COMPUTATIONAL CREATIVITY: GENERATIVE FASHION DESIGN WITH AI-DRIVEN SYSTEM

Team Project

Submitted: April 3rd, 2023

By: Jiaxin Yuan, 1820345
Yunjing Dai, 1820527
Bennur Kaya, 1754738
Tolga Yasar, 1822718
Yi Wang, 1870245
Yiyi Wei, 1873027

Supervisor: Deborah Mateja
Reviewer: Prof. Dr. Armin H. Heinzl

University of Mannheim
Chair of General Management and Information Systems
68131 Mannheim
Phone: +49 (0) 621 181 1691, Fax: +49 (0) 621 181 1692
Homepage: https://www.bwl.uni-mannheim.de/heinzl/

Abstract

The fashion industry has seen a significant shift towards personalized and creative clothing production, leading to a growing demand for effective apparel design tools. In response, our Generative Fashion Design project was developed to enable users to realize their design ideas more effectively. Our project employs Stable Diffusion and DreamBooth as the primary model and algorithm for generating fashion designs, with a focus on exploring the boundaries of models generated by the DreamBooth method. The implementation of DreamBooth in the fashion domain is described in detail, including the underlying knowledge and the customization of the approach for generating creative clothing images. A prototype that is both user-friendly and comprehensive was created to streamline the fashion design process. Additionally, a testing method that is thorough and exhaustive was developed and will be presented. While the generative model has demonstrated promising results in generating innovative designs, certain limitations were identified, and future research directions were proposed. To sum up, fine-tuning a text-to-image model using DreamBooth has the potential to serve as a valuable tool for fashion designers, enhancing their creativity, productivity, and workflows.

Table of Contents

List of Figures

1 Introduction

The clothing market has undergone an Industrial Revolution, providing consumers with an extensive range of clothing options, leading to a growing demand for personalized and creative clothing production (Helen Lewis Brockman, 1967). In light of these developments, our project, Generative Fashion Design, can prove useful in enabling users to realize their apparel design ideas more effectively.

In Chapter 2, we introduce key concepts and terminology related to generative models for image synthesis, such as Generative Adversarial Networks (GANs) (Goodfellow et al., 2020) and Diffusion Models (DMs) (Sohl-Dickstein, Weiss, Maheswaranathan, & Ganguli, 2015). We then discuss Stable Diffusion, an open-source text-to-image model, and its fine-tuning techniques, including Textual Inversion and DreamBooth, explaining their definitions, working mechanisms, benefits, and drawbacks. The passage also explains why Stable Diffusion and DreamBooth were selected as the primary model and algorithm for the generative fashion design task, based on their suitability for the use case and superior performance compared to alternative techniques.

In Chapter 3, we will discuss the user requirements research that we conducted to ensure the functional accuracy of the model. This is divided into two main stages. Firstly, we made reasonable adjustments and improvements to the range of interviewees based on realistic factors and feedback, and secondly, we conducted parallel and vertical comparative analysis of the feedback results, which helped us to better understand user needs and lay the foundation for the feasibility of our model.

Chapter 4 provides a detailed description of the implementation of the DreamBooth method for training models in the fashion domain. It covers the key components of this approach and how it was customized for generating creative clothing images. A prototype was also developed to simplify the method for fashion designers. The chapter also discusses the model evaluation process, which includes a testing strategy that covers multiple dimensions and adheres to the MECE principle (Lee & Chen, 2018). The testing results confirm that the model meets the required standards for the task.

We discuss the limitations and future work for the AI generative model used in fashion design in Chapter 5. While the model showed promising results in generating innovative and unique fashion designs, the study identified several limitations, including reduced performance when introduced to more than eight concepts, occasional generation of unrelated examples, and some imperfections in the generated images. Additionally, the model struggles to maintain consistency in applying fashion elements across different parts of the fashion piece and cannot render legible text. To improve the system, future research could incorporate more diverse training data, include constraints and guidelines, conduct further user studies, explore multi-modal designs, and improve the stability of the model's performance. These efforts could result in a more creative and practical design solution for fashion designers and enthusiasts.

In conclusion, our study aimed to investigate the potential of utilizing AI generative models in the field of fashion design. We conducted a comparative analysis of several models and ultimately selected fine-tuning on Stable Diffusion as the optimal approach to meet the needs of fashion designers. Our findings revealed promising capabilities of the model to comprehend diverse fashion elements, generate unique and innovative designs, and integrate different elements and inspiration concepts. Additionally, we identified certain limitations and proposed future research directions in this area. In summary, we conclude that fine-tuning a text-to-image model using DreamBooth has the potential to serve as a valuable tool for fashion designers, enhancing their creativity and productivity, streamlining their workflows, and saving time.

2 Model Selection Exploration

In this section, we initially present some fundamental concepts and terminology. Subsequently, we delve into the rationale behind selecting Stable Diffusion and DreamBooth as the primary model and algorithm for our generative fashion design task.

2.1 Theoretical Foundations

Comprehending the following terminology is of great importance. In this subchapter, we will illustrate generative models specifically for image synthetic. Then we will explore some of their examples, such as Generative Adversarial Networks (GANs) and Diffusion Models (DMs). Additionally, we will introduce an open-source text-to-image model, Stable Diffusion, along with its fine-tuning techniques such as Textual Inversion and DreamBooth.

Each part will cover the following questions:

- What they are
- How they work
- The benefits and drawbacks associated with their use

2.1.1 Generative models

In contrast to discriminative models, Generative models represent another category of machine learning model that aims to learn and generate new data that exhibits characteristics similar to the training data (Ng & Jordan, 2001). These models can generate new data points that share similar features as the original data and are utilized in various fields, including image and audio processing, natural language processing, and robotics. Generative models function by learning the underlying probability distribution of the input data and then utilizing this knowledge to generate new data points (Creswell et al., 2018). In our situation, we focus specifically on the use case of generating images. Image generation models use complex algorithms to learn patterns and features in existing images and use this knowledge to create new, unique images. Among the various types of generative models, Generative Adversarial Networks (GANs) is one of the most commonly mentioned ones (Goodfellow et al., 2020). In the current state of the art, a majority of the image synthesis models use either a variant of a particular model or a combination of multiple models.

2.1.2 Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) have made remarkable strides in machine learning, particularly in the field of image synthesis. A GAN is a generative model that simultaneously trains two models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G. The training procedure for G is to maximize the probability of D making a mistake. This framework corresponds to a zero-sum game. In the space of arbitrary functions G and D, a unique solution exists, with G recovering the training data distribution and D equal to a half everywhere (Goodfellow et al., 2020).

There exist numerous GAN variations that are widely utilized in computer vision. Some of these have been applied to fashion image generation, such as CAGAN (Jetchev & Research, 2017), which facilitates the swapping of clothing on fashion model photographs, and Attribute-GAN, which investigates clothing matching problems under the cGAN framework and generates clothing images based on semantic attributes, respectively (Liu, Zhang, Ji, & Jonathan Wu, 2019). Additionally, another GAN explores symmetry of generated fashion images by enhancing DCGAN (Makkapati & Patro, 2017; Radford, Metz, & Chintala, 2015), while Poly-GAN generates clothing images conditioned on arbitrary human poses (Pandey & Savakis, 2020).

While these GAN models demonstrate the impressive potential of GANs in fashion image generation, none of them are capable of generating novel and creative fashion designs with artistic merit. Furthermore, while GANs enable efficient sampling of high resolution images with good perceptual quality (Brock, Donahue, & Simonyan, 2018), they pose optimization challenges and are unable to fully capture the distribution. In addition, the constant competition between the generator and discriminator networks in a GAN can lead to instability and slow training (Karras et al., 2020).

2.1.3 Diffusion Models(DMs)

Diffusion Models (DMs) (Sohl-Dickstein, Weiss, Maheswaranathan, & Ganguli, 2015), also known as Diffusion Probabilistic Models, represent another prominent class of generative models. The goal of diffusion models is to learn the latent structure of a dataset by modeling the way in which data points diffuse through the latent space. In computer vision, DMs train a neural network to denoise images blurred with Gaussian noise by learning to reverse the diffusion process (Song et al., 2021; Gu et al., 2022). More specifically, DMs are classified as a type of latent variable model that leverages a fixed Markov chain to map to the latent space (Jonathan, Ajay, & Pieter, 2020).

An important advantage of DMs is their ability to obviate explicit density calculations, thereby increasing their computational efficiency relative to other generative models, such as GANs. Rather than rely on explicit density calculations, DMs utilize a sequence of learned transformations to map a simple distribution to a more complex one.

Despite these advantages, DMs remain computationally demanding due to the need for repeated function evaluations and gradient computations in the high-dimensional space of RGB images (Rombach, Blattmann, Lorenz, Esser, & Ommer, 2022).

2.1.4 Latent Diffusion Models(LDMs)

Latent Diffusion Models (LDMs) are a modification of Diffusion Models (DMs) that operate the diffusion process in a lower-dimensional latent space generated by an autoencoder instead of in the high-dimensional pixel space. By focusing on the important, semantic bits of the data, likelihood-based generative models can be trained more efficiently in this space, resulting in lower training costs and faster inference speeds (Rombach et al., 2022)

In the domain of text-to-image synthesis, Stable Diffusion represents the official implementation of LDMs. Compared to other state-of-the-art models, such as OpenAI's DALL·E 2 and Google's Imagen, Stable Diffusion shows similar performance but is more accessible and flexible because it is open-source.

2.1.5 Fine-tuning Methods for Text-to-image Models

While large text-to-image models, such as those mentioned above, have achieved remarkable results in generating high-quality and diverse images from text prompts, they often lack the ability to accurately mimic the appearance of specific subjects in a given reference set and synthesize novel renditions of them in different contexts (Ramesh et al., 2022; Saharia et al., 2022; Rombach, Blattmann, Lorenz, Esser, & Ommer, 2022). To address this limitation, fine-tuning methods such as Textual Inversion and DreamBooth have been proposed. These techniques enable the model to be fine-tuned on a specific reference set, allowing it to learn and mimic the appearance of the subjects within that set and generate new, context-specific images (Ruiz et al., 2022; Gal et al., 2022).

Textual Inversion

Textual Inversion, implemented on Language Models (LDMs), is a technique utilized to extract novel concepts from a limited number of sample images (Voronov, Khoroshikh, Babenko, & Ryabinin, 2023).. Its primary objective is to enable better control of text-to-image pipelines. This technique involves acquiring new "words" in the embedding space of the text encoder employed by the pipeline. These words can subsequently be incorporated into textual prompts to exercise greater control over the final images produced by the system.

However, the approach has certain limitations, including the inability to learn precise shapes and instead focusing on capturing the "semantic" essence of a concept. Another significant challenge is the extensive training time associated with this methodology, with the learning of a single concept requiring approximately two hours, which can be prohibitively long in certain scenarios (Gal et al., 2022).

DreamBooth

Given a limited number of images featuring a particular subject (typically around 3 to 5), DreamBooth is to embed the subject into the output domain of the model while associating it with a unique identifier. This technique utilizes rare token identifiers to represent the subject and fine-tune a pre-existing text-to-image framework that utilizes a diffusion-based approach. This framework operates in two distinct steps, first generating a low-resolution image from textual prompts and subsequently employing superresolution (SR) diffusion models to enhance the image quality. involves fine-tuning the low-resolution text-to-image model using the available input images, along with textual prompts that contain a unique identifier followed by the subject's class name (e.g., "A [V] dog") (Ruiz et al., 2022).

In contrast to Textual Inversion, which solely trains the embedding without any alterations to the base model, Dreambooth implements fine-tuning of the entire text-to-image model. This approach involves the acquisition of the capacity to associate a distinct identifier with a specific concept, whether it be an object or style. As a result, the generated images are tailored to a greater degree to the specific object or style in question, thus facilitating a more personalized output compared to the results obtained through Textual Inversion (Voronov et al., 2023).

2.2 Final decision

From among the various available techniques, we selected Stable Diffusion and fine-tuned it using the DreamBooth approach. Our decision to use this particular technique was based on a number of factors, which we will discuss in detail in the subsequent section. Ultimately, we determined that this approach was most appropriate for our particular use case and yielded superior results compared to other competing techniques.

