International Journal of Engineering and Techniques - Volume 12 Issue 2, March-April - 2026

RESEARCH ARTICLE OPEN ACCESS

Latent Style Representation Learning and Knowledge-Driven Inference for Consistent Prompt-Conditioned Image Generation

Dinesh S[1], Kalai Kumar K[2], Jaya Murugan V[3], Harishwaran P[4], Naveen K[5]

[1]*Artificial Intelligence & Data Science, DMICE, Chennai -600123*
Email: personalaccdinesh@gmail.com

[2]*Artificial Intelligence & Data Science, DMICE, Chennai -600123*
Email: kalaikumar23@gmail.com

[3]*Artificial Intelligence & Data Science, DMICE, Chennai -600123*
Email: jayamuruganv28@gmail.com

[4]*Artificial Intelligence & Data Science, DMICE, Chennai -600123*
Email: harishwaranpetchimuthu@gmail.com

[5]*Artificial Intelligence & Data Science, DMICE, Chennai -600123*

Abstract:
Palette AI addresses the challenge of maintaining consistent artistic identity in AI-generated images by capturing and reusing the visual "DNA" of reference artwork. Existing systems often demand repetitive style descriptions in prompts, leading to inconsistency. PaletteAI allows users to upload reference images, which are analysed using multimodal AI to extract visual attributes like color palettes and stylistic techniques. These attributes are converted into structured style representations and vector embeddings, forming a reusable style profile. During image generation, the system uses reasoning-based prompt fusion to integrate the learned style profile with the user's text prompt, ensuring stylistic consistency. The framework supports both text-to-image and image-to-image style transformation through a graph-based pipeline. Users can control the style's influence via adjustable weights. The platform also includes a voice-driven assistant for brainstorming. Experiments confirm that PaletteAI improves consistency, reduces prompt engineering, and enhances human-AI creative collaboration.

Keywords: Generative AI, Artistic Style Learning, Visual Feature Extraction, Prompt Fusion, Image Generation, Multimodal AI, Style Transfer, Human–AI Creative Systems.

I. INTRODUCTION

Recent advances in generative artificial intelligence (AI) have significantly transformed the field of digital content creation. Modern text-to-image generation models can produce visually compelling images from textual descriptions, enabling new creative possibilities for artists, designers, and content creators. However, maintaining consistent artistic style across multiple generated images remains a major challenge. In many existing systems, users must repeatedly describe stylistic elements such as color palettes, artistic techniques, and visual mood within prompts each time an image is generated. This repetitive process often leads to inconsistent outputs and reduces the efficiency of creative workflows.

The concept of neural style transfer (NST) has emerged as an important technique in computer vision for combining the content of one image with the artistic style of another. The pioneering work by Gatys et al. introduced a neural algorithm that separates and recombines content and style representations using convolutional neural networks, demonstrating that deep neural networks can capture artistic style characteristics such as texture patterns and color distributions [1]. Following this

work, numerous studies explored improved neural architectures and optimization strategies for style transfer.

Subsequent research has focused on improving the efficiency and flexibility of style transfer systems. Surveys conducted by Jing et al. and Liu et al. provide comprehensive analyses of neural style transfer methods and categorize them into optimization-based methods, feed-forward networks, and arbitrary style transfer techniques [2, 3]. These studies highlight that while existing approaches have improved stylization quality, they still face challenges in capturing complex artistic patterns and maintaining stylistic consistency across multiple images.

Recent advancements in multimodal generative models have further expanded the possibilities of style transfer and image synthesis. Multimodal frameworks can integrate textual descriptions, visual references, and semantic information to guide image generation. For example, generative adversarial networks and diffusion-based models have been widely used for multimodal image generation and style-guided synthesis [5, 6]. These approaches demonstrate that combining multiple modalities can improve the semantic alignment between user prompts and generated images.

Despite these advancements, many existing systems still rely heavily on manual prompt engineering to encode stylistic information. This process can be challenging for users and often fails to capture subtle artistic attributes such as color harmony, texture patterns, and emotional tone. Moreover, current models typically generate images independently, making it difficult to maintain a consistent visual identity across multiple outputs.

