Towards AI-Generated African Textile Patterns with StyleGAN and Stable Diffusion

Christelle Scharff, Samyak Rakesh Meshram, Krishna Mohan Bathula, FNU Kaleemunnisa, Om Gaikhe
Department of Computer Science
Pace University
New York, USA
e-mail: {cscharff, smeshram, kbathula, klnu, ogaikhe}@pace.edu

***Abstract*—Wax are traditional colorful textiles worn across Africa. They are composed of patterns of geometrical and symmetrical shapes that repeat indefinitely. This paper explores and compares the generation of African wax designs using StyleGan2-ADA, StyleGAN3 and Stable Diffusion architectures on a curated synthetic dataset of 2000 1024x1024 images obtained with DALL·E 2. The generated wax designs are evaluated using Fréchet Inception Distance (FID). StyleGAN2-ADA and Stable Diffusion generated better images. StyleGAN2-ADA generated designs diverse in colors, shapes and details with some symmetry and repetition. Stable Diffusion was stronger with symmetry and repetition, but it generated less details. By providing a new tool for creating customizable wax designs, this study has the potential to have an impact on the fashion industry. It is novel as it makes a case for inclusive AI by focusing on applications outside the scope of today's mainstream fashion industry. It also shows that the suggested approaches are promising to produce a variety of plausible and culturally appropriate designs. Our next step is to work with African fashion designers and wax experts to validate the resulting designs.**

Keywords-African wax patterns;* Fréchet Inception Distance (FID); *Stable Diffusion; StyleGAN.

I. Introduction

The fashion industry is one of the fastest growing sectors in Africa. African wax prints, also called Dutch wax prints, were introduced to Africa by Dutch merchants in the 19th century. Vlisco is the most widely known Dutch wax print textile manufacturer worldwide. Wax prints are widely present in West and Central Africa. They are bright colored cotton fabrics. The technique to create wax is inspired from the Indonesian hand-crafted batik method. The name "wax" comes from the technique based on heated wax and colored dyes to obtain the patterns on the fabrics. Wax print fabrics have recently reached a global audience and have been featured in fashion shows of international designers. Wax is recognized as a symbol of African identity worldwide. Wax designs are based on geometrical and special shapes and objects inspired from African culture and carrying meanings and messages. Patterns repeat indefinitely and are usually symmetrical. Some of the original and popular fabrics depict table fans to represent modernity, speed birds to symbolize volatile fortune, and alphabet to indicate literacy.

Creating African wax design is currently done using graphical editing software. Using Generative Artificial Intelligence (AI) would allow manufacturers to produce more designs in less time and guide the creative process with textual descriptions of different granularity of details.

This paper is a first attempt to generate wax designs using currently state-of-the-art generative AI techniques, StyleGAN [1][2] and stable diffusion [3]-[7] specifically. We curated a synthetic dataset of 2000 1024x1024 images generated with DALL·E 2. DALL·E is a general-purpose image generator. Our work focuses on wax generation, and it is expected that our generated wax will be better perceived than DALL·E by fashion professionals. Synthetic generated data are important to tackle problems when data are missing and to tap in new opportunities [8]. Our research relies on identifying the best performing models to generate representative wax designs. It is novel per the focus it takes on African fashion and the case it makes on inclusive AI focusing on applications outside the scope of today's mainstream fashion industry [9].

The paper is organized as follows. Section 2 provides background on StyleGAN and stable diffusion. Section 3 presents the dataset and the methodology we used. Section 4 covers the results we obtained. Section 5 concludes our work and outlines our future work.

II. Background

This section provides the context for this research, rooted in generative AI, and focuses on GAN and Stable Diffusion as generative AI models. It presents research initiatives on generative AI in fashion. It describes methods used to evaluate generated images, with an emphasis on Fréchet Inception Distance (FID).

A. GAN

Generative Adversarial Networks (GAN) [10] have gained popularity commercially in the last ten years, recently illustrated by the launch of the OpenAI DALL.E system that generates images from descriptive text. Part of unsupervised learning techniques, GAN aim at synthesizing huge amounts of images to create realistic new images. GAN reformulate the problem in terms of supervised learning where the generator produces new images and the discriminator classifies them as "real" or "fake".