2.2.1 Why Stable Diffusion?

After a thorough evaluation of our task requirements, which took into account the specific needs of fashion designers, the feasibility of the operation, the significant research support available, and the distinct advantages of Stable Diffusion over its competitors, we devoted a considerable amount of time to reaching a decision regarding its implementation.

Our objective is to create a machine learning (ML) based system for generative fashion design. Following an extensive interview with fashion designers, we discovered that a major challenge they face is finding inspiration during their work. Fortunately, computer vision has been a trending topic for some time now, and several algorithms have demonstrated excellent performance in this field. Among them, Generative Adversarial Networks (GANs) have been recognized as one of the most popular image synthesis algorithms and were therefore our primary consideration. However, GANs are known to have limitations in modeling complex, multi-modal distributions, making it difficult to generate clothes as described in natural language, especially for data with a high degree of variability (Brock et al., 2018; Karras, Laine, & Aila, 2019). Subsequently, we conducted several experiments on the proposed approach, which revealed that the training process posed certain challenges. Specifically, due to the high resource requirements, we were compelled to input low-resolution images and limit the number of epochs, resulting in unsatisfactory outcomes. Meanwhile, although applications such as DALL·E 2 or Imagen have demonstrated exceptional results, we found that the majority of the images generated did not meet our aesthetic and design requirements, specifically with respect to displaying adequate details of silhouette, color, texture, and overall design (Eckman & Wagner, 1995).

In addition, the absence of access to DALL·E 2's parameters and code presents a significant challenge to optimize and implement it in our situation. Initially, we considered using DALL·E 2's approach and inputting clothing data to train our own model until we encountered Stable Diffusion. We discovered that Stable Diffusion is an open-source alternative to DALL·E 2 that produces similar performance in the tasks we prioritize. Furthermore, the kernel model of Stable Diffusion, Latent Diffusion Models (LDMs), operates on a compressed latent space with lower dimensionality, enabling computationally less expensive training and faster inference with almost no loss in synthesis quality. This effectively addresses our resource constraint issue. Therefore, we opted for Stable Diffusion as the preferred choice, considering its practicality and versatility in effectively addressing our specific tasks.

2.2.2 Why DreamBooth?

In order to enhance the performance of the current text-to-image model to meet our needs, we extensively researched various fine-tuning methods and ultimately determined that DreamBooth would be the most suitable choice. This method has demonstrated promising results and the capability to generalize effectively.

Despite demonstrating satisfactory results in image generation, Stable Diffusion has faced similar issues as DALL·E 2 when applied in the fashion domain. Specifically, the model has produced poor results such as twisted human faces, strange colors, and a lack of aesthetics or novelty that do not meet the requirements of fashion designers. Moreover, in the context of fashion design, designers may face difficulties in articulating their visual inspirations using conventional verbal descriptions, and there may be variations in perception and interpretation of the model that generate unexpected results. To address the challenges related to integrating visual inspirations into fashion designs using the Stable Diffusion model, we posit that the utilization of fine-tuning methods can allow designers to incorporate their visual inspirations directly into their designs. By inputting multiple images of a single concept into the Stable Diffusion model, designers can extract specific features, colors, patterns, or textures and then transfer to any fashion piece, thus stimulating creativity and aiding in the design process. This approach can enhance the efficiency and effectiveness of the design process while facilitating the integration of visual inspiration into the final design product.

After thorough experimentation and research, we have selected DreamBooth as our preferred method for fine-tuning Stable Diffusion, over Textual Inversion. Although these methods share similarities in their approach and only require 3-5 images of a user-provided concept as input, DreamBooth proved to be more effective in our experiments. In particular, DreamBooth was able to overcome the significant challenge of Textual Inversion, which is the prolonged training time, making it a more suitable option for our needs.

Overall, DreamBooth produced superior results in our experimentation, which led to our decision to utilize this method for our project. As a consequence, we have opted to employ the DreamBooth method to aid us in constructing our models, which can be utilized in the synthesis of clothing images.

3 User-oriented Requirements

In this section, we outline the approaches we employed to gather user requirements and the methods we employed to analyze them for our project. We begin by elaborating on our three interview designs, highlighting their objectives, significance, and distinctions, as well as explaining their implementation. We subsequently extract relevant information from the interviews to be utilized in subsequent stages.

3.1 User Research Design and Implementation

We conducted three interviews with different objectives. To acquire a comprehensive understanding of user needs, expectations, and preferences, the first interview was a face-to-face session with a professional fashion designer. The purpose of this interview was to identify the challenges encountered by fashion designers and establish well-defined project objectives. The survey questions were primarily geared towards eliciting information on the details of the design process and identifying pain points experienced during the process. During the interview, the participants emphasized the significance of gaining inspiration in the design process.

After training our model and developing an easy-to-use prototype, we conducted the last two interviews to obtain feedback from a user perspective. Initially, considering the niche nature of fashion, we targeted only professional fashion designers who had been in the industry for at least 5 years after graduating from a fashion major. However, according to the feedback we received, most professional fashion designers had reservations about our AI-assisted inspired creation model, stating that AI-generated images are incomparable to their professional designs. Additionally, they did not provide enough useful information in terms of model improvement. As a result, we expanded our audience to fashion design enthusiasts who have an interest in fashion designing and are willing to design their own clothes for real use. This extension helped us gather more useful information concerning our project. In the next subsection, we will provide a detailed explanation of these last two interviews and the insights we gained from them.

3.2 Interviews

The last two interviews were crucial as they provided us with important insights for improving our model. We will discuss the details of the questions we designed, their implementation, and the information we obtained from them individually.

3.2.1 Interviews with Fashion Designer

With the objective of implementing our model in real-world applications and enhancing the performance of clothing image synthesis, we conducted a series of interviews with several professional fashion designers. Each interview was allotted 30 minutes, during which we initiated a discussion regarding their design process to gain insights into the potential applications of our model. Our questionnaire has been structured into three parts: the first part solicits basic information from the interviewees; the second part pertains to the design process, and seeks to elicit responses regarding the challenges faced by the interviewees in this regard; and the third and final part seeks feedback on the generated images. In relation to feedback, we requested the designers to test our prototype by utilizing it to generate new images, and we recorded their feedback pertaining to the prompts, generated images, and their overall perception of the results. In the event that the generated image failed to meet their expectations, we requested them to repeat the process until they were content with the output. These interviews enabled us to acquire valuable feedback on the efficacy of our model in generating design concepts and improving the clothing image synthesis process. The insights gleaned from these discussions helped us identify key areas for improvement, such as expanding the scope of customization options and enhancing the accuracy of the image synthesis process.

It is notable that the responses gathered from the interviews with professional fashion designers highlighted "color matching" and "shape of the clothes" as the primary areas of focus in the fashion design process. This information prompted us to prioritize these elements when generating images, emphasizing their critical importance.

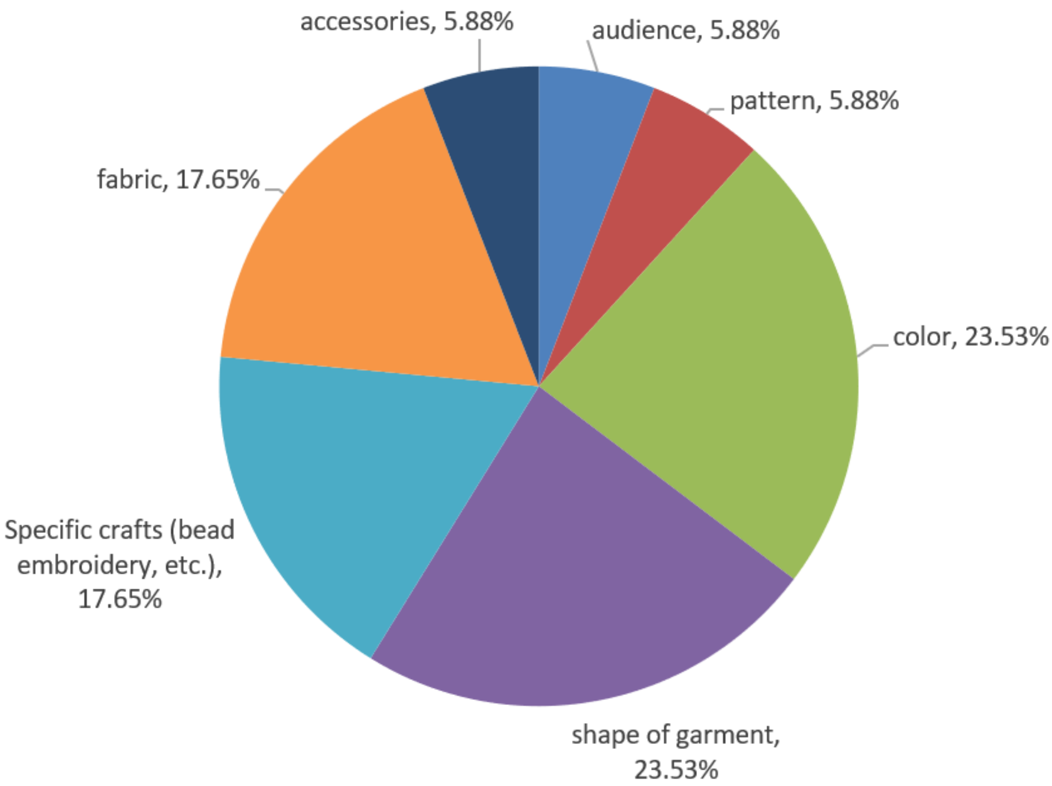


Figure 1. List of Design Elements Considered by Professionals

An additional noteworthy finding is that while 50% of the professional fashion designers surveyed were familiar with AI-driven Image Synthesis systems, none of them had any prior experience with AI assistants, nor expressed willingness to use such systems in the future. In contrast, fashion enthusiasts were found to be more receptive. However, the feedback provided by the professional fashion designers was mostly negative, with a particular emphasis on color combinations not being harmonious and the design styles being uncreative. These designers expressed that the results were of little assistance in stimulating their creativity and had reservations about the project's potential application in the professional field.

3.2.2 Interviews with Fashion Enthusiasts

Interviews with fashion enthusiasts can be regarded as an extension aimed at gathering sufficient information to enhance the performance of our project. Similar to the process used for fashion designers, we also employ questionnaires and testing with this group. However, in this case, our focus is primarily on eliciting feedback from them, considering their limited technical expertise. Therefore, we meticulously document every feedback received for each step involved in generating synthetic images. Their primary challenge lies in their difficulty visualizing amorphous and unstructured concepts in their minds, which our project can effectively address. Consequently, they have exhibited favorable dispositions towards AI-assisted clothing generation models.

Non-professionals and professionals share a common consideration for color matching during the design process. However, non-professionals place a greater emphasis on the presentation of patterns and colors, while professionals tend to overlook the significance of patterns and focus more on practicality and marketability of their designs. Thus, it is crucial to strike a balance between creativity and practicality to cater to both non-professionals and professionals.

