The Euclidean Space is Evil: Hyperbolic Attribute Editing for Few-shot Image Generation

Lingxiao Li[1] Yi Zhang[2] Shuhui Wang[3]*

[1] Columbia University [2] University of Oxford
[3] Institute of Computing Technology, Chinese Academy of Sciences
ll3504@columbia.edu, wolf5965@ox.ac.uk, wangshuhui@ict.ac.cn

Abstract

Few-shot image generation is a challenging task since it aims to generate diverse new images for an unseen category with only a few images. Existing methods suffer from the trade-off between the quality and diversity of generated images. To tackle this problem, we propose Hyperbolic Attribute Editing (HAE), a simple yet effective method. Unlike other methods that work in Euclidean space, HAE captures the hierarchy among images using data from seen categories in hyperbolic space. Given a well-trained HAE, images of unseen categories can be generated by moving the latent code of a given image toward any meaningful directions in the Poincaré disk with a fixing radius. Most importantly, the hyperbolic space allows us to control the semantic diversity of the generated images by setting different radii in the disk. Extensive experiments and visualizations demonstrate that HAE is capable of not only generating images with promising quality and diversity using limited data but achieving a highly controllable and interpretable editing process. Code is available at https://github.com/lingxiao-li/HAE.

1. Introduction

Due to the persistent development of deep learning, the task of image generation has received significant research attention in recent years. Specifically, the Generative Adversarial Networks (GANs) [21] and its variants (*e.g.*, StyleGANv2 [34]) have succeeded in generating high-fidelity and realistic images, requiring a large number of high-quality data for model training. However, considering the long-tail distribution and data imbalance widely exists among different image categories [30], it is difficult for GANs to be trained on categories with sufficient training images to generate new realistic and diverse images for a category with only a few images. This task is referred to as few-shot image generation [10, 26, 30, 28, 29, 27, 15]. A variety of tasks can benefit from improvements in few-shot image generation, for instance, low-data detection [17] and few-shot classification [54, 57].

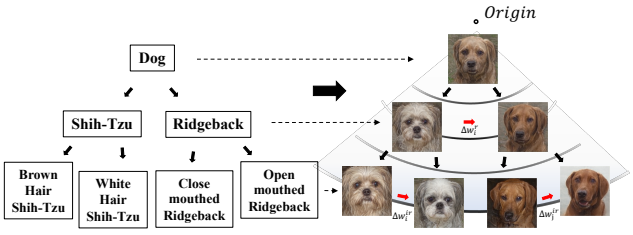

Figure 1: **Illustration of hierarchical attribute editing in hyperbolic space.** Hyperbolic space can naturally and compactly encode semantic hierarchical structures within a large image data corpus. Changing the high-level, *i.e.*, category-relevant attribute $\Delta\mathbf{w}^r$ changes the category of an image. While changing low-level or category-irrelevant attribute $\Delta\mathbf{w}^{ir}$ varies images within categories.

In general, existing GAN-based few-shot image generation mechanisms can be classified into three categories. Transfer-based methods [10, 39] introduce meta-learning or domain adaptation on GANs to generate new images by enforcing knowledge transfer among categories. Fusion-based methods [2, 23, 30, 28] perform feature fusion of multiple input images in a feature space and generate images via decoding the fused features back to image space. However, the output is still highly similar to the source images. Transformation-based methods [1, 29, 27, 15] find intra-category transformations or inject random perturbations to conditional unseen category samples to generate images without tedious fine-tuning. By representing the images in the Euclidean feature space, the above learning mechanisms tend to be over-complicated, and the generated

*Corresponding author

images are often collapsed due to limited diversity.

Similar to the ubiquity of hierarchies in language [44, 56, 13], the semantic hierarchy is also common in images [35, 11]. As Fig. 1 shows, the semantic hierarchies constructed in the language domain can be instantiated with visual images. From the visual perspective, an image can be regarded as a collection of attributes of multiple levels. High-level attributes, *a.k.a.* category-relevant attributes, define the category of an image, such as the shape and color of an animal [15]. For instance, in the middle row of Fig. 1, changing the high-level attributes of the given image of a `Shih-Tzu dog`, the category can be changed to a `Rhodesian Ridgeback Dog`. While the low-level or fine-grained attributes, including expressions, postures, *etc*., that vary within the category as shown at the bottom of Fig. 1, are called category-irrelevant attributes. Therefore, an image can also be viewed as a descendant of another image with the same category-relevant attributes by adding fine-grained category-irrelevant attributes to its parent image. To edit the visual attributes for high-quality image generation, it is crucial to capture the attribute *hierarchy* within the large image data corpus and find a good representation space. Ideally, we aim to construct a hierarchical visual representation in a latent space that allows us to change the category of an image by moving the latent code in a category-relevant direction, and perform few-shot image generation by moving the code in a category-irrelevant direction.

Unfortunately, the Euclidean space and its corresponding distance metrics used by existing GAN-based methods can not facilitate the hierarchical attribute representation, thus the design of complicated attribute disentangling and editing mechanisms seems to be crucial for the generation quality. Inspired by the application of hyperbolic space in images [35] and videos [55], we found that the metrics introduced in hyperbolic geometry can naturally and compactly encode hierarchical structures. Unlike the general affine spaces, *e.g*., the Euclidean space, hyperbolic spaces can be viewed as the continuous analog of a tree since tree-like graphs can be embedded in finite-dimension with minimal distortion [44]. This property of hyperbolic space provides continuous and up to infinite semantic levels for attribute editing, allowing us to robustly generate diverse images with only a few images from unseen categories with simple operations.

Based on the above findings, we propose a simple but effective Hyperbolic Attribute Editing (HAE) method for few-shot image generation. Our method is based on the observation that hierarchical latent code manipulation can be easily implemented in Hyperbolic space. The core of HAE is mapping the latent vectors from the Euclidean space $\mathbb{R}^n$ to a hyperbolic space $\mathbb{D}^n$. We minimize a supervised classification loss function to ensure the images are hierarchically embedded in hyperbolic space. Once we capture the attribute hierarchy among images, we can generate new images of unseen categories by moving the latent code from one leaf to another with the same parents by fixing the radius. Most importantly, the hyperbolic space allows us to control the semantic diversity of generated images by setting different radii in the Poincaré disk. Those operations can well facilitate continually hierarchical attribute editing in hyperbolic space for flexible few-shot image generation with both quality and diversity.

Our contributions can be summarized as follows:

- We propose a simple yet effective method for few-shot image generation, *i.e*., hyperbolic attribute editing. In order to capture the hierarchy among images, we use hyperbolic space as the latent space. To the best of our knowledge, HAE is the first attempt to use hyperbolic latent spaces for few-shot image generation.
- We show that in our designed hyperbolic latent space, the semantic hierarchical attribute relations among images can be reflected by their distances to the center of the Poincaré disk.
- Extensive experiments and visualization suggest that HAE achieves stable few-shot image generation with state-of-the-art quality and diversity. Unlike other few-shot image generation methods, HAE allows us to generate images with better control of diversity by changing the semantic levels of attributes we want to edit.

