School of Computing

Honours Report

Applying AI Image generation methods in the process of Swimwear Design

Pawel Sznura 2023

This report is submitted as part of the requirements for the degree of

BSc (Hons) in Computing (Application Software Development)

at Robert Gordon University, Aberdeen, Scotland

I confirm that the work contained in this Honours project report has been composed solely by myself and has not been accepted in any previous application for a degree. All sources of information have been specifically acknowledged and all verbatim extracts are distinguished by quotation marks.

Signed Pawel Sznura Date 28.04.2023

Contents

Abstract

AI design is a rising area of research, it is commonly used in many industries, nonetheless, in more creative applications like fashion design it is almost non existing. The literature review showed that the necessary technology was already discovered and now it is the time to implement and adapt it. The project utilises pre-trained image to text and text to image models. The results are evaluated using appropriate metric as well as human survey. The image given is transformed into a text description, based on which a state-of-the-art text to image model generates a swimwear design. The final image is evaluated in a human survey. This project aim is to create a system which transforms an image into swimwear.

Introduction

Artificial intelligence has made significant strides in recent years, enabling new applications and capabilities across a wide range of domains. One such domain is the creative industry, where AI techniques have demonstrated great potential in enhancing the design process and generating novel and appealing designs. In this context, the present report explores the development and evaluation of an AI-powered swimwear design system that leverages state-of-the-art AI models and techniques to facilitate the generation of unique and visually appealing swimwear designs.

The primary objective of this project is to build a functional software that employs AI techniques to generate swimwear designs based on a given inspiration image. This involves implementing image-to-text and text-to-image models that can generate descriptive text and corresponding images, respectively. Furthermore, the project aims to assess the effectiveness and appeal of the generated designs through human evaluation surveys and quantitative metrics, such as the Fréchet Inception Distance.

The report is structured as follows: First, a comprehensive literature review is presented, highlighting the current state of AI techniques in creative design and image generation, as well as identifying relevant models and methods for the project. Next, the methodology is described, detailing the implementation of the image-to-text and text-to-image models, and the evaluation process involving human surveys and quantitative metrics. The results and findings from the evaluation process are then discussed, including the successes and limitations of the project in meeting its requirements. Finally, the report concludes by summarising the key insights and potential future work for the project, with a focus on improving the system and expanding its capabilities.

Project Scope

Background

AI in creative applications

The use of artificial intelligence for design is common in many industries. Stretching from user interfaces on websites like Airbnb to unique jar packaging for Nutella. (Philips, 2018) In the fashion industry AI is used in many aspects, like forecasting trends, improving the supply chain or personalising the experience. (Intelistyle, 2022) In 2017, Amazon declared that they would start using an ‘AI fashion designer’ with the aim to replace conventional designers. (Knight, 2017) However, it is an exception, with the majority of companies are still leaving the creative process to human, with AI playing only a support role. (Bisen, 2020)

Figure 1 Unique Nutella jar packaging. (Aouf, 2017)

AI Image Generation

An image is data like any other, stored in a different way, for example as a set of pixels with a red, green and blue value or as a set of vectors. Therefore, working with images is not that different to numerical data. However, an artificial intelligence does not a have any visual concepts, thus it is necessary to train it, based on images from datasets. Many methods and approaches have emerged throughout the years.

Generative Adversarial Networks (GANs)

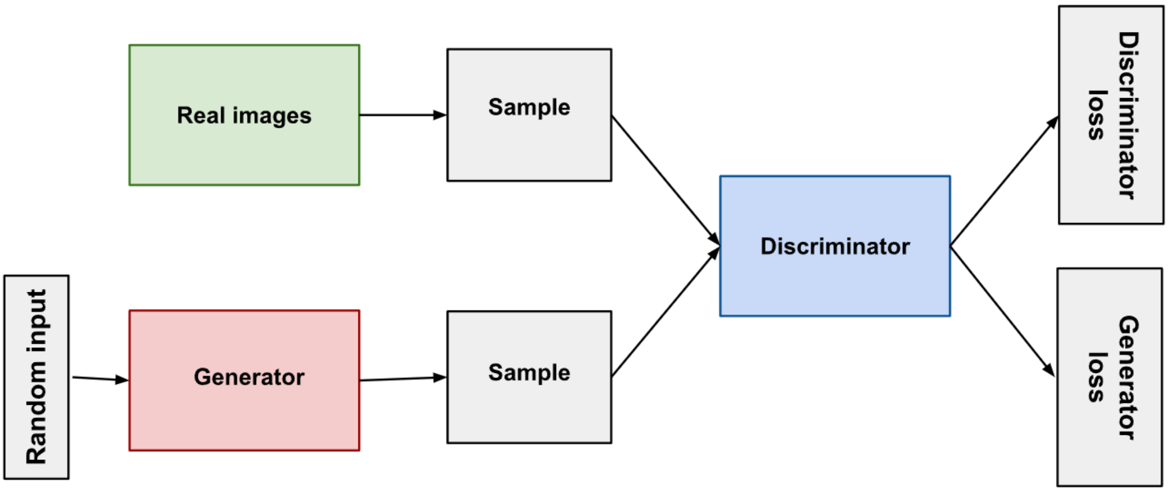


Figure 2 GAN Structure (Google, 2022)

GANs (Generative Adversarial Networks) (Goodfellow et al., 2014) are using a generator and a discriminator which are trained simultaneously and compete. On an example of the use case of image creation, the generator has no access to real images. Only based on the source noise it must create an almost real image. On the other hand, for the discriminator a sample will be drawn from either the synthetic or real stack. It must guess if it is real or generated. If it guesses right, then the generator will improve, if the guess was wrong then the discriminator will change. Ideally the training process will run until the discriminator cannot distinguish the generated images for the real ones and the correct prediction will be 50%, the same value as the discriminator would guess randomly. (Creswell et al., 2017) Nonetheless, they rely on big number of training images. There have been attempts to train GANs with limited data, after all, to achieve a satisfying performance were still hundreds or even thousands of images required. (Yang et al., 2021) One of the drawbacks of GANs is the colossal computation power and time needed for the training of these models, especially when training on a general dataset. The default configuration of StyleGAN was training over 6 days on 8 Tesla V100 GPUs using 1024x1024 pixel images. (Karras et al., 2019)

An extension to GANs are Conditional Generative Adversarial Networks. Conditional Image to Image translation, requires an image from the source domain along with a condition image. The target image should inherit some domain specific features from the conditional image. (J. Lin et al., 2018) The cd-GAN (conditional dual – GAN) approach allows for better controlling the translation result. The authors suggest that it could be extended to applications such as conditional video translation or conditional text translations.

Interpretable GAN Control

Researchers were trying to add user control of the output, on supervised learning and GAN with labelled images. This approach requires manual supervision for every new adjustment to be added. GANSpace proposed a technique for analysing GANs and adding slider controls over image features for the generated output. (Härkönen et al., 2020) It takes existing various pretrained GANs as input. Other studies introduced an unsupervised method to discover meaningful directions in the latent space of GANs by decomposing the model training weights. In a result it was possible to edit the synthetic images in real time. Fig. 3 shows the human – generation model interface which allows for image manipulation. Experiments shown that this approach is applicable to most popular GAN models whether the training dataset. (Shen & Zhou, 2020)

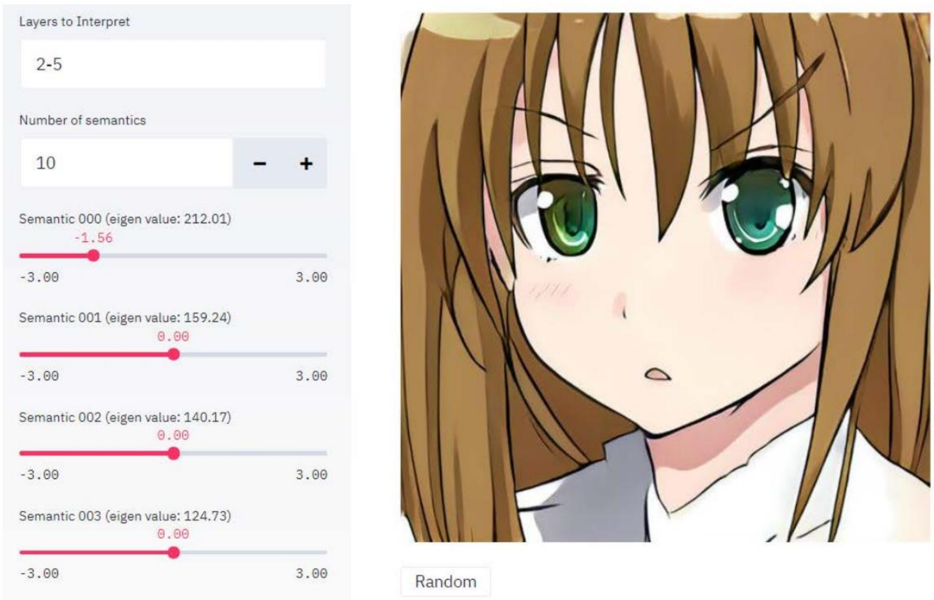


Figure 3 Human-model interaction interface

Text to image generation

Thanks to conditional GANs scientists made a tremendous improvement in generating realistic image based on the given text. A study introduced a model for changing a source image by a text description, with the requirement of maintaining the rest of the image the same and staying realistic. (Dong et al., 2017) It was built upon a conditional GAN framework, which used an encoder-decoder approach, where the source image and text description were encoded to be used by the generator. The discriminator judged based on the text semantic features. The model was evaluated on Caltech-200 bird dataset and Oxford 102 flower dataset, the result completed the requirements and outperformed other methods at that time.

However, the studies focused on real images, not created ones. Therefore, ControlGAN (Li et al., 2019) proposed a controllable text-to-image generative adversarial network, with generates a new image based only on the text input. Also, it allows the user to modification specific elements of the image, with the focus on visual attributes like texture or colour, without interfering with other part of the image. Experiments conducted by the authors indicate that the model has an advantage when compared to StackGAN++ (Zhang et al., 2019) or AttnGAN (Xu et al., 2017) on the tested datasets.

Fashion-Gen introduced a large fashion dataset, almost 300K high definition (1360x1360 pixels) pictures. Fashion items were photographed up to 6 different camera angles, categorised in main and subcategories, additionally professional fashion designers wrote a description for every fashion item alongside with recommendation of matching items, season, designer and brand. Furthermore, the researchers proposed an image generation model conditioned on text description. (Rostamzadeh et al., 2018)

An interesting approach presents MirrorGAN, it is a semantic preserving text to image to text framework. (Qiao et al., 2019) The process starts with a user given text description. The first module, STEM is a semantic text embedding module, its purpose is to recognise local word and global sentence level features from the starting text. Next is GLAM, a global-local collaborative attentive module, which based on these features generates an image. The last module is STREAM, semantic text regeneration and alignment module, which generates a text description for the artificial image and compares it to the starting text. In the experiments based on two benchmark datasets, the MirrorGAN framework achieved superiority over state of the art methods. However, due to the computing power limitation, STREAM module had to be trained separately to the others. Researchers suggest that the text modules could be further improved, because currently only basic methods are in use.

One of the best-known text to image generators is Dall E, mostly thanks to its mini version which is hosted on a website and anyone in matter of minutes could generate images. (Dayma & Cuenca, 2021) However, it is using a simplified model to reduce server computational power. Dall e is a text to image transformer which autoregressive models the text and image tokens as a single stream of data. The whole process in divided into two stages. In stage one, a discrete autoencoder (dVAE), compresses images into a 32 X 32 grid of image tokens. It helps in reducing the memory amount, because generating a token value for every pixel is beyond the current computation power. The second stage links the text with the image tokens, additionally an autoregressive transformer is trained to model the distribution of text and image tokens. Typically, text to image generators were evaluated on rather small datasets. Ramesh et al. (2021) trained the model on 250 million image text pairs collected from the internet. The system achieved a superb image generation quality. Human evaluated the images, in a result, 90% of the time the images generated by Dall E were preferred over prior models. The model presented an unusual high level of abstraction. The study did not provide a revolutionary theory but rather a simple approach; however, the implementation is outstanding. The authors said that the most challenging part of the project was training the model in 16-bit precision past one billion parameters. It does suggest that existing methods can be further improved to match or surpass the state-of-the-art methods.

Image to text generation

Image classification is one of the core problems in computer vision, it is still used often as a benchmark to measure progress in image understanding. Touvron et al. (2020) introduced a data efficient image transformer, it adopts a distilled Vision Transformer (ViT). The model was trained only on the ImageNet dataset, thanks to that it was possible to reduce the training time to less than three days, on a single computer, without sacrificing accuracy. The pretrained and fine tuned image transformer model, together with a model card can be found on the Hugging Face website. (Huggingface, 2021)

Another model which can be count on the website is the BEiT model, which is pretrained in a self supervised fashion on the ImageNet-22K dataset and fine tuned on the same dataset, at the same 224 X 224 resolution. BEiT is a Bidirectional Encoder representation from Image Transformers. (Bao et al., 2021) In the opposite to the initial ViT models, the BEiT model uses relative position embeddings in contrast to absolute position embeddings conduct classification of images by mean pooling the final hidden states of the patches, rather than placing a linear layer on top of the final hidden state of the token. The carried out experiment in the study showed that the model is complementary to supervised pre-training, achieving additional benefits after fine-tuning on ImageNet.

