High Fidelity GAN Inversion via Prior Multi-Subspace Feature Composition

Guanyue Li[1], **Qianfen Jiao**[2], **Sheng Qian**[3], **Si Wu**[1,2,*] and **Hau-San Wong**[2]
[1]School of Computer Science and Engineering, South China University of Technology, Guangzhou, P. R. China
[2]Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong
[3]Huawei Device Company Limited, Shenzhen, P. R. China
csliguanyue007@mail.scut.edu.cn, qjiao4-c@my.cityu.edu.hk, qiansheng3@huawei.com,
cswusi@scut.edu.cn, cshswong@cityu.edu.hk

Abstract

Generative Adversarial Networks (GANs) have shown impressive gains in image synthesis. GAN inversion was recently studied to understand and utilize the knowledge it learns, where a real image is inverted back to a latent code and can thus be reconstructed by the generator. Although increasing the number of latent codes can improve inversion quality to a certain extent, we find that important details may still be neglected when performing feature composition over all the intermediate feature channels. To address this issue, we propose a Prior multi-Subspace Feature Composition (PmSFC) approach for high-fidelity inversion. Considering that the intermediate features are highly correlated with each other, we incorporate a self-expressive layer in the generator to discover meaningful subspaces. In this case, the features at a channel can be expressed as a linear combination of those at other channels in the same subspace. We perform feature composition separately in the subspaces. The semantic differences between them benefit the inversion quality, since the inversion process is regularized based on different aspects of semantics. In the experiments, the superior performance of PmSFC demonstrates the effectiveness of prior subspaces in facilitating GAN inversion together with extended applications in visual manipulation.

Introduction

Generative Adversarial Networks (GANs) (Goodfellow et al. 2014) have achieved considerable success in high-fidelity image synthesis and downstream applications, e.g., data augmentation (Wu et al. 2019; Zhang et al. 2020), image processing (Li et al. 2020; Liu et al. 2020), and so on. A state-of-the-art GAN-based generative model, such as BigGAN (Brock, Donahue, and Simonyan 2019), PGGAN (Karras et al. 2018) and StyleGAN (Karras, Laine, and Aila 2019; Karras et al. 2020), typically has a high capacity, and the training procedure depends on large-scale training data. To reduce data dependence, a few works explore how to utilize a well-trained generic generator for various tasks (Wang et al. 2020; Shen et al. 2020).

GAN inversion is a promising way to understand and utilize the generative capability of a well-trained network. The goal is to reverse a target image back to a status in the latent space. The obtained latent code is decoded into an image, which is expected to approximate the target image to as accurate an extent as possible. As reported in (Gu, Shen, and Zhou 2020), the expressiveness of a single latent code is less satisfactory for the case where the target image is out of the distribution of the GAN's training data. To address this issue, multiple latent codes are used, and the corresponding intermediate feature maps are combined to improve the inversion performance. However, we find that a number of intermediate feature channels are not necessarily important for some aspects of semantics, and feature composition over all the channels may thus lead to reduced emphasis on the corresponding details.

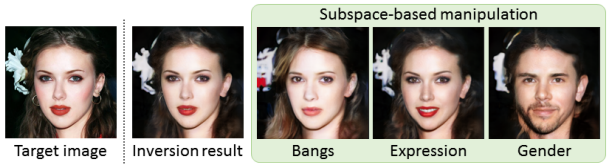

Figure 1: An example to illustrate the capability of the proposed approach to leverage the discovered subspaces over GAN's intermediate feature channels for high fidelity inversion and visual manipulation.

Some recent works (Bau et al. 2018; Shen et al. 2020) discover that the feature maps at an intermediate layer of the generator in a GAN are highly correlated with each other, and they work together to control certain semantics. As shown in Figure 1, we consider that GAN inversion can benefit from this correlation, since the inversion process can be regularized separately based on different aspects of semantics. Toward this end, we apply subspace clustering over GAN's intermediate feature channels. The grouped channels compose meaningful subspaces, which are associated with different visual concepts. We infer latent codes in each subspace, and the resulting separate feature composition serves to enhance the inversion performance.

More specifically, we propose a Prior multi-Subspace feature composition (PmSFC) approach for improving GAN inversion. To explore the degree of correlation among GAN's intermediate feature channels, we divide the generator of a well-trained GAN into two subnetworks by an intermediate layer, and insert a self-expressive layer between them

*Corresponding author.

