Everything is There in Latent Space: Attribute Editing and Attribute Style Manipulation by StyleGAN Latent Space Exploration

Rishubh Parihar  
Indian Institute of Science, Bangalore, India  
rishubp@iisc.ac.in

Ankit Dhiman  
Indian Institute of Science, Bangalore, India  
samsungresearch.india@ankitd@iisc.ac.in

Tejan Karmali  
Indian Institute of Science, Bangalore, India  
tejankarmali@iisc.ac.in

R. Venkatesh Babu  
Indian Institute of Science, Bangalore, India  
venky@iisc.ac.in

Figure 1: Examples of various attribute edits on synthetic faces and art images (Top). Example variations of attribute styles for eyeglasses and hairs generated by FLAME and edits on car dataset (Bottom).

ABSTRACT

Unconstrained Image generation with high realism is now possible using recent Generative Adversarial Networks (GANs). However, it is quite challenging to generate images with a given set of attributes. Recent methods use style-based GAN models to perform image editing by leveraging the semantic hierarchy present in the layers of the generator. We present Few-shot Latent-based Attribute Manipulation and Editing (FLAME), a simple yet effective framework to perform highly controlled image editing by latent space manipulation. Specifically, we estimate linear directions in the latent space (of a pre-trained StyleGAN) that controls semantic attributes in the generated image. In contrast to previous methods that either rely on large-scale attribute labeled datasets or attribute classifiers, FLAME uses minimal supervision of a few curated image pairs to estimate disentangled edit directions. FLAME can perform both individual and sequential edits with high precision on a diverse set of images while preserving identity. Further, we propose a novel task of Attribute Style Manipulation to generate diverse styles for attributes such as eyeglasses and hair. We first encode a set of synthetic images of the same identity but having different attribute styles in the latent space to estimate an attribute style manifold.
Sampling a new latent from this manifold will result in a new attribute style in the generated image. We propose a novel sampling method to sample latent from the manifold, enabling us to generate a diverse set of attribute styles beyond the styles present in the training set. FLAME can generate diverse attribute styles in a disentangled manner. We illustrate the superior performance of FLAME against previous image editing methods by extensive qualitative and quantitative comparisons. FLAME generalizes well on out-of-distribution images from art domain as well as on other datasets such as cars and churches. Project page.

CCS CONCEPTS
• Computer systems organization → Embedded systems; Redundancy; Robotics; • Networks → Network reliability.

KEYWORDS
GANs, Image-Editing, Latent space, Image Manipulation

ACM Reference Format:
Rishubh Parihar, Ankit Dhiman, Tejan Karmali, and R. Venkatesh Babu. 2022. Everything is There in Latent Space: Attribute Editing and Attribute Style Manipulation by StyleGAN Latent Space Exploration. In Proceedings of the 30th ACM International Conference on Multimedia (MM ’22), October 10–14, 2022, Lisboa, Portugal. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3503161.3547972

1 INTRODUCTION
Image synthesis has been one of the long-standing problems in computer vision and graphics. With the advent of deep learning, many methods have been proposed for image synthesis. Among all approaches, Generative Adversarial Networks [12] have shown promising results in generating photorealistic images. Recent StyleGAN models [16, 17] can generate images of diverse categories such as faces, cars, churches, etc., that are often indistinguishable from natural images. Although these networks generate highly realistic images, it is challenging to control this generation process. Interestingly, StyleGAN architectures have layer-wise latent codes and stochastic vectors that control image generation. However, it requires additional methods to find disentangled latent transformations to generate images with given specifications.

The latent space of StyleGAN has rich semantic properties. Methods such as InterFaceGAN [27] and GANSpace [13] demonstrate the existence of directions in latent space that controls the attributes in the generated image. For example, there are directions for pose, age, gender and smile editing. Prior works [3, 4, 13, 27, 30, 33, 35, 39] estimate linear or non-linear paths in the latent space, achieving realistic attribute editing in StyleGAN generated images. GAN encoder models [1, 5, 26, 32] learn the mapping from the image space to the latent space to foster edits on real images. However, the existing methods to estimate the attribute editing directions have two concerns, a) they require supervision from attribute classifiers trained on large data, and b) the estimated attribute directions are entangled with other attributes.