B. GAN in Fashion

GAN have been highlighted in projects related to heritage, tradition, culture and art [11]-[15]. Fashion has generated numerous interesting problems in computer vision and machine learning, including in the use of GAN. Several fashion datasets have been created recently, including 3D datasets [16][17]. GarmentGAN [18] transfers target clothing items to reference body generating realistic images; it permits to see how clothes could fit bodies. Other studies have focused on generating new clothing on an individual using GAN [19].

C. SytleGAN

StyleGAN [1] is an alternative generator architecture for GAN. This architecture leads to an automatically learned, unsupervised separation of high-level attributes and random variation in the new generated images. It improves the state-of-the-art interpolation properties in terms of traditional measures of distribution quality. It permits to generate more photo-realistic high-quality images and to control the style of the generated images by focusing on different types of details. StyleGAN has been extended to StyleGAN2 [2] to reduce water-droplet artifacts appearing in StyleGAN images, StyleGAN2-ADA [5] to train GAN with limited amounts of data, and StyleGAN3 [20] to address image rotation and translation challenges.

D. Stable Diffusion and Latent Diffusion

Stable Diffusion and Latent Diffusion models, such as Stability AI SDXL 1.0 represent significant advancements in generative AI [4][6][21]. They operate by gradually transforming random noise into coherent images, leveraging a deep learning technique known as denoising. This process is guided by textual descriptions, allowing for the generation of detailed and contextually relevant elements. The latent diffusion approach further enhances this by operating in a latent space [22]. This latent space is a compressed representation of data, enabling more efficient and controllable image synthesis. The technique relies on a Variational Autoencoder (VAE) to map images to and from this latent space, and a Denoising Diffusion Probabilistic Model (DDPM) to iteratively refine these images [23]. This combination results in high-quality image generation with remarkable detail and coherence. The development of such models marks a significant improvement in creative AI.

E. Evaluating Generated Images

Evaluating generated images relies on quantitative and qualitative metrics. Inception Score (IS) is one of these metrics; it assesses images based on realism, clarity and diversity. While IS only evaluates the distribution of generated images, FID [24]-[26] compares the distribution of generated images with the distribution of a set of real images, indicating the model's performance in terms of replicating real-image statistics. It is used to evaluate generated images in terms of quality, diversity, and realism. Lowest scores for FID indicate more similarities between the generated and real images, signifying better quality. The approach that FID uses is to have the image embedded in a low-dimensional space using a state-of-the-art image recognition model, the one with the highest accuracy to measure the distribution distance in that space. For analyzing images, FID is typically based on the Inception-v3 mode [27] as it is well suited for GAN imagery. FID has several limitations including its unique application to images, its insensitivity to certain fine-grained details, its subjectivity that does not capture all aspects of human perception and preferences, and its requirements on the preprocessing of the images (scale, cropping and normalization). FID and other evaluation metrics should be coupled with Subject Matter Expert (SME) evaluation to judge the realism and details of generated images.

In addition to FID, Perceptual Path Length (PPL) was introduced as part of StyleGAN to evaluate generated images [2][28]. It measures the smoothness of transitions in the latent space, reflecting on the consistency and coherence of the image transformation. It also evaluates the diversity of generated images, ensuring the production of a wide range of distinct and plausible images, rather than the replication of a limited set of patterns. PPL [2][7], like FID, uses the feature embeddings of deep convolutional neural networks, but uses VGG16 network [29] instead of Inception-v3.

The assessment of FID scores in diffusion models like Stable Diffusion is markedly challenging [3]. FID scores measure the similarity between generated images and real images, based on features extracted by an Inception-v3 model, which is not used in Stable Diffusion. For Stable Diffusion models, the evaluation process is complex due to the time required to generate a sufficient number of images for accurate FID computation, StyleGAN models typically generate more images in the process. In addition to time constraints, the generation of this large number of images is also resource-intensive. Another aspect to consider is that the evaluation of Stable Diffusion models might also require investigating the model's ability to handle conditional image generation tasks, where the diffusion process is guided by textual descriptions or other forms of conditioning to produce images that are aligned with the given inputs. Lastly, researchers often explore the robustness and bias of the model, ensuring it generates diverse and fair outputs across various domains.