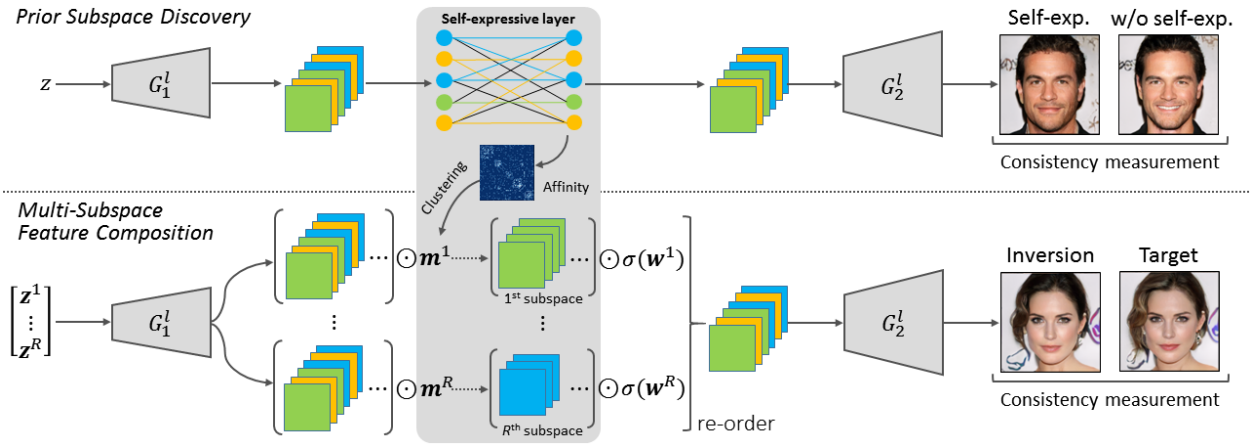

Figure 2: An overview of the proposed PmSFC model for GAN inversion. There are two stages: prior subspace discovery and multi-subspace feature composition. To explore the degree of correlation among GAN's intermediate feature channels, a self-expressive layer is incorporated between two subnetworks G_1^l and G_2^l of a generator. The self-expressiveness property of the channels is leveraged to construct the affinity matrices, and then spectral clustering is adopted to group them into R subspaces. Multiple latent codes are used for feature composition in each subspace. The inversion process benefits from the semantic differences among the subspaces. All the latent codes $\{\boldsymbol{z}^r\}_{r=1}^R$ and composition weights $\{\boldsymbol{w}^r\}_{r=1}^R$ are jointly optimized, and the generator $\{G_1^l, G_2^l\}$ are frozen during training.

as shown in Figure 2. The self-expressiveness property of the channels can be captured during feature reconstruction. Due to the incorporation of sparsity regularization, the layer's weights can be used to compute the affinities among the channels. Further, we adopt spectral clustering to group the channels, and those in each cluster compose a subspace. Instead of the whole feature space, image inversion is performed in multiple subspaces. In each subspace, we assign multiple latent codes and compose the corresponding intermediate features. A nonlinear transformation is applied to the composition weights. The resulting features are further combined and decoded for measuring the consistency with the target image. Since the subspaces may associate with different semantics, the proposed approach is able to focus on various aspects of details. We conduct extensive experiments to verify the effectiveness and superiority of our PmSFC model. In addition to GAN inversion, we also investigate the applicability of PmSFC to a variety of image enhancement tasks.

The main contributions of this work are summarized as follows: (1) We utilize the degree of correlation among GAN's intermediate feature channels to improve its inversion performance. (2) We incorporate a self-expressive layer in the generator of a well-trained GAN to capture the self-expressiveness property, and discover meaningful subspaces via spectral clustering over the channels. (3) There are semantic differences among the subspaces, and different latent codes are thus used together to perform separate feature composition. All the latent codes and composition weights are jointly optimized. (4) Due to the aforementioned improvement techniques, the semantic knowledge learnt by a GAN can be selectively utilized for the inversion task, and better performance can also be achieved on extended image enhancement tasks.

Relative Work

GAN-based Generative Models

GANs have shown superior capability of synthesizing realistic images from latent codes to match the real data distribution. To address the issue of training instability, Wasserstein distance, Lipschitz constraint and other useful techniques have been incorporated into the GAN training process (Arjovsky, Chintala, and Bottou 2017; Gulrajani et al. 2017; Wei et al. 2018). On the other hand, there are a number of works focusing on synthesis quality. Brock et al. (Brock, Donahue, and Simonyan 2019) proposed a BigGAN model, in which larger networks and batches were adopted. The resulting model was capable of synthesizing high-resolution images from complex and large scale datasets. Karras et al. (Karras et al. 2018) progressively increased the capacity of the generator and discriminator, such that the training process can focus on increasingly finer scale details. They also proposed a style-based generator architecture to study latent space disentanglement (Karras, Laine, and Aila 2019). Different from traditional generators, a latent code was mapped to an intermediate feature space. To associate with high-level attributes, the resulting vector was used for adaptive instance normalization at different convolutional layers. By incorporating additional conditions into the generative process, GANs can also be trained for conditional image synthesis (Mirza and Osindero 2014; Nguyen et al. 2017; Miyato and Koyama 2018; Gong et al. 2019).

GAN Inversion

GANs typically have no means of inverting the mapping. To investigate what information are captured by latent codes, GAN inversion have been recently studied. Given a real image, the objective is to infer a latent code to recover the image

to as accurate an extent as possible. Donahue et al. (Donahue, Krahenbuhl, and Darrel 2017) proposed a bidirectional GAN to enable inverse mapping. In addition to a standard GAN framework, a separate encoder is incorporated to explicitly learn the reverse mapping. Similarly, Dumoulin et al. (Dumoulin et al. 2017) proposed to jointly train an encoder and a GAN for inversion. Abdal et al. (Abdal, Qian, and Wonka 2019, 2020) formulated an embedding method to project real images into the latent space of StyleGAN, and various semantic image editing operations were further applied. Perarnau et al. (Perarnau et al. 2016) proposed an invertible conditional GAN for image editing. The semantics can be manipulated by modifying the latent codes. Bau et al. (Bau et al. 2019a,b,c) investigated various strategies of inverting GANs with deeper architectures, and performed image manipulation via a layer-wise inversion method. In (Zhu et al. 2020), a domain-guided encoder was trained not only to encode real images, but to also compete with a domain discriminator to ensure that the reconstructed images are as realistic as possible. To avoid out-of-domain inversion, a semantic regularizer was also applied in the model.

