Deep Curvilinear Editing: Commutative and Nonlinear Image Manipulation for Pretrained Deep Generative Model

Takehiro Aoshima, Takashi Matsubara
Osaka University
1-3 Machikaneyama, Toyonaka, Osaka, 560-8531 Japan.
aoshima@hopf.sys.es.osaka-u.ac.jp, matsubara@sys.es.osaka-u.ac.jp

Abstract

Semantic editing of images is the fundamental goal of computer vision. Although deep learning methods, such as generative adversarial networks (GANs), are capable of producing high-quality images, they often do not have an inherent way of editing generated images semantically. Recent studies have investigated a way of manipulating the latent variable to determine the images to be generated. However, methods that assume linear semantic arithmetic have certain limitations in terms of the quality of image editing, whereas methods that discover nonlinear semantic pathways provide non-commutative editing, which is inconsistent when applied in different orders. This study proposes a novel method called deep curvilinear editing (DeCurvEd) to determine semantic commuting vector fields on the latent space. We theoretically demonstrate that owing to commutativity, the editing of multiple attributes depends only on the quantities and not on the order. Furthermore, we experimentally demonstrate that compared to previous methods, the nonlinear and commutative nature of DeCurvEd facilitates the disentanglement of image attributes and provides higher-quality editing.

1. Introduction

Generating and editing realistic images is one of the fundamental goals in computer vision. Generative adversarial networks (GANs) [20] have emerged as a major image generation approach owing to the quality of generated images [7, 33–36, 48] and provide various real-world applications [1, 13, 18, 53, 63, 68]. Other notable methods include variational autoencoders (VAEs) [24], conditional PixelCNN [61], and diffusion-based models [26, 56], which are collectively called deep generative models. However, deep generative models cannot inherently edit images semantically. They can be viewed as mappings from latent space to image space, and the latent variables determine the generated images. Therefore, several methods have been developed to train deep generative models such that each semantic attribute the user wants to edit is assigned to each element of the latent variable [10, 45, 46] (see the first column of Table 1). However, this approach requires computationally expensive training and can conflict with the quality of image generation. Other studies have developed image-to-image translation that translates images from one domain to another [28, 64, 68]. However, this approach also requires training from scratch and limits image editing to be discontinuous unless combined with latent variable manipulation. Therefore, a general and promising approach is necessary to identify manipulations on latent variables of already trained models that edit images semantically.

A study reported that adding certain vectors to latent variables can modify the corresponding attributes of the object in the generated images [51]. This indicates that the latent space can be regarded as a linear vector space. Some studies have aimed to identify attribute vectors in a supervised or an unsupervised manner [17, 19, 22, 27, 49, 50, 54, 55, 62, 69]. In any case, these studies introduce the strong assumption of linear semantic arithmetic on the latent space (see the second column of Table 1), which limits the quality of image editing. Other studies have proposed methods to determine attribute vectors depending on the position in the latent space, that is, the attribute vector fields (or local attribute coordinates). [2, 11, 12, 17, 29, 37, 44, 52, 58, 60]; these methods edit an image attribute by moving the latent variable nonlinearly along the corresponding attribute vector field. Although this approach is elegant, the edits of different attributes are generally non-commutative (see the third column of Table 1). That is, what we get is different when we edit attributes (denoted by $e_1$ and $e_2$) one after another, or in the reverse order. This property can harm the disentanglement between attributes considering that the relationships among attributes change at different points. In contrast, linear arithmetic ensures that the edits of different attributes are commutative.

To overcome this dilemma, this study proposes **deep**...
Table 1. Comparison of Our Proposal against Related Methods.

<table>
<thead>
<tr>
<th></th>
<th>Training under constraints</th>
<th>Linear arithmetic</th>
<th>Vector fields/Local basis</th>
<th>DeCurvEd (proposed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global coordinate</td>
<td>Cartesian</td>
<td>oblique</td>
<td>(only local)</td>
<td>curvilinear</td>
</tr>
<tr>
<td>No retraining</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Nonlinear edit</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Commutative edit</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
</tr>
</tbody>
</table>

**Conceptual diagram**

---

curvilinear editing (DeCurvEd), a method that determines a set of commuting attribute vector fields in the latent space of a pre-trained deep generative model. The key idea is adopted from the theorem that a set of vector fields is locally expressed as partial derivatives of a coordinate chart if it is linearly independent and commuting [43]. Therefore, we define a curvilinear coordinate system globally [3] by employing a normalizing flow [9, 39], from which we derive commuting vector fields (see the rightmost panel of Table 1). The advantages of DeCurvEd are as follows (see also Table 1):