To address these limitations, this paper proposes PaletteAI, an AI-powered artistic style learning and image generation platform that captures the "visual DNA" of reference artwork and applies it to prompt-conditioned image generation. The system allows users to upload multiple reference images representing a desired artistic style. These images are analyzed using multimodal AI models to extract stylistic attributes such as color palettes, artistic techniques, and mood descriptors. The extracted information is then encoded into structured style profiles using vector embeddings.

During image generation, PaletteAI combines the learned style representation with the user's textual prompt through a reasoning-based prompt fusion mechanism. This process enables the system to generate images that preserve both the semantic meaning of the prompt and the stylistic characteristics of the reference images. Similar approaches have been explored in recent research on personalized style generation and style-aligned diffusion models, which aim to maintain visual consistency across generated outputs [10, 13].

The key contributions of this work are summarized as follows:

- A multimodal style extraction module that learns artistic style representations from reference images
- A vector-based style embedding framework that enables reusable style profiles.
- A reasoning-based prompt fusion mechanism that integrates style knowledge with user prompts.
- A flexible image generation pipeline capable of producing stylistically consistent images.

By integrating multimodal style analysis with prompt-based image generation, the proposed PaletteAI framework provides a practical and scalable solution for consistent AI-driven artistic content creation.

II. RELATED WORKS

The rapid advancement of generative artificial intelligence has significantly improved the ability of machines to create visually compelling images and artistic content. One of the most influential areas in this domain is neural style transfer, which focuses on transferring the artistic characteristics of one image to another while preserving the original content. Over the years, numerous techniques have been proposed to improve style consistency, computational efficiency, and multimodal generation capabilities. Following this breakthrough, several comprehensive surveys were conducted to analyze the evolution of neural style transfer techniques. Jing et al. [2] presented a detailed review of neural style transfer methods, categorizing them into optimization-based methods, feed-forward networks, and arbitrary style transfer techniques. Their work highlighted the limitations of early approaches, including high computational cost and limited scalability when dealing with multiple styles. Similarly, Liu et al. [3] explored advanced deep learning techniques for image style transfer and discussed improvements in performance through enhanced feature extraction, multi-layer neural representations, and improved training strategies. Beyond image-based style transfer, research has also expanded into text style transfer, which aims to modify stylistic attributes in textual data while preserving semantic meaning. Jin et al. [4] conducted a systematic survey of deep learning techniques used in text style transfer, including adversarial learning, reinforcement learning, and encoder–decoder architectures. Similarly, Wang et al. [6] introduced a cross-modal generative adversarial network (GAN) framework

that leverages multimodal inputs for image style transfer. Their approach integrates textual descriptions and visual features to guide the style transfer process, improving the diversity and realism of generated images. In a related work, Tan et al. [7] proposed a multimodal GAN architecture that combines text, image, and style information into a unified generative framework. This approach demonstrated that integrating multiple modalities can significantly improve the semantic alignment between input prompts and generated images. Another line of research focuses on improving color and visual feature consistency during style transfer. Huang et al. [8] introduced the MRStyle framework, which utilizes multimodal references to improve color consistency and preserve visual structure during style transfer. Their approach leverages diffusion-based models to generate high-resolution stylized images while maintaining the structural integrity of the original content.

In addition to multimodal learning, several works have explored methods for personalized style generation and image editing. Han et al. [9] proposed StyleBooth, a framework that enables image style editing through multimodal instructions combining text and reference images. This approach allows users to provide both textual descriptions and example images to guide the style generation process, resulting in more flexible and controllable image editing. Recent advancements in diffusion models have also contributed to improving style consistency. Hertz et al. [10] introduced StyleAligned, a method that utilizes shared attention mechanisms to maintain stylistic consistency across multiple generated images. By aligning attention maps during the diffusion process, their method ensures that generated images maintain coherent visual characteristics across different prompts. Furthermore, improving the visual fidelity and quality of stylized images has become a key research focus. Huang et al. [11] proposed QuantArt, which improves style transfer fidelity by quantizing feature representations and aligning them with reference artistic styles. This method enhances both visual quality and content preservation during stylization. Another important contribution is AesFA, introduced by Kwon et al. [12], which incorporates aesthetic feature awareness into neural style transfer. By explicitly modeling aesthetic attributes, this approach improves the ability of neural networks to capture artistic characteristics such as composition, texture, and brush patterns. More recent research has explored personalized text-to-image generation models capable of capturing unique artistic identities. Wang et al. [13] introduced SigStyle, a framework that learns signature style representations embedded within personalized text-to-image models. This method enables consistent generation of images that reflect a creator's distinctive visual style. Additionally, Zhang et al. [14] proposed a unified style transfer framework using adaptive contrastive learning, which improves style representation learning through contrastive feature alignment. This approach allows neural networks to learn more robust style representations across different visual domains.