2. Related Work

Few-shot image generation. Recently, diverse methods have been proposed for few-shot image generation. The transfer-based methods [10, 39] which introduce meta-learning or domain adaptation on GANs can hardly generate realistic images. While fusion-based methods that fuse the features by matching the random vector with the conditional images [28] or formulating the problem as a conditional generating task [23, 30] suffer from the limited diversity of generated images. Furthermore, transformation-based methods [1, 29, 27, 15] can generate images with only one conditional image by focusing on either capturing the cross-category or intra-category transformations by injecting random perturbations [1]. Nevertheless, the transformation captured by those methods is not very consistent. Ding *et al.* [15, 14] propose the "editing-based" perspective, the intra-category transformation can be modeled as category-irrelevant image editing based on one sample instead of pairs of samples. Most recently, Zhu *et al.* [60] fine-tune powerful diffusion models (DMs) [25] pre-trained on large source domains on limited target data to generate diverse and high quality images. DMs outperform GANs [21] on sample quality with a more controllable training process

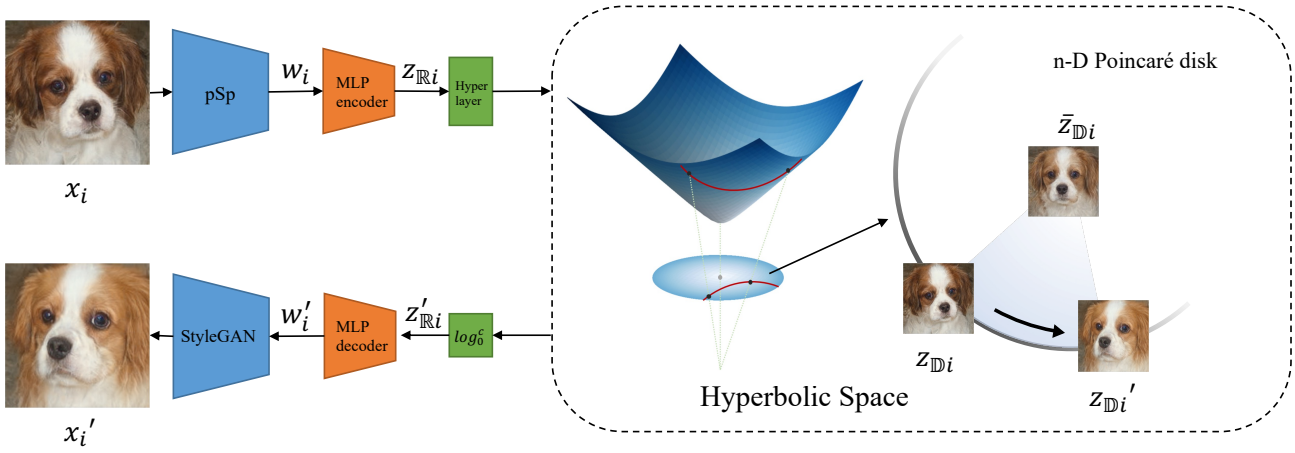

Figure 2: The overview of HAE. The **Hyper layer** is a hyperbolic feedforward layer called *Möbius linear layer* which is used to project the latent code from Euclidean space $\mathbb{R}^n$ to hyperbolic space $\mathbb{D}^n$ [7]. $\bar{z}_{\mathbb{D}i}$ can be viewed as the "parent" or average code of $z_{\mathbb{D}i}$ and $z'_{\mathbb{D}i}$. One can generate diverse images without changing the category by moving the latent code from one child to another of the same parent in the hyperbolic space.

at the cost less flexibility and editability, since they denoise images in the image space rather than operate in the latent space. Furthermore, the inference process of DMs is much slower than GANs [52].

Hyperbolic Embedding. The use of hyperbolic space in deep learning [44, 45, 56, 55, 35] is a pioneering work in recent years. It was first used in natural language processing for hierarchical language representation [44, 45, 56]. The Riemannian optimization algorithms are used to optimize models in hyperbolic space [5, 3]. As hyperbolic space is successfully applied to represent hierarchical data, Ganea *et al.* [18] derives hyperbolic versions of tools in neural networks including multinomial logistic regression, feed-forward, and recurrent neural networks. Following this, hyperbolic geometry is used in image [35], video [55], and graph data [7, 47]. Most recently, Lazcano *et al* [36] shows that hyperbolic space outperforms traditional Euclidean space in image generation using HGAN. However, the hierarchy and controllability of hyperbolic space remain uninvestigated in HGAN, as the generator is still governed by Gaussian samples in Euclidean space.

Latent Code Manipulation. It has been shown that the latent spaces of GANs are able to encode rich semantic information [20, 32, 50]. One of the popular approaches is finding linear directions corresponding to changes in a given binary labeled attributes, which might be difficult to obtain for new datasets and could require manual labeling effort [50, 20, 12]. Others [8, 58, 41, 31, 9] try to find semantic directions in an unsupervised manner. For instance, PCA is applied in the latent space to create interpretable controls for synthesizing images [31, 9]. Most recent works [53, 51] directly compute in the close form to find the meaningful semantic direction without training and optimization. In comparison, our work HAE focuses on attributes in different semantic levels in the latent space rather than trying hard to find disentangled interpretable directions as previous works.

3. Method

The overall framework of HAE is shown in Fig. 2, we first give a detailed explanation of getting the hierarchical representations in the hyperbolic space, and then we introduce the framework of HAE and explain the loss functions.

3.1. Hierarchical Representation

The major issue of our study is how to obtain the hierarchical representation from real images to facilitate editing in different semantic levels, as illustrated in Fig. 1. Therefore, *hyperbolic* space is introduced as the latent space to achieve this goal.

Unlike Euclidean spaces with their zero curvature and spherical spaces with their positive curvature, hyperbolic spaces with negative curvature have been shown that it is more appropriate for learning hierarchical representation [44, 45]. Informally, hyperbolic space can be viewed as a continuous analogy of trees [44]. One important feature of hyperbolic space is that the length grows exponentially with its radius while linearly in Euclidean space. This property allows hyperbolic space to be naturally compatible with hierarchical data [22] including text, images, videos, *etc*.

The n-dimensional hyperbolic space can be formally defined as a homogeneous, simply connected n-dimensional Riemannian manifold, denoted as $\mathbb{H}^n$ with constant negative sectional curvature[1]. We choose to work in the Poincaré disk from five isometric models of hyperbolic

[1]The curvature of the hyperbolic space c is set as -1 in this work.

