GANTASTIC: GAN-based Transfer of Interpretable Directions for Disentangled Image Editing in Text-to-Image Diffusion Models

Yusuf Dalva Hidir Yesiltepe Pinar Yanardag
Virginia Tech
{ydalva, hidir, pinary}@vt.edu

Project webpage: https://gantastic.github.io

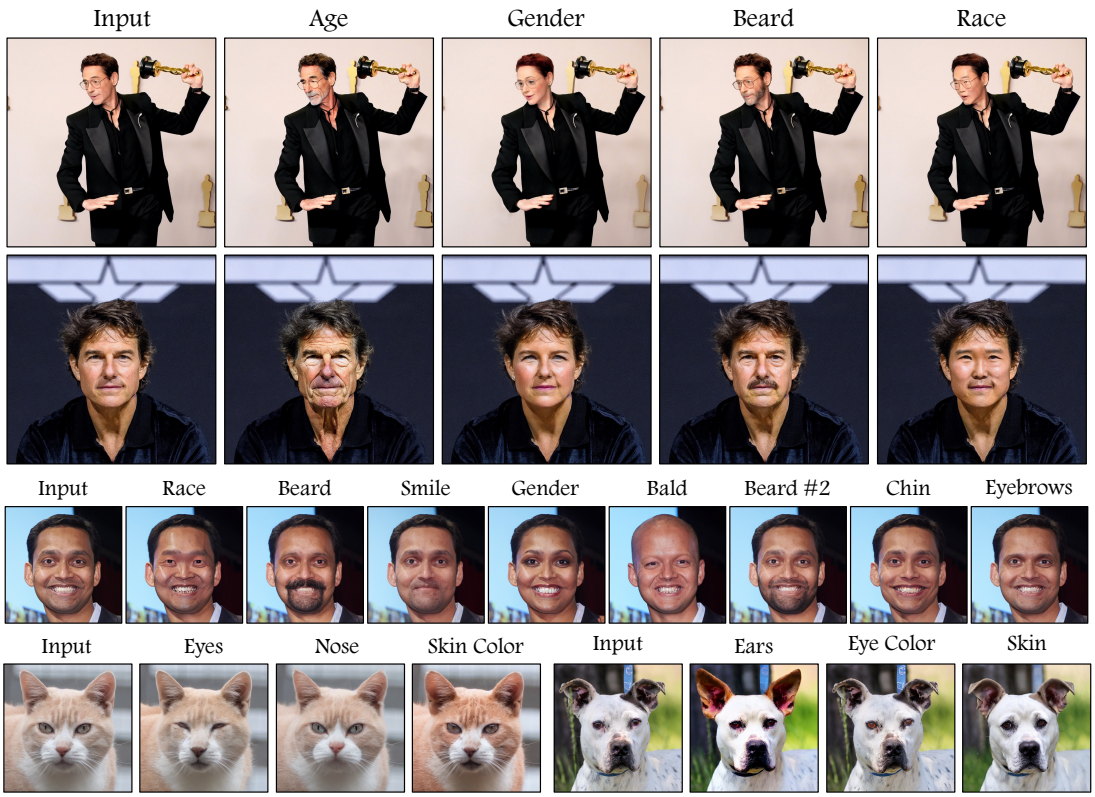

Figure 1. GANTASTIC is a novel framework that transfers interpretable directions from pre-trained GAN models directly into diffusion-based models to enable disentangled and controllable image editing.

Abstract

The rapid advancement in image generation models has predominantly been driven by diffusion models, which have demonstrated unparalleled success in generating high-fidelity, diverse images from textual prompts. Despite their success, diffusion models encounter substantial challenges in the domain of image editing, particularly in executing disentangled edits—changes that target specific attributes of an image while leaving irrelevant parts untouched. In contrast, Generative Adversarial Networks (GANs) have been recognized for their success in disentangled edits through their interpretable latent spaces. We introduce GANTASTIC, a novel framework that takes existing directions from pre-trained GAN models—representative of specific, controllable attributes—and transfers these directions into diffusion-based models. This novel approach not only maintains the generative quality and diversity that diffusion models are known for but also significantly enhances their capability to perform precise, targeted image edits, thereby leveraging the best of both worlds.

1. Introduction

Denoising Diffusion Models (DDMs) [13] and Latent Diffusion Models (LDMs) [32] gained popularity in generative modeling landscape, boasting the capability to generate high-quality, high-resolution images across diverse domains. Their prowess, particularly highlighted by text-to-image models like Stable Diffusion [32], has spurred researchers to leverage them for image editing tasks. These tasks range from text prompt-driven edits to modifications based on scribbles or segmentation maps [47], underpinning a growing interest in harnessing DDMs and LDMs for nuanced image manipulation.

Central to image editing in generative models is the concept of disentangled semantics application, where edits target specific image areas semantically, without altering unintended portions [26, 45]. In this vein, Generative Adversarial Networks (GANs) have shown exceptional proficiency in disentangled image editing, attributed to their structured latent spaces. This has catalyzed extensive research into identifying and utilizing latent directions in GANs for both supervised and unsupervised disentangled editing [10, 35, 46]. Although DDMs and LDMs excel in image generation, GANs surpass them in terms of disentanglement and editing precision, thanks to their more interpretable latent spaces. Identifying semantically meaningful directions within GANs, for instance, through principal component analysis of latent vectors, is a relatively straightforward task [10]. However, mapping similar disentangled directions within diffusion models is inherently more complex due to their design, which involves independent forward noise estimation and the management of numerous latent variables across multiple recursive timesteps.

To bridge this gap, we introduce `GANTASTIC`, a novel approach that marries the disentangled editing capabilities of GANs with the generative excellence of large-scale text-to-image diffusion models. By transferring the well-defined, disentangled directions identified within GANs to diffusion-based models, we present a solution that leverages the best of both worlds. This integration not only enhances the diffusion models' capacity for precise, semantically significant editing but also expands their applicability in domains where generating suitable text prompts might be challenging.

Our contributions are outlined as follows:

- To the best of our knowledge, our approach represents the first study to transfer directions from a pre-trained GAN model to a pre-trained text-to-image diffusion model without finetuning.
- Our approach showcases the capability to transfer a wide range of fine-grained directions spanning various categories, including faces, cats and dogs.
- The directions we have identified are notably disentangled and can be applied together without interfering with each other.
- Our experiments show that our method achieves disentangled editing results that are competitive with state-of-the-art diffusion-based and GAN-based image editing techniques.
- We share our source code along with the discovered directions to enable further research in this area.

2. Related Work

Latent Space Exploration of GANs. Various techniques have been developed that utilize the latent space of GANs for image manipulation, as demonstrated by research such as [5, 6, 30]. Supervised approaches frequently depend on pre-trained attribute classifiers to steer the optimization process, enabling the identification of significant directions within the latent space. Alternatively, these methods might use labeled data to develop classifiers specifically designed to learn desired directions [8, 35]. In contrast, there have been studies showing the feasibility of discovering semantically meaningful directions in the latent space without supervision [16, 33, 39, 41, 46]. More recent investigations into GAN-based latent space exploration are increasingly focusing on the application of image-text alignment techniques such as StyleCLIP [19, 29].

Latent Space Exploration of Diffusion Models. Diffusion-based image generation models, capable of synthesizing images across various domains, encapsulate semantically rich information within their latent representations. To harness this rich semantic content, research efforts have focused on leveraging the latent space's encoded semantics. Building on the exploration of latent spaces, certain studies [21, 42] have explored image editing techniques that alter the backward diffusion path through learned latent variable representations. Specifically, [21] bases its approach on the features learned by the denoising model's bottleneck block, whereas [42] modifies latent variables for specific target domains using stochastic diffusion models. Further advancing this field, [28] introduced a method to identify latent-specific directions that encapsulate various semantics, drawing inspiration from latent space explorations in GANs. Although this method shows promise in identifying directions within single-domain diffusion models like DDPMs, it encounters limitations when applied to large-scale diffusion models, such as Stable Diffusion. Additionally, [24] proposes decomposing images into composable energy functions that represent certain concepts, such as lighting or camera position. A more recent work, NoiseCLR [4] proposes an unsupervised method to find disentangled directions in Stable Diffusion.

Image Editing with Diffusion Models. Interest in using diffusion models for image editing tasks has been on the rise within the image generation field. A typical strategy involves using text prompts to specify the desired edits. However, this often leads to entangled edits, where changes unintentionally affect areas of the image beyond the intended target. Notable exceptions, such as the studies by [11, 47], show enhanced precision in the editing process. For example, ControlNet [47] employs a conditional diffusion model, enabling users to modify specific attributes of an image by setting conditions. Similarly, [40] achieves edits that preserve the original content by finely adjusting the diffusion model to the input image. Furthermore, [9, 15, 27, 43] present techniques for accurate reconstruction of the input image, facilitating edits that preserve content with classifier-free guidance. While these methods excel at maintaining the original image during edits, they require optimization for each image, which hinders their application in real-time editing scenarios. Recently, [42] explored modifying the denoising steps of a stochastic diffusion model for more efficient editing tasks. Although these approaches offer the promise of realistic edits, crafting the perfect editing prompt remains a challenge, often compromising the realism of the edits or their fidelity to the original image. To overcome issues with flexibility, [1, 23] suggested breaking down the editing process into multiple steps. Nonetheless, these approaches struggle with applying multiple edits simultaneously, leading to intertwined outcomes when various modifications are applied to the same image. Recent studies, such as [4], have succeeded in applying disentangled edits in large models such as Stable Diffusion. However, the unsupervised nature of their methodology allows for the identification of only a limited set of directions, thus constraining their versatility.