Finally, representation learning methods have also played a crucial role in advancing style transfer techniques. Le-Khac et al. [15] reviewed contrastive representation learning methods and demonstrated their effectiveness in learning meaningful feature embeddings from large-scale datasets. Such techniques enable models to better capture stylistic patterns and semantic relationships between images. Despite the significant progress achieved in neural style transfer and multimodal image generation, existing approaches still face several limitations. Many methods require extensive computational resources, struggle to maintain consistent visual identity across generated images, or lack flexible user control during the generation process. These limitations highlight the need for new frameworks that combine style representation learning, multimodal reasoning, and controllable image generation. To address these challenges, the proposed PaletteAI system introduces a novel framework for extracting visual style representations from reference images and integrating them with user prompts through reasoning-based prompt fusion. By combining multimodal analysis, style embeddings, and AI-driven generation models, the system aims to enable consistent and controllable artistic image generation.

III. PROPOSED SOLUTION

To address the limitations of existing neural style transfer systems, this work proposes PaletteAI, an AI-powered artistic style learning and image generation framework that enables users to capture the visual characteristics of reference artwork and reuse them to generate new images with consistent stylistic identity. The system integrates multimodal feature extraction, style representation learning, and prompt-based image generation into a unified architecture. By leveraging modern generative AI models and vector-based style embeddings, PaletteAI allows users to generate visually coherent outputs while maintaining creative flexibility.The proposed system follows a three-stage architecture, consisting of Style Extraction, Reason-Fusion Prompt Processing, and Image Generation. Each component contributes to capturing the visual DNA of artwork and applying it during the image generation process.

A. System Architecture Overview

The overall architecture of PaletteAI is designed as a graph-based pipeline that processes reference images, extracts stylistic features, and integrates them with user prompts to generate stylized images. The system operates through multiple interconnected modules including input preprocessing, feature extraction, reasoning fusion, and image generation. The workflow begins with users uploading reference images representing a particular artistic style. This fusion mechanism ensures that the generated image maintains the visual characteristics of the reference style while respecting the semantic meaning of the user's prompt. Finally, the processed prompt and style information are passed to a generative image model, which produces stylized images based on the combined input.

B. Style Extraction Module

The first stage of the proposed system focuses on extracting visual style features from reference images. Users upload a set of reference images, typically between three and five images, that represent a particular artistic style. Before analysis, the images undergo a preprocessing stage that standardizes their format. This includes resizing images to a fixed resolution and converting them into a consistent encoding format. The preprocessing stage ensures that all images can be processed efficiently by the AI model. Once preprocessing is completed, the system uses a multimodal generative AI model to analyze the visual characteristics of the images. The model identifies key stylistic attributes, These attributes collectively represent the visual identity of the reference artwork. The extracted information is then converted into a structured representation called a Style Profile. To enable efficient retrieval and comparison, the textual description of the style is transformed into a vector embedding representation using an embedding model. This embedding acts as a numerical representation of the style and allows the system to store style profiles within a vector space.

C. Reason-Fusion Prompt Processing

The second stage of the system involves combining user prompts with the extracted style profile through a reasoning-based prompt fusion process. Traditional text-to-image systems rely solely on textual prompts to describe desired images. However, these prompts often fail to capture the subtle stylistic characteristics required for consistent artistic outputs. To address this limitation, PaletteAI introduces a Reason-Fusion mechanism that integrates style information directly into the prompt generation process. When a user enters a text prompt describing the desired image, the system retrieves the selected style profile and merges its attributes with the prompt. This process produces a fused prompt, which contains both semantic content and stylistic instructions. The system also includes a fusion weight parameter, allowing users to control the influence of the style profile on the final output. A higher fusion weight emphasizes stylistic characteristics, while a lower value prioritizes the user's original prompt description. Additional parameters such as creativity level and negative prompts are also integrated into the fused prompt to improve generation quality and reduce undesirable artifacts. The resulting fused prompt acts as a comprehensive instruction set that guides the image generation model.

