EXPLORING SEMANTIC VARIATIONS IN GAN LATENT SPACES VIA MATRIX FACTORIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Controlled data generation with GANs is desirable but challenging due to the non-linearity and high dimensionality of their latent spaces. In this work, we explore image manipulations learned by GANSpace, a state-of-the-art method based on PCA. Through quantitative and qualitative assessments we show: (a) GANSpace produces a wide range of high-quality image manipulations, but they can be highly entangled, limiting potential use cases; (b) Replacing PCA with ICA improves the quality and disentanglement of manipulations; (c) The quality of the generated images can be sensitive to the size of GANs, but regardless of their complexity, fundamental controlling directions can be observed in their latent spaces.

1 INTRODUCTION

Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) have enabled significant progress in data synthesis, including the generation of images, video, and text (Karras et al., 2018). GANs consist of two networks: a discriminator that differentiates between real and generated data, and a generator that produces realistic data from random latent vectors. However, the lack of control over the data generated by GANs limits their usage not only for images but also for other types of data. To overcome this limitation, several methods of image manipulation in the latent space of GANs have been developed such as InterFaceGAN (Shen et al., 2020) and GANanalyze (Goetschalckx et al., 2019). Our work is aimed to explore the latent space manipulations in GANs to gain insight into the underlying processes that enable the generation of new and semantically meaningful images. Specifically, we focused on a state-of-the-art (SOTA) unsupervised method - GANSpace (Härkönen et al., 2020), which utilizes Principal Component Analysis (PCA) to find orthogonal directions along which new semantic axes can be located. We performed both a visual and quantitative evaluation of GANSpace on StyleGAN2 (Karras et al., 2020), SOTA large-scale GAN producing realistic images, and GAN from Liu et al. (2021), light-weight GAN that requires significantly fewer resources and less training time, but still produces high-quality images. Finally, we replaced PCA with Independent Component Analysis (ICA) because ICA was found to be a preferred option as it searches for statistically independent components and allows for a more disentangled representation of the data. This work also represents our first attempt towards the goal of exploring larger questions of the semantic organization of latent spaces and their interpretability.

2 METHODS

In this work, we evaluated the transformations obtained by GANSpace (Härkönen et al., 2020) for both StyleGAN2 (Karras et al., 2020) and GAN from Liu et al. (2021) both visually and quantitatively. For StyleGAN2, as was done by Härkönen et al. (2020) we applied GANSpace in $\mathcal{W}$ latent space. For the GAN from Liu et al. (2021), the details of implementation can be found in Section A. We generated 1000 images for both GANs and applied one random transformation from a set found by GANSpace for each image. Further, we calculated the Fréchet inception distance (FID) (Heusel et al., 2017) between the original generated images and transformed ones. The FID score allows to measure the similarity between two sets of images. Finally, we substituted PCA with ICA in the GANSpace method, conducted a similar evaluation of the resulting manipulations, and subsequently compared them to those obtained using the original GANSpace.

3 RESULTS AND DISCUSSION

The examples of manipulations obtained by GANSpace (Härkönen et al., 2020) are shown in Fig. 1 for StyleGAN2 (Karras et al., 2020) and Fig. 2 for GAN from Liu et al. (2021). GANSpace produced high-quality transformations for both GANs, as demonstrated by low FID scores (Table 1). However, we observed that PCA does not fully differentiate entangled directions in the obtained transformations, as it focuses only on maximizing the variance and orthogonality.

To address the problem of entanglement, we have replaced PCA with ICA in the GANSpace processing pipeline. Similar to PCA, ICA requires a number of components to be specified as a hyperparameter. We investigated the effect of the number of components on the obtained manipulations, testing two options for StyleGAN2: 20 and 100 components (Fig. 3, 4). We found that increasing the number of components leads to a significant increase in the disentanglement of the manipulations. In addition, GANSpace with ICA discovers a more diverse set of manipulations than GANSpace with PCA, e.g. it found "weight" (Fig. 4b), "brightness" (Fig. 6b) and "background" directions which were not discovered by the original GANSpace. The FID score improved as the number of components increased. Specifically, when the manipulation strength α was randomly chosen from the range of $[-3, 3]$, the FID score improved from 25.48 to 20.13. Using the larger range of $[-6, 6]$, the FID score also improved from 35.16 to 25.43. These results can be seen in Table 2. This shows that with the increased number of components, the quality of transformations also increases.