We propose a simple yet effective method to obtain disentangled attribute edit directions while using very less data: Few-shot Latent-based Attribute Manipulation and Editing (FLAME). FLAME is able to perform realistic edits for a wide variety of attributes - expression, pose, age, bangs, eye-glasses, hair style, etc. (Fig. 1). Our method requires only ten curated image pairs compared to previous methods requiring large-scale attribute annotations [27] or pre-trained attribute classifiers [3, 4, 9, 21]. Specifically, we create image pairs with a given attribute’s presence and absence. We then compute the difference between the projected latent codes for the images in a pair. Finally, we estimate the dominant direction that aligns closely with these difference directions for all the image pairs in the dataset. Due to the attribute-specific image pairs, we obtain disentangled directions which change only one specific attribute while keeping other attributes unaffected. The estimated edit directions generalize well to diverse identities compared to previous works [3, 4, 9, 21] that estimate instance-specific edits based on attribute scores. Further, we show (in Fig. 1) that directions obtained by FLAME from real images can be applied on out-of-domain artistic images. FLAME also generalizes to edits for other categories such as cars and churches.

Notably, while existing works find a direction for an attribute, they are not able to synthesize diversity within an attribute (for eg. diversity in hairstyles synthesized on a person) and is limited by the extent by which the direction is traversed. We propose a novel task of Attribute Style Manipulation (Fig. 1 Bottom), which aims to create diverse styles of a single attribute without changing other attributes and identity of the image. We propose a method that is a natural extension of our attribute editing framework to estimate the manifold of attribute styles in the latent space of a pre-trained StyleGAN. Sampling from this manifold generates images with variations in attribute styles keeping the identity and other image properties unchanged. We investigate our approach for attribute style manipulation for face images with two important face attributes: eyeglasses and hairstyle. This framework can have wide use in creating synthetic training datasets for training deep learning models for downstream applications. We summarize the main contributions of our work as follows:

- We present a simple yet effective method FLAME that estimates disentangled linear directions in the latent space of StyleGAN using supervision from few (≈ 10) image pairs to perform highly realistic image edits.
- The directions estimated by FLAME generalize to out of domain art images and other categories: cars and churches.
- To the best of our knowledge, we are the first to present a novel task of attribute style manipulation to generate diverse attribute styles, and demonstrate how FLAME can solve it.

2 RELATED WORKS

Image manipulation using GANs: Recent style-based GAN architectures [16, 17] provide hierarchical control in the generated images [37]. Multiple works [3, 7, 9, 13, 27, 31] perform fine-grained image editing by leveraging the rich structure present in the latent space of a pre-trained GAN. Another important direction of research involves training conditional GANs [24] and cycle GANs [42] to perform image editing. MaskGAN [19] learns a mapping between the segmentation mask and the rendered target to edit generated image. [34] conditions the image generator on attribute strengths to perform attribute edits. Although these methods can generate good quality image edits, they require retraining of the GAN model, which is computationally expensive for high resolution images.
In this section we present our method for estimating linear latent directions in the latent space of StyleGAN2 with few image pairs. StyleGAN2 generator \( G \) is composed of a mapping function \( G_m : \mathbb{R}^{512} \rightarrow \mathbb{R}^d \) and a synthesis function \( G_s : \mathbb{R}^d \rightarrow \mathbb{R}^{H \times W \times 3} \). Both the functions are represented as neural networks. Thus, \( G = G_s \circ G_m(z) \) where \( z \sim \mathcal{N}(0^{512}, \text{I}_{512 \times 512}) \) (which is a normal distribution with zero mean and identity covariance), \( G_m(z) \in \mathcal{W}^+ \), which intermediate latent space of StyleGAN that offers disentanglement between different semantic concepts. Given this, we define a linear model for attribute editing as \( w' = w_0 + \alpha d_j \), where \( w', w_0 \in \mathcal{W}^+ \), \( d_j \in \mathbb{R}^d \) is the direction along which attribute \( a_j \) changes; and \( \alpha \) controls the strength of the change. For editing any attribute \( a_j \), we curate a dataset \( D \) consisting of \( n \) image pairs. We describe the dataset creation procedure in detail in Sec 3.1. Our hypothesis is that with image pairs that differ in only a single attribute \( a_j \), we can estimate the direction along which \( a_j \) changes. We demonstrate and validate this hypothesis in Sec 3.2. Finally, we estimate directions for multiple styles of a single attribute and propose an algorithm to approximate the style manifold for that attribute in Sec 3.3.