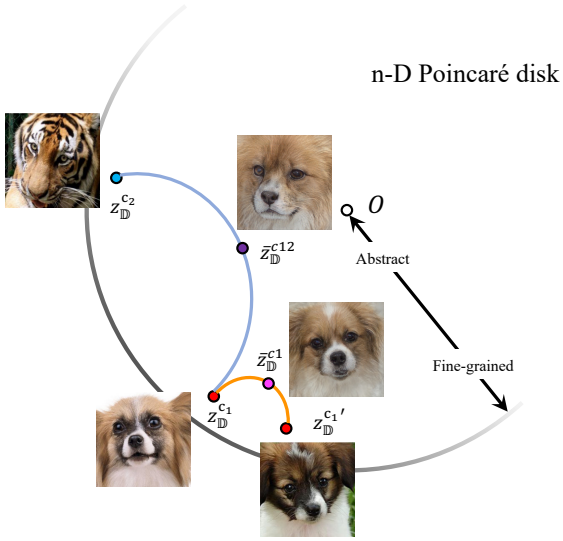

Figure 3: **Illustration of the property of hyperbolic space on the Poincaré disk.** Given two latent codes of Spaniel $z_{\mathbb{D}}^{c1}$ and $z_{\mathbb{D}}^{c1\prime}$(red dots) on the edge of Poincaré disk, the geodesic between these two points is the brown curve rather than a straight line in Euclidean space. Therefore, their average latent code is calculated as $\bar{z}_{\mathbb{D}}^{c1}$(pink dot) which is closer to the center O (still a Spaniel, but less fine-grained). While the latent code of a tiger $z_{\mathbb{D}}^{c2}$(blue dot) locates far from the latent code of a Spaniel. Thus, the hyperbolic average code of tiger and Spaniel $\bar{z}_{\mathbb{D}}^{c12}$(purple dot) is closer to the center O than $\bar{z}_{\mathbb{D}}^{c1}$ which is more abstract (a feline contains features from both tiger and Spaniel).

space defined in [6] since it is commonly used in gradient-based learning [44, 18, 45, 56, 55, 35]. The Poincaré disk model $\left(\mathbb{D}^n, g^{\mathbb{D}}\right)$ is defined by the manifold $\mathbb{D}^n = \{x \in \mathbb{R}^n : \|x\| < 1\}$ equipped with the following Riemannian metric:

$$g_x^{\mathbb{D}} = \lambda_x^2 g^E, \tag{1}$$

where $\lambda_x = \frac{2}{1-\|x\|^2}$, and g^E is the Euclidean metric tensor $g^E = \mathbf{I}^n$. The induced distance between two points $\mathbf{x}, \mathbf{y} \in \mathbb{D}^n$ can be defined by:

$$d_{\mathbb{D}}(\mathbf{x}, \mathbf{y}) = \operatorname{arccosh}\left(1 + 2\frac{\|\mathbf{x}-\mathbf{y}\|^2}{(1-\|\mathbf{x}\|^2)(1-\|\mathbf{y}\|^2)}\right). \tag{2}$$

Recall that a geodesic is a locally minimized-length curve between two points. In the hyperboloid model, the geodesic can be defined as the curve created by intersecting the plane defined by two points and the origin with the hyperboloid [38]. Thus, the mean of two latent codes in hyperbolic space locates at the mid-point of the geodesic that is closer to the origin. This is the key desired feature of hyperbolic space, *i.e.*, the mean between two leaf embeddings is not another leaf embedding, but the hierarchical parent of them [55]. This feature allows us to generate new images by moving the latent code from one leaf to another with the same parents. We can also change the semantic levels of attributes by determining how abstract their parent is.

This unique property is visualized in Fig. 3 on a 2-D Poincaré disk. The image embedding near the edge of the ball (with a large radius) represents a more fine-grained image while the embedding near the center (which has a smaller radius) represents an image with abstract features (an average face).

Although the hyperbolic space shares similar features with trees, it is continuous. In other words, there is no fixed number of hierarchy levels. Instead, there is a continuum from very fine-grained (near the edge of Poincaré disk) to very abstract (near the origin).

3.2. Network Architecture

Although we aim to embed and edit real images in hyperbolic space, the whole network does not need to be implemented in a hyperbolic manner. Instead, we can take advantage of the number of existing GAN inversion models and optimization algorithms that have been fine-tuned for Euclidean space.

To achieve image editing, we need to embed the image back into the latent space. In particular, we select pSp [49] as the backbone of HAE to encode images to the $\mathcal{W}^+$-space of StyleGAN2 [34]:

$$\mathbf{w}_i = \mathrm{pSp}\left(x_i\right), \tag{3}$$

where $\mathbf{w}_i \in \mathbb{R}^{18\times 512}$ is the corresponding latent vector of x_i in the $\mathcal{W}^+$-space.

To manipulate latent code in hyperbolic space, we need to define a bijective map from $\mathbb{R}^n$ to $\mathbb{D}_c^n$ to map Euclidean vectors to the hyperbolic space and vice versa. A manifold is a differentiable topological space that locally resembles the Euclidean space $\mathbb{R}^n$ [37, 38]. For $x \in \mathbb{D}^n$, one can define the tangent space $T_x\mathbb{D}_c^n$ of $\mathbb{D}_c^n$ at x as the first order linear approximation of $\mathbb{D}_c^n$ around x. Therefore, this bijective map can be performed by exponential and logarithmic maps. Specifically, the *exponential map* $\exp_{\mathbf{x}}^c : T_{\mathbf{x}}\mathbb{D}_c^n \cong \mathbb{R}^n \to \mathbb{D}_c^n$, maps from the tangent spaces into the manifold. While the *logarithmic map* $\log_{\mathbf{x}}^c : \mathbb{D}_c^n \to T_{\mathbf{x}}\mathbb{D}_c^n \cong \mathbb{R}^n$ is the reverse map of the exponential map.

