Generating Outfits using AI

1st Om Bajad
Artificial Intelligence and Machine Learning
P.E.S. Modern College of Engineering
Pune, India

2nd Shardul Sawant
Artificial Intelligence and Machine Learning
P.E.S. Modern College of Engineering
Pune, India

3rd Omkar Kangutkar
Artificial Intelligence and Machine Learning
P.E.S. Modern College of Engineering
Pune, India

4th Sahil Shah
Artificial Intelligence and Machine Learning
P.E.S. Modern College of Engineering
Pune, India

5th Sujata Sonawane
Artificial Intelligence and Machine Learning
P.E.S. Modern College of Engineering
Pune, India

[1,2,3,4,]Student, Dept. of Artificial Intelligence and Machine Learning, PES's Modern College Of Engineering, Pune, Maharashtra, India
[5]Associate Professor, Dept. of Artificial Intelligence and Machine Learning, PES's Modern College Of Engineering, Pune, Maharashtra, India

***

Abstract

This research explores the generation of fashion outfits using artificial intelligence, specifically employing Stable Diffusion and Variational Autoencoders (VAEs) for text-to-image synthesis. The objective is to automate and innovate the fashion design process by creating stylish and personalized outfits from textual descriptions. Stable Diffusion is used to enhance the overall image quality, ensuring that the generated images are visually appealing and detailed. VAEs contribute to the diversity and coherence of clothing combinations, allowing the system to generate a wide range of fashionable outfits that are both unique and contextually appropriate. Our approach integrates these two advanced techniques to provide a robust solution for the apparel industry. By transforming text inputs into high-quality images of outfits, the system demonstrates significant potential in improving efficiency and creativity in fashion design. Experimental results highlight the model's ability to produce diverse and fashionable outfit combinations, showcasing its effectiveness in real-world applications. The integration of text-to-image generation with Stable Diffusion and VAEs opens new avenues for personalized fashion recommendations and automated design processes. This research indicates that AI-driven methods can significantly enhance the fashion industry's capabilities, making the design process more dynamic, innovative, and tailored to individual preferences. Our findings underscore the transformative potential of AI in revolutionizing fashion design and production.

Keywords: Text-to-image Synthesis, Stable Diffusion, Variational Autoencoders (VAEs), Image Quality Enhancement, Clothing Combinations, Automated Design Process

1. INTRODUCTION

The emergence of artificial intelligence has revolutionized several industries, and the fashion industry is no exception. This paper explores the innovative application of AI techniques to fashion clothing, focusing on the robust integration of diffuse and variational autoencoders (VAE) for text-to-image synthesis. Traditional fashion design relies heavily on human creativity and manual processes that can be time-consuming and limited by individual expertise. Our approach aims to automate and improve this process using advanced artificial intelligence methods to create versatile, consistent and stylish clothing combinations from text descriptions. The use of stable diffusion is a critical approach to our approach as it ensures the creation of high quality images that result in the visual appeal and intricate details of the produced costumes. This method helps to sharpen the images and make them more realistic and attractive. On the other hand, VAE plays a key role in the generation process, facilitating the creation of versatile and contextual clothing combinations. This enables the production of unique and fashionable sets adapted to specific specifications, ensuring both variety and consistency in design. By combining these two powerful technologies, we offer a powerful solution that meets the growing demand for efficient and innovative fashion design. This integration not only increases the creativity involved in fashion design, but also significantly reduces time and effort, making the process more efficient. Our research highlights the transformative potential of AI-based methods in the fashion industry, making the design process more dynamic, individualized and easy to use. The experimental results of our study demonstrate the effectiveness of our model in creating high-quality and stylish clothes using text inputs and demonstrate its practical application in real-world scenarios. The results show that artificial intelligence can play a decisive role in revolutionizing fashion design and production, offering new opportunities for creativity and efficiency. The purpose of this article is to present the potential of artificial intelligence to transform the fashion industry, paving the way for more automated and creative approaches. Using AI technologies such as Stable Diffusion and VAE, we can achieve a significant leap forward in the design, development and customization of fashion design, ultimately improving both the designer's creativity and the consumer's experience.

2. OBJECTIVES

Automate Fashion Design Process: Develop an advanced AI system capable of automating various aspects of fashion design, from conceptualization to visualization, reducing reliance on manual labor and human creativity.