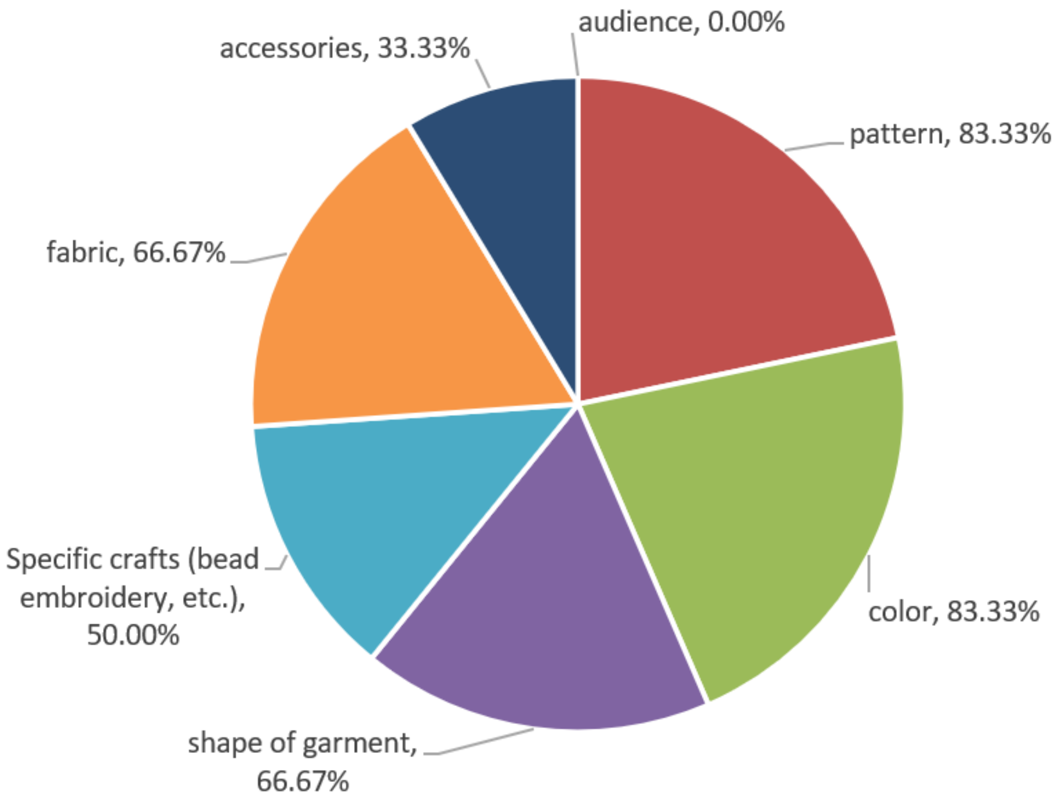


Figure 2. List of Design Elements Considered by Enthusiasts

As previously mentioned, all fashion enthusiasts express a positive attitude towards our model. They believe that creativity is the most crucial aspect of fashion design, and they typically seek inspiration from reading books and admiring paintings. Inspiration can generally be categorized into two types: text-based and image-based. Moreover, they mentioned that lay people still face difficulty in visualizing the output of ambiguous concepts solely through their own efforts, and our model can effectively address this issue, although further optimization could enhance the quality of results.

Figure 3. Enthusiasts Versus Professionals about Design Elements Considering

Overall, the interviews conducted yielded significant insights that informed our focus on key areas for improvement and the efficacy of the model in generating design concepts and enhancing the clothing image synthesis process. Based on feedback from fashion enthusiasts, it became apparent that properly transferring inspiration into fashion pieces and creatively combining them in novel ways is of utmost importance. Thus, this will serve as the main objective for our project going forward.

4 Implementation

This section provides a detailed account of the implementation of the DreamBooth method in training our models. The key components of this approach will be discussed, along with our customized implementation for the fashion domain based on the original DreamBooth paper. Additionally, we developed a prototype that simplifies the method for fashion designers, allowing them to test it through interviews. Finally, we will discuss our model evaluation process, which confirms that our model meets the required standards for our task.

4.1 Training

We utilized the strategy proposed by DreamBooth for fine-tuning our selected clothing images. The objective of our training process is to embed a given subject, consisting of a small set of images (usually 3-5), such as inspirations (e.g., paintings, people, nature, architecture, etc.) or fashion pieces, into the output domain of a model so that it can be synthesized with a unique identifier. This enables us to create customized clothing designs by combining any of the input clothing items or inspirations, with the output controlled by natural language. The training process involves two stages: generating a low-resolution image from text, followed by applying it to high-resolution diffusion models.

We executed this process on a latent text-to-image diffusion model called Stable Diffusion. Specifically, our experiments revealed that the Stable-Diffusion-v1-5 model provided the best results as a base model.

In this section, we will begin by discussing the preliminaries that are necessary to understand the underlying process. Following this, we will present how we implemented the DreamBooth method in practice.

4.1.1 Preliminaries

To provide a comprehensive understanding of the training process, it is necessary to explain the following prior knowledge and emphasize some technical details used in our case.

Latent Diffusion model

As discussed before, the Latent Diffusion Models have been proposed as a solution to reduce computational demands and speed up the inference of diffusion models while maintaining high synthesis quality. The proposed method firstly employs an encoder that learns a space, perceptually equivalent to the image space but with reduced dimensions. Subsequently, a standard Diffusion Model is designed to work with the two-dimensional structure of our learned latent space. Diffusion Models are probabilistic models intended to learn a data distribution p(x) through gradual denoising of a normally distributed variable. This process corresponds to learning the reverse of a fixed Markov Chain of length T.

In our case, a conditional U-Net has been utilized, which takes a noisy sample, a conditional state, and a timestep, and outputs a sample shaped output. The output of the U-Net is later used by a decoder to generate the final images as the output.

Vocabulary Encoding

This process involves transforming a prompt and an image into a vector that can be directly computed. Since text and images belong to different distributions, direct computation of the two together is not feasible. Therefore, considerable research has been conducted in this area. CLIP (Contrastive Language-Image Pre-training) leverages prior work on zero-shot transfer, natural language supervision, and multimodal learning to provide optimized pairs of images and text for us.

In our study, the pre-trained model we used was initialized with the weights of the last version checkpoint and fine-tuned on "laion-aesthetics v2 5+", which took CLIP Image embeddings produced with the OpenAI CLIP model as input, ensuring high-quality visual images and semantic fidelity. After conducting numerous tests, we discovered that the pre-trained CLIP model had integrated almost all fashion-related terminology, including fashion brands such as Gucci and Chanel, as well as fashion-related phrases like jumpsuit, joggers, and woven, among others. As a result, we concluded that there was no need to fine-tune the CLIP model. Moreover, we will use the new subject name and prompt in a standardized CLIP tokenization method.

4.1.2 Training Customized Concept

The customization method offers fashion designers the ability to seamlessly integrate their specific creative ideas into fashion pieces.

Through fine-tuning a pre-trained stable diffusion model using DreamBooth, designers can achieve this by providing just 3-5 images of a particular item and assigning it a unique identifier. This process is akin to adding a few new words into the basic model, teaching it to recognize and replicate the unique design elements of the identified item. Our model enables fashion designers to generate possible garment designs by utilizing human language prompts. Additionally, it allows fashion designers to work with a collection of high-quality fashion pieces that have been incorporated into the model. Furthermore, the model offers extensive customization options, empowering designers to mix and match components and iterate on their designs, ultimately allowing for more streamlined and creative workflows based on their previous work.

To emulate the process of fashion design, we introduced specific fashion pieces, including tops, bottoms, shoes, and accessories, as well as inspirations such as figures, designs, human faces, paintings, natural scenery, and even random subjects like cartoons or pillows. Each introduced item is associated with a concept, either a fashion piece concept or an inspirational concept. By referring to the names we have given them, we can easily retrieve and combine them with other elements.

Through the use of our model, fashion designers can preview their designs in advance, which can aid in identifying potential issues and adjusting details. This approach can ultimately reduce future risks and improve overall efficiency compared to traditional design processes.

4.2 Prototype

In this chapter, we will discuss the prototype developed for our project, which provides a Stable Diffusion model capable of generating photo-realistic images based on a text prompt. However, what sets our prototype apart is that it is not limited to being used solely by fashion designers as a machine learning model. It also allows them to actively participate in the training process. The true power of machine learning models comes from this fine-tuning method, whereby the underlying Stable Diffusion model can be trained on the designers' own fashion pieces, inspirations, and styles, based on real parameter optimization.

Our prototype is a python-based web application developed using the Gradio library. It has several key functionalities. Firstly, a user can train the Stable Diffusion model based on up to five different concepts, either by selecting a model that has not been trained on any concepts or by choosing a model that has been pre-trained with the protoype on other concepts. The prototype logs the training of each concept, as seen in *Figure 4*. If an already pre-trained model is selected for retraining, the new concepts are added to the existing ones, and the prototype displays each concept's name and class (examples for classes: shoe, t-shirt, inspiration, etc.).

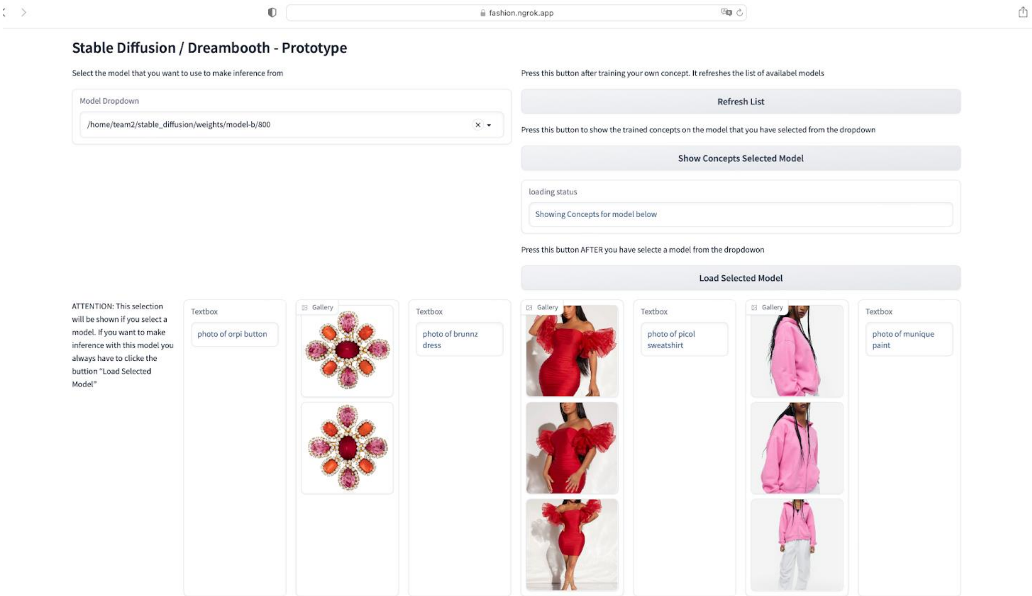


Figure 4. Prototype-Show Concepts of a Pre-trained Model

Next, based on the selected pre-trained model, the prototype provides a prompt generator for the user, allowing them to generate prompts and select the information to be included in prompt to generate new images. This includes selecting a base piece, which can be any fashion piece (e.g. t-shirt, sweater, dress) on which the style is transferred, or a concept introduced by the fashion designer. The user can then decide from which other concepts which aspects should be transferred to the base piece by clicking on the respective boxes. Additionally, they can modify aspects of the base piece independently of the concept (e.g. changing the shape of the base piece to oversize). The generated prompt can be used to generate images directly or can be manually modified by the user if there are special cases not coverd by the prompt generator (Figure 5).

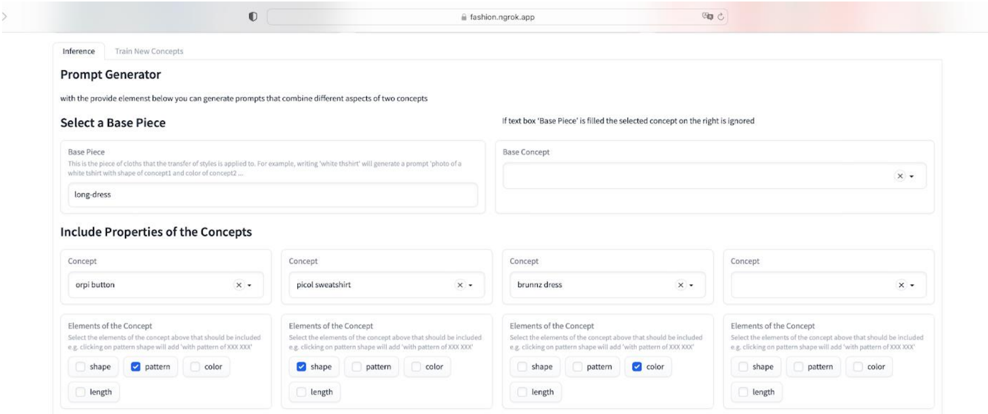


Figure 5. Prototype-Prompt Generator

The prototype also allows the user to generate images based on the prompt, with options for the number of images to be generated, the resolution of the output images, the number of model iterations (higher iterations lead to higher quality results but longer generation time), and some more details.

