Landscape Image Generation using StyleGAN2 ADA

Jyot Makadiya
Institute of Technology
Nirma University
Ahmedabad, India
18bce088@nirmauni.ac.in

Kaushil Mangaroliya
Institute of Technology
Nirma University
Ahmedabad, India
18bce090@nirmauni.ac.in

Priyank Thakkar
Institute of Technology
Nirma University
Ahmedabad, India
priyank.thakkar@nirmauni.ac.in

***Abstract*—In this paper we propose a modified architecture for generating HD quality images based on StyleGAN2 ADA architecture. This paper presents a new architecture which is evaluated on Landscape image dataset. The used landscape dataset is one of the first to be benchmarked on StyleGAN-based architectures. This dataset has potential to become standard test for variety of generative tasks due to wide domain of available images. Our method proposed a generative network with ResNet backbone compared with skip network present in previous StyleGAN architectures.**

***Index Terms*—StyleGAN2 ADA, GAN, Generative algorithms, Landscape images, StyleGAN, High resolution image generation**

I. Introduction

There has been a shift in research work and quality of it since the past decade. This is happening due to availability of large amounts of data and availability of powerful machines to process that data. Combination of huge amounts of data and huge amounts of processing power leads to better research in terms of quality. This is true in case of generative algorithms as well. Generative algorithms have improved a lot since the last many years in terms of the quality of output images. The use of Generative Adversarial Networks (GAN) gave researchers a new path to opt for [1]. Despite of these new advancements, researchers struggle when it comes to understanding the use of features like stochastic features. Understanding about latent space and its demonstration was also not proper [2]–[4].

The research on StyleGAN [5] was something new in its own field because it was based on a style transfer mechanism. The generator changes the known input and adapts to the needed output. Use of noise helps in proper adaptation and getting different output which is necessary in this type of generator model [5]–[8]. When analysed by use of matrices, it was seen that the present generator is way better than the previous approach due to multiple factors. StyleGAN2 was introduced as an improved version of the normal StyleGAN architecture. They focused on fixing small problems and small drawbacks that the initial model they proposed had. Use of Affine transformers was done in this case along with use of Adaptive Instance Normalization (AdaIN) [9]–[11]. In this model relatively low entanglement is seen [12] as a result.

The work in this paper builds on top of recently published paper StyleGAN2 ADA or Adaptive Discriminator Augmentation, which is in turn an improved version of its predecessors StyleGAN and StyleGAN2. The StyleGAN2 ADA focuses on training the GAN with very little amount of data without facing discriminator overfitting problem [13]. Training of discriminator on small datasets with augmentation creates leaking of augmentation in generator and in case of no augmentation the training starts to diverge as a result of overfitting of discriminator on images. The styleGAN2 ADA proposes conditions where no leakage is present while ensuring augmentation during training. This unique setup overcomes the data availability issue and provides StyleGAN2 competing results with substantially small dataset. We present a modified setting of StyleGAN2 ADA which produces similar results and present the results our results on Landscape Image dataset. Thus, this paper proposes two main contributions:

- We propose a modified architecture for StyleGAN2 ADA generator block for suitable training.
- We evaluate our new architecture with baseline on a Landscape image dataset which has previously been not evaluated on any StyleGAN variants. We believe this dataset has the potential to be a benchmark for landscape images as there doesn't exist any standard benchmark despite a few results of evaluation of GAN [2].

II. Previous Work

Derived from parallel work and from results in the case of StyleGAN, it was seen clearly that the traditional GAN generator architectures are restricted in the case of a style-based part [14]. The case is true when compared with conventional evaluation metrics as well, the analysis of other stochastic effects and higher-level features, also using the lenity in case of latent space w proves useful in making advancements related to the capabilities of GAN. During training, the mean path length metric can also be used in place of regularizer without any problem, and maybe an alternative of the well-known separability metric might behave as that.

From previously known experiments, use of whichever augmentation applicable during the training will result in their impact on the images produced [1]. Zhao et al. presented a balanced consistency regularization, an alternative approach which is not expected to expose any set of augmentations to the final images generated during the training [15]. Consistency regularization presents various different series of transformations, performed on the similar images, should result in the same image [16], [17]. Zhao et al. modified regularization terms in case of discriminator loss term, and validated discriminator consistency, whereas no augmentations is performed during the generator training in another case [17]. This proposed solution essentially aims to make the discriminator capable enough by avoiding the set of augmentations. Although, achieving this goal creates an opportunity for exposing augmentations, due to the generator being open to producing images containing augmented images without penalty. The StyleGAN2 ADA paper shows that bCR could also suffers from the described hurdle, and consequently, the effects observed are essentially same as the dataset augmentation [13].