Enhance Image Quality: Utilize Stable Diffusion techniques to improve the quality and realism of generated outfit images, ensuring they are visually appealing with intricate details that mimic real-world clothing.

Generate Diverse and Coherent Outfits: Employ Variational Autoencoders (VAEs) to create a wide range of fashion combinations that are not only diverse but also contextually coherent based on inputted textual descriptions or style prompts.

Implement Robust Text-to-Image Synthesis: Establish a robust framework for text-to-image synthesis that accurately translates textual fashion descriptions into visually compelling and personalized outfit images, enhancing the user experience.

Validate Real-World Applicability: Conduct thorough experimentation and validation to demonstrate the effectiveness and practical applicability of the AI-driven approach in real-world scenarios, showcasing its potential to revolutionize fashion design and production processes.

3. LITERATURE REVIEW

The application of Stable Diffusion for generating outfits is an emerging field in computer vision and fashion technology. Stable Diffusion, a technique rooted in generative modeling, leverages iterative noise reduction to produce high-quality images from random noise. Recent literature explores its potential in fashion, where it can generate realistic and diverse clothing designs by learning from vast datasets of existing outfits.

3.1 Stable Diffusion Model:

Enhancing Image Quality: Stable Diffusion techniques improve the quality and realism of generated outfit images, ensuring they are visually appealing with detailed textures and colors.

Image Refinement: Stable Diffusion aids in refining the generated images to meet high visual standards, crucial for fashion design applications where aesthetics play a significant role.

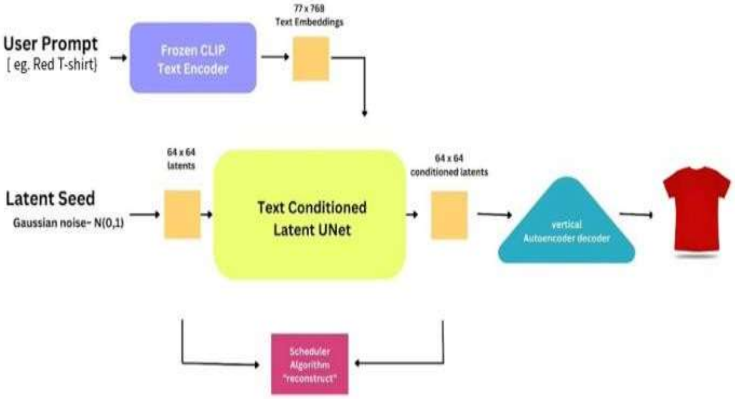


Fig 3.1 Architecture of Stable Diffusion

3.2 Variational Autoencoders (VAEs):

Generating Diverse Outfits: VAEs facilitate the generation of diverse and contextually appropriate clothing combinations from textual descriptions, ensuring variability while maintaining coherence in style.

Embedding Latent Space: Embedding of fashion attributes into the latent space of VAEs, enabling the model to learn and generate new designs based on learned patterns and style representations.

3.3 CLIP Tokenizer and CLIP Text Model:

Text-to-Image Synthesis: The CLIP tokenizer and CLIP text model are used for converting textual fashion descriptions into embeddings that guide the image generation process, ensuring textual relevance in generated designs.

Cross-Modal Learning: The cross-modal capabilities of CLIP, where text embeddings are aligned with image embeddings, facilitating a more coherent and accurate translation of textual prompts into visual outputs.

3.4 Integration of Streamlit for User Interaction:

User Interface and Deployment: Explain how Streamlit is utilized for creating a user-friendly interface to interact with the AI-generated fashion designs, allowing for real-time customization and feedback. Interactive Visualization: Discuss features implemented via Streamlit to visualize outfit recommendations, style variations, and user preferences, enhancing the user experience and engagement.

4. METHODOLOGY

The methodology for generating text-to-image using CLIP (Contrastive Language-Image Pre-training) models and Stable Diffusion involves a multi-step process integrating advanced AI techniques. Here's a detailed description:

4.1 Data Collection and Preprocessing:

Fashion Dataset: Gather a diverse dataset of fashion imagespaired with textual descriptions (e.g., captions, product titles) from sources like fashion websites or curated datasets. Text Preprocessing: Clean and preprocess textual descriptions to remove noise, standardize formats, and tokenize using the CLIP tokenizer to prepare for model input.