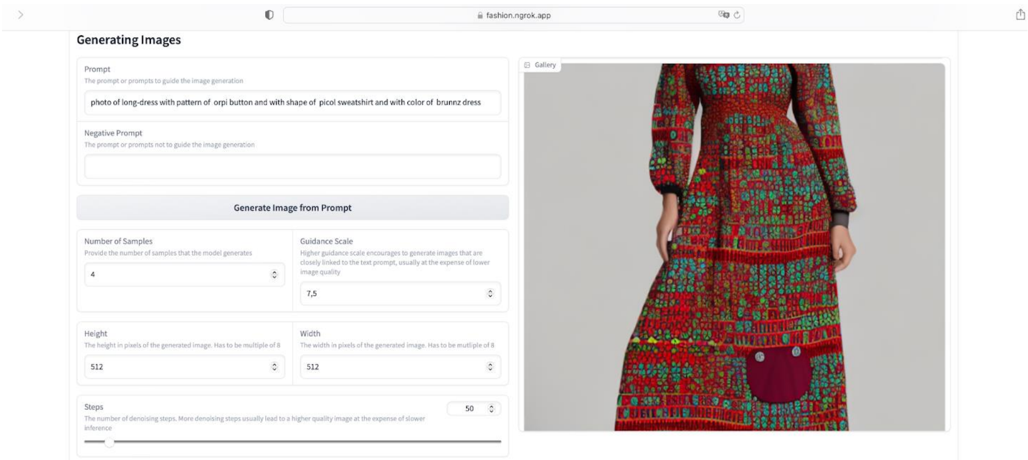


Figure 6. Prototype-Generating Images

Due to the randomness of the image generation process, if there is an image that the user likes but would like to change some aspects of it, the prototype allows the user to reiterate over the generated images. Using the same model, the user can guide the regenerative image generation with a new text prompt (e.g. change the color of the dress to blue. (Figure 7)

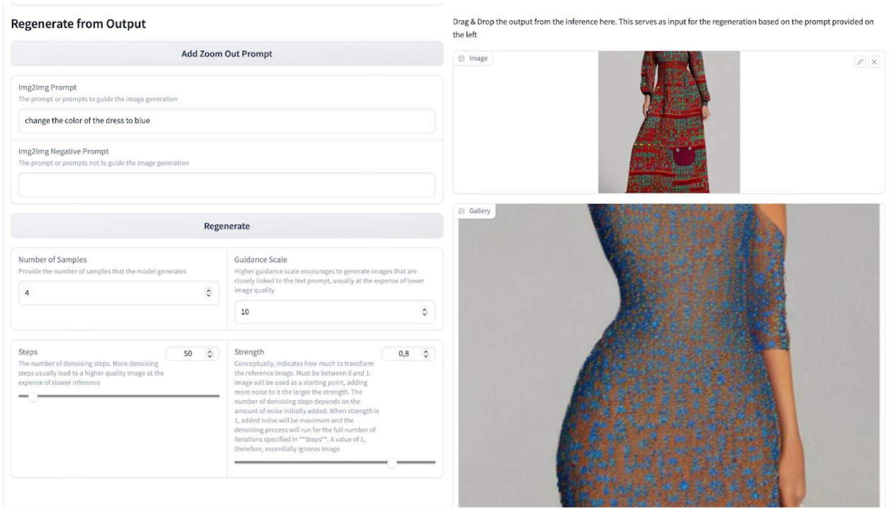


Figure 7. Prototype-Regenerate Images from Output

The prototype is currently accessible via the following *GitHub repository: https://github.com/tyasar53/stable_diffusion* under the branch prod. You can find a detailed usage guide that explains how to use the prototype in more detail with help full tips to achive better results, as well as installation instructions. A lot of guidance are baked into the visual elements of the application which makes it hard to include it into this report.

4.3 Testing

We conducted a comprehensive testing process following our established testing protocol to evaluate the limits and capabilities of our model. This allowed us to gain a better understanding of the model's performance and explore its potential for creating unique and innovative fashion designs.

4.3.1 Testing Strategy

The process of apparel design applies a particular kind of problem solving, consisting of a series of small steps (Cameron, 2009), which is why we create structures ——to help us break down concepts, so that the design process can be traced, interpreted and studied throughout the model (Cross, 2010; Simon, 1996). Based on the above mentioned (Lee & Jirousek, 2015), we introduced varying numbers of concepts to the model to determine the optimal number for achieving high-performance and satisfactory results. These concepts encompass any example we introduce to the model, with each comprising a set of 3-5 images that represent fashion pieces or inspirations. The testing structure is categorized under two main categories, namely Fashion Pieces and Fashion Elements. Fashion Pieces are divided into two classes, with the first representing the broader category of garments such as tops, bottoms, shoes, suits, and accessories, while the second class comprises specific fashion pieces falling under the first class. For instance, shirts and t-shirts belong to the "top" category, while trousers and jeans belong to the "bottom" category under the first class. Fashion Elements, on the other hand, are defined by five dimensions, namely shape, pattern, texture, color, and space, each of which has a unique role in fashion design. Shape refers to the silhouette or outline of a garment or fashion piece, including its contours, curves, and overall form. Pattern encompasses the decorative designs and motifs applied to a garment or fashion piece, such as stripes, polka dots, or florals. Texture describes the tactile quality or surface appearance of a garment or fashion piece, including the feel of the fabric and any details such as embroidery or beading. Color pertains to the hues, shades, and tones of a garment or fashion piece, as well as any color combinations or contrasts used. Space relates to the overall arrangement and distribution of the elements within a garment or fashion piece, including its proportions, balance, and negative space. Positive space pertains to the areas of interest or subject within the fashion piece, such as the silhouette or form of the garment, or the placement of accessories or embellishments. For example, the positive space in a dress design could be the placement of a unique neckline, an intricate pattern, or a striking color contrast. On the other hand, negative space in fashion design refers to the background or the areas surrounding the subject of the work. For example, the negative space created by the asymmetrical neckline shape is sensual and relaxing (Volpintesta, 2014).

In the Fashion Pieces section of our study, we will primarily focus on the capabilities of the model for understanding the features of the fashion piece, shifting within/across the different fashion piece classes, combining different concepts and applying inspirations on fashion pieces. On the other hand, the Fashion Elements category will offer greater opportunities for creativity as we aim to apply the different fashion elements of various products to other fashion pieces, as well as exploring diverse inspirations. This segment will involve a more detailed approach, as we combine the specific fashion elements of different introduced concepts to add unique and distinctive flavors to the fashion designs. Overall, by exploring different scenarios and analyzing the results, we aim to identify the strengths and weaknesses of the model and provide insights for future research and development in this area.

4.3.2 Testing results

Fashion Pieces

Within the Fashion Pieces category, we initially assessed the model's ability to comprehend various types of fashion pieces without the need for introducing any specific concepts. The pre-trained model was able to understand all the fashion pieces. In this model, we use a text-image model and fine-tune it so that it can learn to bind a unique identifier to that particular subject. This identifier can be used to generate realistic images in different scenes. This technique has many practical applications, for example, it can be used for product display in e-commerce, allowing consumers to visualize the product without having to try it on. To implement this technique, we use a semantic prior embedded in the model and a new autoclass prior to preserve the loss. This approach is able to maintain the authenticity of the image while preserving the features and identity of the subject during the synthesis process. Overall, our technique is able to synthesize subjects that do not appear in the reference image under different scenes, poses, perspectives and lighting conditions, which opens up new possibilities for the generation and design of fashion pieces (Ruiz et al., 2022). At the same time, we introduced one concept to evaluate the model's fundamental understanding of a single fashion piece. The model performed well for a single concept, accurately identifying the key features of the fashion piece (color, shape, pattern) and generating new designs inspired by it that were comparable but distinct [Image I-Swarovski button]. As emphasized in the official paper of Dreambooth model, precise selection of a unique identifier followed by the subject's class name (e.g., "A [V] button") significantly enhances the model's capacity to produce more creative and higher-quality results.

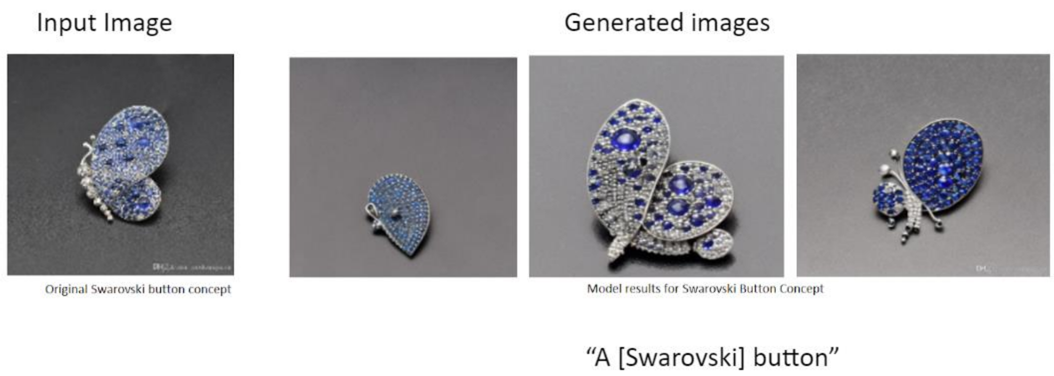


Figure 8. Subject-driven Generation. Given Swarovski (left), it shows good that our approach (right) can synthesize the "Swarovski" with high fidelity and in new contexts (text prompt: “A [Swarovski] button”)

Later on, we introduced various concepts to the model at different times, as this was a crucial step in developing innovative and unique fashion designs. Our findings revealed that by introducing more than eight concepts to the model, it was no longer able to grasp all the concepts and integrate them into designs effectively. For example, when we introduced 24 concepts to the model and called a particular dress in the prompt (i.e., bejflow dress), the model struggled to replicate the given concept and instead generated random dresses without considering the specific characteristics of the original concept. In all the examples presented in the following section, it should be noted that the model was trained with no more than eight concepts.

Following the initial assessment of the model's basic understanding of fashion concepts, we proceeded to evaluate its creativity capabilities in terms of fashion piece concepts and inspiration concepts. Our first objective was to determine if the model could effectively shift between and within classes for the introduced concepts. For example, we tested if the model could shift from a necklace to an earring, which are both in the same "accessory" category, and if it could create high-heeled boots similar to the features of a dress, which are in different categories (i.e., "suit" and "shoes"). The model performed exceptionally well in the within-class results, as demonstrated by Figure 9. However, while the model also produced examples which catches some features of the Orpi Button for the across-first-class, the results exhibited greater dissimilarity compared to transformations within the same class.

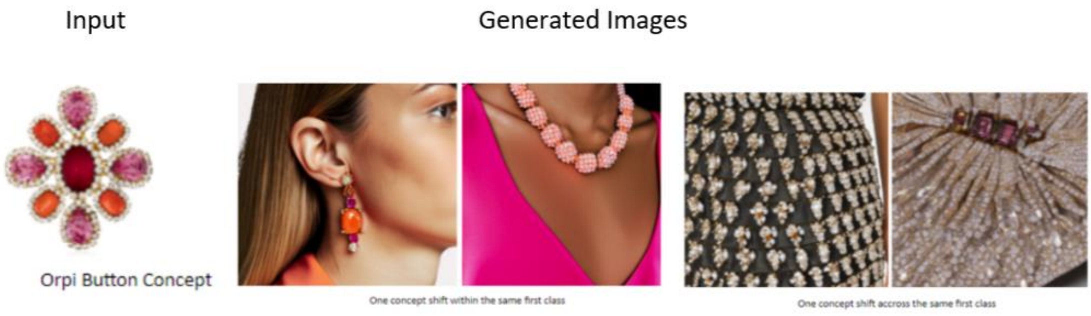


Figure 9. Novel View Compositions, Art Renditions and Property Modifications. We are able to generate novel and meaningful images while faithfully preserving the identity and essence of the subject during same-class, but we lose the main features of the subject during cross-class

In the final stage of our creativity assessment, we introduced inspiration concepts, which were non-fashion products, including art pieces, real-life products (e.g., pillows), patterns, human face photos and more. We applied these inspirations to both fashion pieces with and without introducing a concept, such as on both a [Y] bag or a simple bag. Our observations revealed that art pieces were applied to fashion products successfully. When the prompt began with the name of the introduced art piece and followed by the name of a fashion piece concept, the model tended to generate fashion designs with painting styles inspired by art pieces, such as an impressionist-style dress inspired by Monet as in Figure 10. On the other hand, when the prompt began with the name of the introduced fashion piece concept and was linked to the inspiration concept with prepositions such as "with," "in," "as," "on," or "by," the results produced images of a dress with patterns that reflected the inspiration concept.

