iJRASET
International Journal For Research in
Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 14 **Issue:** VI **Month of publication:** June 2026

**DOI:** https://doi.org/10.22214/ijraset.2026.83610

# AdGenAI- A Unified Framework for Automated Advertisement Generation

Sumaya[1], B. Kranthi Kiran[2]
[1]Post Graduate Student, M. Tech, Department of Computer Science and Engineering, Jawaharlal Nehru Technological University, Hyderabad, India
[2]Professor, Department of Computer Science and Engineering, Jawaharlal Nehru Technological University, Hyderabad, India.

***Abstract: Modern marketing depends extensively on digital advertising, although producing high-quality advertising content often requires a lot of time and skill. The AI-Powered Hybrid Framework for Product Advertisement Generation proposed in this study automatically converts product photos into Advertisement visuals and promotional videos. The proposed framework combines a Flask-based web application, Stable Diffusion XL (SDXL) Inpainting, prompt enhancement, WAN 2.2 image-to-video production and image preprocessing into a single workflow. WAN 2.2 uses the visually appealing commercial images produced by SDXL to create dynamic promotional videos. The interactive platform provided by the Flask-based web application lets users create promotional videos and visuals for advertisements, upload product images, and manage the final output. The generated outputs are evaluated using CLIP Score, SSIM, LPIPS, HPS v2, and Temporal Consistency metrics. Experimental findings show that the system produces high-quality ad content while maintaining video integrity and product characteristics. The performance of the suggested technique for producing realistic and marketing-focused advertising content is further validated by comparative analysis. For AI-driven advertisement generation in digital marketing and e-commerce applications, the suggested system offers a scalable and effective solution.***
***Keywords: Product Advertisement Generation, Digital Advertising, Stable Diffusion XL, WAN 2.2, Prompt Enhancement, Image-to-Video synthesis, SSIM, LPIPS, CLIP Score, HPS v2, Temporal Consistency.***

## I. INTRODUCTION

Promoting products has become a vital aspect of contemporary digital marketing because it enhances product visibility, fosters customer engagement, and boosts brand awareness on online platforms. The requirement for attractive and appealing advertising images and promotional videos has grown significantly due to the rapid growth of e-commerce and social media marketing. In competitive digital markets, strong brand presence and consumer purchasing decisions are greatly influenced by high-quality advertising content. Automated content creation systems that can produce realistic images and videos for a range of commercial applications have been made possible by recent developments in computer vision and generative artificial intelligence [1]. Additionally, generative marketing strategies and intelligent advertisement generation frameworks have shown how AI-driven technology may produce visually appealing and advertising-focused content [2], [3].

Despite these improvements, there still exist a lot of limitations in the way of creating effective systems for generating advertisements. Product extraction, image editing, background design, text enhancement, and video production are just a few of the steps that are frequently needed in traditional advertisement creation workflows. These procedures need a great deal of manual intervention and creative skill, which lengthens manufacturing times and raises expenses. While realistic images and videos can be produced by new generative AI models, many current methods mainly concentrate on creating advertisement graphics or videos as separate tasks. Because of this, in order to produce marketing-ready commercial content, consumers usually rely on a variety of tools and complex workflows. Additionally, it is still difficult for automated advertisement creation systems to preserve product identification, visual quality, prompt alignment, and video consistency [4], [5].

An AI-Powered Hybrid Framework for Product Advertisement Generation is developed to automate the creation of promotional videos and Advertising visuals from a single product image in order to address such issues. Within a single workflow, the suggested system combines depict preprocessing, prompt enhancement, SDXL Inpainting for creating advertising images, WAN 2.2 for creating promotional videos, and a Flask-based web application. The framework seeks to provide attractive advertising content while minimizing manual work during the content generation process and maintaining crucial product characteristics.

The developed framework helps create scalable and effective advertisements for e-commerce and digital marketing applications. Semantic alignment, image quality, visual aesthetics, and video stability are assessed using CLIP Score, SSIM, LPIPS, HPS v2, and Temporal Consistency measures. The proposed architecture provides a feasible strategy for AI-driven advertisement content development by combining image generation, video synthesis, and interactive deployment into a single system.

## II. LITERATURE SURVEY

Generative artificial intelligence has recently attracted considerable interest in the realm of advertisement content creation because of its capability to produce realistic and visually striking images. RefAdGen, a high-fidelity framework for creating advertisement images that maintains product attributes while producing eye-catching advertising scenarios, was introduced by He et al. [1]. The study showed that generative AI is useful for designing commercial advertisements, however it did not address the creation of promotional videos and instead concentrated on image generation.

Approaches based on human feedback have also been investigated to enhance the quality of advertisements. In order to improve visual aesthetics and encourage alignment, Zhang et al. [2] presented a dependable framework for creating advertisement images that integrates human preference signals. The method lacked an integrated multimedia advertisement generating pipeline and mostly relied on feedback-driven optimization, despite improving the quality of advertisements.

Automated advertisement creation has further made significant use of deep learning techniques. An intelligent system for creating advertisements was presented by Rokhade et al. [3]. It makes use of deep learning models to automate the creation of advertising content. Although the framework mostly concentrated on creating advertisement images and did not include promotional video synthesis, the study showed how artificial intelligence may be utilized to create visually appealing ads.

