Improved Attribute Manipulation in the Latent Space of StyleGAN for Semantic Face Editing

Aashish Rai
Electronics and Communication Engineering
National Institute of Technology
Surat, India
aashishrai3799@gmail.com

Clara Ducher
Centre for Intelligent Machines
McGill University
Montreal, Canada
clara.ducher@mail.mcgill.ca

Jeremy R. Cooperstock
Centre for Intelligent Machines
McGill University
Montreal, Canada
jer@cim.mcgill.ca

***Abstract*—With the recent popularization of generative frameworks for producing photorealistic face images, we now have the ability to create a convincing graphical match for any particular individual. It is unrealistic, however, to rely solely on such generative methods to randomly produce the facial characteristics we are seeking. Instead, manipulation of facial attributes in the latent space, enabled by the InterFaceGAN framework, allows us to "tweak" these characteristics in the desired direction to improve the quality of the match. The challenge in this process is that attribute entanglement leads to a change of one feature having an undesirable impact on others. We explore several strategies to improve the results of these manipulations, and demonstrate how the automatic conditioning of attributes can be used to minimize the impact of such entanglement, and further, allow for improved control over complex (non-binary) attributes such as race or face shape.**

***Index Terms*—Generative Adversarial Networks, Latent space manipulation, Face attribute editing**

I. Introduction

Generative Adversarial Networks (GANs) [9] have found application to a wide variety of problems, in particular for the generation, completion, and editing of images [8], [3], [12], [7]. The StyleGAN extension [16], [17] has proven itself to be a powerful tool for efficient synthesis of photorealistic images [5], [15], [16], [17]. Although its results are almost indistinguishable from real images, there is limited understanding of how GANs can be used to map a random latent code to specific facial features.

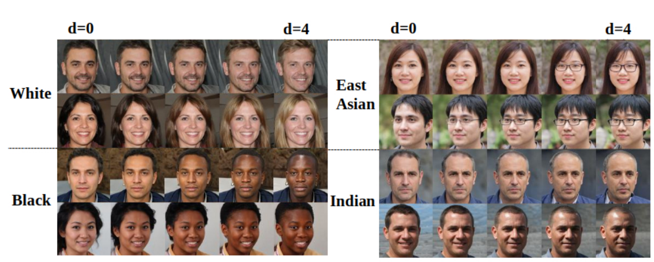

Fig. 1: Results of our framework on the race attribute. Images are synthesized by editing the original latent codes ($d = 0$, shown in the leftmost column) at equal intervals up to a distance of 4 units from the separating hyperplanes.

A. Literature Review

GANs have demonstrated a significant ability to enhance the quality of synthesized images [22], [10], [4], [2], [15], [16], [17]. Diverse frameworks were introduced to improve realism and resolution of face images. The progressive GAN (ProGAN) [15] can generate high-resolution (1024×1024) pixels images using progressive training, but provides only minimal control over specific facial features in the output image. Moreover, tweaking the input latent code can simultaneously affect multiple output facial features, a problem arising from feature entanglement, i.e., one direction in the GAN latent space is unlikely to represent only a single attribute of the output image.

To tackle the entanglement problem, Karras et al. proposed StyleGAN [16], which introduced a disentangled latent space W, and two new sources of randomness: a standalone mapping network and noise layers. However, the results of StyleGAN exhibit several blob-like artifacts, beginning from a resolution as low as 64×64 pixels.

Shen et al. proposed InterFaceGAN [20], a novel framework that controls features in the output images by manipulating the random input vector. They applied their method to the pre-trained PGGAN [15] and StyleGAN [16], obtaining promising results on five attributes in the output image represented by binary semantics: smile, age, gender, pose and eyeglasses. However, due to the entanglement problem, the synthesized features in InterFaceGAN tend to be unpredictable. Their solution, conditional editing, consists of manually performing orthogonal projection of semantic boundaries to decorrelate pairs of attributes, but requires attention to unintended changes in secondary attributes resulting from manipulation of a primary attribute. We posit that this could be improved either by re-implementing the framework on a better image synthesis model or by improving the semantic boundaries. A second problem of InterFaceGAN is that the progressive growing training method of StyleGAN1 tends to produce many blob-shaped artifacts, as noted above. This problem was significantly reduced by the introduction of StyleGAN2 [17], which further improved image quality by normalizing the CNN weights using estimated statistics instead of adaptive instance normalization. Moreover, and importantly for our purposes, StyleGAN2 provides continuity in the latent space.