Figure 10. Inspiration Concept-driven Generation. Given Munique Paint Concept (left), it shows good that our approach (right) can synthesize the "Munique Paint Concept" with high fidelity and in new contexts.

Our assessment also included the use of real human selfies, which we found to be effective in printing faces on fashion products. We also observed that by providing the model with specific product details and a given location, it could produce designs that featured the introduced human with the destination background while wearing the specified fashion piece. Furthermore, the model was also able to understand real-life objects, such as the cutenose pillow concept, and create new designs by analyzing the pillow's color, shape, and animal features. When we provided a contextual background with a specific location, such as a forest or school, the model placed the animal inspired by the pillow in the appropriate concept. We even found that by adding the word "friends" to the prompt, the model could place the pillow character in a school context and print it on a specified fashion piece. As a last example, we explored the use of prompts that included contextual information such as "on runway," which allowed fashion designers to view possible results on the runway and visualize the fashion pieces being worn by models, although the realisticity of such results may be limited.

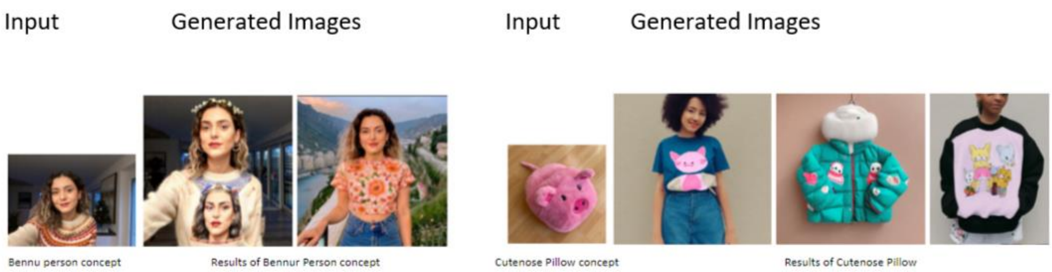


Figure 11. Inspiration Concept-driven Generation. We tried various objectives and generated their concepts.

Fashion Elements

In the Fashion Elements section of our study, we will explore the model's ability to combine specific fashion elements of fashion products. As we did in the previous section, we began by assessing the model's understanding of shape, pattern, texture, color, and space elements. We accept the assumption that each concept contains all fashion elements, regardless of whether it was a fashion piece or an inspiration concept. The model demonstrated a high level of comprehension, even when combining these elements in a single fashion piece.

During the creativity assessment, we applied various fashion elements to the introduced fashion pieces. The model proved to be successful even in combining niche colors, such as mint green with pink and peach orange. It also effectively incorporated specified object shapes, such as trees, moon, and stars, and applied different patterns, including floral and zebra, with various fabrics such as leather and velvet. Subsequently, we explored the model's capacity to understand the specific fashion elements of the introduced concept and apply them to both a fashion piece and introduced fashion piece concept. Applying fashion elements from an introduced concept to an introduced fashion concept proved to be a more challenging task for the model, as it required a higher degree of creativity. However, the resulting designs were more unique and innovative. One example that stood out was the application of Swarovski button (as myswar button) elements to a Zara jacket (as myzara jacket). The goal was to apply the elements of the Swarovski button to the Zara jacket. The model went beyond merely adding the Swarovski button to the jacket as a decorative element. It created new designs for the button, incorporating it as a brooch or chain accessory, or taking inspiration from the color of the specified button and applying it to the jacket. The model even added a new glittering pattern design by arranging stones randomly on different parts of the jacket. Not only did the model consider the shape of the button, but it also incorporated its color, texture, and pattern to create a cohesive and aesthetically pleasing design. As in all previous results, the negative and positive space in the fashion designs were clearly defined. Also, an interesting observation was that when the fabric was specified in the prompt (such as Swarovski button on a *velvet* Zara jacket), the model tended to apply the button across the entire texture of the jacket instead of using it as a single button and the results had a stronger emphasis on the fabric used from closer perspective as shown in Figure 12.

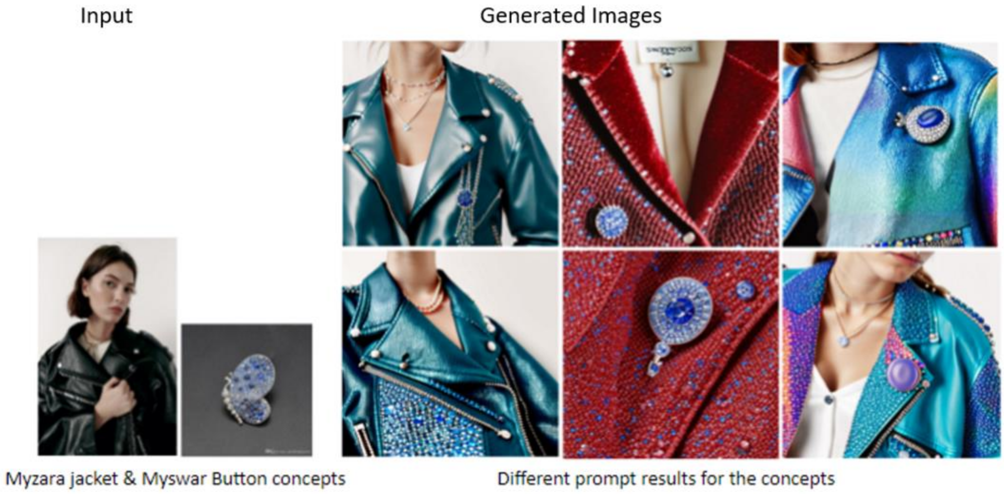


Figure 12. Fashion Pieces-combined Generation. We compared different prompt results, which present a very good combination of results and we note that the model is happy to present the fabric.

To provide fashion designers with greater creative flexibility, we sought to combine elements from multiple concepts. Our first approach was to combine the same elements from multiple concepts on a fashion piece. While the model was generally successful in providing appropriate results, not all results were visually pleasing. The model appeared to find combining patterns relatively easy, but combining colors presented a more significant challenge. Even when the color combination was specified in the prompt, the model often combined the shape of the fashion piece concepts rather than their colors and chose one of the concept's color, as evidenced in Figure 13. However, when we assigned a pattern, such as a zebra pattern, the model was able to combine the colors of multiple concepts within that pattern. When it came to combining two concepts' patterns, the model incorporated clues from both introduced concepts, though the results were not always strong.

Figure 13. Concepts-combined Generation. Our model presents a not-so-good performance when combining the two introduced concepts

Combining the same fashion elements from multiple concepts and applying them to an introduced fashion piece concept resulted in more creative and comprehensive results. As shown in Figure 14 below, the model not only understood and combined the colors but also effectively combined the different fashion elements from both concepts. The texture of the Orpi button, with its shiny surface, was incorporated into the dress and the pink/orange color shapes on the pattern were inspired by the button itself. On the other hand, the pattern was drawn from the Munique art concept by Monet, while the green color was derived from impressionist nature paintings. While this approach offers high creativity opportunities for fashion designers, we found that it is barely possible to combine a specific fashion element while keeping other elements constant.

Figure 14. Advanced Concepts-combined Generation.

In the final part of our exploration, we aimed to apply different fashion elements from different concepts without combining the same elements. This approach was relatively easier than combining since it simply involves gathering different elements together. While the results demonstrated the model's ability to understand some fashion elements, when the prompt included fashion elements from more than two concepts, the model tended to show either some of them or displayed different elements from each concept. Nonetheless, this approach provides fashion designers with more options to incorporate various fashion elements into their designs, expanding the possibilities for creative and innovative fashion concepts. Additionally, we also applied these fashion elements to specific parts of fashion pieces, such as collars, sleeves, etc. This allowed us to explore the model's ability to apply these elements in a more targeted and specific manner. Overall, we found that the model was able to effectively apply the fashion elements to different parts of the fashion pieces, resulting in unique and creative designs. However, we also observed that the model sometimes struggled to maintain consistency in applying the fashion elements across the different parts of the fashion piece.

5 Limitations & Future Work

5.1 Limitations

Our study aimed to explore the creativity capabilities of an AI model in the fashion design process. We conducted comprehensive testing on the model's ability to comprehend various fashion pieces and elements, and its capacity to generate innovative and unique fashion designs. Our study revealed promising prospects for AI models in fashion design with several positive outcomes, including a high level of comprehension in understanding various fashion pieces and elements, effective application of fashion elements from various fashion pieces, and the potential to combine different fashion elements and inspiration concepts to generate innovative and unique designs. However, our findings also highlighted some limitations of the model:

- The model struggles when introduced to more than eight concepts, leading to reduced performance in combining concepts and generating unique designs.
- The model occasionally generates unrelated examples since it focuses on the examples from its pre-trained samples and cannot focus on the specific concept we introduce.
- The generated images may not be entirely realistic and may have some distortions or imperfections.
- The model sometimes struggles to maintain consistency in applying the fashion elements across the different parts of the fashion piece.
- While the model was successful in combining the same elements from multiple concepts, it is challenging to combine a specific fashion element while keeping other elements constant.
- The model is not capable of rendering legible text, which may limit its potential use in certain fashion design applications.
- Randomness is a limitation as the model generates several outputs for the same prompt, and there is no guarantee of the first output being the best.

5.2 Future Research Directions

While the fashion design generation prototype created using Dreambooth has shown promising results, there are several directions that can be explored in future work to further improve and extend the system. Some potential avenues for future research include:

- The current prototype was trained on a limited set of fashion images. Incorporating more diverse training data, such as specific clothes from different cultures and time periods, could help improve the diversity and creativity of the generated designs.
- To make the generated designs more practical and usable, it may be useful to incorporate constraints and guidelines into the system. For example, designers could specify certain design elements or color schemes to be included in the generated designs.
- In order to evaluate the usability and effectiveness of the current system, it would be beneficial to conduct further user studies with fashion designers and enthusiasts. Feedback from users could be used to further refine and improve the system.
- While the current prototype generates fashion designs as static images, it could be interesting to explore the generation of multi-modal designs, such as designs that include both static images and animated 3D models.
- Improving the stability of the model's performance to ensure consistency in generating high-quality designs.

By exploring these avenues for future research, we believe that our fashion design generation system can be further improved and expanded to provide more creative and practical design solutions for fashion designers and enthusiasts.

6 Conclusion

In this study, we explored the potential of AI models in fashion design by using a fashion design generation system based on DreamBooth method. Our testing protocol involved introducing various fashion concepts and inspiration concepts to the model and evaluating its creativity capabilities in generating unique and innovative fashion designs. Our findings demonstrated that the DreamBooth method was capable of understanding various fashion pieces and elements and effectively incorporating them into new designs. However, the model also had some limitations, such as reduced performance when introduced to more than eight concepts, generating unrelated examples since it focuses on the examples from its pre-trained class samples and occasional struggles to maintain consistency in applying fashion elements across different parts of a fashion piece.

Our study has shown that Dreambooth has the potential to meet many of the needs of fashion designers. The model's ability to comprehend various fashion pieces and elements, generate unique and innovative designs, and combine different elements and inspiration concepts makes it a valuable tool for designers. Dreambooth can help democratize the fashion industry by making design solutions more accessible and affordable for emerging designers and small businesses. Despite some limitations, our study suggests that AI models like Dreambooth can enhance designers' creativity and productivity by generating unique and diverse design ideas that may have been overlooked by human designers. Furthermore, our findings indicate that the model can be used to generate designs quickly and efficiently, saving designers time and allowing them to focus on other aspects of the design process. Overall, our research suggests that Dreambooth can be a valuable tool for fashion designers, providing them with a new and innovative approach to the design process, while also streamlining their workflows and saving time.

To further improve and extend the system, future research could focus on incorporating more diverse training data, adding constraints and guidelines to make the generated designs more practical, conducting further user studies with fashion designers and enthusiasts, and exploring the generation of multi-modal designs. By exploring these avenues for future research, we believe that this fashion design generation system can be further refined and expanded to provide more creative and practical design solutions for the fashion industry.