1. Edits of different attributes are always commutative, unlike previous methods that assume attribute vector fields (e.g., [60]). Therefore, an edited image does not depend on the order of editing, but on the amount of editing performed.
2. Edits are nonlinear, which indicates that DeCurvEd provides better editing quality than methods that assume linear arithmetic of attribute vectors (e.g., [62]).
3. DeCurvEd does not require retraining of deep generative models, but identifies attribute vector fields in the latent space of pre-trained models, unlike image-to-image translation and training under constraints (e.g., [10, 28]).
4. We propose CurvilinearGANSpace by combining DeCurvEd with GANs and experimentally demonstrate that the nonlinear and commutative nature disentangles attributes and enables high-quality editing.
5. The key idea is not limited to GANs, and is available for any generative models that are conditioned on latent variables, including VAEs [24], conditional PixelCNN [61], and diffusion-based models [26, 56].

2. Related Work

**Image Editing by Deep Generative Models** Most generative models define a mapping that maps a latent variable to a data sample (that is, an image in this study). Previous studies on deep learning-based generative models have confirmed that manipulating the latent variable can determine the image to be generated [20, 24]. Radford et al. [51] identified a semantically meaningful vector in the latent space based on the difference between two groups. Then, an attribute can be imposed on an image by adding the attribute vector to the latent variable of the image. This discovery attracted wide attention and spurred research into semantic image editing by manipulating the latent variables.

Several studies have developed methods for training models under constraints to easily determine attribute vectors, rather than identifying attribute vectors after training [10, 25, 45, 46]. These methods often make each element of the latent variable as independent as possible [25]. Then, one element is assigned to a group (an attribute) that changes collectively and is independent of other groups in the image [6]. These methods can be viewed as introducing a Cartesian coordinate system to the latent space and assigning one attribute to each axis. However, constraints often conflict with other training criteria (such as likelihood and Gaussian prior) and result in models with inferior quality and diversity. Additionally, the generated models needed to be trained from scratch, which incurs computation costs. See the column “Training under constraints” in Table 1.

Moreover, some studies have developed image-to-image transformation, which maps an image from one domain to another rather than manipulating a latent variable [28, 64, 68]. This approach limits image editing only between domains or needs to be combined with a latent space.

**Discovering Linear Attribute Arithmetic** Several studies have investigated linear manipulations on latent variables in the already trained deep generative models [27, 62]. SeFa and related methods attempted to find semantic directions by analyzing the weight parameters [55, 66, 67]. Voynov and Babenko [62] proposed an unsupervised framework that learns a set of semantic directions. According
to this framework, changing a latent variable along a semantic direction will edit one attribute of the corresponding image, and the degree of change in the attribute will be proportional to the amount of change in the latent variable. GANSpace [27] applied a principal component analysis (PCA) to extract a lower-dimensional subspace of the latent space, assuming each principal component corresponds to an attribute. These methods assume linear arithmetic of the attribute vectors; that is, they introduce an oblique coordinate system to the latent space. See the column “Linear arithmetic” in Table 1.

Because the distribution of real-world data is often biased and skewed, it is unlikely that the latent space is flat and homogeneous. Khrulkov et al. [37] found that different directions correspond to the same attribute at different locations in the latent space. Therefore, the above methods are limited in terms of image editing quality.

Discovering Semantic Vector Fields The direction corresponding to an attribute varies depending on the location in the latent space, thereby indicating that a set of directions corresponding to attributes forms a set of vector fields, rather than linear arithmetic. If so, one can edit an attribute of an image by moving the latent variable nonlinearly along the vector field corresponding to the attribute instead of adding an attribute vector. Tzelepis et al. [60] proposed WarpedGANSpace, which learns a set of vector fields, each of which is defined as a gradient flow of an RBF function on the latent space. Choi et al. [12] learned a local basis at every point of the latent space such that each element of the local basis corresponds to an attribute. StyleFlow [2] and SSFlow [44] used normalizing flows to define a local coordinate system. On an N-dimensional manifold, a local basis, local coordinate system, and a set of N linearly independent vector fields are compatible; such vector fields are called coordinate vector fields (see Example 8.2, [43]). However, because the coordinate system in the above studies is defined only locally, multiple edits may be inconsistent globally. We will demonstrate this in the following section. See the column “Vector fields/Local basis” in Table 1.

