# Distortion Measure Regularized Generative Models for Isometric Representation Learning

Jaehoon Hahm Minjae Kim Hyeryeong Seo Sungsu Hur
Seoul National University
{judol1, minj219, 202222972, sshuh1215}@snu.ac.kr

## Abstract

*Learning a smooth latent manifold is important for representation learning including generative tasks. Smoothness of the latent space implies natural transitions in the image space, and path length regularizer was used in Style-based Generative Adversarial Networks (StyleGANs) [6, 8] to guide the model to learn an isometrical mapping from latents to images. In this work, we reformulate the path length regularizer in StyleGANv2 to a coordinate-invariant form with relaxed distortion measure [4, 19] that releases the preference of scale factor that was explicitly prescribed in previous method with implicit formulation, leading to more natural and stable learning of isometrical mapping. Also, we extend this work to Diffusion models, and propose a new metric, distortion measure, for measuring how much the mapping of a generative model distorts the latent space with the connection to Perceptual Path Length (PPL). This is the first work to apply isometry representation learning regularizer for generative models within our knowledge.*

## 1. Introduction

StyleGANs [6–8, 14] and Diffusion models [3, 5, 15] are exhibiting excellent performances in various generative tasks in the field of vision, audio, natural language, forecasting, and graph. One of the main ideas of StyleGAN was injecting disentangled information via style vectors and this disentanglement was done by injecting intermediate latent code from $\mathcal{W}$ space. By using the method of style transfer, they were able to separate high-level attributes in the latent space, making the basis corresponding to different percpetual features independent. This leads to better interpolation, image manipulation, enabling coherent-looking animations such as random walking in the latent space.

The authors of StyleGAN proposed a metric to measure how well the model disentangled the latent space; Perceptual Path Length (PPL) which measures the perceptual distance between two images which are the mappings from two neighboring latent vectors. Authors interpreted as smaller the PPL, better the disentangled was done and interpreted PPL represents the smoothness of the mapping from the latent space to image space.

As deliberated in [1, 13], learning the structure and semantic relationship between instances or representation points is helpful for model to interpret given images regardless to the domain or style. In light of this, for the generative tasks, smooth or regularized latent space can have similar interpretation with image space in terms of its style or content. It means model might recognize itself semantic information and visual patterns in not only intra domain but also inter domain just as people perceive. With this in mind, for the generative tasks, the model with smooth latent space might be able to synthesize a image including unseen style or even content, which makes the model be more general for various image generation tasks. Furthermore, if we can give more various images to the model, the capability would be more refined as the model learns compact and well-interwined smooth latent space.

As prior works for smoothing out the latent space, [9] used large $l_2$ norm, so that the distribution of styles to be a shrunk Gaussian centered on the origin and compacts the space. [17] proposed the shortest path regularizer by finding a shared latent space. For metrics measuring the smoothness of the mapping, [9] proposed the Perceptual Smoothness(PS) metric, because PPL can be minimized by a collapsed generator. PS consists of the degree of linear alignment and the uniformity using the Gini inequality coefficient. [18] offered quantitative measurement to style transfer by using Base E and C statistics. E statistic evaluate the transformed image has the desired style or not and C statistics evaluate the degree of objectivity in the content images [16].

Additionally, motivated by the arising of Diffusion models, following works discuss about the latent space of diffusion models. Due to excelling generating capability of Diffusion process, there are lots of trials to smooth the latent space. [20] proposed pre-trained DPM AutoEncoder (PDAE) loses the information of latent space. Therefore by


Figure 1. Comparison between images generated by StyleGANv3, path length regularized StyleGANv3, and distortion measure regularized StyleGANv3. Distortion measure based regularizer releases the restricted condition on the scale factor of desired isometry, and leads to stabler and natural convergence of the mapping. Notice body part of animals are absent in the images from previous methods while our distortion measure regularized model fills the missing body parts, while maintaining high generation quality.

using the classifier-guided sampling method, compute an extra mean shift to fill the information gap. [12] proposed conditional DDIM encoder and decoder to learn a rich and smooth latent space. [10] proposed Hierarchical Diffusion Autoencoders (HDAE) that disentangle attributes by adjusting $n$ which is obtained from the weights of the linear classifier for the target attribute.

