Proceedings of the 2nd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/67/20240633

Analyzing abnormalities on age attribute in GAN-control

Haojun Jin[1,†], **Chengju Liu**[2,*,†], **Keyi Liu**[3,†], **Yuman Song**[4,†]

[1] Northcross Shanghai, Shanghai, 201900, China, fret8674@163.com
[2] Department of Computing, The Hong Kong Polytechnic University, Hong Kong, liucjkv@gmail.com
[3] Computer Science, University of Nottingham Ningbo, Ningbo, 31500, China, derekliu827@outlook.com
[4] Department of Mathematics, Imperial College London, London, SW7 2AZ, United Kingdom, colleysong@gmail.com

[†] These authors contributed equally
* Corresponding author: liucjkv@gmail.com

Abstract. The controllable GAN algorithm has the capability of producing high-quality virtual face images. However, the generated results may not always be of optimal quality. The small changes in the sub-vectors of particular parameters may result in big change in the quality and effect of the face images, such as higher occurrence of face distortions and artefect. This paper mainly aims to investigate the reasons behind the suboptimal performance when we alter the age of the image. To achieve this, we analyze the latent space of the outliers and visualize the distance between the outliers using the PCA algorithm. Additionally, this work introduces noise to each sub-vector to gain insights into the potential reasons for the occurrence of outliers. This paper gave a detailed analysis on the inharmonious performance of different features in the latent space, and provided an aspect on the future improvement work in the face synthesis by controllable GAN algorithm.

Keywords: Controllable GAN, PCA, latent space

1. Introduction

GANs have lots of applications in face generation. In a recent study [1], a new method is introduced to disentangle the latent space to get the interpretable explicit control. Using contrastive learning, the latent space can be divided into subspaces and encoded with image properties. Then the MLP (Multi-Layer Perceptron) encoder is designed to correlate specific control parameter values to a related latent subspace. This approach allows for explicit control over each property.

The algorithm is based on the FFHQ set, consisting of 70,000 high-quality PNG images at 1024×1024 resolution and contains considerable variation in age, ethnicity, and image background.

However, the ID of each agent face may still lead to entanglement in age. In other words, there emerge cases where the synthesized face does not grow aging when adjusting the parameter for age (Figure 1). Similar problems exist widely in many other similar models as well, whereas many solutions resort to supervised methods that demand high computational costs.

DOI: 10.54254/2755-2721/67/20240633

Figure 1. One batch with setting age=75.

In this report we try to identify the problem by adding Gaussian noise to each sub-vector. Moreover, applying PCA to project the latent space W, we compare the distances between the latent vectors and the mean vector in the projected space to locate the bad IDs and hence to find out how the difference between those IDs' happens.

2. Background

A group of researchers has recently completed work on an advanced face generation model. The model is based on a variety of existing techniques and approaches, such as StyleGan [2], Relative Control with Entanglement, Conditional GANs and Discrete Control, Disentanglement by Contrastive Learning, and Interpretable Explicit Control.[3]

1. StyleGan:[2] This network consists of two parts, the first part is the Mapping network, in which the hidden variable z is processed and mapped into an intermediate hidden variable w, where w is used to control the style of the generated image. The second part is the Synthesis network, which is the main part of the generated image. The innovation compared to the traditional GAN is that each sub-network is fed with A and B, where A is the affine transformation of w and B is the transformed random noise. It is possible to control the style of the produced image as well as increase the details of the image by adjusting the inputs of A and B. StyleGAN provides a significant improvement over traditional gan in all aspects of image generation under existing standards.

2. Relative Control with Entanglement: The approaches focused on relative control by exploiting GANs' latent space disentanglement [4, 5]. It involves using techniques such as Principal Component Analysis (PCA)[6] to identify boundaries within the latent space. These boundaries separate regions, with each side corresponding to opposite semantic attributes, such as "young" and "old". By navigating latent vectors closer or further away from these boundaries, the strength of the associated attributes can be adjusted. While these methods demonstrate some degree of control, they suffer from entanglement, where changes in one attribute affect others [7]. Therefore, achieving independent control over individual attributes remains challenging.

