. It may not fully capture nuanced design requirements or user intent. These limitations highlight areas for further research and development to enhance the system's reliability and versatility in supporting diverse design workflows [5].

III. METHODOLOGY

The Stable Diffusion model operates through two main components: the Text Representation Generator and the Image Representation Refiner. This stable diffusion model accepts a text prompt from the user.

A. Text Representation Generator:

The Text Representation Generator converts a text prompt into a vector representation, which guides the image generation process. The CLIP (Contrastive Language-Image Pretraining) model consists of a Text Encoder and an Image Encoder. The CLIP model takes the text prompt as input. The text encoder component includes:

Tokenizer: It breaks down sentences into smaller units called tokens using Byte Pair Encoding (BPE), which merges the most frequent pairs of characters or bytes until a defined vocabulary size is achieved. Each token t_i is then mapped to a unique index i in the vocabulary. This process enables the model to efficiently handle text by converting it into a format that is easier to process and understand, facilitating the handling of out-of-vocabulary words and ensuring effective encoding.

***Text Encoder**:* The Text Encoder transforms the tokenized sequences into numerical representations using an embedding layer. This embedding layer maps each token to a dense vector in a high-dimensional space, capturing the semantic meaning of the text. The embedding layer uses a matrix E where each row E_i corresponds to the dense vector representation of token t_i .

For a token sequence $t = [t_1, t_2, .., t_n]$ the embeddings can be represented as: $E_t = [E_{t_1}, E_{t_2}, .., E_{t_n}]$ where $E_{t_i} \in R^d$ is the embedding vector of token t_i in a d-dimensional space. This uses Transformer-based architectures (like BERT or GPT) to convert tokenized sequences into embeddings.

Transformer Architecture:

i) Self-Attention Mechanism: Computes attention scores and weighted sums of token embeddings to capture contextual information. For a token t_i in a sequence, the self-attention mechanism computes:

$$\text{Attention }(t_i)=\text{SoftMax}\left(\frac{Q_i K^T}{\sqrt{d_K}}\right)$$

where Q_i, K, V are the query, key, and value matrices derived from the token embeddings, and d_k is the dimension of the key vectors.

ii) Attention Scores: The attention scores for a pair of tokens t_i, t_j are computed as:

$$\text{score}(t_i, t_j) = \left(\frac{Q_i K^T}{\sqrt{d_K}}\right)$$

This is crucial for generating images that accurately reflect the meaning of the entire text.

iii) Adding Positional Information: Since Transformers do not inherently understand token order, positional encodings P are added to token embeddings to incorporate positional information:

$$Embedding_{Pos} = Embedding + P$$

Each layer of the Transformer includes a feed-forward network that applies linear transformations and activation functions. The text embeddings (after going through the Transformer encoder) are used to condition the image generation process. This conditioning helps the model to generate images that are semantically aligned with the given text prompt.

Text Embeddings*:* The output of the Text Encoder is a sequence of vectors known as text embeddings. These embeddings encapsulate the semantic information of the text prompt and serve as a guide for the image generation process. The Image Encoder (often a CNN or Vision Transformer) extracts features from images, while the Text Encoder processes text. The tokenized text is processed by the Text Encoder to produce a vector representation. Simultaneously, the Image Encoder (which can be a CNN or Vision Transformer) transforms the images into high-dimensional vector representations. The cosine similarity between the text vector e_t and the image vector e_I is computed to ensure alignment between the text and image:

$$\text{Cosine similarity } (e_t, e_I) = \frac{e_t \cdot e_I}{(\|e_t\|)(\|e_I\|)}$$

This similarity metric guides the generation process to produce images that are semantically consistent with the text prompt.

B. Image Representation Refiner:

The Image Representation Refiner takes random noise and transforms it into a high-resolution image over multiple timesteps.

Latent Space: The image generation begins in a latent space, a compressed representation of the image initialized with random noise sampled from a Gaussian distribution.

