Conjugate Gradient for Latent Space Manipulation

Walid Messaoud[1], Rim Trabelsi[2], Adnane Cabani[3] and Fatma Abdelkefi[1]
[1] *Supcom Lab-MEDIATRON, Carthage University, Ariana, Tunisia*
[2] *Hatem Bettaher IResCoMath Laboratory, National Engineering School, University of Gabes, Tunisia*
[3] *Université de Rouen Normandie, ESIGELEC, IRSEEM, 76000 Rouen, France*

Keywords: Generative Adversarial Networks, Latent Space Manipulation, Conjugate Gradient, StyleGAN, Neural Network.

Abstract: Generative Adversarial Networks (GANs) have revolutionized image generation, allowing the production of high-quality images from latent codes in the latent space. However, manipulating the latent space to achieve specific image attributes remains challenging. Existing methods often lack disentanglement, leading to unintended changes in other attributes. Moreover, most of the existing techniques are limited to one-dimensional conditioning, making them less effective for complex multidimensional modifications. In this paper, we propose a novel approach that combines an auxiliary map composed of convolutional layers and Conjugate Gradient (CG) to enhance latent space manipulation. The proposed auxiliary map provides a versatile and expressive way to incorporate external information for image generation, while CG facilitates precise and controlled manipulations. Our experimental results demonstrate better performance compared to state-of-the-art methods.

1 INTRODUCTION

Generative Adversarial Networks (GANs) have emerged as a groundbreaking paradigm in the field of image generation, revolutionizing the way of producing high-quality images. Developed by Ian Goodfellow et al. Goodfellow et al. (2014), GANs have since become a cornerstone of modern machine learning and artificial intelligence research. They achieve remarkable success in image synthesis, producing high-fidelity images from randomly sampled latent codes in the latent space. However, the ability to manipulate the latent space to generate images with specific attributes or properties remains a challenging task, particularly when it comes to achieving multidimensional conditioning.

While existing methods for latent space manipulation have shown promising results, they often face limitations in disentangling manipulations, leading to unintended changes in other attributes. Moreover, many techniques are restricted to one-dimensional conditioning, limiting their applicability for complex modifications that require multidimensional information. In this article, we introduce a groundbreaking methodology for latent space manipulation by harnessing the synergistic potential of an innovative auxiliary map composed of convolutional layers, Swish activation Ramachandran et al. (2017), and group normalization Wu and He (2018). This approach represents a novel paradigm in the realm of latent space manipulation, offering an unparalleled and dynamic avenue to seamlessly integrate external information into the image generation process. Furthermore, we introduce the utilization of CG, which helps overcome the challenges associated with optimization-based methods. By training an auxiliary mapping network that induces a CG method, we enable more precise and disentangled manipulations in the latent space. In order to assess the effectiveness of our proposed method, we conduct a series of quantitative experiments utilizing various metrics to evaluate the disentanglement capabilities of different manipulation techniques using Flickr-Faces-HQ (FFHQ) Karras et al. (2019) and CelebAHQ datasets Karras et al. (2017). Our approach is then compared to the state-of-the-art methods to establish its performance. The results from these quantitative experiments demonstrate the superiority of our method in achieving highly effective disentanglement and precise control over image manipulations. This paper marks a substantial progression in GAN-based image manipulation, ushering in fresh avenues for generating images that incorporate multidimensional information, including keypoints and textual descriptions. By harnessing the power of convolutional layers within an auxiliary map and leveraging the CG methodology, we have propelled the boundaries of latent space manipulation. The main contributions of

Messaoud, W., Trabelsi, R., Cabani, A. and Abdelkefi, F.
Conjugate Gradient for Latent Space Manipulation.
DOI: 10.5220/0012268700003636

In *Proceedings of the 16th International Conference on Agents and Artificial Intelligence (ICAART 2024) - Volume 3*, pages 50-57
ISBN: 978-989-758-680-4; ISSN: 2184-433X

this paper are the followings:

- We propose a pioneering utilization of auxiliary map composed of convolutional layers for latent space manipulation, unlocking innovative possibilities in GAN image generation.
- We leverage CG for precise and controlled latent space manipulations, effectively addressing optimization challenges faced by existing techniques.
- We introduce disentangled and multidimensional conditioning, overcoming limitations with attribute adjustments and supporting complex modifications in GANs.