As the generative models are typically differentiable, Creswell and Bharath (Creswell and Bharath 2019) used gradient descent methods to determine the latent code without explicitly learning a separate mapping from the data space back to latent space. In addition to the pixel-level reconstruction loss, Zhu et al. (Zhu et al. 2016) included a perceptual loss in the latent code optimization process. To prevent the reconstructions from getting stuck, Lipton and Tripathi (Lipton and Tripathi 2017) used a stochastic clipping strategy to modify gradients. In (Ma, Ayaz, and Karaman 2018), Ma et al. theoretically analyzed the invertibility of GANs. In particular, the latent code can be effectively deduced from the network output for the case where a low-complexity GAN is used. The target image may be significantly different from GAN's training data, and the reconstruction quality is thus limited by the expressiveness of a single latent code. Gu et al. (Gu, Shen, and Zhou 2020) proposed a multi-code GAN prior (mGANprior)-based inversion method, in which a target image was associated with multiple latent codes. At an intermediate layer, the corresponding features are composed with channel weights. In the above case, the generators in GANs are fixed, and the coupled discriminators are not involved during inversion. To deal with complex real-world images, Pan et al. (Pan et al. 2020) fine-tuned a pre-trained generator in the inversion process. In addition, the coupled discriminator was also used to construct a feature matching loss to guide the generator.

Differences from the existing works. The most related method to this work is mGANprior. Both mGANprior and the proposed PmSFC infer multiple latent codes by minimizing the standard inversion loss, without any re-training or modification of GANs. However, there are fundamental differences between them. mGANprior performs feature composition over all intermediate feature channels, while our PmSFC discovers meaningful prior subspaces and performs separate feature composition in each subspace. Furthermore, to the best of our knowledge, it is the first attempt to explore the self-expressiveness property of the channels and compose subspaces in a GAN. Inclusion of GAN prior subspaces ensures that the inversion process focuses on various aspects of details.

Revisit mGANprior

We recap the multi-code GAN prior model (Gu, Shen, and Zhou 2020) since it serves as the baseline for the proposed approach. The goal of GAN inversion is to recover an arbitrary input x by finding an appropriate latent code z. The reconstruction quality can be improved by optimizing multiple latent codes, along with combining their corresponding intermediate feature maps. More specifically, let l denote an index of intermediate layer, and the lth layer splits a generator G into two subnetworks: G_1^l and G_2^l. The objective is to determine a set of latent codes $\boldsymbol{z} = \{z_1, z_2, \cdots, z_K\}$ and a set of weighting vectors $\boldsymbol{w} = \{w_1, w_2, \cdots, w_K\}$ for image reconstruction as follows:

$$\widetilde{x} = G_2^l\Big(\sum_{k=1}^{K} G_1^l(z_k) \odot w_k\Big), \tag{1}$$

where the dimension of w_k is the same as the number of channels, and the operation $\odot$ represents channel-wise multiplication. To ensure the consistency between the original input and the reconstruction at both low and high levels, the optimization problem is defined as follows:

$$\min_{\boldsymbol{z},\boldsymbol{w}} \|x - \widetilde{x}\|_2^2 + \|V(x) - V(\widetilde{x})\|_1, \tag{2}$$

where V denotes a pre-trained network for feature extraction.

Proposed Method

In order to analyze the relationship among the intermediate feature channels of G, we incorporate an additional self-expressive layer between G_1^l and G_2^l. We aim to discover a number of meaningful subspaces, in which each channel can be represented as a linear combination of other channels. To facilitate GAN inversion, our strategy is to perform feature composition in different subspaces, such that the inversion process is regularized based on different aspects of semantics.

Explore GAN Prior Subspaces

We flatten each feature map of $G_1^l(z)$ into a vector, and stack all the vectors into columns of a matrix $F(z)$. To explore self-expressiveness characteristics of the feature channels, a coefficient matrix S is learnt to ensure that $F(z) = F(z)S$, and we thus define a reconstruction loss as follows:

$$L_{\texttt{recF}} = \mathbb{E}_{z\sim p_0}[\|F(z) - F(z)S\|_2^2], \tag{3}$$

where p_0 denotes a prior distribution of latent codes. We find that minimizing $L_{\texttt{recF}}$ is not enough to ensure semantic consistency after decoding $F(z)$ and $F(z)S$. Small reconstruction errors in the intermediate feature space may lead to significant visual differences. To avoid this situation, we define another reconstruction loss in the data space as follows:

$$L_{\texttt{recD}} = \mathbb{E}_{z\sim p_0}[\|G_2^l(G_1^l(z)) - G_2^l(\mathcal{T}(F(z)S))\|_2^2], \tag{4}$$

where $\mathcal{T}(\cdot)$ denotes a transformation to reverse the flattening operation, such that the output has the same dimensions as the original feature maps.

On the other hand, S is expected to have a block-diagonal structure to some extent, such that the encoded subspaces will be more separable. After including a sparsity regularizer on S, the optimization of the self-expressive layer is formulated as follows:

$$\begin{aligned} &\min_{S} L_{\texttt{recF}} + \lambda L_{\texttt{recD}} + \mu \|S\|_1, \\ &s.t.\ \ diag(S) = 0, \end{aligned} \tag{5}$$

where λ and μ are weighting factors for adjusting the relative importance of the corresponding terms. Note that there is a trivial solution ($S = I$), and we thus require that $diag(S) = 0$. The self-expressive layer can be implemented via the full connection of size $C \times C$ without applying bias and nonlinear activations, where C denotes the number of channels and each feature channel is taken as a node. Based on the resulting S, we construct an affinity matrix A as $A = (|S| + |S^T|)/2$, and apply the spectral clustering algorithm (Ng, Jordan, and Weiss 2002) to determine a set of subspaces over the intermediate feature channels. We outline the training procedure of the proposed subspace discovery approach in Algorithm 1.