3. Conditional GANs and Discrete Control: Conditional GANs [8, 9] emerged as a popular choice for controllable image generation, it achieves control over attributes by introducing class labeling inference loss. However, this method was limited to discrete control variables, making them unsuitable for continuous attributes such as age or pose. In addition, controlling multiple attributes at the same time poses may lead to coarse control granularity.

4. Disentanglement by contrastive learning: This method is based on StyleGAN2 [10] architecture, which introduced a factorized contrast loss to train the GAN, the loss encourages latent vectors associated with the same attribute to be close to each other, while vectors associated with different attributes are pushed far apart. As a result, the latent space is divided into multiple subspaces, each encoding different

DOI: 10.54254/2755-2721/67/20240633

image attributes. This separation prevents unnecessary interactions between properties and enables more precise control.

5. Interpretable explicit control [3] This method aims to offer precise control over specific attributes by training MLP encoders. To achieve this, the researchers trained N encoders, each mapping a human attribute to a corresponding potential subspace vector w^k. During inference, the geneator can use any combination of these sub-vectors w^k to synthesize an image. The authors have demonstrated through numerous experiments that this method is not only simple and easy to implement but also does not lose much accuracy compared to other complex methods.

3. Methodology

Our method is based on GAN-Control: Explicitly Controllable GANs [3]. We randomly sample the latent space. Then we change the age and find out the the abnormal picture. We also visualize the features based on the the Euclidean distance in the latent space in the stage of PCA method.

3.1. Motivation

When producing images with abnormalities by substituting the ID sub-vectors of some good images with an ID sub-vector extracted from a bad image, Figure 2 shows the result of this experiment: The five images in row 1, with their age attributes set to 75 years old, are of good quality. There are no obvious artifact or abnormality. In row 2, we select an image with abnormality. The age attribute of the image is also set to 75 years old. From observation, we found that the face of the 75-year-old image still looks like a baby. Then we extracted the ID sub-vector from the bad image's latent space w. After that, we replace the ID sub-vectors of the five good images with the bad ID sub-vector. Row 3 are images generated with the modified latent spaces. By observing the images, we found that the five newly generated images are abnormal as their faces look very young even though the age attribute is set to 75 years old.

Figure 2. Substituting the ID sub-vectors of 5 good images with that from a bad image: Row 1 shows five good images with their age attribute set to 75 years old. Row 2 is a bad image with its age attribute set to 75 years old. Row 3 shows images generated by replacing the ID sub-vectors from row 1 with that from row 2.

3.2. Adding Noise to Each Sub-vector

We decided to investigate further into the latent space W to find out what features of images each sub-vector w^k really control. By adding noise to one sub-vector w^k at a time, we can observe the distorted images. Then we can compare the the original images with the distorted images. We can study the

DOI: 10.54254/2755-2721/67/20240633

differences to know more about the disentanglement of this model. We choose to add Gaussian noise to each sub-vector w^k of the latent space w. Equation 1 is the probability density function of Gaussian distribution where σ is the standard deviation and μ is the mean. With a random Gaussian noise vector $X \sim \mathcal{N}(\mu, \sigma^2)$, the distorted sub-vector $w^{'}$ can be represented in 2 where w^k is the original sub-vector. For the Gaussian noise, we set the mean equals to zero and we manipulate the standard deviation to control the intensity of the noise.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \tag{1}$$

$$w^{'} = w^k + X \tag{2}$$

3.3. Principle Component Analysis(PCA)

The PCA method initially analyzes the distribution of the high-dimensional latent space. It then projects the vectors orthogonally onto the principal axis, resulting in a lower-dimensional space that preserves the main features of the original high-dimensional space. We measure the distance (equation 3) between initial latent space in Euclidean space and latent space after changing ages.

$$d(A, B) = ||A - B||_2 = \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2)} \tag{3}$$

Here we try to visualize the latent space into a 2D-figure using PCA to explore the distribution of the vectors. With 1600 vectors mapping into latent space W, Figure 3 shows the shape of distribution, which can be enclosed by a convex curve.