Smoothing and disentangling the latent space can improve the image generation quality by removing the unnecessary and unnatural artifacts in generated images. [9] The work reports that StyleGANv2's intermediate space can generate images with artifacts, and proposes triplet-loss, style regularization(SR) loss and perceptual smoothness(PS) metric. The triplet loss controls the inter-domain distances and preserves disentanglement and the SR loss makes the latent space smooth by compacting the space. As PPL can be reduced by collapsed generator, it defines a new metric that measures the smoothness of the style space. Additionally, there have been numerous works regarding the Diffusion models as a representation learner; [12] proposed conditional DDIM encoder and decoder to learn a rich and smooth latent space with excelling generating capability of Diffusion process.

However, StyleGANv1 did not perform any explicit regularization to decrease the PPL. Afterward, the following work StyleGANv2 added the Path Length Regularizer term that forces the mapping from the latent or the generator of GAN to be isometry up to scale. They had set the constant as a dynamically optimized constant with the exponential moving average of the lengths , which is a heuristic way to find the suitable scale factor. Hence, this method does not explicitly shows how the mapping is far from being a scaled isometry and also has potential to not gurantee an optimal solution. In the original work from StyleGANv2, they report that this regularizer did not quite reached orthogonality, and also their path length regularizer method decreased the FID in FFHQ dataset, but increased the FID in LSUN Car dataset.

We show that this result is due to the coordinate-variant and restricted form of the path length regularizer term. In other words, the formulation of original path length regularizer changes under coordinate transformation of the latent space, while also regularizer restricts the mapping to become an isometry realated to a prescribed scale that is determined in the early stage of training.

Meanwhile, [19] newly proposes a coordinate-invariant functional that measurses how the given mapping is far from being a scaled isometry in a relaxed distortion measure sense. In their work, they added the above distortion measure term to the original autoencoder term, that lead the autoencoder to learn the geomtery-preserving latent space that induces short PPL mapped from the interpolation in the latent space. They explains three types of geometry constraining regularizers: isometry loss, scaled-isometry loss, and area perserving harmonic loss which does not include any dynamically optimized variables.

In this work, we show the coordinate-invariant functional forms of the path length regularizer indeed give sta-