Techniques comparison

Table 1 Techniques comparison

Name	Paper	Type	Notes
GenDA	(Yang et al., 2021)	Image to image	Based on StyleGAN2 model, large-scale source dataset and one-shot domain adaptation
cd-GAN	(J. Lin et al., 2018)	Image to image	Two conditional translation models combined
Fashion-Gen	(Rostamzadeh et al., 2018)	Text to image	Image and description pairs, presents new Fashion dataset
GANSpace	(Härkönen et al., 2020)	Image to image	Finding interpretable controls for pretrained BigGAN or StyleGAN2 models
SeFa	(Shen & Zhou, 2020)	Image to Image	Semantic Factorization, discovers interpretable directions in pretrained GANs (PGGAN, StyleGAN, BigGAN)
Semantic Image Synthesis via Adversarial Learning	(Dong et al., 2017)	Image + text to image	Based on conditional GAN framework, changes images with a text description while maintains other image features
ControlGAN	(Li et al., 2019)	Text to image	Generate and manipulate images based on natural language description
MirrorGAN	(Qiao et al., 2019)	Text to image	Integrates text to image and image to text
Dall E	(Ramesh, Pavlov, Goh, Gray, et al., 2021)	Text to image	Autoregressive transformer, colossal amount of training data
Distilled Data-efficient Image Transformer	(Touvron et al., 2020)	Image to text	Distilled Vision Transformer
BEiT	(Bao et al., 2021)	Image to text	Bidirectional Encoder representation from Image Transformers, pre-trained in a self-supervised fashion, fine-tuned on the same dataset

Evaluation metrics

To assess those models many evaluation metrics were introduced. Inception Score (IS) aims to measure the quality and diversity of an image. (Salimans et al., 2016) At Amazon Mechanical Turk (AMT) human workers are scoring images to measure realism and faithfulness. (Alotaibi, 2020)

Another metric is the Frechet inception distance (FID) which better reflects the similarity of the generated images to the real ones than the Inception Score by being less influenced by noise. (Heusel et al., 2017) It is worth noting that IS and FID metrics are far from perfect, where studies have shown their flaws (Obukhov & Krasnyanskiy, 2020)(Chong & Forsyth, 2020), however they remain very popular.

Conclusion

Overall, the AI design and image generation is an incredibly fast growing field. Where new state of the art methods or approaches are presented every few years. However, the main limitation factor has been and still is hardware performance. Many of the mentioned models were trained with enormous amounts of data and computing power which this project will not be able to reproduce. However, other studies have shown that investing the resources on improving a model can bring a satisfying result, probably even a better one than starting from the beginning. At this point we need to mention the ongoing climate change as well, the carbon footprint of the machine learning field is rising rapidly. (Strubell et al., 2019) Therefore, the emissions of training a new model for this project are not justifiable. Fortunately, many of the previously mentioned models have broadly available pretrained models for download and use.

Requirements analysis

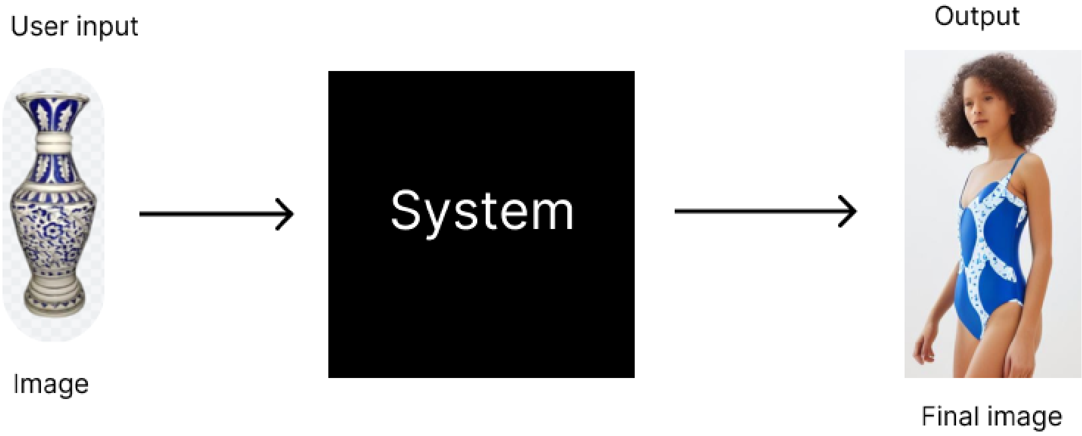


Figure 4 Application diagram

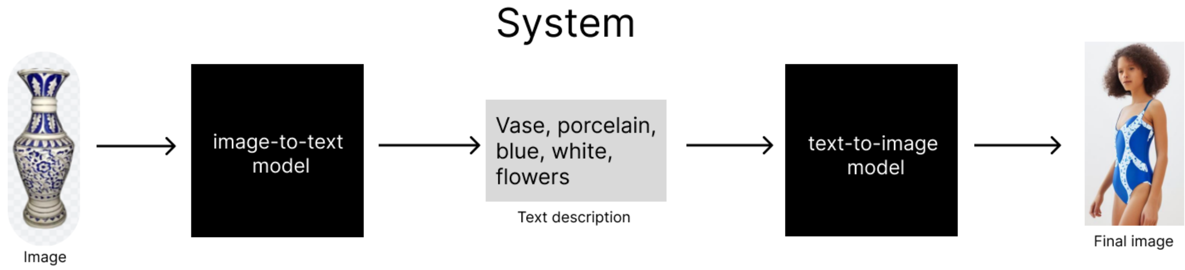


Figure 5 System diagram

Functional Requirements

Must be evaluated in term of whether the image depicts swimwear or not.
Should be evaluated in the terms of design aesthetics, where at least half of the participants in a human survey say that the design is appealing.
Could use Inception Score or Fréchet inception distance for evaluating the model.
The image generation process should last less than 2 minutes in 90% of the cases, on my personal laptop. (AMD Ryzen 7 6800HS, 24GB RAM, AMD Radeon RX 6700S)

Must use an image-to-text model for generating a description based on the image description.
Should use a pre-trained image-to-text model because it will present the best possible results in this project due to the limitation in time and hardware power.
Must take images as starting parameters.
Must transform images into the right format for the model.
Could use different file extensions for the input image provided by the user.

Must use a text-to-image model for generating an image based on the text description.
Should use a pre-trained text-to-image model because the computation power and time required for training are beyond the scope of the project.
Should add relevant keywords to the text description, like “swimsuit”, “swimwear”, to generate the intended design.

Could be able to save the generated image in a user-friendly format as “.png” or “.jpg “
Could allow the user to add a keyword to the text description, for example to add an attribute that is not present in the image.

Could present more than one image at the time to the user.
Could be modified by the user to generate a two-piece swimsuit.
Could have the option for a user text input, when the initial results are not satisfying.

Could have a user interface for selecting and generating images.
Could use Gradio, Dash, Streamlit, Flask or similar framework for building a user interface.
Could be deployed on Hugging Face Space, which would allow to use the model in a web browser.

Will not have the function to edit non-generated images.

Will not be able to retrain the models with a custom dataset.

Will not use Amazon Mechanical Turk for evaluation.

Will not be compared with other image generators.

Will not implement sliders for changing the attributes of the generated image.

Non-functional Requirements

Must use Python 3 for development, various support libraries and big community support make it the best choice for this type of project.

Should be written in Google Colab or Jupyter Notebook environment, including extensive comments to clarify the code.

Should be developed in such a way to be easily reproducible at a point in future.

Should be usable on various operating systems (Windows, MacOS, Linux), as a result of a web based interface.

Should be used by a fashion designer, the system is not developed to be used by the general public.

Must use a version control system like Git, to simplify rolling back changes and prevent potential loss of progress, by regular backups.

The elemental functionality of the model should be ready by mid-January 2023 and the final version by mid-April 2023.

Design

In this section, we discuss the design of the AI-assisted swimwear design system. The user provides an inspiration image, which is then processed by an image-to-text model to generate a textual description. It is enhanced with appropriate keywords to create a text prompt that guides a text-to-image model in generating a swimwear design based on the initial inspiration. The focus of the project was on integrating pre-trained image-to-text and text-to-image models, as well as prompt engineering techniques. The design considerations include selecting the most appropriate pre-trained models, optimisation of the prompts as well as the overall workflow for the task of generating swimwear designs. A key decision in this project was to utilise pre-trained models, as supported by the findings from our literature review. It has been demonstrated that investing resources in improving existing models can yield satisfactory results, potentially even surpassing the performance of newly developed models from scratch.

System Overview

User input

The user provides an inspiration image to serve as the basis for the swimwear design. The image does not necessarily have to be related to swimwear or swimming. It can be any image that the user finds aesthetically pleasing, intriguing, or embodying specific characteristics that they would like to see incorporated into the swimwear. However, it is important to note that in the current version of the system, the image is transformed into a textual description, and some information may be lost during this conversion, especially when dealing with abstract or complex images. Future improvements to the system could focus on enhancing the image-to-text model or switch to an image-to-image model which will increase the ability to understand and describe complex and abstract images more accurately, ensuring that the generated swimwear designs better reflect the user's intended inspiration.

Image to text processing

An image to text model processes the input image and generate a text description representing the key element present in the image. For the purposes of this project over 20 free to use images were selected from the website Unsplash (2023).

Prompt enhancement

The system combines the textual description which relevant keyword to create a text prompt, it will allow to guide a text to image model to generate a swimwear design that reflects the inspiration image. This step plays a crucial role in the design process, as it allows for the most control over the creation, especially since we are working with pre-trained models.

Text to Image generation

The enhanced prompt is fed into a text-to-image model, which generates a swimwear design. The model aims to produce a unique and creative creation that reflects the user's inspiration image.

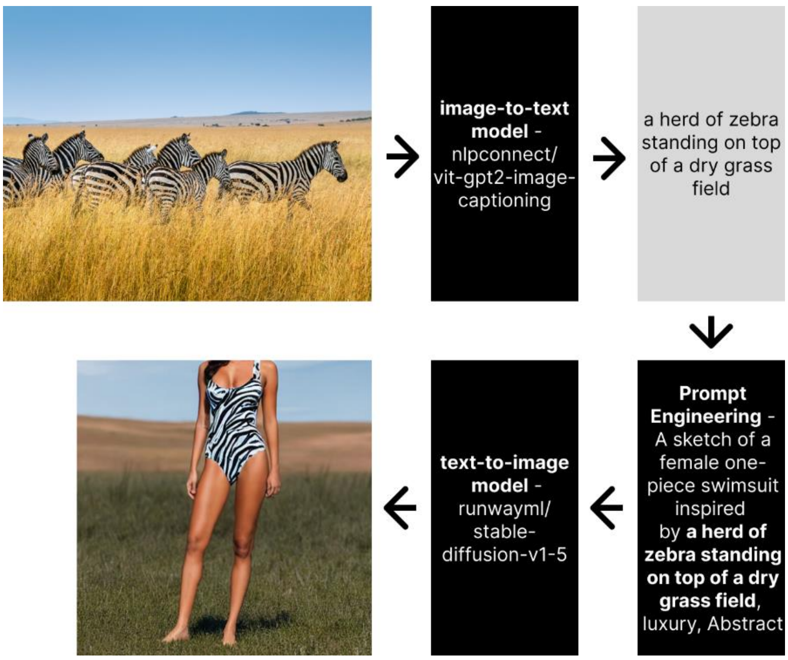


Figure 6 System overview

Software, tools, programming language

The Swimwear Designer project was developed using Python 3 as the primary programming language. Visual Studio Code (VSCode) served as the development environment. Project resources and source code were hosted on GitHub, enabling version control and collaboration. The repository is available at https://github.com/pawelsznura/Swimwear-Designer . These tools and languages effectively supported the development of the AI-powered swimwear design system.

Implementation

In this chapter, we discuss the implementation of all the elements of the AI swimwear designer, especially focusing on the selection and configuration of the pre-trained models, as well as the evolution of the prompt. We also address the challenges and problems encountered during the development. The additional tools developed for the evaluation or data analysis will be discussed in further sections.

Image to text

The image to text component is a well-established and widely researched technique, with numerous pre-trained models available that function in different ways.

Image classifiers

We evaluated the following image classifiers:

- facebook/convnext-base-224 (Facebook, 2022)
- facebook/regnet-y-008(Facebook, 2020)
- google/vit-base-patch16-224 (Google, 2020)
- microsoft/beit-base-patch16-224-pt22k-ft22k (Microsoft, 2021a)
- microsoft/resnet-50 (Microsoft, 2015)
- microsoft/swin-base-patch4-window7-224-in22k (Microsoft, 2021b)
- nvidia/mit-b0 (Nvidia, 2021)

These models were provided by HuggingFace.co, which utilises the Hosted Interface API, allowing calculations to be done on a remote server. However, it was observed that some of these models could yield results with varying levels of certainty. To tackle this challenge, we implemented a multi-model approach, calling multiple classifiers to analyse the images. We then developed a function that interprets the classifiers' responses, taking into account their certainty scores, and extracts the most confident classification.