We focused our experiments by first artificially sub-setting a larger dataset (in our case Landscape image dataset) and observing the results of modified generator architecture. The strength of StyleGAN that is the ability to handle the images generated through the means of style mixing, in other words by inputting a distinct latent vector to distinct neural layers at inference time to obtain effective results dynamics. For the StyleGAN2 ADA baseline, they have considered StyleGAN2 and BigGAN model [18]–[20]. However, we have used StyleGAN2 as baseline and further propose a modified generator architecture in the paper for comparison.

III. Methodology

With the advancement of model architecture in the research domain, state-of-the-art methods are also gradually increasing the quality and innovation expected from new approaches. We prepare the dataset of Landscape first along with suitable preprocessing applied for corresponding model, here StyleGAN2 ADA Model. To start experiments, the larger dataset was first divided into subsets (selecting a subset, Mountain images from larger Landscape image dataset) and observing the resulting dynamics. Based on baseline results, we decided to use the StyleGAN2 ADA. This was because of generated outputs with a very little amount of data while having lower variance due to the adaptive discriminator augmentation mechanism used in the model. For each individual experiment, we modified the architecture and selected the dataset based on the suitability of GPU computation. We have used a preprocessed 256 ×256 version of the Landscape mountains dataset and a smaller-weight configuration compared to the official StyleGAN2 for this dataset. Another set of experiments uses 1024 x 1024 images which are upscaled version of original dataset. The same styleGAN2 ADA preprocessing script was used in both the cases. We have measured the quality of images produced by the generator by computing FID (Frechet Inception Distance) between every 10k generated images by the generator and all provided training images in the dataset, as recommended by Heusel et al. in the given paper, regardless of the subset (although we only used mountains dataset for testing and training) actually used for training [13].

A. Dataset Used

The dataset used was obtained from Kaggle with the title “Intel Image Classification”. It contains 6 different categories: building, forest, glacier, mountain, street, sea. The image size is 150 x 150, although we took the liberty to preprocess the image and upscale it to 256 x 256 for a convenient training pipeline. Another variant of same dataset has set 1024 x 1024 image size during upscaling. The dataset we used contains only mountain images and after scaling and cropping the final dataset is created. To apply the preprocessing, the original StyleGAN2 ADA dataset tool script was used. This allowed us to prepare a suitable dataset for the architecture with even modifications. The total number of images is 2512 for mountains although the landscape image contains around 25,000 images only these 2512 were used. For training and evaluation of FID, the mountains dataset with augmentation is used.

B. Augmentation Used

The solution used by StyleGAN2 ADA is comparable to bCR as they have also applied a series of augmentation used for the given images in the dataset presented for the discriminator. The difference is, in place of appending a different loss term for CR, here they have used to measure the discriminator for the augmented images only. To achieve this, another method for generator training was implemented. Another approach that is also known as stochastic discriminator augmentation is similar in working principles and very straightforward [13]. However, despite this approach having potential, it has received little attention from the research community, possibly because of the reason initially, the fact is not intuitive if the method can work properly: here if discriminator in GAN doesn’t see what the real training dataset images in the dataset actually look like, although the fact is not still obvious whether it can give proper direction to the generator accordingly during the training. We are not going into the details of why and how the above approach works given preconditions, that work is already being described in the StlyeGAN2 ADA paper [13]. Although, the augmentation of the generated images and the full pipeline of transformations is what we have used in our experimentation, we also used default setting present in StyleGAN2 ADA paper for our baseline model evaluation [13]. The underlying assumption that is used here and in StyleGAN2 ADA is that a maximally wide and fully customized series of augmentations are useful, provided the wide known success of RandAugment in image classification tasks [21]. The used augmentation pipeline consists of 18 transformations that are divided into 6 different categories: pixel blitting (which includes x-flips, integer translation and 90 rotations), other general geometric transformations, colour transforms for images, additive noise, image-space filtering, and the last is cutout [19], [22]. The requirement here is that augmentations need to be differentiable due to the fact that here, we perform augmentations also at the time of generator training, the same as StyleGAN2 ADA [13]. This is achieved by performing augmentations

