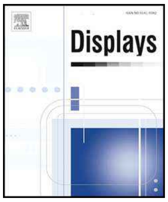
Displays 80 (2023) 102560

Contents lists available at ScienceDirect

Displays

journal homepage: www.elsevier.com/locate/displa



Optimal transport-based unsupervised semantic disentanglement: A novel approach for efficient image editing in GANs[☆]

Yunqi Liu [a], Xue Ouyang [b], Tian Jiang [a], Hongwei Ding [a], Xiaohui Cui [a,*]

[a] *Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, School of Cyber Science and Engineering, Wuhan University, Wuhan, China*

[b] *State Key Laboratory of Information Engineering in Surveying, Wuhan University, Wuhan, China*

ARTICLE INFO

Keywords:
Image editing
Generative adversarial network (GAN)
Optimal transport
Semantic disentanglement
Manifold learning

ABSTRACT

The latent space of pre-trained generative adversarial networks (GANs) is rich in semantic information, which often becomes highly entangled. It is crucial to identify semantic directions within this latent space, as these directions correlate with image attributes and are vital for image editing tasks. Existing methods for semantic discovery usually involve labor-intensive procedures such as manual labeling and training attribute classifiers, which limits their practicality. In response to this issue, the paper proposes the Optimal Transport-based Unsupervised Semantic Disentanglement (OTUSD) algorithm. This novel method efficiently uncovers semantic directions in the latent space of GANs by utilizing the concepts of manifold learning and optimal transport (OT) theory. OTUSD applies singular value decomposition (SVD) to the OT matrix that links latent codes to generated images. This process yields singular vectors that correspond to semantically meaningful directions. Unlike traditional methods, OTUSD bypasses the need for time-consuming labeling and training processes, thus enhancing efficiency and revealing a wider array of semantically meaningful directions. Experimental results demonstrate the effectiveness of OTUSD in discovering semantic directions from several state-of-the-art GAN models, including StyleGAN, StyleGAN2, and BigGAN. This performance emphasizes the potential applicability of OTUSD to image editing and other related tasks, and illuminates its value in harnessing the manifold learning and OT mapping capabilities inherent in GANs for semantic disentanglement. The implementation code is available at https://github.com/LuckAlex/OTUSD.

1. Introduction

Generative adversarial network (GAN) models [1–4] have garnered significant attention in the machine learning community. These models are renowned for their ability to generate photo-realistic samples that rival the quality and authenticity of actual images. Given the impressive prowess of GANs in image generation, a natural progression is image manipulation. This introduces the tantalizing prospect of precisely controlling and modifying specific attributes of an image, such as altering facial age or changing the ambiance of a landscape photo. The pivotal question is: How can such manipulations be effectively achieved?

A direct approach might involve manipulating the pixel values of an image to induce the desired changes. However, this method encounters several challenges:

- **Complexity:** Due to the high-dimensional nature of image data, direct manipulations become intricate and challenging to control.
- **Inconsistencies:** Pixel-level alterations, without a deep understanding of underlying structures, can lead to unrealistic and jarring modifications.
- **Lack of generalization:** An alteration effective for one image might prove unsuitable for another, given the intrinsic variations in attributes.

Contrastingly, the latent space of a GAN provides a more structured and conducive environment for manipulating semantic attributes [5–8]:

- **Compact representation:** The latent space offers a condensed representation of images. Alterations in this domain can effectuate controlled modifications in the resultant image, assuming an understanding of the relationships between different dimensions in the latent space.

[☆] This paper was recommended for publication by Prof Guangtao Zhai.

[*] Corresponding author at: Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, School of Cyber Science and Engineering, Wuhan University, Wuhan, China.

E-mail address: xcui@whu.edu.cn (X. Cui).

https://doi.org/10.1016/j.displa.2023.102560
Received 12 July 2023; Received in revised form 6 October 2023; Accepted 18 October 2023
Available online 21 October 2023

- **Disentanglement potential:** While entanglements are inherent to the latent space, its structured nature permits, with the right techniques, the isolation and disentanglement of specific semantic attributes. Such disentanglement ensures that modifications to one attribute do not inadvertently influence others.
- **Generalizability:** Upon discovering semantic directions in the latent space, they can be consistently applied across various images, ensuring realistic and controlled modifications.

Given this context, multiple methods, categorized primarily into supervised and unsupervised paradigms, have been proposed to manipulate images through semantic disentanglement in the latent space.

Supervised methods are commonly used for latent semantic discovery. They involve converting a large amount of potential code into a set of images. Then, based on the specific semantic direction, use manual or attribute classifiers to label or classify these images. These methods can control the corresponding attribute of the generated image. Several works have explored different techniques for fine-grained image editing using GAN models. Both InterfaceGAN [7] and AdvStyle [9] use GANs to understand the relationship between attributes of images and the latent space, which is a multi-dimensional space inside the model used for generating images. However, their methods for manipulating these attributes differ. InterfaceGAN employs support vector machines [10], a type of machine learning model, to discern boundaries between different binary attributes. This process effectively creates a line of demarcation in the latent space, aiding in understanding what changes can lead to a shift from one attribute to another. On the other hand, AdvStyle adopts an adversarial technique to discover attribute directions within the latent space, using a game-theoretic approach to identify how attributes influence the generated images. More recently, StyleFlow [11] introduced a unique technique called reversible mapping. This method employs normalizing flows and pre-trained classifiers to create a back-and-forth mapping between the latent space and the image space. This results in a more adaptable manipulation of image attributes. However, these approaches still have limitations in capturing the non-linearities and location specific properties of facial attributes. Addressing this issue, Jiang et al. [12] proposed a new concept termed a semantic field. A semantic field is a vector field that explains the specific direction and magnitude of changes for different attributes within the latent space of a GAN. This concept allows for more accurate face editing by moving along a curved trajectory defined by the semantic field. Despite these advancements, supervised methods, which learn from labeled data, face limitations. They require specific target attributes for labeling, which can restrict their ability to discover a wide variety of interpretative directions. As a result, they might not identify new and diverse semantic directions beyond the initial set of attributes, meaning the scope of what they can learn and generate could be constrained by the labeled data on which they are trained.

Unsupervised methods, such as LowRankGAN [13], Style Intervention [14], and SeFa [15], have gained increased attention due to their flexibility in discovering semantic directions. However, these methods also have limitations. LowRankGAN [13] performs low-rank decomposition of the Jacobian matrix, but calculating the matrix is time-consuming and difficult to generalize and apply. The Style Intervention method [14] requires an optimization procedure for each input image and involves operations on each graph, which is not easily generalizable. SeFa [15] identifies edited semantic directions in the latent space by decomposing the pre-trained GAN model weights, but it only considers the weight used in the first transformation step and is constrained by the fixed weight parameters of the pre-trained GAN model. This may result in semantic entanglement and inaccurate control of the discovered semantic direction.

In this study, we introduce a pioneering approach that adeptly uncovers disentangled semantic directions within the GAN latent space. Our innovative algorithm, termed Optimal Transport-based Unsupervised Semantic Disentanglement (OTUSD), leverages optimal transport (OT) theory. This allows it to compute the ideal mapping between two probability distributions, culminating in the generation of an OT matrix bridging the input and output realms of the GAN model. This matrix, rich in information about image variations, forms the nexus between latent and generated spaces, transmuting modifications in the latent codes into discernible variations in the resultant images. Employing singular value decomposition (SVD) on this matrix, OTUSD pinpoints singular vectors intrinsically linked with semantically relevant directions. This methodology not only simplifies the disentanglement process but also expands the range of discernible semantic directions, laying the groundwork for expansive applications in image editing and allied domains. Our main contributions are:

- A groundbreaking unsupervised approach that simplifies the disentanglement process and broadens the scope of identifiable semantic directions.
- An efficient algorithm that surpasses state-of-the-art methods in its range and accuracy.
- Comprehensive experiments showcasing the effectiveness of OTUSD.
- Demonstrated robustness across various GAN models and datasets.

The rest of this article is organized as follows. Section 2 provides an overview of related work in the field. Our OTUSD is introduced in detail in Section 3. Section 4 presents an empirical evaluation of our approach. In Section 5, we engage in a detailed discussion of our findings. Finally, Section 6 encapsulates our conclusions and potential future directions.

2. Related works

2.1. Generative adversarial networks

GANs have revolutionized image processing in recent years [16–18] and have been widely used in the creation of handwritten fonts [19,20], image editing programs [7,21], and image super-resolution [22], etc. Although state-of-the-art models such as StyleGAN2 [17] and BigGAN [23] have made great progress in terms of synthesis quality and training stability, there is still a lack of research on controlling the generation process of GANs.

2.2. OT

OT is the problem of efficiently moving one mass distribution to another [24]. This fundamental problem has numerous applications in mathematical fields [25,26] such as partial differential equations, geometry, functional analysis, optimization, and probability. OT is a powerful tool for studying probability distribution modeling using geometric methods, and it has significant implications for engineering fields [27] such as image processing, data science, economics, and chemical physics. To obtain more comprehensive reviews, the researcher may refer to Refs. [28,29].

2.3. Semantic editing with conditional GANs

In general, unconditional GANs can only generate images randomly [17,30,31]. To control the process of image generation, it is necessary to design the loss function, network structure, or provide extra prior knowledge. Isola et al. [32] proposed pix2pix as a typical conditional GAN that combines two conditional GANs to perform image conversion tasks. Lu et al. [33] generate high-resolution images from low-resolution inputs that satisfy given semantic attributes. However, the quality of images generated by conditional methods is generally inferior to that of unconditional GANs such as StyleGAN [31] and BigGAN [23], and they can only manipulate a few specific attributes. In contrast, exploring semantic directions in the latent space can accurately control image attributes while ensuring image quality.

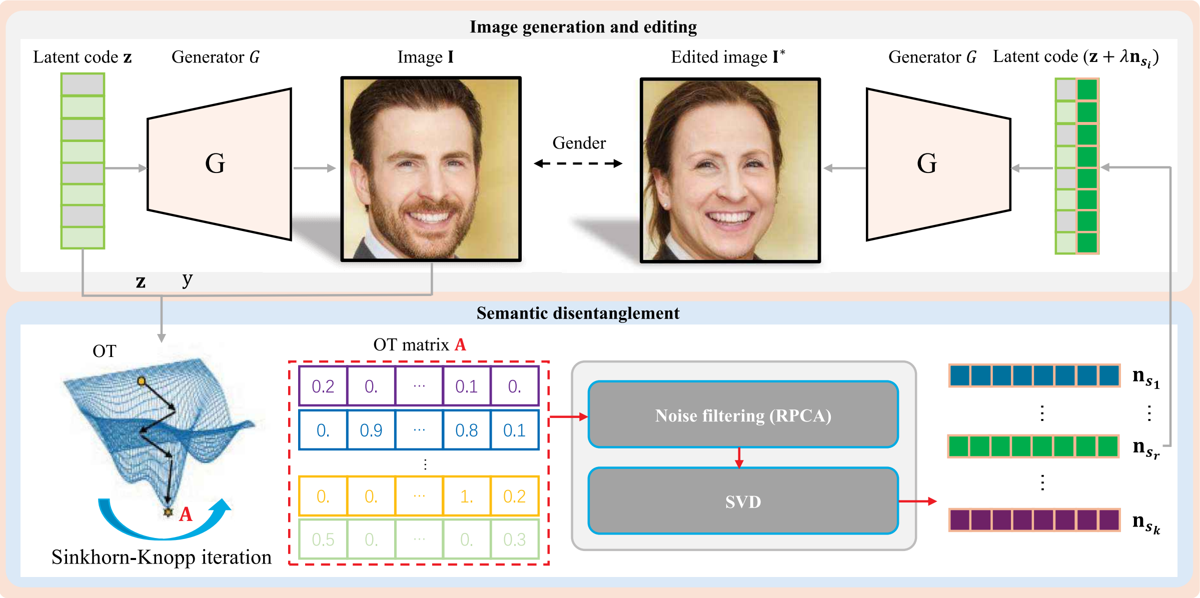


Fig. 1. Overview of the proposed OTUSD. Given a pre-trained GAN model G, we use the OT algorithm to calculate the OT matrix $\mathbf{A}$ between the latent code $\mathbf{z}$ and the generated image $\mathbf{I}$. After noise filtering and SVD of the OT matrix $\mathbf{A}$, the semantic directions $\mathbf{N} = [\mathbf{n}_{s_1}, \ldots, \mathbf{n}_{s_r}, \ldots, \mathbf{n}_{s_k}]$ can be extracted. Using $\mathbf{n}_{s_i}$ can transform the original generated image $\mathbf{I}$ to the edited image $\mathbf{I}^*$, whose gender attribute is changed.

2.4. GAN inversion

The GAN inversion technique can invert a given real image into the latent space of a pre-trained GAN model, which is essential for GAN-based real image editing applications [34]. Existing GAN inversion solutions can be divided into three categories: learning-based [35,36], optimization-based [34], and hybrid-based [37,38]. Generally, learning-based GAN inversion approaches cannot accurately reconstruct image content. Recent advancements in learning-based GAN inversion approaches have focused mostly on accurately reconstructing images, such as incorporating extra face recognition losses during training or iterative feedback. The optimization-based approach can provide superior image reconstruction results, but at a prohibitively high computing cost. Recent optimization-based enhancement techniques have emphasized locating the necessary potential code more quickly and have proposed various initialization and optimizer methodologies. Hybrid-based approaches attempt to balance the aims of reconstruction quality and inference speed, but swiftly finding the correct latent code remains challenging.

3. Methods

In this section, we first illustrate the problem of semantic disentanglement in GANs and then demonstrate how the OT algorithm can be used to compute editable disentangled semantic directions. Fig. 1 illustrates the overall framework of the proposed OTUSD.

3.1. Problem formulation

Simplified model of GANs: Consider a GAN where the generator G accepts a d-dimensional latent code $\mathbf{z}$ as input, typically drawn from a Gaussian distribution, and transforms it into an image as follows:

$$\mathbf{I} = G(\mathbf{z}) = \mathbf{A}\mathbf{z} \tag{1}$$

where $\mathbf{I}$ is the output image with dimensions H (height), W (width), and C (number of channels), such that $\mathbf{I} \in \mathbb{R}^{H \times W \times C}$. And $\mathbf{A}$ is a transformation matrix that encapsulates the mapping from a Gaussian noise distribution to the distribution of generated image data. This linear representation serves as a simplified model to facilitate theoretical analysis, acknowledging that actual GAN models comprise multiple layers of non-linear transformations.

It is important to note that the linear transformation representation in Eq. (1) is a simplification. In actual GAN models, the generator often comprises multiple layers of non-linear transformations instead of a singular linear transformation. However, this simplified model serves as a convenient abstraction to comprehend the fundamental concepts and aid in theoretical analysis.

Mathematical framework for semantic disentanglement: Semantic disentanglement refers to the process of identifying and isolating semantically meaningful directions in the latent space of a GAN, allowing for independent control and manipulation of distinct image attributes in the generated output.

Let Z denote the latent space of a GAN, and S be the set of all possible semantic attributes. For each semantic attribute $s \in S$, we aim to find a direction $\mathbf{n}_s \in Z$ such that modifying the latent code along $\mathbf{n}_s$ results in changes in the attribute s in the generated image.

For a given direction $\mathbf{n}_s$ associated with the semantic attribute s, the edited image $\mathbf{I}^*$ can be represented as:

$$\mathbf{I}^* = G\left(\mathbf{z} + \lambda \mathbf{n}_s\right) = \mathbf{A}\mathbf{z} + \lambda \mathbf{A}\mathbf{n}_s = G(\mathbf{z}) + \lambda \mathbf{A}\mathbf{n}_s \tag{2}$$

where λ is a scalar that modulates the intensity of the attribute s in the edited image $\mathbf{I}^*$.

To edit a specific attribute of an image, such as the eyes or smile, different control parameters can generate the desired effect. For instance, setting λ as λ_1 results in the edited image becoming $\mathbf{I}_1^* = G(\mathbf{z}) + \lambda_1 \mathbf{A}\mathbf{n}_s$. When λ is set as λ_2, the edited image becomes $\mathbf{I}_2^* = G(\mathbf{z}) + \lambda_2 \mathbf{A}\mathbf{n}_s$. The difference between these two edited images is:

$$\mathbf{I}_1^* - \mathbf{I}_2^* = (\lambda_1 - \lambda_2)\mathbf{A}\mathbf{n}_s \tag{3}$$

It is important to emphasize that only the specific attribute being edited changes to varying degrees, while other attributes remain constant.

As depicted in Fig. 2, the latent semantic direction $\mathbf{n}_s$ only undergoes a transformation stretch when acted upon by the transformation matrix $\mathbf{A}$, while its direction remains unaltered. This aligns with the definition of singular vectors, stating that the right singular vector $\mathbf{v}$ of a real-valued matrix $\mathbf{A}$ experiences only stretching or shrinking

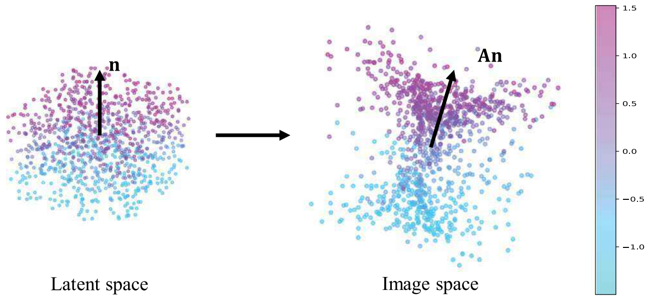


Fig. 2. Illustration of identifying a principal activation direction. Directionality is indicated by a color gradient: blue to purple. Despite spatial orientations due to varying distribution shapes, the color gradient consistently represents the true transformation direction across both spaces.

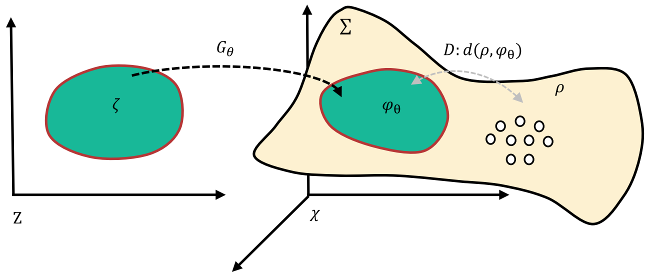


Fig. 3. The working principle of GAN from the perspective of manifold learning. *Source:* Modified from Ref. [39].

by a factor of μ when acted upon by $\mathbf{A}$, with its direction remaining the same. Hence, the latent semantic direction $\mathbf{n}_s$ that satisfies these conditions is the right singular vector of $\mathbf{A}$, i.e., $\mathbf{n}_s = \mathbf{v}$. Consequently, Eq. (3) can be rewritten as:

$$\mathbf{I}_1^* - \mathbf{I}_2^* = (\lambda_1 - \lambda_2)\mathbf{A}\mathbf{n}_s = (\lambda_1 - \lambda_2)\mu\mathbf{A}\mathbf{n}_s \tag{4}$$

where μ denotes the constant singular value.

In conclusion, by adjusting the control parameters, the edited images exhibit changes solely in attributes associated with the latent semantic direction, allowing for precise and efficient control over target images. The solution to this problem involves identifying the transformation matrix $\mathbf{A}$ and subsequently obtaining the set of semantic directions by performing SVD.

3.2. Motivation

In a GAN, the generator, which is central to the model, usually consists of a sequence of non-linear transformations. These transformations map a simple noise distribution to a complex data distribution. Representing this series of transformations as a single matrix, denoted as $\mathbf{A}$, can be complicated due to the intricate, multi-layered structure of the generator. This complexity often necessitates us to resort to simplifications, treating the entire chain of transformations within the generator as a single, effective transformation encapsulated in the matrix $\mathbf{A}$.

In related work, LowRankGAN [13] calculates the Jacobian matrix between the input and output to represent matrix $\mathbf{A}$. However, this method can be computationally intensive, taking more than an hour to calculate a single Jacobian matrix. Another approach, SeFa [15], constructs matrix $\mathbf{A}$ using the weight parameters of the first layer of the pre-trained GAN model. While this method is computationally more efficient, it may not fully capture the entire generative process as it only considers the first layer of the generator.

In contrast, our approach utilizes the OT matrix to represent $\mathbf{A}$. This is based on the observation that GANs, viewed through the lens of manifold learning, perform two main tasks [24,40]:

1. Manifold learning: This involves computing the generative mapping G_θ, where θ represents the parameters of a deep neural network, from the latent space Z to a manifold $\sum$ embedded in the ambient space χ.
2. Probability distribution transformation: This entails the transformation of white noise, typically Gaussian distributed, into a data distribution. The generator of the GAN is essentially computing an OT mapping from the white noise to the data distribution.

These are not independent tasks but are deeply intertwined. The generator and discriminator of the GAN work together to achieve these. While the generator is responsible for both tasks, creating a mapping from the latent space to the data distribution, the discriminator plays a crucial role in guiding this process. It measures the Wasserstein distance [41] between the generated data distribution and the true data distribution. This measurement acts as a feedback signal to the generator, helping it refine the transformation it applies to the latent code to produce a distribution more similar to the true data distribution.

These fundamental processes are depicted in Fig. 3.

OT theory [42] provides a rigorous method for computing the optimal mapping from one probability distribution to another. In the context of a GAN, the generator is essentially computing this OT mapping.

Our approach utilizes the OT matrix as matrix $\mathbf{A}$, aiming to address the limitations of previous methods and offer an alternative solution for precise control and manipulation of semantic directions in GAN models.

3.3. Calculate the transformation matrix

Given the latent code $\mathbf{z}$ of d dimensions, the frequency distribution is captured by weights z_i for each dimension i. Formally:

$$\mathbf{z} = [z_1, z_2, \ldots, z_d] \tag{5}$$

where z_i represents the probability of occurrence of the ith dimension.

The histogram distribution of the image $\mathbf{I}$ is denoted by $\mathbf{y}$ in $\mathbb{R}^l$. Each bin j has a probability weight y_j, representing the likelihood of a specific intensity:

$$\mathbf{y} = [y_1, y_2, \ldots, y_l] \tag{6}$$

where y_j is the probability of the jth intensity value in the histogram of the image.

With these frequency distributions, we approach the OT problem to align the distributions. The problem is to find the transformation matrix $\mathbf{A}$ that minimizes the total transportation cost, subjected to constraints:

$$\begin{aligned} &\mathbf{A} = arg\,min_{\mathbf{A}} \textstyle\sum_i^d \sum_j^l c_{i,j} a_{i,j} \\ &s.t.\ \textstyle\sum_i^d a_{i,j} = z_j;\ \sum_j^l a_{i,j} = y_i;\ a_{i,j} \geq 0 \end{aligned} \tag{7}$$

where $a_{i,j}$ denotes the mass transferred between bins, and $c_{i,j}$ represents the distance (cost) between bins.[1]

However, OT is a linear programming problem that can be solved with time complexity $\mathcal{O}(n^3 logn\ )$, which is quite expensive. This problem can be addressed by adding a regularization term [43]. The use of the regularization term in the OT problem has following effects: (1) it can significantly improve the speed of solving OT problem; (2) it makes the issue convex, ensuring that there is only one solution.

Therefore, the regularized OT problems between $\mathbf{z}$ and $\mathbf{y}$ can be expressed as

$$\begin{aligned} &\mathbf{A} = argmin_{\mathbf{A}} \textstyle\sum_i^d \sum_j^l c_{i,j} a_{i,j}\ + \lambda\Omega(\mathbf{A}) \\ &s.t.\ \textstyle\sum_i^d a_{i,j} = z_j;\ \sum_j^l a_{i,j} = y_i;\ a_{i,j} \geq 0 \end{aligned} \tag{8}$$

[1] *The similarity distance can be measured by commonly used distance measurement methods, and Euclidean distance is chosen in this paper.*

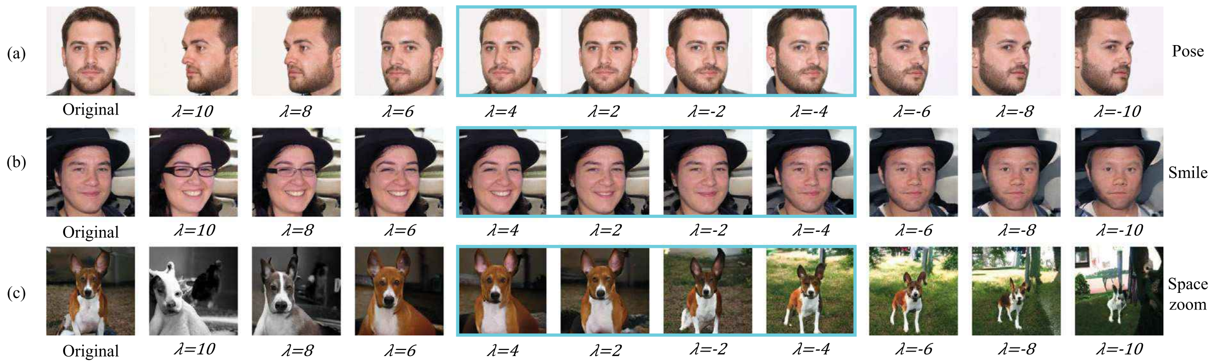


Fig. 4. Generated images under different parameters λ. (a) Pose (StyleGAN [31] trained on CelebA-HQ [46]); (b) Smile (StyleGAN [31] trained on FFHQ [31]); (c) Space zoom (BigGAN [23] trained on ImageNet [47]).

where Ω is the regularization term. Moreover, the regularization term Ω can be expressed by the Entropic regularization [44], i.e.,

$$\Omega(\mathbf{A}) = \sum_{i,j} a_{i,j} log(a_{i,j}) \tag{9}$$

The solution of the resulting optimization problem can be expressed as

$$\mathbf{A} = diag(\mathbf{w})\mathbf{K}diag(\mathbf{v}) \tag{10}$$

where $\mathbf{w}$ and $\mathbf{v}$ are vectors and $\mathbf{K_{i,j}}=exp(-c_{i,j}/\lambda)$. Sinkhorn–Knopp [44] is an alternate projection algorithm that, given high quantities of regularization, can be extremely effective in solving the optimization issue. The detailed derivation of the Sinkhorn–Knopp algorithm can be seen in [44]. After calculation, we can get the OT matrix (transformation matrix) $\mathbf{A}$ between the latent code $\mathbf{z}$ and the generated image $\mathbf{I}$.

3.4. Semantic disentanglement

As previously discussed, the transformation matrix $\mathbf{A}$ has been solved, yet it contains both structural information and noise. Therefore, in order to filter the noise and extract the structural information from $\mathbf{A}$, the robust principal component analysis (RPCA) [45] algorithm is employed to decompose the matrix $\mathbf{A}$ into two additive matrices, that is,

$$\min_{\mathbf{A},\mathbf{E}} \quad rank\,(\mathbf{A}) + \eta\, \mathbf{E}_0 \qquad s.t\ \mathbf{A}{=}\mathbf{A}^*{+}\mathbf{E} \tag{11}$$

where $\mathbf{E}$ is the filtered-out sparse noise matrix and $\mathbf{E}_0$ represents the count of non-zero elements in the matrix $\mathbf{E}$ (i.e., the zero "norm"). $\mathbf{A}^*$ is the low-rank representation of the transformation matrix $\mathbf{A}$, encapsulating significant information about the transformation of latent space semantics into image attributes.

Therefore, the original precise control process of the target image based on Eq. (2) changes as follows

$$\begin{aligned} \mathbf{I}^* &= G\left(\mathbf{z} + \lambda \mathbf{n}_s\right) = \mathbf{A}^*\left(\mathbf{z} + \lambda \mathbf{n}_s\right) \\ &= \mathbf{A}^*\mathbf{z} + \lambda \mathbf{A}^*\mathbf{n}_s = G\left(\mathbf{z}\right) + \lambda \mathbf{A}^*\mathbf{n}_s \end{aligned} \tag{12}$$

where $\mathbf{n}_s$ is a latent semantic direction.

Furthermore, the significant information in matrix $\mathbf{A}$ is needed to extract to discover and obtain disentangled semantic directions. By SVD algorithm, we can get the singular values and singular vectors of $\mathbf{A}$, i.e.,

$$\mathbf{A}^*{=}\mathbf{N}\boldsymbol{\Lambda}\mathbf{N}^\mathbf{T} \tag{13}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix sorted by singular values, $\mathbf{N} = [\mathbf{n}_{s_1}, \ldots, \mathbf{n}_{s_r}, \ldots, \mathbf{n}_{s_k}]$ is an orthogonal matrix composed of the singular vectors of $\mathbf{A}$, which contains the main k important semantic directions. Eventually, we can simply edit a specific attribute of an image by using the singular vector matrix $\mathbf{N}$ based on Eq. (12).

4. Experiments

Our method, OTUSD, has been validated using three state-of-the-art, pre-trained GAN models: BigGAN [23], StyleGAN [31] and StyleGAN2 [17]. These models were initially trained on various datasets, including FFHQ [31], CelebA-HQ [46], anime faces [48], art faces [49], LSUN Car [49], LSUN Cat [49], and ImageNet [47]. The evaluations were carried out on a computer environment equipped with an Intel(R) Core(TM) i5 CPU, operating at 2.90 GHz, and 32 GB RAM. We assessed the performance of OTUSD both qualitatively and quantitatively to demonstrate the versatility and efficiency of our proposed method.

4.1. Implementation details

Algorithm 1 OTUSD Algorithm for Semantic Disentanglement

```
Input: Pre-trained GAN model with generator $G$, latent code $\mathbf{z}$
Output: Identified latent semantic directions $\mathbf{N} = [\mathbf{n}_{s_1}, \ldots, \mathbf{n}_{s_r}, \ldots, \mathbf{n}_{s_k}]$
1: $\mathbf{I} \leftarrow G(\mathbf{z})$                         ▷ Generate image from latent code
2: $\{z_i\}_{i=1}^{d} \leftarrow h(\mathbf{z})$          ▷ Compute occurrence probabilities for latent code
3: $\{y_j\}_{j=1}^{l} \leftarrow h(\mathbf{I})$          ▷ Compute histogram distribution for image
4: $c_{i,j} \leftarrow d(i,j)$                      ▷ Compute cost between bins
5: Solve for $\mathbf{A}$ using regularized OT problem from Eq. (8)
                                                   ▷ Calculate transformation matrix $\mathbf{A}$
6: $\mathbf{A}^*, \mathbf{E} \leftarrow$ RPCA($\mathbf{A}$)          ▷ Apply RPCA to decompose $\mathbf{A}$
7: $\mathbf{N}, \boldsymbol{\Lambda} \leftarrow$ SVD($\mathbf{A}^*$)          ▷ Extract singular vectors and values using SVD
8: return $\mathbf{N}$                         ▷ Return the identified latent semantic directions
```

Algorithmic overview:The proposed method, OTUSD, for semantic disentanglement is presented in Algorithm 1. A brief explanation follows: The process initiates by generating an image from a latent code using a pre-trained GAN model (Step 1). Subsequently, the method computes the occurrence probability of the latent code and the histogram distribution of the generated image (Steps 2–3). A pivotal component is Step 5, where the transformation matrix $\mathbf{A}$ is derived using a regularized OT problem, establishing the relationship between the latent and image spaces. The subsequent step filters noise from $\mathbf{A}$ through RPCA (Step 6) and extracts the semantic direction using SVD (Step 7). The algorithm concludes by returning the identified latent semantic direction, which can be harnessed for precise image manipulations (Step 8). This algorithm embodies the essence of the OTUSD method, facilitating the semantic disentanglement process in GANs.

Selection of input–output pairs: Our approach does not necessitate the computation of an OT matrix for every individual image. This provides a significant computational advantage as, once derived, a single semantic direction can be reused across multiple images, saving on redundant calculations. The semantic directions gleaned from one image can be effectively leveraged to edit either the same image or other images. This feature renders OTUSD as both user-friendly and

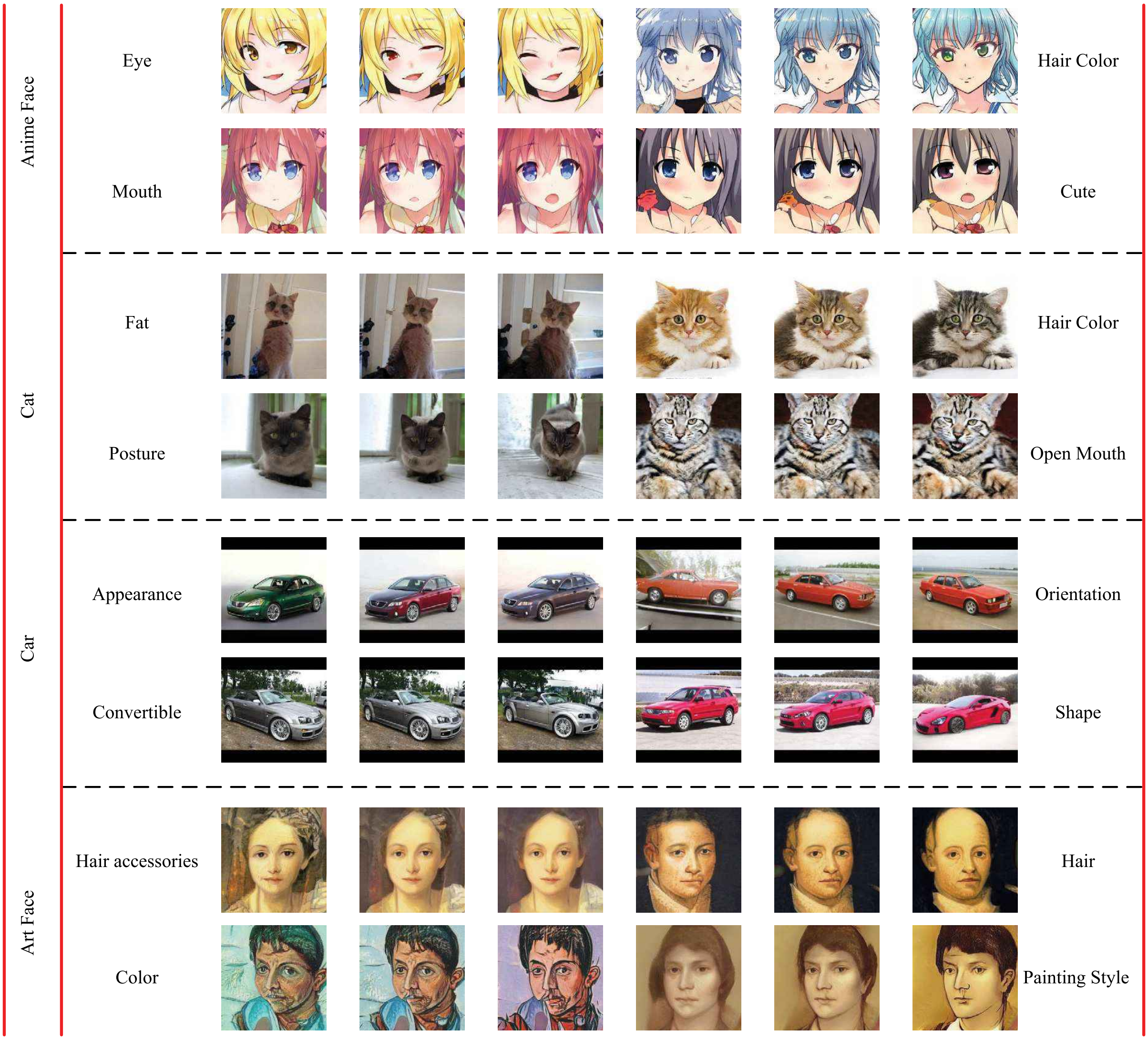


Fig. 5. Results on various models and datasets, with the cat model trained on StyleGAN2 [17] and the others on StyleGAN [31].

versatile. Although semantic directions obtained from a single image can be effectively utilized for editing the same or different images, there are certain attribute features that may not be prevalent across all images. To ensure that the method captures a comprehensive range of image attributes, especially when dealing with diverse image datasets, it might be beneficial to gather multiple input–output pairs, compute their corresponding $\mathbf{A}_i$ matrices, and subsequently obtain a variety of singular vectors.

Semantic direction determination and parameter selection of singular vectors: The unsupervised nature of OTUSD offers a distinct advantage: the capability to identify rare yet valuable semantic directions associated with image manipulations. However, this comes with a challenge: the inability to predict in advance which attributes correspond to each feature vector. Decomposing the singular vectors of the OT matrix yields a set of singular vectors, $\mathbf{N} = [\mathbf{n}_{s_1}, \dots, \mathbf{n}_{s_r}, \dots, \mathbf{n}_{s_k}]$, encompassing k key semantic directions. Yet, even with these directions, the semantics associated with each singular vector remain unclear.

To gain clarity, we apply the decomposed singular vector $\mathbf{n}_{s_r}$ to a sampled latent code. The intensity of changes in image features is regulated by the parameter λ (as detailed in Eq. (12)). To systematically understand its influence, we sample 10 parameter values uniformly within the interval $[-p, p]$. Each sampled value generates a corresponding image. Displaying these images consecutively provides clear insight into the semantic attributes associated with the singular vector, as shown in Fig. 4.

An essential observation from Fig. 4 is the sensitive nature of λ. Surpassing certain thresholds with λ can lead to unintended semantic modifications or degrade image quality. Specifically, λ values above 5 or below −5 tend to yield suboptimal results. Consequently, we suggest maintaining λ within the range [−5,5] to ensure meaningful and high-quality image transformations.

Input spaces for different models: In order to get good results in different models, the input spaces used in different models are given in this section. Firstly, since the significant characteristic of the style-based generator architecture for GANs (such as StyleGAN [31] and StyleGAN2 [17]) is: the input vector $\mathbf{z}\in\mathbf{Z}$ can be translated to the intermediate latent vector $\mathbf{w}\in\mathbf{W}$, which can “unwarp” $\mathbf{W}$ and make the components of variation in the intermediate latent space much more linear. This mapping not only helps the generator to produce realistic images, but also for better analyzing the property of the linear subspace. Moreover, it allows undertaking additional semantic editing in the latent space. Therefore, we focus on the latent space $\mathbf{W}$. In addition, BigGAN [23] is a large-scale GAN model built for conditional generation. The latent code is simultaneously transferred to the initial feature map and given to every convolution layer. For the BigGAN, we choose the input latent space $\mathbf{Z}$.

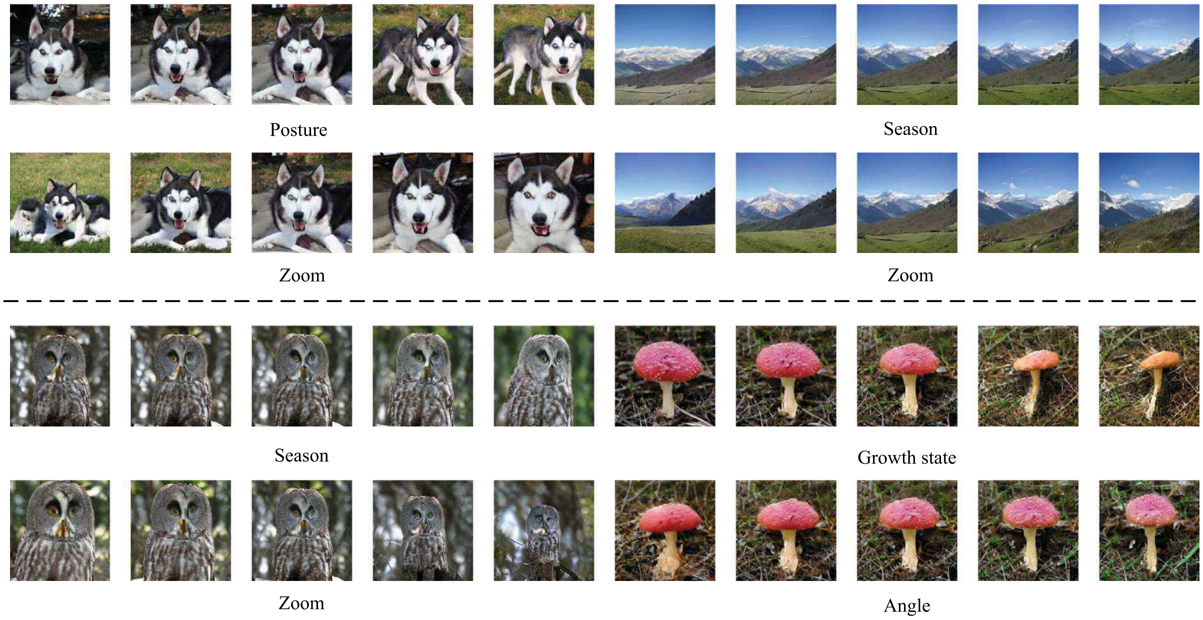


Fig. 6. Different interpretable directions found in the BigGAN [23], which is trained on ImageNet [47].

4.2. Results on different models and datasets

Given that OTUSD is not reliant on specific model structures or datasets, we first evaluated its performance using popular models such as StyleGAN [31] and StyleGAN2 [17], as well as models trained on distinct datasets such as anime faces, cat images, car photos, and artistic portraits. Fig. 5 provides visual examples of different models and datasets. By applying OTUSD to each of these cases, we observed that it successfully decomposes various semantic directions across diverse datasets, allowing for precise editing of specific properties. These results demonstrate that OTUSD exhibits a high degree of generalization, with broad applications in image editing.

Next, we proceeded to analyze the BigGAN model, which was trained on ImageNet [47] using conditional training and a large-scale approach. Specifically, we evaluated the generalization capability of OTUSD using images from a variety of categories. Fig. 6 shows examples of the manipulated images produced by our approach. As demonstrated by these results, OTUSD is able to uncover meaningful semantics that can be used to manipulate images from diverse categories. These findings corroborate our previous results and further support the generalization capability of OTUSD.

4.3. Comparison with supervised approaches

In this section, we applied the proposed OTUSD algorithm to a typical face synthesis model, specifically a StyleGAN model [31] that was pre-trained on the FFHQ dataset [31]. Furthermore, we compared the results obtained using OTUSD with those obtained using state-of-the-art supervised algorithms such as InterFaceGAN [7] and AdvStyle [9].

4.3.1. Qualitative results

In this test, we mainly compared the editing results of OTUSD with InterFaceGAN [7] and AdvStyle [9] for typical attribute directions of the generated images (e.g., gender, glasses, age, smile, and pose), and the results are shown in Fig. 7.

It can be seen that AdvStyle causes significant changes in gender when controlling the semantic attributes of glasses, and is not very precise in controlling other semantic attributes. In contrast, both OTUSD and InterFaceGAN provide precise control over these attributes of the generated face. However, InterFaceGAN introduces a slight variation in the smile attribute when controlling for changes in the gender attribute. OTUSD achieves comparable or even superior control effects compared to supervised methods.

4.3.2. Diversity comparison

For editing characteristic attributes (e.g., race), it is not simply a matter of changing specific organs or facial expressions, but also involves changes to the overall facial characteristics. Therefore, it is a challenging task to modify race attributes while preserving other facial features.

To address this challenge, we varied race, skin color, and beard to different degrees in this experiment, and the results are shown in Fig. 8. This figure presents semantic directions that are absent from the InterFaceGAN and AdvStyle papers. It can be observed that OTUSD can adjust race attributes without producing drastic changes in other facial attributes, which is difficult to achieve. Furthermore, the proposed method enables precise control of skin color and beard attributes. This is because OTUSD is not restricted by the attribute predictor and can extract and apply attribute directions from small samples in the dataset.

Importantly, our method does not require computing the OT matrix for each image. The semantic directions obtained from one image can be used to efficiently edit itself or other images, as illustrated in Fig. 8. This makes OTUSD user-friendly and flexible.

4.3.3. Attribute correlation

Theoretically, different editable attribute directions should be disentangled, i.e., attribute directions are independent of each other [9]. To assess whether they are disentangled, we use cosine similarity to calculate the correlation matrix between different attribute direction vectors and determine their similarity. In particular, the cosine similarity between two attribute direction vectors is independent of their magnitudes, but only of their angle. For example, a cosine similarity closer to 1 indicates a higher correlation between the vectors, closer to 0 indicates a lower correlation, and a negative value indicates a negative correlation, which can usually be considered irrelevant. Fig. 9 shows the correlation matrices calculated by cosine similarity between different attributes of OTUSD, InterFaceGAN [7] and AdvStyle [9]. As shown in Fig. 9(b) and (c), the discovered directions of InterFaceGAN and AdvStyle are relatively entangled. On the contrary, most directions of OTUSD are highly disentangled (see Fig. 9(a)), i.e., they are orthogonal to each other. Therefore, it can be demonstrated that OTUSD can effectively disentangle into different attribute directions.

Table 1
Quantitative comparison with LowRankGAN [13] and GANSpace [50].

Methods	IS	FID	Success rate (%)	User study (%)	Time (minutes)
SeFa [15]	3.35	19.51	92.2	27	1
LowRankGAN [13]	3.46	21.05	88.7	20	68
GANSpace [50]	3.41	20.34	93.0	24	2
OTUSD	3.52	19.48	94.2	29	1

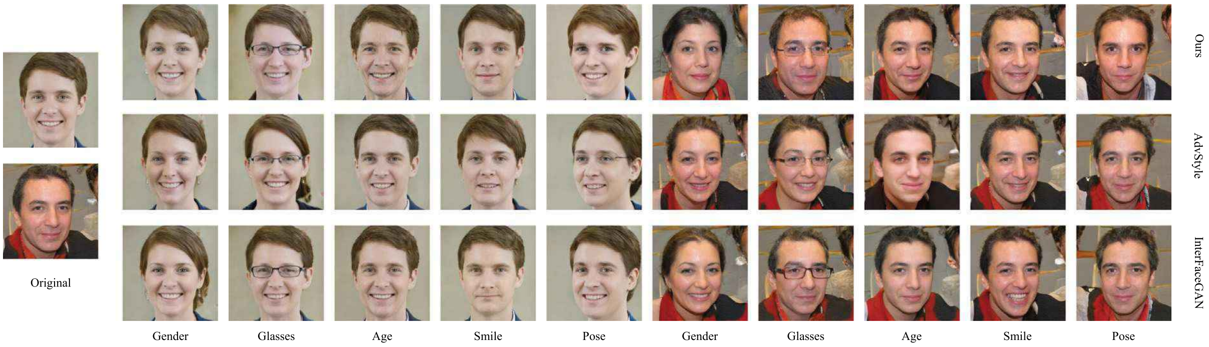


Fig. 7. Qualitative comparison of the latent semantics discovered by OTUSD and supervised methods (InterFaceGAN [7], AdvStyle [9]), including gender, glasses, age, smile, and pose.

4.3.4. *Time comparison*

Labeling data in supervised learning: The labeling process, especially for intricate tasks such as discerning semantic attributes in images, can be quite lengthy. Depending on the complexity of the dataset and other factors, it can take anywhere from days to weeks or even months. Exact quantification is challenging due to the variability in datasets and the labeling process.

Inferring semantic attributes using OTUSD: Our method drastically reduces this time. From sampling noise in the hidden space to calculating and decomposing the OT matrix, the process is completed in approximately one minute. The subsequent process of inferring semantic attributes is detailed in the "Semantic direction determination and parameter selection of singular vectors" subsection of Section 4.1. Typically, this inference concludes within 5 min, regardless of the expertise level of the user.

4.4. *Comparison with unsupervised approaches*

4.4.1. *Qualitative results*

In this test, We compared OTUSD with the unsupervised approaches, including SeFa [15], LowRankGAN [13], GANSpace [50], and the StyleGAN2 [17] model trained on FFHQ [31]. We selected the most relevant vectors that can regulate gender and smile based on their articles, and the results are depicted in Fig. 10. From the results, it can be observed that when gender is edited, SeFa modifies both fat and glass properties, while GANSpace is somewhat less effective and produces a background shift. Moreover, when using SeFa to edit smile, the fat and glass attributes vary, while the hairstyle and background of GANSpace change significantly. The change in the skin and hair color of LowRankGAN is visible to the human eye. On the other hand, OTUSD has a negligible change in other attributes when adding a smile or changing the hairstyle. Therefore, the proposed OTUSD method has a better decoupling effect compared to other unsupervised methods. The gender editing results of LowRankGAN were not presented in Fig. 10 as the code provided by LowRankGAN did not yield the semantic vector required for gender editing. In addition, the article of the study did not report any outcomes of gender editing.

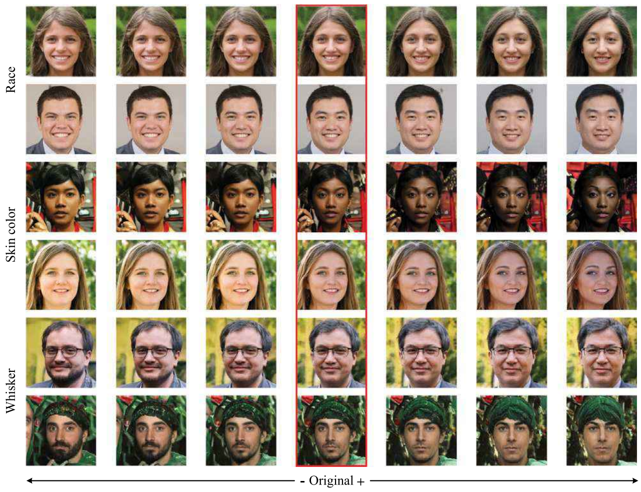
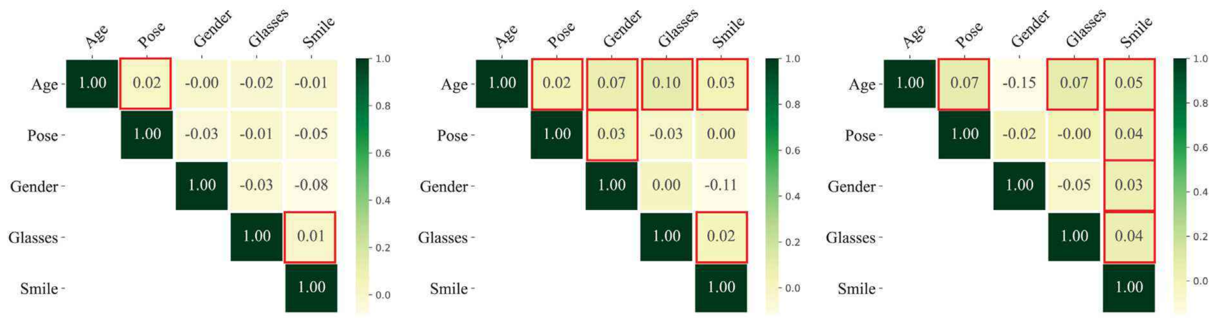


Fig. 8. Results of editing attributes of race, skin color, and whisker, where these images are generated by moving in positive or negative directions.

4.4.2. *Quantitative results*

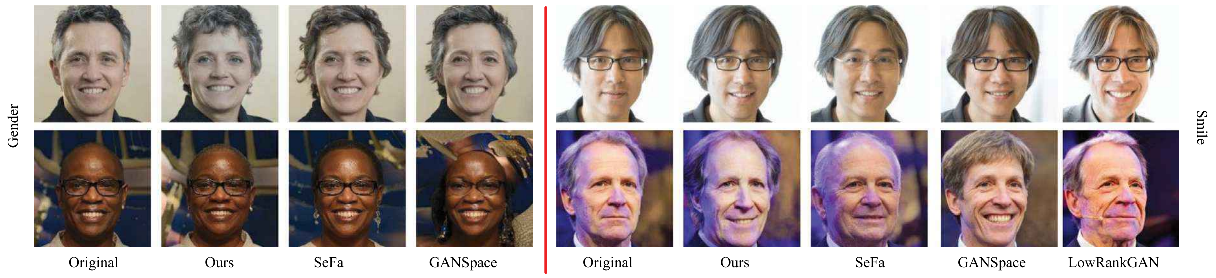
To conduct a more in-depth comparative analysis of our method, OTUSD, we employed a series of established metrics to quantitatively assess the effectiveness and accuracy of image editing, as well as the quality of the edited images and efficiency. Specifically, we selected a set of $2K$ images, edited their smile attributes, and then evaluated the results using the following metrics:

- **Inception score (IS):** IS [51] quantifies the Kullback–Leibler divergence between the conditional and marginal class distributions of generated images. A higher IS signifies enhanced quality and diversity.
- **Fréchet inception distance (FID):** FID [52] evaluates the Wasserstein distance between the multi-dimensional Gaussian distributions fitted to the feature representations of both real and generated images. A lower FID indicates a better resemblance to real images.
- **Success rate:** This metric evaluates the accuracy of semantic manipulations. A subset of images is altered along a specific vector, and a facial recognition tool, such as the LightFace API [53], determines the success of the intended alteration. The success rate signifies the percentage of images that accurately exhibit the intended manipulation.



(a) (b) (c)

Fig. 9. Correlation matrices between different attributes of the (a) OTUSD; (b) InterFaceGAN [7]; (c) AdvStyle [9].



(a) (b)

Fig. 10. Qualitative comparison between SeFa [15], GANSpace [50], LowRankGAN [13] and our method. The StyleGAN2 [17] model trained on FFHQ dataset [31] is used.

- **User study:** Human evaluators specializing in computer vision assess the quality of the edited images, providing insights into the perceptual quality that might escape quantitative metrics [13,15]. Ten scholars specializing in computer vision were enlisted to assess the editing quality of 2K pairs of original and modified images from different methods. The “User study” column denotes the proportion of images that received positive ratings for editing quality.
- **Time:** This metric elucidates the duration required to compute the set of semantic directions.

Table 1 carefully outlines the comparative results between OTUSD and several benchmark methods, including SeFa [15], LowRankGAN [13], and GANSpace [50]. OTUSD surpasses all comparative methods in IS, attaining a score of 3.52, demonstrating its superior ability to edit images with heightened quality and diversity. LowRankGAN is the closest competitor with an IS of 3.46. In the FID metric, OTUSD also excels, registering the lowest value of 19.48, suggesting that the edited images by OTUSD bear a closer resemblance to real images, with SeFa closely following at 19.51. Regarding the success rate, OTUSD leads with 94.2%, with GANSpace following at 93.0%. In the user study, 29% of evaluators found the images edited by OTUSD to be realistic, the highest among the compared methods, with SeFa at 27%. Regarding the time metric, OTUSD exhibits outstanding efficiency, determining the set of semantic directions in merely one minute, equivalent to SeFa, and significantly faster than LowRankGAN, which requires 68 min. In conclusion, the exemplary performance of OTUSD across various metrics unequivocally demonstrates its characteristics of efficiency, accuracy, and high quality in image attribute editing, establishing it as a leading method in the field.

4.5. Real image manipulation

To enable semantic editing of real images, we used the state-of-the-art GAN inversion method [36] to obtain the latent code of a real image. This method projects the real image to the latent space of a pre-trained StyleGAN2 generator and then uses the OTUSD method to obtain the disentangled editable semantic directions of the real image. Fig. 11 shows examples of these directions, such as adding smiles and glasses, which suggests the usability of the OTUSD for real image editing.

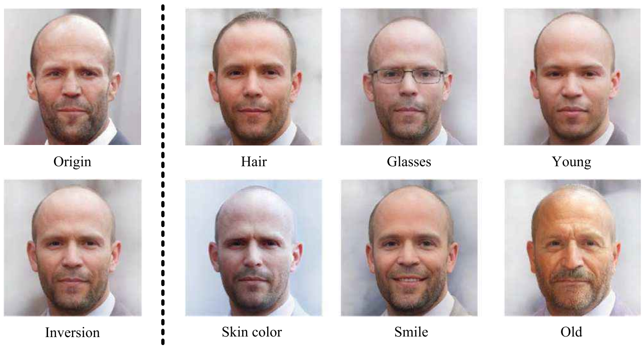


Fig. 11. Real image manipulation.

Therefore, it can be demonstrated that OTUSD can be used for editing not only fake images generated by GANs but also real images, which has practical value in real image editing applications.

5. Discussion

In this section, we elucidate the distinctive merits of OTUSD in comparison to other methods, highlight its practical applications in fields such as computer graphics, medicine, and security, and address its limitations while suggesting potential avenues for future research.

The merits of OTUSD over existing methods: OTUSD leverages OT theory to effectively disentangle various semantic attributes in the latent space of GANs. This approach ensures that the modification of one attribute does not inadvertently affect others, leading to more accurate and efficient image editing. Compared with the supervised training way, our method eliminates the need for labeled data, reducing the resource requirements, and making it more adaptable to various datasets. Additionally, OTUSD demonstrates greater robustness to variations in training data and can uncover more nuanced and diverse attribute representations. Compared with other unsupervised methods, our method exhibits superior performance in terms of image quality and editing accuracy, as demonstrated by lower FID scores and higher user study satisfaction rates.

Real-world applications and practical implications: OTUSD unlocks numerous applications, particularly in computer graphics, medicine, and security. It enables artists to create realistic and diverse content through precise image editing. In medicine and security, OTUSD is pivotal for discovering semantic directions of rare data points like medical anomalies. This ability facilitates synthetic data augmentation by generating additional data points, enriching datasets, and enhancing the robustness and performance of machine learning models in various scenarios.

Limitations and prospects for future work: While OTUSD makes notable advancements in unsupervised semantic disentanglement, it faces limitations with uncommon semantic meanings, requiring users to locate specific sample images, a time-consuming task. This adds complexity to discovering the semantic direction. Future work could explore strategies to streamline the identification of rare semantic meanings, potentially through more efficient search algorithms or automatic categorization of uncommon semantics.

6. Conclusion

In this work, we introduced the OTUSD algorithm, a novel approach to interpreting the latent space of GANs for image attribute manipulation. Extensive experiments underscored the robust performance of OTUSD and its precise control in generation with well-trained GAN models, highlighting its potential across various applications owing to its generality and efficiency. OTUSD is distinguished by its adaptability, ability to handle diverse attribute representations, and superiority over other unsupervised methods. However, it encounters challenges with uncommon semantic meanings, which signals avenues for future exploration to streamline the identification of such semantics. Despite these challenges, the advancements made by OTUSD pave the way for significant real-world applications and lay the groundwork for further research in this domain.

CRediT authorship contribution statement

Yunqi Liu: Conceptualization, Methodology, Validation, Investigation, Data curation, Writing – original draft, Writing – review & editing. **Xue Ouyang:** Conceptualization, Methodology, Writing – review & editing. **Tian Jiang:** Validation, Formal analysis. **Hongwei Ding:** Investigation, Visualization. **Xiaohui Cui:** Supervision, Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

I shared a link to my code in the body of the article.

Acknowledgments

The numerical calculations in this paper have been done on the super-computing system in the Supercomputing Center of Wuhan University. This work was supported in part by the Science and Technology Department of Yunnan Province 2022KZ00125, in part by Wuhan University Specific Fund for Major School-level International Initiatives.

References

[1] Z. Liu, M. Li, Y. Zhang, C. Wang, Q. Zhang, J. Wang, Y. Nie, Fine-grained face swapping via regional GAN inversion, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 8578–8587.
[2] H. Pehlivan, Y. Dalva, A. Dundar, Styleres: Transforming the residuals for real image editing with stylegan, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 1828–1837.
[3] H. Ding, L. Chen, L. Dong, Z. Fu, X. Cui, Imbalanced data classification: A KNN and generative adversarial networks-based hybrid approach for intrusion detection, Future Gener. Comput. Syst. 131 (2022) 240–254.
[4] Q. Song, G. Li, S. Wu, W. Shen, H.S. Wong, Discriminator feature-based progressive GAN inversion, Knowl.-Based Syst. 261 (2023) 110186.
[5] C. Yang, Y. Shen, B. Zhou, Semantic hierarchy emerges in deep generative representations for scene synthesis, Int. J. Comput. Vis. 129 (5) (2021) 1451–1466.
[6] D. Jiang, D. Song, R. Tong, M. Tang, StyleIPSB: Identity-preserving semantic basis of stylegan for high fidelity face swapping, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 352–361.
[7] Y. Shen, C. Yang, X. Tang, B. Zhou, Interfacegan: Interpreting the disentangled face representation learned by gans, IEEE Trans. Pattern Anal. Mach. Intell. (2020).
[8] A. Voynov, A. Babenko, Unsupervised discovery of interpretable directions in the gan latent space, in: International Conference on Machine Learning, PMLR, 2020, pp. 9786–9796.
[9] H. Yang, L. Chai, Q. Wen, S. Zhao, Z. Sun, S. He, Discovering interpretable latent space directions of gans beyond binary attributes, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 12177–12185.
[10] J. Jäger, R.V. Krems, Universal expressiveness of variational quantum classifiers and quantum kernels for support vector machines, Nature Commun. 14 (1) (2023) 576.
[11] R. Abdal, P. Zhu, N.J. Mitra, P. Wonka, Styleflow: Attribute-conditioned exploration of stylegan-generated images using conditional continuous normalizing flows, ACM Trans. Graph. 40 (3) (2021) 1–21.
[12] Y. Jiang, Z. Huang, X. Pan, C.C. Loy, Z. Liu, Talk-to-edit: Fine-grained facial editing via dialog, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 13799–13808.
[13] J. Zhu, R. Feng, Y. Shen, D. Zhao, Z.J. Zha, J. Zhou, Q. Chen, Low-rank subspaces in GANs, in: Advances in Neural Information Processing Systems, 2021.
[14] Y. Liu, Q. Li, Q. Deng, Z. Sun, Towards spatially disentangled manipulation of face images with pre-trained StyleGANs, IEEE Trans. Circuits Syst. Video Technol. (2022) 1.
[15] Y. Shen, B. Zhou, Closed-form factorization of latent semantics in gans, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 1532–1540.
[16] L. Zhang, H. Yang, T. Qiu, L. Li, AP-GAN: Improving attribute preservation in video face swapping, IEEE Trans. Circuits Syst. Video Technol. PP (99) (2021) 1.
[17] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, T. Aila, Analyzing and improving the image quality of stylegan, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 8110–8119.
[18] Y. Wu, R. Wang, M. Gong, J. Cheng, Z. Yu, D. Tao, Adversarial UV-transformation texture estimation for 3D face aging, IEEE Trans. Circuits Syst. Video Technol. 32 (7) (2022) 4338–4350.
[19] M. Fahim Sikder, Bangla handwritten digit recognition and generation, in: Proceedings of International Joint Conference on Computational Intelligence, Springer, 2020, pp. 547–556.
[20] B. Ji, T. Chen, Generative adversarial network for handwritten text, 2019, arXiv preprint arXiv:1907.11845.
[21] S. Khodadadeh, S. Ghadar, S. Motiian, W.A. Lin, L. Bölöni, R. Kalarot, Latent to latent: A learned mapper for identity preserving editing of multiple face attributes in stylegan-generated images, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 3184–3192.
[22] Y. Wang, Y. Hu, J. Yu, J. Zhang, Gan prior based null-space learning for consistent super-resolution, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 37, No. 3, 2023, pp. 2724–2732.
[23] A. Brock, J. Donahue, K. Simonyan, Large scale GAN training for high fidelity natural image synthesis, in: International Conference on Learning Representations, 2018.
[24] D. An, AE-OT: A new generative model based on extended semi-discrete optimal transport, in: Proceedings of the 8th International Conference on Learning Representations, 2020.
[25] A. Mondino, S. Suhr, An optimal transport formulation of the Einstein equations of general relativity, J. Eur. Math. Soc. 25 (3) (2022) 933–994.
[26] S. Eckstein, M. Nutz, Convergence rates for regularized optimal transport via quantization, Math. Oper. Res. (2023).
[27] I. Redko, N. Courty, R. Flamary, D. Tuia, Optimal transport for multi-source domain adaptation under target shift, in: The 22nd International Conference on Artificial Intelligence and Statistics, PMLR, 2019, pp. 849–858.

[28] N. Bonneel, J. Digne, A survey of optimal transport for computer graphics and computer vision, in: Computer Graphics Forum, Vol. 42, No. 2, Wiley Online Library, 2023, pp. 439–460.

[29] B. Taşkesen, S. Shafieezadeh-Abadeh, D. Kuhn, Semi-discrete optimal transport: Hardness, regularization and numerical solution, Math. Program. 199 (1–2) (2023) 1033–1106.

[30] T. Karras, M. Aittala, S. Laine, E. Härkönen, J. Hellsten, J. Lehtinen, T. Aila, Alias-free generative adversarial networks, Adv. Neural Inf. Process. Syst. 34 (2021) 852–863.

[31] T. Karras, S. Laine, T. Aila, A style-based generator architecture for generative adversarial networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 4401–4410.

[32] P. Isola, J.Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1125–1134.

[33] Y. Lu, Y.W. Tai, C.K. Tang, Attribute-guided face generation using conditional cyclegan, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 282–297.

[34] W. Xia, Y. Zhang, Y. Yang, J.H. Xue, B. Zhou, M.H. Yang, Gan inversion: A survey, IEEE Trans. Pattern Anal. Mach. Intell. (2022).

[35] Y. Alaluf, O. Patashnik, D. Cohen-Or, Restyle: A residual-based stylegan encoder via iterative refinement, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 6711–6720.

[36] T. Wang, Y. Zhang, Y. Fan, J. Wang, Q. Chen, High-fidelity GAN inversion for image attribute editing, 2021, arXiv e-prints, arXiv–2109.

[37] J. Zhu, Y. Shen, D. Zhao, B. Zhou, In-domain gan inversion for real image editing, in: European Conference on Computer Vision, Springer, 2020, pp. 592–608.

[38] A. Cherepkov, A. Voynov, A. Babenko, Navigating the gan parameter space for semantic image editing, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 3671–3680.

[39] N. Lei, D. An, Y. Guo, K. Su, S. Liu, Z. Luo, S.T. Yau, X. Gu, A geometric understanding of deep learning, Engineering 6 (3) (2020) 361–374.

[40] X. Gu, N. Lei, S.T. Yau, Optimal transport for generative models, in: Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision, Springer, 2021, pp. 1–48.

[41] J. Mi, C. Ma, L. Zheng, M. Zhang, M. Li, M. Wang, WGAN-CL: A Wasserstein GAN with confidence loss for small-sample augmentation, Expert Syst. Appl. 233 (2023) 120943.

[42] C. Villani, Optimal Transport: Old and New, Vol. 338, Springer, 2009.

[43] Z. Zeng, S. Zhang, Y. Xia, H. Tong, PARROT: Position-aware regularized optimal transport for network alignment, in: Proceedings of the ACM Web Conference 2023, 2023, pp. 372–382.

[44] M. Cuturi, Sinkhorn distances: Lightspeed computation of optimal transport, Adv. Neural Inf. Process. Syst. 26 (2013).

[45] D. Paul, S. Chakraborty, S. Das, Robust principal component analysis: A median of means approach, IEEE Trans. Neural Netw. Learn. Syst. (2023).

[46] T. Karras, T. Aila, S. Laine, J. Lehtinen, Progressive growing of GANs for improved quality, stability, and variation, in: Proceedings of the International Conference on Learning Representations, 2018.

[47] J. Deng, W. Dong, R. Socher, L.J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255.

[48] G. Branwen, Danbooru2019: A large-scale crowdsourced and tagged anime illustration dataset, 2019.

[49] F. Yu, A. Seff, Y. Zhang, S. Song, T. Funkhouser, J. Xiao, Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop, 2015, arXiv preprint arXiv:1506.03365.

[50] E. Härkönen, A. Hertzmann, J. Lehtinen, S. Paris, Ganspace: Discovering interpretable gan controls, Adv. Neural Inf. Process. Syst. 33 (2020) 9841–9850.

[51] R. Mishra, K. Sharma, R. Jha, A. Bhavsar, NeuroGAN: image reconstruction from EEG signals via an attention-based GAN, Neural Comput. Appl. 35 (12) (2023) 9181–9192.

[52] A.V. Nadimpalli, A. Rattani, ProActive DeepFake detection using GAN-based visible watermarking, ACM Trans. Multimed. Comput. Commun. Appl. (2023).

[53] S.I. Serengil, A. Ozpinar, HyperExtended LightFace: A facial attribute analysis framework, in: 2021 International Conference on Engineering and Emerging Technologies, ICEET, IEEE, 2021, pp. 1–4.