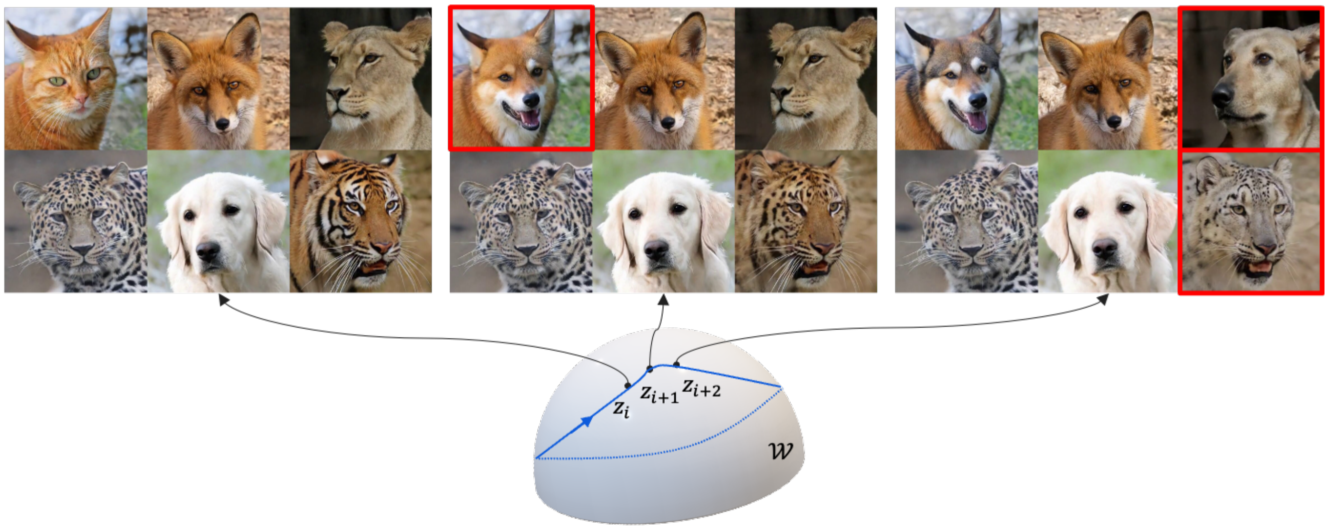


Figure 2. Mappings from three neighboring latents $z_i$, $z_{i+1}$, $z_{i+2}$ from the latent manifold $\mathcal{M}$ trained with path length regularizer. These are the latent interpolations of two endpoints. Since the mapping of the generative model, or the generator of the GAN in the figure, is not ideally a scaled-isometry, the magnitude or angle between displacement vectors in the latent manifold is not preserved after the mapping to the images. This implies the gradient of the mapping is inhomogeneous, and points having steep gradient exists, which leads to drastic perceptual change during latent interpolation. For example, during $z_i \rightarrow z_{i+1}$, the upper left cat drastically changes to a fox, while at $z_{i+1} \rightarrow z_{i+2}$, the right column's tigers drastically change to dog or its pattern.

bler training process that leads to decrease of PPL as well as FID that was unsuccessful in original StyleGANs. This will be the first work to apply coordinate invariant distortion measure to generative model and guide it to learn an isometric representation learning. Also we extend this formulation to iterative generative models such as Diffusion models to guide the score model to learn an isometrical mapping. Finally, we propose a new metric to quantify the smoothness of the latent mapping, and show its advantageous consistency over PPL.

## 2. Perceptual Path Length Regularizer

Isometry is a mapping that preserves distance and angle. Scaled-isometry is a relaxed version of isometry that it may preserve distance and angle up to a global constant scale. We may interpret scaled-isometry as a homogeneous stretching of latent manifold. In generative tasks, we may use the trained mapping from latent to image as a decoder, and using the learned latent space, image interpolation, temperature manipulation, image reconstruction becomes possible. We hypothesize that learning a smooth mapping is important for generative models to learn a well-conditioned latent space in a sense that such space is advantageous for stable training and sampling.

Figure 2 shows the result of latent interpolation via mapping learned from the original perceptual length regularizer in StyleGANv2. Since the mapping of the generative model, or the generator of the GAN in the figure, is not ideally a scaled-isometry, the magnitude or angle between displacement vectors in the latent manifold is not preserved after the mapping to the images. This implies the gradient of the mapping is inhomogeneous, and points having steep gradient exists, which leads to drastic perceptual change during latent interpolation. For example, during $z_i \rightarrow z_{i+1}$, the upper-right cat drastically changes to a fox, while at $z_{i+1} \rightarrow z_{i+2}$, the right column's tigers drastically change to dog or its color. This shows that original path length regularizer may decrease PPL because it guides the mapping to scaled-isometry but not in an ideal form. This leads to the existence of points exhibiting high gradients causing drastic perceptual change in small $\epsilon$-step in the latent space.

### 2.1. Regularizers for scaled-isometry

The authors of StyleGANv2 proposed a following path length regularizer ($L_pl$) that guides the mapping to become a scaled-isometry associated with a dynamically computed global scale $c$. $J_w$ is the Jacobian of the mapping $f$ at $w$.