The remainder of this paper is organized into the following sections. In Section II, we discuss related works and examine different techniques for latent space manipulation in GANs. We provide an overview of the existing attribute manipulation approaches. Section III presents our proposed methodology, detailing the use of auxiliary mapping, classifier, and CG for latent space manipulation. In Section IV, we depict the evaluation metrics and the implementation details of our extensive experiments. Section V evaluates the performance of our approach on facial attribute editing. We compare our results with the state-of-the-art methods and discuss the outcomes and their implications, highlighting the significant performance of our method in achieving accurate and visually appealing attribute manipulations. Finally, in Section VI, we conclude and highlight the main perspectives of this work.

2 RELATED WORK

Within this section, we delve into the world of Generative Adversarial Networks (GANs) and their remarkable impact on image synthesis. Going deeper, we unveil the intricacies of manipulating latent spaces through cutting-edge techniques.

2.1 Generative Adversarial Networks

Recent advancements in generative models have led to remarkable progress in generating high-quality, photo-realistic images. Notably, various GAN models, such as PG-GAN Karras et al. (2017), BigGAN Brock et al. (2018), StyleGAN Karras et al. (2019), have demonstrated their ability to synthesize realistic face images with diverse attributes, expressions, ages, backgrounds, and viewing angles. These GANs encode rich semantic information in intermediate features and latent spaces, enabling high-quality face image generation. However, one limitation of GANs is the absence of inference functions or encoders, restricting the application of latent space manipulations solely to GAN-generated images and not real-world images. To address this limitation, GAN inversion methods have been proposed, such as the works investigated by Abdal et al. Abdal et al. (2019) and Xia et al. Xia et al. (2021), allowing the inversion of real images back into the latent space of pre-trained GAN models. This process bridges the gap between real and fake face image domains, resulting in improved quality of the generated face images. However, existing GAN inversion methods are often specific to individual GAN architectures which can limit their broader application.

2.2 Manipulation on Latent Vector

Vector arithmetic in the latent space, as introduced by early GAN research Radford et al. (2015), enables semantic editing of generated images and has been further explored in recent works. However, a comprehensive understanding of a well-trained GAN's latent space and its encoded semantics remains incomplete.

Concurrent research by Jahanian et al. Jahanian et al. (2019) and Yang et al. Yang et al. (2021) has made significant contributions to exploring latent semantics in GANs. In [10], authors proposed Interface-GAN, a novel framework for semantic face editing, interpreting latent semantics learned by GANs. They discovered that the well-trained generative models learn a disentangled representation through linear transformations in the latent space, enabling precise control of facial attributes. Interface-GAN demonstrated the manipulation of various facial attributes, including gender, age, expression, eyeglasses, face pose, and artifact fixing. This method even allows for real image manipulation when it is combined with GAN inversion methods or encoder-involved models.

GANSpace, proposed by Härkönen et al. Härkönen et al. (2020), provides a simple technique to analyze GANs and create interpretable controls for image synthesis. They identify important latent directions using Principal Component Analysis (PCA) applied either in the latent space or feature space. By perturbing the latent space along these principal directions, GANSpace allows for diverse and interpretable image edits, such as changes in viewpoint, aging, lighting, and time of day. The researchers extended their approach to control BigGAN with layer-wise inputs in a StyleGAN-like manner, achieving impressive results across different GAN architectures and datasets.