Combining GAN and Diffusion Models. [38] aims to use a pre-trained text-to-image diffusion model as a training objective for adapting a GAN generator to another domain. However, they do not transfer directions between GAN models and diffusion models, but show that generators can be shifted into new domains indicated by text prompts. On the other hand, previous studies, such as $\mathcal{W}+$ Adapter [22] and Concept Sliders [7], have also explored leveraging the disentangled image editing capabilities of the StyleGAN model. However, these approaches significantly differ than ours in terms of both their main goals and methodologies. Concept Sliders [7] focuses on finetuning the Stable Diffusion model using LoRAs [14] in order to learn specific concepts. This process involves using paired images or text prompts to train separate LoRA models for every concept. One of the use-cases they demonstrated is using a pre-trained StyleGAN to generate paired images, and then training LoRA models with these images.

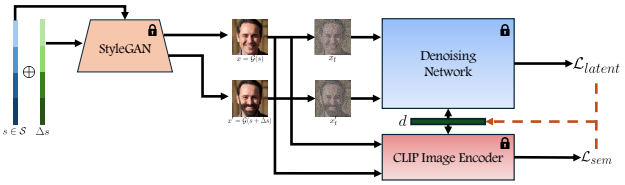

Figure 2. **GANTASTIC framework.** After generating a set of N images using StyleGAN, denoted as $G(s)$, and their edited versions, denoted as $G(s + \Delta s)$, our framework learns a latent direction d that reflects the edits introduced by Δs (e.g. beard) to the pre-trained diffusion model. To effectively learn such a latent direction, we utilize both the denoising network used by the diffusion model, and the CLIP [31] Image Encoder.

Unlike this method, our approach involves transferring directions from StyleGAN to a single pre-trained Stable Diffusion model, eliminating the need for finetuning or training separate LoRA models. This allows our model to apply any transferred direction without the computational burden of maintaining several finetuned LoRA models. The $\mathcal{W}+$ Adapter [22] seeks to utilize the StyleGAN model for image editing on individual images. Unlike our approach, which learns directions applicable to any given image, the $\mathcal{W}+$ Adapter finetunes the Stable Diffusion so that it can edit an image on $w+$ vectors of StyleGAN and *transfer the edit per image* to diffusion model. In contrast, our method can *transfer directions* from StyleGAN to diffusion model which can perform edits within diffusion model.

3. Method

In this section, we describe out proposed method GANTASTIC, in which we utilize in order to transfer semantic directions observed in GAN-based generative models.

3.1. StyleGAN

The generation process of StyleGAN2 consists of several latent spaces, namely $\mathcal{Z}$, $\mathcal{W}$, $\mathcal{W}+$ and $\mathcal{S}$. More formally, let $\mathcal{G}$ denote a generator acting as a mapping function $\mathcal{G} : \mathcal{Z} \rightarrow \mathcal{X}$ where $\mathcal{X}$ is the target image domain. The latent code $\mathbf{z} \in \mathcal{Z}$ is drawn from a prior distribution $p(\mathbf{z})$, typically chosen to be Gaussian. The $\mathbf{z}$ vectors are transformed into an intermediate latent space $\mathcal{W}$ using a mapper function consisting of 8 fully connected layers. The latent vectors $\mathbf{w} \in \mathcal{W}$ are then transformed into channel-wise style parameters, forming the *style space*, denoted $\mathcal{S}$, which is the latent space that determines the style parameters of the image. This particular space provides extensive editing possibilities, with each style channel governing a specific attribute modification, such as *smile, eye color,* or *hair type*. Essentially, this means that targeted adjustments to the channel-wise style parameters can facilitate precise, disentangled alterations to

an image. In our work, we use the directions in style space $\mathcal{S}$ identified by previous work [36, 44].

3.2. Denoising Probabilistic Diffusion Models

Diffusion models, as detailed in works by [13, 32, 37], are a type of generative model that creates data samples via an iterative denoising procedure, commonly referred to as the reverse process. This process operates over a sequence of noise levels $t \in \{1, ..., T\}$, with $\epsilon^t = \alpha^t \epsilon$ where ϵ is drawn from a normal distribution $\mathcal{N}(0, 1)$. The role of the denoising network, denoted as ϵ_θ, is to predict the noise component ϵ in the noised image x_t during the reverse process. Here, x_t symbolizes the noised variant of the original image x_0, subjected to a noise level of ϵ^t. The training of such a denoising network revolves around an objective function that is structured as follows:

$$\mathcal{L}_{DM} = \mathbb{E}_{x_0, \epsilon^t \sim \mathcal{N}(0,1), t}\left[||\epsilon^t - \epsilon_\theta(x_t, t)||_2^2\right] \quad (1)$$

To produce an image with the denoising network ϵ_θ, the reverse process begins with an initial input x_T, which is sampled from a normal distribution $\mathcal{N}(0, 1)$. During the reverse diffusion procedure, the variable x_t undergoes a series of iterative denoising steps to gradually approach x_0, for each noise level t ranging from 1 to T. This iterative denoising process is mathematically represented by Eq. 2, which is defined for a given step size γ and a specific timestep t.

$$x_{t-1} = x_t - \gamma \epsilon_\theta(x_t, t) + \xi, \; \xi \sim \mathcal{N}(0, \sigma_t^2 I) \quad (2)$$

Classifier-free guidance, introduced by [12], facilitates conditioned sampling by making nuanced modifications to both the forward and reverse diffusion processes based on a specific condition c. By adapting the training of the denoising network ϵ_θ to be compatible with classifier-free guidance, it becomes feasible to generate images conditionally. This is achieved by adjusting the standard noise prediction $\epsilon_\theta(x_t)$ to incorporate the condition, resulting in a conditional noise prediction denoted as $\tilde{\epsilon_\theta}(x_t, c)$. For the sake of clarity, the notation $\epsilon_\theta(x_t)$ is used here to indicate the predicted noise at timestep t, with the understanding that t is implicitly indicated by the variable x_t. The formulation for the noise prediction under classifier-free guidance, $\tilde{\epsilon_\theta}(x_t, c)$, is given by:

$$\tilde{\epsilon_\theta}(x_t, c) = \epsilon_\theta(x_t, \phi) + \lambda_g(\epsilon_\theta(x_t, c) - \epsilon_\theta(x_t, \phi)) \quad (3)$$

where λ_g is guidance scale and ϕ is null-text.

3.3. Learning Objective

Our proposed method GANTASTIC aims to learn a latent direction d formulated as a conditional embedding, that represents the edit performed by Δs on the style space of StyleGAN. In order to achieve this task, we initially populate a dataset consisting N image pairs corresponding to images generated by StyleGAN, $\mathcal{G}(s)$, and their edited co-variants, $\mathcal{G}(s + \Delta s)$. Throughout our framework we label the these image sets as $\mathcal{X}_{input} = \{x_1, \cdots, x_N\}$ and $\mathcal{X}_{edited} = \{x'_1, \cdots, x'_N\}$. Using these image sets, we formulate our overall loss function to learn the latent direction d with two objectives, which targets both the semantic level differences and latent level differences between the image pairs. These two objectives are described below.

Latent Alignment Loss. With the objective of learning a latent direction d that effectively reflects the difference between the image sets $\mathcal{X}_{input}$ and $\mathcal{X}_{edited}$, we utilize a latent alignment objective $\mathcal{L}_{latent}$ based on the denoising network included in the pre-trained latent diffusion model, which is Stable Diffusion in our case. As a preliminary step, we first perform the forward diffusion step to the image pair (x, x') for $t \in \mathcal{U}(1, T)$ timesteps to obtain (x_t, x'_t), where $x \in \mathcal{X}_{input}$ and $x' \in \mathcal{X}_{edited}$. Followingly, we formulate our latent alignment objective s.t. the direction d maximizes the difference between the noise predictions performed on x_t and x'_t. Furthermore, we formulate $\mathcal{L}_{latent}$ as the difference between the loss predictions $\epsilon_\theta(x_t, d)$ and $\epsilon_\theta(x'_t, d)$ with a negative sign, to maximize this difference. We formulate our latent alignment objective in Eq. 4.

$$\mathcal{L}_{latent} = -\mathbb{E}_{x_0, \epsilon^t \sim \mathcal{N}(0,1), t}\left[||\epsilon_\theta(x'_t, d) - \epsilon_\theta(x_t, d)||_2^2\right] \quad (4)$$

Semantic Alignment Loss. In addition to the objective of obtaining a latent direction that maximizes the difference between $\mathcal{X}_{input}$ and $\mathcal{X}_{edited}$ from the latent representations, we also desire to learn a latent direction that semantically aligns with the difference between these image sets. To achieve this objective, we utilize CLIP [31] Image Encoder (E_I) in our semantic alignment objective. Fundamentally, for a latent direction that represents the difference between the image sets $\mathcal{X}_{input}$ and $\mathcal{X}_{edited}$, the similarity between $x' \in \mathcal{X}_{edited}$ and the direction d should be maximized whereas it should be minimized for $x \in \mathcal{X}_{input}$. To reflect this behavior in our overall optimization objective, we use the difference between the similarity values of x' and x with direction d, where we maximize the similarity value for x' and minimize for x. This way, we enforce our method to learn the desired semantic change only from the paired images generated by the StyleGAN generator. We provide the semantic alignment loss $\mathcal{L}_{sem}$ in Eq. 5.