More recently, Hou et al. introduced GuidedStyle [11], which emphasizes a novel learning framework that leverages the knowledge from face attribute classifiers to guide the image synthesis process of StyleGAN. Their approach allows a sparse attention mechanism to select a single layer of the generator for semantic face editing. As a result, it can disentangle various semantics and achieve good editing results on six face attributes. Abdal et al. presented StyleFlow [1], a robust solution to the conditional exploration of the latent space of StyleGAN to deal with semantic editing. They reviewed two important sub-problems of attribute-conditioned sampling and attribute-controlled editing on StyleGAN using conditional continuous normalizing flows, and found that this method can sample high-quality images from the latent space given a set of attributes. Despite these advantages, the approach does not offer a solution to deal with complex attributes, such as face shape or race, that cannot be efficiently represented in the latent space by a single hyperplane [20], [11], [1].

B. Contribution

This paper contributes a solution to address the shortcomings described above, which we achieve by applying the InterFaceGAN framework to the improved StyleGAN2 [17]. We extend this framework to cover the additional facial attributes of beard, chubby face, pointy nose, big nose, race, hair style and face shape. We further investigate the correlation between these attributes in the latent space, demonstrating how manipulation of one such parameter may affect others. Building on the concept of orthogonal projection of semantic boundaries [20], we propose an automated means of obtaining the conditional boundary that minimizes the impact of each attribute on the others. Our exploration reveals that the quality of the semantics relies heavily on the accuracy of the classification algorithms used to generate the labels of the GAN outputs.

II. Architecture

We extended the InterFaceGAN framework on a pre-trained StyleGAN2 model [17] to synthesize state-of-the-art facial images, while avoiding the artifacts resulting from progressive growing in earlier models [15], [16]. The framework is based on the hypothesis that there exists a hyperplane in the latent space separating a binary attribute, such as age, into two classes, e.g., young and old faces. Ideally, if all attributes were independent, this hyperplane would remain unchanged as other attributes were modified. This hypothesis can be verified by training a linear SVM over the attributes and evaluating it on test data [20]. Greater SVM accuracy results from hyperplanes that better separate faces with a positive response to the considered attribute from those with a negative response.

A. Boundary Synthesis

We now explain how semantics are obtained and how they can be used to edit images. For each selected facial attribute, a binary classifier is trained using a labeled database to determine whether the considered attribute is present in any face picture. The trained classifier is then applied to randomly sampled images from the GAN to generate a score that defines, for each image, its level of "belonging" to the attribute. For instance, for the smile attribute, GAN outputs with scores of ± 1 are associated with happy and sad faces, while scores of 0, lying on the hyperplane boundary, represent neutral faces. Next, the latent codes corresponding to the scored images are used as inputs to train a linear SVM, using as labels the scores obtained from the previous step. The weights of the resulting SVM, which define a hyperplane in the latent space, are used as a semantic boundary.

The InterFaceGAN framework also allows for manual adjustments of the generated boundary to disentangle attributes. For instance, while editing images along the smile semantic, we must be concerned with the possibility of simultaneous changes along the gender boundary, with which it is correlated (see Table II). To prevent this, we can take the orthogonal projection of the smile boundary with respect to the gender boundary, for which our pipeline proposes an automation of conditioning based on the highest correlated attribute with the desired one. We reused the pipeline with multi-class classifiers and linear SVM to compute categorical semantics for complex attributes such as race, hair style and face shape.

B. Latent Code Manipulation

We tested our implementation by generating boundaries for the five attributes of the original framework, four additional simple attributes, and three complex ones.

