TAILOR

Amgad Abdelbasset
Computer science department
Ain shams university
20191700132

Andrew Nasssef Amin
Computer science department
Ain shams university
20191700144

Andre Adel Ibrahim
Computer science department
Ain shams university
20191700145

Maria Maurice Ayoub
Computer science department
Ain shams university
20191700473

Mayve Ehab Rasmy
Computer science department
Ain shams university
20191700485

Mina Samy Anwar
Computer science department
Ain shams university
20191700676

I. ABSTRACT

TAILOR AI Art Generator is a mobile application designed provide users with the ability to generate unique and creative fashion designs using artificial intelligence.
The project's motivation stems from the need to offer users an efficient and accessible way to generate custom images of wedding dresses or casual wear based on their own text inputs. The traditional process of designing custom clothing can be time-consuming and expensive, often requiring the skills of a professional designer or costly software. The AI Art Generator aims to address these challenges by leveraging AI technology to generate personalized clothing designs quickly and easily.
The results of the developed system are highly promising. The AI model successfully generates images that closely match the user's input descriptions of wedding dresses or casual wear. Users can create unique and personalized designs quickly and easily, saving both time and money compared to traditional methods. The system offers accessibility to a wider range of people, including those who may not have access to high-end fashion designers or a large budget for expensive clothing. Furthermore, the system has the potential to contribute to the fashion industry by providing opportunities for young people to develop their skills and creativity.
Through this comprehensive documentation, readers will gain a thorough understanding of the AI Art Generator application, its significance in the fashion design domain, and its potential for further development and innovation.

II. INTRODUCTION

There is a need for an efficient and accessible way for users to generate custom images of wedding dresses or casual wear based on their own text inputs. Traditionally, designing custom clothing can be a time-consuming and expensive process, requiring the skills of a professional designer or costly software. This project aims to provide a solution that is accessible to a wider range of users, allowing them to generate unique and personalized clothing designs quickly and easily. By using AI technology to generate these designs, the project also aims to improve the accuracy and consistency of the design process, reducing the need for manual adjustments and revisions.

The main motivation for us was concentrated in the following points:

Time and cost savings: the average cost of a wedding dress design depending on whom you select to make your dress but typically it can range from $2,000 - $8,000, and the average cost of casual wear design $200 - $250, which can be a significant expense for many people. By providing a tool that allows users to create their own unique designs, that can help save them time and money.
Customization: According to a survey by The Wedding Report, 83% of brides want a personalized wedding dress. By using an AI text to image generator, users can create a dress that is tailored to their specific preferences and style, rather than having to settle for a pre-made design.
Accessibility: Not everyone has access to high-end fashion designers or the budget to purchase expensive clothing. By providing a tool that allows users to create their own designs, you can help make fashion more accessible to a wider range of people.
Additionally, according to a report by the UNDP, Egypt has a high youth unemployment rate, which can make it difficult for young people to afford high-end fashion or to find employment in the fashion industry. By providing a tool that allows users to create their own designs, your AI art generator can help make fashion more accessible and provide opportunities for young people to develop their skills and creativity.

III. RELATED WORK

A. DreamBooth

Fine Tuning Text-to-Image Diffusion Models for Subject Driven Generation

DreamBooth uses a few input images and a corresponding class name to fine-tune a text-to-image model. The model can then generate diverse instances of the subject by implanting a unique identifier in different sentences. The approach involves two steps.

1) Fine-tuning the text-to-image model with the input images and class name.

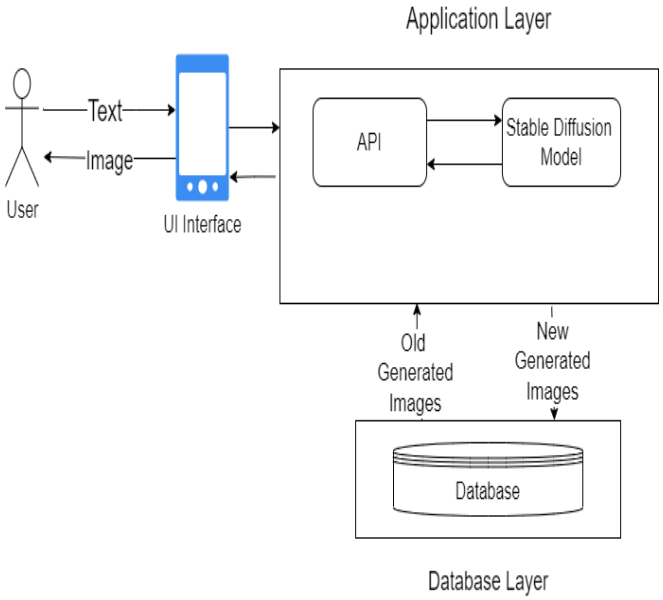
2) Generating diverse instances of the subject by implanting a unique identifier in different sentences.

B. ERNIE-ViLG 2.0

Improving Text-to-Image Diffusion Model with Knowledge-Enhanced Mixture-of-Denoising-Experts

ERNIE-ViLG 2.0 is a large-scale Chinese text-to-image diffusion model that progressively upgrades the quality of generated images by incorporating fine-grained textual and visual knowledge of key elements in the scene and utilizing different denoising experts at different denoising stages.