For the GAN from Liu et al. (2021), ICA was applied at the $conv1$ layer as was done for the original GANSpace (Section A provides more details). But as this layer has 16384 components, ICA converged only for 500 and 1000 components. The resulting set of manipulations can be characterised as more diverse than those obtained with PCA and similar to one produced by ICA with StyleGAN2. However, the increased number of ICA components only slightly increases the disentanglement and does not change the manipulation quality (Fig. 5, 6, Table 3).

By substituting PCA with ICA in GANSpace, the found manipulations may become significantly more disentangled and the set of manipulations becomes more diverse (Table 2 and Fig. 4). This observation can be attributed to the fact that PCA focuses only on maximizing the variance, while ICA pays attention to the statistical independence of components. However, ICA does not always achieve disentanglement as its results highly depend on the chosen number of components. Moreover, the disentanglement of the original feature space also affects ICA results. For instance, $\mathcal{W}$ space StyleGAN2 is known as disentangled latent space (Wu et al., 2021) which makes achieving the disentanglement possible with ICA. However, the $conv1$ layer of the GAN from Liu et al. (2021) does not have underlying independent components, making ICA much less efficient than for StyleGAN2. Overall, our results suggest that ICA may help achieve disentangled image manipulations in the latent space of GANs, but its effectiveness depends on the underlying disentanglement of the feature space and the careful choice of the number of components.

We also noticed that, despite the difference in the number of parameters and dimensionality of the feature spaces, both the original GANSpace and ICA discover the same set of manipulation directions for both GANs. For example, for both StyleGAN and the GAN from Liu et al. (2021), GANSpace found "pose", "age", "skin colour" and "smile" transformations (Fig. 3, 4) and ICA discovered additional "age", "smile", "weight", "skin colour" and "brightness" manipulations (Fig. 5, 6). Thus, the latent spaces of both GANs retain similar fundamental moving directions.

4 CONCLUSION

In conclusion, this study examined image manipulation in the latent space of GANs, focusing on GANSpace applied to two GANs: StyleGAN2 and the GAN from Liu et al. (2021). GANSpace demonstrates effective image manipulations on both GAN models, however, the manipulations exhibit a high degree of entanglement. To address this limitation, ICA was tested as a replacement for PCA in GANSpace. The findings reveal that ICA can potentially produce more disentangled manipulations and a wider variety of transformations compared to PCA. However, the number of components needs to be carefully selected to achieve disentanglement, and the disentanglement of manipulations depends on whether the original latent or feature space has the underlying independent components. The quantitative evaluation using FID scores supported the visual evaluation results. Furthermore, our results suggest that the primary controlling directions within the latent space remain consistent regardless of the complexity of the investigated GAN models.

URM STATEMENT

The authors acknowledge that at least one key author of this work meets the URM criteria of ICLR 2023 Tiny Papers Track.

REFERENCES

Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. In *ICLR*, 2019.

Lore Goetschalckx, Alex Andonian, Aude Oliva, and Phillip Isola. Ganalyze: Toward visual definitions of cognitive image properties. In *ICCV*, 2019.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. In *NeurIPS*, 2014.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *NeurIPS*, 2017.

Erik Härkönen, Aaron Hertzmann, Jaakko Lehtinen, and Sylvain Paris. Ganspace: Discovering interpretable gan controls. In *NeurIPS*, 2020.

Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. In *ICLR*, 2018.

Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *CVPR*, 2020.

Bingchen Liu, Yizhe Zhu, Kunpeng Song, and Ahmed Elgammal. Towards faster and stabilized gan training for high-fidelity few-shot image synthesis. In *ICLR*, 2021.

Yujun Shen, Jinjin Gu, Xiaoou Tang, and Zhou Bolei. Interpreting the latent space of gans for semantic face editing. In *CVPR*, 2020.

Zongze Wu, Dani Lischinski, and Eli Shechtman. Stylespace analysis: Disentangled controls for stylegan image generation. In *CVPR*, 2021.