The original latent codes can be manipulated by linear interpolation on either side of the semantic boundary. However, latent code samples exceeding some "critical distance" will generate faces that lack the characteristics of the original. Consider an original latent code z, an edited latent code, z', and a vector β orthogonal to a semantic boundary, each of 512 dimensions. To interpolate between z and z', we apply an orthogonal displacement of factor α, $\alpha \in R$: $z' = z + \alpha\beta$.

Once we have the edited $512D$ latent codes, we can map them to the photo-realistic images using a generator function G. We used the original StyleGAN2 [17] model, trained on the FFHQ database, which can be employed directly to synthesize high resolution (1024×1024) images X from the corresponding edited latent codes z' as $X_i = G(z'_i)$.

III. Implementation and Analysis

Through several experiments, we applied the Tensorflow implementation of the framework described in Section II to the StyleGAN2 model, to

1) illustrate the results obtained on the original InterFaceGAN framework, applied to StyleGAN2,
2) demonstrate that the framework can be used to train new semantics for a wider range of facial attributes than the five original ones used by InterFaceGAN [20],
3) highlight the need for conditioning boundaries on one or more attributes, and demonstrate the results of automating this process,

4) examine how the accuracy of the classifier and the SVM impact the edited images, and
5) compute recursive binary semantics and categorical semantics for complex attributes.

TABLE I: Ethnicity distribution of synthesized database and SVM training data. The database synthesized using StyleGAN2 generator is highly unbalanced and may result in biased boundaries. To avoid such bias, we picked the samples with highest scores from each group and gathered a balanced database of 2000 images for SVM training.

Database	**Black**	**East Asian**	**Indian**	**Middle East**	**White**
Synthesized	1.28%	8.6%	1.05%	1.28%	86.69%
SVM Training	18.3%	20.4%	18.7%	19.2%	23.4%

Each semantic boundary was trained using a balanced SVM database of 2000 out of a total of 100 000 samples from the StyleGAN2 synthesized database. These samples are chosen as those with lowest and highest scores from a trained CNN classifier, and therefore represent the extrema for each semantic label, such as, for the "smile" attribute, very sad or happy faces, thus avoiding ambiguity in the results. Moreover, this helps avoid biased boundaries resulting from unbalanced datasets, as can be seen in the comparison of ethnicity distributions of the balanced SVM training database and the StyleGAN2 synthesized database, as shown in Table I. If the synthesized database were used directly for boundary synthesis, this may similarly result in a highly biased boundary [13], a problem avoided by choosing the best 2000 samples for SVM training, resulting in proper boundaries.

A. *InterFaceGAN from StyleGAN1 to StyleGAN2*

Fig. 2 shows edited images of both versions of StyleGAN along the five attributes proposed in the original InterFaceGAN paper [20]. As can be seen in the results, the manipulation of one attribute has a stronger impact on other attributes in StyleGAN1 than in StyleGAN2, confirming that the latter suffers less from entanglement. The quality of the edited StyleGAN2 images is superior, exhibiting fewer flaws, due to correction of artifacts that were present in StyleGAN1. This is particularly evident in the images at the extrema (leftmost and rightmost columns). Despite these improvements, the latent space in StyleGAN2 nevertheless remains entangled, such that manipulation of a single attribute often has an undesirable impact on secondary attributes as well. This is most obvious by noting that many of the attributes demonstrate a shift of gender across the hyperplane boundary.

B. *InterFaceGAN on more attributes*

We were also interested in testing the framework on more fine-grained attributes (e.g., beard, pointy nose, and chubby face), to cover more subtle facial features, so as to allow for editing of the images at varying levels of refinement. Fig. 3 illustrates the results for four attributes on which we trained binary semantics. Despite the model correctly learning the semantic boundaries of these new attributes, we can still