Additionally, the Surrogate Gradient Field (SGF) method proposed by Wang et al. Li et al. (2021) enables manipulation with multidimensional conditions, such as keypoints and captions. The algorithm searches for a new latent code that satisfies the target conditions by leveraging the Surrogate Gradient Field. This approach opens up possibilities for controlling and manipulating GAN-generated images based on complex and high-dimensional attributes, providing a powerful tool for creative image synthesis. In addition to these works, attribute manipulation in generated images has been studied using both supervised and unsupervised methods. Shen et al. in Shen et al. (2018) employed a linear Support Vector Machine (SVM) as an additional classifier to label properties and adjust attributes in the latent space. On the other hand, self-supervised learning Voynov and Babenko (2020), directly discover semantically meaningful directions. These diverse approaches collectively contribute to the understanding and the capability to control and interpret the latent representations learned by GANs for image synthesis and editing. While current approaches to latent space manipulation have displayed encouraging outcomes, they frequently encounter challenges in effectively disentangling manipulations, which can result in inadvertent alterations to other attributes. Furthermore, numerous methodologies are confined to single-dimensional conditioning, curtailing their utility for intricate modifications necessitating multidimensional information.

3 ENHANCED LATENT VECTOR SYNTHESIS VIA CG METHOD

In this section, we delve into the methodology that underpins our approach, which enables the synthesis of optimized latent vectors through the application of the CG method. By employing this technique, we address the challenge of latent space manipulation within GANs with the aim of achieving specific image attributes.

3.1 Problem Statement

Latent space manipulation within Generative Adversarial Networks (GANs) entails the deliberate adjustment of latent vectors to achieve specific attributes in the resulting generated images. In the context of a pretrained GAN generator $G : Z \to X$, where Z is a d-dimensional latent space and X represents the space of generated images, a classifier network $C : X \to C$ predicts semantic properties c from generated images x. Given an initial latent vector $z_0 \in Z$ associated with properties $c_0 = \Phi(z_0)$ and a desired target property c_1, the primary objective is to determine a new latent vector $z_1 \in Z$ such that when fed through the generator G, it produces an image whose predicted properties $\Phi(z_1)$ align precisely with the target property c_1. This process enables a controlled transformation of latent vectors to achieve precise image attribute manipulation.

3.2 Learning the Auxiliary Mapping

StyleGAN2, renowned for its exceptional power, effortlessly generates a diverse array of images with the desired properties c_1. However, our primary objective is to achieve these properties with minimal undesired alterations to the image. Intuitively, we can perturb the vector z_0 in the Z space to obtain a new vector z_1 that closely aligns with z_0. Nevertheless, empirically, the existing gradient field is unsuitable for this purpose. Hence, our focus lies in introducing a novel gradient field that effectively addresses this limitation.

To achieve this, we introduce an auxiliary mapping $F : Z \times C \to Z$, satisfying the condition:

$$F(z, \Phi(z)) = z, \quad \forall z \in Z \tag{1}$$

where F is implemented as a multi-layer neural network and trained using a simple reconstruction loss.

3.3 Deriving the CG Update Process

The CG method introduces a powerful iterative approach for dynamically refining latent vectors within the latent space. This iterative update process is formalized as follows:

$$x_{n+1} := x_n + \alpha_n d_n, \tag{2}$$

$$d_{n+1} := -\nabla f(x_{n+1}) + \beta_{n+1} d_n. \tag{3}$$

In this specific context, let's delve into the significance of each variable:

- x_n is the current latent vector being adjusted within the latent space.
- f is the loss function, and $\nabla f(x_{n+1})$ represents the gradient of the loss function concerning the latent vector at the subsequent iteration. This gradient guides the optimization process by indicating the direction of steepest ascent.
- d_n symbolizes the search direction in the latent space. It guides the update direction for the latent vector, providing a trajectory that facilitates convergence.

- α_n, where $n \in \mathbb{N}$, is a sequence of step sizes. These step sizes control the magnitude of the updates applied to the latent vector. They balance the trade-off between faster convergence and overshooting.
- $\beta_{n+1} \in \mathbb{R}^+$ is a coefficient that modulates the combination of the current search direction d_{n+1} and the previous search direction d_n. This coefficient is crucial for maintaining conjugacy and efficient convergence.