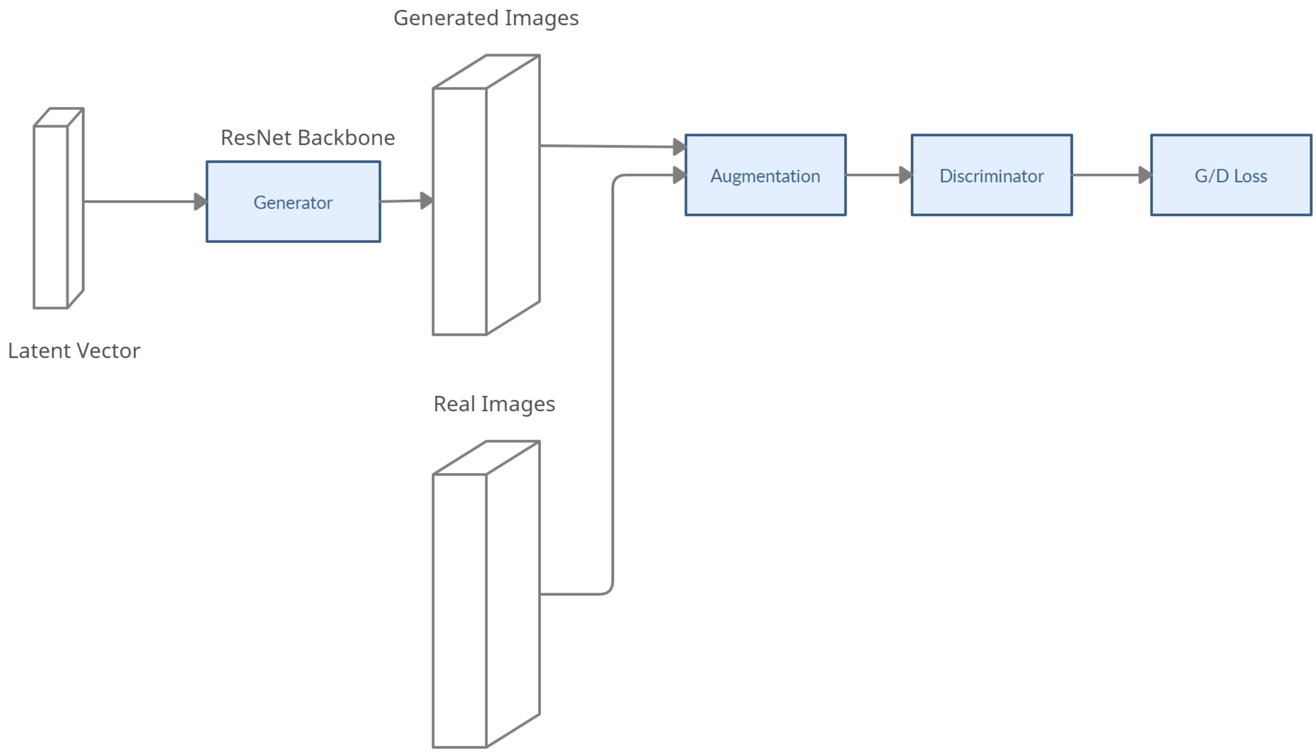

Fig. 1. StyleGAN2 ADA with modified ResNet backbone in Generator network architecture.

using the given standard differentiable primitives from the underlying DL framework. During the training, each individual image is presented to the discriminator by the method of an already predefined series of transformations in a determined manner and to achieve the randomized nature the augmentation strength is applied using a scalar p $\in$ [0,1]. This p represents the probability with which each transformation is applied during the training. The used randomization process is applied distinctively for the individual augmentation step along with an individual image in the minibatch as proposed in StyleGAN2 ADA [13]. The adaptive discriminator augmentation is used as it is from the StyleGAN2 ADA architecture. However, we used a different generator architecture that is modified for our dataset and further experiment results are also provided in the Evaluation Results section. The modified architecture includes ResNet as backbone architecture in the synthesis network block of the StyleGAN2 model. Furthermore, we also experimented with a complete augmentation transformation setting which improved the results but at the cost of training time.