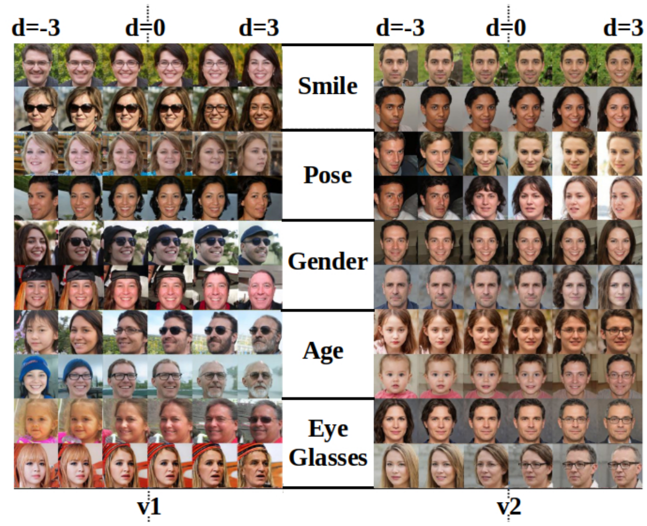

Fig. 2: Comparison of manipulated results of the original InterFaceGAN framework on StyleGAN v1 (left) and v2 (right). The dotted vertical line splitting each set of results at the midpoint represents the hyperplane separating positive and negative samples. Each row represents the results of a different latent code, with the images corresponding to $d = \pm 3$ shown on the extreme left and right. Intermediate thumbnail images are transitional results. Note that unlike other figures in this paper, here, we illustrate a progression of distances both on the *left* and *right* of the separating hyperplane.

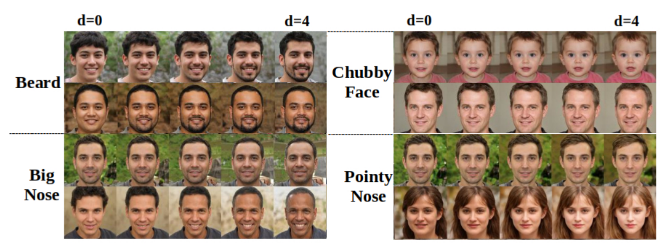

Fig. 3: Results of additional simple attributes. Each row illustrates edited versions (d=1,2,3,4) of the original face (d=0), which appears in the leftmost column.

observe the effects of entanglement. For instance, changes in the pointy nose attribute are correlated with unintended changes in hair color after a few interpolations. However, our ability to decorrelate these features is limited to some extent by the fact that some correlations exist between these attributes in the real images that are used to train StyleGAN models. For example, older people tend to wear eyeglasses, and children more often than adults, have blonde hair.

C. *Automatic boundary conditioning and multi-conditioning*

1) Single attribute conditioning: The correlation between different semantic boundaries in the latent space is described in Table II. The length of the vector projection of one attribute onto other attributes is an estimate of the relationship between their semantic boundaries. For instance, all the race boundaries

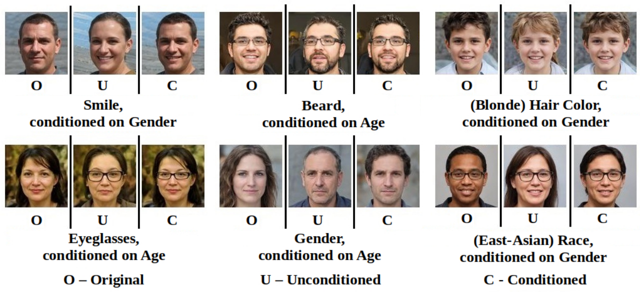

Fig. 4: Examples of (O)riginal, (U)nconditioned, and (C)onditioned on the most entangled attribute, the corresponding edited sample.

were conditioned on age, the attribute with which they are most correlated, while eyeglasses was conditioned on age, a sensible correlation given that older individuals are more likely to wear glasses.

For each primary attribute p, and maximally correlated secondary attribute, s, we find the conditional boundary b by subtracting from the primary boundary its projection onto the secondary boundary. [20].

$$b = n_p - (n_p^T \cdot n_s)n_s \tag{1}$$

Figure 4 illustrates the results obtained after automated conditioning of the attribute with maximum correlation. As can be seen, the results of this conditioning are new images in which the primary attribute has been modified as desired, with minimal impact on the secondary attribute with which it would otherwise be entangled.