The determination of the direction d_{n+1} at the $(n+1)$-th iteration embodies insights from both the current gradient $\nabla f(x_{n+1})$ and the historical search direction d_n. This amalgamation of information facilitates a balanced and informed adjustment process that enhances efficiency and efficacy.

3.4 Algorithm: CG Method for Latent Space Manipulation

The CG method for latent space manipulation algorithm aims to optimize a latent vector in the context of GANs. Given a Generator (G), Classifier (C), and Auxiliary mapping (F), the algorithm iteratively updates the latent vector to achieve a desired target attribute.

Algorithm 1: CG Method for Latent Space Manipulation.

Require: Generator G, Classifier C, Auxiliary mapping F, Initial latent vector z_0, Target attributes c_1, Number of iterations n, Learning rate α
Ensure: Optimized latent vector z_n

```
c_0 ← C(G(z_0))
δc ← α(c_1 − c_0)
c^(0) ← c_0
d^(0) ← 0                                   ▷ Initialize search direction
z^(0) ← z_0
for i = 1 to n do
    g^(i−1) ← ∇_z C(G(z^(i−1)))
    β^(i) ← |g^(i−1)|^2 / |g^(i−2)|^2       ▷ Compute CG coefficient
    d^(i) ← −g^(i−1) + β^(i) d^(i−1)        ▷ Compute search direction
    z^(i) ← z^(i−1) + α d^(i)               ▷ Update latent vector using search direction
    c^(i) ← C(G(z^(i)))
    if c^(i) is close to c_1 then
        return z^(i)
    end if
end for
return z_n
```

The proposed *Algorithm 1* takes as input the initial latent vector z_0, target attributes c_1, number of iterations n, and learning rate α. It starts by computing the initial attribute value c_0 using the classifier on the generated image from z_0. The difference between c_1 and c_0 is scaled by the learning rate α to obtain δc. In each iteration, the algorithm updates the latent vector z using the auxiliary mapping F, the previous latent vector $z^{(i-1)}$, and the current attribute value $c^{(i)}$. The search direction d is computed based on the gradient of the classifier's output with respect to the latent vector and a CG coefficient β. The latent vector is then updated by adding the search direction scaled by the learning rate α. The classifier is applied to the updated latent vector to obtain the current attribute value $c^{(i)}$. If $c^{(i)}$ is close to c_1, the algorithm terminates and returns the optimized latent vector $z^{(i)}$. The algorithm repeats this process for the specified number of iterations n. If no satisfactory attribute value is achieved, it returns the last latent vector z_n. The CG method as depicted in Fig.2 and *algorithm 1* offers an efficient and effective approach for optimizing the latent space in GANs, enabling precise manipulation of the generated images' properties. We opt to use the CG method for latent space manipulation in Generative Adversarial Networks (GANs) stems from its exceptional efficiency Shewchuk et al. (1994). Unlike conventional gradient-based techniques, the CG method excels in navigating non-convex optimization landscapes, a characteristic vital for precise and controlled latent vector adjustments Powell (1984). The CG method's ability to iteratively refine latent vectors while determining optimal step sizes enhances our manipulation process's quality and efficiency. In contrast to evolutionary algorithms that may suffer from slow convergence, the CG method strikes a balance between convergence speed and computational feasibility. This choice not only ensures the accuracy and stability but also broadens the horizons of GAN-based image generation applications

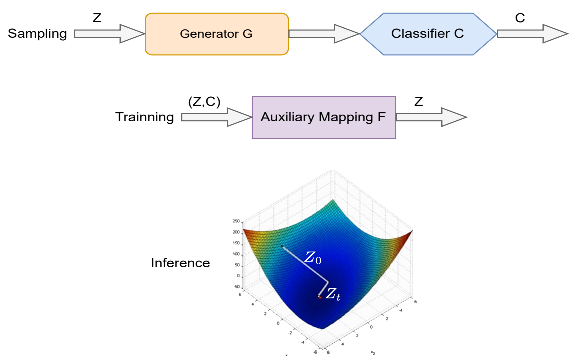

Figure 1: Principle of the CG method.