**GAN encoder models:** GAN encoder models are used to learn mapping between real images to the latent space which can then be modified to perform image edits on real images. Multiple StyleGAN encoder models \([1, 2, 5, 6, 8, 26, 32, 36]\) are proposed in the literature based on the use case of editability vs reconstruction. For StyleGAN models, the original \( \mathcal{Z} \) space entangles multiple semantic concepts compared to the learned \( \mathcal{W} \) space, which is more disentangled \([17]\). Furthermore, \( \mathcal{W}^+ \) space provides more flexibility as it allows separate latent codes for each generator layer. Most GAN encoder models map the input image to this immense \( \mathcal{W}^+ \) space to obtain realistic reconstructions. Domain GAN inversion \([41]\) first performs inversion using an encoder followed by an optimization step which has a good reconstruction quality and is also semantically meaningful for editing tasks. PIE \([30]\) proposed a non-linear iterative optimization scheme to embed images in the latent space. Xu et al. \([36]\) propose an encoder model for videos that uses optical flow. Chai et al. \([8]\) trained the encoder with masked images, which results in the latent code corresponding to images while preserving the unmasked content in the input image.

**3 METHODOLOGY**

In this section, we present our method for estimating linear latent directions in the latent space of StyleGAN2 with few image pairs. StyleGAN2 generator \( G \) is composed of a mapping function \( G_m : \mathbb{R}^{512} \rightarrow \mathbb{R}^d \) and a synthesis function \( G_s : \mathbb{R}^d \rightarrow \mathbb{R}^{H \times W \times 3} \). Both the functions are represented as neural networks. Thus, \( G = G_s \circ G_m(z) \) where \( z \sim \mathcal{N}(0^{512}, \text{I}_{512 \times 512}) \) (which is a normal distribution with zero mean and identity covariance), \( G_m(z) \in \mathcal{W}^+ \), which intermediate latent space of StyleGAN that offers disentanglement between different semantic concepts. Given this, we define a linear model for attribute editing as \( w' = w_0 + \alpha d_j \), where \( w', w_0 \in \mathcal{W}^+ \), \( d_j \in \mathbb{R}^d \) is the direction along which attribute \( a_j \) changes; and \( \alpha \) controls the strength of the change. For editing any attribute \( a_j \), we curate a dataset \( D \) consisting of \( n \) image pairs. We describe the dataset creation procedure in detail in Sec 3.1. Our hypothesis is that with image pairs that differ in only a single attribute \( a_j \), we can estimate the direction along which \( a_j \) changes. We demonstrate and validate this hypothesis in Sec 3.2. Finally, we estimate directions for multiple styles of a single attribute and propose an algorithm to approximate the style manifold for that attribute in Sec 3.3.
3.1 Synthetic Pair Creation

For a given attribute $a_j$, we find a direction $d_j \in \mathcal{W}^+$ such that traversing along $d_j$ alters only $a_j$ while keeping other attributes intact. To estimate the direction, we use a dataset consisting of $n$ image pairs $\mathcal{D}$. An image pair $\mathcal{D}^k = \{I_p^k, I_n^k\}$ has a positive image $I_p^k$, which contains the attribute $a_j$, and a negative image $I_n^k$ which does not contain $a_j$. Due to lack of availability of such paired datasets which have variation along a single attribute, we create a synthetic dataset which satisfies this property.