Despite the convenience of using these classifies through the online service there were challenges associated with this approach. Firstly, the API calls could be slow, occasionally interrupted, or fail altogether, as the Interface API loads models on demand. This impacted the system's reliability, suggesting that running the models locally might be a better solution. Another issue with the image classifiers by design is their tendency to focus on a single element within the picture. These classifiers are often trained on the ImageNet dataset and can only choose one of the predefined categories. As a result, they might not capture the full range of elements and features in the inspiration image. After initial testing, it become evident that a different method was needed to address this limitation and better represent the input images in the text description.

Image captioning

To address the limitations encountered with image classifiers, we explored image captioning models, which are designed to generate more comprehensive and accurate textual descriptions of images. We focused on two pre-trained models: ydshieh/vit-gpt2-coco-en (Shieh, 2022) and nlpconnect/vit-gpt2-image-captioning (NLP Connect, 2022).

vit-gpt2-coco-en

This model combines the ViT architecture with GPT-2 for image captioning purposes. It is trained on the COCO dataset (Lin et al., 2015), a widely used resource for image captioning tasks, which includes a diverse range of images and associated captions. Although it is not a state-of-the-art model, it produces coherent and contextually relevant text descriptions of images.

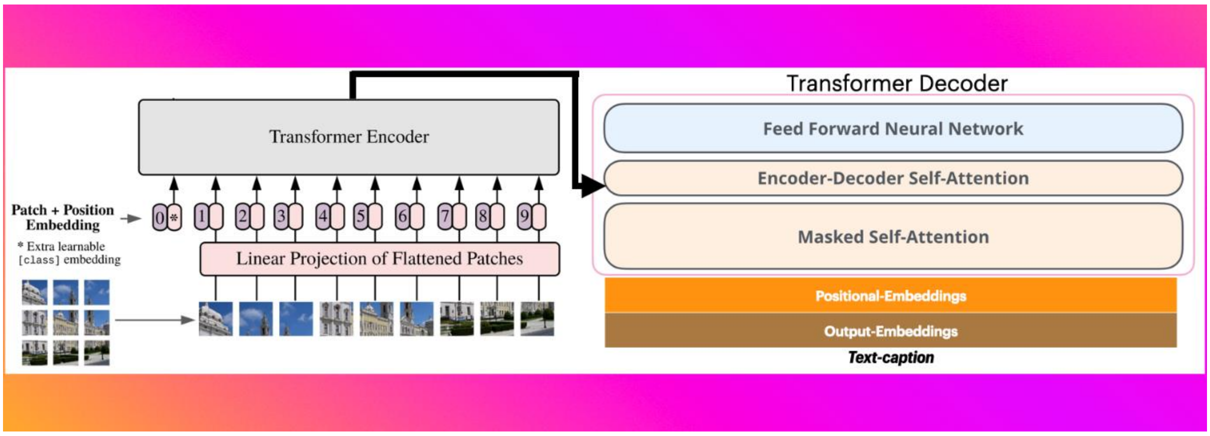


Figure 7 Vision Encoder Decoder Architecture (Kumar, 2022)

nlpconnect/vit-gpt2-image-captioning

The nlpconnect/vit-gpt2-image-captioning model is a PyTorch version of the previous model which was trained in Flax (Huggingface, 2022). It also leverages the ViT and GPT-2 architectures it is trained on the COCO dataset.

To integrate the models into our system we used the Hugging Face (Transformers, 2023) library, which provides a user-friendly interface for working with pre-trained models, streamlining the process of loading models and processing input images. The models were implemented according to the documentation. For each of the image captioning models, we first loaded the pre-trained model, feature extractor, and tokenizer using the transformers library. Then, we pre-processed the input images according to the model's requirements, converted them into the appropriate format, and extracted the image features using the ViTFeatureExtractor. After that, we used the VisionEncoderDecoderModel to generate text descriptions for the inspiration images by feeding the extracted features to the model.

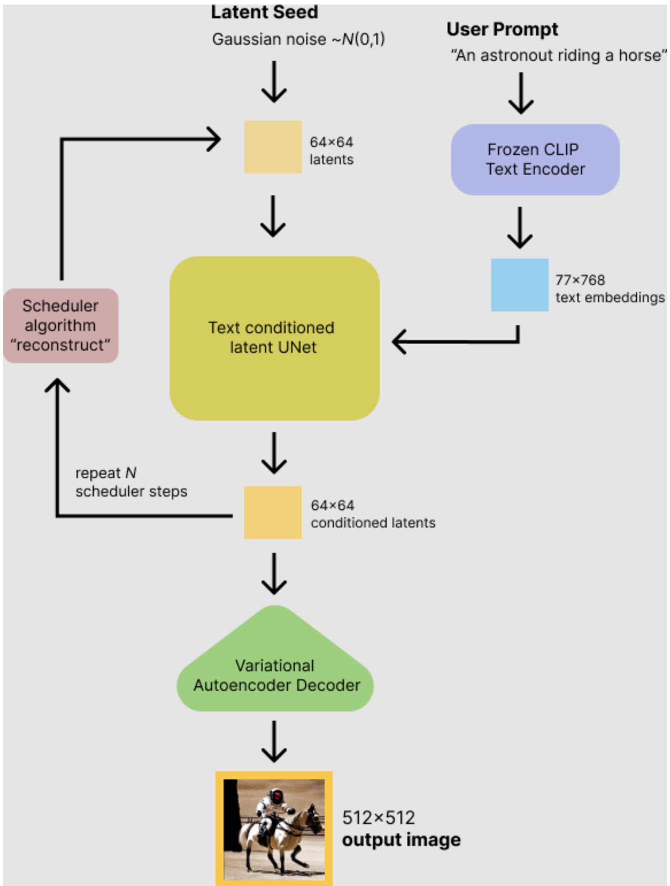


Figure 8 Stable Diffusion logical flow (Patil et al., 2022)

Prompt engineering

Prompt engineering is a critical step that significantly influences the final generated image in our design system. The system combines textual descriptions from the Image-to-Text model with relevant keywords to create an enhanced text prompt, guiding the Text-to-Image model in generating swimwear designs that reflect the inspiration image.

It is a relatively new field with no universally accepted best practices. During the exploration of various tactics, the Stable Diffusion search engines – Lexica (2023) and PromtHero (2023) proved to be very useful. It presents images alongside corresponding prompts, model specifications, and parameters. This resource allows for the identification of current trends and suggests changes for the iterative process of prompt improvement.

Prompts may consist of keywords or complete sentences. Keywords enable quick and relevant results but may lack context or produce ambiguous responses. Complete sentences provide more context, resulting in more accurate responses, though they may require additional tokens. In our case, we need to incorporate the text description from the Image-to-Text model, which is in the form of a sentence. For example, "a herd of zebra standing on top of a dry grass field." We add text to the front, such as "A sketch of a female one-piece swimsuit inspired by," to create a logical sentence, and append keywords like "luxury, abstract" at the end.

To identify the best prompt combination, we created a list of possible words and iterated through them using a for loop. The length of a prompt can play a vital role in its effectiveness. Longer prompts provide more context and details for the AI model. However, according to various sources, the maximum usable length of a Stable Diffusion text prompt is 75 tokens (Andrew, 2023; Maks-s, 2023).

Striking a balance between length and context is essential for optimal results. Although Stable diffusion uses the CLIP tokenizer (Patil et al., 2022), we can use GPT-3 tokenizer on OpenAI's website (*OpenAI API*, 2023) to get a rough estimate of the token count. Our example prompt, "A sketch of a female one-piece swimsuit inspired by a herd of zebra standing on top of a dry grass field, luxury, abstract," is equivalent to 29 tokens. It is important to remember that the text description may vary depending on the starting image, leaving room for prompt improvement without making it too long.

Negative prompts describe what we do not want to see in the picture, allowing for more refined output. These prompts may not be required, however, based on our project experience, negative prompts significantly improve results. Based on the finding from Stable Diffusion search engines, we selected some universal negative prompts and applied them to every generation process. While our project focused on specific prompting techniques, others exist, such as keyword weighting or special syntax, that were not covered in this project. Future iterations and improvements could explore these additional techniques to enhance prompt engineering and the resulting generated images.

Text to image

The Text-to-Image generation section focuses on the process of converting the enhanced text prompts into visually appealing swimwear designs. The was to utilise the available dedicated hardware (GPU AMD Radeon™ RX 6700S Mobile Graphics) to generate images using a pre-trained text-to-image model like Stable Diffusion. However, this process encountered several challenges, as detailed below. Running the selected models locally on the AMD GPU proved to be more challenging than initially anticipated. Consumer AMD graphics cards do not have official support for machine learning, and the available unofficial support, such as AMD ROCm (2023), only provides drivers the Linux operating system. Despite reports from some users (Poliquin, 2022) that it could work, the project encountered difficulties in using the AMD GPU for Stable Diffusion models.

Given the challenges faced, alternative solutions were explored to enable the generation of images using the available hardware. One possible solution involved using Microsoft DirectML (2023) and converting the Stable Diffusion model to ONNX format, as demonstrated in a tutorial (Averad, 2022). This approach allowed the model to run on the AMD GPU, but it had limitations. The generated images had poor quality, and the hardware constraints required to a low resolution. Ultimately, it was determined that using the CPU (AMD Ryzen™ 7 6800HS) for image generation would yield better results. Although this may have led to longer processing times, it ensured better image quality and a more reliable generation process. With more system memory the model was able to generate the standard resolution image of 512x512 pixels.

Hugging Face (2023) is a leading platform for AI models, offering a comprehensive selection of pre-trained models and associated Python libraries for implementation. To streamline the image generation process, we employed the StableDiffusionPipeline from diffusers library, which enabled us to efficiently assign all the necessary parameters for the model.

In addition to generating and saving the images, the project also captured and stored relevant information about the parameters used in each image generation instance. This included details such as the input image, text prompt, models used, and execution time. This record-keeping ensured that the project maintained a comprehensive history of the generation process, providing valuable insights for future improvements and iterations of the system.

Testing & Results

Input images and Image to Text Models

For this project, over 20 free-to-use images were selected from Unsplash (2023) to test the system. The image-to-text models, ydshieh/vit-gpt2-coco-en (Shieh, 2022) and nlpconnect/vit-gpt2-image-captioning (NLP Connect, 2022), were compared for their performance in generating textual descriptions. Both models produced similar results, which were satisfactory but not perfect, however quality level should allow us to meet the project requirements. The vit_gpt2_image_captioning model was chosen for the rest of the project.

Text to Image Models

Stable Diffusion v1.5 (Runway, 2022) and v2.1 (Stability AI, 2022) models were explored for their ability to generate high-quality, innovative swimwear designs, which correspond to the user provided inspiration image. The differences in generated image quality were subtle, but the computation time varied significantly. The v2.1 model took over 30 minutes for one generation on the available hardware, while the v1.5 model produced similar quality designs in less time, 5-7 minutes, when the parameters were identical for both models. Consequently, the v1.5 model was chosen for the project. The number of steps taken during the image generation process affected the output quality and computation time. With 30 steps, the process took about 4 minutes but resulted in notably lower-quality images. 50 steps proved to be a good balance between speed and quality, with additional steps not yielding evident improvements.

Figure 9 Generated Image 86.png – 30 steps

Figure 10 Generated Image 87.png – 50 steps

Surprisingly, there was a significant difference in the system's performance on Windows 11 22H2 and Ubuntu 22.04 Linux. Execution times on Linux ranged from 5 to 7 minutes, while Windows took between 11 and 13 minutes for one design generation. The cause of this disparity might be attributed to more background processes on Windows occupying the CPU, thus reducing available computational resources. Consequently, the images were generated on Linux for this study.

One of the drawbacks of the system was the generation of only one side of the swimwear design, ideally the system would generate all sides of the swimsuit. Attempts to generate both front and back sides of swimwear designs were unsuccessful. It was tried to guide the model to the desired output by adapting the prompt. In some cases, the system did not divide the image, while in others, image 78, the main part of the picture is the one side of the swimwear, and only a small portion of the image is dedicated to the other side, but it is barely visible, from the short fragment of the image we can only assume that it is the same swimwear design. In other instances, image 83, the image is divided in half as we wanted, however one part is showing the back side of the swimwear, and the other half is completely unrelated.

Figure 11 Generated Image 78.png

Figure 12 Generated Image 83.png

Stable diffusion has a safety filter which prevent generating explicit images (Rando et al., 2022), a few generated images were entirely black, indicating that the filter was triggered. Moreover, some images displayed only parts of the swimsuit. To address the issues the following phrases were added to the negative prompts “NSFW” and “body out of frame”.

Evaluation

The generated designs are evaluated using objective metrics (FID) and human survey. The human surveys involve participants assessing the realism, creativity, and appeal of the swimwear design, as well as the connection between the inspiration image and the final swimwear design.