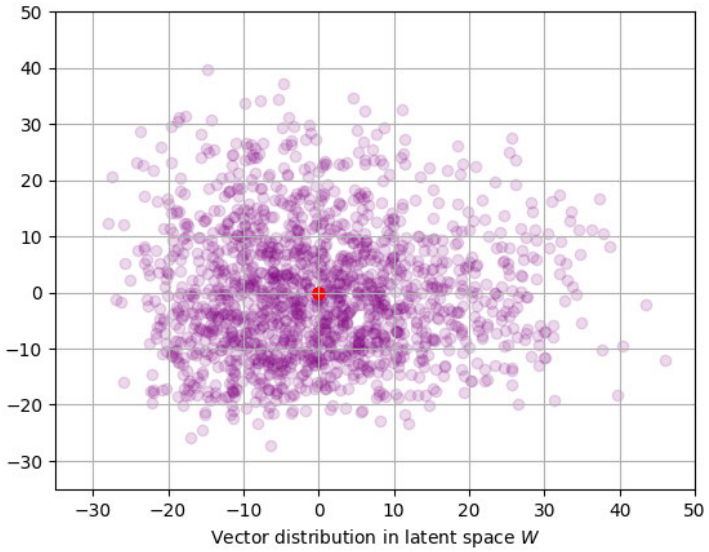


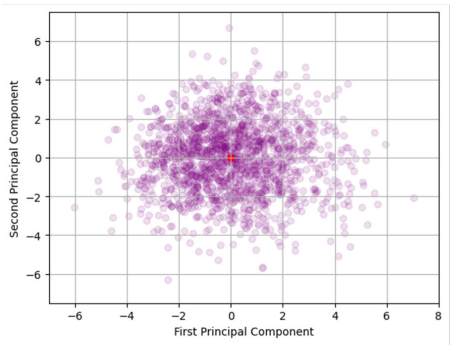
Figure 3. Visualize latent space W using PCA

As the ID attribute is generally independent of other attributes except for age attribute, we aim to focus on the performance of the single ID attribute to further study its entanglement with the age attribute. Similar to the above method, we get the projected figure for the vectors composing the ID attribute.

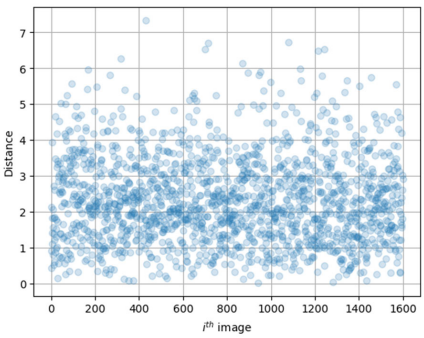
DOI: 10.54254/2755-2721/67/20240633

Interestingly, the distribution in ID was almost sphere(Figure 4(a)), which may represent the distribution of features within the image content.

It can be better observed through the distances between the ID vectors and the mean. In the next section we will analyze the ID attribute in the latent space W in-depth.



(a) Visualizing the ID attribute vectors.



(b) Euclidean distance distribution for all 1600 vectors(only focusing on ID attributes).

Figure 4. ID vectors visualization and related distances to the mean: red dot (at origin) in (a) is the mean, which is rounded to 0 under PCA

4. Experimental Results

4.1. Results Analysis for Adding Noise

We randomly generate the initial five images(Figure 5). Each time, we add Gaussian noise to one sub-vector w^k of their latent space w to examine the differences.

Figure 5. Five randomly sampled images: they are the original images before adding noise

In Figure 6, we add noise to the ID sub-vectors of the five original images. When we set the standard deviation of the noise to 1, we can observe changes in the regions of eyes, noise and mouth. The distorted images still seem somewhat realistic. When we set the standard deviation of the noise to 1.3, the shapes of eyes, noise and mouth transform drastically. The overall shapes of the faces also change. We also observe that image features like background and color of the hat also change after adding noise to the ID sub-vector.

DOI: 10.54254/2755-2721/67/20240633

(a) Original **(b)** $\sigma = 2$

(c) $\sigma = 1.3$

Figure 6. Distorted images generated by adding noise to ID sub-vector: (a) shows the original images. In (b) we add noise sampled from $\mathcal{N}(0,\ 1^2)$. In (c) we add noise sampled from $\mathcal{N}(0,\ 1.3^2)$.

In Figure 7, we add noise to the head orientation (pose) sub-vectors of the five original images. When the standard deviation of the noise is 1, we noticed that the differences between the original images and the distorted images are nearly indistinguishable. When the standard deviation is 1.5, we observed some minor changes in the background, color, and head orientation.