3.5 Implementation Details of the Auxiliary Mapping F

The auxiliary map embodies a sophisticated mechanism designed to facilitate intricate transformations within the latent space of StyleGAN2. This map acts as an intermediary between latent vectors and associated label information, orchestrating a multi-step process of latent vector manipulation.

Comprising a series of convolutional layers, normalization procedures, and activation functions, the auxiliary map endeavors to reshape latent vectors in a purposeful manner. These convolutional layers serve as adaptable filters, allowing the map to identify and emphasize specific features within the latent vectors. The subsequent normalization steps ensure that the transformed vectors maintain a balanced distribution, preventing distortions during the manipulation process. Activation functions like Swish are strategically applied to induce non-linearity, introducing complex relationships between latent vector components Ramachandran et al. (2017). The basic network architecture of the auxiliary mapping network is illustrated in Fig.3. As the manipulation unfolds, the auxiliary map collaborates with the provided label information. This interaction serves as a guiding force, steering the map's transformations toward desired attribute changes. By iteratively applying these transformations, the map progressively molds the latent vectors to adhere to the intended alterations. The resulting manipulated latent vectors can then be used to generate images that exhibit the targeted attributes. In essence, the auxiliary map acts as a dynamic bridge between the latent space and the generated images, enabling deliberate control over specific image attributes. Its ability to delicately adjust latent vectors based on label cues opens up innovative avenues for generating images with customized characteristics while preserving the essential attributes encoded within the latent space.

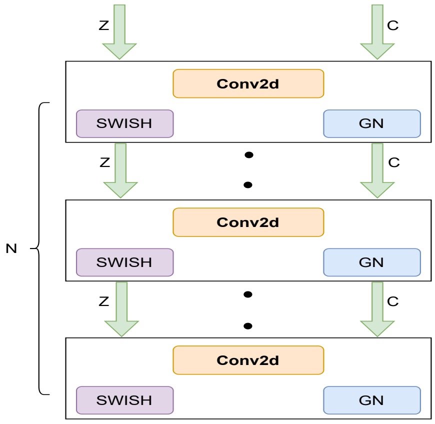

Figure 2: The architecture of the auxiliary mapping network.

4 EXPERIMENTS

4.1 Generator Models and Datasets

To assess facial attributes editing, we utilize a pretrained FFHQ StyleGAN2 Karras et al. (2020) and CelebAHQ StyleGAN2 Karras et al. (2017) alternatively as the generator. For the classifier, we finetuned a pretrained SEResNet50 Hu et al. (2018) model obtained from the VGGFaces2 Cao et al. (2018) dataset. FFHQ represents a premium collection of facial data, comprising 70,000 high-definition face images in PNG format, each boasting a resolution of 1024×1024. This dataset exhibits a wide range of age groups, ethnicities, and image backgrounds, which results in prominent variations in facial attributes. As a consequence, it presents an excellent opportunity for the development of face attribute classification or face semantic segmentation models. The CelebA-HQ dataset is a high-quality version of CelebA that consists of 30,000 images at 1024×1024 resolution.