Streamlit Image gallery

To facilitate a more user-friendly experience while exploring the generated swimwear designs, their descriptions, and the input images, a simple Streamlit web interface was developed. Streamlit is an open-source framework that enables developers to create custom web applications. The web interface allows users to view and interact with the generated images and their associated information seamlessly. This approach greatly simplifies the evaluation process, enabling to quickly identify which prompt tactics are effective and which ones need improvement. Previously, the images were stored in a folder while their descriptions were stored separately, making it challenging to quickly find and correlate the relevant information. The web interface also provides the ability to filter images by user scores, making it easier to focus on the most successful designs. Future work could include integrating a feature that allows users to directly score images within the UI, further streamlining the evaluation process. The Web interface is available at: https://pawelsznura-swimwear-designer-streamlitfront-j144vx.streamlit.app/

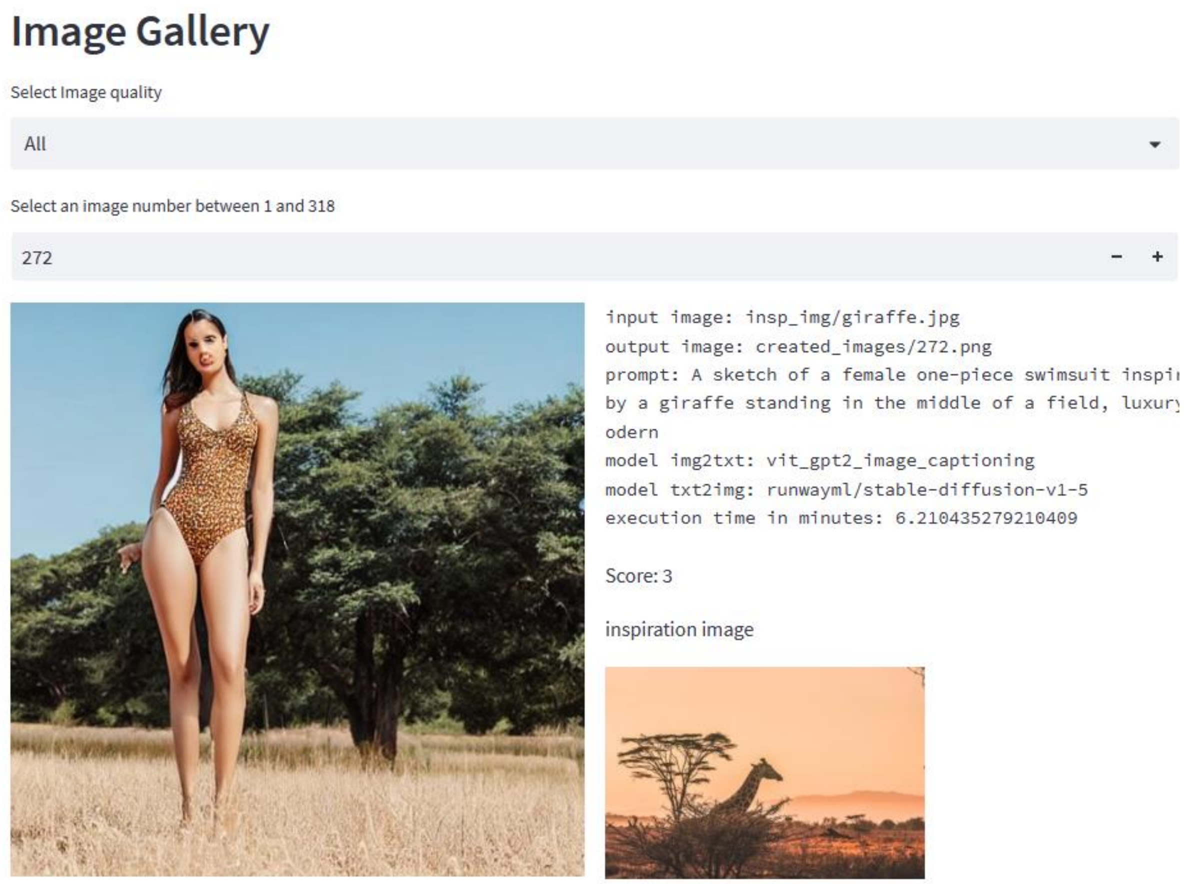


Figure 13 Screenshot of the Web UI

Fréchet inception distance

The Fréchet Inception Distance (FID) is a widely used evaluation metric for assessing the quality and diversity of generated images in generative adversarial networks (GANs) and other generative models. However, there are certain limitations and considerations, FID requires generated and real images in the ratio 1:1. We have over 300 generated samples, and we can use the ImageNet class “tank suit” with almost 700 images. However, it is recommended to use a minimum of 10,000 samples to calculate the FID otherwise the score will be underestimate (Heusel et al., 2018). Additionally, to avoid a calculational error, the feature layer value needs to be less than the number of samples, by default it is 2048, but in our case, we will use 64. In this project we use the FID implementation from Torch Metrics (2023).

With all the changes and limitations, we cannot objectively compare the FID scores with other systems. Although it can be useful to decide if our changes are improving the system, by calculating the FID for all generated images, and for the last 200, by which we are excluding the testing phase of different models, prompt techniques. The results are suggesting an improvement in the system. In future work, given more time and computational power we could increase the sample size and compare the FID metric with other system.

Table 2 FID results

Number of samples	FID
318 (All)	4.8265
200 (last)	4.3164

Inception Score

The Inception Score (IS) is another popular evaluation metric used to assess the quality and diversity of generated images in GANs. We have implemented it with the help of the inception_v3 pre-trained model from torchvision (PyTorch, 2023). However, it allowed us only to calculate the IS for 100 images at once, more was not possible due to insufficient system memory. The results were between 1.2 and 1.4. The IS metric is designed to evaluate the diversity and quality of generated images by comparing them to known object classes in the ImageNet dataset. (Barratt & Sharma, 2018) It assumes that a generative algorithm produces a diverse set of images, which is not the case in our project, therefore the IS is an inadequate metric for evaluation of our system.

Human Survey

To evaluate the effectiveness of our AI swimwear design system, we conducted a human survey targeting a general adult audience, in accordance with the ethics form provided in Appendix D. The survey aimed to assess the realism, creativity, appeal of the generated designs compared to real images and to discover if the connection between the starting image and generate is noticeable. The survey was distributed via social media, within the university, and among friends.

Image Selection

The generated images were scored on a scale of 0-5 based on their perceived usefulness for a fashion designer, their innovativeness, and the clarity of the design (where 0 represents not useful at all). We excluded images with scores of 0 and 1 from the survey and selected a fair distribution of image quality: 4 images with a score of 5, 3 with a score of 4, 3 with a score of 3, and 2 with a score of 2. The scores should not be treated as an objective metric, but rather as a guide to assist in the selection process for the survey. In addition, a few real images were included in the survey for comparison. In total, 12 generated images from our system and 3 real images from Unsplash (2023) were used.

Survey Method

The survey was conducted using Google Forms. Participants were informed about the purpose of the research, its voluntary and anonymous nature, and they had to confirm that they understood the information and were at least 18 years old. Each participant was shown an image of a swimwear design and asked if they thought the image was AI-generated, with response options of "yes," "no," or "not sure." They were then asked to rate the design on a scale of 1-5 for realism, creativity, and personal appeal. On the following page, participants were presented with the same swimwear image alongside its corresponding inspiration image and asked if they could see a connection between the swimwear design and the second image, also there was a text field for additional comments.

Survey Results

Between March 8th and 21st, we collected 30 responses. To get a broad overview of the results, we compared the mean ratings of each category for real and generated images. As the connection between the inspiration image and generated design is irrelevant in the case of real images, the data for this category is not included in the plot below. The high mean realistic value for generated images suggests that the system can generate images with realistic swimwear designs. Even the non-generated images did not achieve the highest possible score in this metric. Some generated images were so convincing that only around half of the participants could correctly identify the images as AI-generated. Overall, it can be concluded that out system has succeeded in creating realistic designs.

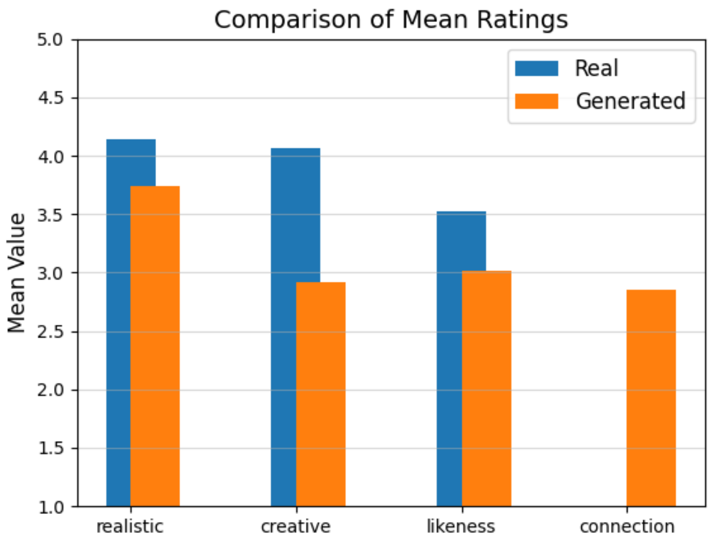


Figure 14 Comparison of Mean Ratings for Real vs. Generated Images

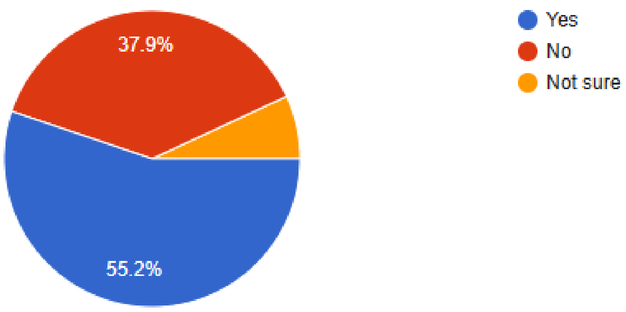


Figure 15 Question 8 – Image 155.png, Is that image AI generate?

Figure 16 Generated Image 155.png

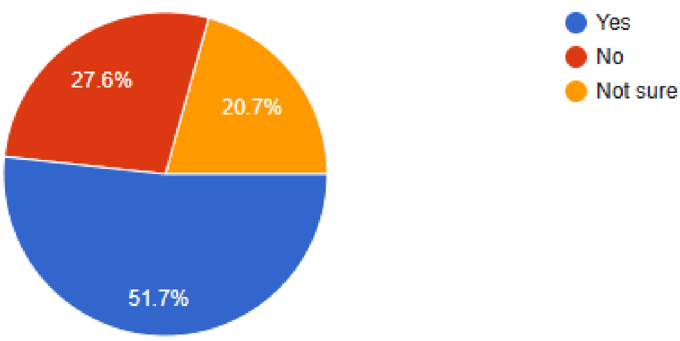


Figure 17 Question 9 – Image 17.png, Is that image AI generate?

Figure 18 Generated Image 17.png

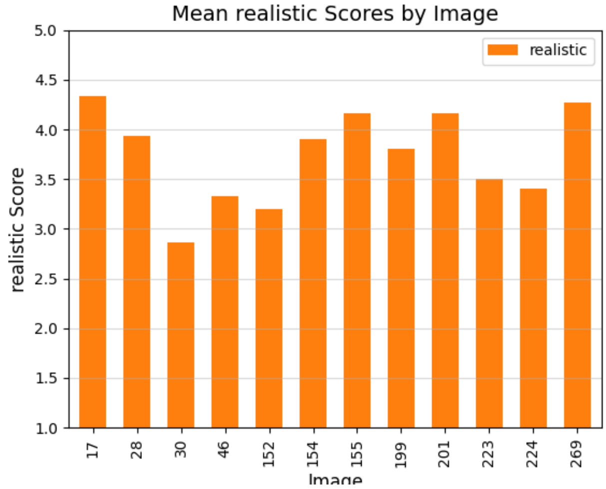


Figure 19 Mean realistic Score for Generated Images

One of the project requirements was to evaluate the aesthetics of the designs, with the goal of having at least half of the participants find the design appealing. Instead of asking a simple yes or no question, we provided a rating scale from 1 to 5. For the purposes of this requirement, we will assume that the mean rating in the upper half of the spectrum (>=3) means that the participants found the design appealing, while a rating in the lower half (<3) indicates that they did not.

When calculating the mean likeness value, we found that seven of the images were considered appealing and five were not. While this outcome meets the requirement, there is still significant room for improvement. However, it is worth noting that the values for each image were consistent, suggesting that the model's performance in generating aesthetically appealing designs was relatively stable.

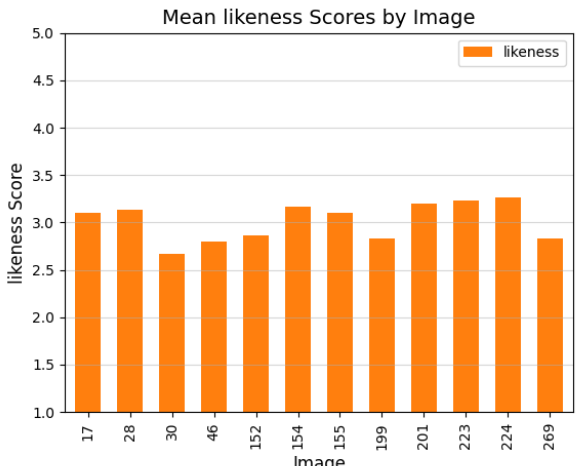


Figure 20 Mean likeness Score for Generated Images

Upon examining the plot comparing mean ratings between real and generated images, the creative and likeness values appear to be quite similar. One might initially assume that the results for individual images would also be similar; however, this assumption would be incorrect. In reality, the values exhibit considerable fluctuation, leading to the conclusion that the system is capable of producing very interesting designs, but at times also generates rather dull ones. This observation highlights the variability of the model's performance in terms of creativity and visual appeal. To address this issue, future work should focus on refining the model to enhance its consistency in generating captivating designs while minimising the occurrence of prosaic ones.