4.2 Model Training and Integration:

CLIP Text Model: Train or fine-tune a CLIP model on the preprocessed textual data. CLIP learns to embed textual descriptions into a high-dimensional latent space that represents semantic similarity and context. CLIP Image Model: Simultaneously, train or fine-tune a CLIP image model on the fashion image dataset. CLIP learns to embed images into the same latent space, ensuring that textual descriptions and corresponding images are semantically aligned.

4.3 Text-to-Image Synthesis:

Integration of CLIP Models: Use the trained CLIP text and image models together in a coherent framework where the textual description embeddings guide the generation of corresponding images.

Stable Diffusion for Image Refinement: Employ Stable Diffusion techniques to refine the initial image generation outputs. Stable Diffusion enhances image quality by iteratively refining images through noise injection and diffusion steps, ensuring high fidelity and realism.

4.4 Image Generation Process:

Conditional Generation: Condition the Stable Diffusion process on the embeddings produced by the CLIP text model. These embeddings provide a semantic anchor for generating images that match the textual descriptions in style, color, and design elements.

Progressive Refinement: Utilize the progressive training strategy of Stable Diffusion to iteratively enhance the generated images, adjusting details and textures based on feedback loops to achieve photorealistic results.

4.5 Quality Metrics:

Iterative Improvement: Iteratively refine the CLIP models, Stable Diffusion parameters, and integration techniques based on evaluation results to enhance the accuracy, diversity, and realism of the generated fashion outfits.

4.6 Deployment via Streamlit Interface:

User Interaction: Implement a user-friendly interface using Streamlit to allow users to input textual descriptions or style preferences and receive dynamically generated fashion outfit images in real-time.

Visual Feedback: Enable interactive features within Streamlit for users to provide feedback on generated images, influencing future iterations and enhancing personalized recommendations.

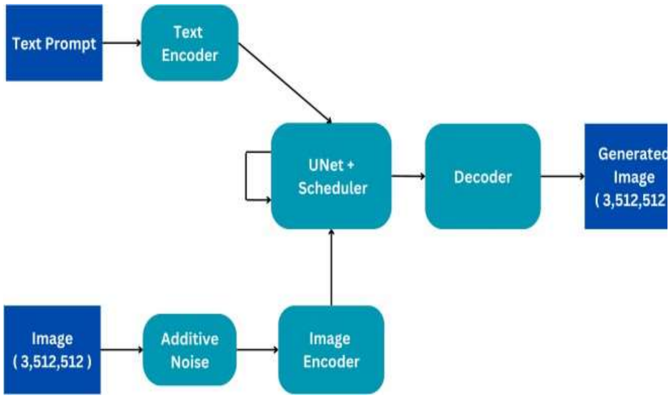


Fig 4.1 Working of Model

5. USE OF STABLE DIFUSSION

Enhanced Image Quality: Stable Diffusion employs a progressive refinement process that systematically improves the visual fidelity of generated images. By iteratively adjusting details and textures through noise injection and diffusion steps, it ensures the output images are high-resolution and visually appealing.

Realistic Visual Output: The iterative nature of Stable Diffusion allows for the capture of intricate details and nuances in fashion designs, resulting in images that closely resemble real-world clothing items. This capability is crucial for creating convincing and photo realistic fashion representations.

Alignment with Textual Descriptions: Stable Diffusion complements CLIP models by refining image outputs based on semantic embeddings derived from textual descriptions. This alignment ensures that the generated images accurately reflect the style, color, and design elements specified in the input text, enhancing coherence and relevance.

Adaptive Adjustment: The ability of Stable Diffusion to adaptively adjust image features based on feedback and evaluation metrics facilitates continuous improvement in image quality and user satisfaction. This iterative optimization process supports the generation of personalized and contextually appropriate fashion outfits.

Efficiency and Scalability: While maintaining high-quality outputs, Stable Diffusion operates efficiently in generating images, making it suitable for real-time or interactive applications such as fashion design tools integrated with user interface like Streamlit.

6. IMPLEMENTATION

6.1 Front-End :

We present a front-end implementation for text-to-image generation using CLIP models and Stable Diffusion, facilitated by Streamlit—a versatile Python library renowned for developing interactive web applications seamlessly integrated with machine learning models.