2) Conditioning on Multiple Attributes: Despite the improvements seen above, Table II demonstrates that multiple attributes may be correlated with the primary attribute in the latent space. Unfortunately, we are limited in applying the above technique repeatedly: if we keep taking orthogonal projections onto further hyperplanes corresponding to other semantic boundaries, the influence of the primary attribute is continually reduced in the output images, resulting in edited images that are minimally changed relative to the original.

D. Complex attributes

Here, we test the framework on three complex attributes of race, hairstyle , and face shape.

1) Split semantics for complex attributes: We first consider the possibility that multiple hyperplanes can be found in the latent space to partition the values of each of the complex attributes. Accordingly, to manipulate semantics of complex attributes, we consider all possible $n(n-1)$ pairs, where n represents the number of values of each attribute. For example, in the case of race attributes, we train a separate SVM classifier for each pair of White-Black, White-MiddleEastern, White-EastAsian, White-Indian, Black-MiddleEastern, etc., that can be used to manipulate the features. However, this approach suffers from the fact that the pairwise comparisons do not fully characterize an attribute with more than two possible values. For example, an Asian person is neither White nor Black, but the labeling process assigns either of these labels to every one of the 100k samples in the model.

2) Categorical semantics for complex attributes: To overcome this limitation, we propose to use a multi-class CNN for predicting scores and a multi-class linear SVM for the semantic boundaries of the complex attributes. Results shown in Figures 1 and 5 were synthesized applying this approach. In this manner, partitioning of the entire space allows for detection of boundary crossing between classes of these complex attributes. We also compared the two methods in manipulating non-binary attributes and found that categorical semantics

TABLE II: Correlation of attributes in the latent space against the basic (first five) attributes. Values are shown in bold when the maximum correlation is with either the *age* or *gender*.

	Smile	**Age**	**Gender**	**Pose**	**Eyeglasses**
Smile	—	0.06	**0.26**	0.08	0.00
Age	0.06	—	**0.38**	0.14	0.35
Gender	0.26	**0.38**	—	0.06	0.12
Pose	0.08	**0.14**	0.06	—	0.00
Eyeglasses	0.00	**0.35**	0.12	0.00	—
Beard	0.20	**0.40**	0.21	0.05	0.16
Pointy Nose	0.03	**0.42**	0.36	0.00	0.20
Big Nose	0.11	**0.43**	0.36	0.01	0.28
Chubby	0.02	**0.56**	0.40	0.03	0.32
Race					
Black	0.09	**0.23**	0.11	0.17	0.08
White	0.08	0.07	**0.29**	0.12	0.11
East- Asian	0.01	0.06	**0.29**	0.03	0.24
Indian	0.00	**0.24**	0.21	0.07	0.06
Hair Style					
Blonde	0.11	0.36	**0.45**	0.13	0.19
Brown	0.14	0.37	**0.44**	0.12	0.20
Black	0.09	0.34	**0.36**	0.04	0.17
Wavy	0.07	0.07	0.11	*0.54*	0.03
Short	0.07	**0.36**	0.25	0.00	0.28
Face Shape					
Square	0.09	**0.31**	0.18	0.00	0.18
Heart	0.03	0.18	**0.21**	0.00	0.05
Round	0.00	**0.36**	0.18	0.01	0.12
Long	0.18	**0.24**	0.21	0.03	0.00
Oval	0.18	0.20	**0.28**	0.15	0.11

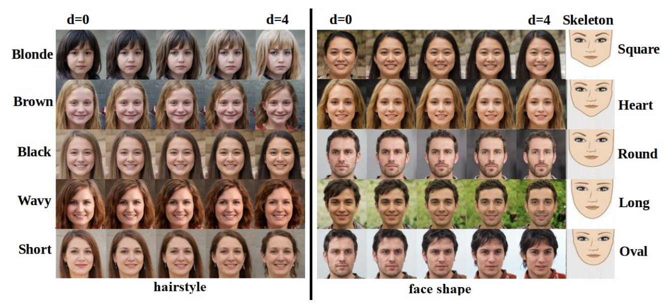

Fig. 5: Results of editing of hairstyle attribute and face shape attribute after automated conditioning. Moving from left to right in each row starts with the original latent code, followed by progressively edited transitional images. Face shapes are illustrated by the sketches to the right of face shape manipulations.

