Final thesis video presentation

LJMU Masters in ML and AI

MARCH 2023

Details

- Student Name: Rohit Kale
- Student Id: 1038020
- Project Name: Image to Image translation using deep learning techniques for visualisation in education industry for artists.
- Thesis supervisor Name: Ankan Dutta

Background

- Art has been used to express the imagination and creativity of artists and to communicate messages to the audience. In the past, creating art required a high level of skill and training, as artists had to master various techniques and materials to create their works.
- However, with the advent of technology and the development of artificial intelligence (AI) and machine learning (ML) algorithms, the creation of art has become more accessible and inclusive
- This technology allows artists to express their imagination directly into their work, without the need for extensive technical skills or training.
- Specifically, the study will investigate the accuracy and effectiveness of these algorithms in capturing the nuances of the artist's imagination and converting them into visual representations.

Problem statement

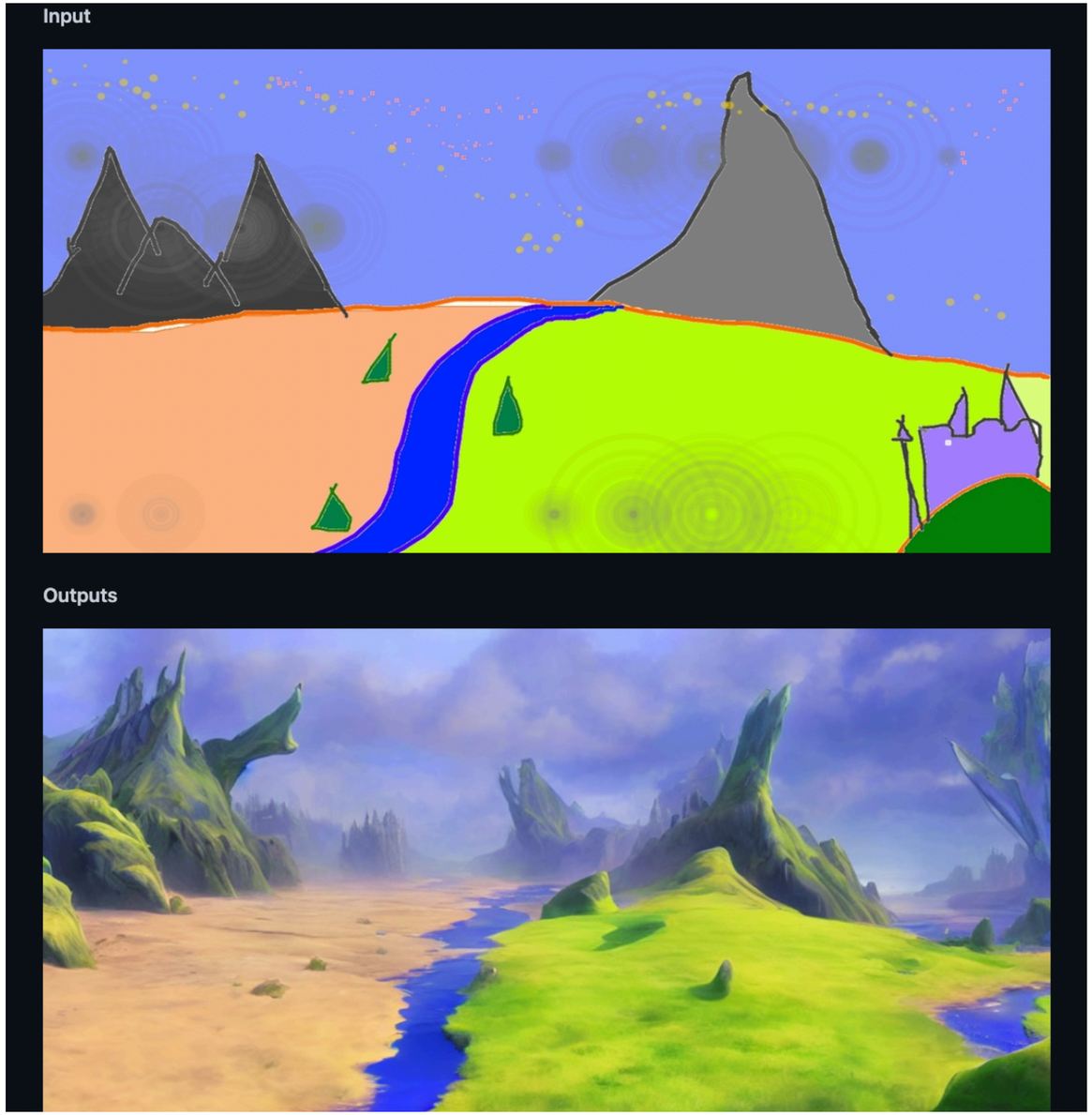
- The advent of artificial intelligence (AI) and machine learning (ML) technologies has the potential to address these issues by providing a means for individuals to create art without requiring extensive training or skill. Specifically, the use of AI and ML algorithms for the translation of sketches into images offers a way for artists to directly express their imagination into their work, bypassing the need for technical proficiency.
- However, despite the promise of this technology, there are several challenges that must be addressed. First, the accuracy and effectiveness of the AI and ML algorithms in capturing the nuances of the artist's imagination and translating them into visual representations must be evaluated. Second, the potential impact of this technology on the traditional art-making process must be considered, including the potential displacement of skilled artists and the implications for the art market.
- Therefore, the problem that this study seeks to address is how to effectively use AI and ML algorithms to translate sketches into images for art purposes, while also addressing the potential challenges and implications of this technology for the field of art. The study aims to contribute to the development of a more inclusive and accessible art community, while also preserving the value of traditional art-making processes.

Aim of thesis

The aim of this thesis is to translate sketches into images for artistic purposes, enabling individuals to express their imagination directly into their work without requiring extensive technical skill.

To achieve this aim, the following objectives will be pursued:

- To review the existing literature on the use of AI and ML in art creation, with a particular focus on the translation of sketches into images.
- To develop a prototype system that uses AI and ML algorithms to translate sketches into images for artistic purposes.
- To investigate the potential impact of this technology on the traditional art-making process, including the potential displacement of skilled artists and the implications for the art market.
- To explore the potential applications of this technology in promoting inclusivity and diversity in the field of art, by enabling individuals with limited technical skills to express their imagination and creativity through art.



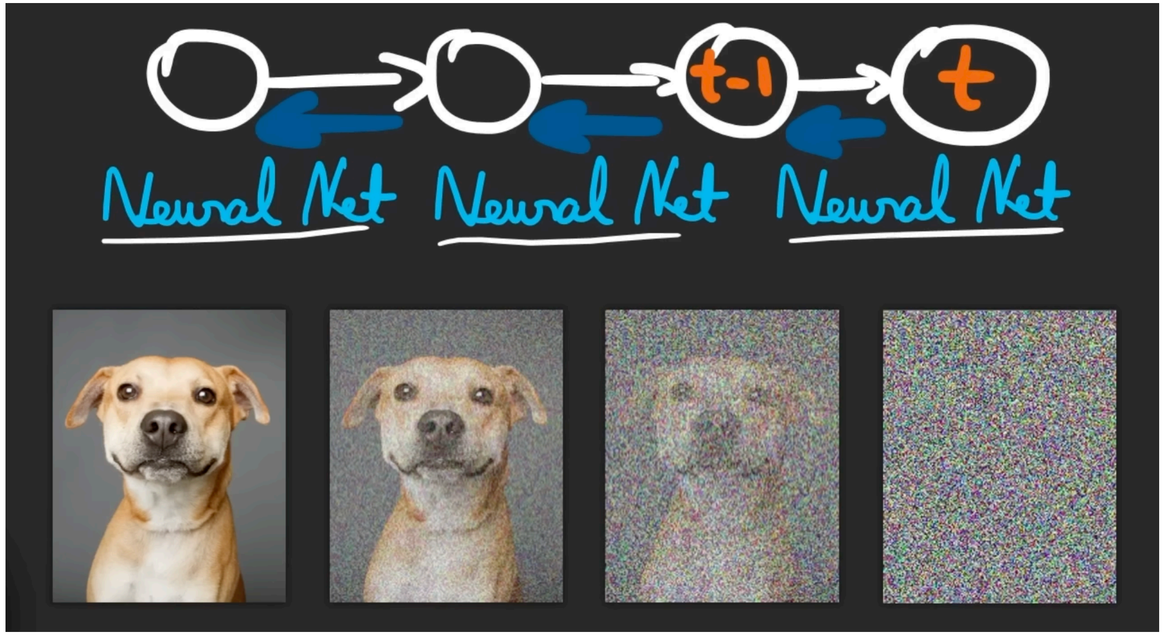
Literature Review

- Generative adversarial networks (GANs): GANs are a type of machine learning model that consist of two neural networks: a generator network and a discriminator network. The generator network is trained to generate synthetic images that are like the training data, while the discriminator network is trained to distinguish between real and synthetic images. GANs have been used successfully for sketch-to- image translation, as they are able to generate realistic images based on simple sketches.
  - CycleGAN: This algorithm uses a GAN to learn a mapping between two different image domains, such as photographs and paintings.
  - StyleGAN: This algorithm uses a GAN to generate high-quality synthetic images, including portraits and landscapes.

- Convolutional neural networks (CNNs): CNNs are a type of neural network that are particularly well-suited to image processing tasks. They have been used for sketch-to-image translation by training them on a large dataset of images and sketches, and then using them to generate synthetic images based on new sketches.
- Variational autoencoders (VAEs): VAEs are a type of neural network that can learn a compact representation of a dataset, known as a latent space. They have been used for sketch-to-image translation by training them on a dataset of images and sketches, and then using them to generate synthetic images by sampling from the latent space.
- Deep reinforcement learning: This is a type of machine learning that involves training an agent to perform a task by rewarding it for achieving certain goals. Deep reinforcement learning has been used for sketch-to-image translation by training an agent to generate images that match a given sketch, and then rewarding it for generating images that are more realistic or that better match the sketch.

Research Methodology

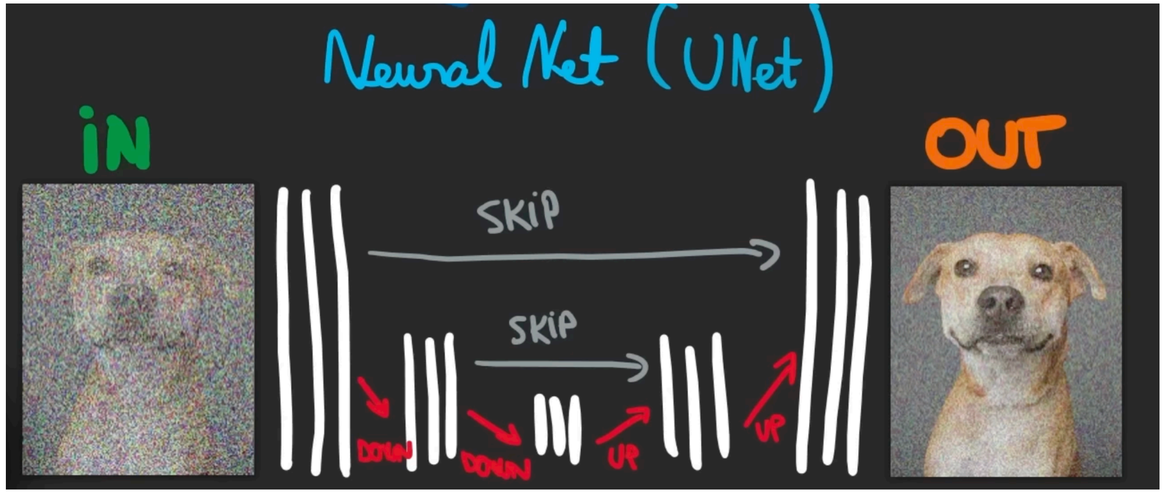
- Stable diffusion is a latent diffusion model. Diffusion process is where you diffuse more and more noise into your image. Take an image, and in t steps gradually add more noise to it until at the last timestep t the image is approximately just noise. Diffusion model follows the backward diffusion process. Follow each of the t steps and reduce the noise gradually, step by step. For this use the same neural network, usually a U-Net to go from t to t-1, t-1 to t-2 unto the first image.
- U-Net: Convolution-based neural network that is down sampling an image into lower dimensional representation and reconstructs it during up sampling. The down sampling and up sampling stacks of layers communicate through skip connections. Input: Image at step t Output:
- Total noise that should be subtracted from the noisy version of the image at step t to reconstruct the original image. The diffusion process is all about going from a little noise to more noise. The backward diffusion process is the reverse. The U-Net that we apply at each step gets a noise image at step t and predicts the whole noise that the image contains, and not just the noise we need to subtract to get from t to t-1. There is no trust for the model to just subtract this whole noise in one go, so at each step, we extract just a fraction of the total noise from the image at timestep t.
-



Research Methodology

It is easier to inject textual information gradually than injecting all at once. How process of injection works:

- Input the diffusion model by concatenating the text representation coming from a language transformer to the image input
- Cross-attention, letting the U-Net attention later attend to the text tokens. The Idea of stable diffusion was introduced in this CVPR paper When trying to generate a large image such as 1024*1024-dimensional noise grid and produce an image out of it. As you can imagine this can become expensive for one diffusion step and one has to do it t times where t can be something like 150. To circumvent this problem as we have seen in GLIDE is to train their diffusion model on much smaller images like 256*256. And then have an extra neural network that learned to up sample and sharpen 256*256 to higher resolution.
- LDMs are similar but cleverer to previous work that GLIDE did. GLIDE down sampled to 256*256 and stayed in image space. Then ran diffusion and up sampled the result to get a high- resolution image. What makes stable diffusion so special and fit for art generation, is that it was an LDM trained on a core dataset consisting of LAION-Aesthetics, a soon to be released subset of LAION-5B . LAION-Aesthetics was created with a new CLIP based model that filtered LAION-5B based on how beautiful images are. This different training data makes it different from DALL-E2, or Imagen.



Dataset training

- LAION-5B is a dataset of hand-drawn sketches that was created for use in sketch-to-image translation research. The dataset consists of over 5 million sketches in a variety of categories, including animals, people, plants, and objects. The sketches in the dataset were drawn by a diverse group of artists and are of varying styles and quality.
- One of the main features of the LAION-5B dataset is its size, which makes it one of the largest available datasets for sketch-to-image translation research. The large number of sketches in the dataset allows researchers to train and evaluate machine learning models on a diverse and representative set of data, which can help to improve the generalizability and performance of the models.
- In addition to the sketches, the LAION-5B dataset also includes corresponding images for each sketch, which can be used as reference images for the sketch- to-image translation process. These images were selected to match the style and content of the sketches as closely as possible, and they provide a benchmark against which the performance of the machine learning models can be evaluated.
- The LAION-5B dataset is a valuable resource for researchers working on sketch-to-image translation, as it provides a large and diverse set of data that can be used to train and evaluate machine learning models. It is also a useful resource for artists and designers who are interested in using sketch-to-image translation algorithms in their work.

Dataset sampling

- ImageNet-Sketch is a dataset of hand-drawn sketches that has been widely used in research on sketch-to-image translation and other related areas. Here are a few examples of papers that have cited ImageNet-Sketch in their research:

- This dataset contains over 50,000 pairs of sketches and corresponding photographs from the ImageNet dataset. The sketches were created by human annotators using the "Quick, Draw!" game developed by Google, where the user is given a prompt and has 20 seconds to draw a picture of it.
  The photographs in the dataset are from a subset of the ImageNet dataset and are labeled with the same categories as the sketches. The goal of this project is to use machine learning algorithms to translate the sketches into realistic images, allowing users to bring their imaginative sketches to life.

- the dataset required some cleaning before it could be used in our implementation. We first had to pick out a small number of samples from the dataset, since processing the entire dataset would be time-consuming and resource-intensive. To do this, we randomly selected a few hundred sketches from different categories and saved them in separate folders based on their category names.

- Next, we uploaded these sample sketch folders to our Google Drive, which would allow us to easily access them from the Colab notebook where we would be implementing Stable Diffusion. The dataset cleaning process was critical in ensuring that our implementation was efficient and focused only on the sketches and categories that were relevant to our research question.

Implementation

- The implementation of the model using the from_pretrained method from the StableDiffusionImg2ImgPipeline class. The StableDiffusionImg2ImgPipeline is a class from the diffusers library that provides an implementation of the Stable Diffusion process for image to image translation.
- Here, we first open the image using the Image class from the PIL library and convert it to the RGB format. We then resize the image to (768, 512) to make it compatible with the model's input size. Next, we set the generator for the model, which is used to generate random numbers during the Stable Diffusion process.
  - **model_name_or_path**
  - **image_size**
  - **diffusion_steps**
  - **timestep_respacing**
  - **start_scale**
  - **num_resolutions**
  - **rescale_timesteps**
  - **use_fp16**

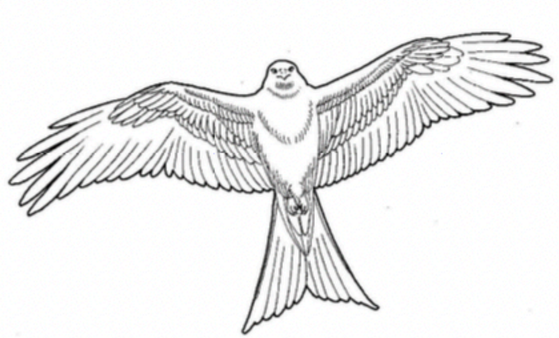
Hyperparameter tuning

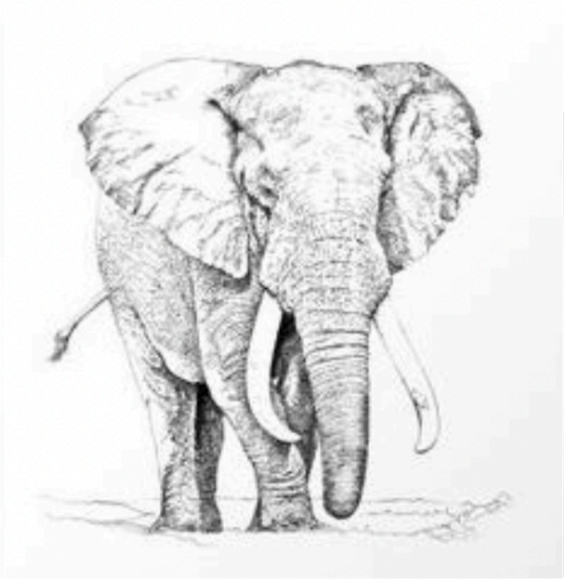
It takes the following hyperparameters:

- **model_name_or_path**: This parameter can be a string representing the model name or path to a checkpoint directory.
- **image_size**: This parameter represents the size of the input image. For our implementation, we set it to (768, 512).
- **diffusion_steps**: This parameter represents the number of diffusion steps to take during the Stable Diffusion process. We set it to 1000.
- **timestep_respacing**: This parameter represents the time step spacing for the Stable Diffusion process. We set it to 100.
- **start_scale**: This parameter represents the start scale for the Stable Diffusion process. We set it to 0.
- **num_resolutions**: This parameter represents the number of resolutions for the Stable Diffusion process. We set it to 4.
- **rescale_timesteps**: This parameter represents whether to rescale the time steps between resolutions. We set it to True.
- **use_fp16**: This parameter represents whether to use 16-bit precision for the Stable Diffusion process. We set it to True.

Result and Conclusion

- In this chapter, we present the results and discussions of our experiments using StableDiffusionImg2ImgPipeline to convert a sample dataset of sketches to an image dataset. Stable diffusion is a technique for image denoising and generation that has gained popularity in recent years due to its ability to handle heavy-tailed noise and generate high-quality samples.
- Firstly, the current implementation of StableDiffusionImg2ImgPipeline is computationally expensive and requires a significant amount of memory. This limits its practical applications in real-world scenarios.
- Secondly, the performance of StableDiffusionImg2ImgPipeline heavily depends on the quality of the initial sketches. Therefore, improving the quality of the sketches can lead to better results.
- StableDiffusionImg2ImgPipeline in the domain of image-to-image translation. Further research is required to address its limitations and explore its practical applications in real-world scenarios.





Evaluation of methodology

- The evaluation of the results obtained from the experiments conducted using the StableDiffusionImg2ImgPipeline class reveals that the performance of the model varies significantly with changes in the strength and guidance scale variables.
- In particular, we observed that increasing the strength variable led to significant improvements in the quality of the generated images. This is because the strength variable controls the amount of smoothing applied to the image during the diffusion process, and a higher strength value leads to smoother and more realistic images.
- However, we also observed that increasing the guidance scale variable beyond a certain point can lead to the model producing images that deviate significantly from the original sketch. In some cases, the generated images seemed way beyond imagination, indicating that the model had introduced too much noise or information into the image.
- On the other hand, in some cases, we observed that the generated images changed the whole perspective of the original imagination. This indicates that the model had learned to capture the essence of the sketch and generate images that were not only faithful to the original but also creatively extended the idea.
- Overall, our evaluation suggests that the StableDiffusionImg2ImgPipeline class is a powerful tool for translating hand-drawn sketches into digital images, but careful tuning of the strength and guidance scale variables is necessary to ensure high-quality and faithful results. Moreover, the model has the potential to generate images that go beyond the original imagination and lead to new and exciting creative directions in design and art.

Future recommendations

- Test the Stable Diffusion model on a larger dataset.
- Compare the results with other generative models.
- Explore the impact of different hyperparameters on the performance of the model.
- The use of different prompts for generating images can be investigated to determine the effect of prompts on the quality of generated images.

THANK YOU