4.2 Evaluation Metrics

4.2.1 Manipulation Disentanglement Score (MDS)

evaluates the balance between accuracy and disentanglement in achieving a manipulation goal. It involves altering an image to reach a desired outcome. To compute the MDS, we plot a Manipulation Disentanglement Curve (MDC) by gradually increasing manipulation strength and recording accuracy and disentanglement measures. A method with an MDC closer to 1 indicates superior overall disentanglement. In experiments involving attributes manipulation with N samples, manipulation accuracy is the success rate of achieving the target attributes. Manipulation disentanglement is calculated based on the number of attributes, other than the target attribute, that have changed during manipulation. We consider an attribute which is changed if the score changes more than 0.5 during the manipulation. Suppose that there are N_s samples which successfully have their attributes changed to the target attributes. The manipulation accuracy is then the success rate N_s/N. For sample i, if n_i attributes other than the target attribute have changed, we can use the following sum to quantify the manipulation disentanglement:

$$\frac{1}{N}\sum_{i=1}^{N}\left(1-\frac{n_i}{M-1}\right), \quad (4)$$

Where, N represents the total number of samples, and M is the total number of attributes. The term $\frac{n_i}{M-1}$ calculates the ratio of attributes other than the target attribute that have changed for sample i. This sum evaluates the average manipulation disentanglement across all samples."

4.2.2 Mean Square Error loss

Given a dataset with n samples, let y_i be the observed or ground truth value, and $\hat{y}_i$ be the predicted or estimated value for the i-th sample. The Mean Square Error (MSE) is computed as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2. \quad (5)$$

A lower MSE indicates that the predicted values are closer to the ground truth values, suggesting better accuracy and performance. Conversely, a higher MSE indicates larger discrepancies between the predicted and actual values, implying lower accuracy and higher error in the predictions.

4.3 Implementation Details

The auxiliary map F is constructed of a convolutional model, composed of a series of convolutional layers that facilitate latent space manipulation. This innovative model architecture consists of multiple layers. Each convolutional layer is strategically integrated with the group normalization and Swish activation function, promoting disentanglement and attribute-conditioned mappings. The convolutional model is trained with a batch size of 8 using the Adam optimizer with a learning rate of 0.0002. This novel architecture successfully achieves the disentanglement of attributes and latent codes, facilitating effective latent space manipulation while ensuring network stability. Moreover, we incorporate the CG method to enhance latent space manipulation. This iterative algorithm adjusts latent vectors to achieve the desired attribute alterations while adhering to specific constraints. By integrating the CG method into our framework, we achieve precise and controlled manipulations within the latent space, further enhancing the versatility of our approach for various image generation tasks.

5 RESULTS AND DISCUSSION

In this section, we present the results of our comprehensive analysis and evaluations. We delve into a detailed comparison of MDS for facial attribute editing using various methodologies on both the FFHQ-Attributes and CelebAHQ-Attributes datasets. Additionally, we explore the quality of the generated outputs through the MSE comparisons. These results shed light on the performance of our proposed method in comparison to state-of-the-art techniques. We discuss the implications and significance of these findings, highlighting the potential of our approach for advancing the field of generative models and its practical applicability in real-world scenarios.

5.1 MDS Comparison

Table 1 shows a comprehensive comparison of MDS for facial attribute editing on the FFHQ-Attributes dataset. Notably, our proposed approach achieves an outstanding MDS score of **0.855**, surpassing all other reference methods. GANSpace, InterfaceGAN, and Surrogate Gradient Field attain MDS scores of 0.531, 0.721, and 0.837, respectively. These results highlight the superior performance of our method, which harnesses the power of the auxiliary mapping and leverages CG to manipulate the latent space. The significantly higher MDS score attained by our approach signifies its exceptional ability to disentangle and manipulate facial attributes effectively, outperforming state-of-the-art techniques. This demonstrates the potential of our approach in achieving more accurate and visually appealing facial attribute editing results.

Table 1: MDS comparison on facial attribute editing on FFHQ dataset.

Method	MDS
GANSpace Härkönen et al. (2020)	0.531
InterfaceGAN Jahanian et al. (2019)	0.721
Surrogate Gradient Field Li et al. (2021)	0.837
CG method (ours)	**0.855**

Table 2 presents a similar comparison of MDS for facial attribute editing on the CelebAHQ-Attributes dataset. InterfaceGAN and Surrogate Gradient Field achieve MDS scores of 0.758 and 0.876, respectively. However, our CG method outperforms both with an MDS score of **0.88**. These results further demonstrate the consistency and robustness of our proposed approach across different datasets.