Algorithm 1 Pseudo-code of subspace discovery over GAN's intermediate feature channels.

Initialize: Pre-trained generator $\{G_1^l, G_2^l\}$, self-expressive layer S, number of subspaces R, learning rate ς, and number of training iterations Γ.
for $t = 1$ to Γ **do**
  Randomly sample latent code $z \sim p_0$, and feed it to G_1^l to obtain the intermediate feature maps $G_1^l(z)$.
  Flatten the feature maps and convert them to a data matrix $F(z)$.
  Convert $F(z)S$ back to the form of feature maps, and decode them via G_2^l.
  Optimize S by using Adam: $S \leftarrow \texttt{Adam}(\nabla_S(L_{\texttt{recF}} + \lambda L_{\texttt{recD}} + \mu\|S\|_1), S, \varsigma)$.
end for
Construct an affinity matrix $A = (|S| + |S^T|)/2$.
Apply spectral clustering with A.
Compose R subspaces, and represent them via binary vectors $\boldsymbol{m}$.
Return $\boldsymbol{m}$.

Multi-Subspace Feature Composition

We partition the intermediate feature channels into R subspaces for GAN inversion, and the resulting subspaces can be represented by $\boldsymbol{m} = \{m^1, m^2, \cdots, m^R\}$, where $m^r \in \{0,1\}^C$ denotes a binary indicator vector for the rth subspace. The component $m^r(c)$ is set to either 1 or 0, which corresponses to whether the cth feature channel is selected or not.

To reconstruct the original image x, we assign multiple latent codes $\boldsymbol{z}^r = \{z_1^r, z_2^r, \cdots, z_K^r\}$ for each subspace, and each latent code z_k^r is also associated with a weighting vector w_k^r. In the rth subspace, we combine the intermediate features as follows:

$$\boldsymbol{f}^r = \frac{\sum_{k=1}^{K}(m^r \odot G_1^l(z_k^r)) \odot \sigma(w_k^r)}{\sum_{k=1}^{K}\sigma(w_k^r)}, \tag{6}$$

where σ denotes the activation function $tanh(\cdot)$ to regularize the weights into the range of $(-1, 1)$. We consider that different subspaces are associated with different visual concepts, and image reconstruction can thus be improved when combining the corresponding feature maps composed in the subspaces. Toward this end, the original image is approximated as follows:

$$\widehat{x} = G_2^l(\texttt{re-order}(\boldsymbol{f}^1, \boldsymbol{f}^2, \ldots, \boldsymbol{f}^R)), \tag{7}$$

where $\texttt{re-order}(\cdot)$ represents an operation to place the feature maps according to the inherent channel order of G_1^l. To ensure the consistency between the original and reconstructed images at both pixel-level and perceptual-level, we formulate the corresponding optimization problem as follows:

$$\min_{\{\boldsymbol{z}^r\}_{r=1}^R, \{\boldsymbol{w}^r\}_{r=1}^R} \|x - \widehat{x}\|_2^2 + \|V(x) - V(\widehat{x})\|_1. \tag{8}$$

Although Eq.(2) and Eq.(8) are similar in form, there is a significant difference between mGANprior and the proposed model. The former directly learns a set of latent codes and weighting factors in the intermediate feature space, while we explore the meaningful subspaces followed by jointly optimizing the associated latent codes and weighting factors. Inclusion of these subspaces ensures more emphasis on different aspects of semantics, which benefits the inversion quality.

Applicability to Unsupervised Image Enhancement

Image reconstruction is a fundamental application of the proposed subspace-based GAN inversion approach. Based on a well-trained GAN, our model can be further applied to multiple extended tasks, including image colorization, inpainting and super-resolution. In these tasks, a given image x is to be processed for restoration or enhancement. When inputting x, our model is required to output an image $\widehat{x}$ with the desired properties. After adopting the corresponding post-processing for $\widehat{x}$, the resulting image should be visually close to x. Therefore, in the unsupervised case, the proposed inversion model can be trained by minimizing the difference between the output and the original image, and the corresponding optimization problem is formulated as follows:

$$\min_{\{\boldsymbol{z}^r\}_{r=1}^R, \{\boldsymbol{w}^r\}_{r=1}^R} \|x - \mathcal{P}(\widehat{x})\|_2^2, \tag{9}$$

where $\mathcal{P}$ denotes the post-processing function for a specific task, e.g., graying for colorization and downsampling for super-resolution. The training procedure for these tasks is the same as that for image reconstruction.

Experiments

Extensive experiments are conducted to evaluate the proposed PmSFC model on a variety of datasets, including CelebA-HQ (Karras et al. 2018) and LSUN (Yu et al. 2015). These datasets are widely used for image synthesis. In this section, we first introduce the implementation details and experiment configurations. Next, we analyze the effectiveness of subspace

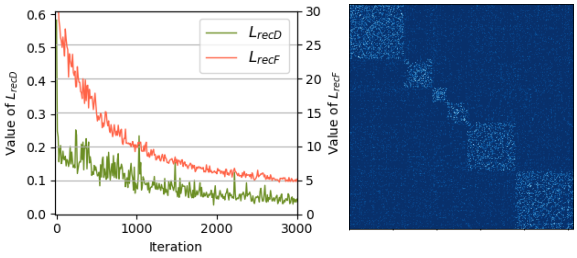

Figure 3: (*left*) The plots of the loss functions used for training the self-expressive layer. (*right*) Visualization of the affinity matrix between channels.

discovery over GAN's intermediate feature channels, and also seek insights regarding what visual concepts are associated with them. Further, we compare the proposed approach with the main competing inversion methods on various tasks.