We use exponential and logarithmic maps at origin $\mathbf{0}$ for the transformation between the Euclidean and hyperbolic representations. After getting $\mathbf{w}_i$ in the $\mathcal{W}^+$-space, we first use a Multi-layer Perceptron (MLP) encoder to reduce the dimension of latent vectors in Euclidean space. Then we apply an exponential map to project the Euclidean latent code $z_{\mathbb{R}i}$ to hyperbolic space. After that, we use the hyperbolic feed-forward layer as [18] to obtain the final hierarchical representation $z_{\mathbb{D}}$ as shown in Fig. 2:

$$z_{\mathbb{D}i} = f^{\otimes_c}(\exp_{\mathbf{0}}^c(\mathtt{MLP}_E(\mathbf{w}_i))), \tag{4}$$

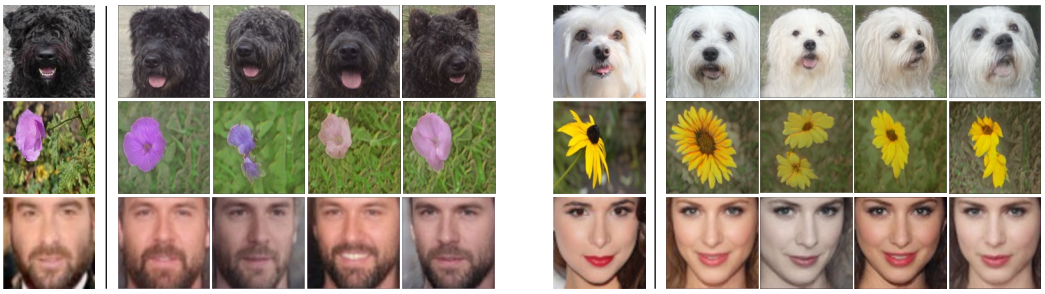

Figure 4: One-shot image generation from HAE on Animal Faces, Flowers, and VGGFaces.

where $f^{\otimes_c}$ is the Möbius translation of feed-forward layer f as the map from $\mathbb{D}_c^n$ to $\mathbb{D}_c^m$, denoted as *Möbius linear layer*.

Finally, the hyperbolic representation $z_{\mathbb{D}}$ needs to be projected back to the $\mathcal{W}^+$-space of StyleGAN2. In practice, this is achieved by applying a logarithmic map followed by an MLP decoder:

$$\mathbf{w}_i' = \texttt{MLP}_D(\log_{\mathbf{0}}^c(z_{\mathbb{D}i})), \tag{5}$$

and $\mathbf{w}_i'$ will be fed into a pre-trained StyleGAN2's generator G to reconstruct the image x_i'.

3.3. Loss Function

The loss function of HAE consists of two parts: the *Hyperbolic loss* ensures to get the hierarchical representation in the hyperbolic space and the *reconstruction loss* guarantees the quality of reconstruction images.

Hyperbolic Loss. To learn the semantic hierarchical representation of real images in hyperbolic space, we minimize the distance between latent codes of images with similar categories and attributes while pushing away the latent codes from different categories. We choose the supervised approach to achieve this. In order to perform multi-class classification on the Poincaré disk defined in Sec. 3.1, one needs to generalize multinomial logistic regression (MLR) to the Poincaré disk defined in [18]. An extra linear layer needs to be trained for the classification and the softmax probability can be computed as: Given K classes and $k \in \{1, \ldots, K\}, p_k \in \mathbb{D}_c^n, a_k \in T_{p_k}\mathbb{D}_c^n \backslash \{\mathbf{0}\}$:

$$p(y=k \mid x) \propto \exp\left(\frac{\lambda_{p_k}^c \left\|a_k\right\|}{\sqrt{c}} \sinh^{-1}\left(\frac{2\sqrt{c}\left\langle -p_k \oplus_c x, a_k\right\rangle}{\left(1-c\left\|-p_k \oplus_c x\right\|^2\right)\left\|a_k\right\|}\right)\right), \quad \forall x \in \mathbb{D}_c^n. \tag{6}$$

where $\oplus_c$ denotes the Möbius addition defined in [35] with fixed sectional curvature of the space, denoted by c.

After getting the softmax result for each class, one can use *negative log-likelihood loss* (NLL Loss) to calculate the hyperbolic loss:

$$\mathcal{L}_{\text{hyper}} = -\frac{1}{N}\sum_{n=1}^{N}\log(p_n), \tag{7}$$

where N is the batch size and p_n is the probability predicted by the model for the correct class.

As mentioned in Sec. 3.1, the distance between points grows exponentially with their radius in the Poincaré disk. In order to minimize Eq. (7), the latent codes of fine-grained images will be pushed to the edge of the ball to maximize the distances between different categories while the embedding of abstract images (images have common features from many categories) will be located near the center of the ball. Since hyperbolic space is continuous and differentiable, we are able to optimize Eq. (7) with stochastic gradient descent, which learns the hierarchy of the images.

Reconstruction Loss. In order to guarantee the quality of the generated images, we first use the $\mathcal{L}_2$ loss and LPIPS loss used in pSp [49], given image x_i:

$$\mathcal{L}_2(x_i) = \|x_i - \texttt{HAE}(x_i)\|_2. \tag{8}$$

$$\mathcal{L}_{\text{LPIPS}}(x_i) = \|F(x_i) - F(\texttt{HAE}(x_i))\|_2, \tag{9}$$

where $F(\cdot)$ denotes the perceptual feature extractor.

Since the pSp encoder and StyleGAN2 generator are pre-trained, we only train the neural layers between the encoder and generator of HAE. To further guarantee the network to better project back to the $\mathcal{W}^+$-space, the reconstructed $\mathbf{w}_i'$ should be the same as the original $\mathbf{w}_i$:

$$\mathcal{L}_{\text{rec}}(w_i) = \|\mathbf{w_i} - \mathbf{w}_i'\|_2, \tag{10}$$

where $\mathbf{w}_i'$ can be calculated by Eq. (4) and Eq. (5).

The **overall loss function** is:

$$\mathcal{L} = \mathcal{L}_2(x_i) + \lambda_1\mathcal{L}_{\text{LPIPS}} + \lambda_2\mathcal{L}_{\text{rec}} + \lambda_3\mathcal{L}_{\text{hyper}}, \tag{11}$$

where λ_1, λ_2 and λ_3 are trade-off adaptive parameters. This curated set of loss functions ensures the model learns the hierarchical representation and reconstructs images.

Figure 5: Images generated by HAE by adding the same perturbation on the latent code of a given image with different hyperbolic radii on Animal Faces and Flowers.

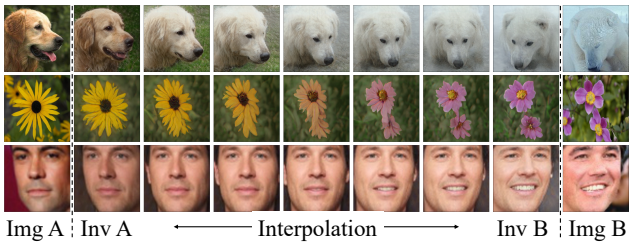

Figure 6: Interpolations in hyperbolic space along the edge of the Poincaré disk (with $r_{\mathbb{D}} = 6.2126$) on three datasets.

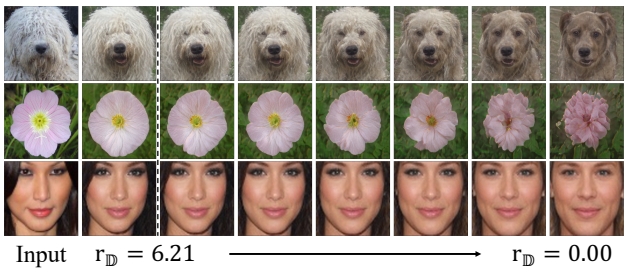

Figure 7: Interpolations by moving the latent codes from the edge to the center of the Poincaré disk (from fine-grained to abstract) on three datasets.