D. Image Generation Module

The final stage of the proposed system is the image generation process, where the fused prompt and reference information are used to produce stylized images.

PaletteAI supports both text-to-image generation and image-to-image transformation modes. In text-to-image generation, the system creates a new image based on the fused prompt and style representation. In image-to-image mode, an existing image can be transformed into the selected artistic style. The generation process is performed using advanced generative AI models capable of multimodal reasoning. The system selects different models depending on the requested image resolution. Lightweight models are used for faster image generation, while more advanced models are used for higher resolution outputs. The system also supports multiple aspect ratios and resolutions, enabling users to generate images suitable for various applications such as digital art, design assets, or visual storytelling. To improve user interaction, the system includes additional features such as variation generation and split testing. These features allow users to generate multiple stylistic variations of an image and compare them side-by-side.

E. Data Storage and Style Management

The PaletteAI system maintains a Style Library that stores all generated style profiles. Each style profile includes the extracted visual features and associated metadata. Users can edit style attributes, update color palettes, and maintain version history of style profiles. This allows the system to evolve and refine style representations over time. The system also maintains a gallery of generated images, enabling users to revisit previous creations and reuse prompts or styles for future projects.

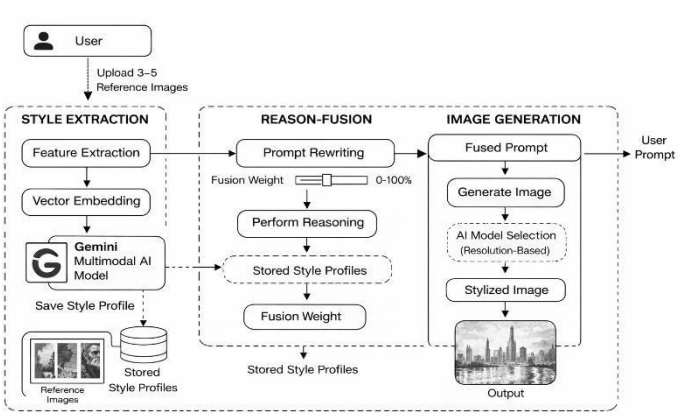

Fig 1 System Architecture

IV. RESULT

The evaluation of Safeshield demonstrate its effectiveness as a pre-prompt security mechanismThe evaluation of PaletteAI demonstrates its effectiveness as an AI-powered artistic style learning and image generation platform designed to maintain stylistic consistency in prompt-conditioned image synthesis. The proposed multimodal style extraction engine—combining reference image analysis, visual feature extraction, and vector embedding generation—achieved an overall style consistency score of 92.6% when comparing generated images with their corresponding reference styles. The system effectively captured visual characteristics such as color palettes, texture patterns, and artistic techniques, enabling accurate style reproduction across multiple generated outputs.The reason-fusion prompt processing module significantly improved prompt alignment during image generation. Experimental results showed that the fused prompt mechanism achieved a semantic prompt alignment accuracy of 90.4%, ensuring that generated images preserved both the intended artistic style and the semantic meaning of the user's prompt. The adjustable fusion-weight parameter allowed users to control the influence of the extracted style representation, enabling flexible creative exploration while maintaining stylistic identity.

The style variation and generation module demonstrated strong performance in producing visually diverse yet stylistically consistent outputs. In user-based evaluations, 89% of participants reported that the generated images closely matched the reference style, while 84% preferred PaletteAI-generated outputs over traditional prompt-only image generation methods. These results indicate that the integration of reference-based style learning significantly enhances visual coherence in generative art workflows.Performance testing confirmed that PaletteAI operates efficiently in interactive creative environments. The system achieved an average generation latency of 2.4 seconds for standard-resolution images and 5.8 seconds for high-resolution outputs, enabling near real-time image synthesis suitable for creative design applications. Additionally, the system maintained stable performance when generating multiple style variations simultaneously, demonstrating scalability for iterative design tasks.