Experimental Setup

All the experiments are based on a pre-trained PGGAN. Unless otherwise indicated, we divide the generator in PGGAN into two subnetworks by the 3rd layer. For prior subspace discovery, the size of full connection in the self-expressive layer is 512×512. We train the layer for 3000 iterations using the Adam optimizer (Kingma and Ba 2015) with learning rate of 0.0001 and momentum parameters (0.9, 0.999). To reach a balance among the terms in the overall training loss, the weighting factors λ and μ in Eq.(5) are set to 10 and 1, respectively. For the inversion and extended tasks, the settings are the same as above, but the number of iterations increases to 7000.

Evaluation protocol. There are several metrics for quantitatively assessing the reconstructed image quality. The Peak Signal-to-Noise Ratio (PSNR), Structure SIMilarity (SSIM) (Wang et al. 2004) and Naturalness Image Quality Evaluator (NIQE) (Mittal, Soundararajan, and Bovik 2012) are used to measure the low/mid-level similarity between the original and reconstructed images. A high-level metric is the Learned Perceptual Image Patch Similarity (LPIPS) (Zhang et al. 2018a), which is more consistent with human perception.

Subspace Discovery

In this experiment, PGGANs are pre-trained on CelebA and LSUN, respectively. We adopt a self-expressive layer to capture the self-expressiveness property of intermediate feature channels. In Figure 3 (*left*), we plot the values of the loss functions $L_{\texttt{recF}}$ and $L_{\texttt{recD}}$ during training. The reconstruction loss in both feature and data spaces is rapidly reduced. The result indicates that the features at one channel can be effectively reconstructed by those at other channels without incurring changes in semantics. We also visualize the affinities between channels in Figure 3 (*right*), and a block-diagonal structure can be observed. Based on the affinity, we apply spectral clustering to determine a set of subspaces, and investigate what visual concepts are associated with them. Specifically, we specify a subspace, and exchange the corresponding feature maps of the paired images. The second

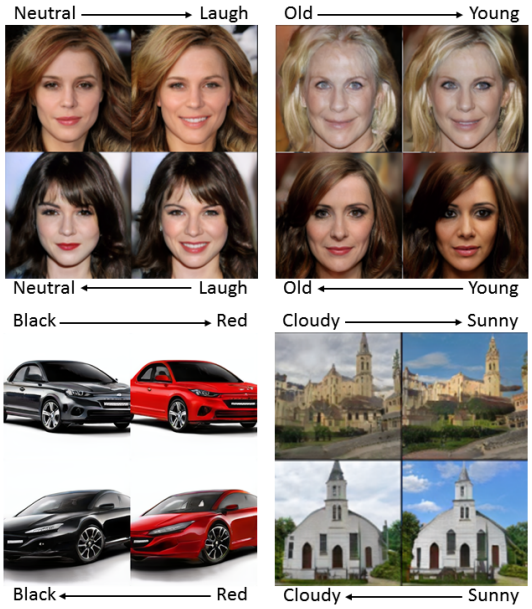

Figure 4: Examples of exchanging feature maps in the subspaces associated with different attributes: expression (*top left*), age (*top right*), color (*bottom left*), and weather (*bottom right*). In each example, diagonal images are the inversion results.

subnetwork G_2^l is used to decode the resulting features. As shown in Figure 4, we find that the obtained subspaces control meaningful attributes, such as expression, age, color and weather. Note that we discover subspaces over intermediate feature channels without a disentangling process, since the generator is frozen during training.

GAN Inversion

We further evaluate the proposed PmSFC model on the GAN inversion task. mGANprior serves as a state-of-the-art inversion method. In addition to mGANprior, we also compare the proposed approach with a number of recent inversion methods, including InvertGenNet (Ma, Ayaz, and Karaman 2018), GenVisMani (Zhu et al. 2016) and InvertLayers (Bau et al. 2019b).

On each dataset, the methods are applied to reconstruct the given images via a pre-trained PGGAN. For a quantitative comparison, all the methods are tested on 300 randomly selected images. The setting is the same as (Gu, Shen, and Zhou 2020). For PmSFC, we group intermediate feature channels into 6 subspaces, and assign 5 latent codes for each one. Different from mGANprior which combines all feature maps, each subspace is only associated with a subset of them, and the number of parameters in PmSFC is thus smaller than that in mGANprior. Table 1 summarizes the performance comparison of the competing methods in terms of LPIPS and PSNR. The results of other competing methods are obtained from the existing literature, and our experimental settings are compatible with them. The proposed PmSFC model achieves lower LPIPS and higher PSNR scores than other inversion meth-

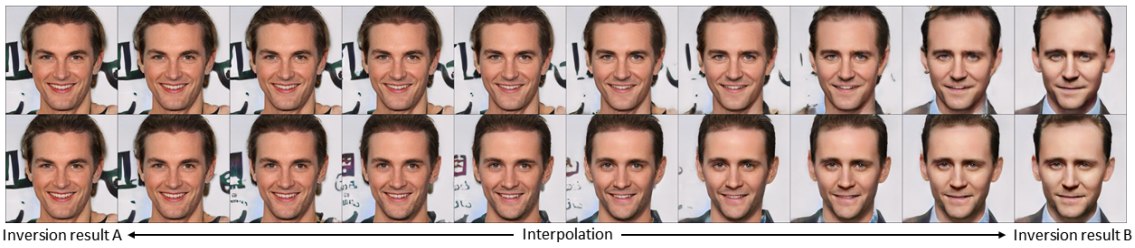

Figure 6: An example of interpolation path visualization: (*top row*) mGANprior and (*bottom row*) PmSFC. PmSFC produces a smooth transformation along the interpolation path, while mGANprior fails.