$$\mathcal{L}_{sem} = 1 - cossim(E_I(x'), d) + cossim(E_I(x), d) \quad (5)$$

We formulate the overall learning objective $\mathcal{L}$ to learn the latent direction d, representing the translation from $\mathcal{X}_{input}$

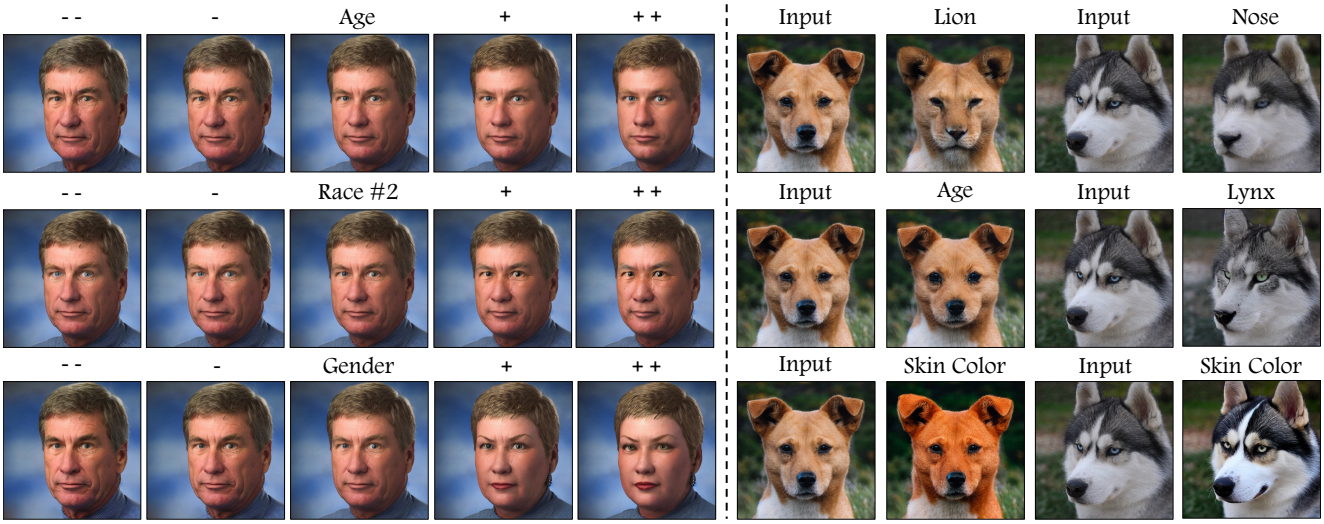

Figure 3. **Capabilities of GANTASTIC.** The proposed framework can successfully learn latent directions from a variety of domains including human faces and dog images. Additionally, GANTASTIC enables users to adjust the intensity of the editing effect through a scaling parameter. This functionality gives users the flexibility to either tone down or intensify the impact of a given editing direction. For instance, in the case of the *Gender* edit, users can lessen the effect for a more *masculine* appearance or enhance it for a more *feminine* look by applying a negative or positive scale, respectively.

to $\mathcal{X}_{edited}$ in Eq. 6.

$$\mathcal{L} = \mathcal{L}_{sem} + \mathcal{L}_{latent} \tag{6}$$

which combines semantic alignment loss and latent alignment loss to transfer the GAN direction.

3.4. Image Editing

Given a latent direction d_i, we perform image editing in a way that it successfully reflects the desired semantic to the input image in a disentangled manner. Our editing scheme relies mainly on the findings of classifier-free guidance, which can utilize condition c to perform conditional image generation as defined in Eq. 3. To extend classifier-free guidance such that one can perform image editing with minimal adjustments, we add an editing term based on d, which is the latent direction we learn during the optimization process. Briefly, we add an additional classifier-free guidance term corresponding to editing in which we formulate as Eq. 7 for the timesteps where the edit is going to be applied. We denote the editing scale with λ_e, the image condition with c and the editing direction with d.

$$\bar{\epsilon_\theta}(x_t, c, d) = \tilde{\epsilon_\theta}(x_t, c) + \lambda_e(\epsilon_\theta(x_t, d) - \epsilon_\theta(x_t, \phi)) \tag{7}$$

Editing with multiple directions. To extend our editing scheme for multiple edits, we perform a sum over multiple editing directions with their corresponding editing scales. For a set of directions $D = \{d_1, \cdots, d_k\}$, which are going to be applied at a given timestep, we expand Eq. 7 with a summation term over direction set D in Eq. 8.

$$\tilde{\epsilon_\theta}(x_t, c) + \sum_{i=1}^{|D|} \lambda_{e_i}(\epsilon_\theta(x_t, d_i) - \epsilon_\theta(x_t, \phi)) \tag{8}$$

Real Image Editing. In addition to performing edits on generated images, the directions learned by GANTASTIC can also be used to edit real images. To do so, we adopt DDPM Inversion [15] to obtain an initial latent representation x_T instead of sampling from $\mathcal{N}(0, I)$. After inverting x_T unconditionally, we reformulate the overall noise prediction $\bar{\epsilon_\theta}(x_t, d)$ for real image editing, where d denotes the editing direction, in Eq. 9.

$$\bar{\epsilon_\theta}(x_t, d) = \epsilon_\theta(x_t, \phi) + \lambda_e(\epsilon_\theta(x_t, d) - \epsilon_\theta(x_t, \phi)) \tag{9}$$

Note that the approach for editing with multiple directions is also applicable to real image editing, by replacing the edit guidance term $\lambda_e(\epsilon_\theta(x_t, d) - \epsilon_\theta(x_t, \phi))$ with the summation provided in Eq. 8.

4. Experiments

To evaluate the effectiveness of GANTASTIC in transferring semantically meaningful latent directions and to demonstrate our method's generalizability, we conducted assessments across various domains, such as human faces, cats and dogs. Additionally, we benchmarked our method

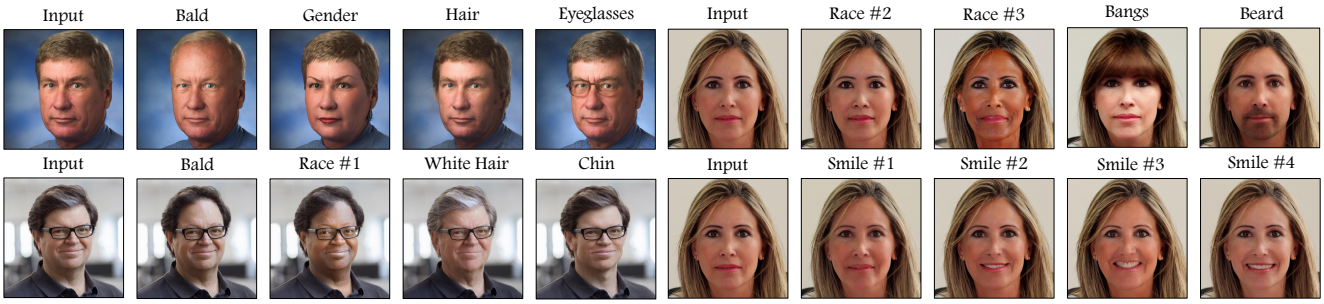

Figure 4. **Qualitative Results**. GANTASTIC successfully transfers editing directions that modify the overall look, including changes in *race* or *aging*, as well as more detailed edits that target specific facial attributes, such as *eyeglasses* or a *beard*. GANTASTIC can also distinguish among various edits for the same feature underlines the versatility of our approach, providing users with an extensive selection of editing options for individual characteristics, like multiple smile designs (see row 2) or styles of baldness (as shown in Rows 1 and 2).

against state-of-the-art image editing methods in diffusion-based models and compared with latent direction methods in both diffusion-based and GAN-based methods.

Experimental Setup. We used Stable Diffusion-v1.5 for all of our experiments. We used StyleGAN2 [18] trained on several diverse datasets including FFHQ [17], AFHQ-Cats [3], and AFHQ-Dogs [3]. In our default setting, we train GANTASTIC with $N = 100$ samples for each latent direction we would like to transfer. To optimize the directions, we use a learning rate of 5e-3 and batch size of 8 for AdamW optimizer [25] for 1000 iterations. To ensure the reproducibility of our experiments, we conduct all experiments with a fixed random seed of 0. Moreover, training our method on a single domain requires approximately 30 minutes to transfer a direction, and once trained, performing any edit in a zero-shot manner takes about 5 seconds using a single NVIDIA L40 GPU.

4.1. Ablation Studies

We conducted ablation studies focusing on the number of timesteps, the quantity of samples from the GAN direction intended for transfer, and the distinct components of our loss function.

Ablation on timesteps. Existing research, such as [43] and [11], has identified timesteps as critical elements influencing the ability of Stable Diffusion to achieve disentangled edits. Our approach involves adjusting the noise prediction across a specific range of timesteps to attain more distinct edits. Fig. 8 (a) presents an ablation study on timesteps for the *Asian* and *Young* directions. The results indicate that applying the edit across all timesteps leads to significant changes in facial structure, even though the edit itself is successful. However, implementing the edit at $0.4T$ not only successfully applies the edit but does so in a highly disentangled manner.

Ablation on sample size. We also conducted an ablation study to assess the impact of the number of images sampled from GAN directions during the learning phase (see Fig. 8 (b)). Our findings reveal that GANTASTIC can successfully learn directions with as few as $N = 10$ samples. Moreover, increasing the sample size to $N = 100$ appears to yield slightly more disentangled results.