IV. SYSTEM ARCHITECTURE



The software architecture is divided into two layers: the application layer and the database layer. The application layer consists of a user, a user interface, an API, and a stable diffusion model. The user interacts with the user interface, which sends requests to the API. The API accesses the stable diffusion model. The stable diffusion model is used to generate images for the user based on the text prompt user provided to the user interface. Each user can store his/her generated images into the database and retrieve it any time

Datasets*:*

1) First Dataset*:*

- Fashion product dataset, Kaggle. 21K of fashion products images divided into two subcategories (top wear, bottom wear)
- 10 features describing the product such as gender, base color, year,

2) Second Dataset*:*

- 500 collected wedding dresses images with detailed description.
- Preprocessing done on images: face segmentation, background.
- Preprocessing done on text: summarize the text prompt by getting the most important sentences in the description, enter the output of step 1 to T5-transformer to summarize the description according to the meaning of the entered sentences.

V. RESULTS

Testing*:*

We used Fréchet Inception Distance which is a measure of similarity between two datasets of images. It was shown to correlate well with the human judgment of visual quality and is most often used to evaluate the quality of samples of Generative Adversarial Networks. FID is calculated by computing the Fréchet distance between two Gaussians fitted to feature representations of the Inception network.

These two datasets are essentially the dataset of real images and the dataset of fake images (generated images in our case). FID is usually calculated with two large datasets. However, we couldn't use large dataset with the wedding dresses model.

In the casual wear model, we had a huge dataset, so we split the dataset and we choose unique prompts selected randomly from the entire dataset, while in the dresses model, we had to collect another 60 images of wedding dresses as it is a self-collected dataset and used them.

VI. CONCLUSION AND FUTURE WORK

A. Conclusion

In this project, we have developed a virtual clothing customization system that allows users to generate realistic clothing images based on textual descriptions of desired attributes. Throughout the project, we have discussed the various components, methodologies, and techniques employed to design and implement the system. Now, let us provide a complete summary of the project, including the results obtained.

1) The primary objective of the project was to create a user-friendly and interactive platform where users could effortlessly customize clothing items by simply describing their desired attributes. To achieve this, we designed and implemented a system that combines natural language processing techniques with state-of-the-art image generation models.

2) We began by conducting a thorough analysis of user requirements and designing the system architecture accordingly. The system architecture consists of components such as the user interface, text processing module, image generation module, and database. These components work together seamlessly to provide a smooth and intuitive user experience.

3) The image generation module employs a diffusion model and algorithm, which leverages deep learning and probabilistic modeling to generate visually coherent and contextually relevant clothing images. By utilizing diffusion processes, the model captures both global structure and fine-grained details, resulting in high-quality image synthesis.

4) Throughout the implementation phase, we rigorously tested and optimized the system to ensure its reliability and efficiency. We also integrated the system with a database to store and manage user data and previously generated images.

B. Future Work

1) The project opens possibilities for further enhancements and extensions, such as incorporating additional clothing categories, refining the user interface, and integrating more advanced image generation models. With continuous development and improvement, the system has the potential to revolutionize the way users interact with virtual clothing customization.

VII. References

A. Blattmann, A., Rombach, R., Oktay, K. and Ommer, B., 2022. Retrieval-Augmented Diffusion Models. arXiv preprint arXiv:2204.11824.

B. Feng, Z., Zhang, Z., Yu, X., Fang, Y., Li, L., Chen, X., Lu, Y., Liu, J., Yin, W., Feng, S. and Sun, Y., 2022. ERNIE-ViLG 2.0: Improving Text-to-Image Diffusion Model with Knowledge-Enhanced Mixture-of-Denoising-Experts. arXiv preprint arXiv:2210.15257.

C. Li, R., Li, W., Yang, Y., Wei, H., Jiang, J. and Bai, Q., 2022. Swinv2-Imagen: Hierarchical Vision Transformer Diffusion Models for Text-to-Image Generation. arXiv preprint arXiv:2210.09549.

D. Balaji, Y., Nah, S., Huang, X., Vahdat, A., Song, J., Kreis, K., Aittala, M., Aila, T., Laine, S., Catanzaro, B. and Karras, T., 2022. ediffi: Text-to-image diffusion models with an ensemble of expert denoisers. arXiv preprint arXiv:2211.01324.

E. Yu, J., Xu, Y., Koh, J.Y., Luong, T., Baid, G., Wang, Z., Vasudevan, V., Ku, A., Yang, Y., Ayan, B.K. and Hutchinson, B., 2022. Scaling autoregressive models for content-rich text-to-image generation. arXiv preprint arXiv:2206.10789.

F. Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M. and Aberman, K., 2023. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 22500-22510).

G. Gal, R., Alaluf, Y., Atzmon, Y., Patashnik, O., Bermano, A.H., Chechik, G. and Cohen-Or, D., 2022. An image is worth one word: Personalizing text-to-image generation using textual inversion. arXiv preprint arXiv:2208.01618.

H. Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, Kfir Aberman., 2022 DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation using textual inversion. arXiv preprint arXiv:2208.12242v2