3.4. Image Generation

To study the generating quality of the model, a straightforward way is to generate new images via interpolation between two designated images, or random perturbation.

In hyperbolic space, the shortest path with the induced distance between two points is given by the geodesic defined in Eq. (2). The geodesic equation between two embeddings $z_{\mathbb{D}\mathbf{i}}$ and $z_{\mathbb{D}\mathbf{j}}$, denoted by $\gamma_{z_{\mathbb{D}i} \to z_{\mathbb{D}j}}(t)$, is given by

$$\gamma_{z_{\mathbb{D}i} \to z_{\mathbb{D}j}}(t) = z_{\mathbb{D}i} \oplus_c t \otimes_c \left((-z_{\mathbb{D}i}) \oplus_c z_{\mathbb{D}j} \right), \; t \in [0, 1], \tag{12}$$

where $\oplus_c$ denotes the Möbius addition with aforementioned sectional curvature c, with details in supplementary material.

We adopt the following method to achieve generating via perturbation: For a given image x_i, we first rescale its embedding $z_{\mathbb{D}i}$ to the desired radius $r_{\mathbb{D}}$. Then we sample a random vector from seen categories in $z_{\mathbb{D}j}$ with radius $r_{\mathbb{D}}$ fixed and take the geodesic as the direction of perturbation to generate images.

4. Experiment

4.1. Implementation Details

In the training stage, we first train a StyleGAN2 [33] and pSp [49] with seen categories. Given a trained pSp, the MLP encoder $\texttt{MLP}_E$ is an 8-layer MLP with a Leaky-ReLU activation function. The dimension of the latent code in hyperbolic space is chosen to be 512. More details can be found in the supplementary.

4.2. Datasets

We evaluate our method on Animal Faces [40], Flowers [46], and VGGFaces [48] following the settings described in [15].

Animal Faces. We randomly select 119 categories as seen for training and leave 30 as unseen categories for testing.

Flowers. The Flowers [46] dataset is split into 85 seen categories for training and 17 unseen categories for testing.

VGGFaces. For VGGFaces [48], we randomly select 1802 categories for training and 572 for testing.

4.3. Analysis of Hierarchical Feature Editing

We analyze the properties of the learned hierarchical representations and how the levels of attributes relate to their locations of latent codes in hyperbolic space.

As we mentioned in Sec. 3.1, there is a continuum from fine-grained attributes to abstract attributes, corresponding to the points from the peripheral to the center of the ball. We define the hyperbolic radius $r_{\mathbb{D}}$[2] as the hyperbolic dis-

[2]The radius of the Poincaré disk in our experiment is about 6.2126

Method	Settings	Flowers		Animal Faces		VGG Faces*	
		FID(↓)	LPIPS(↑)	FID(↓)	LPIPS(↑)	FID(↓)	LPIPS(↑)
DAWSON [39]	3-shot	188.96	0.0583	208.68	0.0642	137.82	0.0769
MatchingGAN [28]	3-shot	143.35	0.1627	148.52	0.1514	118.62	0.1695
F2GAN [30]	3-shot	120.48	0.2172	117.74	0.1831	109.16	0.2125
LoFGAN [23]	3-shot	79.33	0.3862	112.81	0.4964	**20.31**	0.2869
DeltaGAN [27]	1-shot	109.78	0.3912	89.81	0.4418	80.12	0.3146
Disco-FUNIT [29]	1-shot	90.12	0.4436	71.44	0.4511	-	-
AGE [15]	1-shot	45.96	0.4305	28.04	0.5575	34.86	0.3294
SAGE [14]	1-shot	**43.52**	0.4392	27.43	0.5448	34.97	0.3232
HAE (Ours)	1-shot	50.10	**0.4739**	**26.33**	**0.5636**	35.93	**0.5919**

Table 1: FID(↓) and LPIPS(↑) of images generated by different methods for unseen categories on three datasets. **Bold** indicates the best results and underline indicates the second best results. VGGFaces is marked with * because different methods report different numbers of unseen categories on this dataset(*e.g.* 552 in LoFGAN, 96 in DeltaGAN, 497 in L2GAN, and 572 in AGE and SAGE). Note that: Disco-FUNIT [29] does not provide pre-trained models on VGG Faces [48] dataset.

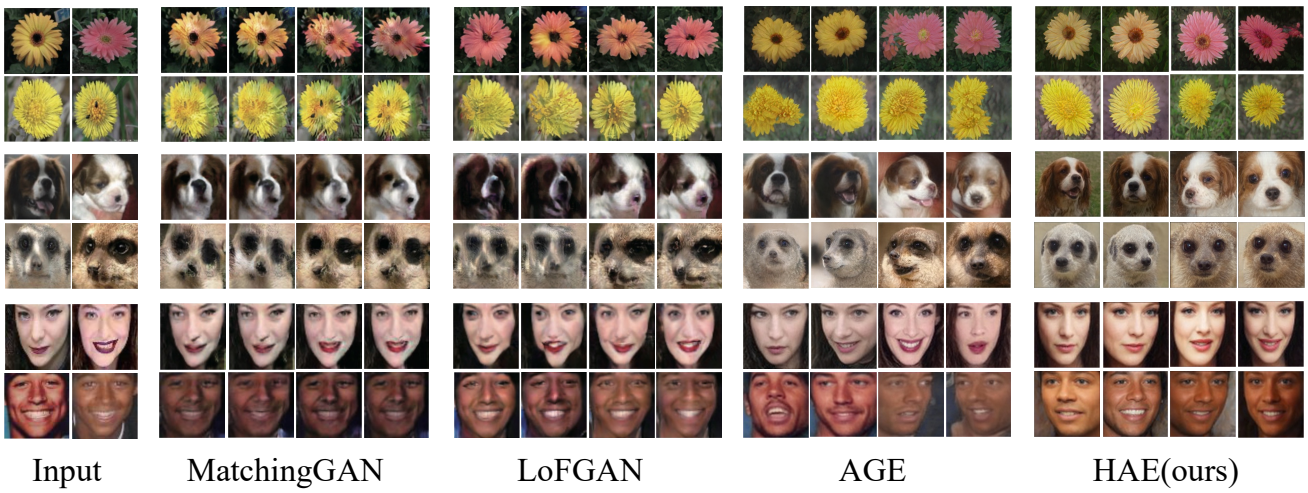

Figure 8: Comparison between images generated by MatchingGAN, LoFGAN, AGE, and HAE on Flowers, Animal Faces, and VGGFaces. Zoom in to see the details. Note that: SAGE [14] has not released code and pre-trained models.

tance of the given latent code to the center of the Poincaré disk. To study the influence of the radius of embeddings in hyperbolic space, we run several experiments with different choices of $r_{\mathbb{D}}$.