Qualitative evaluation is often necessary to complement quantitative metrics when evaluating generated images. It is based on human involvement to assess the realism and artistic quality of the outputs. The involvement of subject matter experts is crucial as some fine-grained elements related to the domain may not have been captured by the model.

III. Methodology

Our approach is based on experimenting with StyleGAN2-ADA, StyleGAN3 and Stable Diffusion on a curated synthetic dataset composed of wax pattern images.

A. Dataset

We curated a synthetic dataset (2K-Dataset) consisting of 2000 1024x1024 images collected from DALL·E 2. Per the Content Policy & Terms of DALL·E, we own the generated images. The reason for this synthetic dataset is that there is no dedicated wax dataset available. When evaluated qualitatively by fashion experts, DALL·E generated images have some similarities with wax but there is room for improvement. DALL·E is a general-purpose image generator. Our work focuses on wax generation, and it is expected that our generated wax will be better perceived by fashion professionals. These images were carefully selected to represent a diverse range of African wax designs, capturing different patterns, colors, and cultural motifs. To avoid biases in the images collected, such as color, design, pattern, and to ensure that the generated images are not duplicates, we used carefully selected predefined prompts (e.g., Green African wax', 'African wax textile pattern with traditional African symbols of Adire brown color', 'African wax textile pattern design inspired by modern styles with flowers in red'). Preprocessing steps, such as normalization, were applied to the 2K-Dataset to ensure optimal training performance. Figure 1 presents samples of our dataset.

Figure 1. Sample Synthetic DALL·E 2 Generated Wax 1024x1024 Images Used in the 2K-Dataset.

B. Experimentation

We used StyleGAN2-ADA, StyleGAN3 and Stable Diffusion as architectures and experimented with the 2K-Dataset. The training was done on the university High Performance Computer (HPC) on two nodes. Each node is a 2x Intel(R) Xeon(R) Gold 6136 CPU @ 3.00GHz (24 cores total), 384GB RAM and 3x Nvidia Tesla V100 GPUs.

1) StyleGAN2-ADA and StyleGAN3

In StyleGAN2-ADA and StyleGAN3, we used mirroring to increase the dataset size and hyperparameters in the augmentation pipeline, such as blit, geometry, color, filter and noise, and applied factors, such as scaling, rotation, brightness and contrast. We saved intermediate results as snapshots to monitor model training without storing its weights. We assigned 2 to map_depth, 0.25 to the general learning rate, and 0.25 to the discriminator learning rate. We started by initializing the model's weights and optimized them to minimize the FID.

We reached 35.57 as the FID for StyleGAN2-ADA and 65.27 for StyleGAN3 as the lowest FID. Figures 2 and 3 show samples of wax designs generated for these FID. When training StyleGAN, "kimgs" (thousands of images) serves as a measure of how much experience the model has gained with a particular type of data. This exposure is crucial for the model to understand and capture the unique characteristics, colors, textures, and designs inherent in images, in our case African wax patterns. The more kimgs the model is trained with, the more refined and accurate its generated patterns tend to be. In our experiment, we went up to 5000 kimgs. In practice, experiments have shown that 5000 kimgs is a good benchmark to show strong performance on a variety of datasets. As a result, it has become a somewhat standard training unit. There is however a balance to consider as overtraining (exposing the model to too many images) can lead to overfitting, where the model performs well on the training data but poorly on new, unseen data. In the context of StyleGAN and African wax patterns, the goal is to train the model with enough kimgs to effectively capture the diversity and complexity of these patterns, enabling it to generate new, high-quality designs that are both novel and representative of the style. Figure 4 shows the evolution of the FID in StyleGAN2-ADA and StyleGAN3 with the learning progress of the models, showing that StyleGAN2-ADA permits to obtain a better FID and that a plateau is obtained after 4000 kimgs.

Figure 2. Generated patterns of African wax designs using StyleGAN2-ADA with 1024x1024 images (FID 35.57).