Forward Diffusion Process: Gaussian noise is progressively added to the image over a series of timesteps. The noise added at each timestep follows a Gaussian distribution with mean 0 and a certain variance. It is mathematically represented as:

$$x_{t+1} = x_t + \sqrt{\beta_t} \cdot \epsilon_t$$

Where x_t is the image at timestep t, β_t is the noise variance at timestep t, and ϵ_t Gaussian noise sampled from $N(0, \sigma)$.The addition of Gaussian noise is controlled by a variance schedule, which dictates how much noise is added at each timestep. This helps in learning a diverse set of image representations.

Backward diffusion process:

U-Net: The U-Net (a neural network) used in Stable Diffusion consists of an encoder-decoder structure with skip connections. The conditioning layers integrate text embeddings into the model, allowing the U-Net to generate images that are conditioned on the text prompt.

A U-Net is used to predict and subtract the noise from the noisy image, gradually reconstructing the original image. The denoised image at each timestep t is given by:

$$x_{t+1} = x_t - \sqrt{\beta_t} \cdot \epsilon_t$$

where x_t is the current noisy image and ϵt is the predicted noise. The U-Net architecture takes the noisy image x_t and the text embeddings as inputs and predicts the noise ϵ_t .

U-Net Architecture: The initial input to the U-Net is the noisy image x, at each timestep of the backward diffusion process. This image has undergone the forward diffusion process and contains added noise.

i) Encoder (Down sampling Path):

Convolutional Blocks*:* Each block consists of two or more convolutional layers, followed by a non-linear activation function (e.g., ReLU). These blocks extract high-level features from the noisy image.

Pooling Layers: Max pooling or average pooling is used to down sample the feature maps, reducing spatial dimensions while preserving important features.

Down sampling: The encoder progressively reduces the spatial dimensions of the feature maps while increasing the number of feature channels. This allows the model to capture and abstract complex features at different levels of granularity.

ii) Bottleneck: Latent Representation: At the bottleneck, the feature maps are the most compressed. Convolutional operations are applied to refine the features in this latent space.

iii) Decoder (Up sampling Path)*:*

Up sampling Blocks: Each block upsamples the feature maps from the latent space using transposed convolutions (deconvolutions) or other upsampling techniques. Upsampling restores the spatial dimensions of the feature maps to match the original image size and helps in generating detailed output.

Skip Connections: Skip connections provide a direct path for features to pass from the encoder to the decoder, ensuring that detailed spatial information is preserved.

Concatenation with Skip Connections: The upsampled feature maps are concatenated with corresponding feature maps from the encoder (skip connections). This helps retain fine-grained details.

Conditioning Layers: It is enhanced with conditioning layers that integrate text embeddings. These layers adjust the skip connections within the U-Net to combine text embeddings with image features at each stage. This integration ensures that the text prompt influences the reconstruction process throughout, guiding the model to produce images that align with the described content.

Final Output Layer**:** A final convolutional layer is applied to map the feature maps back to the desired output image size and number of channels (e.g., RGB image).

Final Image Generation*:* After iteratively refining the image through the backward diffusion process, the final representation is upscaled to produce a high-resolution image. This final image is expected to closely match the content described by the text prompt, both in visual features and semantic meaning.

C. Evaluation metrics:

SSIM (Structural Similarity Index Measure):

SSIM is used to measure the similarity between two images, focusing on changes in structural information. It compares luminance, contrast, and structure between the images.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y+C_1)(2\sigma_x\sigma_y+C2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)}$$

PSNR (Peak Signal-to-Noise Ratio):

PSNR measures the peak error between two images. It is commonly used to evaluate image quality and is especially useful for images with high bit depth.

$$\text{PSNR}=10\, log_{10}\left(\frac{\mathbb{R}^2}{MSE}\right)$$

where, R is the maximum possible pixel value MSE is the Mean Squared Error between the original image and the generated image.

D. Fine Tuning method:

Progressive fine tuning:

Initially, all layers in the model may be frozen, meaning their parameters are not updated during training. As training progresses, more layers are progressively unfrozen to allow their parameters to be updated. This approach helps the model first learn fundamental features with a limited set of layers and then gradually adapt to more complex features. Layers are frozen by setting their gradient updates to zero. At a designated epoch, gradients for these layers are enabled, allowing their parameters to be updated. This involves adjusting the training process to include these layers in gradient computation. This strategy can stabilize training and enhance the performance of the fine-tuned model. This method of fine-tuning enhanced the proposed model's accuracy to 85% and precision to 83%.