We start by randomly sampling a set of $m$ negative images ($I_n^k$) and $m$ source images ($I_p^k$) (having $a_j$ present) with their part-wise segmentation mask ($M^k$) from CelebAMask-HQ dataset [20]. Thereafter, we use a simple cut and paste approach to create the corresponding positive image ($I_p^k$). Specifically, given a $I_p^k$ and $M^k$, we choose the part-mask $M^k(j)$ that contains the regions corresponding to the attribute $a_j$. For example, mouth and hair region contains the attributes of expressions and bangs respectively. We blend $I_p^k$ and $I_n^k$ using $M^k(j)$ to obtain the positive image $I_p^k$ using Eq. 1 and as shown in 2-I. Note that we do not have to perform alignment of images ($I_p^k$ and $I_n^k$) as CelebAMask-HQ dataset has all eye-aligned images. The resulting positive image $I_p^k$ differs from the negative image $I_n^k$ only in $a_j$; all other attributes are unchanged. Finally, from the generated pairs (30) image pairs $\mathcal{D}$, we manually select 10 image pairs by discarding unnatural looking positive images.

$$I_p = (1 - M) \circ I_n + M \circ I_s$$

We create synthetic image pairs using this method for all the attributes located at different parts of the face (e.g., eyeglasses, smile, wearing-hat, adding-hair, bangs, facial hair, eye-close). However, not all the attributes can be transferred in this way and therefore we obtain positive-negative image pairs for these attributes in a different manner. We flip the negative image for the pose attribute and the negative images respectively and $k \in \{1, 2, \ldots, n\}$, we project it into the $\mathcal{W}^+$ latent space using StyleGAN2 encoder $E$ [8].

After projection, we obtain a dataset $L$ consisting of $n$ pairs of latent codes of the form $L_k = (w_p^k, w_n^k)$, where $w_p^k = E(I_p^k)$ and $w_n^k = E(I_n^k)$. We then compute the difference direction for each latent pair as $d_p^k = w_p^k - w_n^k$ and normalize it to unit length. Note that, all the $d_p^k$ vectors correspond to the same attribute edit but from different image pairs. We want to estimate a direction $d_j$ which aligns closely with all of these difference vectors $d_p^k$.

$$d_j = \arg\max_k \sum_{k=1}^n (d_p^k, d_j)^2$$

To solve the above optimization problem, we create a matrix $A$ by stacking all the $d_p^k$ vectors as the rows. We then compute the Singular Value Decomposition of the matrix $A$ to obtain: $A = U \Sigma V^T$. The column vector $v_1$ of $V$ matrix associated with the highest singular value will maximize the given optimization function (see supplementary material).

3.2 Semantic Direction Estimation

We give an overview of our method for direction estimation in Fig. 2-II. As shown, our method relies on creation of attribute specific positive-negative image pairs, which we have described in Sec. 3.1. Having such a pair $D_k = \{I_p^k, I_n^k\}$ where $I_p^k$ and $I_n^k$ are the positive

Figure 3: a) Attribute Style Manifold: All the attribute style directions $v_j$’s lie on the unit sphere (pink shaded) are projected onto the tangent hyperplane $P$ at $v_1$. b) Hyperplane $P$ where the primitive directions $v_k$’s are estimated by taking a difference between $p_k$’s and $p^*$

3.3 Attribute Style Manipulation

We introduce a novel task of Attribute Style Manipulation and propose an algorithm to perform such attribute style edits with high fidelity. Current editing methods are limited to adding/removing any attribute or changing the attribute’s strength such as age. However, for certain attributes such as hair, multiple styles exist, but the current method only alters the length of the hairs. To this end, we estimate a manifold for various styles for a given attribute to generate different styles. Images having diverse styles of an attribute can be sampled from this manifold, while keeping the others unchanged.