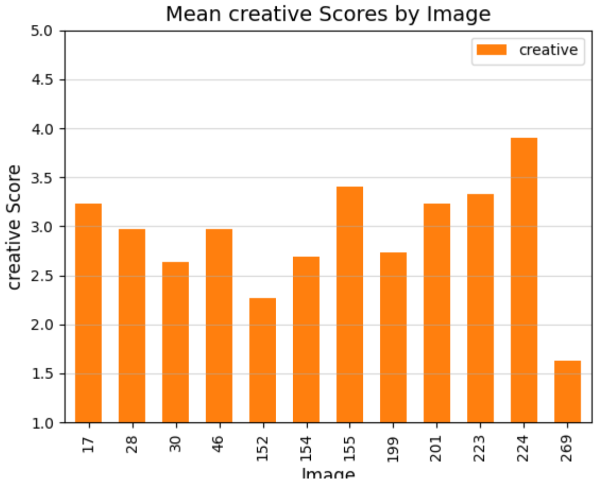


Figure 21 Mean creative Score for Generated Images

Figure 22 Generated Image 224.png – most creative.

Figure 23 Generated Image 269.png – least creative.

As for the connection between the inspiration image and the generated swimwear designs, we observed one outstanding example, while the remaining images were average or exhibited very weak connections. This is concerning, as establishing a strong connection between the inspiration and the final output is a fundamental aspect of our system. Consequently, future work should emphasize increasing the influence of the inspiration image on the generated designs.

It is important to note that survey participants were instructed to focus solely on the swimwear, rather than the background. In future iterations, refining the model to prioritise the connection between the inspiration image and the swimwear design may lead to more consistent and compelling results that better align with the initial vision of the project.

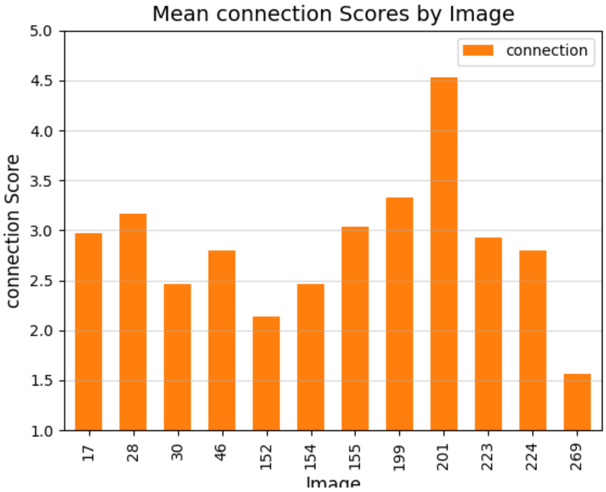


Figure 24 Mean connection Score for Generated Images

Figure 25 Generated Image 201.png – connection is most visible.

prompt: A sketch of a female one-piece swimsuit inspired by a herd of zebra standing on top of a dry grass field, luxury, Abstract

Figure 26 Inspiration Image for 201.png and prompt

Figure 27 Generated Image 152.png – one of the least visible connections

prompt: A drawing of a female one-piece swimsuit inspired by a train crossing a bridge over a lush green hillside, luxury, Modern

Figure 28 Inspiration Image for 152.png and prompt

Conclusion & Future work

Review of requirements

In this project, the primary aim was to create a functioning software for designing swimwear, powered by artificial intelligence. Most of the requirements were successfully met, while others were not achieved or partially fulfilled. This review aims to provide an overview of the current state of the project.

The project successfully implemented an image-to-text model for generating descriptions based on the input image, and a text-to-image model for generating images based on the text description. The chosen models, which were both pre-trained, aligned with the project's scope and limitations in terms of time and computational power. The image generation process utilised appropriate prompt techniques, adding relevant keywords like "swimsuit" and "swimwear" to guide the output towards the desired design.

The human survey evaluated whether the images depicted swimwear or not and assessed the design aesthetics. While the results showed that the system can generate realistic swimwear designs and at least half of the participants found the designs appealing. The connection between the inspiration image and the generated designs needs improvement, as the majority of the outputs demonstrated weak connections to the inspiration.

In terms of metrics, the Inception Score and Fréchet Inception Distance were calculated for model evaluation. However, due to the current state of the project and its limitations, it is not possible to directly compare these metrics with other systems. The generated images were saved in user-friendly format ".png". A user interface for exploring generated images was built using Streamlit, but generating new images through the interface is not possible.

On the other hand, the project failed to meet some requirements. The image generation process took longer than the desired two minutes, mostly because due to working on a CPU instead of a GPU. The project did not test or implement functionalities for modifying the swimwear type (e.g., two-piece swimsuits) or for allowing user text input when initial results were unsatisfying. Additionally, the project was not deployed on Hugging Face Spaces or similar platforms for web-based usage.

In summary, this project demonstrates that it is relatively easy to generate realistic images of swimwear using current AI techniques. The human evaluation results show that the AI generated designs are perceived as realistic, although they tend to lack creativity. Additionally, they suggest that the connection between the design and inspiration image is not always noticeable. Despite achieving a good FID score, the sample size is too small to compare the results to other models.

Future work

Future work on this project should focus on several key areas. First and foremost, it is crucial to keep up with the rapidly changing field of AI and generative models. As new research and methods emerge, they should be investigated and implemented to ensure optimal performance. During the project, various browser interfaces such as AUTOMATIC1111 (2023) gained sizable popularity with almost 70,000 stars on GitHub. These interfaces can operate on local hardware or make use of cloud services. Future work could investigate these solutions for possible improvements in user experience. In terms of performance improvement, future work should explore cloud-based GPU services or investing in dedicated machine learning hardware. This could help enhance the efficiency and performance of the Text-to-Image generation process, leading to faster generation times.

Collaboration with fashion designers and experts is another area that requires attention. Working closely with these experts would allow for better adjustment of the system to their needs and provide valuable input for evaluation. It is important to recognise that human expertise is still necessary for creating reliably good swimwear designs, and the system should be seen as a supplementary tool rather than a complete replacement. Future work may also involve utilising image-to-image models for better connection with the initial image and inpainting techniques to provide designers with more control. The techniques could allow for modifying the image and provide users with more control over the generated designs. Overall, the project lays the groundwork for applying AI in creative swimwear design while highlighting areas for improvement.

References

Alotaibi, A. (2020). Deep Generative Adversarial Networks for Image-to-Image Translation: A Review. *Symmetry 2020, Vol. 12, Page 1705*, *12*(10), 1705. https://doi.org/10.3390/SYM12101705

AMD ROCm. (2023). *Introduction to ROCm Installation Guide for Linux*. https://docs.amd.com/bundle/ROCm-Installation-Guide-v5.4/page/Introduction_to_ROCm_Installation_Guide_for_Linux.html

Andrew. (2023). *Stable Diffusion prompt: a definitive guide - Stable Diffusion Art*. https://stable-diffusion-art.com/prompt-guide/

Aouf, R. S. (2017). *Algorithm designs seven million different jars of Nutella*. https://www.dezeen.com/2017/06/01/algorithm-seven-million-different-jars-nutella-packaging-design/

AUTOMATIC1111. (2023). *AUTOMATIC1111/stable-diffusion-webui: Stable Diffusion web UI*. https://github.com/AUTOMATIC1111/stable-diffusion-webui

Averad. (2022). *Stable Diffusion on AMD GPUs on Windows using DirectML*. https://gist.github.com/averad/256c507baa3dcc9464203dc14610d674

Bao, H., Dong, L., Piao, S., & Wei, F. (2021). *BEiT: BERT Pre-Training of Image Transformers*. https://doi.org/10.48550/arxiv.2106.08254

Barratt, S., & Sharma, R. (2018). *A Note on the Inception Score*. https://arxiv.org/pdf/1801.01973.pdf

Bisen, V. S. (2020). *How AI is Changing Fashion: Impact on the Industry with Use Cases*. https://medium.com/vsinghbisen/how-ai-is-changing-fashion-impact-on-the-industry-with-use-cases-76f20fc5d93f#:~:text=Artificial%20Intelligence%20Impact%20on%20Fashion%20Industry&text=Customers%20can%20take%20a%20photo,fast%20through%20automated%20warehousing%20management.

Chong, M. J., & Forsyth, D. (2020). *Effectively Unbiased FID and Inception Score and Where to Find Them* (pp. 6070–6079). https://openaccess.thecvf.com/content_CVPR_2020/html/Chong_Effectively_Unbiased_FID_and_Inception_Score_and_Where_to_Find_CVPR_2020_paper.html

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2017). *Generative Adversarial Networks: An Overview*. https://doi.org/10.48550/arXiv.1710.07035

Dayma, B., & Cuenca, P. (2021). *Craiyon, formerly DALL-E mini*. https://www.craiyon.com/#faq

Dong, H., Yu, S., Wu, C., & Guo, Y. (2017). Semantic Image Synthesis via Adversarial Learning. *Proceedings of the IEEE International Conference on Computer Vision*, *2017-October*, 5707–5715. https://doi.org/10.48550/arxiv.1707.06873

Facebook. (2020). *facebook/regnet-y-008 · Hugging Face*. https://huggingface.co/facebook/regnet-y-008

Facebook. (2022). *facebook/convnext-base-224 · Hugging Face*. https://huggingface.co/facebook/convnext-base-224

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). *Generative Adversarial Nets*. http://www.github.com/goodfeli/adversarial

Google. (2020). *google/vit-base-patch16-224 · Hugging Face*. https://huggingface.co/google/vit-base-patch16-224

Google. (2022). *Overview of GAN Structure | Machine Learning | Google Developers*. https://developers.google.com/machine-learning/gan/gan_structure

Härkönen, E., Hertzmann, A., Lehtinen, J., & Paris, S. (2020). GANSpace: Discovering Interpretable GAN Controls. *Advances in Neural Information Processing Systems*, *2020-December*. https://doi.org/10.48550/arxiv.2004.02546

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. *Advances in Neural Information Processing Systems*, *30*. https://arxiv.org/abs/1706.08500

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2018). *bioinf-jku/TTUR: Two time-scale update rule for training GANs*. https://github.com/bioinf-jku/TTUR

Hugging Face. (2023). *Hugging Face – The AI community building the future.* https://huggingface.co/

Huggingface. (2021). *Distilled Data-efficient Image Transformer (base-sized model)*. https://huggingface.co/facebook/deit-base-distilled-patch16-224

Huggingface. (2022). *transformers/examples/flax/image-captioning at main · huggingface/transformers*. https://github.com/huggingface/transformers/tree/main/examples/flax/image-captioning

Intelistyle. (2022, June). *AI in Fashion: An Extensive Guide To All Applications For Retail - Intelistyle*. https://www.intelistyle.com/ai-fashion-retail-applications/

Karras, T., Laine, S., & Aila, T. (2019). *A Style-Based Generator Architecture for Generative Adversarial Networks Timo Aila NVIDIA*. https://github.com/NVlabs/stylegan

Knight, W. (2017, August 24). *Amazon Has Developed an AI Fashion Designer | MIT Technology Review*. https://www.technologyreview.com/2017/08/24/149518/amazon-has-developed-an-ai-fashion-designer/

Kumar, A. (2022). *The Illustrated Image Captioning using transformers - Ankur NLP Enthusiast*. https://ankur3107.github.io/blogs/the-illustrated-image-captioning-using-transformers/

Lexica. (2023). *Lexica*. https://lexica.art/

Li, B., Qi, X., Lukasiewicz, T., & Torr, P. H. S. (2019). Controllable Text-to-Image Generation. *Advances in Neural Information Processing Systems*, *32*. https://github.com/mrlibw/ControlGAN.

Lin, J., Xia, Y., Qin, T., Chen, Z., & Liu, T.-Y. (2018). *Conditional Image-to-Image translation*. https://arxiv.org/abs/1805.00251

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., & Dolí, P. (2015). *Microsoft COCO: Common Objects in Context*.