Hyperbolic Perturbation and Interpolation. As mentioned in Sec. 3.4, we demonstrate the results of perturbation and interpolation. In addition to the choice of perturbation, we can set the intensity of the perturbation by controlling both the step distance and radius as shown in Fig. 5. The results show that level of semantic attributes is highly related to $r_{\mathbb{D}}$. With the radius becoming smaller, the attributes become more abstract. We further visualize this property of hyperbolic space by moving the latent codes of the given image from the edge of the Poincaré disk to the center. As Fig. 7 shows, the images change from very fine-grained to very abstract (the average of all images). The results in Fig. 6 show that we can achieve smooth interpolation in hyperbolic space without any distortion. The results demonstrate that with HAE, we can freely control the editing geodesically and hierarchically.

4.4. Few-shot Image Generation

As Fig. 5 shows, the image categories will be changed when $r_{\mathbb{D}}$ is smaller than about 4, and the category-irrelevant attributes of images will be changed when $r_{\mathbb{D}}$ is larger than about 5. The embeddings of Animal Faces are visualized in 2-D Poincaré disk using UMAP [43] shown in Fig. 9. As Fig. 6 shows, the posture and the angle of the images will be changed at the early stage of interpolation without changing the category. Thus, the images can be generated by moving the latent code of a given image to some randomly selected semantic direction within the cluster of the category. In practice, we select $r_{\mathbb{D}} = 6.21$ and step size of perturbation as 8 to achieve few-shot image generation as

Figure 9: UMAP visualization of hyperbolic 2-D embeddings of Animal Faces dataset. We observe that similar categories are clustered and positioned near the boundary, while ambiguous samples are located near the center. Zoom in to see the details.

Fig. 4 shows the diverse images generated by adding random perturbations from seen categories. We conduct three experiments to show that HAE can achieve promising few-shot image generation. More examples of generated images are available in the supplementary.

Quantitative Comparison with State-of-the-art. We calculate the FID [24] and LPIPS [59] to evaluate the fidelity and diversity of the generated images following one-shot settings in [15, 14]. The comparison results are shown in Tab. 1. Our method achieves the best scores on most of the FID and LPIPS metrics compared with state-of-the-art few-shot image generation methods, which indicates that our method not only improves the model from the semantic aspect but also achieves state-of-the-art performance on the traditional evaluation metrics. Specifically, the LPIPS score of HAE beats SOTA model SAGE on all three datasets since HAE can generate more diverse images.

Qualitative Evaluation. We qualitatively compare our method with MatchingGAN [28], LoFGAN [23], DeltaGAN [27] and AGE [15]. As shown in Fig. 8, HAE can generate images with diversity and fine-grained details. More importantly, the newly generated images have more semantic diversity than others. For instance, the shadow and skin color of the generated faces change with the light condition, and this effect looks more natural. We further conduct a user study by randomly selecting 60 (20 from each dataset) images with generated variants using AGE and HAE. 50 users from different backgrounds are asked to rate the results only based on diversity and quality external information. This is achieved by randomly shuffling the order of images pairwisely and inside any pair. HAE won by a ratio of 58.1% $(1743/3000)$ over AGE (more details in supplementary).

Figure 10: Manipulate images from different categories with the same perturbation (Target 1&2).

Transferability. If we move latent codes at category-irrelevant levels, the target perturbation is transferable across all categories. We edit the images from three categories with the same editing direction, the output images are shown in Fig. 10. It demonstrates that HAE achieves a highly controllable and interpretable editing process.

Method	Flowers		Animal Faces		VGG Faces	
	FID	LPIPS	FID	LPIPS	FID	LPIPS
SAGE [14]	**43.52**	0.4392	27.43	0.5448	**34.97**	0.3232
HAE(Euc)	54.62	0.4293	**25.27**	0.5129	38.46	0.5908
HAE(Hyp)	50.10	**0.4739**	26.33	**0.5636**	35.93	**0.5919**

Table 2: FID(↓) and LPIPS(↑) of images generated by HAE in different geometries for unseen categories on three datasets. **Bold** indicates the best results and underline indicates the second best results.

4.5. Ablation Study

HAE in Euclidean. We re-trained HAE models in Euclidean space with the NLL loss to validate the performance gain in Tab. 1 is due to the hierarchical hyperbolic representation rather than the disentanglement caused by Eq. (7). The quantitative comparison is shown in Tab. 2. It shows that the hyperbolic space boosts the performance, especially for the LPIPS score, since the latent code is more disentangled in hyperbolic space [19]. This finding is also supported by the UMAP visualization in Fig. 9. More details can be found in the supplementary material.

Hyperbolic Radius versus Truncation. StyleGAN [33] uses truncation trick [42, 4, 33, 34] in $\mathcal{W}$-space to achieve the balance between the image quality and diversity. The experiments in [33, 34] also show that the truncation level in $\mathcal{W}^+$-space control the level of abstraction of the generated images. We conduct the experiments in Sec. 3.4 using truncation to validate the gains of hyperbolic space. The results are illustrated in Fig. 11 and Fig. 12. As Fig. 11 shows, the category of the image changes along with the posture of the dog as the truncation gets smaller, while the category-relevant attributes do not change when the hyperbolic radius

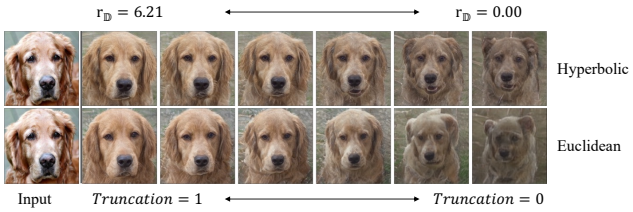

Figure 11: **Top**: Interpolations by moving the latent codes from the edge to the center in hyperbolic space. **Bottom**: Interpolation with different truncation in Euclidean space. Zoom in to see the details.

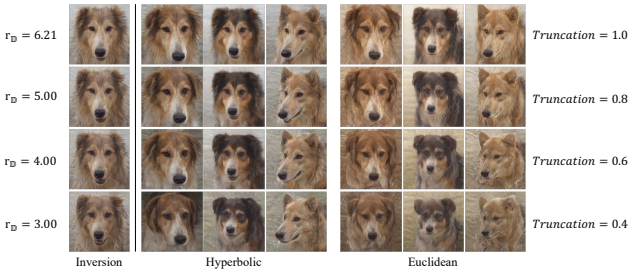

Figure 12: Images generated by HAE(Hyp) and HAE(Euc) by adding the same perturbation on the latent code of a given image with different settings of hyperbolic radius and truncation. Zoom in to see the details.

($r_{\mathbb{D}}$) is large. This can also be proved in Fig. 12. The category remains the same after adding perturbation when $r_{\mathbb{D}}$ is large, while the truncation can not control semantic-level editing. This shows that Euclidean space can only capture scale-based hierarchy rather than the semantic hierarchy.