Table 2: MDS comparison on facial attribute editing on CelebAHQ dataset.

Method	MDS
InterfaceGAN Jahanian et al. (2019)	0.758
Surrogate Gradient Field Li et al. (2021)	0.876
CG method (ours)	**0.88**

5.2 MSE Loss Comparison

The quality of generated outputs from various methods is critically evaluated through MSE values, as presented in Tables 3 and 4. MSE serves as a fundamental metric for quantifying the discrepancy between generated images and their corresponding ground truth images. Lower MSE values indicate a higher level of image fidelity and alignment with the desired attributes.

Table 3: MSE Loss comparison using FFHQ dataset.

Method	MSE
CG method (ours)	**7.8 e-05**
SGF	9 e-04

Table 4: MSE Loss comparison using CelebAHQ dataset.

Method	MSE
CG method (ours)	**9 e-05**
SGF	4 e-04

Examining the FFHQ dataset results (Table 3), we observe a distinct contrast in MSE values between our proposed CG method and the SGF method. Our CG method stands out with an impressively low MSE of 7.8×10^{-5}, which reflects its proficiency in generating images that closely resemble the intended attributes. In comparison, the SGF method yields a relatively higher MSE of 9×10^{-4}, signifying a comparatively greater deviation from the target attributes. This underscored difference emphasizes that our CG method excels in not only preserving the desired attributes but also maintaining a high level of precision and accuracy in the generated images. Turning our attention to the CelebAHQ dataset (Table 4), the superiority of our CG method persists, exhibiting an MSE of 9×10^{-5}. In contrast, the SGF method records an MSE of 4×10^{-4}. This consistent performance across datasets reaffirms the robustness of our approach in consistently generating images that align well with the target attributes, while maintaining realistic and visually coherent appearances. The exceptional capability of our proposed CG method in minimizing MSE values can be attributed to the synergistic integration of the auxiliary mapping and CG technique. Leveraging the auxiliary mapping enhances the model's capacity to capture intricate latent space patterns related to the desired attributes. Meanwhile, the CG technique facilitates an informed exploration of the latent space, leading to better optimization and consequently, more faithful image generation. Beyond the numerical aspect, the implications of lower MSE values extend to the perceptual quality of the generated images. A lower MSE signifies more visually realistic and accurate images, essential for tasks like image editing, synthesis, and manipulation. Such attributes are integral for real-world applications where the quality of generated content is paramount. In summary, our CG method demonstrates exceptional prowess in multiple dimensions: facial attribute editing, latent space manipulation, and image generation. The combination of significantly lower MSE values and superior MDS scores underscores the pragmatic effectiveness of our approach. These findings not only contribute to the advancement of generative models but also lay the groundwork for enhancing the precision and visual appeal of facial attribute editing applications. The established superiority of our method indicates a promising trajectory for future advancements, positioning it as a compelling contender for diverse real-world applications and catalyzing innovation within the realm of computer vision and image processing.

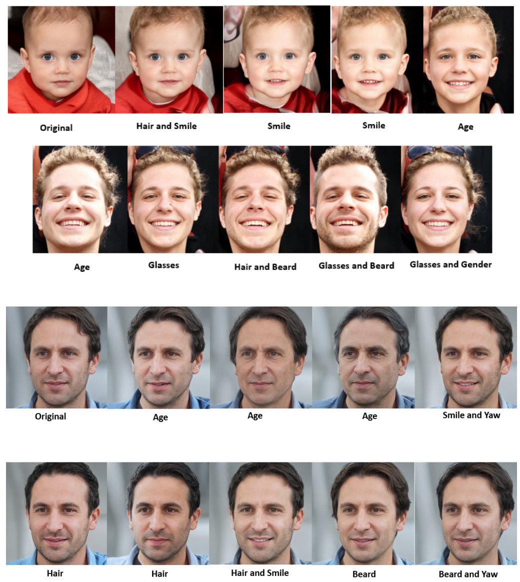

Figure 3: A performed Conjugate Gradient Method to efficiently edit images via manipulating the latent codes of GANs.