Figure 3. Generated patterns of African wax designs using StyleGAN3 with 1024x1024 images (FID 65.27).

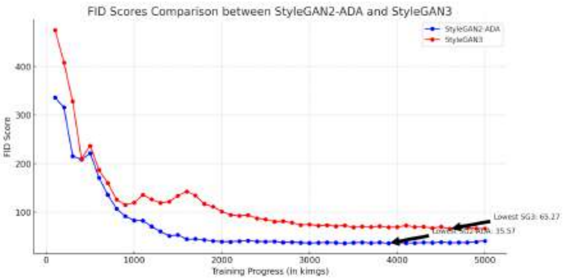

Figure 4. Training Progress of StyleGAN2-ADA and StyleGAN3 in Kimgs Depicting FID.

2) Stable Diffusion

The training of the Stable Diffusion model (Stability AI/SDXL 1.0) was conducted using the DreamBooth training method [30]. DreamBooth is an approach distinct from traditional GAN methods introducing unique training elements, such as personalized image generation, regularization techniques, and fine-tuning of pre-trained models. Unlike StyleGAN that focuses on generating diverse images from a large dataset, DreamBooth tailors the generation process to create images that are specific to a smaller, targeted dataset. It employs regularization techniques to maintain generalization abilities of the model while learning from a limited number of images. Instead of training the model from scratch like, it fine-tunes a pre-trained diffusion model on a small set of target images. Figure 5 presents resulting wax designs obtained with Stable Diffusion with FID of 226.75. In our approach, FID scores were calculated by generating 1000 images using our trained Diffusion Model, followed by comparing these images to our original dataset of 2000 images. This method, though resource-intensive, provided a feasible approach to gauge the model's performance.

Figure 5. Generated patterns of African wax designs using Stable Diffusion (FID 226.75) Prompt: Prompt: "Afwapa, beautiful african wax pattern with blue and black designs". Afwapa is a class name required by Dreambooth meaning AFrican WAx PAttern.

IV. RESULTS

In the considered application of generating African wax designs, StyleGAN2-ADA, StyleGAN3, and the Stable Diffusion (Stability AI SDXL) models showcased varied performances and nuances in the production of the generated images. StyleGAN models are dedicated to learn and mimic complex distributions from training data. When trained on the 2K-Dataset of African wax patterns, StyleGAN2-ADA and StyleGAN3 models attempt to capture the high-level abstract features and the distribution of colors, shapes, and patterns that define the African wax style. Since they are GAN, the training process involves the generator trying to produce new images that the discriminator cannot distinguish from real African wax designs. The models do this by internalizing the nuances of the training images, ranging from geometric repetitions to the intricate designs typical of African wax prints. StyleGAN architectures were capable of capturing global and local patterns, which are important for the coherent generation of such designs. However, in general, the quality of the generated images can vary depending on factors, such as the diversity and size of the training dataset, the alignment of images, the model capacity, and the specifics of the training data.

In the case of Stable Diffusion models, when fine-tuned on African wax designs, the model leveraged the characteristics of these patterns more directly throughout the image generation process. This could result in a higher fidelity to the specific overall style of African wax patterns, especially if the fine-tuning process is meticulously guided to preserve the defining features of the prints.

The StyleGAN2-ADA model achieved the lowest FID score of 35.57, indicating a high similarity between the generated images and the training dataset. This suggests that StyleGAN2-ADA effectively captures the intricate designs and color patterns characteristic of African Wax Patterns, possibly due to its robust architecture and data compatibility. The StyleGAN3 model, with an FID score of 65.27, did not perform as well as StyleGAN2-ADA in this specific task. Possible reasons for this might include the fact that StyleGAN3 might not be as optimized and sensitive as StyleGAN2-ADA to specific smooth transitions, textures and patterns present in African wax designs. We also noticed that some colors appeared dominant in the StyleGAN3 generated images, such as green and orange. On the other hand, despite its higher FID score of 226.75, the Stable Diffusion model is better in symmetry, image quality, and granular control through text-to-image guidance. This suggests a different focus of the model, prioritizing creative control and qualitative aspects of image generation over strict adherence to the training dataset's statistical properties. Its higher FID score could be indicative of a trade-off between creative diversity and statistical accuracy.