Maks-s. (2023). *Maks-s/sd-akashic: A compendium of informations regarding Stable Diffusion (SD)*. https://github.com/Maks-s/sd-akashic

Microsoft. (2015). *microsoft/resnet-50 · Hugging Face*. https://huggingface.co/microsoft/resnet-50

Microsoft. (2021a). *microsoft/beit-base-patch16-224-pt22k-ft22k · Hugging Face*. https://huggingface.co/microsoft/beit-base-patch16-224-pt22k-ft22k

Microsoft. (2021b). *microsoft/swin-base-patch4-window7-224-in22k · Hugging Face*. https://huggingface.co/microsoft/swin-base-patch4-window7-224-in22k

Microsoft DirectML. (2023). *microsoft/DirectML: DirectML is a high-performance, hardware-accelerated DirectX 12 library for machine learning. DirectML provides GPU acceleration for common machine learning tasks across a broad range of supported hardware and drivers, including all DirectX 12-capable GPUs from vendors such as AMD, Intel, NVIDIA, and Qualcomm.* https://github.com/microsoft/DirectML

NLP Connect. (2022). *nlpconnect/vit-gpt2-image-captioning · Hugging Face*. https://huggingface.co/nlpconnect/vit-gpt2-image-captioning

Nvidia. (2021). *nvidia/mit-b0 · Hugging Face*. https://huggingface.co/nvidia/mit-b0

Obukhov, A., & Krasnyanskiy, M. (2020). Quality Assessment Method for GAN Based on Modified Metrics Inception Score and Fréchet Inception Distance. *Advances in Intelligent Systems and Computing*, *1294*, 102–114. https://doi.org/10.1007/978-3-030-63322-6_8/FIGURES/7

OpenAI API. (2023). https://platform.openai.com/tokenizer

Patil, S., Cuenca, P., Lambert, N., & von Platen, P. (2022, August 22). *Stable Diffusion with 🧨 Diffusers*. Stable Diffusion with 🧨 Diffusers. https://huggingface.co/blog/stable_diffusion

Philips, M. (2018). *The Present and Future of AI in Design (with infographic) | Toptal*. https://www.toptal.com/designers/product-design/infographic-ai-in-design

Poliquin, M. (2022). *Stable Diffusion on AMD GPUs using ROCm on Linux | videogames.ai*. https://www.videogames.ai/2022/11/06/Stable-Diffusion-AMD-GPU-ROCm-Linux.html

PromtHero. (2023). *PromptHero - Search prompts for Stable Diffusion, ChatGPT & Midjourney*. https://prompthero.com/

PyTorch. (2023). *Inception_v3 | PyTorch*. https://pytorch.org/hub/pytorch_vision_inception_v3/

Qiao, T., Zhang, J., Xu, D., & Tao, D. (2019). *MirrorGAN: Learning Text-To-Image Generation by Redescription*. 1505–1514. https://arxiv.org/abs/1903.05854

Ramesh, A., Pavlov, M., Goh, G., & Gray, S. (2021). *DALL·E: Creating Images from Text*. https://openai.com/blog/dall-e/

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., & Sutskever, I. (2021). *Zero-Shot Text-to-Image Generation*. https://github.com/openai/DALL-E

Rando, J., Paleka, D., Lindner, D., Heim, L., & Tramèr, F. (2022). *Red-Teaming the Stable Diffusion Safety Filter*.

Rashad, F. (2022). *ClothingGAN - a Hugging Face Space by mfrashad*. https://huggingface.co/spaces/mfrashad/ClothingGAN

Rostamzadeh, N., Hosseini, S., Boquet, T., Stokowiec, W., Zhang, Y., Jauvin, C., & Pal, C. (2018). *Fashion-Gen: The Generative Fashion Dataset and Challenge*. https://doi.org/10.48550/arxiv.1806.08317

Runway. (2022). *runwayml/stable-diffusion-v1-5 · Hugging Face*. https://huggingface.co/runwayml/stable-diffusion-v1-5

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X., & Chen, X. (2016). Improved Techniques for Training GANs. *Advances in Neural Information Processing Systems*, *29*. https://github.com/openai/improved-gan.

Shen, Y., & Zhou, B. (2020). Closed-Form Factorization of Latent Semantics in GANs. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1532–1540. https://doi.org/10.48550/arxiv.2007.06600

Shieh, Y.-D. (2022). *ydshieh/vit-gpt2-coco-en · Hugging Face*. https://huggingface.co/ydshieh/vit-gpt2-coco-en

Stability AI. (2022). *stabilityai/stable-diffusion-2-1 · Hugging Face*. https://huggingface.co/stabilityai/stable-diffusion-2-1

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. *Journal of International Marketing*, *53*(1), 3645–3650. https://doi.org/10.48550/arxiv.1906.02243

Torch Metrics. (2023). *Frechet Inception Distance (FID) — PyTorch-Metrics 0.11.4 documentation*. https://torchmetrics.readthedocs.io/en/stable/image/frechet_inception_distance.html

Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., Jégou, H., & Ai, F. (2020). *Training data-efficient image transformers & distillation through attention*. https://arxiv.org/abs/2012.12877

Transformers. (2023). *🤗 Transformers*. https://huggingface.co/docs/transformers/index

Unsplash. (n.d.). *Beautiful Free Images & Pictures | Unsplash*. 2023. Retrieved April 23, 2023, from https://unsplash.com/

Xu, T., Zhang, P., Huang, Q., Zhang, H., Gan, Z., Huang, X., & He, X. (2017). AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1316–1324. https://doi.org/10.48550/arxiv.1711.10485

Yang, C., Shen, Y., Zhang, Z., Xu, Y., Zhu, J., Wu, Z., & Zhou, B. (2021). *ONE-SHOT GENERATIVE DOMAIN ADAPTATION Large-scale pre-training*. https://genforce.github.io/genda.

Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., & Metaxas, D. N. (2019). StackGAN++: Realistic Image Synthesis with Stacked Generative Adversarial Networks. *IEEE Transactions on*

Pattern Analysis and Machine Intelligence, *41*(8), 1947–1962. https://doi.org/10.1109/TPAMI.2018.2856256

Appendices:

Appendix A: Project Proposal

Detailed Project Proposal

First Name:	Pawel
Last Name:	Sznura
Student Number:	1910628
Supervisor:	Dr Pam Johnston

Defining your Project

1.1 Project title

Help: *a brief statement about what you are going to do.*

Applying AI Image generation methods in the process of Swimwear Design

1.2 Background

Help: *Provide the background to your project. This section should highlight the main topics in the area you are going to research. Essentially what is the project about, what has been done before and why is this project important? ~500 words*

The use of artificial intelligence for design is common in many industries. Stretching from UI on websites like Airbnb to jar packaging for Nutella. (Philips, 2018) In the fashion industry AI is used in many aspects, like forecasting trends, improving the supply chain or personalising the experience (Intelistyle, 2022), it follows as well the trend to use AI in design. In 2017, Amazon declared that they would start using an 'AI fashion designer' with the aim to replace conventional designers. (Knight, 2017)

The Generation of images was done before. For example, recently a lot of popularity gained Dall E when it launched the public mini version. It was one of the main inspirations for the project. Dall E is a text to image generator, which is using a dataset of text-image pairs. (Ramesh, Pavlov, Goh, & Gray, 2021)

However, this project is aiming for an image-to-image generation approach. GANs (Generative Adversarial Networks) (Goodfellow et al., 2014) are using a generator and a discriminator to create almost identical images compared to real ones. They rely on big number of training images. There have been attempts to train GANs with limited data, nonetheless, to achieve a satisfying performance were still hundreds or even thousands of images required. (Yang et al., 2021)

To assess those models many evaluation metrics were introduced. Inception Score (IS) aims to measure the quality and diversity of an image. (Salimans et al., 2016) At Amazon Mechanical Turk (AMT) human workers are scoring images to measure realism and faithfulness. (Alotaibi, 2020) Another metric is the Frechet inception distance (FID) which better reflects the similarity of the generated images to the real ones than the Inception Score by being less influenced by noise. (Heusel et al., 2017)

With fashion being such a wide topic, having in mind the complexity and the time constraints of research the project will only focus on swimwear. The goal of this project is not to replace fashion designers rather, shift the emphasis from creating to focusing more on inspiration and adjusting the design. The designer will select images for the colour, texture, style of the outfit. The model will combine this data and output preferably multiple designs. Similar how Dall E mini is generating 9 images from the same text input. During the second phase using the UI the designer will modify the creation. In an identical way as in ClothingGAN (Rashad, 2022)using sliders to change various attributes of the design.

References:

Alotaibi, A. (2020). Deep Generative Adversarial Networks for Image-to-Image Translation: A Review. *Symmetry 2020, Vol. 12, Page 1705*, *12*(10), 1705. https://doi.org/10.3390/SYM12101705

AMD ROCm. (2023). *Introduction to ROCm Installation Guide for Linux*. https://docs.amd.com/bundle/ROCm-Installation-Guide-v5.4/page/Introduction_to_ROCm_Installation_Guide_for_Linux.html

Andrew. (2023). *Stable Diffusion prompt: a definitive guide - Stable Diffusion Art*. https://stable-diffusion-art.com/prompt-guide/

Aouf, R. S. (2017). *Algorithm designs seven million different jars of Nutella*. https://www.dezeen.com/2017/06/01/algorithm-seven-million-different-jars-nutella-packaging-design/

AUTOMATIC1111. (2023). *AUTOMATIC1111/stable-diffusion-webui: Stable Diffusion web UI*. https://github.com/AUTOMATIC1111/stable-diffusion-webui

Averad. (2022). *Stable Diffusion on AMD GPUs on Windows using DirectML*. https://gist.github.com/averad/256c507baa3dcc9464203dc14610d674

Bao, H., Dong, L., Piao, S., & Wei, F. (2021). *BEiT: BERT Pre-Training of Image Transformers*. https://doi.org/10.48550/arxiv.2106.08254

Barratt, S., & Sharma, R. (2018). *A Note on the Inception Score*. https://arxiv.org/pdf/1801.01973.pdf

Bisen, V. S. (2020). *How AI is Changing Fashion: Impact on the Industry with Use Cases*. https://medium.com/vsinghbisen/how-ai-is-changing-fashion-impact-on-the-industry-with-use-cases-76f20fc5d93f#:~:text=Artificial%20Intelligence%20Impact%20on%20Fashion%20Industry&text=Customers%20can%20take%20a%20photo,fast%20through%20automated%20warehousing%20management.

Chong, M. J., & Forsyth, D. (2020). *Effectively Unbiased FID and Inception Score and Where to Find Them* (pp. 6070–6079). https://openaccess.thecvf.com/content_CVPR_2020/html/Chong_Effectively_Unbiased_FID_and_Inception_Score_and_Where_to_Find_CVPR_2020_paper.html

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2017). *Generative Adversarial Networks: An Overview*. https://doi.org/10.48550/arXiv.1710.07035

Dayma, B., & Cuenca, P. (2021). *Craiyon, formerly DALL-E mini*. https://www.craiyon.com/#faq

Dong, H., Yu, S., Wu, C., & Guo, Y. (2017). Semantic Image Synthesis via Adversarial Learning. *Proceedings of the IEEE International Conference on Computer Vision*, *2017-October*, 5707–5715. https://doi.org/10.48550/arxiv.1707.06873

Facebook. (2020). *facebook/regnet-y-008 · Hugging Face*. https://huggingface.co/facebook/regnet-y-008

Facebook. (2022). *facebook/convnext-base-224 · Hugging Face*. https://huggingface.co/facebook/convnext-base-224

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). *Generative Adversarial Nets*. http://www.github.com/goodfeli/adversarial

Google. (2020). *google/vit-base-patch16-224 · Hugging Face*. https://huggingface.co/google/vit-base-patch16-224

Google. (2022). *Overview of GAN Structure | Machine Learning | Google Developers*. https://developers.google.com/machine-learning/gan/gan_structure

Härkönen, E., Hertzmann, A., Lehtinen, J., & Paris, S. (2020). GANSpace: Discovering Interpretable GAN Controls. *Advances in Neural Information Processing Systems*, *2020-December*. https://doi.org/10.48550/arxiv.2004.02546

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. *Advances in Neural Information Processing Systems*, *30*. https://arxiv.org/abs/1706.08500

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2018). *bioinf-jku/TTUR: Two time-scale update rule for training GANs*. https://github.com/bioinf-jku/TTUR

Hugging Face. (2023). *Hugging Face – The AI community building the future.* https://huggingface.co/

Huggingface. (2021). *Distilled Data-efficient Image Transformer (base-sized model)*. https://huggingface.co/facebook/deit-base-distilled-patch16-224

Huggingface. (2022). *transformers/examples/flax/image-captioning at main · huggingface/transformers*. https://github.com/huggingface/transformers/tree/main/examples/flax/image-captioning

Intelistyle. (2022, June). *AI in Fashion: An Extensive Guide To All Applications For Retail - Intelistyle*. https://www.intelistyle.com/ai-fashion-retail-applications/

Karras, T., Laine, S., & Aila, T. (2019). *A Style-Based Generator Architecture for Generative Adversarial Networks Timo Aila NVIDIA*. https://github.com/NVlabs/stylegan

Knight, W. (2017, August 24). *Amazon Has Developed an AI Fashion Designer | MIT Technology Review*. https://www.technologyreview.com/2017/08/24/149518/amazon-has-developed-an-ai-fashion-designer/

Kumar, A. (2022). *The Illustrated Image Captioning using transformers - Ankur NLP Enthusiast*. https://ankur3107.github.io/blogs/the-illustrated-image-captioning-using-transformers/

Lexica. (2023). *Lexica*. https://lexica.art/

Li, B., Qi, X., Lukasiewicz, T., & Torr, P. H. S. (2019). Controllable Text-to-Image Generation. *Advances in Neural Information Processing Systems*, *32*. https://github.com/mrlibw/ControlGAN.

Lin, J., Xia, Y., Qin, T., Chen, Z., & Liu, T.-Y. (2018). *Conditional Image-to-Image translation*. https://arxiv.org/abs/1805.00251

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., & Dolí, P. (2015). *Microsoft COCO: Common Objects in Context*.