In addition to the quantitative evaluations, the qualitative results of our proposed method further affirm its efficacy in enhancing latent space manipulation within GANs. As depicted in Fig.1, we present a selection of meticulously manipulated face images, showcasing attribute transformations from one to another. Our pioneering approach leverages convolutional layers within an auxiliary mapping network, along with the integration of the CG method. These results vividly demonstrate the precision and control our method offers in manipulating diverse attributes, including but not limited to smile, age, glasses, gender, hair, and beard. The seamless transitions from one attribute to another highlight the strength of our method in disentangling complex interactions within the latent space. These qualitative results provide compelling visual evidence of the versatility and potential of our approach for generating images that effectively reflect desired attribute changes while maintaining the essential characteristics of the underlying latent vectors.

6 CONCLUSION

In this paper, we presented a pioneering approach to enhance latent space manipulation in GANs using convolutional layers and CG. Our method leverages the power of auxiliary mapping to disentangle latent semantics and achieve precise attribute manipulations. Through extensive experiments on facial attribute editing, we demonstrated the effectiveness and superiority of our approach, surpassing existing state-of-the-art methods in manipulation disentanglement and image quality. The combination of auxiliary mapping and CG offers a promising direction for advancing GAN-based image generation and opens possibilities for more sophisticated applications, such as generating images with multidimensional information.

REFERENCES

Abdal, R., Qin, Y., and Wonka, P. (2019). Image2stylegan: How to embed images into the stylegan latent space? In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4432–4441.

Brock, A., Donahue, J., and Simonyan, K. (2018). Large scale gan training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*.

Cao, Q., Shen, L., Xie, W., Parkhi, O. M., and Zisserman, A. (2018). Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pages 67–74. IEEE.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27.

Härkönen, E., Hertzmann, A., Lehtinen, J., and Paris, S. (2020). Ganspace: Discovering interpretable gan controls. *Advances in neural information processing systems*, 33:9841–9850.

Hu, J., Shen, L., and Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141.

Jahanian, A., Chai, L., and Isola, P. (2019). On the" steerability" of generative adversarial networks. *arXiv preprint arXiv:1907.07171*.

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

Karras, T., Laine, S., and Aila, T. (2019). 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.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. (2020). Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119.

Li, M., Jin, Y., and Zhu, H. (2021). Surrogate gradient field for latent space manipulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6529–6538.

Powell, M. J. (1984). Nonconvex minimization calculations and the conjugate gradient method. In *Numerical Analysis: Proceedings of the 10th Biennial Conference held at Dundee, Scotland, June 28–July 1, 1983*, pages 122–141. Springer.

Radford, A., Metz, L., and Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.

Ramachandran, P., Zoph, B., and Le, Q. V. (2017). Searching for activation functions. *arXiv preprint arXiv:1710.05941*.

Shen, Y., Zhou, B., Luo, P., and Tang, X. (2018). Facefeat-gan: a two-stage approach for identity-preserving face synthesis. *arXiv preprint arXiv:1812.01288*.

Shewchuk, J. R. et al. (1994). An introduction to the conjugate gradient method without the agonizing pain.

Voynov, A. and Babenko, A. (2020). Unsupervised discovery of interpretable directions in the gan latent space. In *International conference on machine learning*, pages 9786–9796. PMLR.

Wu, Y. and He, K. (2018). Group normalization. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19.

Xia, W., Yang, Y., Xue, J.-H., and Wu, B. (2021). Tedigan: Text-guided diverse face image generation and manipulation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2256–2265.

Yang, C., Shen, Y., and Zhou, B. (2021). Semantic hierarchy emerges in deep generative representations for scene synthesis. *International Journal of Computer Vision*, 129:1451–1466.