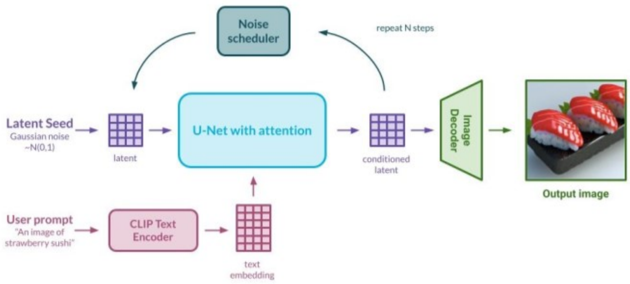


Fig. 1. Stable Diffusion Architecture

IV. WORKFLOW

The workflow begins with collecting textual data and tokens from Hugging Face, which are then preprocessed and augmented. The CLIP encoder translates textual prompts into latent representations, guiding the storyboard generation process.

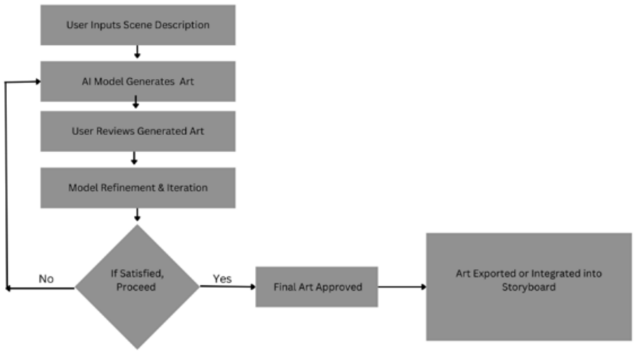


Fig. 2. Artistic fusion flow diagram

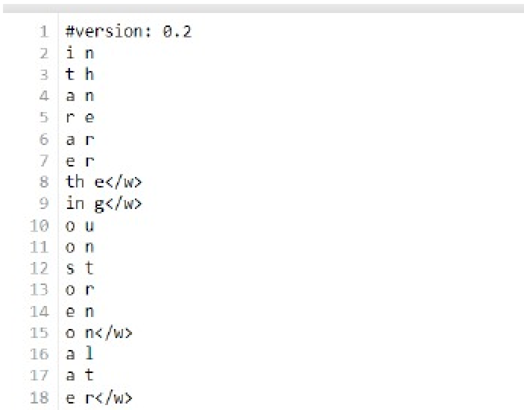
Stable diffusion models generate noisy images, denoised using the UNET architecture with text prompts. Iterative refinement enhances storyboard quality. Evaluation assesses the generated storyboards using metrics and user feedback, ensuring alignment with filmmakers' creative visions. This systematic approach streamlines the storyboard creation process, fostering collaboration and efficiency in filmmaking.

V. IMPLEMENTATION

A. Data Collection and Preprocessing:

By collecting textual data and tokens from Hugging Face along with relevant metadata, the data collection process ensures the availability of diverse and high-quality resources for training and evaluating the AI-powered storyboard generation system. This comprehensive approach enhances the system's capability to generate accurate, contextually relevant, and visually compelling storyboards for the film industry.

1) Text Encoding with CLIP Encoder: The CLIP encoder, trained on a vast text and image dataset, encodes text prompts into latent representations, capturing the nuanced meanings of the prompts for subsequent image generation.



```
#version: 0.2
i n
t h
a n
r e
a r
e r
th e</w>
in g</w>
o u
o n
s t
o r
e n
o n</w>
a l
a t
e r</w>
```

Fig. 3. Sample data and tokens

2) Noisy Image Generation using Diffusion Model: Employing the diffusion model, a generative approach, a noisy image is generated. The diffusion step hyperparameter controls the gradual denoising process, influencing the level of noise within the image[10].