Some studies have attempted to define a (Riemannian) metric on the latent space [4, 5, 8]. These methods successfully interpolate between two images by nonlinearly moving latent variables along the geodesic; however, they are insufficient for attribute editing. Some others attempted complex and dynamic editing specified by text rather than attributes [59]; nevertheless, such methods cannot be directly compared to ours.

3. Theoretical Background

We introduce the theoretical background of the proposed and related methods introduced in Section 2. Theorems in this paper are basic knowledge about manifolds; readers unfamiliar with this topic are referred to the reference [43]. Remarks are our findings.

Let \( \mathcal{X} \) and \( \mathcal{Z} \) denote an image space and a latent space of a deep generative model, respectively. The generator \( G \) (also called decoder) of the deep generative model is a mapping from the latent space \( \mathcal{Z} \) to the image space \( \mathcal{X} \); given a latent variable \( z \in \mathcal{Z} \), the generator produces an image \( x = G(z) \). We assume the latent space \( \mathcal{Z} \) to be an N-dimensional space diffeomorphic to a Euclidean space.

Let \( \{z^i\}^N_{i=1} \) denote the coordinate system (i.e., the basis) on a neighborhood of the point \( z \in \mathcal{Z} \). Let \( Z_k \) denote a vector field on the latent space \( \mathcal{Z} \) indexed by \( k \), that is, \( Z_k : \mathcal{Z} \rightarrow \mathcal{T}_z \mathcal{Z} \), where \( \mathcal{T}_z \mathcal{Z} \) is the tangent space (i.e., the space of tangent vectors or velocities) of the latent space \( \mathcal{Z} \) at point \( z \). At point \( z \), the coordinate system on tangent space \( \mathcal{T}_z \mathcal{Z} \) is denoted by \( \{\frac{\partial}{\partial z_i}\}^N_{i=1} \), and a vector field \( Z_k \) is expressed as \( Z_k(z) = \sum_{i=1}^N Z^i_k(z) \frac{\partial}{\partial z_i} \) for smooth functions \( Z^i_k : \mathcal{Z} \rightarrow \mathbb{R} \). A method that assumes attribute vector fields [12, 52, 60] edits an attribute \( k \) of an image \( x \) by integrating a latent variable \( z \) along the vector field \( Z_k \) that corresponds to attribute \( k \); the edited image is \( x' = G(z') \) for

\[
\begin{align*}
z' &= z + \int_0^t Z_k(z(\tau)))d\tau, \\
\end{align*}
\]

where \( z(0) = z \), and \( t \in \mathbb{R} \) denotes the change amount of attribute \( k \). \( t \) may be positive or negative. We rewrite the above equation using a flow, denoted by \( \phi^k_t : \mathcal{Z} \rightarrow \mathcal{Z} \) for \( t \) as:

\[
\begin{align*}
z' &= \phi^k_t(z). \\
\end{align*}
\]

Then, we define the commutativity of editing as follows:

**Definition 1 (Commutativity).** Edits of attributes \( k \) and \( l \) are commutative if and only if the corresponding flows \( \phi^k_t \) and \( \phi^l_t \) are commuting, that is, it holds that \( \phi^l_s \circ \phi^k_t = \phi^k_t \circ \phi^l_s \) for any \( s, t \in \mathbb{R} \) at any point \( z \) on the latent space \( \mathcal{Z} \).

Intuitively, making a person smile and then wear sunglasses results in the same image as making the person wear sunglasses and then smile if the vector fields corresponding to smiling and wearing sunglasses are commuting. Else, edits in different orders produce different images.

**Remark 1.** A method that assumes linear attribute arithmetic (e.g., [27, 55, 62]) is a special case of a method that assumes attribute vector fields, and its edits are commutative.

See Appendix A for formal proofs of any remarks in this manuscript. Therefore, we can discuss a method that assumes linear attribute arithmetic in the same context. We introduce the following theorem.

**Theorem 1 (Commuting Vector Fields, Theorem 9.44, [43]).** Two flows \( \phi_k \) and \( \phi_l \) are commuting if and only if the underlying vector fields \( Z_k \) and \( Z_l \) are commuting.
This theorem suggests the following remark:

**Remark 2.** In general, edits by a method that assumes attribute vector fields (e.g., [12, 52, 60]) are non-commutative.

In addition, we introduce the following theorem.