Downstream Task. We conduct data augmentation via HAE for image classification on Animal Faces [48]. We randomly select 30, 35, and 35 images for each category as train, val, and test, respectively. Following [23], a ResNet-18 backbone is initialized from the seen categories. Then the model is fine-tuned on the unseen categories referred to as the *baseline*. 60 images are generated for each unseen category as data augmentation. The result is presented in Tab. 3. The diversity and quality of generated images are primarily controlled by the hyperbolic radii $r_{\mathbb{D}}$. As the radius becomes smaller, HAE generates images of higher diversity, but categories (referring to high-level attributes) also gradually change to others. $r_{\mathbb{D}} = 6$ achieves the best performance on the classification experiment. However, the performance drops when the radius is smaller than 4.5. This is because the semantic attributes change too much and thus mislead the classifier.

4.6. Limitations and Future Work

Although HAE achieves reliable hierarchical attribute editing in hyperbolic space for few-shot image generation, there are several limitations.

First, the boundary of category changing is hard to be quantified since the hierarchical levels are continuous in the hyperbolic space. Users need to find a "safe" boundary by trying different radii and step sizes of perturbation before generating new images. However, from another perspective, this continuity of hierarchy provides flexibility for users to set different boundaries for different downstream tasks as they need.

Hyperbolic Radius	Accuracy	FID(↓)	LPIPS(↑)
baseline	58.67	-	-
6.0	**60.10**	**46.89**	0.4520
5.5	59.52	48.68	0.4651
5.0	59.05	52.08	0.4823
4.5	59.14	60.87	0.5174
4.0	58.57	65.83	0.5386
3.5	56.86	68.44	0.6034
3.0	54.14	69.40	**0.6316**

Table 3: Ablation of same perturbation on different radii on Animal Faces.

Second, the performance of HAE is limited by the pretrained styleGAN and the inversion method. If the input image can not be well embedded, the editing will also fail. This problem can be solved by changing more powerful backbones, *e.g.*, ViT [16], in future work.

Finally, we use supervised learning to get the hierarchical embedding in hyperbolic space. However, the number of images in existing datasets with labels for generation tasks is relatively small, which makes the embeddings in the hyperbolic space not evenly distributed. The solution to this problem is simple, use unsupervised learning with large-scale high-quality datasets.

5. Conclusion

In this work, we propose a simple yet effective method HAE to edit hierarchical attributes in hyperbolic space. After learning the semantic hierarchy from images, our model is able to edit continuous semantic hierarchical features of images for flexible few-shot image generation in the hyperbolic space. Experiments demonstrate that HAE is capable of achieving not only stable few-shot image generation with state-of-the-art quality and diversity but a controllable and interpretable editing process. Future work includes the combination of HAE and large pretrained models and applications to more downstream tasks.

Acknowledgement. This work was supported in part by the National Key R&D Program of China under Grant 2018AAA0102000, in part by National Natural Science Foundation of China: 62022083 and 62236008.

References

[1] Antreas Antoniou, Amos Storkey, and Harrison Edwards. Data augmentation generative adversarial networks. *arXiv preprint arXiv:1711.04340*, 2017. 1, 2

[2] Sergey Bartunov and Dmitry P. Vetrov. Few-shot generative modelling with generative matching networks. In *AISTATS*, 2018. 1

[3] Silvère Bonnabel. Stochastic gradient descent on riemannian manifolds. *IEEE Transactions on Automatic Control*, 58(9):2217–2229, 2013. 3

[4] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. In *ICLR*, 2019. 8

[5] Gary Bécigneul and Octavian-Eugen Ganea. Riemannian adaptive optimization methods. In *ICLR*, 2019. 3

[6] James W Cannon, William J Floyd, Richard Kenyon, and Walter R Parry. Hyperbolic geometry. *Flavors of geometry*, 31:59–115, 1997. 4

[7] Ines Chami, Rex Ying, Christopher Ré, and Jure Leskovec. Hyperbolic graph convolutional neural networks. In *NeurIPS*, page 4868–4879, 2019. 3

[8] Anton Cherepkov, Andrey Voynov, and Artem Babenko. Navigating the gan parameter space for semantic image editing. In *CVPR*, pages 3670–3679, 2021. 3

[9] Jaewoong Choi, Junho Lee, Changyeon Yoon, Jung Ho Park, Geonho Hwang, and Myungjoo Kang. Do not escape from the manifold: Discovering the local coordinates on the latent space of gans. In *ICLR*, 2022. 3

[10] Louis Clouâtre and Marc Demers. Figr: Few-shot image generation with reptile. *arXiv:1901.02199*, 2019. 1, 2

[11] Jiali Cui, Ying Nian Wu, and Tian Han. Learning joint latent space ebm prior model for multi-layer generator. In *CVPR*, pages 3603–3612, 2023. 2

[12] Emily Denton, Ben Hutchinson, Margaret Mitchell, and Timnit Gebru. Detecting bias with generative counterfactual face attribute augmentation. *CoRR*, abs/1906.06439, 2019. 3

[13] Bhuwan Dhingra, Chris Shallue, Mohammad Norouzi, Andrew Dai, and George Dahl. Embedding text in hyperbolic spaces. *arXiv preprint arXiv:1806.04313*, 2018. 2

[14] Guanqi Ding, Xinzhe Han, Shuhui Wang, Xin Jin, Dandan Tu, and Qingming Huang. Stable attribute group editing for reliable few-shot image generation. *arXiv preprint arXiv:2302.00179*, 2023. 2, 7, 8

[15] Guanqi Ding, Xinzhe Han, Shuhui Wang, Shuzhe Wu, Xin Jin, Dandan Tu, and Qingming Huang. Attribute group editing for reliable few-shot image generation. In *CVPR*, pages 11184–11193, 2022. 1, 2, 6, 7, 8

[16] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*, 2021. 9

[17] Kun Fu, Tengfei Zhang, Yue Zhang, Menglong Yan, Zhonghan Chang, Zhengyuan Zhang, and Xian Sun. Meta-ssd: Towards fast adaptation for few-shot object detection with meta-learning. *IEEE Access*, 7:77597–77606, 2019. 1

[18] Octavian-Eugen Ganea, Gary Bécigneul, and Thomas Hofmann. Hyperbolic neural networks. In *NeurIPS*, pages 5345–5355, 2018. 3, 4, 5

[19] Songwei Ge, Shlok Mishra, Simon Kornblith, Chun-Liang Li, and David Jacobs. Hyperbolic contrastive learning for visual representations beyond objects. In *CVPR*, 2023. 8

[20] Lore Goetschalckx, Alex Andonian, Aude Oliva, and Phillip Isola. Ganalyze: Toward visual definitions of cognitive image properties. In *CVPR*, page 5744–5753, 2019. 3