Hartmann et al. [4] looked into the potential of generative models to provide excellent visual marketing content, which further studied the use of generative AI in marketing. Their results demonstrated how generative AI can increase marketing efficacy while lowering the amount of work required to create content. However, the study did not offer a comprehensive framework for automated commercial picture and video creation, instead concentrating primarily on the creation of visual marketing content.

Stable Video Diffusion, a diffusion-based framework that can produce realistic videos from image inputs, was proposed by Blattmann et al. [5]. The approach showed encouraging outcomes in temporal consistency and video synthesis. However, the framework lacked advertisement-focused image production techniques and was never developed with product advertisement generation in consideration.

Overall, research has demonstrated that generative AI is useful for creating visual marketing content, creating advertisement images, and synthesizing videos. However, the majority of methods focus on discrete tasks and lack an integrated framework that can produce promotional videos and advertisement images from a single product image. In addition, issues related to timely alignment, visual quality, product fidelity, and temporal consistency are still not adequately addressed. These drawbacks emphasize the necessity for a single, automated framework for creating advertisements that can effectively create multimedia content that is ready for marketing.

## III. METHODLOGY

### *A. System Overview*

The proposed system is a framework for automated product advertisement creation driven by artificial intelligence. In addition to performing images preprocessing, mask generation, prompt enhancement, advertisement image generation, and promotional video synthesis, the framework takes a product image as input. Image preprocessing techniques increase product visibility and produce the masks needed to create content.

In order to produce comprehensive advertising descriptions that raise the standard of produced outputs, prompt improvement is implemented out. SDXL Inpainting is used to create visually appealing images for advertisements while maintaining crucial aspects of the product. The WAN 2.2 image-to-video creation methodology then utilizes the created advertisement image to produce dynamic promotional videos. Image quality, semantic alignment, visual aesthetics, and video stability are assessed using CLIP Score, SSIM, LPIPS, HPS v2, and Temporal Consistency criteria. An effective and scalable solution for automatic advertisement production in digital marketing and e-commerce applications is provided by the integration of the entire framework into a Flask-based web application.

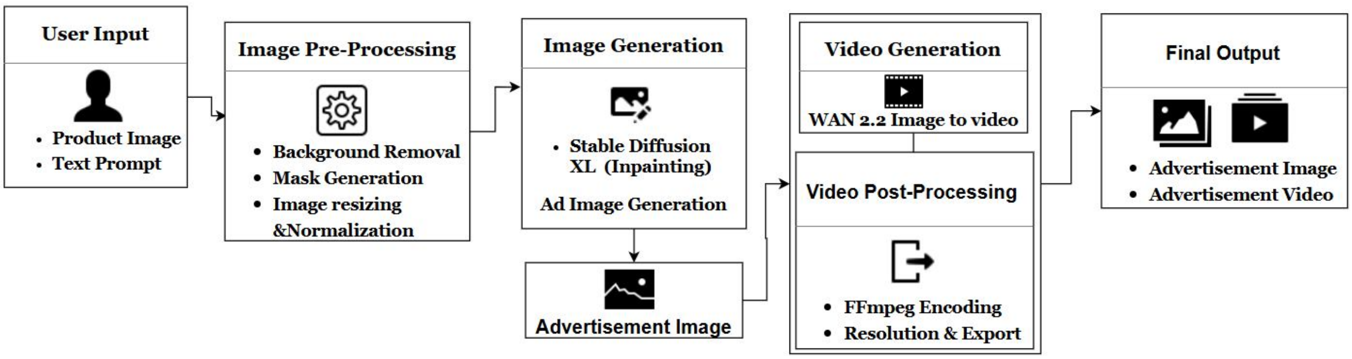



Fig. 1 System Architecture

Fig. 1 illustrates the overall architecture of the proposed framework for creating product advertisements. The system combines SDXL Inpainting, WAN 2.2 video production, image preprocessing, and a Flask-based web application for automatic ad creation.

### *B. Data Acquisition and Preprocessing*

In order to generate advertisements, the proposed structure takes two inputs: a product image and a text prompt. To enhance product visibility and maintain consistency with the content generation pipeline, the images uploaded perform background removal, mask generation, image resizing, and standardizing. Additionally, the Gemini Flash API is used to improve the user-provided prompt in order to produce thorough and marketing-oriented descriptions that are appropriate for creating advertisements.

### *C. Generative Advertisement Framework*

Stable Diffusion XL (SDXL) Inpainting is used to create advertisement images. The model is provided with the generated mask, the improved prompt, and the preprocessed product image. SDXL creates aesthetically pleasing advertising backgrounds while upholding the integrity of the product and key features. Realistic backgrounds and promotional aesthetics appropriate for digital marketing applications are included in the generated outputs.

### *D. Promotional Video Generation*

The WAN 2.2 image-to-video generation model receives the advertisement image produced by SDXL. By adding realistic motion and temporal transitions while preserving visual consistency between frames, WAN 2.2 creates dynamic promotional videos. Static ad images are converted into engaging marketing content with this approach.