Reference List

Brock, A., Donahue, J., & Simonyan, K. (2018). Large Scale GAN Training for High Fidelity Natural Image Synthesis. Retrieved from arXiv.org website: https://arxiv.org/abs/1809.11096

Cameron, M. (2009). Review Essays: Donald A. Schön, The Reflective Practitioner: How Professionals Think in Action. New York: Basic Books, 1983. ISBN 0—465—06874—X (hbk); ISBN 0—465—06878—2 (pbk). Qualitative Social Work: Research and Practice, 8(1), 124–129. https://doi.org/10.1177/14733250090080010802

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative Adversarial Networks: An Overview. IEEE Signal Processing Magazine, 35(1), 53–65. https://doi.org/10.1109/msp.2017.2765202

Cross, N. (2010). Designerly ways of knowing. Springer London Ltd.

Eckman, M., & Wagner, J. (1995). Aesthetic Aspects of the Consumption of Fashion Design: the Conceptual and Empirical Challenge. ACR North American Advances, NA-22. Retrieved from https://www.acrwebsite.org/volumes/7825/volumes/v22/NA

Gal, R., Alaluf, Y., Atzmon, Y., Patashnik, O., Bermano, A. H., Chechik, G., & Cohen-Or, D. (2022). An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion. *ArXiv:2208.01618 [Cs]*. Retrieved from https://arxiv.org/abs/2208.01618

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., … Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, *63*(11), 139–144. https://doi.org/10.1145/3422622

Gu, S., Chen, D., Bao, J., Wen, F., Zhang, B., Chen, D., … Guo, B. (2022). Vector Quantized Diffusion Model for Text-to-Image Synthesis. *ArXiv:2111.14822 [Cs]*. Retrieved from https://arxiv.org/abs/2111.14822

Jetchev, N., & Research, Z. (2017). *The Conditional Analogy GAN: Swapping Fashion Articles on People Images*. Retrieved from https://openaccess.thecvf.com/content_ICCV_2017_workshops/papers/w32/Jetchev_The_Conditional_Analogy_ICCV_2017_paper.pdf

Jonathan, H., Ajay, J., & Pieter, A. (2020). Denoising Diffusion Probabilistic Models. *Advances in Neural Information Processing Systems*, *33*. Retrieved from https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html

Karras, T., Laine, S., & Aila, T. (2019). A Style-Based Generator Architecture for Generative Adversarial Networks. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr.2019.00453

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). Analyzing and Improving the Image Quality of StyleGAN. Retrieved from openaccess.thecvf.com website: https://openaccess.thecvf.com/content_CVPR_2020/html/Karras_Analyzing_and_Improving_the_Image_Quality_of_StyleGAN_CVPR_2020_paper.html

Lee, J. S., & Jirousek, C. (2015). The development of design ideas in the early apparel design process: a pilot study. *International Journal of Fashion Design, Technology and Education*, *8*(2), 151–161. https://doi.org/10.1080/17543266.2015.1026411

Liu, L., Zhang, H., Ji, Y., & Jonathan Wu, Q. M. (2019). Toward AI fashion design: An Attribute-GAN model for clothing match. *Neurocomputing*, *341*, 156–167. https://doi.org/10.1016/j.neucom.2019.03.011

Makkapati, V., & Patro, A. (2017). Enhancing Symmetry in GAN Generated Fashion Images. *Artificial Intelligence XXXIV*, *10630*, 405–410. https://doi.org/10.1007/978-3-319-71078-5_34

Ng, A., & Jordan, M. (2001). On Discriminative vs. Generative Classifiers: A comparison of logistic regression and naive Bayes. Retrieved April 2, 2023, from Neural Information Processing Systems website: https://papers.nips.cc/paper_files/paper/2001/hash/7b7a53e239400a13bd6be6c91c4f6c4e-Abstract.html

Pandey, N., & Savakis, A. (2020). Poly-GAN: Multi-conditioned GAN for fashion synthesis. *Neurocomputing*, *414*, 356–364. https://doi.org/10.1016/j.neucom.2020.07.092

Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. Retrieved from arXiv.org website: https://arxiv.org/abs/1511.06434

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Openai, M. (2022). *Hierarchical Text-Conditional Image Generation with CLIP Latents*. Retrieved from https://cdn.openai.com/papers/dall-e-2.pdf

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., … Sutskever, I. (2021, July 1). Zero-Shot Text-to-Image Generation. Retrieved from proceedings.mlr.press website: https://proceedings.mlr.press/v139/ramesh21a.html

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-Resolution Image Synthesis with Latent Diffusion Models. *ArXiv:2112.10752 [Cs]*, *2*. Retrieved from https://arxiv.org/abs/2112.10752

Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., & Aberman, K. (2022). DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation. *ArXiv:2208.12242 [Cs]*. Retrieved from https://arxiv.org/abs/2208.12242

Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., … Norouzi, M. (2022, October 31). Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. Retrieved April 2, 2023, from openreview.net website: https://openreview.net/forum?id=08Yk-n5l2Al

Simon, H. A. (1996). *The Sciences of the Artificial, third edition*. MIT Press.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., & Ganguli, S. (2015, June 1). Deep Unsupervised Learning using Nonequilibrium Thermodynamics. Retrieved April 2, 2023, from proceedings.mlr.press website: https://proceedings.mlr.press/v37/sohl-dickstein15.html

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., & Poole, B. (2021). Score-Based Generative Modeling through Stochastic Differential Equations. *ArXiv:2011.13456 [Cs, Stat]*. Retrieved from https://arxiv.org/abs/2011.13456

Volpintesta, L. (2014). The language of fashion design: 26 principles every fashion designer should know. In *Colorado Mountain College*. Beverly, MA: Rockport Publishers. Retrieved from https://cmc.marmot.org/Record/.b42267080

Voronov, A., Khoroshikh, M., Babenko, A., & Ryabinin, M. (2023). Is This Loss Informative? Speeding Up Textual Inversion with Deterministic Objective Evaluation. *ArXiv:2302.04841 [Cs]*. Retrieved from https://arxiv.org/abs/2302.04841

Yan, H., Zhang, H., Liu, L., Zhou, D., Xu, X., Zhang, Z., & Yan, S. (2022). Toward Intelligent Design: An AI-based Fashion Designer Using Generative Adversarial Networks Aided by Sketch and Rendering Generators. *IEEE Transactions on Multimedia*, 1–1. https://doi.org/10.1109/TMM.2022.3146010

Lee, C.-Y., & Chen, B.-S. (2018). Mutually-exclusive-and-collectively-exhaustive feature selection scheme. *Applied Soft Computing*, *68*, 961–971. https://doi.org/10.1016/j.asoc.2017.04.055

Appendix A

1. Interview Feedback

https://urlzs.com/4tQNi

2. Prototype Github Repository

https://github.com/tyasar53/stable_diffusion

3. Link of Test Structure

https://urlzs.com/vmWpW

4. Printed Version of Test Structure with Prompts

Testing Strategy

COLOR GUIDE

(green)	it works
(red)	it doesn´t work
(grey)	not testable

Fashion Pieces			understanding						creativity check			
				introducing the concept	fashion piece concepts						inspirations: person, figure, geometric shape, painting, carton	
					one concept				multiple concepts		inspiration	
	first class	second class	fashion piece check	one concept check	one concept shift within the same first class	one concept shift across the first class	on a fashion piece	on our introduced fashion piece concept	on a fashion piece	on our introduced fashion piece concept	inspiration(s) on a fashion piece	inspiration(s) on our introduced
	top	shirt, t-shirt, jacket, bikini	photo of a t-shirt photo of a shirt	photo of a picol sweatshirt	picol sweatshirt tshirt	dress like picol sweatshirt dress	rhot clothing on blue jeans	rhot clothing on lovelz pants	picol sweatshirt on a skirt with moop bag	picol sweatshirt with an orpi button and a flowre necklace	- long sleeve pullover with bennu person print - bennu person wears flower crop top in Amalfi coast	munique paint with rhot clothing
	bottom	trousers, jeans, shorts, skirt	photo of pants photo of a skirt	photo of a bluski skirt	bluski skirt pants	crop top like lovelz pants	bluski skirt under a long jacket	bluski skirt under mymon coat	lovelz pants with orpi button under a crop top	bluski skirt with orpi button under rhot clothing	bluski skirt with onne design	lovelz pants with onne design
	shoes	sneakers, high heels boots	photo of high heels boots photo of sneakers	photo of a crocod boots	sneakers like crocod boots crocod boots sneaker crocod shoes summer shoe	rocod boots jacket jacket like crocod boots jacket as crocod boots jacket with crocod boots style	crocod boots on orange socks	crocod boots and mavibe dress	crocod boots with yellow dress with flowre necklace	nky sneaker with flowre necklace and orpi button	boots with munique art	nky sneaker with yummz pattern
	suit	dress, jumpsuit	photo of a dress photo of a jumpsuit	photo of a bejflow dress	bruny dress jumpsuit	boots like bruny dress	bejflow dress on pink shoes	bruny dress with flowre necklace on it	bejflow dress on blue high heels shoes with orpi button	bejflow dress with flowre necklace and mymon coat and moop bag and crocod boots	cutenose pillow bejflow dress	xsuns view on bejflow dress
	accessories	button, scarf, belts, brosche, purse	photo of a scarf photo of a belt	photo of a flowre necklace	orpi button necklace	skirt like orpi button	moop bag with pink buttons	swarovski button on zara jacket	orpi button belt on rainbow myfav dress	swarovski button on crocod boots with lovelz pants	purse with Munique art	xsuns view on moop bag

Fashion Elements			understanding	creativity								iterative test
				take elements from no concept	take elements from one concept			take elements from multiple concepts				deviate from input image / deviate from prompt
								same element from multiple concepts		different elements from different concepts		
	class	example	fashion element check	on an introduced fashion piece concept	on a fashion piece	on our introduced fashion piece concept	on our introduced fashion piece concept with specification	on a fashion piece	on our introduced fashion piece concept	on a fashion piece	on our introduced fashion piece concept	original prompt: photo of orpi button
	shape	slip, oversize, tunic, balloon	oversize shaped shirt balloon shaped dress	pink orange flower pattern on mycomfy clothing	flowre necklace shaped earrings	orpi button shape and color on necklace on green color myfav dress		bruny dress shape and flowre necklace shape on a t-shirt	bruny dress shape and flowre necklace shape on a rhot clothing	xsuns view color and bluski skirt shaped shorts	xsuns view color and orpi button shape accessory on mavibe dress	change the button shape to triangle
	pattern	stripe, animal print, floral	zebra pattern on a shirt pants with floral pattern	highlighted mint green pink orange color on mycomfy clothing	sallz cartoon colored bag	xsuns view pattern on mymon jacket	swarovski button on red color velvet fabric myzara jacket	mixture of brunny dress color and picol sweatshirt color as zebra pattern on a tshirt	brunnz dress with orpi button pattern and munique paint pattern on	crop top with brunnz dress shape and yummz pattern	brunnz dress with picol sweatshirt color and orpi button pattern belt	change the button pattern to floral
	fabric (texture)	leather, velvet, fleece, others	fleece fabric sweatshirt leather fabric jacket	green tree shape on brown color mycomfy clothing pullover	yummz pattern on a necklace	swarovski button as a pattern on leather fabric zara jacket	flowre necklace on pink satin fabric bejflow dress	bruny dress fabric and bejflow dress fabric on a shirt	bruny dress fabric and bejflow dress fabric on mymon coat	orpi button accessory on picol sweatshirt color tshirt	woman wears brunnz dress color myfav dress with orpi button pattern necklace	change the fabric to velvet
	color	all	pink color bikini blue color shirt	leather fabric zara jacket with pink color scarf	lalala sweatshirt fabric on a pullover	xsuns view color on rectangle shaped moop bag		bruny dress color and picol sweatshirt color dress	xsuns view color and moop bag color on rhot clothing	sallz cartoon color and lalala sweatshirt fabric pullover	mymon coat fabric and orpi button shaped brooche on myfav dress	change the color to blue
	space	positive background, negative background										