IV. Evaluation Results

We have used the training from the scratch method for all of the experiments, an alternative approach could be transfer learning. Training from scratch was suitable for our dataset as well as for the ADA method. Transfer learning from StyleGAN2 holds potential for providing better results over short amounts of time given the similarity in the dataset. The used landscape dataset was different as we tested the StyleGAN2 ADA model on Landscape images (precisely mountain images) than the previous dataset that was used in StyleGAN and StyleGAN2. We focus on providing quality images with little data using a modified StyleGAN2 ADA model. Without augmentations as seen in the previous StyleGAN model, the gradients received from the discriminator by the generator became extremely simple over time — consequently, the discriminator begins to consider a few features, and the corresponding generator is thus starting to generate nonsensical unseen images. Since the use of ADA, this gradient field found during training stays much more detailed compared to the previous cases which prevent such degradation of quality. We proposed a modified version of ADA with ResNet as the backbone of the generator model. This allowed matching the accuracy and architectural similarity of discriminator as it was also using ResNet architecture previously. The addition of new architecture option with similar performance opens up new gateways for training StyleGAN based models using ADA. We also found that using a skip network as a backbone will speed up the training process and it can be observed that FID is greater when compared to ResNet architecture. Table I presents our experimental results with StyleGAN2 ADA architecture.

The landscape image dataset results are quantitatively evaluated based on Fréchet Inception distance and various archi-

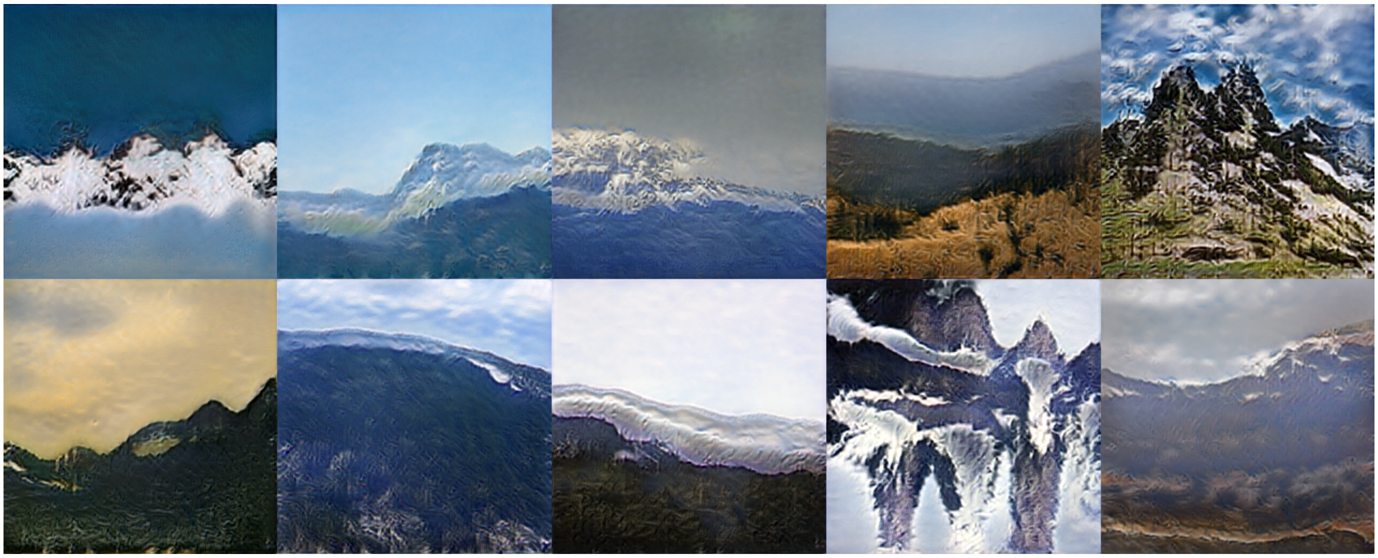

Fig. 2. Sample of Landscape Images Generated by Modified StyleGAN2 ADA Model

TABLE I
RESULTS OF BASELINE AND MODIFIED STYLEGAN2 ADA MODEL ON THE LANDSCAPE IMAGES

Model architecture of StyleGAN2 ADA	Metric: fid50k_full (Fréchet Inception distance)	Training Time (in hours)	Input dataset image size	Kimg (training duration unit)
Baseline	45.12	8	256	400
Baseline - full augmentation	72.63	7	256	250
Modified with Resnet backbone	78.83	8	256	420
Modified with Resnet backbone - full augmentation	82.10	7	256	300
Baseline	43.01	25	1024	200
Modified with Resnet backbone	34.29	28	1024	200

tectures were used. Finally, we noticed an improvement in performance using full augmentation (all 6 transformations) for the same architecture. The synthesis network architecture was made similar to that of the discriminator to evaluate the performance results.