3)Image Denoising with Unet Architecture: The Unet architecture, a specialized CNN, plays a crucial role in denoising the generated image. Its U-shaped structure, consisting of contracting and expanding branches, effectively combines up sampling and down sampling while utilizing skip connections for multi-scale feature learning.

4)Text-Guided Denoising Process: The text prompt functions as a guide for denoising. Up sampled feature maps from the contracting branch are integrated with skip connections, and the text prompt assists in refining the image, ensuring consistency between the generated visuals and the intended meaning[11].

5)Iterative Refinement for Quality: Steps 2 and 3 are iteratively repeated until the desired image quality is achieved. The number of diffusion steps required is determined by the image complexity and the realism level sought[14].

6)Output of the Generated Image: The final step involves outputting the generated image, which can be saved to a file or displayed on a screen. This marks the culmination of the image generation process guided by text prompts and the collaborative efforts of the CLIP encoder, diffusion model, and Unet architecture[15].

B. StoryboardIQ Portal

A web application has been developed using HTML, CSS, and JavaScript for the frontend and Django for the backend to facilitate the generation and display of images based on user prompts. The frontend ensures an interactive and responsive user interface, allowing users to input text prompts and receive generated images in real-time.

Django, a high-level Python web framework, was chosen for the backend due to its robustness, scalability, and ease of maintenance. Django's comprehensive features and architecture support the development of a reliable and maintainable application, making it an ideal choice for handling backend processes in our web application [19].

A key component of the application is the integration of the diffusion model. This model is designed to manage complex text-to-image tasks and leverages advanced machine learning techniques to produce high-quality images. The diffusion model's architecture is optimized to interpret user prompts accurately and generate corresponding images efficiently.

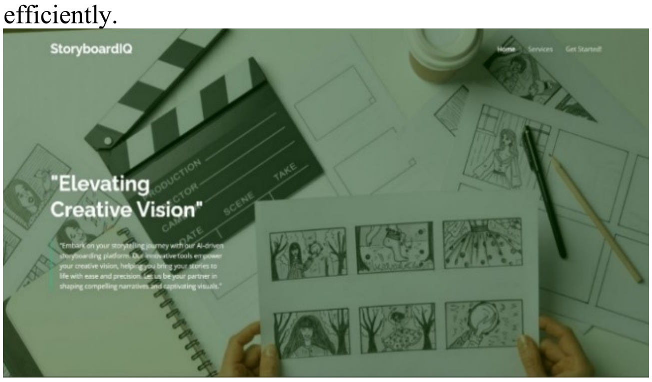


Fig. 4. Home Page

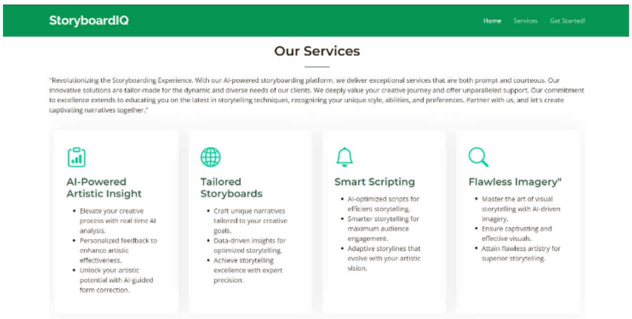


Fig. 5. Services provided.

VI. Results and Discussions

Comparative analysis:

In the landscape of creative project management, Krock.io and the AI-enhanced storyboarding both address different facets of the creative process. Krock.io excels in collaborative video review and project organization, offering centralized management and visual feedback tools. Its cloud-based platform streamlines workflows for creative teams, enhancing productivity and quality control.

In contrast, the AI-powered storyboarding tool revolutionizes the pre-production phase of filmmaking. By utilizing stable diffusion models, it automates the traditionally labor-intensive storyboarding process, addressing challenges in manual creation, artistic skill requirements, and dynamic scene representation. The comparative analysis underscores that the AI-enhanced storyboarding project stands out for its emphasis on efficiency and innovation in creative visualization. Unlike traditional storyboarding methods, which often rely on manual input and can be time-consuming, this AI-driven approach leverages advanced algorithms to streamline the process and enhance creativity.