### *E. Video Post-Processing and Output Generation*

FFmpeg is used to further process the produced videos for export, format conversion, and video encoding. The final outcome is a promotional video and a high-quality advertisement image that are shown via a Flask-based web application. The produced content can be viewed, downloaded, and managed by users for social media promotion, e-commerce, and digital marketing.

## IV. RESULTS AND DISCUSSION

### *A. Performance Evaluation Metrics*

The performance of the proposed advertisement generation framework is evaluated using:

CLIP Score: The semantic similarity between the created advertisement image and the matching text prompt is measured by the CLIP Score. Improved image-text alignment is indicated by higher values.

$$CLIP\ Score(I,T) = \frac{E_I \cdot E_T}{||E_I||\ ||E_T||} \quad (1)$$

Structural Similarity Index Measure (SSIM): The generated image and the input image's structural similarity is assessed by SSIM. Better preservation of the visual content and image structure is indicated by higher values.

$$SSIM(x,y) = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\sigma_{xy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \quad (2)$$

Learned Perceptual Image Patch Similarity (LPIPS): LPIPS requires deep feature representations to quantify perceptual similarity. Better image resemblance and perceived quality are indicated by lower values.

$$LPIPS(x,y) - \sum_{l} \frac{1}{H_l W_l} \sum_{hw} \left\| w_l \odot \left( \hat{x}^{l}_{hw} - \hat{y}^{l}_{hw} \right) \right\|_2^2 \tag{3}$$

Temporal Consistency (TC): In the produced promotional video, temporal consistency quantifies the visual coherence between consecutive frames. Smoother motion and better video stability are indicated by higher values.

$$TC = \frac{1}{N-1} \sum_{t=1}^{N-1} CosSim(F_t, F_{t+1}) \tag{4}$$

## *B. Quantitative Performance Analysis*

The performance of the proposed SDXL-based framework was compared with SD 1.5, ControlNet, and InstructPix2Pix (IP2P) using CLIP Score, SSIM, LPIPS, and HPS v2 metrics. The proposed SDXL Inpainting framework provides superior CLIP Score and HPS v2 values, suggesting increased semantic alignment and advertisement quality, according to the data shown in Table 1 and Fig.2. The system creates new advertisement backgrounds while maintaining the original product attributes, which explains why SSIM and LPIPS values are somewhat lower. Overall, the findings show how successful the suggested strategy is at producing realistic and marketing-focused advertising content.

TABLE I

MODEL COMPARISON

| Model | CLIP Score | SSIM | LPIPS | HPSv2 |
|---|---|---|---|---|
| SD 1.5 | 0.6214 | 0.4783 | 0.6521 | 0.7012 |
| ControlNet | 0.6732 | 0.4516 | 0.6984 | 0.6635 |
| IP2P | 0.4628 | 0.7415 | 0.5234 | 0.6241 |
| SDXL Inpainting (proposed) | 0.7943 | 0.6327 | 0.5816 | 0.7428 |

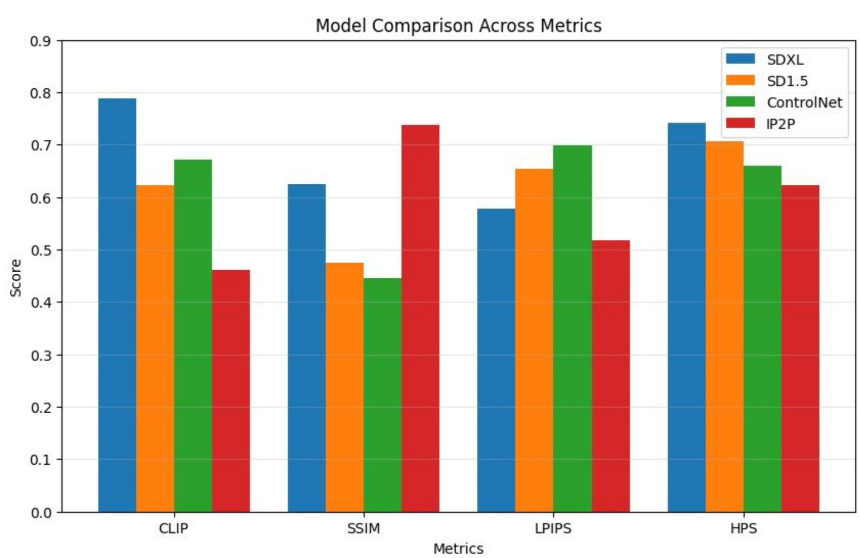



Fig. 2 Comparison of Generation Models across the metrics

## *C. Temporal Consistency Analysis*

The promotional videos created with WAN 2.2 were assessed for motion stability and visual coherence using temporal consistency. Higher values indicate smoother motion, less flickering, and better visual content preservation throughout the output video sequence. The metric quantifies the similarity between consecutive video frames

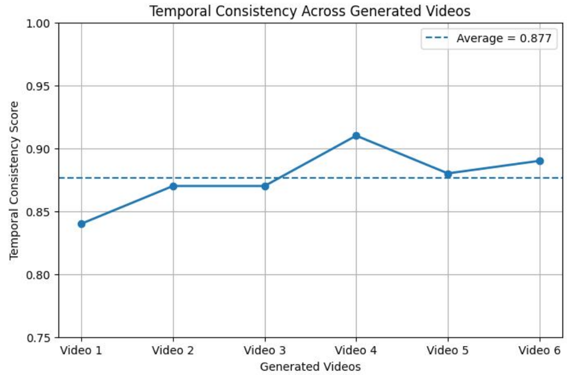



Fig. 3 Temporal Consistency Across Generated Videos

Fig.3 displays the produced promotional videos temporal consistency scores, exhibiting steady visual coherence and smooth frame transitions.