(a) Original **(b)** $\sigma = 1$

(c) $\sigma = 1.5$

Figure 7. Distorted images generated by adding noise to head orientation sub-vector: (a) shows the original images. In (b) we add noise sampled from $\mathcal{N}(0,\ 1^2)$. In (c) we add noise sampled from $\mathcal{N}(0,\ 1.5^2)$.

In Figure 8, we add noise to the age sub-vectors of the five original images. When the standard deviation of the noise is 1, we can see minor changes in skin texture. When the standard deviation of the noise is set to 1.5, we can see more wrinkles on the faces. The color of the pupils of the second image from the left also changes. The color of the background from the last image from the left also changes significantly.

(a) Original **(b)** $\sigma = 1$

(c) $\sigma = 1.5$

Figure 8. Distorted images generated by adding noise to age sub-vector: (a) shows the original images. In (b) we add noise sampled from $\mathcal{N}(0,\ 1^2)$. In (c) we add noise sampled from $\mathcal{N}(0,\ 1.5^2)$.

DOI: 10.54254/2755-2721/67/20240633

In Figure 9 we add noise the the sub-vector that controls hair color. We do not observe significant changes in hair color, however, some other attributes change drastically. For example, the overall color tone of the first image form the left becomes green when the standard deviation of the noise is 1.5. The color of the hat of the second image form the left changes instead of the color of the hair. The background color of the other three images changes significantly adding noise to the hair color sub-vector.

(a) Original **(b)** $\sigma = 1$ **(c)** $\sigma = 1.5$

Figure 9. Distorted images generated by adding noise to hair color sub-vector: (a) shows the original images. In (b) we add noise sampled from $\mathcal{N}(0,\ 1^2)$. In (c) we add noise sampled from $\mathcal{N}(0,\ 1.5^2)$.

We show that the disentanglement of this model is not perfect. The disentanglement learned for the head orientation sub-vector shows satisfying results, while the disentanglement of the other sub-vectors are relatively weak. We also identify that the ID sub-vector controls the shape of the face.

4.2. Results Analysis for PCA

By K-means clustering algorithms we then cluster the w vectors into 3 main types, where most of the bad ID(resulting in abnormalities under age changing) are densely distributed. We show these 3 types on the distance distribution (Figure 10), where the bands denote the area of 95% confidence interval centered at the red line corresponding to the distances. To better fit the main bad IDs' density, we add white noise to control the fluctuation in the predictions.

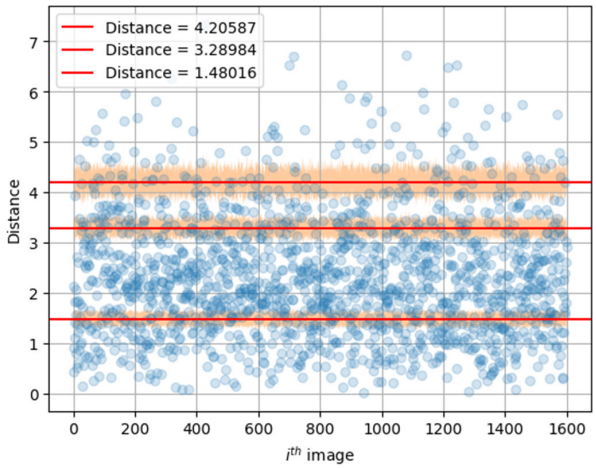


Figure 10. Locating the Bad IDs clusters on distance distribution.

It is easily observed that the bands centered at 4.20587 are wider than the other two types, which means the variance is larger, possibly as a result of the widespread dots in this area. Relatively, the group with

DOI: 10.54254/2755-2721/67/20240633

distance=1.48016 has the smallest variance, but from the graph, there are a lot of vectors around the area within this distance interval, which makes it a little harder to locate and compare the bad IDs.

4.3. Experimental Analysis

After locating the bad IDs in 3 different groups, we get the experimental results.

There are large amounts of artifacts when generating faces at age= 5, showing irregular color lumps on the faces and malformed hands, while the performance goes better off during age growing. From Figure 11, the cases of abnormalities in growing older emerge more frequently than the other two groups. In contrast, the effect of this kind of abnormality is the slightest in the third group(Figure 13), which may be the result of highly overlapped IDs in the area such that it is more difficult to split the bad IDs from the whole groups.