Metrics	Accuracy	Precision	Recall	F1-score	Inception Score
Baseline (Single task)	65%	68%	69%	69%	4.5
Dreambooth	64%	68%	70%	72%	4.8
LoRA	75%	73%	77%	79%	5.0
Proposed Method	82%	85%	83%	84%	5.5

Table VI.1 *Comparative analysis with other models*

The results indicate a significant leap forward in efficiency, creativity, and collaboration within the film industry. The success of the AI-powered storyboard generator, enhanced by insights from stable diffusion models and strategic hyperparameter adjustments, promises a transformative impact on storytelling in visual media. This project exemplifies the boundless possibilities when advanced AI techniques are seamlessly integrated into the cinematic storytelling process.

The structural similarity index measure (SSIM) is a technique used to assess the perceived quality of digital images and videos, including television and cinematic pictures. Additionally, it serves as a metric for comparing the similarity between two images.

The SSIM Scores, graph of sample prompts given from the user to the model is:

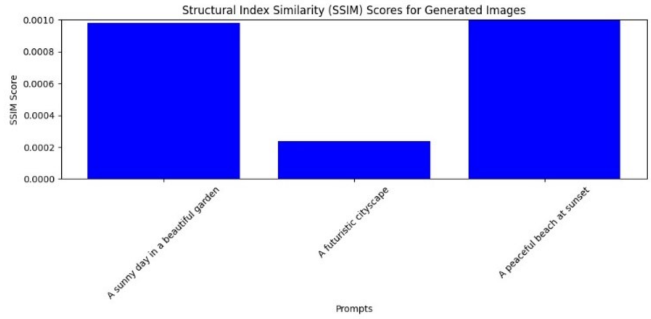


Fig. 6. SSIM graph

Peak Signal-to-Noise Ratio (PSNR) quantifies the ratio between the maximum possible power of a signal and the power of noise that affects its representation. It is a widely used metric for assessing image quality, particularly when comparing the performance of different compression codecs. PSNR is typically used alongside other image quality assessment tools like Structural Similarity Index (SSIM), Mean Square Error (MSE), and Universal Image Quality Index (UIQI) [13].

The SSIM Scores and graph of the sample prompts given from the user to the model is:

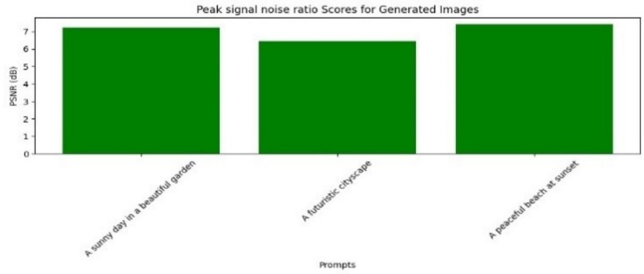


Fig. 7. PSNR Graph

VII. CONCLUSION

In conclusion, this research emphasizes the benefits of integrating Diffusion Models into traditional film storyboarding techniques. By utilizing this technology, the study. Enhances visualization processes, surpassing the constraints of approaches and generating more engaging and diverse scenes. The use of Stable Diffusion Models enhances image realism. Promotes collaboration, through improvements while fine tuning hyperparameters enhances model performance and stability. Through an analysis it is evident that adopting AI powered storyboarding not only boosts efficiency and creativity in comparison to traditional methods but also establishes a new benchmark in pre-production workflows. The findings showcase progress in accuracy, precision and overall efficacy representing an advancement in merging AI with cinematic narrative creation.

VIII. REFERENCES

[1] Salome Kazeminia et al, " GAN for medical image analysis", *Artificial Intelligence in Medicine,* vol 109*,* 2020.

[2] Shijie Cheng et al, "SUGAN: A Stable U-Net Based Generative Adversarial Network", *Journal of Sensors*, vol. 23, 2023.

[3] Shizhe Chen et al, "Neural Storyboard Artist: Visualizing Stories with Coherent Image Sequences", *arXiv preprint arXiv:1911.10460,* 2019.