### *D. Advertisement Image Generation*

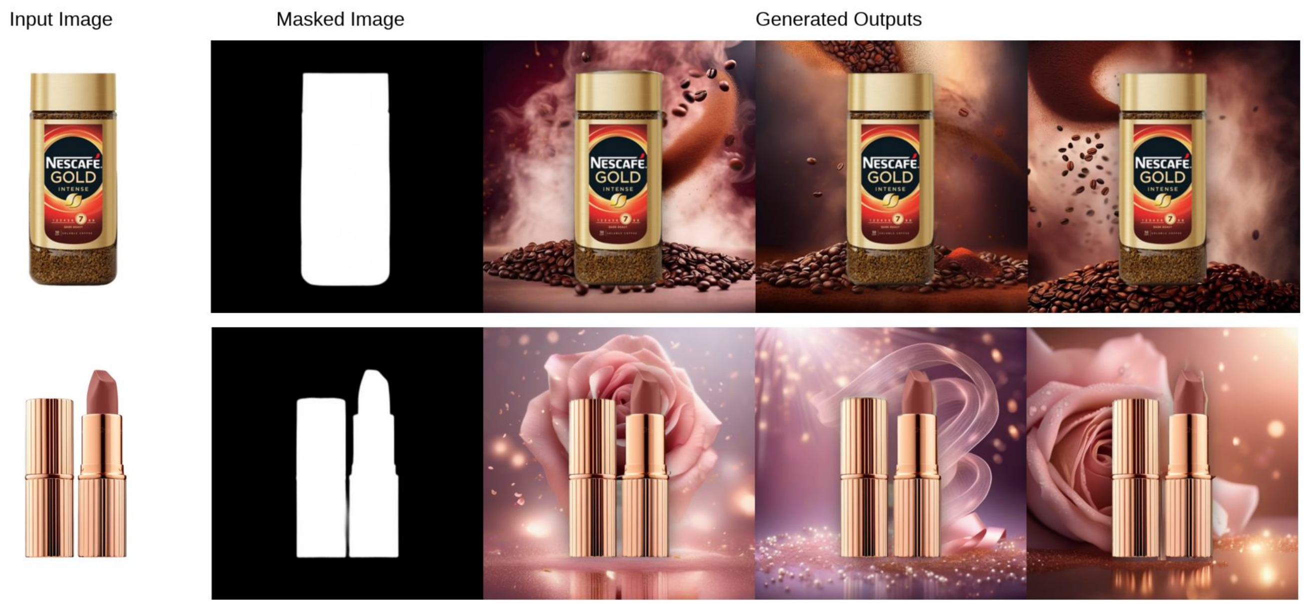



Fig. 4 Advertisement Image Generation Results of the Proposed Framework

The Fig. 4 illustrates the results of the creation of the advertisement image. The produced outputs provide visually appealing advertisement scenarios appropriate for marketing applications while maintaining product features.

### *E. Promotional Video Generation*


Fig.5 Promotional Video Output Visualization

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 14 Issue VI June 2026- Available at www.ijraset.com*

Fig.5 indicates the results of the creation of advertising videos. The retrieved frames preserve product attributes throughout the video sequence while exhibiting smooth motion production and visual coherence.