**Theorem 2** (Canonical Forms for Commuting Vector Fields, Theorem 9.46, [43]). Let vector fields $Z_1, Z_2, \ldots, Z_N$ on an $N$-dimensional space $Z$ be linearly independent and commuting on an open set $U \subset Z$. At each $z \in U$, there exists a smooth coordinate chart $\{ \frac{\partial}{\partial z^i} \}^N_{i=1}$ centered at $z$ such that $\frac{\partial}{\partial z^i} = Z_i$.

Furthermore, given a smooth coordinate chart $\{ \frac{\partial}{\partial z^i} \}^N_{i=1}$, vector fields $Z_i = \frac{\partial}{\partial z^i}$ are commuting. A coordinate chart is a nonlinear bijective mapping to Euclidean space. Therefore, intuitively, a set of $N$ linearly independent and commuting vector fields on an $N$-dimensional space $Z$ is equivalent to a set of $N$ vector fields along the axes of a coordinate system up to geometric transformation.

**4. Method**

**4.1. DeCurvEd**

Given the theoretical background, we propose DeCurvEd, as shown in Fig. 1. Intuitively, we consider the case where the open set $U$ in Theorem 2 is not a proper subset but equal to the latent space $Z$.

We prepare an $N$-dimensional Euclidean space $V$ and call it the Cartesianized latent space, whose coordinate system $\{ v^i \}^N_{i=1}$ is a global Cartesian coordinate system. Let $e_k$ denote the $k$-th element of the standard basis of the tangent space, that is, $e_k := \frac{\partial}{\partial v^k}$. Then, the vector field $Z_k$ corresponding to attribute $k$ is defined as

$$
\tilde{Z}_k := e_k.
$$

As discussed in the previous section, vector fields $\tilde{Z}_k$ and $\tilde{Z}_i$ defined in this way are commuting for any $k$ and $l$. The flow $\psi^k_l : V \to V$ that arises from the vector field $\tilde{Z}_k$ is given by

$$
\psi^k_l(v) := v + \int_0^l e_k \mathrm{d}\tau = v + t e_k.
$$

The flows $\psi^k_l$ are commuting because $(\psi^k_l \circ \psi^l_l)(v) = v + t e_k + s e_l = (\psi^l_l \circ \psi^k_l)(v)$. We introduce a smooth bijective mapping $f : Z \to V, z \mapsto v$ that corresponds to the coordinate chart in Theorem 2. The mapping $f$ can be implemented using a normalizing flow; however, it is not limited to [9, 21, 39]. We define a flow $\phi^k$ that edits attribute $k$ on the latent space $Z$ as:

$$
\phi^k := f^{-1} \circ \psi^k_l \circ f.
$$

See also the left half of Fig. 1. We redefine the edit as Algorithm 1 in Appendix B.

Subsequently, one can generate an edited image $x' = G(z')$ using generator $G$. Deep generative models such as GANs do not have an inherent way of inferring a latent variable $z$ from an image $x$; this is outside the scope of this study. Interested readers can refer to this survey [65].

**4.2. Theoretical Analysis**

The pushforward $f_*$ is a mapping naturally induced by the mapping $f$, which maps a tangent vector (or a basis) on the latent space $Z$ to that on the Cartesianized latent space $V$. Also, the pushforward $(f^{-1})_*$ maps the Cartesian coordinate system on the Cartesianized latent space $V$ and implicitly defines a coordinate system on the latent space $Z$ [43]. A coordinate system defined by a bijective transformation of a Cartesian coordinate is called a curvilinear coordinate [3]. Therefore, we name this method deep curvilinear editing (DeCurvEd). Because the mapping $f$ is defined globally between spaces $Z$ and $V$, the curvilinear coordinate system is also defined globally. The pushforward $(f^{-1})_*$ can define commuting vector fields on $Z$ by push-forwarding the coordinate vector fields on $V$. Therefore, we make the following remarks.

**Remark 3.** Using DeCurvEd, any edits of attributes in the latent space $Z$ can be nonlinear and commutative.

**Remark 4.** DeCurvEd can define vector fields on the latent space $Z$ and is a special case of a method that assumes attribute vector fields (e.g., [12, 52, 60]).
**Remark 5.** A method that assumes linear attribute arithmetic (e.g., \[27, 55, 62\]) is a special case of DeCurvEd, with a linear mapping \(f\).

Therefore, DeCurvEd enjoys the advantages of both attribute arithmetic and vector fields. All theories and remarks are not dependent on the properties of particular models. Thus, we make the following remark.

**Remark 6.** DeCurvEd offers attribute editing for any generative models conditioned on latent variables, including GANs \[20\], VAEs (see Fig. 4 of \[24\]), conditional PixelCNN \[61\], and diffusion-based models (see Fig. 8 of \[26\] and Fig. 4 of \[56\]).

### 4.3. CurvilinearGANSpace

Attribute editing by DeCurvEd is available for any deep generative models and for both supervised and unsupervised learning. This study adopted the unsupervised training framework for GANs proposed by Voynov and Babenko \[62\], as shown in the right half of Fig. 1. Following previous studies, we call it CurvilinearGANSpace.

Given a latent variable \(z\), CurvilinearGANSpace randomly edits index \(k\) by \(\epsilon\) and produces an edited one \(z'\). In some cases, only the first \(N'\) indices of all \(N\) indices are candidates for editing. We prepare a neural network called reconstructor \(R\), which accepts the pair of generated images \(x = G(z)\) and \(x' = G(z)\) and regresses the edited index \(k\) and the change amount \(\epsilon\). In particular, one output is an \(N'\)-dimensional vector \(\hat{k}\) to regress the edited index \(k\): the loss function is the classification error \(L_{\text{cls}}(k; \hat{k})\), which is defined as the cross-entropy. As the mapping \(f\) minimizes this error, image editing of index \(k\) becomes easier for the reconstructor \(R\) to distinguish from image editing of other indices \(l \neq k\), thereby assigning one attribute to each vector field and facilitating the disentanglement between attributes. The other output \(\hat{\epsilon}\) is a scalar regressing the change amount \(\epsilon\); the loss function is the regression error \(L_{\text{reg}}(\epsilon; \hat{\epsilon})\) defined as the absolute error. As this error is minimized, the change in the latent variable continuously matches the semantic change in the image.

Additionally, we introduce a regularization term \(L_{n1}\) to be minimized for the mapping \(f\):

\[
L_{n1}(z) = (\log \det |\frac{\partial f}{\partial z}|)^2. \tag{6}
\]

The Jacobian determinant \(\det |\frac{\partial f}{\partial z}|\) of the mapping \(f\) indicates the extent to which the latent space \(\mathcal{Z}\) is stretched by the mapping \(f\); when it is 1.0, the mapping \(f\) is isometric. Subsequently, this term \(L_{n1}\) avoids extreme deformation of the latent space \(\mathcal{Z}\) by the mapping \(f\). The final objective function is defined as:

\[
\mathbb{E}_{z,k,\epsilon} \left[ L_{\text{cls}}(k; \hat{k}) + \lambda L_{\text{reg}}(\epsilon; \hat{\epsilon}) + \alpha L_{n1}(z) \right], \tag{7}
\]

where \(\lambda, \alpha \in \mathbb{R}\) are hyperparameters weighing objectives. See also Algorithm 2 in Appendix B for more details.

### 5. Experiments

#### 5.1. Experimental Setting

**Datasets, Backbones, and Comparison Methods** We examined CurvilinearGANSpace and related methods using combinations of datasets, GANs, and reconstructors, as summarized in Table 2. \(N\) denotes the number of dimensions of the latent space \(\mathcal{Z}\), and \(N'\) denotes the number of dimensions used for training. For StyleGAN2, \(W\) space was used as the latent space. For ILSVRC and CelebA-HQ, we used pre-trained models from their official repositories. These experimental settings are identical to those in previous studies \[60, 62\]. See Appendix C.1 and the references \[60, 62\] for more details.

For comparison, we used a method that assumes linear arithmetic \[62\] and a method that assumes attribute vector fields called WarpedGANSpace \[60\]. To clarify the difference, we hereafter refer to the former method as LinearGANSpace. We used their pre-trained models for all but the LSUN Car dataset and used our own trained models for the LSUN Car dataset, each trained in the same framework.

#### Architectures and Hyperparameters

As the bijective mapping \(f\), we used a continuous normalizing flow with six concat-squash layers \[21\]. We set the number of hidden units equal to the input dimension and used hyperbolic tangent function as its activation function. See Appendix C.2 for more introduction. We used Adam optimizer \[38\] with a constant learning rate of \(10^{-4}\). We used \(\lambda = 0.25\), which is equivalent to that used by previous studies \[60, 62\]. For simplicity, we used \(\alpha = 1\).

#### Evaluation Metrics

For CelebA-HQ, we measured the attribute scores of generated images using separate pre-trained attribute predictors \(A_k : \mathcal{X} \rightarrow \mathbb{R}^5\). FairFace measured age, gender, and race (skin color) attributes \[32\], and CelebA-HQ attributes classifier measured smile, beard, and bangs attributes \[30\] from 0 to 1; Hopenet measured face

---

1. [https://github.com/anvoynov/GANLatentDiscovery](https://github.com/anvoynov/GANLatentDiscovery) for LinearGANSpace.
2. [https://github.com/chi0tsp/WarpedGANSpace](https://github.com/chi0tsp/WarpedGANSpace) for WarpedGANSpace and attribute predictors.
directions, yaw and pitch, in degree [16]. Additionally, ArcFace measured the identity score $I(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \rightarrow [0, 1]$ to evaluate whether two images are of the same person [15]. We selected index $k$ as the one corresponding to that attribute if the measured attribute score has the largest covariance with the change amounts of index $k$. The above procedure is identical to that of the previous study [60].

The amount $t$ by which attribute $k$ of latent variable $z$ is edited differs from the amount by which the corresponding attribute score $A_k$ of the generated image $x = G(x)$ is changed. For a fair comparison, we normalized the change amount $t$ such that the measured attribute score changes by 5 degrees for the pitch and yaw attributes and 0.1 for others, and denoted the amount by $t = 0.1$.

After index identification and normalization, we used several evaluation metrics. We defined commutativity error of attributes $k+l$ to evaluate how commutative the image editing is by measuring the difference in the attribute score between images with edits of attributes $k$ and $l$ applied in different orders. We defined side effect error to evaluate the disentanglement between attributes by measuring how much an edit of the target attribute $k$ changes the other attributes $l \neq k$ as undesired side effects. We also defined identity error to evaluate the disentanglement by measuring how much an edit of the target attribute $k$ reduces the identity score. The errors in edits of multiple attributes are defined similarly.

Owing to the availability of attribute predictors, these evaluations were performed only for CelebA-HQ. See Appendix C.3 for the detailed procedures and definition. For other datasets, we manually selected the index $k$, following previous studies [60,62].

### Table 3. Commutativity Errors [%] of StyleGAN2.

<table>
<thead>
<tr>
<th></th>
<th>$A+G$</th>
<th>$R+P$</th>
<th>$B+Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LinearGANSpace</td>
<td>0.01 / 0.05</td>
<td>0.02 / 0.07</td>
<td>0.02 / 0.15</td>
</tr>
<tr>
<td>WarpedGANSpace</td>
<td>11.40 / 6.62</td>
<td>3.15 / 3.46</td>
<td>1.28 / 2.22</td>
</tr>
<tr>
<td>CurvilinearGANSpace (ours)</td>
<td>0.07 / 0.35</td>
<td>0.05 / 0.62</td>
<td>0.08 / 0.55</td>
</tr>
</tbody>
</table>


### 5.2. Experimental Results

**Commutativity of Editing** Table 3 shows the commutativity errors for CelebA-HQ+StyleGAN2 with $t = 0.1$. Those of LinearGANSpace and CurvilinearGANSpace were always less than 0.7 %; even though they were not exactly zero due to numerical and rounding errors, the errors were negligible. The errors of WarpedGANSpace were between 1.2 % and 11.4 %. Therefore, as expected, the image editing by WarpedGANSpace is non-commutative, whereas that by LinearGANSpace and CurvilinearGANSpace is commutative.

We edited image attributes sequentially so that the total amount of change is zero and summarized the results in Fig. 2. The images generated after sequential editing by LinearGANSpace or CurvilinearGANSpace look identical to the originals, which indicates that their image editing is commutative. When editing a human face by WarpedGANSpace, the face’s yaw rotation and image brightness were not restored. Also for an AnimeFaces image, the hair color was not restored. These results indicate that image editing by WarpedGANSpace is non-commutative.

A closer look at each edit reveals that the editing of a human face by LinearGANSpace does not properly edit the smile attribute, and the edit of yaw rotation changes the hairstyle as well. When WarpedGANSpace edits the pitch or yaw rotation of the human face, it changes hair color, skin color, and brightness as well. For the AnimeFaces image, the editing by LinearGANSpace is of inferior quality. WarpedGANSpace’s edits of the hair color and hair length change the face (i.e., identity). CurvilinearGANSpace’s editing is of excellent quality without severe side effects.

These results indicate that CurvilinearGANSpace provides commutative editing and significantly improves the disentanglement between attributes. The training framework used [62] leads the editing methods to learn disentanglement between attributes by classifying indices. WarpedGANSpace takes advantage of nonlinearity to allow better editing; however, there is no mechanism to further facilitate disentanglement. CurvilinearGANSpace assumes...
that attribute vector fields are locally linearly independent, and hence, always assigns a different direction to each attribute, which facilitates the disentanglement. We describe disentanglement in the following section.

**Disentanglement of Attributes**  Table 4 shows the side effect errors. We defined a “severe side effect” as a change in another attribute by 0.09 or more when editing a target attribute by \( t = 0.1 \), and highlighted it in bold red. All three editing methods confounded the age and bangs attributes (i.e., editing one impacted the other) owing to their high correlation, caused by unsupervised learning. CurvilinearGANSpace has no other severe side effects, whereas LinearGANSpace and WarpedGANSpace have many severe side effects; for example, WarpedGANSpace’s edit of the race attribute rather changes the bangs attribute.

Table 5 shows the identity errors. CurvilinearGANSpace produced the lowest errors for two of the six attributes, the second lowest errors for the remaining, and the lowest average error. LinearGANSpace and WarpedGANSpace produced severe identity errors in some cases.

These results indicate that only CurvilinearGANSpace selectively edits target attributes and preserves as much other information as possible, that is, it is of excellent quality for the disentanglement of attributes.

**Visualization of Disentanglement**  We summarized the identity errors and generated images when editing attributes in Fig. 3. The larger the attribute editing, the greater the increase in identity error. However, CurvilinearGANSpace has the lowest identity errors. When WarpedGANSpace edited the age attribute of a StyleGAN2 image, it also altered the skin color, hair color, and facial expression. LinearGANSpace also lost the identity. WarpedGANSpace’s edit of the smile attribute of a ProgGAN image altered the pitch and yaw rotations. CurvilinearGANSpace can edit only specific attributes with a smaller loss of identity.

We summarized the results of other models when editing specific attributes in Fig. 4. The image editing by CurvilinearGANSpace was as intended with the least side effects; however, the image editing by LinearGANSpace

---

**Table 4. Side Effect Errors [%] of StyleGAN2.**

<table>
<thead>
<tr>
<th>Target</th>
<th>A</th>
<th>G</th>
<th>R</th>
<th>B</th>
<th>P</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>LinearGANSpace [62]</td>
<td>100</td>
<td>59</td>
<td>37</td>
<td>63</td>
<td>41</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>52</td>
<td>71</td>
<td>24</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>71</td>
<td>90</td>
<td>43</td>
<td>76</td>
<td>100</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>58</td>
<td>55</td>
<td>43</td>
<td>94</td>
<td>36</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>51</td>
<td>63</td>
<td>111</td>
<td>59</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>100</td>
<td>94</td>
<td>124</td>
<td>236</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>63</td>
<td>64</td>
<td>100</td>
<td>131</td>
<td>73</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>41</td>
<td>44</td>
<td>30</td>
<td>80</td>
<td>100</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>30</td>
<td>22</td>
<td>97</td>
<td>23</td>
<td>100</td>
</tr>
<tr>
<td>WarpedGANSpace [60]</td>
<td>100</td>
<td>51</td>
<td>63</td>
<td>111</td>
<td>59</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>100</td>
<td>94</td>
<td>124</td>
<td>236</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>63</td>
<td>64</td>
<td>100</td>
<td>131</td>
<td>73</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>41</td>
<td>44</td>
<td>30</td>
<td>80</td>
<td>100</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>30</td>
<td>22</td>
<td>97</td>
<td>23</td>
<td>100</td>
</tr>
<tr>
<td>CurvilinearGANSpace (ours)</td>
<td>100</td>
<td>51</td>
<td>63</td>
<td>111</td>
<td>59</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>100</td>
<td>94</td>
<td>124</td>
<td>236</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>63</td>
<td>64</td>
<td>100</td>
<td>131</td>
<td>73</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>41</td>
<td>44</td>
<td>30</td>
<td>80</td>
<td>100</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>30</td>
<td>22</td>
<td>97</td>
<td>23</td>
<td>100</td>
</tr>
</tbody>
</table>


**Table 5. Identity Errors [%] of StyleGAN2.**

<table>
<thead>
<tr>
<th>A</th>
<th>G</th>
<th>R</th>
<th>B</th>
<th>P</th>
<th>Y</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LinearGANSpace [62]</td>
<td>26.1</td>
<td>5.5</td>
<td>19.1</td>
<td>47.4</td>
<td>26.4</td>
<td>24.7</td>
</tr>
<tr>
<td>WarpedGANSpace [60]</td>
<td>27.6</td>
<td>56.2</td>
<td>33.6</td>
<td>6.3</td>
<td>84</td>
<td>8.4</td>
</tr>
<tr>
<td>CurvilinearGANSpace (ours)</td>
<td>21.1</td>
<td>15.4</td>
<td>25.3</td>
<td>6.0</td>
<td>18.9</td>
<td>9.6</td>
</tr>
</tbody>
</table>

Linear
Warped
Ours
(a) MNIST, "width".
(b) MNIST, "thickness".
(c) AnimeFaces, "hair color".
(d) AnimeFaces, "hair length".
(e) ILSVRC, "object size".
(f) ILSVRC, "vertical position".
(g) ProgGAN, "yaw".
(h) ProgGAN, "smile".
(i) LSUN Car, "color".
(j) LSUN Car, "rotation".

Figure 4. Visualization results. The models and edited attributes $k$ are shown below the panels. The image in the center is the original, the images on the right have attributes added, and the images on the left have attributes subtracted, as is the case with Fig. 3.

or WarpedGANSpace exhibited severe side effects. Panels (a) and (b) show the edits of the width of digit 0 and the thickness of digit 6 in the MNIST dataset, respectively. LinearGANSpace and WarpedGANSpace additionally rotated the digits, while CurvilinearGANSpace maintained the original direction. Panels (c) and (d) show the results of the AnimeFaces dataset. The edit of the hair color by LinearGANSpace or WarpedGANSpace altered the face (i.e., loses the identity). When LinearGANSpace and WarpedGANSpace lengthened the hair, they paradoxically reduced the face's shading. Panels (e) and (f) demonstrate that, when LinearGANSpace and WarpedGANSpace enlarged or vertically moved dogs in photos, they changed the orientations and backgrounds. Panel (g) shows that, when editing the yaw attribute, LinearGANSpace also edits the hairstyle, and WarpedGANSpace edits the skin color. Panel (h) shows that WarpedGANSpace's edit of the smile attribute altered the pitch and yaw rotations, as in Fig. 3. Panels (i) and (j) show similar tendencies for LSUN Car.

Therefore, we conclude that the nonlinear and commutative nature of DeCurvEd contributes to the disentanglement between attributes and high-quality editing. We also provide additional results in Appendix D.

6. Conclusion

This study proposed deep curvilinear editing (DeCurvEd), which defines a curvilinear coordinate on the latent space of generative models and edits images along axes of the coordinate. DeCurvEd's edits of semantic attributes are theoretically nonlinear and commutative. Combined with pre-trained GANs, we proposed CurvilinearGANSpace and experimentally demonstrated that it is superior to previous methods whose edits are linear or non-commutative in terms of the disentanglement between attributes and the preservation of identity. Future work will focus on a combination of DeCurvEd with other deep generative models in supervised and unsupervised learning, such as Khrulkov et al. [37].

Limitations: Because DeCurvEd assumes a continuous change in attribute, it is unavailable for a discrete attribute, such as "wearing sunglasses." A combination with discrete attributes remains a topic for future research.

Acknowledgements: This work was supported by JST PRESTO (JPMJPR21C7), CREST (JPMJCR1914), and JSPS KAKENHI (19K20344, 19H04172), Japan.
References


Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. \textit{Alien-Free Generative Adversarial Networks}. In \textit{Advances in Neural Information Processing Systems}, 2021. 1


Yann LeCun and Corinna Cortes. \textit{The mnist database of handwritten digits}. 2005. 5, 13


Aditya Ramesh, Youngduck Choi, and Yann LeCun. \textit{A Spectral Regularizer for Unsupervised Disentanglement}. \textit{ArXiv}, pages 1–17, 2018. 1, 2, 3, 4


[56] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-Based Generative Modeling through Stochastic Differential Equations. In International Conference on Learning Representations (ICLR), pages 1–32, 2021. 1, 2, 5


[65] W. Xia, Y. Zhang, Y. Yang, J. Xue, B. Zhou, and M. Yang. IEEE Transactions on Pattern Analysis and Machine Intelligence. 4


[67] Jiapeng Zhu, Yujun Shen, Yinghao Xu, Deli Zhao, and Qifeng Chen. Region-Based Semantic Factorization in GANs. In International Conference on Machine Learning, 2022. 2