significantly outperforms split semantics. The former achieved an average accuracy of 95.56% for race, 92.52% for face shape, and 94.92% for hairstyle, while split semantics achieved an average accuracy of 74.44%, 66.66% and 81.26% for the same attributes, respectively.

E. Impact of accurate boundaries and non linear kernels

To overcome the challenges of entanglement and obtain a good semantic boundary for both simple and complex attributes, training of the classifier to obtain high prediction accuracy is paramount. For this purpose, we modified the VGG-16 (D configuration) neural network model [21], by removing the last block of three convolution layers, followed by a max-pooling layer. We started with 32 filters and doubled them in each successive block until reaching 512, the size of the last three dense layers. This reduced the trainable parameters by 78.8%, thereby significantly improving training time for each model. To improve generalization, we also applied regularization in each layer. Performance of the CNN classifiers is provided in Table III for the binary and non-binary semantics we use.

TABLE III: Details of CNN and SVM classifiers (linear kernel).

Attribute	Dataset	CNN Train Accuracy (%)	CNN Val. Accuracy (%)	SVM Val. Accuracy (%)
Chubby Face	CelebA	98.98	98.20	95.5
Gender	CelebA	99.46	99.27	94.0
Pose	Custom	97.24	96.00	93.1
Beard	CelebA	99.46	98.95	93.0
Smile	CelebA	98.77	98.68	92.2
Eyeglasses	CelebA	99.17	98.94	89.3
Pointy Nose	CelebA	98.19	97.98	86.5
Big Nose	CelebA	98.62	96.26	84.0
Age	Custom	98.00	96.40	82.0
Hair-Style	CelebA	98.90	98.70	88.5
Race	FairFace	98.40	97.60	87.8
Face Shape	Custom	97.70	97.40	82.3

To prepare the training database for nine simple attributes (smile, pose, age, gender, and eyeglasses, beard, chubby face, pointy nose, and big nose), we used images from the CelebA dataset [19] and a custom dataset synthesized using InterFaceGAN [20], whereas for the complex attributes, we used the CelebA [19], FairFace [14] and a custom dataset. In all cases, images were resized to 256×256 pixels. After applying data augmentation to these datasets, we trained the classifiers and used them to determine the prediction scores for the 100k images. To train the binary and multi-class CNN classifiers, we applied the cross-entropy loss with adaptive moment estimation (ADAM) [18], a popular gradient descent algorithm for non-convex optimization.

The classification accuracy of each SVM determines the quality of the semantic boundary for each associated attribute. We observe that the gender semantic could synthesize faces with a clear progression in gender even without conditioning. Thus, SVM classification accuracy of 94% (Table III) appears

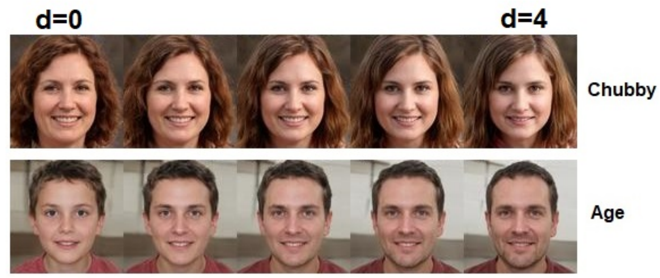

Fig. 6: Images are synthesized by editing the original latent codes (leftmost) at equal intervals up to a distance of 4 units (rightmost) from the separating hyperplanes.

to be a reasonable target to obtain convincing semantics. However, this level was reached for only a few simple attributes. In some other cases, visual inspection of the manipulated results reveals lower perceptual quality than expected. For instance, the chubby face semantic is 13.5% more accurate than the age semantic. Nevertheless, age manipulations are quite convincing while moving towards a chubby face does not necessarily produce obvious variations in this attribute. (Figure 6)