*F. System Interface*

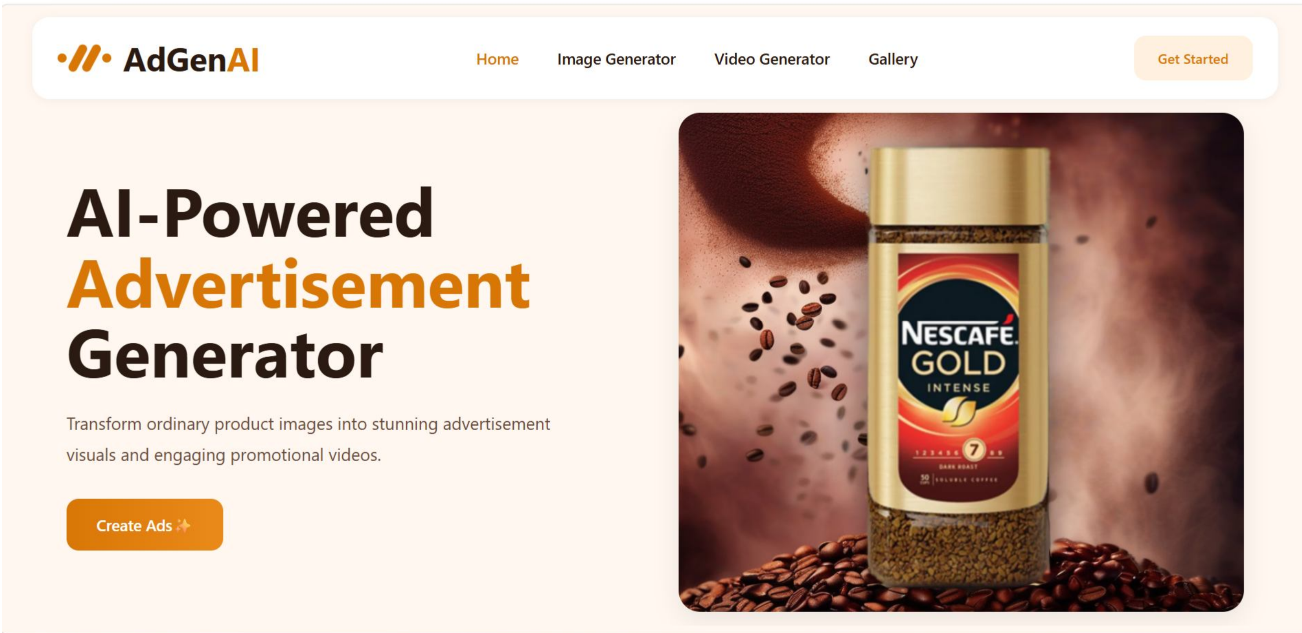



Fig. 6 Home Page of the Advertisement Generation System

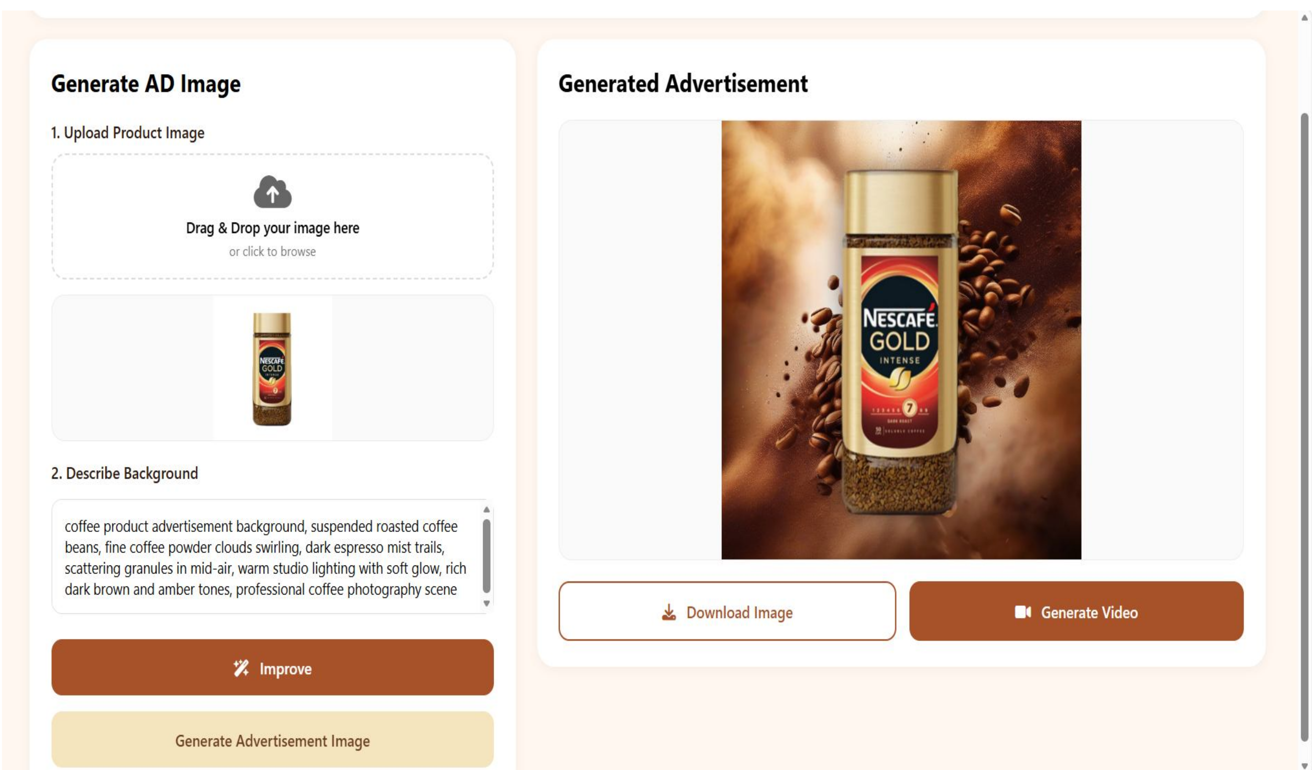



Fig. 7 mage Advertisement Generation

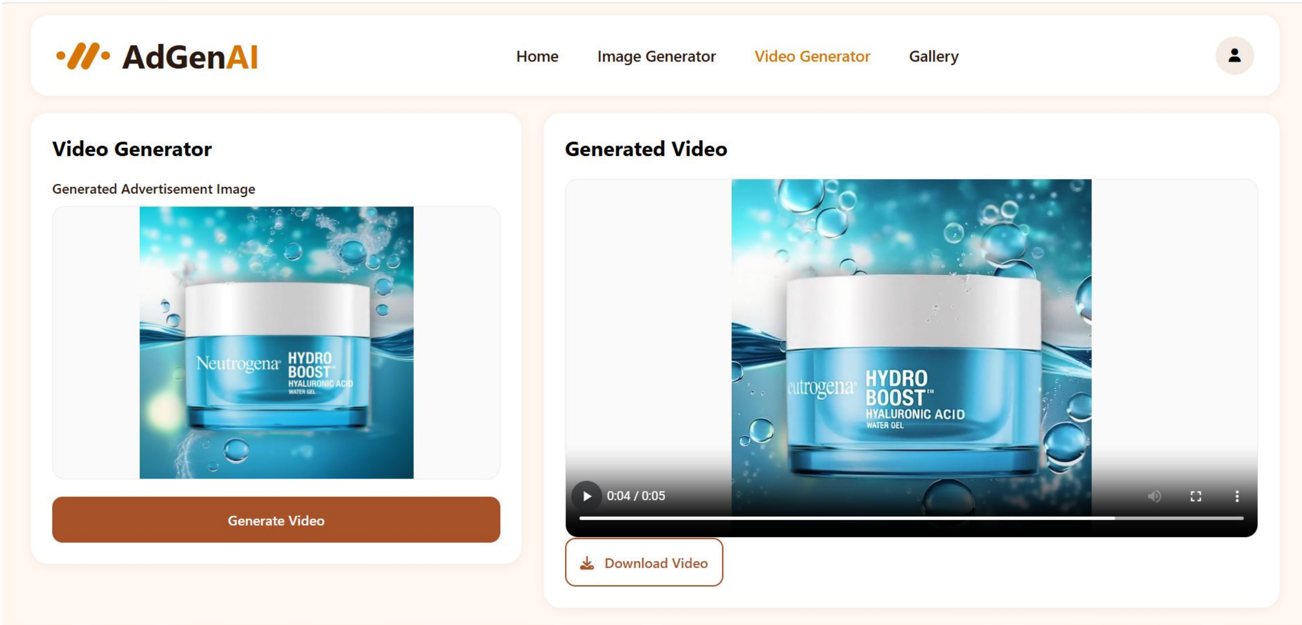



Fig. 8 Promotional Video Generation

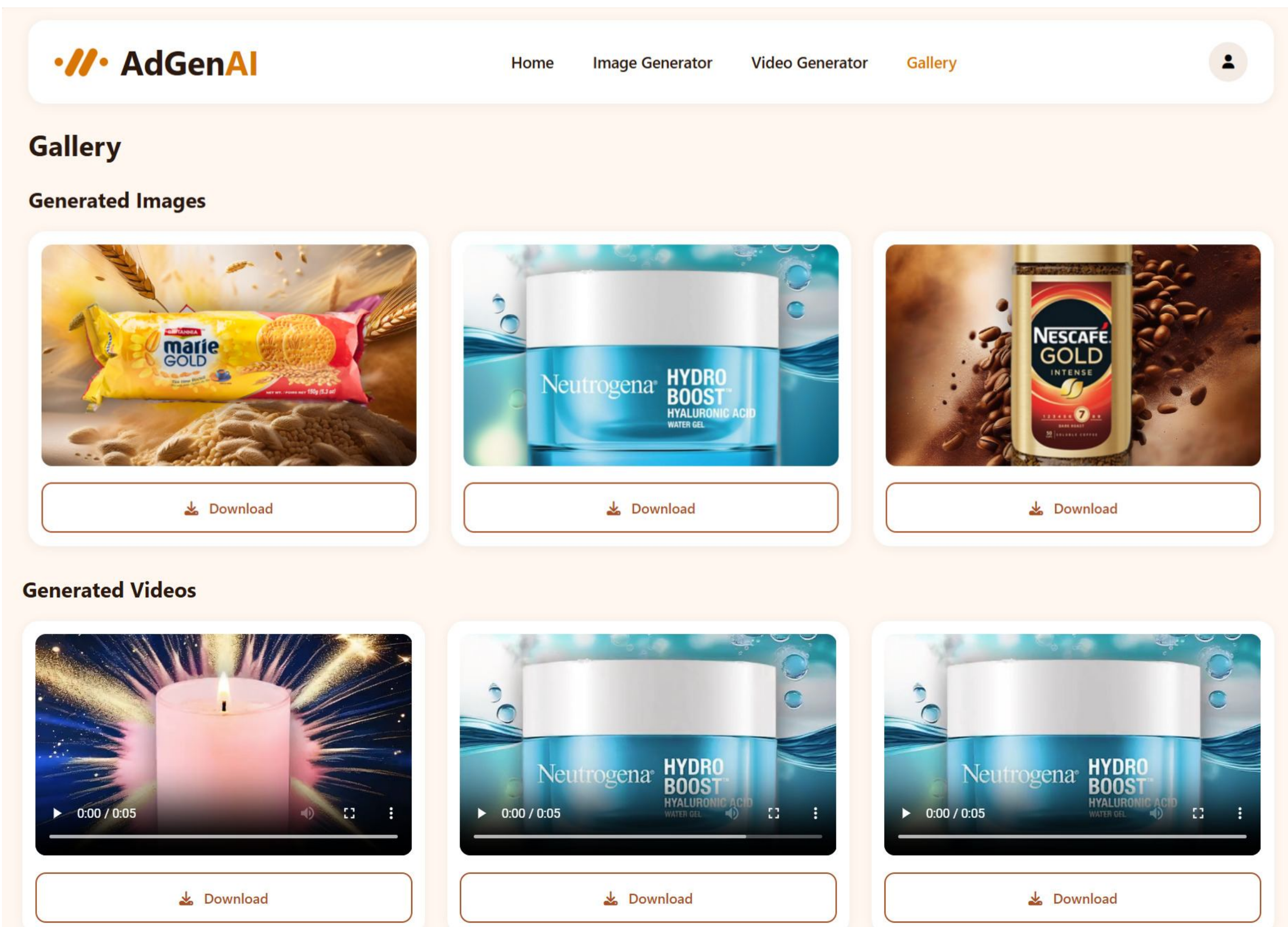



Fig. 9 Gallery page

## V. CONCLUSION

Developing an AI-powered hybrid framework for automated product advertisement generation was the main goal of this study. The suggested solution combines WAN 2.2 for promotional video synthesis, SDXL Inpainting for creating advertising images, Gemini Flash prompt enhancement, and picture preprocessing into a single workflow. The framework successfully converts common product photos into eye-catching commercial images and captivating promotional videos while maintaining key product attributes, according to the results. The capacity of the suggested method to produce high-quality and semantically relevant advertising content is confirmed by quantitative evaluation utilizing CLIP Score, SSIM, LPIPS, HPS v2, and Temporal Consistency.

With all factors assessed, the suggested framework offers an effective, scalable, and useful solution for AI-driven advertisement production, decreasing the quantity of manual design work while assisting with digital marketing, social media promotion, and e-commerce applications.