Maks-s. (2023). *Maks-s/sd-akashic: A compendium of informations regarding Stable Diffusion (SD)*. https://github.com/Maks-s/sd-akashic

Microsoft. (2015). *microsoft/resnet-50 · Hugging Face*. https://huggingface.co/microsoft/resnet-50

Microsoft. (2021a). *microsoft/beit-base-patch16-224-pt22k-ft22k · Hugging Face*. https://huggingface.co/microsoft/beit-base-patch16-224-pt22k-ft22k

Microsoft. (2021b). *microsoft/swin-base-patch4-window7-224-in22k · Hugging Face*. https://huggingface.co/microsoft/swin-base-patch4-window7-224-in22k

Microsoft DirectML. (2023). *microsoft/DirectML: DirectML is a high-performance, hardware-accelerated DirectX 12 library for machine learning. DirectML provides GPU acceleration for common machine learning tasks across a broad range of supported hardware and drivers, including all DirectX 12-capable GPUs from vendors such as AMD, Intel, NVIDIA, and Qualcomm.* https://github.com/microsoft/DirectML

NLP Connect. (2022). *nlpconnect/vit-gpt2-image-captioning · Hugging Face*. https://huggingface.co/nlpconnect/vit-gpt2-image-captioning

Nvidia. (2021). *nvidia/mit-b0 · Hugging Face.* https://huggingface.co/nvidia/mit-b0

Obukhov, A., & Krasnyanskiy, M. (2020). Quality Assessment Method for GAN Based on Modified Metrics Inception Score and Fréchet Inception Distance. *Advances in Intelligent Systems and Computing*, *1294*, 102–114. https://doi.org/10.1007/978-3-030-63322-6_8/FIGURES/7

OpenAI API. (2023). https://platform.openai.com/tokenizer

Patil, S., Cuenca, P., Lambert, N., & von Platen, P. (2022, August 22). *Stable Diffusion with 🧨 Diffusers*. Stable Diffusion with 🧨 Diffusers. https://huggingface.co/blog/stable_diffusion

Philips, M. (2018). *The Present and Future of AI in Design (with infographic) | Toptal*. https://www.toptal.com/designers/product-design/infographic-ai-in-design
Poliquin, M. (2022). *Stable Diffusion on AMD GPUs using ROCm on Linux | videogames.ai*. https://www.videogames.ai/2022/11/06/Stable-Diffusion-AMD-GPU-ROCm-Linux.html
PromtHero. (2023). *PromptHero - Search prompts for Stable Diffusion, ChatGPT & Midjourney*. https://prompthero.com/
PyTorch. (2023). *Inception_v3 | PyTorch*. https://pytorch.org/hub/pytorch_vision_inception_v3/
Qiao, T., Zhang, J., Xu, D., & Tao, D. (2019). *MirrorGAN: Learning Text-To-Image Generation by Redescription*. 1505–1514. https://arxiv.org/abs/1903.05854
Ramesh, A., Pavlov, M., Goh, G., & Gray, S. (2021). *DALL·E: Creating Images from Text*. https://openai.com/blog/dall-e/
Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., & Sutskever, I. (2021). *Zero-Shot Text-to-Image Generation*. https://github.com/openai/DALL-E
Rando, J., Paleka, D., Lindner, D., Heim, L., & Tramèr, F. (2022). *Red-Teaming the Stable Diffusion Safety Filter*.
Rashad, F. (2022). *ClothingGAN - a Hugging Face Space by mfrashad*. https://huggingface.co/spaces/mfrashad/ClothingGAN
Rostamzadeh, N., Hosseini, S., Boquet, T., Stokowiec, W., Zhang, Y., Jauvin, C., & Pal, C. (2018). *Fashion-Gen: The Generative Fashion Dataset and Challenge*. https://doi.org/10.48550/arxiv.1806.08317
Runway. (2022). *runwayml/stable-diffusion-v1-5 · Hugging Face*. https://huggingface.co/runwayml/stable-diffusion-v1-5
Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X., & Chen, X. (2016). Improved Techniques for Training GANs. *Advances in Neural Information Processing Systems*, *29*. https://github.com/openai/improved-gan.
Shen, Y., & Zhou, B. (2020). Closed-Form Factorization of Latent Semantics in GANs. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1532–1540. https://doi.org/10.48550/arxiv.2007.06600
Shieh, Y.-D. (2022). *ydshieh/vit-gpt2-coco-en · Hugging Face*. https://huggingface.co/ydshieh/vit-gpt2-coco-en
Stability AI. (2022). *stabilityai/stable-diffusion-2-1 · Hugging Face*. https://huggingface.co/stabilityai/stable-diffusion-2-1
Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. *Journal of International Marketing*, *53*(1), 3645–3650. https://doi.org/10.48550/arxiv.1906.02243
Torch Metrics. (2023). *Frechet Inception Distance (FID) — PyTorch-Metrics 0.11.4 documentation*. https://torchmetrics.readthedocs.io/en/stable/image/frechet_inception_distance.html
Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., Jégou, H., & Ai, F. (2020). *Training data-efficient image transformers & distillation through attention*. https://arxiv.org/abs/2012.12877
Transformers. (2023). *🤗 Transformers*. https://huggingface.co/docs/transformers/index
Unsplash. (n.d.). *Beautiful Free Images & Pictures | Unsplash*. 2023. Retrieved April 23, 2023, from https://unsplash.com/
Xu, T., Zhang, P., Huang, Q., Zhang, H., Gan, Z., Huang, X., & He, X. (2017). AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1316–1324. https://doi.org/10.48550/arxiv.1711.10485
Yang, C., Shen, Y., Zhang, Z., Xu, Y., Zhu, J., Wu, Z., & Zhou, B. (2021). *ONE-SHOT GENERATIVE DOMAIN ADAPTATION Large-scale pre-training*. https://genforce.github.io/genda.
Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., & Metaxas, D. N. (2019). StackGAN++: Realistic Image Synthesis with Stacked Generative Adversarial Networks. *IEEE Transactions*

on Pattern Analysis and Machine Intelligence, *41*(8), 1947–1962. https://doi.org/10.1109/TPAMI.2018.2856256

1.3 Aim & Objectives

Help: *Outline what are the main things your project is going to do and what steps or milestones will be used to achieve this aim. The Aim is unlikely to change throughout your project; however, the objectives are likely to adapt to your ongoing research and development.*

Example:
Aim: To create a functioning attendance application that efficiently automates the taking of class registers.
Objective 1: Study existing register system in place at RGU and identify weaknesses
Objective 2: Research existing automation technology's and identify and evaluate those that may be appropriate to taking in class registers
Objective 3: Implement chosen technologies to create prototype application
Objective 4: Conduct user trials to evaluate capabilities of prototype application
Objective 5: Create a refined application incorporating feedback from user trials

Aim: To create a functioning software for designing swimwear, powered by artificial intelligence.
Objective 1: Research existing technologies and identify useful approaches for this project
Objective 2: Find a suitable dataset and prepare it for use
Objective 3: Train and test a machine learning model
Objective 4: Evaluate and improve the model
Objective 5: Implement the user interface
Objective 6: Evaluate the final product

1.4 Tools & Technologies

Help: *Perform some initial research into the area and outline what tools and techniques you expect to be using in your project.*

Python and libraries like: Numpy, Scikit-learn, Pandas, PIL, PyTorch
Google Colab – to write and execute the code for the machine learning model
GAN – Generative adversarial network, for training the image-to-image model
ImageNet classification – to benchmark the result
Gradio or Streamlit – for building a User Interface
Hugging Face - for deploying the model

1.5 Project Plan

Help: *This is the project plan as to how you will go about achieving your objectives over the timescale of the Honours Project. At a minimum this can be a month-by-month plan.*

	October	November	December	January	February	March	April
Literature Review							
Dataset search and preparation							
Training and Testing the Model							
Evaluating and Improving the Model							
Implementing the UI							
Evaluating the final product							

1.6 Ethics Form

You must include in your signed ethics form in this submission, or you will not be able to continue the project.

Appendix B: Poster

Applying AI Image generation methods in the process of Swimwear Design

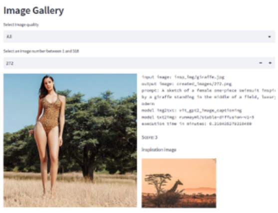
Using pre-trained Image-to-Text and Text-to-Image models

Pawel Sznura & Dr Pam Johnston

Aims & Objectives

AI design is a rising area of research, commonly used in many industries. Nonetheless, in more creative applications such as fashion design it is almost non existing. The primary aim of this project is to create a functioning software for designing swimwear, powered by artificial intelligence. The literature review showed that the necessary technology was already discovered and now it is the time to implement and adapt it. The objectives are as follows:

- Research existing technologies and identify useful image-to-text and text-to-image generation methods that can be effectively applied to this project.
- Develop a software system that utilises state-of-the-art pre-trained AI models to transform input images into swimwear designs.
- Test the system's performance and enhance the generation process by applying appropriate prompt engineering techniques based on the results.
- Evaluate the generated designs using appropriate metrics, including human survey to assess the realism, creativity, and appeal of the swimwear design.
- Implement a user interface to allow easy interaction with the AI-powered swimwear design software.



AI Generated Swimwear Gallery: tinyurl.com/AISwimwearDesign

Methods

image-to-text model - nlpconnect/vit-gpt2-image-captioning

a herd of zebra standing on top of a dry grass field

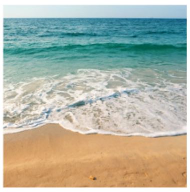
Prompt Engineering - A sketch of a female one-piece swimsuit inspired by **a herd of zebra standing on top of a dry grass field**, luxury, Abstract

text-to-image model - runwayml/stable-diffusion-v1-5

The AI-assisted swimwear design system is based on a user-provided inspiration image, which is processed through image-to-text (vit-gpt2-image-captioning) and text-to-image (stable-diffusion-v1-5) models. The design process consists of the following steps:

- User Input: The user provides an inspiration image, which can be any aesthetically pleasing, intriguing, or characteristic image they would like to incorporate into the swimwear design.
- Image-to-Text Processing: The input image is processed by an image-to-text model, generating a text description representing the key elements present in the image.
- Prompt Enhancement: The system combines the textual description with relevant keywords to create an enhanced text prompt, guiding the text-to-image model in generating a swimwear design reflecting the inspiration image.
- Text-to-Image Generation: The enhanced prompt is fed into a text-to-image model, producing a unique and creative swimwear design based on the user's inspiration image.

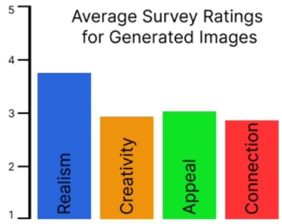
The project focuses on integrating pre-trained image-to-text and text-to-image models, as well as optimizing prompts and overall workflow for swimwear design generation. The generated designs are evaluated using objective metrics (FID) and human survey. The human surveys involve participants assessing the realism, creativity, and appeal of the swimwear design, as well as the connection between the inspiration image and the final swimwear design.






Conclusions

This project demonstrates that it is relatively easy to generate realistic images of swimwear using current AI techniques. The human evaluation results show that the AI-generated designs are perceived as realistic, although they tend to lack creativity. Additionally, they suggest that the connection between the design and inspiration image is not always noticeable. Despite achieving a good FID score, the sample size is too small to compare the results to other models. Future work may involve utilising image-to-image models for better connection with the initial image and inpainting techniques to provide designers with more control. Overall, the project lays the groundwork for applying AI in creative swimwear design while highlighting areas for improvement.



Acknowledgments

I would like to express my deepest gratitude to my supervisor, Pam, for her guidance and mentorship throughout the project. I am also incredibly grateful to my wife, Wiktoria, not only for her constant support and encouragement throughout the project but also for her valuable input as a fashion management student. Her initial scoring of the swimsuits provided essential insights and helped refine the AI-generated designs. Finally, I appreciate the survey participants whose feedback was crucial in evaluating the generated swimwear designs.