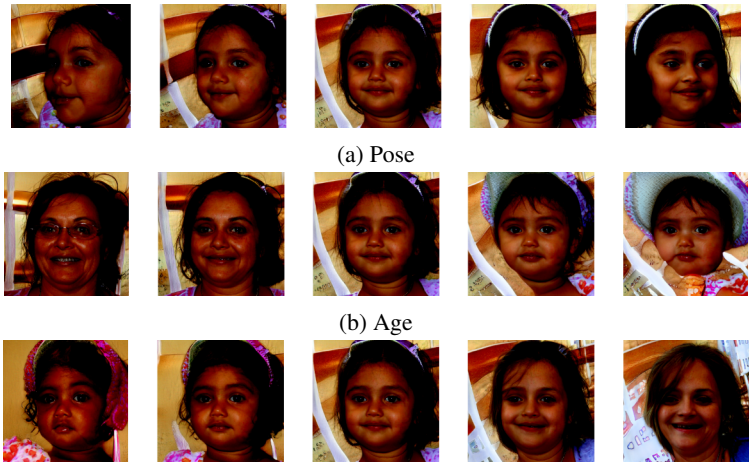
A IMPLEMENTATION FOR THE GAN FROM LIU ET AL.

We faced the problem of applying GANSpace (Härkönen et al., 2020) to the GAN from Liu et al. (2021), as the authors originally applied it to StyleGAN2 (Karras et al., 2020) using $\mathcal{W}$ space and to BigGAN (Brock et al., 2019) using the first linear layer. However, GAN from Liu et al. (2021) has neither of them. To resolve this issue, we performed the approach analogical to what the authors did to BigGAN (Brock et al., 2019):

1. Sampled N random latent vectors $z_{1:N}$. where N=1000;
2. Processed these latent vectors through $conv1$ layer of the GAN to get N feature vectors $y_{1:N}$;
3. Applied PCA to these feature vectors to get the principal component matrix V;
4. Projected feature vector onto the PCA components: $x_j = V^T(y_j - \mu)$, where μ is the mean feature vector;
5. Obtained the matrix of transformation in the original latent space using the least squares method: $U = \arg\min \sum_j \|Ux_j - z_j\|$.

The reason why we chose $conv1$ layer is that, according to Härkönen et al. (2020), for BigGAN the first linear layer allows to get more useful moving directions than other layers.

B EXAMPLES OF MANIPULATIONS AND FID SCORES



(a) Pose

(b) Age

(c) Skin colour+age

Figure 1: Sequences of image manipulations performed using directions from GANSpace applied to StyleGAN2

(a) Gender+smile

(b) Age+pose

(c) Pose+smile

Figure 2: Sequences of image manipulations performed using directions from GANSpace, applied to GAN from Liu et al. (2021)

Table 1: FID scores for original GANSpace

Manipulation strength \ Name of GAN	StyleGAN	GAN from Liu et al. (2021)
$\alpha \sim U[-3, 3]$	25.26	20.41
$\alpha \sim U[-6, 6]$	33.33	24.33

(a) Smile+weight

(b) Skin colour

(c) Age+smile

Figure 3: Sequences of image manipulations performed using directions from GANSpace with ICA applied to StyleGAN2 (number of components is 20)

(a) Age+smile

(b) Weight

(c) Age

Figure 4: Sequences of image manipulations performed using directions from GANSpace with ICA applied to StyleGAN2 (number of components is 100)

Table 2: FID scores for GANSpace for StyleGAN2 with ICA

Manipulation strength \ Number of components	20	100
$\alpha \sim U[-3, 3]$	25.48	20.13
$\alpha \sim U[-6, 6]$	35.16	25.43

(a) Age

(b) Pose+brightness

(c) Age+pose

Figure 5: Sequences of image manipulations performed using directions from GANSpace with ICA applied to GAN from Liu et al. (2021) (number of components is 500)

(a) Background

(b) Smile+brightness

(c) Age+pose

Figure 6: Sequences of image manipulations performed using directions from GANSpace with ICA applied to GAN from Liu et al. (2021) (number of components is 1000)

Table 3: FID scores for GANSpace for GAN from Liu et al. (2021) with ICA

Manipulation strength \ Number of components	500	1000
$\alpha \sim U[-3, 3]$	20.44	20.92
$\alpha \sim U[-6, 6]$	24.13	23.82