Top Text

Tom Jacobs and James Boogaard

December 15th 2022
Module: SEAR
Venlo, Limburg, Netherlands

Abstract

This is the abstract.

Contents

List of Figures

1 Introduction

During the year of 2022 there has been a massive increase in the popularity and use of text-to-image generators, an artificial intelligence technique that translates human text into a computer-generated image. This has led to many different generators being made, which on the one hand has given artist, a person that uses a text-to-image generator to create an image with text, many more options but has also made the landscape difficult to navigate with its abundance of choice. These new options are great for the field as a whole, however it also means that artists who are not experienced may struggle to find the generator which suits their needs. This paper makes an effort to try and alleviate this problem by picking four of the most used text-to-image generators and comparing them based on their accuracy and process so we can help people gain some perspective on which text-to-image generator is best for their needs.

A research paper was published on the nineteenth of February 2020 called: "A survey and taxonomy of adversarial neural networks for text-to-image synthesis" that makes use of surveys to compare models for text-to-image generations. This paper aimed to show which text-to-image technique yielded the most realistic results in a wide variety of categories. One issue with this paper was that is was written before the explosion of consumer text-to-image generators happened, which is why we will instead focus on modern consumer-grade generators which are available today.

Our goal during the course of this research paper is to find out what is the best text-to-image generator, out of the four we chose, for artists who are experienced as well as artists who are not experienced in artificial art generation.

Another question we take a look at is to find the best technique out of the four applications that we use and then analysing why that technique is so effective for artificial intelligence text-to-image generation. The techniques we will take a look at are latent diffusion, the V4 model, and VOGAN and CLIP.

The text-to-image generators we will be testing during the course of this paper are: Stable diffusion, Midjourney, DallE and Dream by Wombo. Stable diffusion was released in 2022 and makes use of the ai text-to-image synthesis technique known as latent diffusion. Midjourney which is currently in open beta and also makes use of latent diffusion to generate its images. DallE was released in January of 2021 and makes use of the V4 model to generate its images. Finally is dream by Wombo which was released in late 2021 and makes use of a combination of VOGAN and CLIP for text-to-image generation.

The best text-to-image generator out of the four we chose will be stable diffusion. This is due to the fact that stable diffusion offers its users more tools and techniques like inpainting (allows you to regenerate a specific area of an image without regenerating the whole image) which allows the user more freedom to fine-tune their result. Given enough time this will yield a more accurate result.

2 Methods

Our goal during the course of this research paper is to find out what the best text to image generator is, out of the four we chose, for artists who are experienced as well as artists who are not experienced in AI art generation.

A secondary research question that we take a look at in this paper is to find out the strengths of the different ai text to image generation techniques used in the applications we have chosen.

This paper is used as an attempt to prove that stable diffusion is the best application out of the four, we chose. Our x value is the array of different applications that we use and the y value is the result of that (the distance) for each iteration on the road to trying to recreate an image using these applications. The intervening variable in this case would be experience using these applications.

We have chosen an image that we find sufficiently covers all the hard parts of recreating an image using ai art generation such as reflections and complex composition. We then use each of the four artificial intelligence applications to try and recreate that image within a given time frame (30 minutes) and after each iteration we take that image and compile it into a list of the iterations for the process of recreating that image for that specific application. With those sets of data, we then plug the images (each iteration) into an AI application that can give you a numerical value for how close one image is to another and see how close each iteration is to the image we are trying to recreate. this then gives us insight into how fast the application is (how efficient its technique is) and how many iterations you have to do to achieve a certain result. Once the data is gathered, we can create a graph to compare the applications more easily. The people that carried out this experiment also consisted of one novice and one experienced individual. This is to see which is the best application for each group of people and thus, the results will be conveyed on two separate graphs. In terms of our secondary research question this method also gives us insight if we are aware of the techniques each of them use we can notice if there is a pattern of success or failure with each.

3 Results

4 Discussion