F. Results on fine attributes and sequential editing

We further tried implementing some fine attributes on the framework and obtained reasonable results, as shown in Figure 3. Note that implementing fine attributes may demand more well-trained decision boundaries; otherwise, one may obtain a semantic boundary that separates coarse attributes rather than the desired one. As an example, Figure 1 shows the results of the race attribute. Interestingly, the framework could synthesize Indian faces even though there are only a few such samples in the entire database of 100k images. It may help us understand that the framework attempts to learn from the specific iterating features in the latent space rather than merely reflecting the entire image. Figures 1,3, and 5 show the results after automated conditioning.

Most other image-to-image translation frameworks, such as InterFaceGAN, generate poor results, or worse, may fail to produce photo-realistic images, when the generated images are re-introduced to the network with a new set of target attributes. In some cases, the generated outputs simply cannot be reused. For instance, if a first output of the StarGAN network [6] is reused to edit a second attribute, the input image is simply reproduced, independent of the second targeted attribute. A similar trend is observed when using InterFaceGAN with StyleGAN1. Upon editing images more than twice, the model generates artifacts, and the identity of the original image is progressively lost due to entanglement. However, as seen in Figure 7, when InterFaceGAN is implemented on StyleGAN2 with improved boundary quality, we could perform several editing rounds while preserving the identity. Similar results are observed under different evaluation metrics (Table IV), including Fréchet Inception Distance (FID), Sliced Wasserstein Distance (SWD), Cosine Similarity (CS), and Euclidean

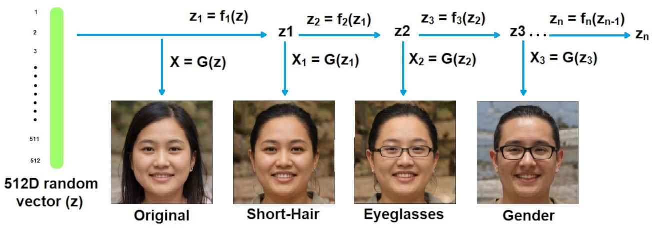

Fig. 7: Demonstration of sequential editing, obtained by manipulating the original latent code with different boundaries in a serial fashion. Functions f1(.), f2(.), and f3(.) correspond to short-hair, eyeglasses, and gender boundaries, respectively. G(.) is the StyleGAN2 generator.

Distance (ED). We compared a randomly edited set of 500 samples of InterFaceGAN on StyleGAN1 against our approach along five semantics (eyeglasses, smile, pose, age, and gender). We did not use new semantics since these were not available in InterFaceGAN.

TABLE IV: Qualitative comparison of InterFaceGAN on StyleGAN1 and our approach. (↓ or ↑ indicates a lower or higher score, respectively, is considered superior.)

Model	FID ↓	SWD ↓	CS ↑	ED ↓
InterFaceGAN	51.48	270.16	0.58	0.88
Our	44.59	254.07	0.63	0.81

IV. Conclusions and Future Work

We revisited the InterFaceGAN [20] framework, and proposed modifications to overcome the issues caused by the entanglement problem. By implementing the framework on StyleGAN2 and enhancing the quality of semantic boundaries, we improved the quality of synthesized images. We were further able to implement fine attributes, including face shape, hair color, and nose shape, and in addition, made progress in addressing the challenges of complex attributes, i.e., those with more than two distinct values.

However, the semantic boundaries, both of simple (binary) and complex attributes, remain imperfect, a challenge remaining to be tackled in future work in order to further improve the quality of the output. Similarly, we found that conditioning on multiple attributes may reduce the effectiveness of the primary attribute in the output. It would be highly desirable to overcome this limitation. Finally, we noticed that the framework occasionally generates outputs on the "wrong side" of the hyperplane boundary as specified by the user. This suggests a possible lack of optimality in the semantic boundaries.