Let an image of $f$ be a manifold. Then $J_w$ becomes a vector that spans a tangent space at $w$. Suppose the tangent space is perturbed small as $y \sim N(0, I)$. It means the amount of perturbation of the tangent space is $< \frac{\partial f}{\partial w}, y >= J_w^T y$ in each direction. That is, the role of the Jacobian is to provide a local approximation for each point $w$ and $||J_w^T y||$ means how much the image $y$ moves in the latent space. If there is any direction that changes abruptly, then the latent space will be unstable. Also $L_pl$ can be treated as variance estimator. Since the most import role of pl regularizer is stretching $W$ space orthogonally and wanted to keep same scale regardless of the direction. Therefore, we set $||J_w^T y||$ as a constant to estimate the variance as constant $'c'$ in any direction. Since we set $||J_w^T y||$ as a constant, perturbing the image by a small amount at a local point in any direction, there will be no unstable abrupt changes as long as $||y||$ is same.

In addition, constant $c$ is determined by exponential moving average of $|J_w^T y|_2$ with a period of 100 evaluations. By dynamically determining the scaled-isometry's scale with moving average, the regularizer guides $f$ to become a $c$ scaled-isometry.

$$L_{pl} = E_{w,y\sim N(0,I)}(||J_w^T y||_2 - c)^2 \tag{1}$$

However, because the global scale is determined in a dynamic sense, the training process becomes an iterative trial of finding a appropriate scale, and the sequence of mapping has potential to converge to isometry with different scale for every distinct training. Also, preference of specific global scale $c$ should not exist because the important relation that mapping should satisfy is the conservation of distance and angle up to scale, while not restricted to any desired scale.

Formally, for two mapping $f$ and $f'$ satisfying $J_w^T = c' J'^T$ for some constant $c'$, the desired scaled-isometry regularizer would equally penalize both mappings, but the defined path length regularizer will prefer mapping with the smaller Jacobian. This implies the regularizer form does not take account the equivalence relation between mapping when its Jacobians are equivalent up to scale.

In [19], the equivalence relation is stated in more generalized sense, also considering the Riemannian metric $H$ of the image space. We suppose Euclidean geometry and set the Riemannian metric to become identity.

$$L(f) = L(f') \quad \text{if} \quad J^T H J = c' J'^T H J' \\ \text{, for some} \quad c' > 0 \tag{2}$$

This motivates the use of generalized formulation of the path length regularizer that takes account of the desired equivalence relation as well as the elimination of dynamic training process rooting from explicit calculation of global scale $c$.

### 2.2. Relaxed Distortion Measure

Meanwhile, [19] proposes a coordinate-invariant functional that measure how much the function is far from being a scaled-isometry. The authors named it relaxed distortion measure, and one of its formulation is given as the following.

$$L_{dm} = \int_{\mathbb{R}^m} \sum_{i=1}^{m} \left( \frac{\lambda_i(z)}{\int_{\mathcal{M}} \sum_{j=1}^{m} \frac{\lambda_i(z')}{m} dv(z')} - 1 \right)^2 dv(z) \tag{3}$$

By following the proposition in [19] and using the Hutchinson's trace estimator to estimate the trace value for efficient computation, it leads to the following formulation. $(H(z) = J^T(z)J(z))$

$$L_{dm} = \frac{E_{z\sim p_\theta}[Tr(H^2(z)]}{E_{z\sim p_\theta}[Tr(H(z)^2]} = \frac{E_{z,v\sim N(0,I)}[v^T H^2 v]}{E_{z,v\sim N(0,I)}[v^T H v]^2} \tag{4}$$

Using this distortion measure, we may measure how much the mapping from latent to image is far from being a scaled-isometry and this was the exact task $L_{pl}$ tried to achieve. $L_{dm}$ does not explicitly guide the mapping to be associated with a prescribed scale $c$, but implicitly guides the mapping to beceome one of the cases. Also, the convergence of $L_{dm}$ will imply satisfaction of the equivalence relation, that was not achieved with $L_{pl}$. These reasons sum up to explain the advantages of distortion measure regularizer over the original path length regularizer.

## 3. Experiments

We investigate on the application of scaled-isometry regularizer in two powerful generative models, StyleGANs and Diffusion models. We first studied on the effect of substituting the original path length regularizer to distortion measure based regularizer in StyleGAN. We also want to study the effect of the scaled-isometry regularizer when applied to iterative models such as Diffusion models, but since Diffusion models sample the generating image in iterative manner, different formulation is required for the guidance toward mapping to become close to a scaled-isometry.