Figure 11. Distance= 4.20587, variance= σ_1, age= $5, 18, 50, 75$

Figure 12. Distance= 3.28984, variance= σ_2, age= $5, 18, 50, 75$

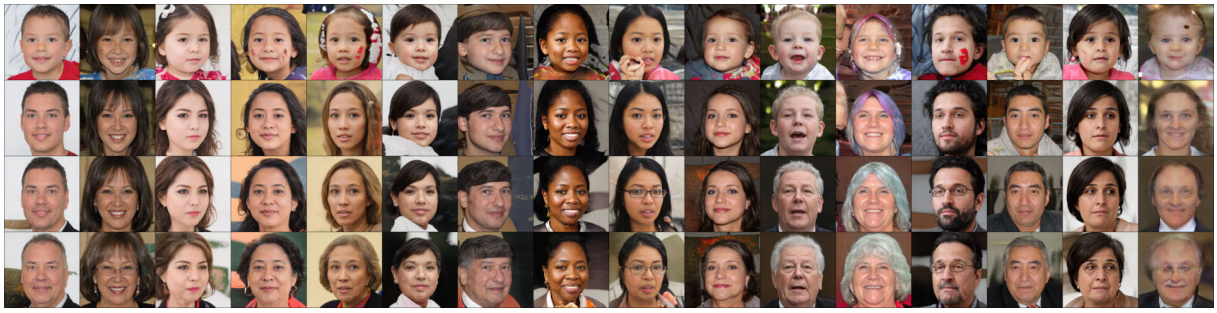


Figure 13. Distance= 1.48016, variance= σ_3, age= $5, 18, 50, 75$

5. Conclusion

5.1. Discussion

Although the paper of GAN-Control [1] mentions the model isolated each attribute according to the usage of the subspace of latent spaces, by experiment in 4.1, we prove that each attribute does not

DOI: 10.54254/2755-2721/67/20240633

separate perfectly. It is hard to separate each attribute into each sub-space. We try to classify the groups corresponding to bad IDs by looking at the distances to the mean in the latent space. The PCA result shows some correlations and can give a good explanation but splitting all the bad IDs from the whole may be hard work and hence the results may not be quite convincing by only watching the latent space.

5.2. Future work

Future work can be done to compare good IDs' latent space and bad IDs' latent space more rigorously to find out which clusters of the ID latent space cause the bad performance of generated image when we change the ages. Also, more work can be done to identify the underrepresented regions in the FFHQ dataset [2] to strengthen the dataset.

Acknowledgments

Chengju Liu, Haojun Jin, Keyi Liu, and Yuman Song contributed equally to this work and should be considered co-first authors.

References

[1] Alon Shoshan, Nadav Bhonker, Igor Kviatkovsky, and Gerard Medioni. Gan-control: Explicitly controllable gans, 2021.

[2] 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.

[3] Alon Shoshan, Nadav Bhonker, Igor Kviatkovsky, and Gerard Medioni. Gan-control: Explicitly controllable gans. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 14083–14093, 2021.

[4] Erik Härkönen, Aaron Hertzmann, Jaakko Lehtinen, and Sylvain Paris. Ganspace: Discovering interpretable gan controls. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 9841–9850. Curran Associates, Inc., 2020.

[5] Yujun Shen, Jinjin Gu, Xiaoou Tang, and Bolei Zhou. Interpreting the latent space of gans for semantic face editing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9243–9252, 2020.

[6] Chi-Hieu Pham, Saïd Ladjal, and Alasdair Newson. Pca-ae: Principal component analysis autoencoder for organising the latent space of generative networks. *Journal of Mathematical Imaging and Vision*, 64(5):569–585, 2022.

[7] Yujun Shen, Ping Luo, Junjie Yan, Xiaogang Wang, and Xiaoou Tang. Faceid-gan: Learning a symmetry three-player gan for identity-preserving face synthesis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 821–830, 2018.

[8] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*, 2018.

[9] Augustus Odena, Christopher Olah, and Jonathon Shlens. Conditional image synthesis with auxiliary classifier gans. In *International conference on machine learning*, pages 2642–2651. PMLR, 2017.

[10] 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.