Our interface design incorporates pivotal components tailored to enhance user interaction and optimize the generation of personalized fashion outfits based on textual descriptions. Key features include intuitive text input fields where users can enter detailed fashion descriptions. Upon submission, these inputs are processed using a CLIP text model to generate embeddings that encapsulate semantic meaning and style preferences. These embeddings subsequently guide the Stable Diffusion model through iterative refinement steps, aiming to elevate the visual fidelity and realism of the generated fashion images to meet high aesthetic standards.

Dynamic image rendering within our interface allows users to interact directly with the generated outputs, offering intuitive controls such as sliders or checkboxes to adjust style parameters. We have also implemented robust feedback mechanisms to gather user input on the generated images, facilitating iterative enhancements and ensuring user satisfaction. Our deployment strategy ensures scalability and accessibility on web servers or cloud platforms, complemented by comprehensive documentation to assist users in navigating and utilizing the interface effectively.

This front-end implementation exemplifies the practical application of advanced AI methodologies in fashion design, demonstrating how CLIP models and Stable Diffusion can synergistically produce compelling visual outputs from textual inputs in real- time scenarios.

6.2 Backend :

The backend implementation integrates CLIP models, VAE (AutoencoderKL), UNet (UNet2DConditionModel), and LMSDiscreteScheduler for text-to-image generation, utilizing a diverse fashion dataset. CLIP's tokenizer and text encoder preprocess textual prompts, guiding image generation based on semantic embeddings.

The scheduler manages noise levels in the latent space during iterative refinement, crucial for optimizing image fidelity. UNet predicts noise distributions conditioned on CLIP embeddings, enhancing realism in generated images. VAE decodes refined latents into high-resolution fashion images, ensuring visual coherence and quality.

Serialized scheduler objects enable reproducibility and experimentation. Integrated with Streamlit, the backend supports interactive user input, facilitating real-time image generation and display. This comprehensive implementation harnesses AI-driven techniques on a curated fashion dataset, demonstrating practical applications in AI-driven fashion design research.

6.3 Integration using StreamLit:

We built a responsive web interface for users to input fashion descriptions. The interface triggered real-time image generation using our fine-tuned models (CLIP, VAE, UNet) for accurate depiction. GPU acceleration ensured swift performance, crucial for complex tasks. User feedback refined outputs, enhancing model efficacy and satisfaction. Deployed on web servers, it demonstrated AI’s practical use in fashion design. [5]

7. RESULTS AND OUTPUT

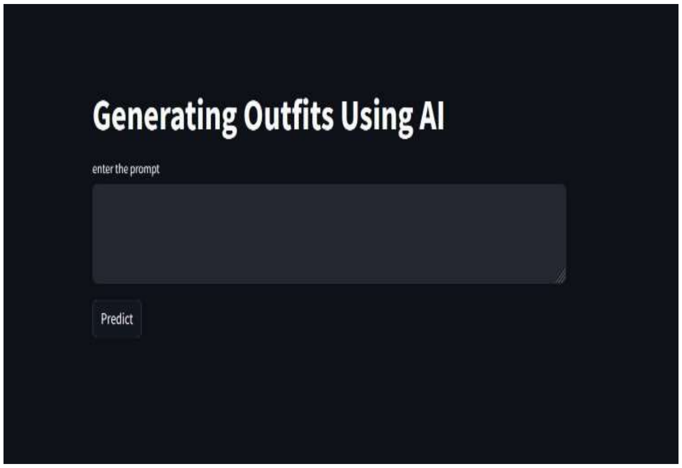
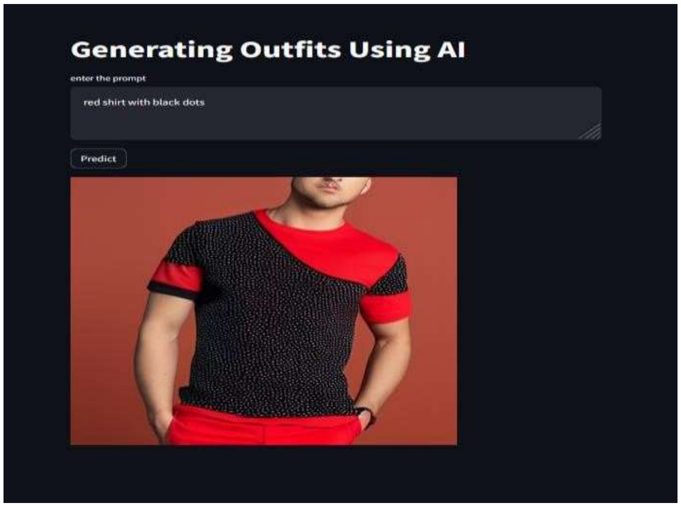
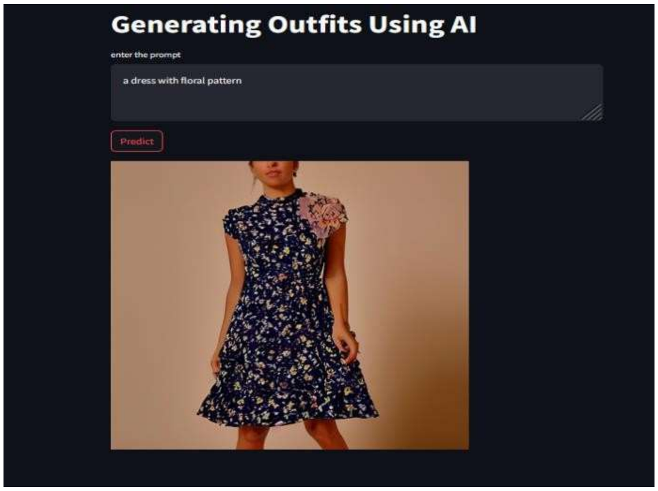