### 3.1. Distortion measure regularizer in StyleGAN

We experimented the effect of distortion measure regularization on StyleGAN by comparing three metrics in different configurations. The metrics are FID, computed with 50k samples, PPL, which stands for perceptual path length metric, and DM, which stands for relaxed distortion measure that computes the distance between the trained mapping and scaled-isometry. The newly proposed DM metric will represent how much the mapping is distorted from scaled-isometry, or how much the mapping is close to scaled-isometry.

### 3.2. Preparing for datasets

For our image dataset, we used public image datasets FFHQ(Flickr-Faces-HQ) and AFHQ(Animal-Faces-HQ). FFHQ dataset consits of 70,000 high-quality PNG images at


Figure 3. Comparison between images generated by StyleGANv3, path length regularized StyleGANv3, and distortion measure regularized StyleGANv3 on FFHQ dataset.

| **Configuration** | AFHQ, 512x512 | | FFHQ, 256x256 | |
|---|---|---|---|---|
| | FID ↓ | PPL ↓ | FID ↓ | PPL ↓ |
| A = StyleGAN3 [6] | 4.68 | 963.8 | 13.90 | 2136.4 |
| B = A + Path length regularization [8] | 5.64 | 319.4 | 16.46 | 378.2 |
| C = A + Distortion measure regularization | 4.59 | 652.1 | 9.37 | 1010.74 |

Table 1. We report the best metrics measured at checkpoints with the best FID score after training with 400,000 images. FID is measured by sampling 50k images from the trained generator. PPL stands for perceptual path length metric, which is computed from average perceptual distance between pair of images corresponding to randomly sampled pair of latents. DM stands for relaxed distortion measure, that we have used to generalize the regularizer that guides the mapping to become a scaled-isometry.

1024×1024 resolution and contains considerable variation in terms of age, ethnicity and image background. AFHQ dataset consists of 15,000 high-quality images at 512×512 resolution. The dataset includes three domains of cat, dog, and wildlife, each providing 5000 images. After downloading the original datasets, we downsampled FFHQ dataset to 256x256 pixel size.

### 3.3. Experiment settings

To conduct our experiments, we calcuated FID and PPL scores using three configurations. config A is a baseline StyleGAN3 checkpoint obtained from the work of [6]. The authors released their code and models, including the baseline checkpoint on github. We used the baseline trained with config stylegan3-t, which is a translation equivalent configuration. For config B, we conducted a transfer learning experiment with the application of path length regularizer from StyleGAN2, using config A as basline. Lastly, for configuration C, we conducted another transfer learning experiment with the baseline config A, using scaled-isometry regularizer. Table 1 compares results of three configurations experimented with FFHQ256x256 and AFHQ512x512 datasets.

For calculation of FID score and PPL score, we used the metric fid50k and ppl2_wend implemented in the official Pytorch implementation of StyleGAN3. The metric fid50k [11] refers to Fréchet inception distance against the full dataset, and ppl2_wend [8] refers to perceptual path length in W, endpoints, full image. It is known that lower FID and lower PPL score both improves image quality. These two metrics differ in the fact that lower FID is relatively more related to realistic textures, where lower PPL is more related to improving semantic consistency of the images. For each config, we trained the model for 80,000 images and calculated the metrics.

### 3.4. Results

While path length regularizer decreased PPL and smoothed out the mapping in both AFHQ and FFHQ dataset, the FID score significantly increased from the score of pretrained model. This implies the tradeoff between PPL and FID, sacrificing FID score in order to increase the smoothness of the mapping. On the other hand, using the distortion measure regularizer, PPl decreased with a smaller step compared to the setting of path length regularized training, but the decrease of FID during training was significantly reduced.