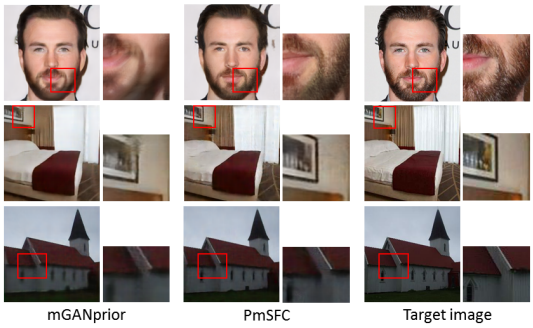

Figure 5: Visual comparison of mGANprior and our PmSFC in GAN inversion.

	CelebA		Bedroom		Church	
Method	LPIPS↓	PSNR↑	LPIPS↓	PSNR↑	LPIPS↓	PSNR↑
InvertGenNet	0.5797	19.17	0.5897	17.19	0.5339	17.15
GenVisMani	0.6992	11.18	0.6247	11.59	0.5961	11.58
InvertLayers	0.5321	20.33	0.5201	18.34	0.4789	17.81
mGANprior	0.4432	23.59	0.1578	25.13	0.1799	22.76
PmSFC	**0.4068**	**27.88**	**0.1300**	**27.59**	**0.1501**	**29.67**

Table 1: Comparison of the proposed PmSFC model and competing methods on the GAN inversion task.

ods. On CelebA, PmSFC surpasses mGANprior by about 4 percentage points in LPIPS. Figure 5 shows a number of the inverted images by mGANprior and our PmSFC. Note that mGANprior is implemented in our configuration according to the open-source code. The zoomed-in regions highlight that the capability of PmSFC to restore more details of the target images. The results suggest that inclusion of subspaces over GAN's intermediate feature channels enhances the inversion performance.

Model Analysis

Interpolation. We also conduct an interpolation experiment on CelebA to highlight the advantage of using multiple GAN subspaces. For the paired face images, our PmSFC model is used to infer the corresponding latent codes and then obtain the feature maps after feature composition. We apply linear interpolation to construct an interpolation path between the feature maps of paired images. Figure 6 shows a number of resulting images by decoding the interpolated feature maps

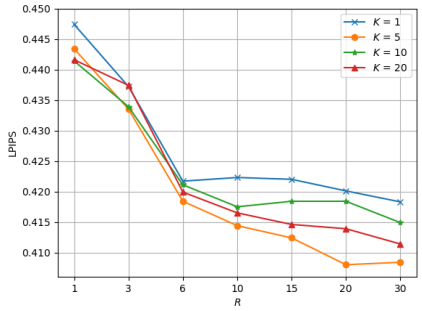

Figure 7: An experiment to investigate the impact of the hyper-parameters, the number of subspaces R and the number of latent codes K, on the inversion performance.

along the path. There is an abrupt change in the generated images of mGANprior, while our PmSFC model is able to produce continuously changing images. In addition, we can observe that the interpolated face images have reasonable structures and realistic appearance. We consider that the proposed approach is able to find an effective feature composition, such that the interpolation path stays more closely to the underlying data structure.

Impact of hyper-parameters. The proposed inversion approach mainly benefits from subspaces over GAN's intermediate feature channels. The number of subspaces R and the number of latent codes per subspace K control the representation capability of intermediate feature composition. We investigate the impact of different values of R and K on inversion quality on CelebA. The results shown in Figure 7 demonstrate that the inversion quality can be improved when the number of subspaces increases, but the relative improvement becomes smaller. In addition, we find that multiple latent codes can lead to better inversion quality than a single latent code in all the cases, and the best overall performance is reached when using 5 latent codes. We consider that our model with $R = 6$ and $K = 5$ can attain a balance between inversion quality and model complexity.

Extended Applications

As described in the previous section, the proposed GAN inversion approach is capable of performing various unsupervised image enhancement tasks. On each task, we specify the task-specific post-processing function $\mathcal{P}$, and our PmSFC model can be trained according to Eq.(9). The experiment shows

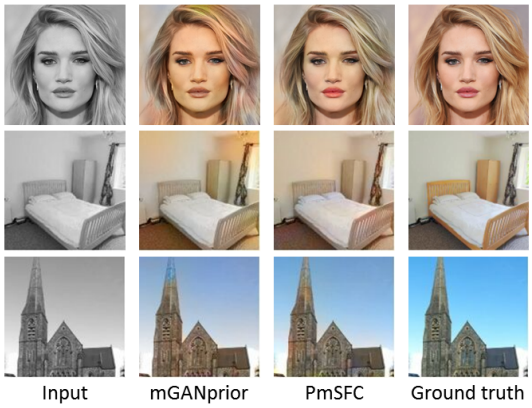

Figure 8: Visual comparison of mGANprior and our PmSFC in transforming images from grayscale to color.