While StyleGAN2-ADA shows a strong ability to accurately replicate the statistical properties of the training dataset for African Wax Patterns, the StyleGAN3 model's performance indicates a potential mismatch between its optimization objectives and the specific characteristics of the dataset. The Stable Diffusion model, despite a higher

FID score, offers significant advantages in terms of creative versatility and control, making it a valuable tool for tasks prioritizing these aspects. Based on our results, StyleGAN2-ADA generated African wax designs diverse in colors, shapes and details with some symmetry and repetition. Stable Diffusion was stronger with symmetry and repetition, but it generated less details. The choice among these models should be guided by the specific requirements of the task, whether it is the faithful reproduction of a dataset, creative image generation, or a balance between the two.

V. Conclusion and Future Work

This paper explored and compared the generation of African wax designs using StyleGan2-ADA, StyleGAN3 and Stable Diffusion architectures on a curated dataset of 2000 1024x1024 images generated with DALL·E 2. Our work is documented at [31].

StyleGAN2-ADA generated designs diverse in colors, shapes and details with some symmetry and repetition. Stable Diffusion was stronger with symmetry and repetition, but it generated less details. StyleGAN3 did not perform well on the dedicated task. The generated wax designs were evaluated using FID. While FID makes a lot of sense when used for StyleGAN, it is not the most adapted metrics for the nature of Stable Diffusion. We also evaluated the generated images qualitatively within the research team and based on our experience. DALL·E is a general-purpose image generation system; we focused on a very specific problem of wax pattern generation and obtained a higher level of details and, sometimes, symmetry. While we presented only a selection of the images, more images in the FID-range could have been provided to support our case. We plan to involve subject matter experts to refine the results, including by creating a significant synthetic dataset of African wax designs from the generated images to be used in the fashion sector. Fashion is a widely studied area of AI. This work has potential to impact fashion, one, by focusing on African fashion applications outside the scope of today's mainstream fashion industry, and, second, by providing designers with a new tool for creating customizable wax designs. In addition, we used a synthetic set of images obtained with DALL·E 2, but plan to work with wax print textile manufacturers on their wax patterns to compare the results.

References

[1] T. Karras, T., S. Laine and T. Aila. (2018). A Style-Based Generator Architecture for Generative Adversarial Networks. ArXiv. https://doi.org/10.48550/arXiv.1812.04948

[2] T. Karras et al., "Analyzing and Improving the Image Quality of StyleGAN," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020 pp. 8107-8116. doi: 10.1109/CVPR42600.2020.00813

[3] A. Borji. (2022). Generated faces in the wild: Quantitative comparison of stable diffusion, midjourney and dall-e 2. arXiv preprint arXiv:2210.00586.

[4] J. Ho, A. Jain and P. Abbeel. 2020. Denoising diffusion probabilistic models. In Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS'20). Curran Associates Inc., Red Hook, NY, USA, Article 574, 6840–6851.

[5] T. Karras, M. Aittala, S. Laine, E. Härkönen, J., Hellsten, J. Lehtinen and T. Aila. (2021). Alias-Free Generative Adversarial Networks. ArXiv. https://arxiv.org/abs/2106.12423

[6] J. Song, C. Meng and S. Ermon. (2020). Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502.

[7] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, E. and O. Wang. (2018). The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In Proceedings - 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2018 (pp. 586-595). Article 8578166 (Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition). IEEE Computer Society. https://doi.org/10.1109/CVPR.2018.00068

[8] S. Aydore, Q., Zhaozhi and M. Van Der Schaar. Synthetic Data Generation with Generative AI. Workshop at NeurIPS 2023, December 2023.

[9] C. Scharff, J. Brusseau, K. M. Bathula, Kaleemunnisa, S. M. Meshram, O. Gaikhe. Inclusion Ethics in AI: Use Cases in African Fashion. AAAI-24 Spring Symposium on Increasing Diversity in AI Education and Research.