We interpret the result as the path length regularizer was constraining the model to learn a isometrical mapping with a prescribed scale constant, in a hard manner, and forces the sacrifice of FID during training. In distortion measure training, the hard constrain is released, in a constant-free, coordinate invaraint manner, and we release the tradeoff relation between PPL and FID. In the place of PPl, one could substitute an arbitrary distortion measure of the mapping.

In particular, the generated result from model trained with FFHQ dataset shows an interesting property. Since face images from the dataset is randomly cropped, baseline model generates faces that is randomly positioned inside the given frames. On the other hand, using the isometry regularizer, the regularizer forces the generated images to have constant distances with respect to arbitrary latent directions, and generates faces positioned at the vicinity of the center of the frames.

### 3.5. Isometry regularizer in Diffusion models

Recently, Diffusion models are performing excellent generation quality in various domains including vision, audio, natural language, and scientific applications such as molecule docking and climate event forecasting. Hence, it is natural to expect that one may want to extend the formulation of distortion measure regularizer in the language of diffusion models, in order to give guidance to isometric representation learning, expecting a smooth and well-conditioned latent space of Diffusion models.

(5) gives the Poisson Flow ODE's reverse diffusion process [15] where trained Diffusion model, or the score model outputs the partially denoised image when given the noisy image as input. When isometry regularizer is applied to every step of reverse process in diffusion model, we may anticipate that the overall reverse process will be isometric. Since ODE reverse diffusion process is deterministic, we can formulate it as a function $f$.

$$f : x \to x + dx = x + [f(x,t) - \frac{g^2(t)}{2}\nabla_x logp_t(x)]dt \quad (5)$$

Then we may calculate its Jacobian as shown in (6), or more explicitly as in (7).

$$J = \frac{df}{dx'} = I + \nabla_{x'}[f(x,t) - \frac{g^2(t)}{2}\nabla_x logp_t(x)]dt \quad (6)$$

$$J_{ij} = \frac{df_i}{dx'_j} = \delta_{ij} + [\frac{\partial f_i(x,t)}{\partial x_j} - \frac{g^2(t)}{2}\frac{\partial^2 logp_t(x)}{\partial x_j \partial x_i}]dt$$
$$\text{, where } \quad \frac{\partial^2 logp_t(x)}{\partial x_j \partial x_i} = \frac{\partial s_\theta(x,t)}{\partial x_j} \quad (7)$$

$\frac{\partial s_\theta(x,t)}{\partial x_j}$ is a Jacobian of score function, or in another words, Hessian of the data distribution. That is to say, we may explicitly use the 2nd order derivative of the data distribution to apply distortion measure regularizer to Diffusion model in order to force the model to learn an isometric mapping from $i$th denoised manifold to $i+1$ th denoised manifold.

Diffusion models using information of higher order gradients of the data distribution was shown to have advantages through accelerating the sampling process [2]. We only give theoretical formulation of the distortion measure regularizer here, and leave the investigation to Isometric Diffusion models to future work.

## 4. Conclusions and future work

In this work, we applied distortion measure regularizer to StyleGANv3 to show its effect on forcing the generative model to learn an isometric mapping from latents to images. This isometric mapping leads to training of smooth mapping which fills out the absent parts of generated images compared to baseline StyleGANv3 and smoother transition during latent interpolations. Isometric mapping implies constant magnitude of image change in arbitrary latent directions which inhibits drastic image change during random walking in the latent space.

StyleGANv2 previously proposed Path length regularizer to achieve the goal of smoothing out the latent space, which involved a prescribed constant that explicitly restricted the model to learn a scaled isometry related to the specifically prescribed scale constant. We proposed distortion measure that relax this condition and calculates the degree of distortion in a constant-free, coordinate-invariant sense.

Qualitative comparison between images generated from StyleGANv3s trained with different regularizers show that applying isometry regularizers yields a smoother mapping from latent space, generating a smoother transition during latent random walking. Path length regularizer decreases PPL but due to its explicit regularizing, FID increases, as a result of tradeoff with PPL. Our distortion measure is constant-free, coordinate invariant and is a relaxed measure of the smoothness of the mapping. With our relaxed regularizer, PPL decreases without tradeoff with FID score.