Method	CelebA	Bedroom	Church
Grayscale Input	87.33*	88.02	85.50
SemPhoMani	-	85.41	86.10
DIP	-	84.33	83.31
ColHallucination	-	88.55	89.13
mGANprior	90.35*	90.02	89.43
PmSFC	**91.50**	**91.34**	**90.44**

Table 2: Comparison of the proposed PmSFC model and competing methods on the image colorization task in terms of AuC. $*$ indicates our implementation.

the applicability of our PmSFC model to different tasks.

Colorization. The proposed approach is used to transform a given image from grayscale to color. We follow the setting of (Gu, Shen, and Zhou 2020), and the experiments are performed on the PGGANs pre-trained on CelebA, LSUN-Bedroom and LSUN-Church, respectively. By inversion, we determine the latent codes and weighting vectors associated with each subspace. The reconstructed image is in color and is expected to be close to the ground-truth color image. The proposed approach is compared with mGANprior, SemPhoMani (Bau et al. 2019a), DIP (Ulyanov, Vedaldi, and Lempitsky 2018) and ColHallucination (Zhang, Isola, and Efros 2016). Table 2 shows the results of our approach and competing methods in terms of AuC (the area under the curve of cumulative error distribution) over the CIELAB color space excluding the lightness channel (Zhang, Isola, and Efros 2016). The proposed PmSFC model achieves the best performance in restoring the color information from the grayscale input on all the three datasets. Figure 8 shows some representative colorized images.

Super-resolution. Based on a PGGAN pre-trained on CelebA, the proposed approach can also be applied to recover a high resolution face image from a given low resolution one. The training procedure is similar to that of the colorization task, and the only difference is to use downsampling as the post-processing function. The original image is of size

Figure 9: Visual comparison of mGANprior and our PmSFC in transforming images from low-resolution to high-resolution.

Method	PSNR↑	LPIPS↓	NIQE↓
DIP	26.87	0.4236	4.66
RCAN	28.82	0.4579	5.70
ESRGAN	25.26	0.3862	3.27
mGANprior	26.93	0.3584	3.19
PmSFC	**28.90**	**0.3450**	**2.50**

Table 3: Comparison of the proposed PmSFC model and competing methods on the image super-resolution task.

$64 \times 64 \times 3$, and we aim to upscale the image to that with $16\times$ resolution. We compare our model with mGANprior, DIP, RCAN (Zhang et al. 2018b) and ESRGAN (Wang et al. 2018). In Table 3, the results of the competing methods are listed in terms of PSNR, LPIPS and NIQE. We also visualize the representative results in Figure 9. Our PmSFC model is able to produce better-resolved images with more details, when compared to mGANprior.

Conclusion

In this paper, we explore how to improve feature composition in an intermediate feature space for GAN inversion. Considering that the intermediate feature channels are highly correlated with each other, we incorporate a self-expressive layer into the generator to capture their self-expressiveness property, and the affinities among them are thus determined. Spectral clustering over the channels leads to a number of meaningful subspaces. We leverage the semantically differences among the subspaces by performing feature composition in each of them, and all the parameters are jointly optimized. Experimental results demonstrate that GAN prior subspaces benefit the inversion and extended applications. In our future work, we would investigate whether the proposed inversion approach preserves the GAN's latent space.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Project No. 62072189, U1611461), in part by the Research Grants Council of the Hong Kong Special Administration Region (Project No. CityU 11201220), and in part by the Natural Science Foundation of Guangdong Province (Project No. 2020A1515010484).

References

Abdal, R.; Qian, Y.; and Wonka, P. 2019. Image2StyleGAN: how to embed images into the StyleGAN latent space? In *Proc. IEEE International Conference on Computer Vision*.

Abdal, R.; Qian, Y.; and Wonka, P. 2020. Image2StyleGAN++: how to edit the embedded images? In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.

Arjovsky, M.; Chintala, S.; and Bottou, L. 2017. Wasserstein GAN. In *arXiv:1701.07875*.

Bau, D.; Strobelt, H.; Peebles, W.; Wulff, J.; Zhou, B.; Zhu, J.-Y.; and Torralba, A. 2019a. Semantic photo manipulation with a generative image prior. *ACM Transactions on Graphics* 38(4): 1–11.

Bau, D.; Zhu, J.-Y.; Strobelt, H.; Zhou, B.; Tenenbaum, J. B.; Freeman, W. T.; and Torralba, A. 2018. Gan dissection: Visualizing and understanding generative adversarial networks. *arXiv:1811.10597* .

Bau, D.; Zhu, J.-Y.; Walff, J.; Peebles, W.; Strobelt, H.; Zhou, B.; and Torralba, A. 2019b. Inverting layers of a large generator. In *Proc. ICLR Workshop*.

Bau, D.; Zhu, J.-Y.; Walff, J.; Peebles, W.; Strobelt, H.; Zhou, B.; and Torralba, A. 2019c. Seeing what a GAN cannot generate. In *Proc. IEEE International Conference on Computer Vision*.

Brock, A.; Donahue, J.; and Simonyan, K. 2019. Large scale GAN training for high fidelity natual image synthesis. In *Proc. International Conference on Learning Representation*.

Creswell, A.; and Bharath, A. A. 2019. Inverting the generator of a generative adversarial networks. *IEEE Transactions on Neural Networks and Learning Systems* 30(7): 1967–1974.

Donahue, J.; Krahenbuhl, P.; and Darrel, T. 2017. Adversarial feature learning. In *Proc. International Conference on Learning Representation*.