## VI. FUTURE WORK

The proposed framework can be enhanced by incorporating advanced generative models for better marketing visuals and longer promotional videos.AI-based voice-over synthesis, multilingual advertisement creation, and automated advertisement text generating are possible future advances. Personalized marketing, improved video realism, and interaction with social media and e-commerce platforms can all be areas for future development.

## VII. ACKNOWLEDGMENT

I express my sincere gratitude to my guide, Dr. B. Kranthi Kiran, Professor, Department of Computer Science and Engineering, Jawaharlal Nehru Technological University Hyderabad, for his valuable guidance, continuous support, and constructive suggestions throughout the development of this research work.

I also extend my heartfelt thanks to the faculty members of the Department of Computer Science and Engineering for their support and motivation during the course of this study.

## REFERENCES

[1] He, Y., Yang, S., Zhang, H., Wang, J., & Chen, W., RefAdGen: High-Fidelity Advertising Image Generation, pp. 1–8, 2024, https://doi.org/10.48550/arXiv.2508.11695

[2] Blattmann, A., Dockhorn, T., Kulal, S., Mendelevitch, D., Kilian, M., & Lorenz, D., Stable Video Diffusion: Scaling Latent Video Diffusion Models to Large Datasets, arXiv preprint arXiv:2311.15127, 2023,https://arxiv.org/abs/2311.15127

[3] Du, Z., Feng, W., Wang, H., Li, Y., Wang, J., Li, J., Zhang, Z., Lv, J., Zhu, X., Jin, J., Shen, J., Lin, Z., Shao, J., "Towards Reliable Advertising Image Generation Using Human Feedback", Computer Vision – ECCV 2024, Lecture Notes in Computer Science (LNCS), Vol. 15078, pp. 399–415, Springer, Cham, 2025, https://doi.org/10.1007/978-3-031-72661-3_23

[4] A. A. Rokhade, A. Deivam, A. J. V. Shettigar, T. M. and V. R. B. Prasad, "Intelligent Advertisement Generation: Harnessing Deep Learning Techniques," Proceedings of the 8th International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE,pp.1127–1132,2024,ISSN:2694-2090, https://doi.org/10.1109/ICAAIC60222.2024.10575640

[5] J. Hartmann, Y. Exner, and S. Domdey, "The power of generative marketing: Can generative AI create superhuman visual marketing content?," International Journal of Research in Marketing, vol. 42, no. 3, pp. 245–263, 2025, Elsevier, ISSN: 0167-8116, https://doi.org/10.1016/j.ijresmar.2024.09.002