This work affords fine-grained control over face generation, but remains heavily dependent on the contents of the datasets on which the frameworks are trained. To address this issue, in part, while generating the semantic boundaries, we redistributed the original dataset to improve diversity and avoid racial and gender bias in the end results. As a final comment, we note that tools such as ours could be used for malign purposes, i.e., unethical applications of "deep fakes". Given the proliferation and rapid growth in power of these technologies, it is imperative that the research community develop guidelines to address the ethical implications of these technologies.

References

[1] Abdal, R., Zhu, P., Mitra, N., Wonka, P.: Styleflow: Attribute-conditioned exploration of StyleGAN-generated images using conditional continuous normalizing flows. arXiv e-prints (2020)

[2] Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: Precup, D., Teh, Y.W. (eds.) Proceedings of the 34th International Conference on Machine Learning. Proc. Machine Learning Research, vol. 70, pp. 214–223. Sydney, Australia (Aug 2017)

[3] Bao, J., Chen, D., Wen, F., Li, H., Hua, G.: CVAE-GAN: fine-grained image generation through asymmetric training. In: Proc. Conf. on Computer Vision. pp. 2745–2754. IEEE (2017)

[4] Berthelot, D., Schumm, T., Metz, L.: Began: Boundary equilibrium generative adversarial networks. arXiv preprint arXiv:1703.10717 (2017)

[5] Brock, A., Donahue, J., Simonyan, K.: Large scale GAN training for high fidelity natural image synthesis. arXiv preprint arXiv:1809.11096 (2018)

[6] Choi, Y., Choi, M., Kim, M., Ha, J.W., Kim, S., Choo, J.: Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In: Proc. Conf. on Computer Vision and Pattern Recognition. pp. 8789–8797. IEEE (2018)

[7] Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., Greenspan, H.: GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing **321**, 321–331 (2018)

[8] Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., Greenspan, H.: Synthetic data augmentation using GAN for improved liver lesion classification. In: Int. Symp. on Biomedical Imaging. pp. 289–293. IEEE (2018)

[9] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Proc. Neural Information Processing Systems, Vol. 2. p. 2672–2680. MIT Press, Cambridge, MA, USA (2014)

[10] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of Wasserstein GANs. In: Advances in neural information processing systems. pp. 5767–5777 (2017)

[11] Hou, X., Zhang, X., Shen, L., Lai, Z., Wan, J.: Guidedstyle: Attribute knowledge guided style manipulation for semantic face editing. arXiv preprint arXiv:2012.11856 (2020)

[12] Iizuka, S., Simo-Serra, E., Ishikawa, H.: Globally and locally consistent image completion. ACM Transactions on Graphics **36**(4), 1–14 (2017)

[13] Johnson, J.M., Khoshgoftaar, T.M.: Survey on deep learning with class imbalance. Journal of Big Data **6**(1), 1–54 (2019)

[14] Kärkkäinen, K., Joo, J.: Fairface: Face attribute dataset for balanced race, gender, and age. arXiv preprint arXiv:1908.04913 (2019)

[15] Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of GANs for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196 (2017)

[16] Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: Proc. Conf. on Computer Vision and Pattern Recognition. pp. 4401–4410. IEEE (2019)

[17] Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.: Analyzing and improving the image quality of stylegan. In: Proc. Conf. on Computer Vision and Pattern Recognition. pp. 8110–8119. IEEE/CVF (2020)

[18] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

[19] Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: Proceedings of the IEEE international conference on computer vision. pp. 3730–3738 (2015)

[20] Shen, Y., Gu, J., Tang, X., Zhou, B.: Interpreting the latent space of gans for semantic face editing. In: Proc. Conf. on Computer Vision and Pattern Recognition. pp. 9243–9252. IEEE/CVF (2020)

[21] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: Bengio, Y., LeCun, Y. (eds.) Int. Conf. on Learning Representations. San Diego, CA, USA (May 2015)

[22] Zhang, H., Goodfellow, I., Metaxas, D., Odena, A.: Self-attention generative adversarial networks. In: Int. Conf. on Machine Learning. pp. 7354–7363 (2019)