Dumoulin, V.; Belghazi, I.; Poole, B.; Mastropietro, O.; Lamb, A.; Arjovsky, M.; and Courville, A. 2017. Adversarially learned inference. In *Proc. International Conference on Learning Representation*.

Gong, M.; Xu, Y.; Li, C.; Zhang, K.; and Batmanghelich, K. 2019. Twin auxiliary classifiers GAN. In *Proc. Neural Information Processing Systmes*.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative adversarial nets. In *Advances in neural information processing systems*, 2672–2680.

Gu, J.; Shen, Y.; and Zhou, B. 2020. Image processing using multi-code GAN prior. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.

Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; and Courville, A. C. 2017. Improved training of Wasserstein GANs. In *Proc. Neural Information Processing Systmes*.

Karras, T.; Aila, T.; Laine, S.; and Lehtinen, J. 2018. Progressive growing of GANs for improved quality, stability, and variation. In *Proc. International Conference on Learning Representation*.

Karras, T.; Laine, S.; and Aila, T. 2019. A style-based generator architecture for generative adversarial networks. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.

Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; and Aila, T. 2020. Analyzing and improving the image quality of StyleGAN. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.

Kingma, D.; and Ba, J. 2015. Adam: a method for stochastic optimization. In *Proc. International Conference on Learning Representation*.

Li, B.; Qi, X.; Lukasiewicz, T.; and Torr, P. H. S. 2020. ManiGAN: Text-guided image manipulation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.

Lipton, Z. C.; and Tripathi, S. 2017. Precise recovery of latent vectors from generative adversarial networks. In *Proc. ICLR Workshop*.

Liu, D.; Long, C.; Zhang, H.; Yu, H.; Dong, X.; and Xiao, C. 2020. ARShadowGAN: shadow generative adversarial network for augmented reality in single light scenes. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.

Ma, F.; Ayaz, U.; and Karaman, S. 2018. Invertibility of convolutional generative networks from partial measurements. In *Proc. Neural Information Processing Systmes*.

Mirza, M.; and Osindero, S. 2014. Conditional generative adversarial nets. *arXiv:1411.1784* .

Mittal, A.; Soundararajan, R.; and Bovik, A. C. 2012. Making a completely blind image quality analyzer. *IEEE Signal Processing Letters* 20(3): 209–212.

Miyato, T.; and Koyama, M. 2018. cGANs with projection discriminator. In *Proc. International Conference on Learning Representation*.

Ng, A. Y.; Jordan, M. I.; and Weiss, Y. 2002. On spectral clustering: Analysis and an algorithm. In *Advances in neural information processing systems*, 849–856.

Nguyen, A.; Clune, J.; Bengio, Y.; Dosovitskiy, A.; and Yosinski, J. 2017. Plug & play generative networks: conditional iterative generation of images in latent space. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.

Pan, X.; Zhan, X.; Dai, B.; Lin, D.; Loy, C. C.; and Luo, P. 2020. Exploiting deep generative prior for versatile image restoration and manipulation. In *Proc. European Conference on Computer Vision*.

Perarnau, G.; van de Weijer, J.; Raducanu, B.; and Alvarez, J. M. 2016. Invertible conditional GANs for image editing. In *Proc. NeurIPS Workshop*.

Shen, Y.; Gu, J.; Tang, X.; and Zhou, B. 2020. Interpreting the latent space of gans for semantic face editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9243–9252.

Ulyanov, D.; Vedaldi, A.; and Lempitsky, V. 2018. Deep image prior. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.

Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; and Loy, C. C. 2018. ESRGAN: enhanced super-resolution generative adversarial networks. In *Proc. European Conference on Computer Vision Workshop*.

Wang, Y.; Gonzalez-Garcia, A.; Berga, D.; Herranz, L.; Khan, F. S.; and van de Weijer, J. 2020. MineGAN: effective knowledge transfer from GANs to target domains with few images. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 13(4): 600–612.

Wei, X.; Gong, B.; Liu, Z.; Lu, W.; and Wang, L. 2018. Improving the improved training of Wasserstein GANs: a consistency term and its dual effect. In *Proc. International Conference on Learning Representation*.

Wu, S.; Lin, S.; Wu, W.; Azzam, M.; and Wong, H.-S. 2019. Semi-supervised pedestrian instance synthesis and detection with mutual reinforcement. In *Proc. IEEE International Conference on Computer Vision*.

Yu, F.; Seff, A.; Zhang, Y.; Song, S.; Funkhouser, T.; and Xiao, J. 2015. LSUN: construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv:1506.03365* .

Zhang, R.; Isola, P.; and Efros, A. A. 2016. Colorful image colorization. In *Proc. European Conference on Computer Vision*.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018a. The unreasonable effectiveness of deep features as a perceptual metric. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.

Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; and Fu, Y. 2018b. Image super-resolution using very deep residual channel attention networks. In *Proc. European Conference on Computer Vision*.

Zhang, Y.; Tsang, I. W.; Luo, Y.; Hu, C.; Lu, X.; and Yu, X. 2020. Copy and paste GAN: face hallucination from shaded thumbnails. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.

Zhu, J.; Shen, Y.; Zhao, D.; and Zhou, B. 2020. In-domain GAN inversion for real image editing. In *Proc. European Conference on Computer Vision*.

Zhu, J.-Y.; Krahenbuhl, P.; Shechtman, E.; and Efros, A. A. 2016. Generative visual manipulation on the natural image manifold. In *Proc. European Conference on Computer Vision*.