5. Printed Version of Test Structure with Assessments

Testing Strategy

COLOR GUIDE

- it works
- it doesn't work
- not testable

<table>
<tr><td rowspan="9">Fashion Pieces</td><td colspan="2" rowspan="2"></td><td colspan="4">understanding</td><td colspan="6">creativity check</td></tr>
<tr><td rowspan="3">fashion piece check</td><td>introducing the concept</td><td colspan="6">fashion piece concepts</td><td colspan="2">inspirations: person, figure, geomatric shape, painting, carton</td></tr>
<tr><td rowspan="2">first class</td><td rowspan="2">second class</td><td rowspan="2">one concept check</td><td colspan="4">one concept</td><td colspan="2">multiple concepts</td><td colspan="2">inspiration</td></tr>
<tr><td>one concept shift within the same</td><td>one concept shift across the first</td><td>on a fashion piece</td><td>on our introduced fashion piece concept</td><td>on a fashion piece</td><td>on our introduced fashion piece concept</td><td>inspiration(s) on a fashion piece</td><td>inspiration(s) on our introduced fashion piece concept</td></tr>
<tr><td>top</td><td>shirt, t-shirt, jacket, bikini</td><td rowspan="5">works very well for every specific fashion pieces, the model knows exactly what these words refer to</td><td rowspan="5">works very well for every specific fashion pieces, the model understands each concept we introduced</td><td rowspan="2">works very well, the model transfer the color and texture to new piece</td><td rowspan="2">works well, the model understands color, but sometimes generates random general fashion piece</td><td rowspan="5">works well, the model understands all fashion elements, but sometimes generates fashion piece not as required. It sometimes makes mistakes on the color, texture or the shape</td><td rowspan="2">works well, the model understands color, texture and shape, but sometimes put the color in the background not in the specified fashion piece</td><td rowspan="5">works well, the model understands all fashion pieces, but sometimes it makes mistakes on color, or put them in wrong place</td><td rowspan="5">works well, the model understands all fashion pieces, but sometimes it makes mistakes on color, or put them in wrong place</td><td>works very well, the model understands the concept we introduced and generate image as directed properly</td><td>works very well, the model understands inspiration and fashion piece we introduced</td></tr>
<tr><td>bottom</td><td>trousers, jeans, shorts, skirt</td><td>works well, the model understands the concept we introduced, but sometimes forgets to generate fashion piece</td><td>works, the model can understand the shape and pattern, but sometimes doesn't understand the design</td></tr>
<tr><td>shoes</td><td>sneakers, high heels boots</td><td>the model forgets target fashion piece, always very similar to the introduced concept</td><td>the model always created boots and jacket seperately</td><td>the model doesn't transform the correct color but keeps the shape the same</td><td rowspan="3">works very well, the model understands the inspiration we introduced and generate image as directed properly</td><td>works, the model understands introduced concepts, but not always generate in a proper way</td></tr>
<tr><td>suit</td><td>dress, jumpsuit</td><td rowspan="2">works well, the model understands color, texture, shape, but sometimes generates unexpected fashion piece</td><td rowspan="2">works well, the model understands color or shape, but sometimes makes mistakes on either fashion piece we specified or the concept we introduced</td><td rowspan="2">works very well, the model understands all fahsion elements and put them into the right position</td><td rowspan="2">works very well, the model all the elements and the inspiration we introduced, and mixtures them in an aesthetic way</td></tr>
<tr><td>accessories</td><td>button, scarf, belts, brosche, purse</td></tr>
</table>

<table>
<tr><td rowspan="9">Fashion Elements</td><td rowspan="4">class</td><td rowspan="4">example</td><td rowspan="2">understanding</td><td colspan="8">creativity</td><td colspan="2">iterative test</td></tr>
<tr><td rowspan="2">take elements from no concept</td><td colspan="3" rowspan="2">take elements from one concept</td><td colspan="4">take elements from multiple concepts</td><td rowspan="3">deviate from input image</td><td rowspan="3">deviate from prompt</td></tr>
<tr><td rowspan="2">fashion element check</td><td colspan="2">same element from multiple concepts</td><td colspan="2">different elements from different concepts</td></tr>
<tr><td>on an introduced fashion piece concept</td><td>on a fashion piece</td><td>on our introduced fashion piece concept</td><td>on our introduced fashion piece concept with specification</td><td>on a fashion piece</td><td>on our introduced fashion piece concept</td><td>on a fashion piece</td><td>on our introduced fashion piece concept</td></tr>
<tr><td>shape</td><td>slip, oversize, tunic, balloon</td><td rowspan="4">works very well, the model understands every fashion elements</td><td rowspan="4">works very well, the model understands every fashion elements and mixtures with the concept we introduced properly</td><td rowspan="4">works very well, the model takes the fashion element from the concepts we introduced properly</td><td rowspan="4">works very well, the model takes the fashion element from the concepts we introduced properly, and makes modifications as required based on the introduced concepts</td><td rowspan="4">works very well, the model takes the fashion element from the concepts we introduced properly, and makes modifications as required based on the introduced concepts and prompt</td><td>works very well, the model take the fashion element from multiple concepts properly</td><td rowspan="4">works very well, the model take the fashion element from multiple concepts properly, but sometimes forgets the target fashion piece</td><td rowspan="2">works very well, the model take the elements from requied concepts properly</td><td rowspan="2">works very well, the model take the elements from required concepts properly</td><td rowspan="4">works, the model extends on overall shape of the input image, but varies a lot</td><td rowspan="4">works very well, the model modifies the specific element correctly according to the prompt given with creativity</td></tr>
<tr><td>pattern</td><td>stripe, animal print, floral</td><td>works very well, the model mix colors of multiple concepts properly</td></tr>
<tr><td>fabric (texture)</td><td>leather, velvet, fleece, others</td><td>the model forgets the target fashion piece</td><td>works, the model take parts of the elements as required</td><td rowspan="2">works well, the model take the elements from required concepts properly, but sometimes sticks too much on original fahsion piece</td></tr>
<tr><td>color</td><td>all</td><td>the model can't mix colors but can mix shapes</td><td>works very well, the model take the elements from requied concepts properly</td></tr>
<tr><td>space</td><td>positive background, negative background</td><td colspan="11"></td></tr>
</table>

6. Test Results Corresponding to Test Structure

Testing Strategy

COLOR GUIDE
- it works
- it doesn´t work
- not testable

			understanding				creativity check					
			fashion piece check	introducing the concept	fashion piece concepts						inspirations: person, figure, geomatric shape, painting, carton	
					one concept				multiple concepts		inspiration	
Fashion Pieces	first class	second class		one concept check	one concept shift within the	one concept shift across the first	on a fashion piece	on our introduced	on a fashion piece	on our introduced	inspiration(s) on a fashion	inspiration(s) on our introduced
	top	shirt, t-shirt, jacket, bikini	D10	E10	F10	G10	H10	I10	J10	K10	L10	M10
	bottom	trousers, jeans, shorts, skirt	D11	E11	F11	G11	H11	I11	J11	K11	L11	M11
	shoes	sneakers, high heels boots	D12	E12	F12	G12	H12	I12	J12	K12	L12	M12
	suit	dress, jumpsuit	D13	E13	F13	G13	H13	I13	J13	K13	L13	M13
	accessories	button, scarf, belts, brosche, purse	D14	E14	F14	G14	H14	I14	J14	K14	L14	M14

			understanding	creativity								iterative test
			fashion element check	take elements from no concept	take elements from one concept			take elements from multiple concepts				deviate from input image / deviate from prompt
								same element from multiple concepts		different elements from different concepts		
Fashion Elements	class	example		on an introduced fashion piece concept	on a fashion piece	on our introduced fashion piece concept	on our introduced fashion piece concept with specification	on a fashion piece	on our introduced fashion piece concept	on a fashion piece	on our introduced fashion piece concept	original prompt: photo of orpi button
	shape	slip, oversize, tunic, balloon	D20	E20	F20	G20	H20	I20	J20	K20	L20	M20
	pattern	stripe, animal print, floral	D21	E21	F21	G21		I21	J21	K21	L21	M21
	fabric (texture)	leather, velvet, fleece, others	D22	E22	F22	G22		I22	J22	K22	L22	M22
	color	all	D23	E23	F23	G23		I23	J23	K23	L23	M23
	space	positive background, negative background										

D10

D11

D12

D13

D14

E10

input:

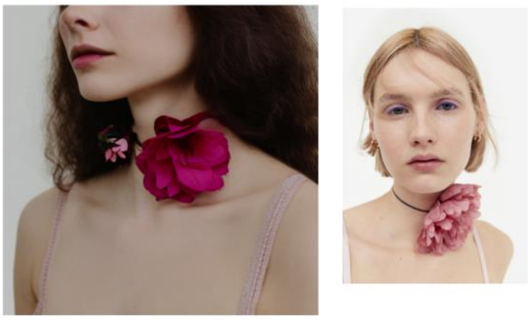
E11

input:

E12 Input

E13 Input

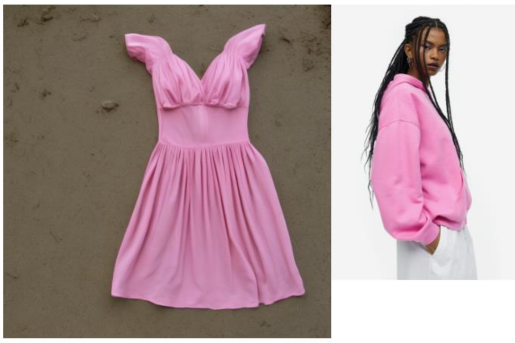
E14 Input



F10 Input

F14 Input

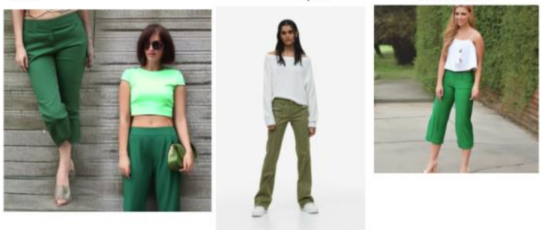
G10 Input



F13 Input Bad Results

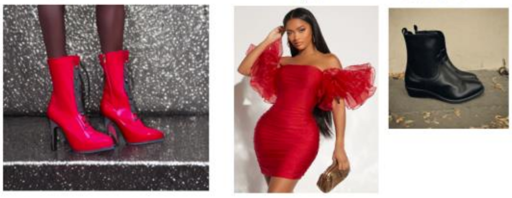
F12 Input Bad Results

G11 input bad results



G12 Input bad results:

G13 input bad results



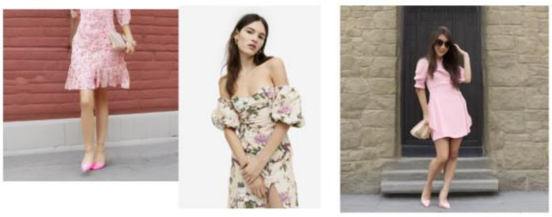
G14 input bad results

H10 input bad results

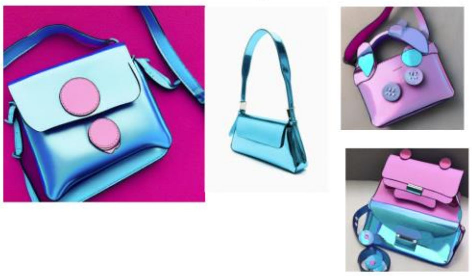
H11 input bad results

H12 input bad results

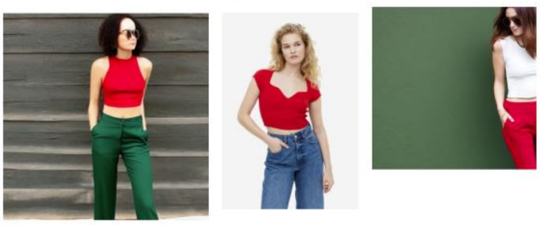
H13 input bad results



H14 input bad results



I10 input bad results

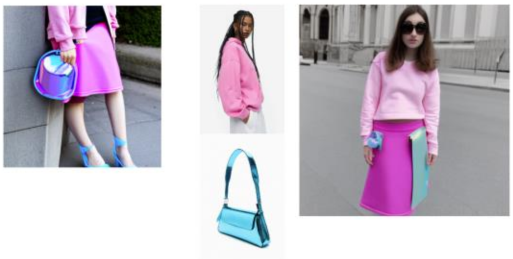


I12 input bad results

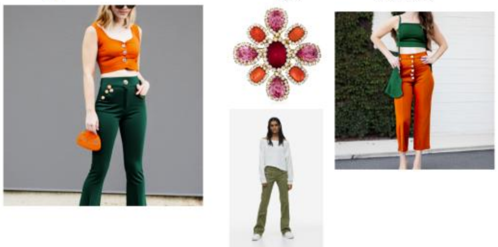
I13 input bad results

I14 input bad results

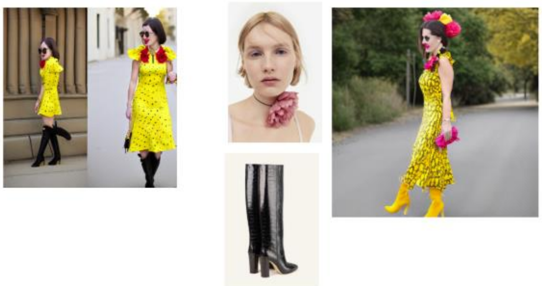
J10 input bad results



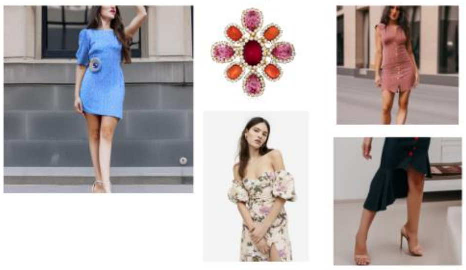
J11 input bad results



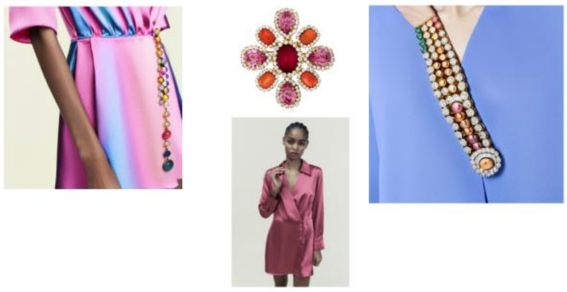
J12 input bad results



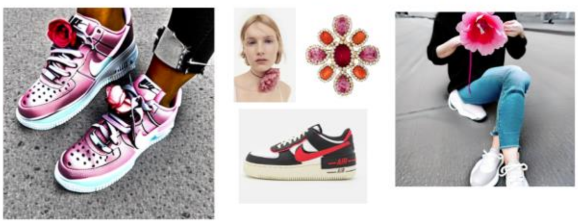
J13 input bad results



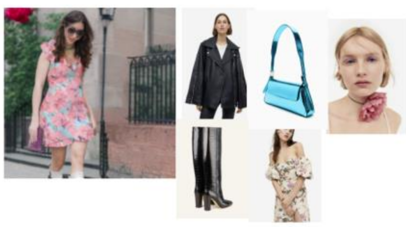
J14 input bad results



K12 input bad results

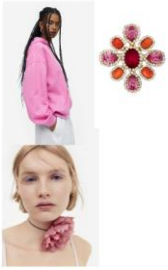


K13 input



K10

input



K11

input

bad results

L11

input

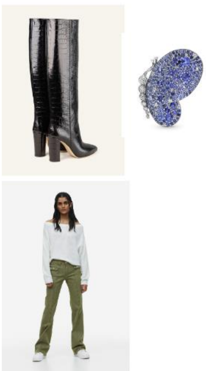
bad results

L12

input

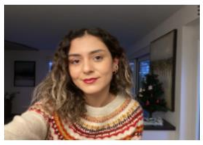
K14

input



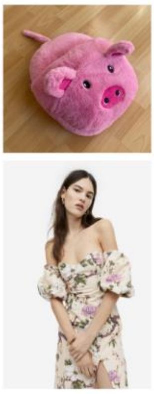
L10

input



L13

input



L14

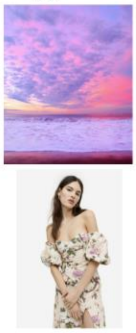
input

M10 input bad results

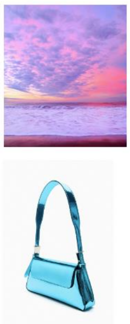
M11 input bad results

M12 input bad results

M13 input bad results

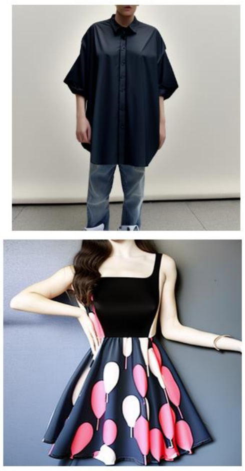


M14 input

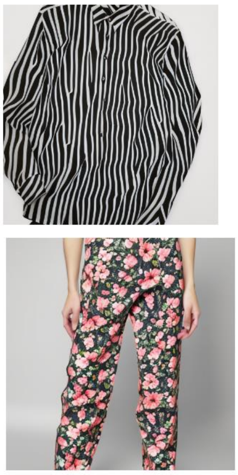


M10 input

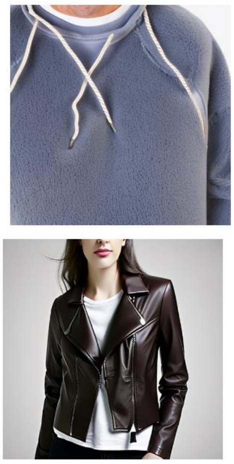
D20



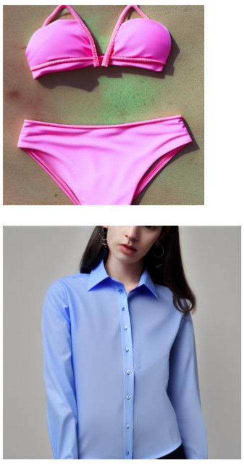
D21



D22:



D23



E20

Input

E21

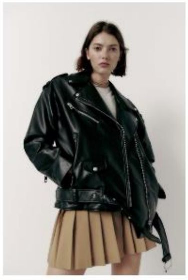
Input

E22

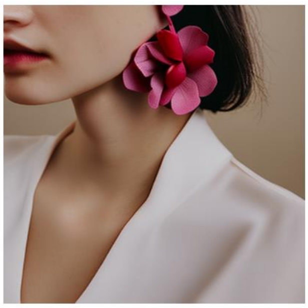
Input

E23

Input



F20



Input

F21

Input

F22

Input

F23

Input

G20

Input

G21

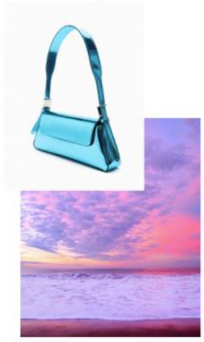
Input

G22

Input

G23

Input



I23

Input

Bad results

I22

Input

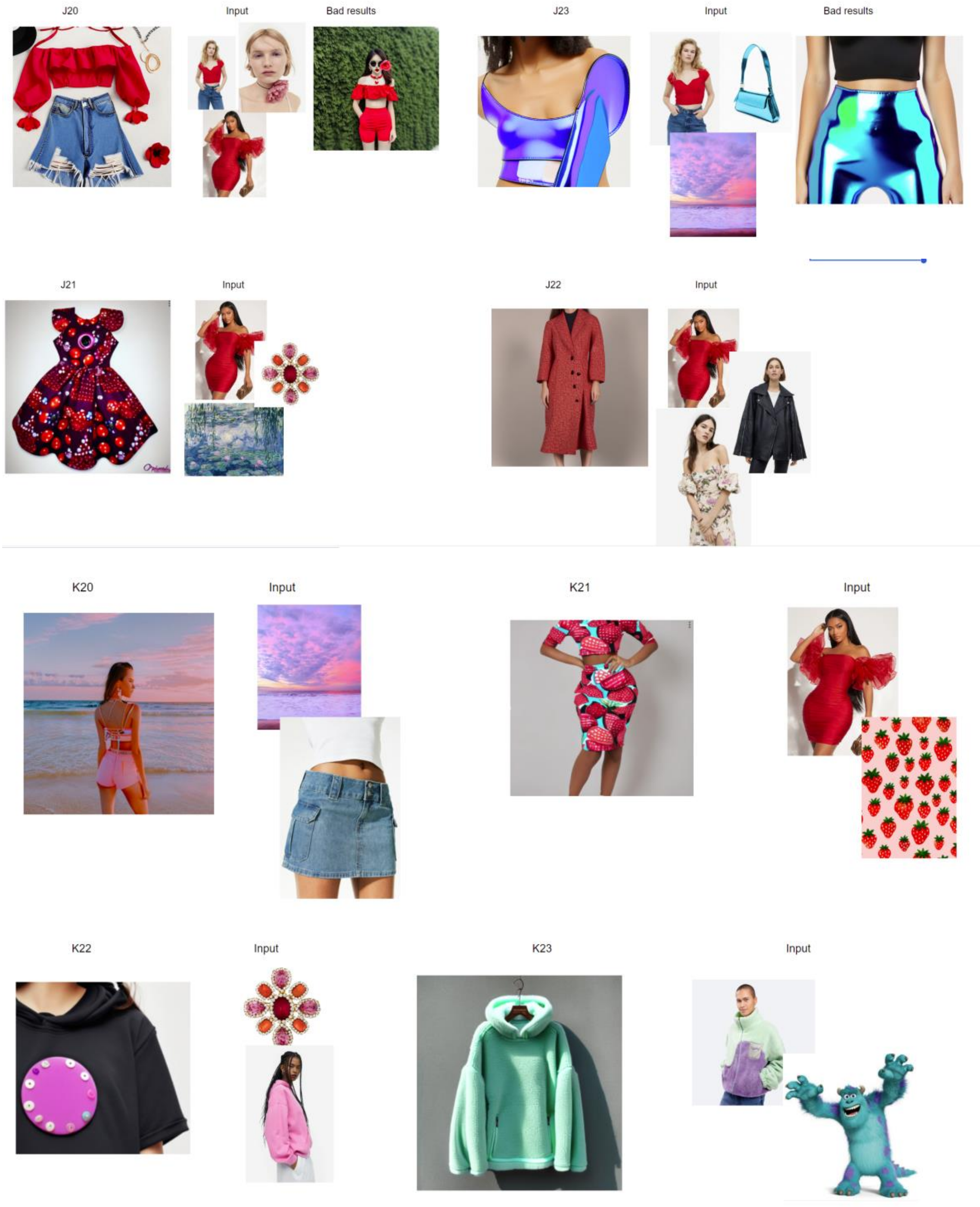
Bad results

H20

Input

I20

Input


J20
Input
Bad results
J23
Input
Bad results
J21
Input
J22
Input
K20
Input
K21
Input
K22
Input
K23
Input

L20

Input

L21

Input

L22

Input

Bad resutls

L23

Input

Bad resutls

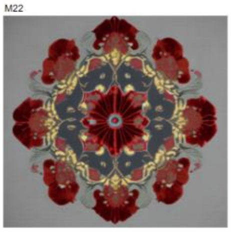
M20

source

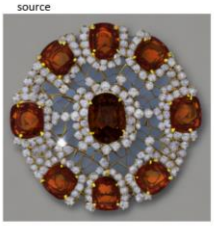
M21

source

M22

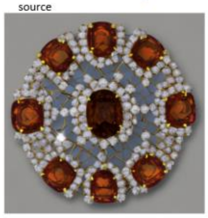


source



M23

source



Affidavit

I hereby declare that I have developed and written the enclosed seminar paper / bachelor thesis / master thesis entirely on my own and have not used outside sources without declaration in the text. Any concepts or quotations applicable to these sources are clearly attributed to them. This seminar paper / bachelor thesis / master thesis has not been submitted in the same or a substantially similar version, not even in part, to any other authority for grading and has not been published elsewhere. This is to certify that the printed version is equivalent to the submitted electronic one. I am aware of the fact that a misstatement may have serious legal consequences.

I also agree that my thesis can be sent and stored anonymously for plagiarism purposes. I know that my thesis may not be corrected if the declaration is not issued.

Mannheim, April 3, 2023

Jiaxin Yuan, Yunjing Dai, Bennur Kaya, Tolga Yasar, Yi Wang, Yiyi Wei,