[10] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio. (2014). Generative Adversarial Networks. ArXiv. https://doi.org/10.48550/arXiv.1406.2661

[11] S. S Chen, H. Cui, P. Tan, X. Sun, X., Y. Ji and H. Duh. (2020). Cantonese porcelain image generation using user-guided generative adversarial networks. IEEE Computer Graphics and Applications, 40(5), 100–107. https://doi.org/10.1109/mcg.2020.3012079

[12] N. Matsumura, H. Tokura, Y. Kuroda, Y. Ito and K. Nakano. (2018). Tile art image generation using conditional generative adversarial networks. 2018 Sixth International Symposium on Computing and Networking Workshops (CANDARW). https://doi.org/10.1109/candarw.2018.00047

[13] S. S. Nasrin and R. I. Rasel. (2020). Hennagan: Henna art design generation using deep convolutional generative Adversarial Network (DCGAN). 2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE). https://doi.org/10.1109/wiecon-ece52138.2020.9398005

[14] S. Shahriar and K. Hayawi. (2022). NFTGAN: Non-fungible token art generation using generative adversarial networks. 2022 7th International Conference on Machine Learning Technologies (ICMLT). https://doi.org/10.1145/3529399.3529439

[15] Y. Yin, Z. Chen, Y. Zhao, J. Li and K. Zhang. (2020). Automated Chinese seal carving art creation with Ai Assistance. 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). https://doi.org/10.1109/mipr49039.2020.00086

[16] H. Zhu. Y. Cao, H. Jin, W. Chen, D. Du, Z. Wang, S. Cui and X. Han. (2020). Deep Fashion3D: A Dataset and Benchmark for 3D Garment Reconstruction from Single Images. ArXiv. https://doi.org/10.48550/arXiv.2003.12753

[17] X. Zou, X. Kong, W. Wong, C. Wang, Y. Liu and Y. Cao. (2019). FashionAI: A Hierarchical Dataset for Fashion Understanding. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 2019, pp. 296-304, doi: 10.1109/CVPRW.2019.00039.

[18] A. H. Raffiee and M. Sollami. (2020). GarmentGAN: Photo-realistic Adversarial Fashion Transfer. ArXiv. https://doi.org/10.48550/arXiv.2003.01894

[19] S. Zhu, S. Fidler, R. Urtasun, D. Lin and C. C. Loy. (2017). Be Your Own Prada: Fashion Synthesis with Structural Coherence. ArXiv. https://doi.org/10.48550/arXiv.1710.07346

[20] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen and T. Aila. (2020). Training Generative Adversarial Networks with Limited Data. ArXiv. https://doi.org/10.48550/arXiv.2006.06676

[21] P. Dhariwal and A. Nichol. (2021). Diffusion models beat gans on image synthesis. Advances in neural information processing systems, 34, 8780-8794.

[22] R. Rombach, A. Blattmann, D. Lorenz, P. Esser and B. Ommer. (2022). High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10684-10695).

[23] D. P. Kingma and M. Welling. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.

[24] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler and S. Hochreiter. (2017). GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. ArXiv. https://doi.org/10.48550/arXiv.1706.08500

[25] E. J. Nunn, P. Khadivi and S. Samavi. (2021). Compound Frechet Inception Distance for Quality Assessment of GAN Created Images. ArXiv. https://doi.org/10.48550/arXiv.2106.08575

[26] M. Soloveitchik, T. Diskin, E. Morin and A. Wiesel. (2021). Conditional Frechet Inception Distance. ArXiv. https://doi.org/10.48550/arXiv.2103.11521

[27] C. Szegedy et al. (2015) Going deeper with convolutions. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015 pp. 1-9. doi: 10.1109/CVPR.2015.7298594

[28] J. Sohl-Dickstein, E. À. Weiss, N. Maheswaranathan and S. Ganguli. 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37 (ICML'15). JMLR.org, 2256–2265.

[29] K. Simonyan and A. Zisserman. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. http://arxiv.org/abs/1409.1556

[30] N. Ruiz, Y. Li, V., Jampani, Y. Pritch, M. Rubinstein and K. Aberman. (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).

[31] Scharff, C., Meshram, S. R., Bathula, K. M., Kaleemunnisa and Gaikhe, O. https://github.com/researchpace/waxfashion.