V. CONCLUSION

We have proposed an advanced method for generating images that is capable of producing high-quality images, based on a well-known model StyleGAN2 ADA. In this work, we evaluate the modified generator architecture of StyleGAN2 ADA on Landscape dataset containing around 2500 mountain images. The landscape images generated in this process seem to have acceptable quality considering the amount of training done. Training the StlyleGAN2 ADA model with a smaller dataset gave better results because it uses an adaptive discriminator augmentation mechanism [13]. The use of Modification with the ResNet backbone has increased the training time. Although the baseline score is better than full augmentation modified architecture, we found that modified architecture produces more vivid images qualitatively. The results are summarized in the results section with a detailed comparison of their performance. The modifications were made in the synthesis network which includes ResNet as the backbone and using an augmentation pipeline for training. The advancement in StyleGAN2 ADA along with our experiments prove the advantages of using smaller datasets which are capable of yielding the same quality image.

REFERENCES

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, and B. Xu, "D. warde-f arley, s. o zair, a. c ourville, y. b engio. generative adversarial networks," *arXiv preprint arXiv:1406.2661*, 2014.

[2] M. Afifi, M. A. Brubaker, and M. S. Brown, "Histogan: Controlling colors of gan-generated and real images via color histograms," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021.

[3] V. Dumoulin, E. Perez, N. Schucher, F. Strub, H. d. Vries, A. Courville, and Y. Bengio, "Feature-wise transformations," *Distill*, vol. 3, no. 7, p. e11, 2018.

[4] S. Laine, "Feature-based metrics for exploring the latent space of generative models," 2018.

[5] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," *Advances in neural information processing systems*, vol. 30, 2017.

[6] T. Sainburg, M. Thielk, B. Theilman, B. Migliori, and T. Gentner, "Generative adversarial interpolative autoencoding: adversarial training on latent space interpolations encourage convex latent distributions," *arXiv preprint arXiv:1807.06650*, 2018.

[7] K. Kurach, M. Lucic, X. Zhai, M. Michalski, and S. Gelly, "The gan landscape: Losses, architectures, regularization, and normalization," 2018.

[8] M. Lucic, K. Kurach, M. Michalski, S. Gelly, and O. Bousquet, "Are gans created equal? a large-scale study," *Advances in neural information processing systems*, vol. 31, 2018.

[9] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 1501–1510.

[10] V. Dumoulin, J. Shlens, and M. Kudlur, "A learned representation for artistic style," *arXiv preprint arXiv:1610.07629*, 2016.

[11] Y. Li, C. Fang, J. Yang, Z. Wang, X. Lu, and M.-H. Yang, "Universal style transfer via feature transforms," *Advances in neural information processing systems*, vol. 30, 2017.

[12] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," *arXiv preprint arXiv:1710.10196*, 2017.

[13] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila, "Training generative adversarial networks with limited data," *Advances in Neural Information Processing Systems*, vol. 33, pp. 12 104–12 114, 2020.

[14] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4401–4410.

[15] Z. Zhao, S. Singh, H. Lee, Z. Zhang, A. Odena, and H. Zhang, "Improved consistency regularization for gans," *arXiv preprint arXiv:2002.04724*, 2020.

[16] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," *arXiv preprint arXiv:1610.02242*, 2016.

[17] M. Sajjadi, M. Javanmardi, and T. Tasdizen, "Regularization with stochastic transformations and perturbations for deep semi-supervised learning," *Advances in neural information processing systems*, vol. 29, 2016.

[18] A. Brock, J. Donahue, and K. Simonyan, "Large scale gan training for high fidelity natural image synthesis," *arXiv preprint arXiv:1809.11096*, 2018.

[19] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of stylegan," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 8110–8119.

[20] E. Schonfeld, B. Schiele, and A. Khoreva, "A u-net based discriminator for generative adversarial networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8207–8216.

[21] E. D. Cubuk, B. Zoph, J. Shlens, and Q. V. Le, "Randaugment: Practical automated data augmentation with a reduced search space," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 702–703.

[22] T. DeVries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," *arXiv preprint arXiv:1708.04552*, 2017.