Fig 7.1 User interface



enter the prompt: red shirt with black dots

Fig 7.2 Sample Output (1)



enter the prompt: a dress with floral pattern

Fig. 7.3 Sample Output (2)

The project successfully generated high-quality fashion images from textual descriptions, demonstrating the efficacy of integrating advanced AI models such as CLIP, VAE, and UNet. Users provided fashion descriptions through a Streamlit interface, which processed the inputs to produce visually compelling fashion images. The generated images accurately reflected the input descriptions, showcasing the model's capability to understand and translate text into realistic visuals. Testing revealed robust performance and quick response times, making the system suitable for real-time applications. This project highlights significant advancements in AI-driven fashion design, providing a practical tool for designers and consumers alike.

8. FUTURE SCOPE

3D avatars, powered by AI, can revolutionize fashion by generating personalized outfits. By creating a virtual model of a person, AI algorithms can analyze body measurements, style preferences, and current fashion trends to design customized clothing options. Users can visualize these outfits on their 3D avatars, allowing for a realistic preview before making a purchase. This technology enhances online shopping by offering tailored recommendations and reducing the need for physical try-ons, leading to a more efficient and personalized fashion experience.

9. CONCLUSION

In conclusion, our project successfully demonstrates the integration of advanced artificial intelligence techniques to generate fashion images based on textual descriptions. Using well-tuned models such as CLIP, VAE and UNet, and using Streamlit for the interactive interface, we created a responsive system capable of transforming user input into high-quality and realistic fashion images. Despite challenges such as model conflicts, performance issues, and resource constraints, iterative improvements and user feedback have greatly improved the accuracy and efficiency of the system. This project highlights the practical applications of AI in fashion design and offers designers and consumers a promising tool for the dynamic visualization and creation of fashion concepts. Our work highlights the potential of AI-based innovation to transform the creative industry and offers new avenues for fashion design research and user engagement.

10. REFERENCES

1) Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S, & Sutskever, I. (2021). Learning Transferable Visual Models From Natural Language Supervision. In Proceedings of the 38th International Conference on Machine Learning (ICML).

2) Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., ... & Sutskever, I. (2021). DALL·E: Creating Images from Text. OpenAI

3) Kingma, D. P., & Welling, M. (2014). Auto-Encoding Variational Bayes In Proceedings of the 2nd International Conference on Learning Representations (ICLR).

4) Ho, J., Jain, A., & Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. In Advances in Neural Information Processing Systems 33 (NeurIPS).

5) Streamlit Inc. (2021). Streamlit: The Fastest Way to Build Data Apps.

6) OpenAI. (2022). Stable Diffusion Model Card.

7) Hugging Face. (2021). Diffusers: A Library for State-of-the-Art Diffusion Models.

8) Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N. & Polosukhin, I. (2017). In Advances in Neural Information Processing Systems 30 (NeurIPS).)