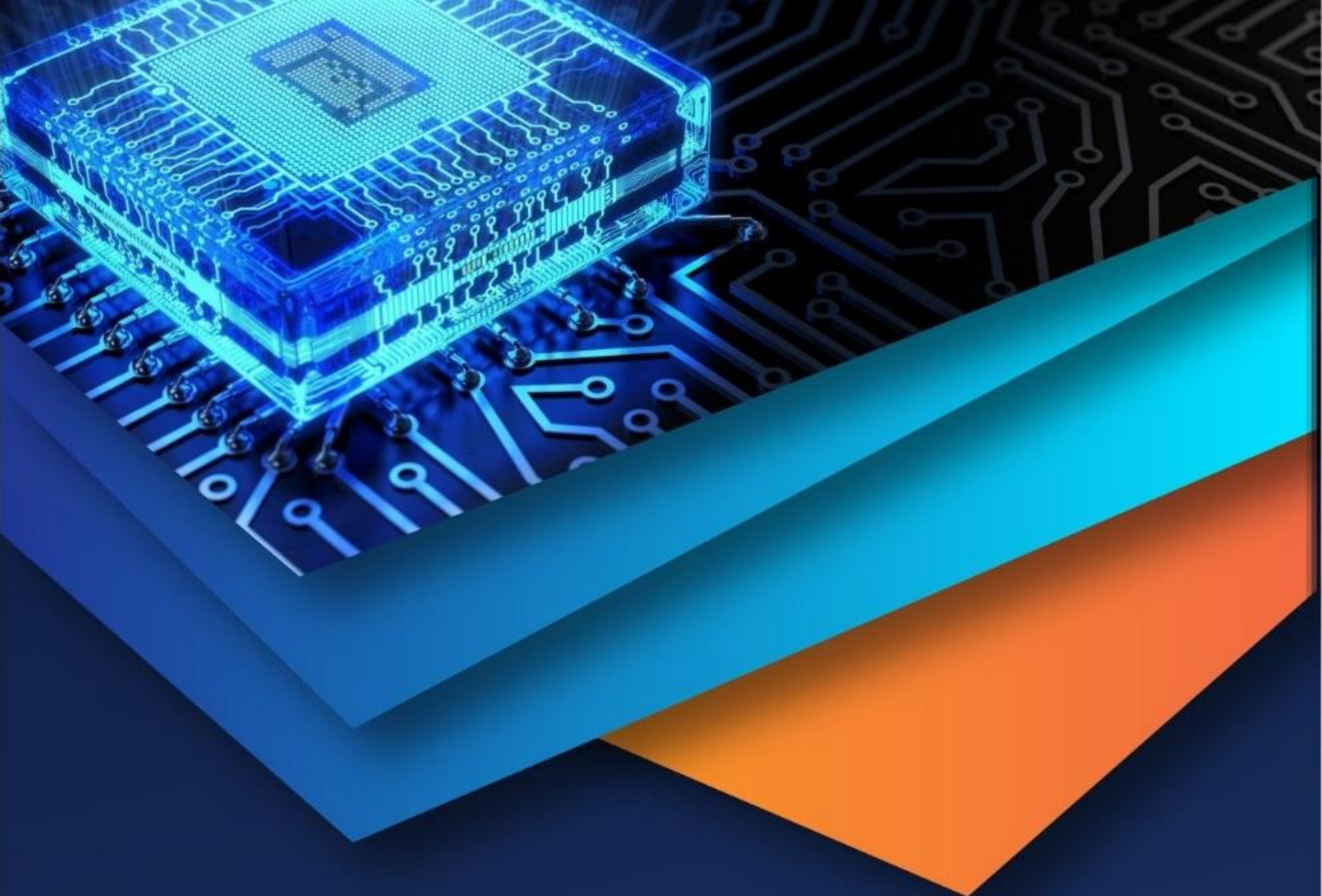
[6] Zheng, X., Qiao, X., Cao, Y., and Lau, R. W. H. "Content-Aware Generative Modeling of Graphic Design Layouts." ACM Transactions on Graphics (TOG), vol. 38, no. 4, Article 133, pp. 1–15, July 2019. ISSN: 0730-0301, https://doi.org/10.1145/3306346.3322971

[7] Liu, K., Li, W., Yang, C., and Yang, G. "Intelligent Design of Multimedia Content in Alibaba." Frontiers of Information Technology & Electronic Engineering, vol. 20, no. 12, pp. 1657–1668, 2019. Zhejiang University and Springer-Verlag GmbH Germany, part of Springer Nature. ISSN 2095-9184 (print), ISSN 2095-9230 (online), https://doi.org/10.1631/FITEE.1900580

[8] Blattmann, A., Rombach, R., Ling, H., Dockhorn, T., Kim, S. W., Fidler, S., and Kreis, K., "Align Your Latents: High-Resolution Video Synthesis with Latent Diffusion Models," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 22563–22575, 2023, https://doi.org/10.1109/CVPR52729.2023.02161

[9] Yingqing He, Tianyu Yang, Yong Zhang, Ying Shan, Qifeng Chen, "Latent Video Diffusion Models for High-Fidelity Long Video Generation," arXiv preprint arXiv:2211.13221, 2023, https://doi.org/10.48550/arXiv.2211.13221

[10] Yu, S., Sohn, K., Kim, S., and Shin, J., "Video Probabilistic Diffusion Models in Projected Latent Space," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 18456–18466, 2023, https://doi.org/10.1109/CVPR52729.2023.01770



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR:
7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY