FAU Friedrich-Alexander-Universität
Department Artificial Intelligence
in Biomedical Engineering

IDEA Lab
Image Data Exploration
and Analysis Lab

Evaluating the Performance of StyleGAN2-ADA on Natural Landscapes

A Validation of the Model's Results and Experiment on Custom Dataset

Term Paper

in the context of the seminar "Project Representation Learning"

at Friedrich-Alexander-Universität Erlangen-Nürnberg
at the Department Artificial Intelligence in Biomedical Engineering (AIBE)
Image Data Exploration and Analysis (IDEA) Lab

Principal Supervisor:	Prof. Dr. Bernhard Kainz
Associate Supervisor:	Mischa Dombrowski
Author:	Krishnachander Govindarajan Falkenstr.25 91056 Erlangen +49 175 6853865 krishnachander.govindarajan@fau.de 23032054
Submission:	19th March 2023

Abstract

Training Generative Adversarial Networks (GANs) is a computationally expensive and challenging task that typically demands large datasets. However, the StyleGAN2-ADA offers a solution to this problem by enabling faster training and requiring smaller datasets. The paper first reproduces and validates the results of the original StyleGAN2-ADA paper - Karras et al. (2020). Then, a model was trained on a landscapes dataset with 1349 images created using Flickr and Stable Diffusion. The results demonstrate that StyleGAN2-ADA can produce high quality and visually realistic images of natural landscapes and has potential for various applications such as virtual reality, gaming and content creation. The resultant model also performed well on GAN specific evaluation metrics - Fréchet Inception Distance (FID) and Kernal Inception Distance (KID). FID was observed to be consistent with human perceptual evaluation of generated synthetic images. Finally, the work proves that StyleGAN2-ADA consistently produces good quality models without significant hyper-parameter search or frequent retraining.

Contents

Figures

Tables

Abbreviations

ADA	Adaptive Discriminator Augmentation
AFHQ	Animal-Faces-HQ
API	Application Programming Interface
FFHQ	Flickr-Faces-HQ
FID	Fréchet Inception Distance
GANs	Generative Adversarial Networks
GPUs	Graphics processing units
KID	Kernal Inception Distance

1 Introduction

Generative Adversarial Networks (GANs) are a popular class of deep learning methods which are focused towards studying a collection of training examples and learning the probability distribution that generated them. GANs are then able to generate more images from that probability distribution. (Goodfellow et al., 2020) Although GANs have shown good performance in various application areas such as medical imaging and satellite imagery, there are some limitations - high computational requirements, large amount of training data, and training complexity. (Nesvold & Mukerji, 2021; Woodland et al., 2022)

When training with a low amount of data using GANs, the discriminator tends to over fit, which causes the training to diverge. StyleGAN2-ADA is model developed by Nvidia which aims to stabilize training in limited data regimes. (Karras et al., 2020) The authors of Karras et al. (2020) developed an extension to the original StyleGAN-2 which does not change the loss function or the network architectures.

Despite its success, there is limited research on its ability to be used on general purpose applications, where small datasets can be prepared quickly and models can be trained on them using low cost GPUs and in a shorter time frame. The paper aims to fill this gap by evaluating the ability of StyleGAN2-ADA to generate high quality images of natural landscapes(day and night time) and assessing its visual appeal and realism.

Main contribution of the research are as follows

- Validation and reproduction of StyleGAN2-ADA model results
- Creation of a limited dataset from scratch using a mix of Flickr images and stable diffusion
- Training of STYLEGAN-ADA model training on the limited dataset, reaching a state of the art FID
- Evaluation of the synthetic images generated from the model
- Providing evidence that FID is consistent with the human perceptual evaluation of images
- Latent space exploration of images that were created from the model

2 Methods

In this project, the aim was to test the performance of StyleGAN2-ADA model on a limited dataset consisting of landscape images, including mountains, beaches, and forests, in both day and night settings. FID score and KID score were used as the quantitative metric for evaluation. In this article, we will dive deeper into the image sources used, the generative modeling process, and the evaluation metrics used to assess the quality of the generated images.

2.1 Custom Dataset using Flickr and Stable Diffusion

The aim of the project required a limited dataset but with landscape images(mountains, beaches, forests, etc.) of both day and night. At first, flickr was used to fetch this dataset but there was an observation that, only day images contained the required features. Night images mostly consisted of images containing stars(without mountains, forests or beaches). So, stable diffusion was used to create the remainder of the dataset.

2.1.1 Flickr API

To create a customized dataset, Flickr's API was utilized. The API allows the input of a search term and the desired image resolution. (Heaton, 2020) It then retrieves all relevant results and downloads the images to a specified file location. In this project, the following parameters were selected to obtain the required dataset:

Parameter Name	Input
search	landscapes
min_width	256
min_height	256
crop_square	True
max_download	3000

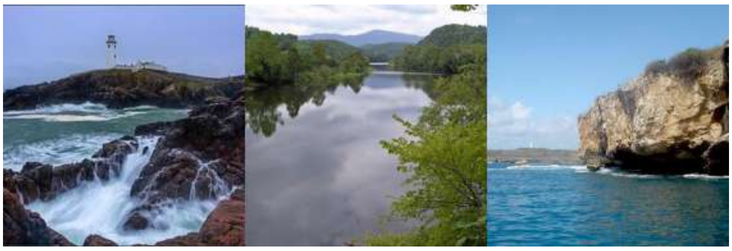


Figure 1 Flickr API Paramaters and Output Samples

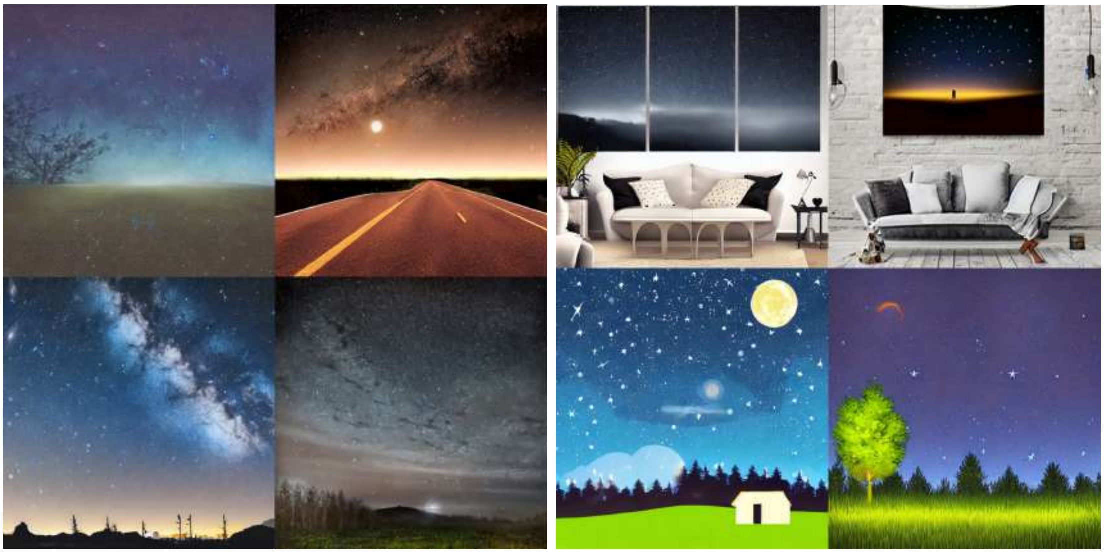
After performing the search using the specified search term, some images obtained were found to be vague and were removed manually. Furthermore, there were only a few images that contained night landscapes, and for these images, stable diffusion was utilized.

2.1.2 Stable Diffusion

"By decomposing the image formation process into a sequential application of denoising autoencoders, diffusion models (DMs) achieve state-of-the-art synthesis results on image data and beyond." (Rombach et al., 2022) In this particular case, stable diffusion was utilized to create night landscape images, using the following configuration:

Parameter Name	Input
Library	Huggingface
model_id	CompVis/stable-diffusion-v1-4
prompt	realistic night landscape with stars and moon

Table 1 Configuration for Stable Diffusion



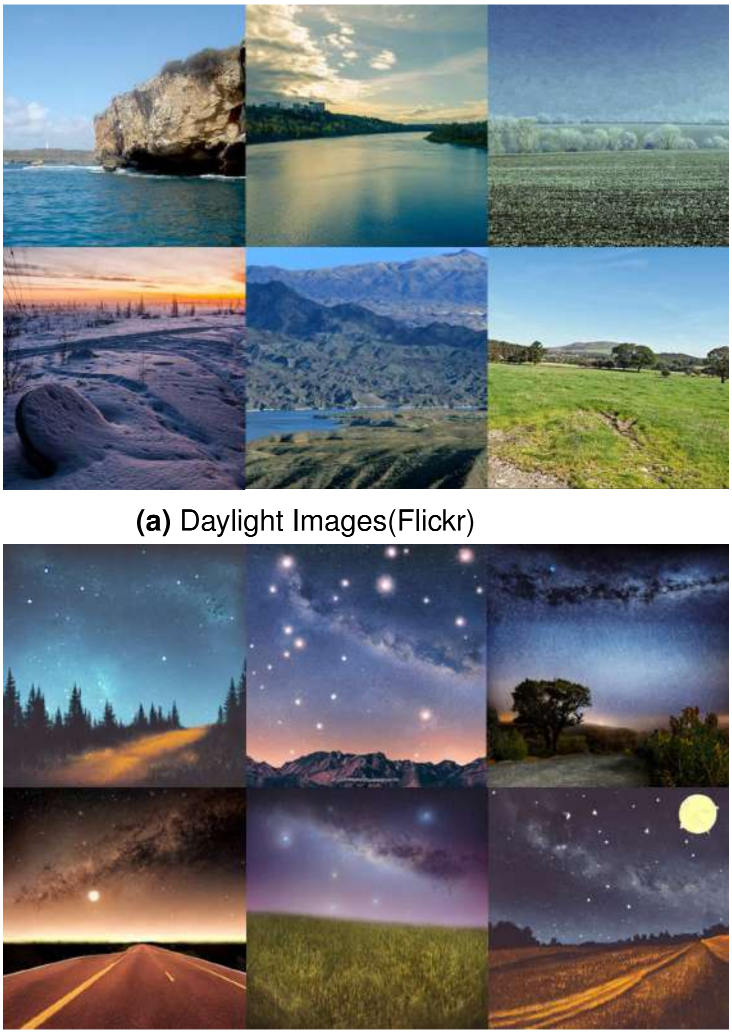
(a) Relevant Images(Included) **(b)** Irrelevant Images (Rejected)

Figure 2 Output Samples - Stable Diffusion

It should be noted that some of the images generated using the selected prompt were irrelevant and they were manually excluded from the dataset, as highlighted in Figure 2. In Figure 10 of Appendix A, I carried out an analysis to determine the effect of removing specific terms from the query prompt on the quality of the generated output samples.

2.1.3 Landscapes Dataset

The created landscapes dataset consists of 1349 images with a resolution of 256x256 pixels. Among these images, 808 depict daylight landscapes and 541 depict night-time landscapes.



(a) Daylight Images(Flickr)

(b) Night-time Images(Stable Diffusion)

Figure 3 Landscapes Dataset

2.2 Datasets for StyleGAN2-ADA results validation

For the validation and reproduction of the project, the following datasets were used –

- FFHQ Dataset(Flickr-Faces-HQ): The dataset has images of faces with variations in terms of age, ethnicity and image background. It also has a good coverage in terms of accessories such that eyewear, hats, cloths etc. These images have been collected from Flickr. (NVlabs, 2022)

- MetFaces: This dataset contains art from over 5000 years from around the world – present in two iconic sites in New York(The Met Fifth Avenue and The Met Cloisters). It has a total of only 1336 images, which makes it perfect to test to performance of StyleGAN2-ADA on limited datasets. (NVlabs, 2021)
- BreCaHAD: The data set contains 162 breast cancer histopathology images, namely the breast cancer histopathological annotation and diagnosis dataset (BreCaHAD which allows researchers to optimize and evaluate the usefulness of their proposed methods. The dataset includes various malignant cases. These 162 images are cropped and pre-processed before starting the model training. Each of these images are cropped to form 12 overlapping images of the resolution 512x512. The final dataset contains 1944 images. (Aksac et al., 2019)
- AFHQ: Animal Faces-HQ is a dataset of animal faces consisting of 15,000 high-quality images at 512 × 512 resolution. The dataset includes three domains of cat, dog, and wildlife, each providing 5000 images.(Choi et al., 2020)

2.3 Generative Modeling - StyleGAN2-ADA

StyleGAN2-ADA (Adaptive Discriminator Augmentation) is an improved version of StyleGAN2, a generative adversarial network (GAN) architecture developed by NVIDIA for generating high-quality images.

StyleGAN2-ADA improves upon the original StyleGAN2 by introducing a new training technique called Adaptive Discriminator Augmentation. This technique involves dynamically adjusting the augmentations applied to the real images during the discriminator training process. By doing so, the discriminator becomes more robust to different image transformations, resulting in more stable and higher quality image generation.

In addition, StyleGAN2-ADA also introduces a number of other improvements, such as a more efficient architecture, better regularization, and improved synthesis quality. StyleGAN2-ADA is considered to be one of the state-of-the-art methods for generating high-quality images using GANs, and has been used in a variety of applications such as image synthesis, image editing, and video synthesis. (Karras et al., 2020)

The hyperparameters outlined in 2, as described in Karras et al. (2020), were employed to train the model on the landscapes dataset.

Hyperparameter	**Value**
Optimizer	Adam
Learning Rate	0.0025
Resolution	256
Snapshot Ticks	30
Batch Size	32

Table 2 Hyperparameters for StyleGAN2-ADA training

2.4 Evaluation Metrics

The following GAN specific metrics were considered to evaluate the trained model -

2.4.1 Fréchet Inception Distance (FID)

The FID is the standard for state of the art GAN evaluation in natural imaging. (Borji, 2019) FID compares the probability distribution of synthetic images with the distribution of a set of real images, which are the ground truth. The authors of FID promote its ability to distinguish synthetic images from real images, agreement with human perceptual evaluations, sensitivity to blurriness, and its computational efficiency.

$$FID(r,g) = \|\mu_r - \mu_g\|_2^2 + \mathrm{Tr}\left(\Sigma_r + \Sigma_g - 2\left(\Sigma_r\Sigma_g\right)^{\frac{1}{2}}\right),$$

where (μ_r, Σ_r) and (μ_g, Σ_g) are the mean and covariance of the real data and model distributions, respectively. “Lower FID means smaller distances between synthetic and real data distributions.” (Heusel et al., 2018)

2.4.2 Kernal Inception Distance (KID)

Introduced first in Bińkowski et al. (2021), KID is another metric used to evaluate the quality of generated images in the context of GANs. KID measures the dissimilarity between two probability distributions P_r and P_g using samples drawn independently from each distribution.

$$KID = MMD\left(f_{\text{real}}, f_{\text{fake}}\right)^2,$$

where MMD is the maximum mean discrepancy and f_{real}, f_{fake} are extracted features from real and fake images. (Kynkäänniemi et al., 2019)

3 Results

First, I try to validate the reproduce the results present in the orignal Stylegan2-ADA paper. Following experiments were performed-

- Verifying FID score of pre-trained model
- Generating random seeds using Nvidia's pre-trained models

Then, I will present the findings of my study on the performance of StyleGAN2 ADA model on landscape images. The objective was to evaluate the model's ability to generate high-quality images of natural landscapes(day and night time) and to determine whether it could produce images that are visually appealing and realistic.

3.1 Reproduction and validation of StyleGAN2-ADA model results

StyleGAN2-ADA contains an adaptive discriminator augmentation mechanism that significantly stabilizes training in limited data regimes. The authors promise that good results are now possible with only few thousand images. The authors trained several models on a wide variety of small data sets.

3.1.1 Validating FID Score of Pre-Trained model

StyleGAN2-ADA's pre-trained model on the AFHQ Cat dataset is designed to achieve an FID score of 3.55. Upon further training of the pre-trained model on the same dataset, I observed an FID of 3.11 after 40 training iterations, which is in close proximity to the promised FID. The generated images accurately represent the class, demonstrating the model's ability to produce high-quality and realistic images.

Evaluation Metric	**StyleGAN2-ADA Original Paper Result**	**Calculated Result**
Fréchet Inception Distance (FID)	3.55	3.11
Kernal Inception Distance (KID)	0.00066	0.00066

Table 3 AFHQ Cat Dataset Pre-Trained Model - Metrics

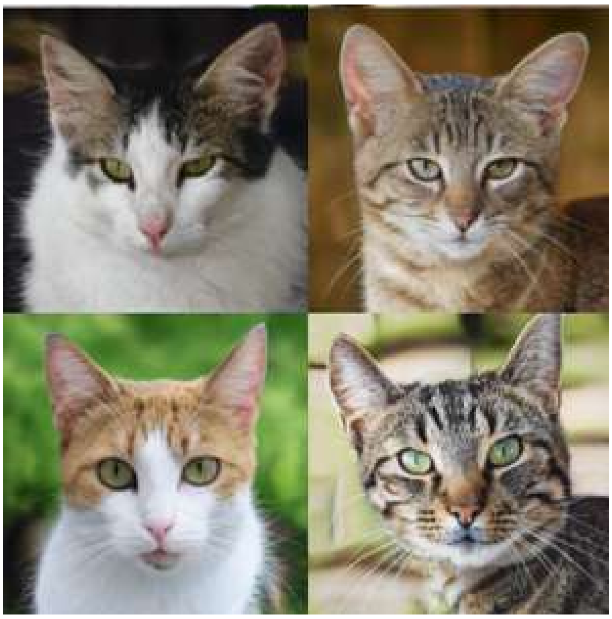
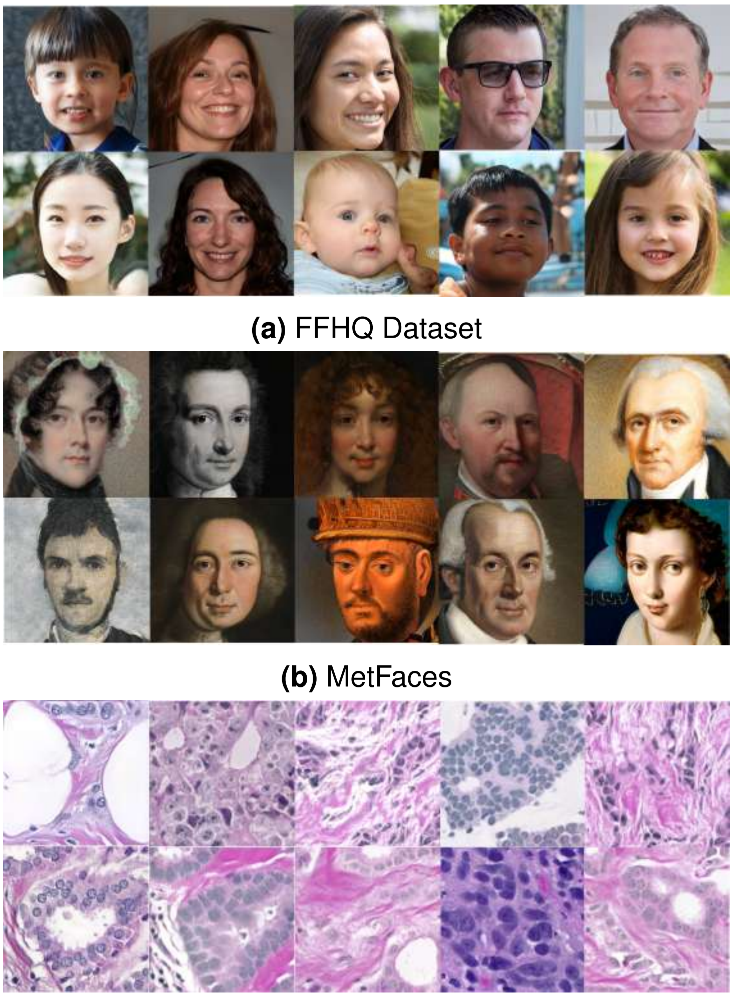


Figure 4 AFHQ Cat Dataset Pre-Trained Model - Random Seeds

3.1.2 Synthetic Image Generation from StylegGAN2-ADA pretrained models

After loading a pre-trained StyleGAN2-ADA model trained on 256x256 resolution, the generator can be used to create a 3x256x256 image from a 1x512 latent vector. This vector can be manipulated according to specific requirements to produce desired changes in the final image output. To generate random samples of images, one can use numpy's random seed generation function to create vectors of the desired dimensions, and then utilize the generator to convert them into image outputs.

The images generated by the pre trained models are displayed in Figure 4 and Figure 5. They demonstrate a strong representation of their respective classes, despite being generated randomly.



(a) FFHQ Dataset

(b) MetFaces

(c) BrCaHad

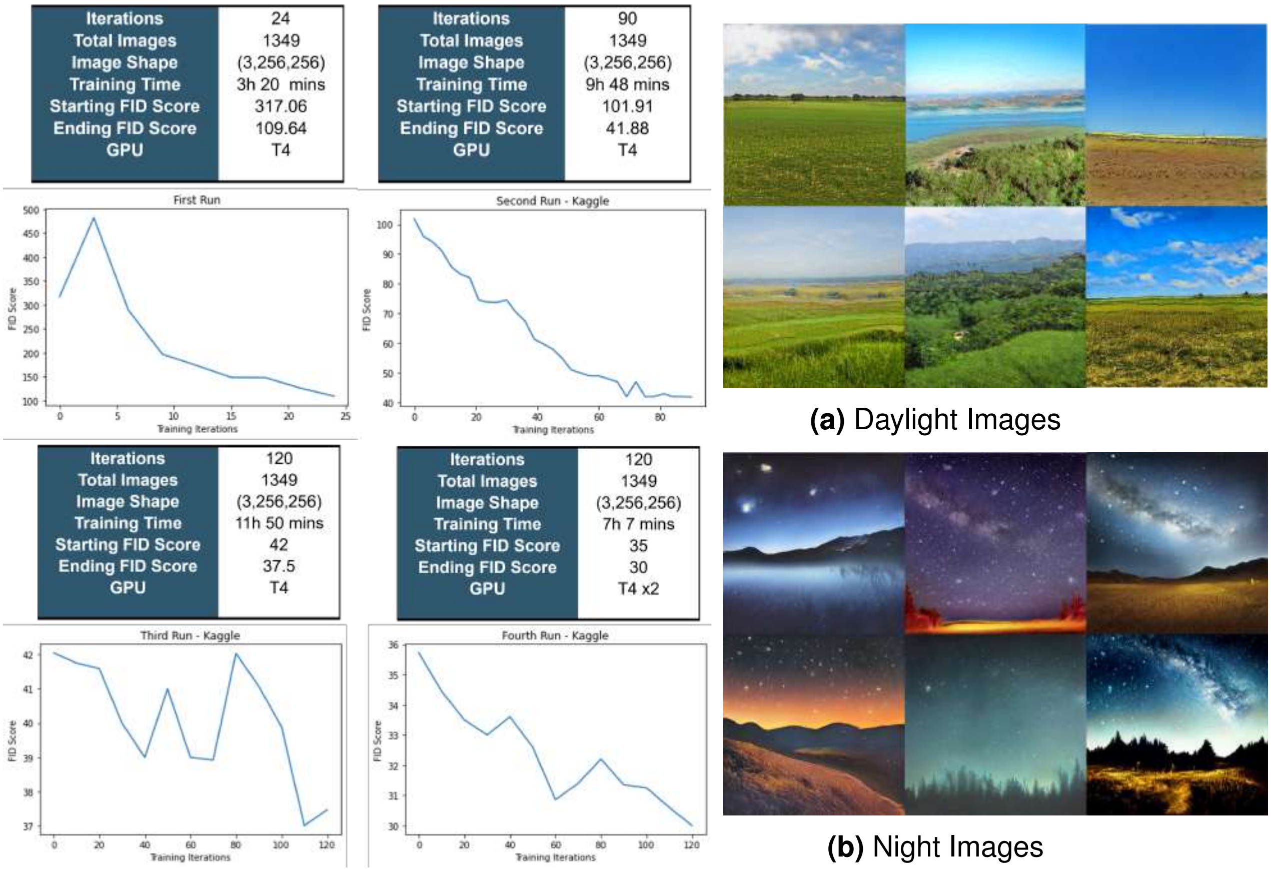
Figure 5 Using Nvidia's Pre Trained Models to Generate Random Seeds

3.2 Training model from start on landscapes dataset

After conducting four training runs on the custom dataset comprising landscape images, I achieved an FID score of 30. According to Karras et al. (2020), this FID score indicates that the model is capable of producing high-quality images. Specifically, the paper suggests that models that converge to an FID of below 40 on limited datasets would generate good quality images.

Evaluation Metric	**Score**
Fréchet Inception Distance (FID)	30.0418
Kernal Inception Distance (KID)	0.00750

Table 4 Metrics: Landscape Dataset



Iterations	24
Total Images	1349
Image Shape	(3,256,256)
Training Time	3h 20 mins
Starting FID Score	317.06
Ending FID Score	109.64
GPU	T4

Iterations	90
Total Images	1349
Image Shape	(3,256,256)
Training Time	9h 48 mins
Starting FID Score	101.91
Ending FID Score	41.88
GPU	T4

Iterations	120
Total Images	1349
Image Shape	(3,256,256)
Training Time	11h 50 mins
Starting FID Score	42
Ending FID Score	37.5
GPU	T4

Iterations	120
Total Images	1349
Image Shape	(3,256,256)
Training Time	7h 7 mins
Starting FID Score	35
Ending FID Score	30
GPU	T4 x2

(a) Daylight Images

(b) Night Images

Figure 6 Training FID v/s Iterations (left) and Output Samples(Right)

3.2.1 Discussion on Training

The training of the model utilized the T4 GPU from Google Colab and T4 x2 GPU from Kaggle, which lasted for around 32 hours and successfully attained a state-of-the-art FID score. The final model exhibited a well-balanced distribution of images portraying both daytime and nighttime scenes, as evidenced by the FID and KID scores in Table 4. It is worth noting that all resources utilized for training were readily accessible and free to use. The use of traditional GANs without ADA in this environment would have been difficult due to restrictions on GPU availability and storage.

3.3 Latent Space Exploration: Interpolating between two images

By calculating the difference between each channel of the (1x512) sized latent vector of two images generated by a StyleGAN2-ADA model and linearly interpolating them, we can create transition images that display a gradual change between the original images. An instance of this technique is demonstrated in Figure 7, where a daytime image is transformed into a nighttime image. The fifth image in the sequence depicts the final nighttime version of the original image.

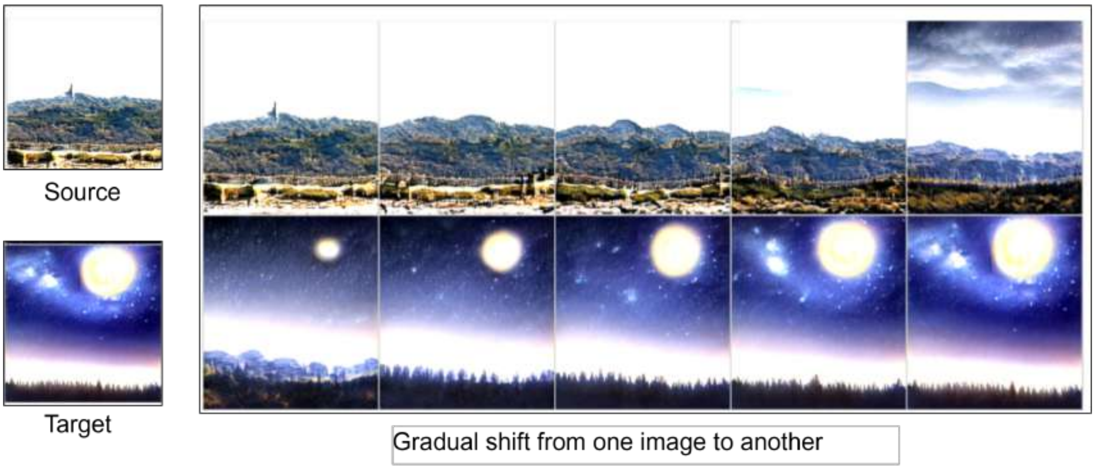


Figure 7 Interpolation between two synthetic images

3.3.1 Convert real daylight image to night image

It is possible to project a real image to StyleGAN2-ADA model's latent space. This can be done by using the projector function available in the model which starts from a given seed and calculates loss against the target image. (Karras et al., 2020) A descent approach minimizing the loss by modifying the latent vector with 1000 iterations can provide a result vector which represents the real image. The real image's latent vector can then be modified to change few characteristics of the image.

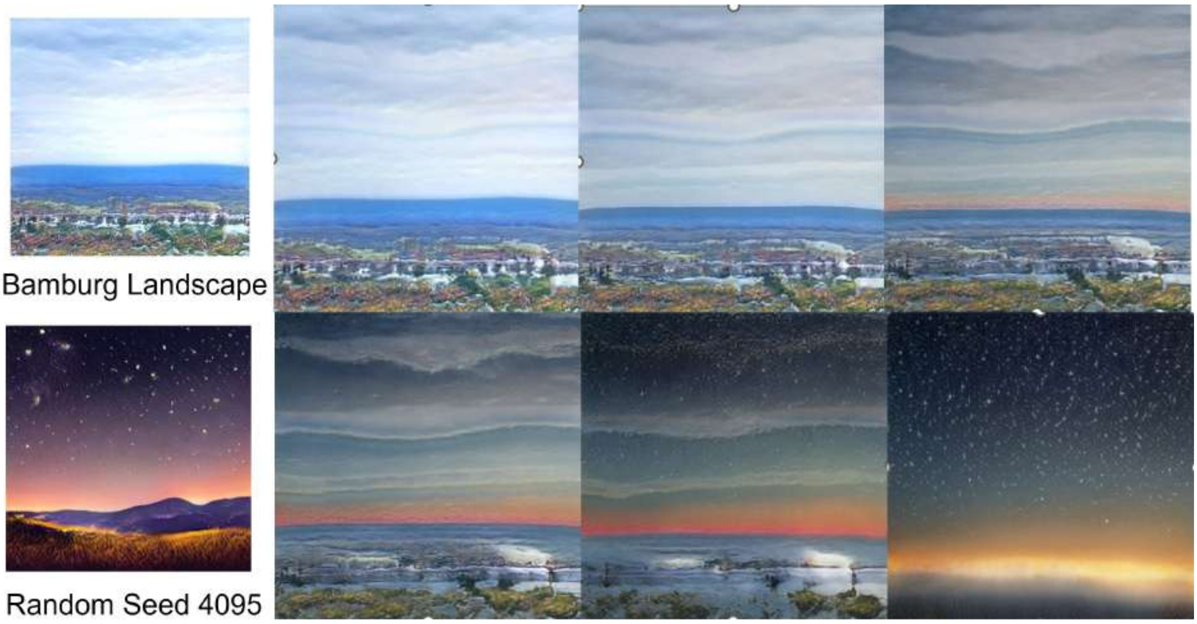
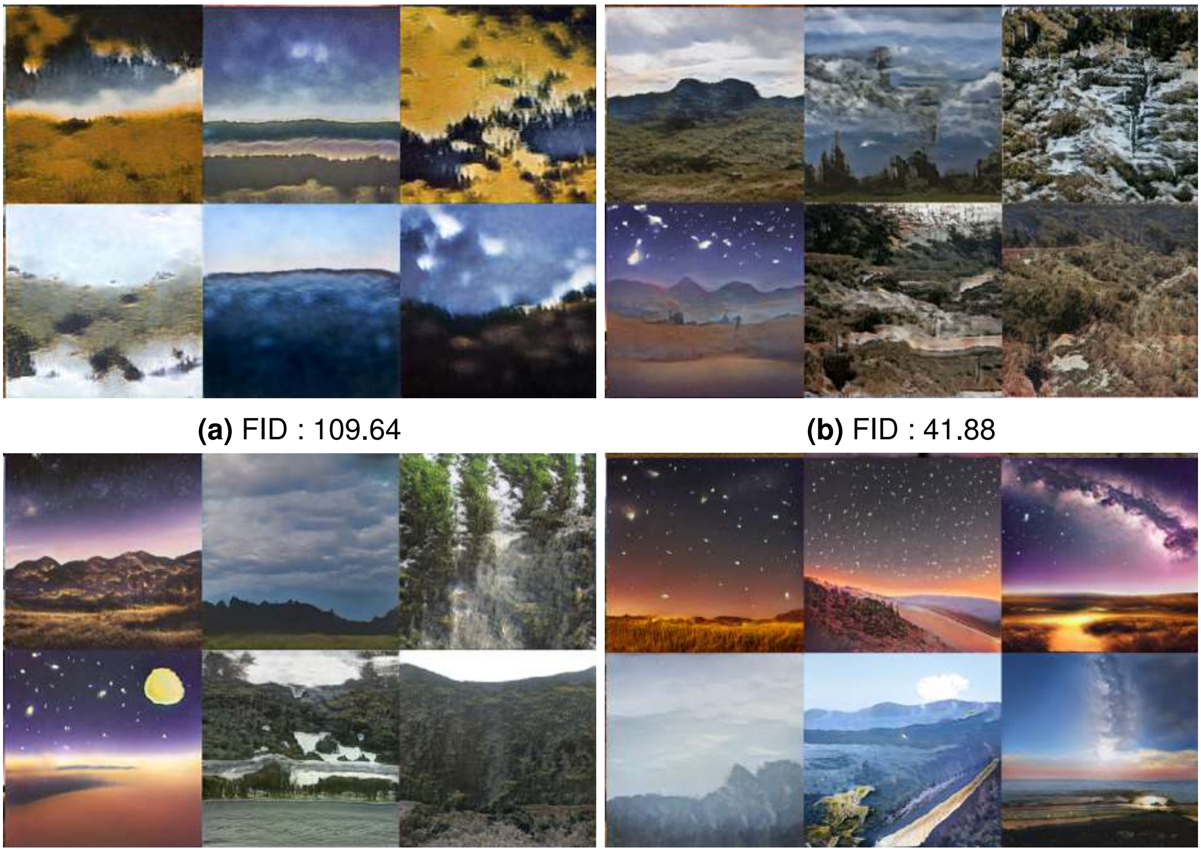


Figure 8 Transitions between Bamberg day landscape and synthetic night landscape

In Figure 8, a picture of Bamberg landscape was projected to StyleGAN2-ADA landscape model's latent space. Using a similar landscapes image but shot during the night as a target, the latent space of the Bamberg image was modified such that we can see different levels of night landscapes on it.

3.4 Consistency of FID with human perceptual evaluation of images

During the training of the StyleGAN2-ADA model, the synthetic images were periodically evaluated and the Frechet Inception Distance (FID) metric was used to measure their realism. After four training sessions, the FID metric was found to be consistently correlated with the perceptual quality of the synthetic images. In Figure 9, it is shown that as the FID decreases, the quality of the synthetic images improves.



(a) FID : 109.64

(b) FID : 41.88

(c) FID : 37.5

(d) FID : 30

Figure 9 Synthetic Images created with Model FIDs

4 Conclusion and Future Work

The StyleGAN2-Adaptive Discriminator Augmentation (ADA) model is an effective generative model for producing high-quality images. In this study, I first validated the results of the model as presented in the original paper. The FID scores obtained from pre-trained models on different datasets were in close proximity to the promised scores, and the generated images accurately represented the respective classes. The generated random samples of images using the pre-trained models and obtained realistic and visually appealing outputs.

Subsequently, I trained the StyleGAN2-ADA model from scratch on a custom dataset of landscape images. The FID score obtained from the model after four training runs was 30, indicating the model's ability to generate high-quality images on limited datasets. We also obtained visually appealing and realistic images of natural landscapes from the trained model.

Finally, I demonstrated the ability to create transition images between two generated images by linearly interpolating the difference between their respective latent vectors. Overall, the StyleGAN2-ADA model has proven to be an effective generative model for producing high-quality images, and this study contributes to the growing body of research on its capabilities. The results of the paper support the claim that StyleGAN-2 ADA can be used or general purpose applications in the fields of content creation, marketing, gaming, and virtual reality.

In the future, the landscapes dataset can be improved by adding more diverse and high-quality landscape images. This would help the model to learn more about the nuances of natural landscapes and produce better quality images. Also, performance of StyleGAN2-ADA can be compared with **Few Shot Learning**, which was implemented in Xiao et al. (2022) and designed to work on limited datasets.

References

Aksac, A., Demetrick, D. J., Özyer, T., & Alhajj, R. (2019). BreCaHAD: A Dataset for Breast Cancer Histopathological Annotation and Diagnosis. https://doi.org/10.6084/m9.figshare.7379186.v3

Bińkowski, M., Sutherland, D. J., Arbel, M., & Gretton, A. (2021). Demystifying mmd gans.

Borji, A. (2019). Pros and cons of gan evaluation measures. *Computer Vision and Image Understanding*, *179*, 41–65. https://doi.org/https://doi.org/10.1016/j.cviu.2018.10.009

Choi, Y., Uh, Y., Yoo, J., & Ha, J.-W. (2020). Stargan v2: Diverse image synthesis for multiple domains.

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

Heaton, J. (2020). Jeffheaton/pyimgdata. https://github.com/jeffheaton/pyimgdata

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2018). Gans trained by a two time-scale update rule converge to a local nash equilibrium.

Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., & Aila, T. (2020). Training generative adversarial networks with limited data. *Proc. NeurIPS*.

Kynkäänniemi, T., Karras, T., Laine, S., Lehtinen, J., & Aila, T. (2019). Improved precision and recall metric for assessing generative models.

Nesvold, E., & Mukerji, T. (2021). Simulation of fluvial patterns with gans trained on a data set of satellite imagery [e2019WR025787 2019WR025787]. *Water Resources Research*, *57*(5), e2019WR025787. https://doi.org/https://doi.org/10.1029/2019WR025787

NVlabs. (2021). Nvlabs/metfaces-dataset. https://github.com/NVlabs/metfaces-dataset

NVlabs. (2022). Nvlabs/ffhq-dataset: Flickr-faces-hq dataset (ffhq). https://github.com/NVlabs/ffhq-dataset

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models.