We sample $S$ positive images with different styles for the attribute $a_j$. We then estimate the direction for each style following the procedure given in 3.1 and 3.2. We denoted the estimated directions
for $S$ attribute styles (for $a_j$) as $v_k$ for $k \in \{1, 2, \ldots, S\}$ and use them to find a manifold for different styles. After this, we estimate the dominant direction $v^*$ that aligns with all the normalized $v_k$'s by solving the optimization problem similar to Eq. 2. As the $v_k$'s and $v^*$ are normalized to unit length, we shift them to origin so that they lie on the surface of a unit sphere as shown in Fig. 3-a. To find a new attribute style, we wish to sample vectors on the surface of this sphere in the neighborhood of $v_k$'s. However, it is challenging to directly sample a vector from the desired region on the sphere.

To this end, we first compute a tangent hyperplane $P$ to the sphere at point $v^*$ and extend all the $v_k$ vectors up to $P$ to obtain the intersection points $p_k$ and $p^*(=v^*)$ as shown in Fig. 3-a. To sample a point on the sphere, we can sample a point on the hyperplane $P$ and then project it back onto the sphere’s surface by normalizing it to a unit length. Hence, we estimate the primitive vectors $u_k$’s lying on $P$ using Eq. 3 by subtracting the intersections $p_k$’s from $p^*$ as shown in Fig. 3-b. We take a linear combination of $u_k$’s to sample a point $b$ on the hyperplane $P$ as given in Eq. 4. $b$ is then projected back onto the unit sphere by normalizing it and thus, it is now a new sampled point on the sphere surface. Note that we wish to sample from the neighborhood of the vectors $v_k$ hence we sample small values for the weights $\lambda_i$ from the range of $(-e, e)$. Finally, for any desired image $I$ for which attribute style variation is to be generated, we first project it to $W^+$ as $w = E(I)$, and manipulate it as $w' = w + z b$. We explore other method to sample $b$ - convex combination of $mathbb{v}_k$ and modifying the attribute strengths $mathbb{v}_k$ in the supplementary material.

$$u_k = p_k - p^* \quad k \in \{1, 2, \ldots, S\}$$

(3)

$$b = \sum_{k=1}^{S} \lambda_k u_k$$

(4)

### 4 EXPERIMENTS

This section will present the results and experiments to evaluate our method for attribute editing and attribute style manipulation. We use CelebAMask-HQ [20] dataset and test set of StyleFlow [3] for all of our experiments on face images. For art images we used Metfaces dataset [15], LSUN cars [38] for cars and LSUN church [38] for churches. For creating the synthetic dataset as explained in Sec. 3.1, we used segmentation mask from CelebAMask-HQ [20] along with the attribute labels from [23]. We use StyleGAN2 [17] model, trained on facial images to generate images and a pre-trained encoder from [8] for mapping real images to latent codes.

Building from the intuition that each layer of StyleGAN Generator controls different hierarchical properties [37], we define a set of layers for each attribute editing as follows: for hair and hat $0-6$, eyeglasses $0-9$, smile $5-6$, pose $0-4$, facial hair $6,7$ and $10$, lighting $7-18$ and eye-close $5-7$. We have empirically found that modifying only the above-selected layer for editing any attribute performs the best. This is not uncommon practice to alter only few layers for editing of any given attribute and all the state-of-the-art methods follow this approach [3, 13, 39].

#### 4.1 Attribute Editing

We show results for a diverse set of face images and Out-of-Domain (OOD) art images from Metfaces edited using random sequential attribute editing in Fig. 4 and Fig. 5 respectively. The edited images from our approach look realistic and coherent even though we use only ten synthetic image pairs. We observe that the edited images closely resemble the original image, maintaining a person’s identity. Also, note that while editing any attribute, all the other attributes are unchanged, proving that our edit directions are largely disentangled. Additionally, FLAME does not modify the background and the skin tone during the edits. Interestingly for art images, FLAME is able to preserve the identity and the painting style during the editing. Note that, the edit directions are obtained from the real image pairs and they do generalize really well on art images. Additionally, we have also performed editing on real face images by first encoding the input image into the $W^+$ latent space using encoder [8] and using the obtained latent code for editing. Results for real image editing is shown in Fig. 4 (Bottom). One can observe that FLAME results in realistic attribute editing on real images and the edits are disentangled.