For future work, we leave the extension of distortion measure to Diffusion models. Because diffusion models iteratively samples the images, different formulation of distortion measure needs to be made. We give a theoretical proposal of distortion measure in Diffusion models and leave its application to future work. Because diffusion model showed drastics changes in latent random walking in preceding works, we look forward to see the effect of distortion measure regularizer to Diffusion Models.

## References


[1] Francesco Barbato, Marco Toldo, Umberto Michieli, and Pietro Zanuttigh. Latent space regularization for unsupervised domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2835–2845, 2021. 1

[2] Tim Dockhorn, Arash Vahdat, and Karsten Kreis. Genie: Higher-order denoising diffusion solvers. *arXiv preprint arXiv:2210.05475*, 2022. 6

[3] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020. 1

[4] Cheongjae Jang, Yung-Kyun Noh, and Frank Chongwoo Park. A riemannian geometric framework for manifold learning of non-euclidean data. *Advances in Data Analysis and Classification*, 15(3):673–699, 2021. 1

[5] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *arXiv preprint arXiv:2206.00364*, 2022. 1

[6] Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Alias-free generative adversarial networks. *Advances in Neural Information Processing Systems*, 34:852–863, 2021. 1, 5

[7] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019. 1

[8] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020. 1, 5

[9] Yahui Liu, Enver Sangineto, Yajing Chen, Linchao Bao, Haoxian Zhang, Nicu Sebe, Bruno Lepri, Wei Wang, and Marco De Nadai. Smoothing the disentangled latent style space for unsupervised image-to-image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10785–10794, 2021. 1, 2

[10] Zeyu Lu, Chengyue Wu, Xinyuan Chen, Yaohui Wang, Yu Qiao, and Xihui Liu. Hierarchical diffusion autoencoders and disentangled image manipulation. *arXiv preprint arXiv:2304.11829*, 2023. 2

[11] Thomas Unterthiner Bernhard Nessler Sepp Hochreiter Martin Heusel, Hubert Ramsauer. Gans trained by a two timescale update rule converge to a local nash equilibrium. *arXiv preprint arXiv:1706.08500*, 2017. 5

[12] Konpat Preechakul, Nattanat Chatthee, Suttisak Wizadwongsa, and Supasorn Suwajanakorn. Diffusion autoencoders: Toward a meaningful and decodable representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10619–10629, 2022. 2

[13] Kuniaki Saito, Donghyun Kim, Stan Sclaroff, and Kate Saenko. Universal domain adaptation through self supervision. *Advances in neural information processing systems*, 33:16282–16292, 2020. 1

[14] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. *Advances in neural information processing systems*, 29, 2016. 1

[15] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020. 1, 6

[16] Na Su and Yun Pan. Research on evaluation method of image style transfer algorithm. In *2022 3rd International Conference on Information Science, Parallel and Distributed Systems (ISPDS)*, pages 430–434. IEEE, 2022. 1

[17] Shaoan Xie, Yanwu Xu, Mingming Gong, and Kun Zhang. Unpaired image-to-image translation with shortest path regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10177–10187, 2023. 1

[18] Mao-Chuang Yeh, Shuai Tang, Anand Bhattad, Chuhang Zou, and David Forsyth. Improving style transfer with calibrated metrics. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3160–3168, 2020. 1

[19] LEE Yonghyeon, Sangwoong Yoon, MinJun Son, and Frank C Park. Regularized autoencoders for isometric representation learning. In *International Conference on Learning Representations*, 2022. 1, 2, 4

[20] Zijian Zhang, Zhou Zhao, and Zhijie Lin. Unsupervised representation learning from pre-trained diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 35:22117–22130, 2022. 1


# A. Appendix

We give more comparison of generation results from baseline StyleGANv3, path length regularized StyleGANv3, and distortion measure regularized StyleGANv3.


Figure 4. Comparison between images generated by StyleGANv3, path length regularized StyleGANv3, and distortion measure regularized StyleGANv3.