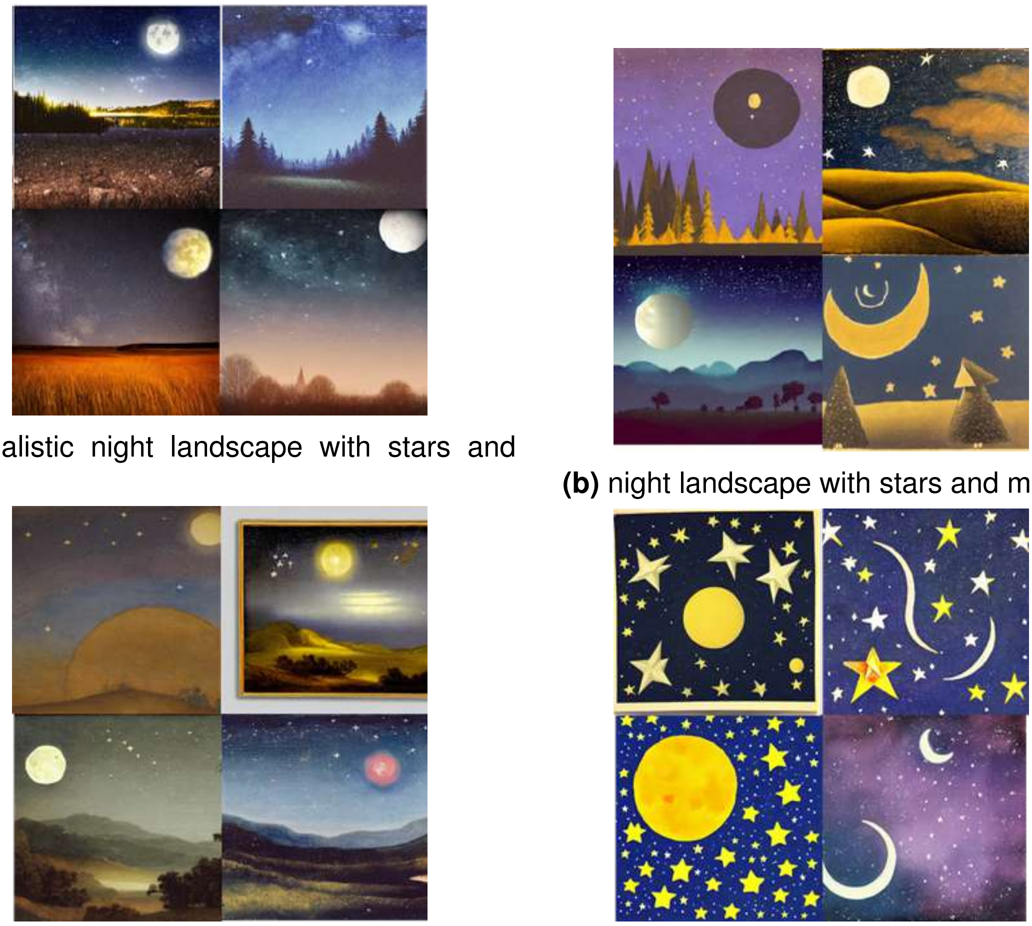
Woodland, M., Wood, J., Anderson, B. M., Kundu, S., Lin, E., Koay, E., Odisio, B., Chung, C., Kang, H. C., Venkatesan, A. M., Yedururi, S., De, B., Lin, Y.-M., Patel, A. B., & Brock, K. K. (2022). Evaluating the performance of stylegan2-ada on medical images. In C. Zhao, D. Svoboda, J. M. Wolterink, & M. Escobar (Eds.), *Simulation and synthesis in medical imaging* (pp. 142–153). Springer International Publishing.

Xiao, J., Li, L., Wang, C., Zha, Z.-J., & Huang, Q. (2022). Few shot generative model adaption via relaxed spatial structural alignment.

Appendices

A Stable Difusion: Ablation Study on Search Phrase

For creating the custom landscape dataset, the stable diffusion model query was employed using the phrase "realistic night landscape with stars and moon". Various queries were attempted, and the impact of removing certain terms on the generated output samples is illustrated in Figure 10.



(a) realistic night landscape with stars and moon

(b) night landscape with stars and moon

(c) landscape with stars and moon

(d) stars and moon

Figure 10 Stable Diffusion: Output samples as per specified prompt

A.1 Code for generating images

```
%pip install --quiet --upgrade diffusers transformers scipy mediapy
    accelerate
!huggingface-cli login
from diffusers import PNDMScheduler, DDIMScheduler, LMSDiscreteScheduler

scheduler = PNDMScheduler(beta_start=0.00085, beta_end=0.012,
    beta_schedule="scaled_linear", skip_prk_steps=True)
import mediapy as media
import torch
from torch import autocast
from diffusers import StableDiffusionPipeline

model_id = "CompVis/stable-diffusion-v1-4"
device = "cuda"
remove_safety = False


pipe = StableDiffusionPipeline.from_pretrained(model_id, scheduler=
    scheduler, torch_dtype=torch.float16, revision="fp16", use_auth_token
    =True)
if remove_safety:
  pipe.safety_checker = lambda images, clip_input: (images, False)
pipe = pipe.to(device)
prompt = "realistic night landscape with stars and moon"
num_images = 2
prompts = [ prompt ] * num_images

for i in range (1,12):
  with autocast("cuda"):
      images = pipe(prompts, guidance_scale=7.5, num_inference_steps=50)
          .images
  media.show_images(images)
```

Listing 1 Python Code for generating images using stable diffusion

The code specific in Listing 1 can be used to generate images using stable diffusion. (Rombach et al., 2022)

Declaration of Academic Integrity

I hereby declare that this term paper and the work presented in it is entirely my own. Where I have consulted the work of others, this is always clearly attributed. Where I have quoted from the work of others, the source is always given. I am aware that the thesis in digital form can be examined for the use of unauthorised aid and in order to determine whether the thesis as a whole or in parts may amount to plagiarism. I am aware that a false assurance fulfils the elements of fraud in accord with § 10 and § 13 ABMPO/TechFak and will result in the consequences proclaimed there. This paper was not previously presented to another examination board and has not been published.

Erlangen, 19th March 2023

Krishnachander Govindarajan