**Ablation on image pair selection:** To evaluate the robustness of our method against the pairs selected for direction estimation, we perform an experiment with 5 novel expert volunteers. The volunteers were asked to select the most natural looking 10 image pairs $(D^P)$ from generated 100 image pairs. We then estimate the dominant direction $(\hat{d}_j)$ for each of these set of image pairs as explained in Sec. 3.2 and computed the pair-wise cosine similarity between them. Tab. 1 shows the histogram of similarity scores. We can observe that most of the directions are highly correlated as evident in the distribution which is skewed towards large values. This suggests that a new user can easily create the required image pairs with minimal efforts to find edit directions and our method is robust to the choice of specific image pairs used.

#### 4.2 Comparison with state-of-the-art methods

We compare FLAME quantitatively and quantitatively with three recent face editing methods - InterFaceGAN [27], GANSpace [13] and StyleFlow [3]. InterFaceGAN and StyleFlow are supervised methods, whereas GANSpace is an unsupervised method. For InterFaceGAN, we use latent directions for expression, pose, age and eyeglass attributes from the provided implementation on StyleGAN2 [17]. We use the implementation provided by the authors for GANSpace to estimate the PCA and manually select those principal components which correlate with expression, pose, age and eyeglass attributes. For StyleFlow, we use the original codebase for editing images for the above set of attributes. In StyleFlow and GANSpace original implementation only a subset of layers is modified for editing and InterFaceGAN modify all the layers as they train a SVM. We have kept the same configuration during this experiment for a fair comparison. In this experiment, we estimate the attribute edit directions with 10 synthetic image pairs. We use the test set of StyleFlow for evaluation purposes as any of the methods did not use it during

---

**Table 1: Distribution of pair-wise cosine similarity between directions obtained by multiple image-pair sets selected by novel volunteers.** The statistics is aggregated over three attributes: pose, age and eyeglass.

| Cosine Similarity $\uparrow$ | 0.0 − 0.7 | 0.7 − 0.8 | 0.8 − 0.9 | 0.9 − 1.0 | Mean $\uparrow$ |
|-----------------------------|-----------|-----------|-----------|-----------|----------------|
| Normalized Frequency        | 0         | 0.133     | 0.300     | 0.567     | 0.893          |

---
training. For comparison, we perform individual and sequential edits for expression, pose and age attribute editing as these are the common attributes in all four methods.

**Qualitative Comparison:** We compare FLAME against other methods by performing sequential image edits by iteratively changing attributes using the sequence: expression, pose, age, and eyeglasses. Fig. 7 shows the visual results for sequential edits. We observe that our method retains the identity and face structure well even after multiple edit operations. InterFaceGAN and StyleFlow erroneously change the gender while editing the age attribute in both the examples and GANSpace alters the gender while adding glasses. Note that GANSpace entangles multiple attributes, inducing a change in lighting and skin tone. StyleFlow generates realistic edits and doesn’t change most attributes but alters identity after a few sequential edits.

**Quantitative Comparison:** We compare the FID (Fréchet Inception Distance) scores of the sequentially edited images and individual edited images from all four methods to quantify the quality of edits. For sequential editing, we applied the following edit sequence: expression, pose, age, and eyeglasses.
Table 3: Comparison for individual attribute editing

| Attribute | Metric | InterfaceGAN | GANSpace | StyleFlow | FLAME |
|-----------|--------|--------------|----------|-----------|-------|
| Expression | FID ↓ | 36.45 | 36.32 | 34.01 | 33.98 |
|           | CS ↑  | 0.98 | 0.98 | 0.99 | 1.00 |
|           | ED ↓  | 0.32 | 0.29 | 0.23 | 0.15 |
| Pose      | FID ↓ | 34.53 | 34.51 | 34.34 | 30.81 |
|           | CS ↑  | 0.97 | 0.97 | 0.97 | 0.98 |
|           | ED ↓  | 0.38 | 0.36 | 0.36 | 0.28 |
| Age       | FID ↓ | 36.69 | 36.24 | 47.82 | 34.11 |
|           | CS ↑  | 0.93 | 0.95 | 0.89 | 0.95 |
|           | ED ↓  | 0.55 | 0.47 | 0.70 | 0.48 |