[21] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *NeurIPS*, 2014. 1, 2

[22] Michael Gromov. Hyperbolic groups. In *Essays in group theory*, 1987. 3

[23] Zheng Gu, Wenbin Li, Jing Huo, Lei Wang, and Yang Gao. Lofgan: Fusing local representations for fewshot image generation. In *ICCV*, 2021. 1, 2, 7, 8, 9

[24] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *NeurIPS*, 2017. 8

[25] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *ICLR*, 2020. 2

[26] Yan Hong, Li Niu, Jianfu Zhang, Jing Liang, and Liqing Zhang. Deltagan: Towards diverse few-shot image generation with sample-specific delta. In *CVPR*, 2020. 1

[27] Yan Hong, Li Niu, Jianfu Zhang, Jing Liang, and Liqing Zhang. Deltagan: Towards diverse few-shot image generation with sample-specific delta. In *ECCV*, 2022. 1, 2, 7, 8

[28] Yan Hong, Li Niu, Jianfu Zhang, and Liqing Zhang. Matchinggan: Matching-based few-shot image generation. In *2020 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, 2020. 1, 2, 7, 8

[29] Yan Hong, Li Niu, Jianfu Zhang, and Liqing Zhang. Few-shot image generation using discrete content representation. In *Proceedings of the 30th ACM International Conference on Multimedia*, MM '22, page 2796–2804, New York, NY, USA, 2022. Association for Computing Machinery. 1, 2, 7

[30] Yan Hong, Li Niu, Jianfu Zhang, Weijie Zhao, Chen Fu, and Liqing Zhang. F2gan: Fusing-and-filling gan for few-shot image generation. In *Proceedings of the 28th ACM International Conference on Multimedia*, MM '20, page 2535–2543. Association for Computing Machinery, 2020. 1, 2, 7

[31] Erik Härkönen, Aaron Hertzmann, Jaakko Lehtinen, and Sylvain Paris. Ganspace: Discovering interpretable gan controls. In *NeurIPS*, 2020. 3

[32] Ali Jahanian, Lucy Chai, and Phillip Isola. On the" steerability" of generative adversarial networks. In *ICLR*, 2020. 3

[33] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *CVPR*, pages 4217–4228, 2019. 6, 8

[34] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *CVPR*, pages 8107–8116, 2020. 1, 4, 8

[35] Valentin Khrulkov, Leyla Mirvakhabova, Evgeniya Ustinova, Ivan Oseledets, and Victor Lempitsky. Hyperbolic image embeddings. In *CVPR*, pages 6417–6427, 2020. 2, 3, 4, 5

[36] Diego Lazcano, Nicolás Fredes Franco, and Werner Creixell. Hgan: Hyperbolic generative adversarial network. *IEEE Access*, 9:96309–96320, 2021. 3

[37] John M Lee. Riemannian manifolds: an introduction to curvature. *Springer Science & Business Media*, 176, 2006. 4

[38] John M Lee. *Introduction to Smooth Manifolds*. Springer, 2013. 4

[39] Weixin Liang, Zixuan Liu, and Can Liu. Dawson: A domain adaptive few shot generation framework. *arXiv preprint arXiv:2001.00576*, 2020. 1, 2, 7

[40] Ming-Yu Liu, Xun Huang, Arun Mallya, Tero Karras, Timo Aila, Jaakko Lehtinen, and Jan Kautz. Few-shot unsueprvised image-to-image translation. In *ICCV*, 2019. 6

[41] Yu-Ding Lu, Hsin-Ying Lee, Hung-Yu Tseng, and Ming-Hsuan Yang. Unsupervised discovery of disentangled manifolds in gans. *arXiv preprint arXiv:2011.11842*, 2020. 3

[42] Marco Marchesi. Megapixel size image creation using generative adversarial networks. *CoRR*, abs/1706.00082, 2017. 8

[43] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Grossberger. Umap: Uniform manifold approximation and projection. *The Journal of Open Source Software*, 3(29):861, 2018. 7

[44] Maximillian Nickel and Douwe Kiela. Generative visual manipulation on the natural image manifold. In *ECCV*, 2017. 2, 3, 4

[45] Maximillian Nickel and Douwe Kiela. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. In *ICML*, 2018. 3, 4

[46] Maria-Elena Nilsback and Andrew Zisserman. Automated flower classification over a large number of classes. In *2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing*, pages 722–729, 2008. 6

[47] Jiwoong Park, Junho Cho, Hyung Jin Chang, and Jin Young Choi. Unsupervised hyperbolic representation learning via message passing auto-encoders. In *CVPR*, pages 5512–5522, 2021. 3

[48] Omkar M. Parkhi, Andrea Vedaldi, and Andrew Zisserman. Deep face recognition. In *British Machine Vision Conference*, 2015. 6, 7, 9

[49] Elad Richardson, Yuval Alaluf, Or Patashnik, Yotam Nitzan, Yaniv Azar, Stav Shapiro, and Daniel Cohen-Or. Encoding in style: a stylegan encoder for image-to-image translation. In *CVPR*, pages 2287–2296, 2021. 4, 5, 6

[50] Yujun Shen, Jinjin Gu, Xiaoou Tang, and Bolei Zhou. Interpreting the latent space of gans for semantic face editing. In *CVPR*, pages 9240–9249, 2020. 3

[51] Yujun Shen and Bolei Zhou. Closed-form factorization of latent semantics in gans. In *CVPR*, pages 1532–1540, 2021. 3

[52] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *ICLR*, 2021. 3

[53] Nurit Spingarn-Eliezer, Ron Banner, and Tomer Michaeli. Gan "steerability" without optimization. In *ICLR*, 2021. 3

[54] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip H.S. Torr, and Timothy M. Hospedales. Learning to compare: Relation network for few-shot learning. In *CVPR*, pages 1199–1208, 2018. 1

[55] Dídac Surís, Ruoshi Liu, and Carl Vondrick. Learning the predictability of the future. In *CVPR*, pages 12602–12612, 2021. 2, 3, 4

[56] Alexandru Tifrea, Gary Bécigneul, and OctavianEugen Ganea. Poincaré glove: Hyperbolic word embeddings. In *ICLR*, 2019. 2, 3, 4

[57] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Koray Kavukcuoglu, and Daan Wierstra. Matching networks for one shot learning. In *NeurIPS*, 2016. 1

[58] Andrey Voynov and Artem Babenko. Unsupervised discovery of interpretable directions in the gan latent space. In *ICML*, pages 9786–9796, 2020. 3

[59] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, pages 586–595, 2018. 8

[60] Jingyuan Zhu, Huimin Ma, Jiansheng Chen, and Jian Yuan. Few-shot image generation with diffusion models. *arXiv preprint arXiv:2211.03264*, 2022. 2