Ablation of loss terms. Finally, we conducted an ablation study focusing on the individual loss terms, as illustrated in Figure 8 (c). Employing solely the semantic alignment loss (indicated as 'w/o $\mathcal{L}_{latent}$') demonstrates the capability to learn the desired edit. However, this approach results in entangled outcomes, significantly altering the facial structure. On the other hand, utilizing both loss terms in conjunction (indicated with 'with $\mathcal{L}_{latent}$') leads to highly disentangled edits, ensuring the edits are consistent with the facial structure and background.

4.2. Qualitative Results

Our method leverages a single pre-trained diffusion model to transfer latent directions across different domains. Given the significant variability in facial features and the popularity of face editing in both GAN and diffusion-based models, we initially explore the face editing potential of directions uncovered by GANTASTIC. As illustrated in Fig. 1 and Fig. 4, our technique showcases a range of editing capabilities, from comprehensive changes that alter the face's overall appearance, such as *race* or *aging*, to more fine-grained adjustments targeting specific facial features, like *eyeglasses* or a *beard*. Our approach can transfer a variety of editing directions to the same region, for example, different styles of bangs, hairstyles, or degrees of baldness. Notably, our method is capable of distinguishing between very detailed variations within the same editing task; for instance, when provided with four distinct GAN directions for varying smile edits, GANTASTIC successfully learned to differentiate between them. This highlights our method's

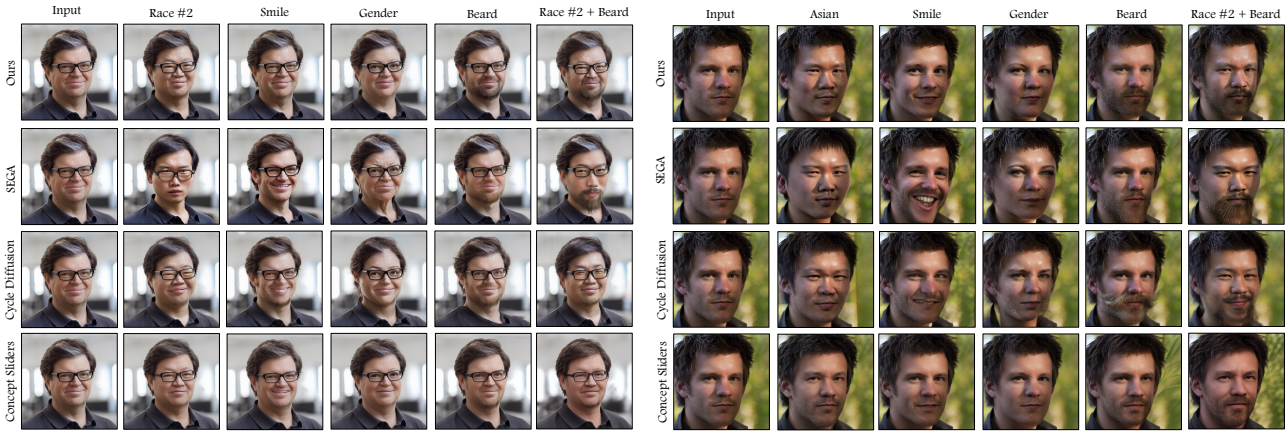

Figure 5. **Qualitative Comparison with Diffusion-based Image Editing Methods** We compare our approach with Concept Sliders [7], SEGA [1], Cycle-Diffusion [42]. The qualitative outcomes demonstrate that GANTASTIC outperforms the aforementioned methods in achieving disentangled image edits and in identifying detailed latent directions.

adaptability, offering users a wide range of options for a single attribute, such as various smile types (row 2) or baldness styles (refer to rows 1 and 2). We also note that a key feature of our edits is their high degree of disentanglement, ensuring that only the targeted modifications are made without altering other unrelated aspects. Beyond facial editing, our method's effectiveness extends to other domains, including cats and dogs. The qualitative results depicted in Fig. 3 demonstrate GANTASTIC's ability to grasp and apply a wide array of semantic variations across different domains.

Interpolating Edits. Our technique provides users with the ability to fine-tune the editing effect via a scale parameter. Demonstrated in Fig. 3, this capability facilitates edits across both negative and positive scales, enabling users to either mitigate or enhance the editing direction's impact. In the context of the 'Gender' edit, for instance, users have the option to lessen the feminine traits or accentuate them by applying a scale in a positive direction. Furthermore, our method accomplishes these interpolations in a distinct and isolated manner, ensuring that edits, whether in positive or negative directions, accurately preserve the essence of the original image. It's worth mentioning that our approach necessitates only one of the directions from GAN, either positive or negative, eliminating the need to generate both scales applied to a GAN direction for transfer. Despite this, our method effectively transfers the specified GAN direction and is capable of interpolating between positive and negative scales.

Qualitative Comparison with Diffusion-based Image Editing Methods. We benchmark our approach against several recent image editing methods, including Concept Sliders [7], SEGA [1], Cycle-Diffusion [42]. Notably, SEGA often result in substantial alterations to the input image for edits like 'Asian'. Similarly, Cycle-Diffusion faces challenges in achieving disentangled edits, noticeably altering the age of the input image in attempts to edit features such as 'Smile' and 'Beard'. As illustrated in Fig. 5, GANTASTIC surpasses these alternatives in maintaining semantic accuracy and in its ability to execute disentangled edits. Concept Sliders face challenges in applying multiple edits simultaneously, such as combining Race and Beard modifications, resulting in significant deviations from the original input image.

Qualitative Comparison with Diffusion-based Latent Direction Methods. We conduct comparisons with recent methods that identify latent directions within the Stable Diffusion model. Diffusion Pullback [28] introduces an unsupervised approach for discovering latent directions in diffusion-based models, employing the pullback metric for this purpose. Another recent method, NoiseCLR [4], employs a contrastive learning framework to uncover directions without supervision. Given the unsupervised nature of both methods, we selected three overlapping directions for comparison: 'Gender', 'Old', and 'Race'. Fig. 6 showcases a comparative analysis between our approach, Diffusion Pullback (referred to as D-Pullback), and NoiseCLR. The comparison reveals that Diffusion Pullback often results in significant alterations to the input image (such as Race edit), as acknowledged in their study [28]. While NoiseCLR shows performance on par with our method, its unsupervised approach inherently restricts it to a limited number of discoverable directions.

Qualitative Comparison with GAN-based Latent Direction Methods. GAN-based editing techniques are recognized for their exceptional editing abilities, attributed to their disentangled latent spaces [45]. In our comparative

Figure 6. **Qualitative Comparison with Diffusion-based Latent Direction Methods** We compare our method with Diffusion Pullback [28] and NoiseCLR [4].

analysis, we evaluate GANTASTIC alongside leading GAN-based methods capable of identifying directions within the latent space, including StyleCLIP [29] that finds directions via text-based prompts, and unsupervised methods LatentCLR [46], GANSpace [10], and SeFa [34] (see Fig. 7). The comparison reveals that our diffusion-based approach delivers results that are on par with those of GAN-based models.

4.3. Quantitative Results

For the comparison of edit effectiveness, we selected the semantic attributes "Asian," "Smile," "Gender," and "Beard" which were chosen from among the capabilities common to all competing methods. We utilize LPIPS [48] scores to evaluate the degree of similarity retained with the original image distribution after editing. The results for various edits are presented in Table 1. The LPIPS metrics reveal that our method consistently performs lower scores compared to other methods, indicating superior fidelity in maintaining the original image's integrity throughout the editing process.

Re-scoring Analysis. Following [4, 35], we conduct a re-scoring analysis to assess how CLIP classification probabilities for specific attributes change following an edit. The rows in Table 2 correspond to various editing directions—Asian, Smile, Gender, and Beard—applied to 100 images, while the columns show the resulting shifts in CLIP scores. Consistent with expectations, performing a targeted edit boosts the probability of the image being classified under that attribute. For example, an Asian edit enhances the image's likelihood of being identified as Asian by 53.6%, with similar increases observed for other attributes, as detailed in the diagonal entries of Table 2. Additionally, some edits naturally impact related attributes; for instance, enhancing the Gender attribute towards femininity notably reduces the Beard attribute probability. The interaction between Beard and Smile attributes also demonstrates a degree of interdependence, which can be attributed to inherent biases within the SD model, where adding beard diminish the presence of smile attribute. Our approach notably supports disentangled editing, as it minimally affects the scores of unrelated attributes when making specific edits.

Method	Asian	Smile	Gender	Beard
SEGA [1]	0.23	0.22	0.18	0.17
Cycle Diff. [42]	**0.09**	0.06	**0.11**	0.12
Sliders [7]	0.19	0.14	0.14	0.22
Ours	**0.09**	**0.05**	**0.11**	**0.11**

Table 1. **LPIPS scores (lower is the better).** Our method is able to achieve lower LPIPS than the other methods, indicating greater content preservation during image editing.

	Asian	Smile	Gender	Beard
Asian	**53.6**	15.8	-12.1	-3.5
Smile	-20.4	**41.2**	11.8	-7.2
Gender	2.0	-4.3	**94.7**	-19.8
Beard	-2.6	-8.1	-0.05	**28.3**

Table 2. **Re-scoring Analysis.** GANTASTIC can perform edits efficiently on several attributes. The attributes edited are shown as the rows whereas the measured attributes are shown as the columns.

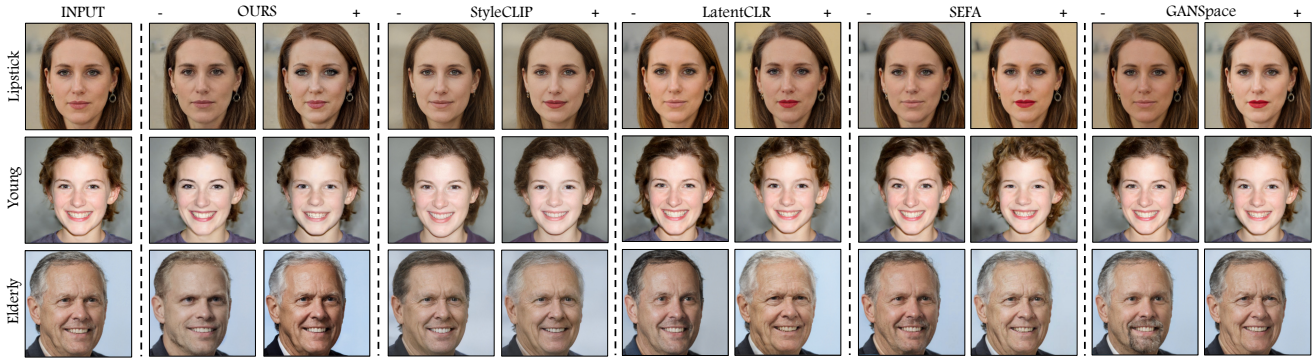

Figure 7. **Qualitative Comparisons with GAN-based Latent Direction Discovery.** Additionally, GANTASTIC is evaluated against methods for discovering latent directions in GANs. Our findings show that the editing and direction discovery capabilities of GANTASTIC are on par with those of GAN-based approaches, especially in the context of detailed face editing.

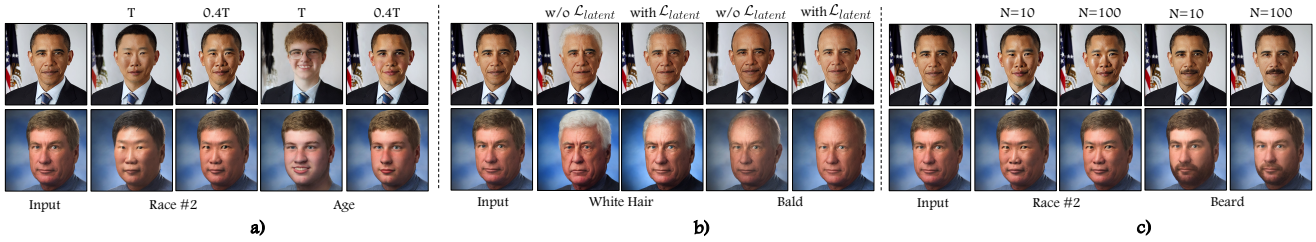

Figure 8. **Ablation Study.** We perform ablations to assess the effectiveness of three different variables, which are the selection of editing timesteps (a), effect of proposed loss terms (b) and the number of sample used for training (c). For each of our ablations we present qualitative results on two different edits, where their corresponding labels are provided for easy understanding.

User Study. Furthermore, we conducted a user study with 50 participants on the Prolific.com. Participants were shown a series of edits made using common semantics by each method under comparison. They were then asked to judge whether they deemed the edit successful in conveying the intended semantic and if the edit was executed in a disentangled manner. Participants rated each question on a scale from 1 to 5, with 5 representing the highest level of satisfaction. Our method earned an average rating of **3.36** and achieves the highest score in comparison to the competing methods, where SEGA [1], Cycle Diffusion [42] and Concept Sliders [7] have earned the scores 1.96, 2.65 and 3.25, respectively.

5. Limitations and Broader Impact

The effectiveness of our method is contingent on the quality of the directions transferred from GANs. Furthermore, due to the inherent biases present in both CLIP and Stable Diffusion, our model tends to learn entangled edits for the Age attribute. For example, when the model transfers directions for white hair or eyeglasses, it often conflates these features with older age, a correlation also noted in previous research [44]. This tendency underscores the challenge of disentangling certain attributes due to the biases embedded within the training data of CLIP and Stable Diffusion, highlighting a critical area for improvement in our method. Similarly to other image synthesis and editing technologies, our method also carries the risk of being misused for malicious purposes a concern widely discussed in the context of tools like those described by [20]. Despite these potential drawbacks, our approach significantly enhances the precision of edits on images. This increased control opens up new avenues for creative expression, potentially democratizing image editing for a broader audience.

6. Conclusion

In this paper, we introduce a novel approach that capitalizes on the strengths of GANs, known for their disentangled latent spaces and powerful manipulation capabilities, and harmonizes them with the exceptional image generation abilities of diffusion models. Our method aims to bring the best of the both worlds, transferring directions from GAN models to exploit the generative capacity of text-to-image diffusion models like Stable Diffusion. This strategic combination not only delivers editing capabilities that are competitive in both diffusion-based and GAN-based image editing techniques but also significantly refines the precision of the image generation process.

References

[1] Manuel Brack, Felix Friedrich, Dominik Hintersdorf, Lukas Struppek, Patrick Schramowski, and Kristian Kersting. SEGA: Instructing text-to-image models using semantic guidance. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. 3, 7, 8, 9

[2] Manuel Brack, Felix Friedrich, Katharina Kornmeier, Linoy Tsaban, Patrick Schramowski, Kristian Kersting, and Apolinário Passos. Ledits++: Limitless image editing using text-to-image models. *arXiv preprint arXiv:2311.16711*, 2023. 1, 3, 4, 5

[3] Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. Stargan v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020. 6

[4] Yusuf Dalva and Pinar Yanardag. Noiseclr: A contrastive learning approach for unsupervised discovery of interpretable directions in diffusion models. *arXiv preprint arXiv:2312.05390*, 2023. 2, 3, 7, 8

[5] Yusuf Dalva, Said Fahri Altındiş, and Aysegul Dundar. Vecgan: Image-to-image translation with interpretable latent directions. In *European Conference on Computer Vision*, pages 153–169. Springer, 2022. 2

[6] Yusuf Dalva, Hamza Pehlivan, Oyku Irmak Hatipoglu, Cansu Moran, and Aysegul Dundar. Image-to-image translation with disentangled latent vectors for face editing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 2

[7] Rohit Gandikota, Joanna Materzyńska, Tingrui Zhou, Antonio Torralba, and David Bau. Concept sliders: Lora adaptors for precise control in diffusion models. *arXiv preprint arXiv:2311.12092*, 2023. 3, 7, 8, 9, 1, 4, 5

[8] Lore Goetschalckx, Alex Andonian, Aude Oliva, and Phillip Isola. Ganalyze: Toward visual definitions of cognitive image properties. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5744–5753, 2019. 2

[9] Ligong Han, Song Wen, Qi Chen, Zhixing Zhang, Kunpeng Song, Mengwei Ren, Ruijiang Gao, Yuxiao Chen, Di Liu, Qilong Zhangli, et al. Improving negative-prompt inversion via proximal guidance. *arXiv preprint arXiv:2306.05414*, 2023. 3

[10] Erik Härkönen, Aaron Hertzmann, Jaakko Lehtinen, and Sylvain Paris. Ganspace: Discovering interpretable gan controls. *arXiv preprint arXiv:2004.02546*, 2020. 2, 8

[11] Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Prompt-to-prompt image editing with cross attention control. *arXiv preprint arXiv:2208.01626*, 2022. 3, 6

[12] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022. 4

[13] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 2, 4

[14] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021. 3

[15] Inbar Huberman-Spiegelglas, Vladimir Kulikov, and Tomer Michaeli. An edit friendly ddpm noise space: Inversion and manipulations. *arXiv preprint arXiv:2304.06140*, 2023. 3, 5

[16] Ali Jahanian, Lucy Chai, and Phillip Isola. On the" steerability" of generative adversarial networks. *arXiv preprint arXiv:1907.07171*, 2019. 2

[17] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019. 6

[18] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020. 6

[19] Umut Kocasari, Alara Dirik, Mert Tiftikci, and Pinar Yanardag. Stylemc: multi-channel based fast text-guided image generation and manipulation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 895–904, 2022. 2

[20] Pavel Korshunov and Sébastien Marcel. Deepfakes: a new threat to face recognition? assessment and detection. *arXiv preprint arXiv:1812.08685*, 2018. 9

[21] Mingi Kwon, Jaeseok Jeong, and Youngjung Uh. Diffusion models already have a semantic latent space. *arXiv preprint arXiv:2210.10960*, 2022. 2

[22] Xiaoming Li, Xinyu Hou, and Chen Change Loy. When stylegan meets stable diffusion: a $\mathcal{W}_+$ adapter for personalized image generation. *arXiv preprint arXiv:2311.17461*, 2023. 3, 1, 2

[23] Nan Liu, Shuang Li, Yilun Du, Antonio Torralba, and Joshua B Tenenbaum. Compositional visual generation with composable diffusion models. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVII*, pages 423–439. Springer, 2022. 3

[24] Nan Liu, Yilun Du, Shuang Li, Joshua B. Tenenbaum, and Antonio Torralba. Unsupervised compositional concepts discovery with text-to-image generative models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2085–2095, 2023. 2

[25] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 6

[26] Emile Mathieu, Tom Rainforth, Nana Siddharth, and Yee Whye Teh. Disentangling disentanglement in variational autoencoders. In *International conference on machine learning*, pages 4402–4412. PMLR, 2019. 2

[27] Ron Mokady, Amir Hertz, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Null-text inversion for editing real images using guided diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6038–6047, 2023. 3

[28] Yong-Hyun Park, Mingi Kwon, Jaewoong Choi, Junghyo Jo, and Youngjung Uh. Understanding the latent space of diffusion models through the lens of riemannian geometry. *arXiv preprint arXiv:2307.12868*, 2023. 2, 7, 8

[29] Or Patashnik, Zongze Wu, Eli Shechtman, Daniel Cohen-Or, and Dani Lischinski. Styleclip: Text-driven manipulation of stylegan imagery. *arXiv preprint arXiv:2103.17249*, 2021. 2, 8

[30] Hamza Pehlivan, Yusuf Dalva, and Aysegul Dundar. Styleres: Transforming the residuals for real image editing with stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1828–1837, 2023. 2

[31] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021. 3, 4

[32] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 2, 4

[33] Yujun Shen and Bolei Zhou. Closed-form factorization of latent semantics in gans. *arXiv preprint arXiv:2007.06600*, 2020. 2

[34] Yujun Shen and Bolei Zhou. Closed-form factorization of latent semantics in gans. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1532–1540, 2021. 8

[35] Yujun Shen, Ceyuan Yang, Xiaoou Tang, and Bolei Zhou. Interfacegan: Interpreting the disentangled face representation learned by gans. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 2, 8

[36] Enis Simsar, Umut Kocasari, Ezgi Gülperi Er, and Pinar Yanardag. Fantastic style channels and where to find them: A submodular framework for discovering diverse directions in gans. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 4731–4740, 2023. 4

[37] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020. 4

[38] Kunpeng Song, Ligong Han, Bingchen Liu, Dimitris Metaxas, and Ahmed Elgammal. Stylegan-fusion: Diffusion guided domain adaptation of image generators. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 5453–5463, 2024. 3

[39] Paul Upchurch, Jacob Gardner, Geoff Pleiss, Robert Pless, Noah Snavely, Kavita Bala, and Kilian Weinberger. Deep feature interpolation for image content changes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7064–7073, 2017. 2

[40] Dani Valevski, Matan Kalman, Eyal Molad, Eyal Segalis, Yossi Matias, and Yaniv Leviathan. Unitune: Text-driven image editing by fine tuning a diffusion model on a single image. 42(4), 2023. 3

[41] Andrey Voynov and Artem Babenko. Unsupervised discovery of interpretable directions in the gan latent space. In *International Conference on Machine Learning*, pages 9786–9796. PMLR, 2020. 2

[42] Chen Henry Wu and Fernando De la Torre. A latent space of stochastic diffusion models for zero-shot image editing and guidance. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7378–7387, 2023. 2, 3, 7, 8, 9

[43] Qiucheng Wu, Yujian Liu, Handong Zhao, Ajinkya Kale, Trung Bui, Tong Yu, Zhe Lin, Yang Zhang, and Shiyu Chang. Uncovering the disentanglement capability in text-to-image diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1900–1910, 2023. 3, 6

[44] Zongze Wu, Dani Lischinski, and Eli Shechtman. Stylespace analysis: Disentangled controls for stylegan image generation. *arXiv preprint arXiv:2011.12799*, 2020. 4, 9

[45] Weihao Xia, Yulun Zhang, Yujiu Yang, Jing-Hao Xue, Bolei Zhou, and Ming-Hsuan Yang. Gan inversion: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3121–3138, 2022. 2, 7

[46] Oğuz Kaan Yüksel, Enis Simsar, Ezgi Gülperi Er, and Pinar Yanardag. Latentclr: A contrastive learning approach for unsupervised discovery of interpretable directions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14263–14272, 2021. 2, 8

[47] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3836–3847, 2023. 2, 3

[48] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 8

GANTASTIC: GAN-based Transfer of Interpretable Directions for Disentangled Image Editing in Text-to-Image Diffusion Models

Supplementary Material

In this appendix we include our supplementary material as follows:

- Additional qualitative results in Sec. S.1 and Sec. S.3.
- Comparisons with $\mathcal{W}+$ Adapter [22] and Concept Sliders [7] in Sec. S.2.
- Comparisons of edits learned and reference edits in StyleGAN in Sec. S.4.
- The training algorithm in Sec. S.5.

S.1. Additional Qualitative Results

We provide additional qualitative results in Fig. S.9 where Age, Gender, Beard and Race directions learned by our method are presented. To demonstrate the effectiveness our method, we include examples of edits performed on full-body images (Row 1), and images including multiple faces (Row 2). Even though our method does not use any masks to specify the target of the desired edits, our method successfully performs such edits on faces present in the provided images and does not interfere with the details of the input image that are irrelevant to the performed edit.

S.2. Comparisons with W+ Adapter and Concept Sliders

We provide a qualitative comparison between the $\mathcal{W}+$ Adapter and our proposed method in Fig. S.10. Our approach effectively learns latent directions from StyleGAN, such as Gender or Age, and applies these directions to any given image, as demonstrated in Fig. S.10. In contrast, the $\mathcal{W}+$ Adapter requires finetuning the Stable Diffusion model by training a separate $\mathcal{W}+$ Adapter for each image in Fig. S.10. Our results demonstrate that our edits more accurately preserve the integrity of the input image while implementing the desired edits, such as changes in Gender or Age.

Moreover, we provide qualitative comparisons with Concept Sliders [7] in Fig. S.11, Fig. S.12 and Fig. S.13 where our method surpasses [7] both in terms of disentanglement capabilities and representation quality without the need of training separate LoRA models for each direction.

S.3. Comparisons with Other Methods

We provide additional comparisons with a recent editing method, LEDITS++ [2] that requires text prompts to perform edits within Stable Diffusion model. We perform these comparisons on Beard, Gender and Race semantics, which we provide in Fig. S.11, S.12 and S.13. As can be seen from the results, our method performs more disentangled edits compared to LEDITS++.

S.4. StyleGAN Edits vs. Transferred Edits

Additionally, to demonstrate how the representations that `GANTASTIC` learns align with the latent space of StyleGAN, we demonstrate the edits performed by the directions learned by our framework and input-edit pairs sampled from StyleGAN in Fig. S.14.

S.5. Algorithm

To further clarify our training procedure for learning a latent direction d, we provide the training algorithm of `GANTASTIC` in Alg. 1.

Figure S.9. **Supplementary Editing Examples.** We provide additional editing examples to further demonstrate the effectiveness of the directions learned by GANTASTIC. Here, we demonstrate edits performed by GANTASTIC framework on full-body images (Row 1), and images with multiple faces (Row 2). Additionally, for the example with multiple faces (Row 2), editing the gender attribute amplifies the feminine traits in both of the faces.

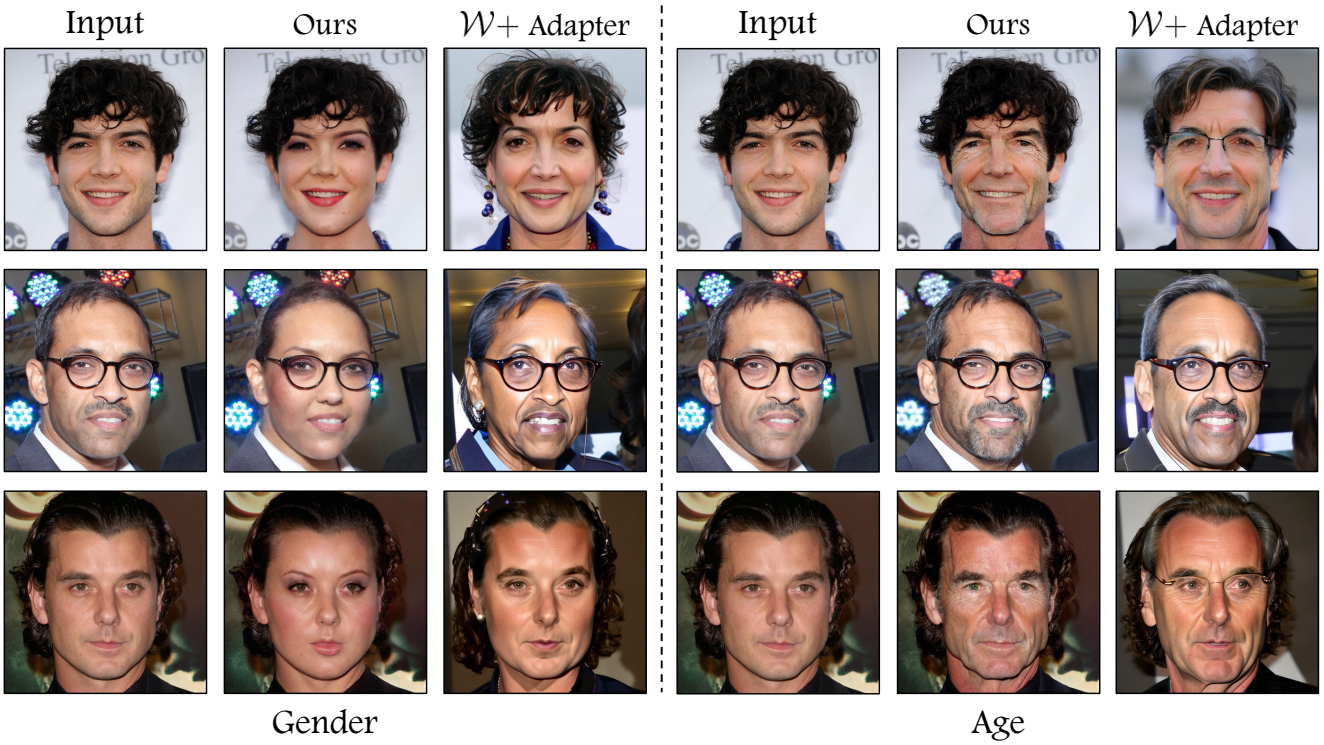

Figure S.10. **Comparisons with $\mathcal{W}+$ Adapter [22].** We compare our method with [22] as a competing approach on face editing task. Apparent from the results we provide, our method succeeds over in terms of capabilities such as content preservation (preserving the identity and details irrelevant to the edit) and disentangled editing (such as disentangling attributes like eyeglasses and age).

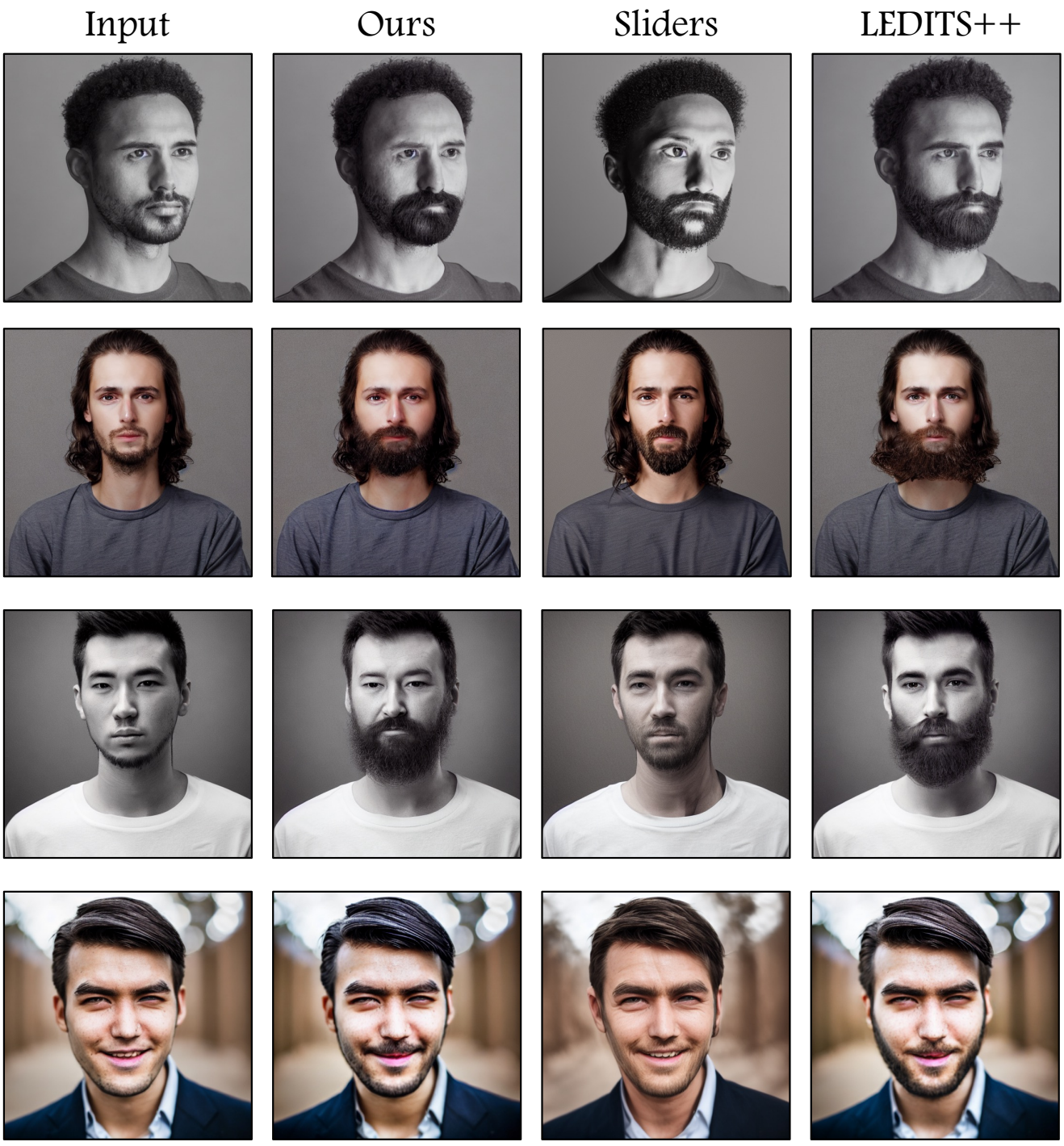

Figure S.11. **Additional Comparisons on Beard Edit.** We provide additional comparisons on beard editing task with Concept Sliders [7] and LEDITS++ [2] where we use image sliders in [7] for a fair comparison. Our editing results succeeds over the competing approaches both in terms of editing quality and content preservation. Note that [7], which learns the edit based on reference images, struggle when it attempts to add a beard to sample without any traces of the attribute.

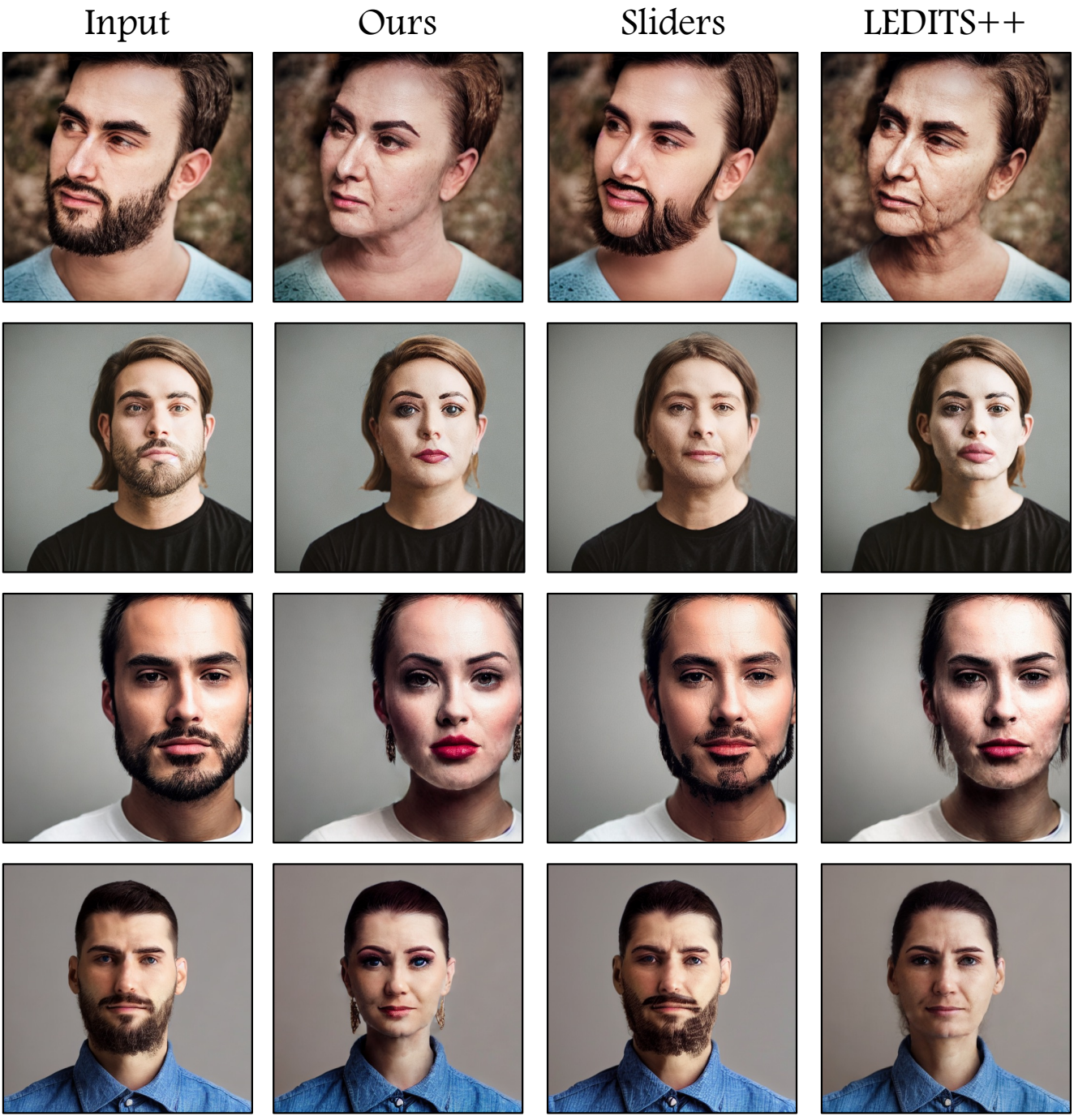

Figure S.12. **Additional Comparisons on Gender Edit.** To demonstrate the effectiveness of the gender editing direction learned by GANTASTIC, we provide additional comparisons with Concept Sliders [7] and LEDITS++ [2]. Notably, both [7] and [2] struggle with artifacts while performing such an edit that changes the overall appearance of the face, where [7] experiences it more severely. Directions learned by our method can both perform such edits without sacrificing from generation quality and in a disentangled manner.

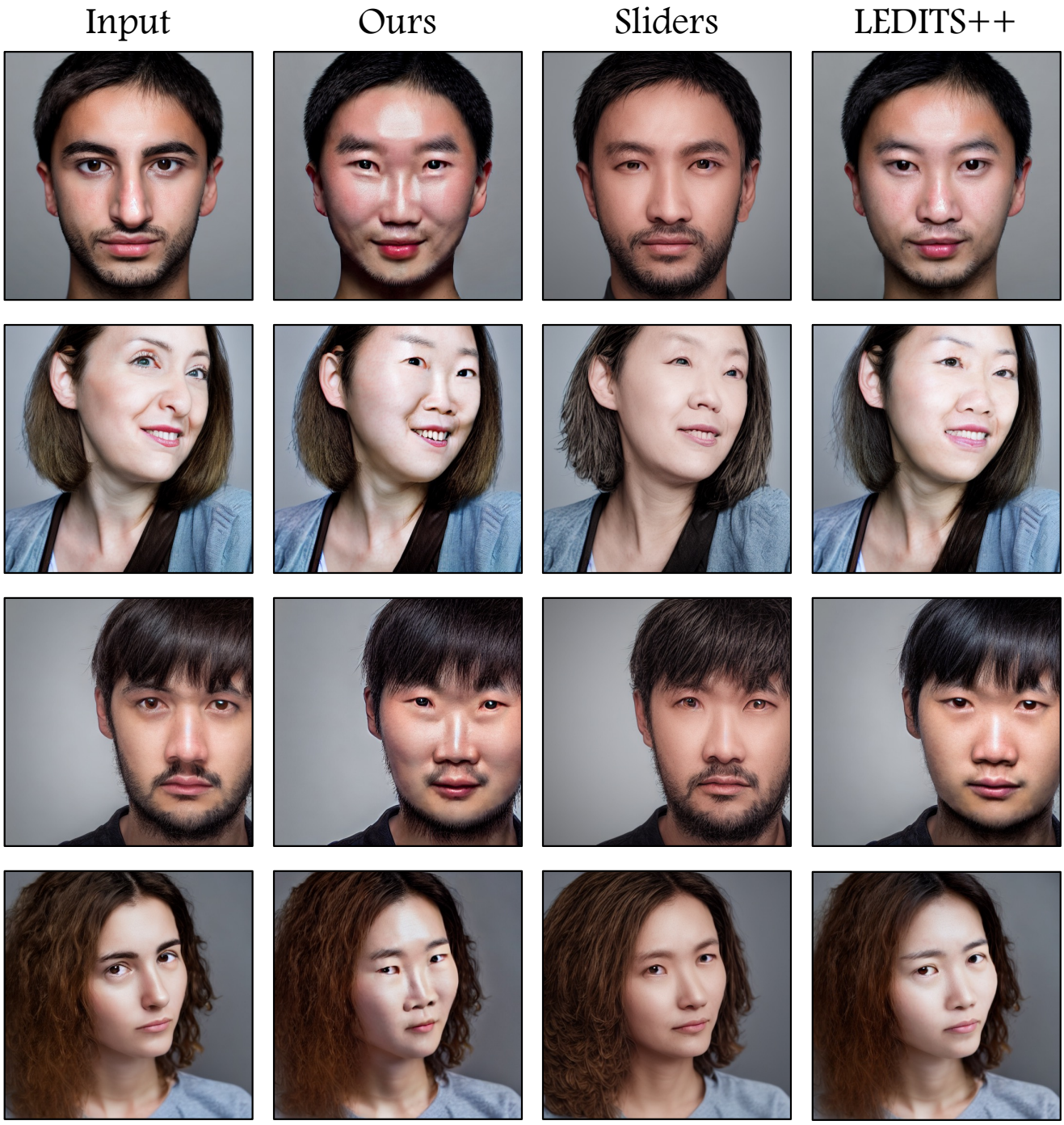

Figure S.13. **Additional Comparisons on Race #2 Edit.** Supplementary to the results we present in the main paper, we provide additional qualitative comparisons on the attribute Race #2 with Concept Sliders [7] and LEDITS++ [2]. As observed from the provided examples, our method successfully reflects the edit while preserving the identity of the input image. Note that with LoRA based approaches such as [7], image quality is sacrificed in order to apply the edit where significant changes to the input are present in the corresponding edits.

Algorithm 1 Learning direction d with GANTASTIC

Require: Pre-trained diffusion model $\epsilon_\theta(x_t, c)$, pre-trained CLIP Image Encoder $E_I(x)$, training dataset containing N latent codes $\{s_1, \cdots, s_N\}$, editing direction Δs, StyleGAN generator $\mathcal{G}(s)$, randomly initialized conditional embedding d, learning rate λ.

while training **do**:

for $i = 1, \cdots, N$ **do**

Sample $x_i = \mathcal{G}(s_i)$, $x_i' = \mathcal{G}(s_i + \Delta s)$

Initialize Gaussian Noise $\epsilon \sim \mathcal{N}(0, 1)$

Initialize a noise $\epsilon^t = \alpha^t \epsilon$ at timestep $t \sim \mathcal{U}(1, T)$

$x_{i,t} = x_i + \epsilon^t$, $x'_{i,t} = x_i + \epsilon^t$

Compute $\epsilon_{input} = \epsilon_\theta(x_{i,t}, d)$

Compute $\epsilon_{edited} = \epsilon_\theta(x'_{i,t}, d)$

$\mathcal{L}_{latent} = -||\epsilon_{edited} - \epsilon_{input}||_2^2$

$\mathcal{L}_{sem} = 1 - cossim(E_I(x_i'), d) + cossim(E_I(x_i), d)$

$\mathcal{L} = \mathcal{L}_{sem} + \mathcal{L}_{latent}$

$d = d - \lambda \nabla_{d,\mathcal{L}}$

end for

end while

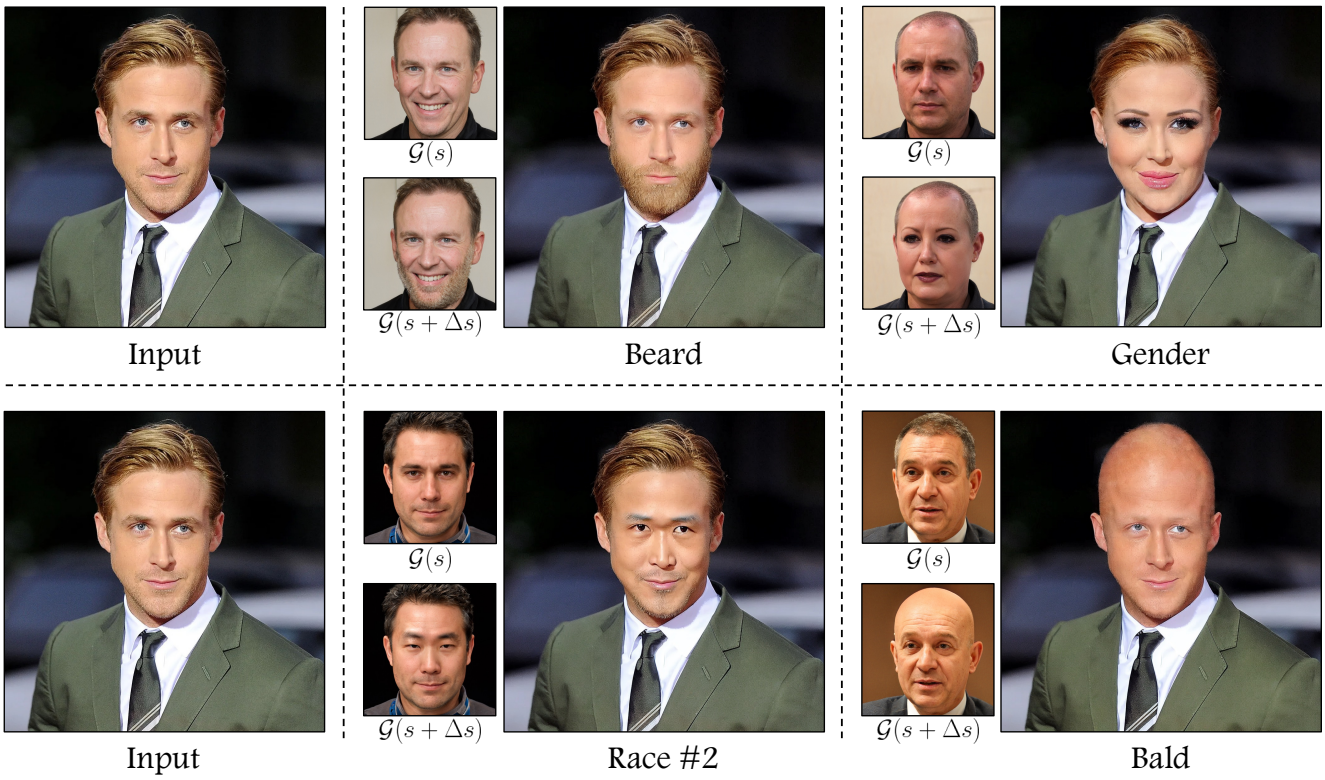

Figure S.14. **Edits learned by GANTASTIC.** We demonstrate the beard, gender, race # 2 and baldness edits above, along with reference images from the training datasets constructed using StyleGAN. Above, we show the images generated by StyleGAN as $\mathcal{G}(s)$ and their edited counter-parts as $\mathcal{G}(s + \Delta s)$, respectively. As our qualitative results also show, edits learned by GANTASTIC successfully translates the disentangled directions performed on StyleGAN to Stable Diffusion.