Results based on identity preservation and overall visual quality. We use the following sequence of operations expression, pose, age and eyeglasses to generate editing results. Tab. 2 compiles the results from this user study and shows that FLAME was selected most of the time (54.59%) followed by InterFaceGAN (20.40%), StyleFlow (16.31%) and GANSpace (8.70%).

**Qualitative Results on Car and Church categories.** To show the generalization ability of our proposed method, we performed image editing on two additional datasets of cars and churches. For cars, we performed three edits: pose-change, background-change and background removal as shown in top three rows in Fig. 6. For churches, we performed day-to-night editing which is shown in the bottom row in Fig. 6. We use ten curated image pairs (See pairs in supplementary material) and pre-trained StyleGAN encoder models for cars and churches provided by [8]. For cars, our method preserves all the fine details such as the orientation of wheel-rim, color, head and tail lights in the pose and background change tasks for cars. Similarly, it preserves the structure for day-night editing for churches. The wheel rim has changed for the background removal edit in cars, but all other fine details are unchanged. These results substantiate that our approach works effectively for other classes.

**4.3 Attribute Style Manipulation**

Fig. 8 presents the generated diverse style variations for hair and eyeglass attributes. We empirically found the following values for hyper-parameters works best: $\lambda_1 \in (-0.35, 0.35)$, and edit strength $\alpha \in (0.36, 0.46)$ and $\beta \in (0.48, 0.58)$ for eyeglass and hair, respectively. As shown in Fig. 8, our methods can generate diverse frame shapes ranging from frameless to big frames for eyeglasses. The generated results also include sunglasses with varying transparency in the lens. Similarly, our method generates diverse structures and appearances for hairstyles, as shown in Fig. 8. Observe that, in the third original image, the forehead was partially hidden by the hair. Still, new hairstyles are generated in some of the generated images where the forehead is completely visible.

All the generated attributes styles look realistic and match the face and the image’s background well. Note that most of the other image properties like identity are unchanged during style manipulation, while lighting and background do not change significantly. However, there are very subtle changes in expressions but are majorly unnoticeable.

We compare the embeddings from the face-recognition network [10] to quantitatively evaluate the identity preservation in the generated samples. We generated 100 attribute style variations for six sets of images for both eyeglass and hair. Then, we computed the CS and ED between the original and style-edited image. We conducted a user study to compare FLAME with InterFaceGAN, GANSpace, StyleFlow, in which 24 images were presented to 25 participants. The participants were shown the original image and the final sequentially edited image along with intermediates edited images in the sequence for the four methods in random order. The volunteers were asked to select the best editing
obtained a CS score of 0.976 and ED score of 0.34, and for eyeglass, we obtained a CS score of 0.956 and ED score of 0.457. These results imply that our method well preserves identity in generated images.

5 DISCUSSION AND CONCLUSION

In this work, we propose a simple yet effective approach FLAME for face attribute editing by discovering disentangled linear directions in the latent space of the pre-trained StyleGAN model. Our method requires only a few synthesized image pairs to obtain attribute edit directions. We show extensive results for both qualitatively and quantitatively for our method. One limitation of our work is curating synthetic image pairs which can be difficult in some cases, such as gender editing. Similar to existing editing works, our method can also be potentially misused for malicious purposes. Additionally, we propose a novel method to generate attribute style variations for glasses and hairstyles, keeping other attributes unchanged. The proposed framework of attribute style manipulation can be used to generate synthetic image datasets for multiple downstream tasks.

Acknowledgements. Rishubh Parihar acknowledges the support from Prime Minister’s Research Fellowship (PMRF).