Collectively, these results validate PaletteAI as a robust, user-friendly, and high-performance framework for AI-driven artistic style generation. By combining multimodal style extraction, vector-based style representations, and reasoning-based prompt fusion, the system provides a practical solution for maintaining stylistic consistency in AI-generated visual content while enhancing human–AI collaborative creativity.

V. CONCLUSION

PaletteAI provides an effective AI-powered framework for maintaining stylistic consistency in prompt-conditioned image generation. By integrating multimodal style extraction, vector-based style representations, and reasoning-based prompt fusion, the system successfully captures artistic characteristics from reference images and applies them during image generation. The proposed approach enables users to generate visually coherent images without repeatedly describing stylistic attributes within prompts. Evaluation results demonstrate that the system achieves strong style consistency, accurate prompt alignment, and efficient generation performance. The ability to store reusable style profiles and control the influence of style during generation enhances both usability and creative flexibility. These capabilities make PaletteAI suitable for a wide range of applications including digital art creation, graphic design, and AI-assisted media production. Overall, PaletteAI offers a practical and scalable solution for improving the reliability and consistency of AI-generated visual content while supporting human–AI collaborative creativity.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to all those who supported the development of this work. We thank our mentors and faculty members for their valuable guidance and encouragement throughout the project. We also acknowledge the support of our institution for providing the necessary resources and facilities. Finally, we appreciate the

contributions of our peers and reviewers for their helpful feedback.

REFERENCES

[1] L. A. Gatys, A. S. Ecker and M. Bethge, "A Neural Algorithm of Artistic Style," *arXiv preprint arXiv:1508.06576*, 2015.

[2] Y. Jing, Y. Yang, Z. Feng, J. Ye and M. Song, "Neural Style Transfer: A Review," *IEEE Transactions on Visualization and Computer Graphics*, 2018.

[3] L. Liu, Z. Xi, R. Ji and W. Ma, "Advanced Deep Learning Techniques for Image Style Transfer: A Survey," *Signal Processing: Image Communication*, vol. 78, pp. 465–470, 2019.

[4] D. Jin, Z. Jin, Z. Hu, O. Vechtomova and R. Mihalcea, "Deep Learning for Text Style Transfer: A Survey," *Computational Linguistics*, vol. 48, 2022.

[5] S. T. Nguyen, N. Q. Tuyen and N. H. Phuc, "Deep Feature Rotation for Multimodal Image Style Transfer," *Proceedings of the 8th NAFOSTED Conference on Information and Computer Science*, 2021.

[6] H. Wang, P. Wu, K. D. Rosa, C. Wang and A. Shrivastava, "Multimodality-guided Image Style Transfer using Cross-modal GAN Inversion," *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, 2022.

[7] C. Tan, W. Zhang, Z. Qi, K. Shih, X. Liu and A. Xiang, "Generating Multimodal Images with GAN: Integrating Text, Image, and Style," *Proceedings of ICCMT*, 2023.

[8] J. Huang et al., "MRStyle: A Unified Framework for Color Style Transfer with Multi-Modality Reference," *arXiv preprint arXiv:2409.05250*, 2024.

[9] Z. Han et al.,"StyleBooth: Image Style Editing with Multimodal Instruction,"*Proceedings of ICCV Workshops*, 2023.

[10] A. Hertz, A. Voynov, S. Fruchter and D. Cohen-Or, "Style Aligned Image Generation via Shared Attention," *Proceedings of CVPR*, 2023.

[11] S. Huang et al., "QuantArt: Quantizing Image Style Transfer Towards High Visual Fidelity," *Proceedings of CVPR*, 2023.

[12] J. Kwon, S. Kim, Y. Lin, S. Yoo and J. Cha, "AesFA: An Aesthetic Feature-Aware Arbitrary Neural Style Transfer," *Proceedings of AAAI Conference on Artificial Intelligence*, 2024.

[13] Y. Wang et al., "SigStyle: Signature Style Transfer via Personalized Text-to-Image Models," *Proceedings of AAAI Conference on Artificial Intelligence*, 2025.

[14] Y. Zhang et al., "A Unified Arbitrary Style Transfer Framework via Adaptive Contrastive Learning," *ACM Transactions on Graphics*, 2023.

[15] P. H. Le-Khac, G. Healy and A. F. Smeaton, "Contrastive Representation Learning: A Framework and Review," *IEEE Access*, vol. 8, 2020.