[4] Thomas Brett Talbot, Katherine Elizabeth Thiry, Michael Jenkins, "Storyboarding the Virtuality: Methods and Best Practices to Depict Scenes and Interactive Stories in Virtual and Mixed Reality", *Advances in Intelligent Systems and Computing,* vol. 1217, 2020.

[5] Zhaohui Liang et al, "StoryDiffusion: How to Support UX Storyboarding with Generative-AI", *arXiv preprint arXiv:2407.07672*, 2024.

[6] Federico Manuri, Andrea Sanna, Marco Scarzello, Francesco De Pace, "A Novel Approach to 3D Storyboarding", *Intelligent Technologies for Interactive Entertainmentl of Sensors*, vol. 360, 2024.

[7] Nitin Hariharan, Gerard Deepak, "SSAT: Scientific Storyboarding Framework Using Artificial Intelligence Techniques", *Data Science and Security*, vol. 922, 2024.

[8] S. Tan, D. Liu, M. Wong, "Generic Image Application Using GAN," *International Journal of Computer Vision*, vol. 42, no. 3, pp. 234-245, 2023.

[9] R. Sharma, L. Zhang, E. Kim, "Mobile-UNet GAN: A Single-Image Dehazing Model," *IEEE Transactions on Image Processing*, vol. 31, pp. 567-578, 2023.

[10] K. E. Hedges, "Artificial Intelligence (AI) Art Generators in the Architectural Design Curricula," in *2023 ASEE Annual Conference & Exposition*, June 2023.

[11] S. Jo, S. Shin, S.-W. Kim, "Interactive Storyboarding System Leveraging Large-Scale Pre-Trained Model," 2023.

[12] Tori S. Simenec et al "A Novel Blueprint Storyboarding Method Using Digitization for Efficient Cultural Adaptation of Prevention Programs to Serve Diverse Youth and Communities", *Prevention Science,* 2022.

[13] Shwetha Rajaram, Franziska Roesner, Michael Nebeling, "An Augmented Reality Storyboarding Tool for Character-Driven Analysis of Security & Privacy Concerns", 2023.

[14] S. Lee, B. Hoover, H. Strobelt, Z. J. Wang, S. Peng, A. Wright, ... & D. H. Chau, "Diffusion Explainer: Visual Explanation for Text-to-image Stable Diffusion," *arXiv preprint arXiv:2305.03509*, 2023.

[15] Aarthi, M., and A. Bhuvaneshwaran. "IoT based drainage and waste management monitoring and alert system for smart city." *Annals of the Romanian Society for Cell Biology (2021)*: *6641-6651*.

[16] Narasimha Mallikarjunan, K., A. Bhuvaneshwaran, K. Sundarakantham, and S. Mercy Shalinie. "DDAM: detecting DDoS attacks using machine learning approach." In *Computational Intelligence: Theories, Applications and Future Directions-Volume I: Icci-2017*, pp. 261-273. Singapore: Springer Singapore, 2018..

[17] Mallikarjunan, K. Narasimha, S. Mercy Shalinie, and A. Bhuvaneshwaran. "BAIT: behaviour aided intruder testimony technique for attacker intention prediction in business data handling." *International Journal of Business Intelligence and Data Mining* 14, no. 1-2 (2019): 177-198.

[18] Bhuvaneshwaran, A., and C. Deisy. "IoT Defect Prediction Using AI Framework: A Review." In *2023 Third International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS)*, pp. 376-383. IEEE, 2023.

[19] Sudharson, D., A. Bhuvaneshwaran, T. R. Kalaiarasan, and V. Sushmita. "A multimodal AI framework for hyper automation in industry 5.0." In *2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA)*, pp. 282-286. IEEE, 2023.

[20] Bhuvaneshwaran, A., P. Manickam, M. Ilayaraja, K. Sathesh Kumar, and K. Shankar. "Clustering Based Cybersecurity Model for Cloud Data." *Cybersecurity and Secure Information Systems: Challenges and Solutions in Smart Environments* (2019): 227-240.