RGU SCHOOL OF COMPUTING

BSc (Hons) Computing (Application Software Development)

Appendix C: Project Log

commit df599dc7b14920df7d7f0016dcfec1fdd87f7085
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date: Fri Apr 28 01:47:18 2023 +0100

    minor changes to data analisys code

commit 565fcb41c485120a187f911b5df7a73546c3f462
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date: Wed Mar 29 21:59:01 2023 +0100

    survey data analysis

commit de2b576998f820ffdef550bb885d129eb5789332
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date: Wed Mar 22 14:51:51 2023 +0000

    some minor comments

commit 9ea2af36eca956c076e25e85e402b82b44cec331
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date: Tue Mar 21 14:44:17 2023 +0000

    small typo

commit bf71e69f98e2840484143a69bd1c662c1ad86475
Author: Pawel <p.sznura@rgu.ac.uk>
Date: Mon Mar 20 19:48:54 2023 +0000

    added comments to FID

commit 4c792e287da105b27020545b150349a0ab5e39fe
Author: Pawel <p.sznura@rgu.ac.uk>
Date: Mon Mar 20 18:39:51 2023 +0000

    frechet inception distance

commit 2bfef15b7729dd8c16b97ac1087d6b39561cbde4
Author: Pawel <p.sznura@rgu.ac.uk>
Date: Mon Mar 20 17:45:13 2023 +0000

    added imagenet "tank suit" class to the repository

commit 37bd1d34e11c9c88403020feb223d08c75123544
Author: Pawel <p.sznura@rgu.ac.uk>
Date: Mon Mar 20 14:24:02 2023 +0000

    inception score

```
commit 3c9a36830e8258291f453e0a204f94291da1154a
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date:   Mon Mar 6 21:49:05 2023 +0000

    updated reqs

commit 4a581a0f7018b7d5d1a1a471c5f3d383b76de886
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date:   Mon Mar 6 21:45:22 2023 +0000

    new streamlit version

commit 76310bac21ec72f04c00b31aa084ecc722c0f5c8
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date:   Mon Mar 6 21:27:49 2023 +0000

    added reqs file

commit ad301ab4d3ca717b308dad855d087aa90ec40261
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date:   Mon Mar 6 21:17:23 2023 +0000

    added score

commit 7fe5eb309e144c85c72dc1074b1bc23434cd3736
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date:   Mon Mar 6 20:57:02 2023 +0000

    select images by quality, display quality

commit 0a70e0793efff1cfd05f762039d9285ee6a4c8e8
Author: Pawel <p.sznura@rgu.ac.uk>
Date:   Mon Mar 6 19:12:12 2023 +0000

    new images

commit 2e86b031bc032f507238bdfd83ed9eaf5d333acb
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date:   Wed Mar 1 19:32:13 2023 +0000

    simple streamlit image gallery

commit b3c576d5339c730d48cf906b317d2bf5283b9b8e
Author: Pawel <p.sznura@rgu.ac.uk>
Date:   Wed Feb 22 11:51:45 2023 +0000

    result of 19h generating
```

commit aff20b7e8c729929f6889b5fae977f6ab6caf2d4
Author: Pawel <p.sznura@rgu.ac.uk>
Date: Tue Feb 21 17:00:02 2023 +0000

testing number of steps

commit fe95c68f5acaf9cc0ad8dcf7e5690883adc561f3
Author: Pawel <p.sznura@rgu.ac.uk>
Date: Mon Feb 20 14:12:57 2023 +0000

code cleanup

commit e12aee70933e03103491299ab698ee379d013eb0
Author: Pawel <p.sznura@rgu.ac.uk>
Date: Sat Feb 18 20:06:37 2023 +0000

next gen

commit 86f18668eb103bc1e82bf7725e2a1a7942946f28
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date: Mon Feb 13 23:44:39 2023 +0000

next gen of images

commit bdc5a0a25071170be7639ca74636a3ab58300989
Author: Pawel <p.sznura@rgu.ac.uk>
Date: Mon Feb 13 17:10:07 2023 +0000

testing with different RAM config

commit f9ba0f9e04e39b1b7358829f4c064760a2516208
Author: Pawel <p.sznura@rgu.ac.uk>
Date: Fri Feb 10 14:34:08 2023 +0000

still working with prompts

commit 65775a6802c3bf699560afecdcd28f9a17babdf2
Author: Pawel <p.sznura@rgu.ac.uk>
Date: Fri Feb 10 13:46:57 2023 +0000

added prompt options

commit de6869b2283e79c04a7f5d8834c8b124b6a39ecc
Author: Pawel <p.sznura@rgu.ac.uk>
Date: Wed Feb 8 22:06:48 2023 +0000

more images

```
commit c3e4fb6dd052e288c14a673993c309e885309d37
Merge: adb95e7 07dfe82
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date:   Wed Feb 8 18:43:34 2023 +0000

    Merge pull request #1 from pawelsznura/linux

    working on linux

commit 07dfe82427cf6ab9eb667dc89bb88c0adb603f39
Author: Pawel Sznura <p.sznura@rgu.ac.uk>
Date:   Wed Feb 8 15:20:48 2023 +0000

    new img

commit 535773575f1401e82f5d1c0edf4f58e36f97de0d
Author: Pawel Sznura <p.sznura@rgu.ac.uk>
Date:   Wed Feb 8 15:14:24 2023 +0000

    created a few images

commit b8d0a50d9a5a8a20f579944764d31c5c527cdf7c
Author: Pawel Sznura <p.sznura@rgu.ac.uk>
Date:   Tue Feb 7 17:07:09 2023 +0000

    working on linux

commit adb95e7de4abe86e6cc03487200fb6015f6faa57
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date:   Tue Feb 7 16:49:07 2023 +0000

    added timer to image description

commit c095a50d864837f464accbb893250b3a2915875b
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date:   Mon Feb 6 20:16:22 2023 +0000

    added run file to execute and new inspo images

commit 969bdc8c50300a41a8aca875b8406dcb009ffd59
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date:   Fri Feb 3 14:54:04 2023 +0000

    negative prompts

commit f50ff9278c9d001835ebd2ba6be1f2f653996b59
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
```

Date: Wed Feb 1 16:47:08 2023 +0000

    gpu onnx model can be accessed from main.py

commit 4131e53a00dc69ec8261d30087b354194c905c7e
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date: Fri Jan 27 14:43:51 2023 +0000

    SD 2.1 is so sloooow, over 30 minutes for image

commit ac60f22b57fb2315bdb79e4d5195de5ff150a043
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date: Fri Jan 27 13:22:43 2023 +0000

    tried out tabilityai/stable-diffusion-2-1

commit a607e58f1d431589c737fc28c293b8f3dcfa2680
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date: Thu Jan 26 12:35:55 2023 +0000

    save input img, output img, prompt, model txt2img, model img2txt

commit 5267c87024a22bb7fe63be263ce8bdab70e69a88
Merge: f1d9ab0 92a8302
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date: Sat Jan 21 15:12:15 2023 +0000

    Merge branch 'main' of https://github.com/pawelsznura/Swimwear-Designer

commit f1d9ab09be49ef7812b85e2dc155675748f2a892
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date: Sat Jan 21 15:11:40 2023 +0000

    table Diffusion for AMD GPUs on Windows using DirectML

    https://gist.github.com/averad/256c507baa3dcc9464203dc14610d674
    followed this tutorial

commit 92a830239d750c46593e5aa1f7749c27195c7f7b
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date: Wed Jan 18 13:53:37 2023 +0000

    Create README.md

commit a0af640352afe21026faaf13017ec308d71d5d44
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date: Wed Jan 18 13:42:10 2023 +0000

saves generated images in directory

commit 834d6b9d0e6c8e5ae98ee1453cbeceb41d74b54c
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date: Wed Jan 4 19:29:48 2023 +0000

added another image captioning

commit ea8f5948ca77bd7ef4699361fed53c1d907fa6a8
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date: Wed Jan 4 15:12:13 2023 +0000

image description

commit 19bdbe64268ed7d03e985d449e2ca87087b1741c
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date: Tue Jan 3 21:06:42 2023 +0000

new prompt for stable diffusion

commit 5a0425733a11c000dd6c5224184157cfecb9bcb3
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date: Tue Dec 27 19:48:16 2022 +0000

text2img stable diffusion

commit 474bb3a5df01d0bf970c448f3bee180785a1c6c6
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date: Tue Dec 27 17:18:05 2022 +0000

should work now, couldn't replicate the bug

commit 252a20f7a90645c0517a8950ac3df75b9f1bf137
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date: Tue Dec 27 16:35:50 2022 +0000

tried to resolve model not loaded bug, still not working

commit 321854e63652b240c06713014ca9ae481ecb2439
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date: Tue Dec 27 16:02:13 2022 +0000

fixed bug with multi and single word labels

commit 158f35fd2e6dc042896e2707c5a42b1e981d2573
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date: Tue Dec 27 15:51:35 2022 +0000

```
    returns labels with > 0.5 score

commit f23e0702f5c21476f2cd3a2e68055380a2a4c962
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date:   Tue Dec 27 12:36:11 2022 +0000

    different img2txt models

commit 20338898da3207d3f12ffd19e32061ab4bad485d
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date:   Sun Dec 18 21:19:36 2022 +0000

    code for easy img to class label prediction and added some insp img

commit 1d4a7e16670217e1cc2d196d0435ed88290818c4
Author: pawelsznura <91138465+pawelsznura@users.noreply.github.com>
Date:   Sun Dec 18 20:53:11 2022 +0000

    Initial commit
```

Appendix D: Ethics Form

STUDENT PROJECT ETHICAL REVIEW (SPER) FORM

The aim of the University's *Research Ethics Policy* is to establish and promote good ethical practice in the conduct of academic research. The questionnaire is intended to enable researchers to undertake an initial self-assessment of ethical issues in their research. Ethical conduct is not primarily a matter of following fixed rules; it depends on researchers developing a considered, flexible and thoughtful practice.

The questionnaire aims to engage researchers discursively with the ethical dimensions of their work and potential ethical issues, and the main focus of any subsequent review is not to 'approve' or 'disapprove' of a project but to make sure that this process has taken place.

The *Research Ethics Policy* is available at www.intranet.rgu.ac.uk/credo/staff/page.cfm?pge=7060

Student Name	Pawel Sznura
Supervisor	Dr Pam Johnston
Project Title	Applying AI Image generation methods in the process of Swimwear Design
Course of Study	Computing Application Software Development
School/Department	School of Computing

Part 1 : Descriptive Questions			
1	Does the research involve, or does information in the research relate to:	Yes	No
	(a) individual human subjects		X
	(b) groups (e.g. families, communities, crowds)		X
	(c) organisations		X
	(d) animals?		X
	Please provide further details:		
	The research involves conducting an anonymous survey with individual human subjects, all of whom are over 18 years old. These are research participants rather than subjects.		
2	Will the research deal with information which is private or confidential?	Yes	No
		X	
	Please provide further details:		
	The research will collect anonymous responses from participants, and their privacy will be maintained throughout the study.		

Part 2: The Impact of the Research			
3	In the process of doing the research, is there any potential for harm to be done to, or costs to be imposed on	Yes	No
	(a) research participants?		X
	(b) research subjects?		X
	(c) you, as the researcher?		X
	(d) third parties?		X
	Please state what you believe are the implications of the research:		
4	When the research is complete, could negative consequences follow:	Yes	No
	(a) for research subjects		X
	(b) or elsewhere?		X
	Please state what you believe are the consequences of the research:		

Part 3: Ethical Procedures			
5	Does the research require informed consent or approval from:	Yes	No
	(a) research participants?	X	
	(b) research subjects		X
	(c) external bodies		X
	If you answered yes to any of the above, please explain your answer: Participants will be informed about the purpose, procedures, and potential risks of the study, and their consent will be obtained before they take part in the survey. Participant responses will be held as long as is necessary to complete the project and then deleted.		
6	Are there reasons why research subjects may need safeguards or protection?	Yes	No
			X
	If you answered yes to the above, please state the reasons and indicate the measures to be		
7	Has PVG membership status been considered?		
	(a) PVG membership is not required.	X	
	(b) PVG membership is required for working with children.		X
	(c) PVG membership is required for working with protected adults.		X
	(d) PVG membership is required for working with both children and protected		X
	If you answered yes to (b), (c) or (d) above, please give details:		
8	Are specified procedures or safeguards required for recording, management, or storage of data?	Yes	No
		X	
	If you answered yes to the above, please outline the likely undertakings: Data collected from participants will be stored and managed securely. Access		

	to the data will be limited to the researcher and authorized personnel only. All data will be anonymized to ensure the confidentiality of participants.

Part 4: The Research Relationship			
9	Does the research require you to give or make undertakings to research participants or subjects about the use of data?	Yes	No
		X	
	If you answered yes to the above, please outline the likely undertakings: Research participant data will be anonymised as far as is possible, and will be stored in the anonymised format and its use limited for this research project only.		
10	Is the research likely to be affected by the relationship with a sponsor, funder or employer?	Yes	No
			X
	If you answered yes to the above, please identify how the research may be affected:		

Part 5: Other Issues			
11	Are there any other ethical issues not covered by this form which you believe you should raise?	Yes	No
			X

Statement by Student I believe that the information I have given in this form is correct, and that I have addressed the ethical issues as fully as possible at this stage.			
Signature	Szmura Pawel	Date	26.04.2023

If any ethical issues arise during the course of the research, students should complete a further Student Project Ethical Review (SPER) form.

The *Research Ethics Policy* is available at
www.intranet.rgu.ac.uk/credo/staff/page.cfm?pge=7060

Part 6: To be completed by the supervisor			
12	Does the research have potentially negative implications for the University?	Yes	No X
	If you answered yes to the above, please explain your answer:		
13	Are any potential conflicts of interest likely to arise in the course of the research?	Yes	No X
	If you answered yes to the above, please identify the potential conflicts:		
14	Are you satisfied that the student has engaged adequately with the ethical implications of the work? [In signifying agreement, supervisors are accepting part of the ethical responsibility for the project]	Yes X	No
	If you answered no to the above, please identify the potential issues:		
15	**Appraisal:** Please select one of the following		
	The research project should proceed in its present form – no further action is required	X	
	The research project requires ethical approval by the School Ethics Review Panel		
	The research project needs to be returned to the student for modification prior to further action		
	The research project requires ethical review by an external body. If this applies please give details		
	Title of External Body providing ethical review		
	Address of External Body		
	Anticipated date when External Body may consider project		

Affirmation by Supervisor			
I have read the student's responses and have discussed ethical issues arising with the student. I can confirm that, to the best of my understanding, the information presented by the student is correct and appropriate to allow an informed judgement on whether further ethical approval is required.			
Signature	P. Johnston	**Date**	28th April 2023

Appendix E: Project Code Repository

https://github.com/pawelsznura/Swimwear-Designer