High-Resolution GAN Inversion for Degraded Images in Large Diverse Datasets

Yanbo Wang¹*, Chuming Lin², Donghao Luo², Ying Tai², Zhizhong Zhang¹†, Yuan Xie¹

¹School of Computer Science and Technology, East China Normal University
²Tencent Youtu Lab
51205901021@stu.ecnu.edu.cn, {chuminglin, michaelluo, yingtai}@tencent.com, {zizzhang, yxie}@cs.ecnu.edu.cn

Abstract

The last decades are marked by massive and diverse image data, which shows increasingly high resolution and quality. However, some images we obtained may be corrupted, affecting the perception and the application of downstream tasks. A generic method for generating a high-quality image from the degraded one is in demand. In this paper, we present a novel GAN inversion framework that utilizes the powerful generative ability of StyleGAN-XL for this problem. To ease the inversion challenge with StyleGAN-XL, Clustering & Regularize Inversion (CRI) is proposed. Specifically, the latent space is firstly divided into finer-grained sub-spaces by clustering. Instead of initializing the inversion with the average latent vector, we approximate a centroid latent vector from the clusters, which generates an image close to the input image. Then, an offset with a regularization term is introduced to keep the inverted latent vector within a certain range. We validate our CRI scheme on multiple restoration tasks (i.e., inpainting, colorization, and super-resolution) of complex natural images, and show preferable quantitative and qualitative results. We further demonstrate our technique is robust in terms of data and different GAN models. To our best knowledge, we are the first to adopt StyleGAN-XL for generating high-quality natural images from diverse degraded inputs. Code is available at https://github.com/Booooooooooo/CRI.

Introduction

With the mushroom development of imaging devices, images are now characterized by increasingly high resolution and diverse scenes. However, the quality of images on the Internet is uneven. For example, we may get images with low-resolution, images with missing parts, or gray-scale images. Therefore we hope to generate high-quality images directly from the degraded raw natural images.

One feasible approach is to leverage the generative ability of generative adversarial networks (GAN) and conduct GAN inversion. The main idea is to “invert” the corrupted image back to the latent space of the pre-trained GAN and then reconstruct a restored image from the optimal latent (Xia et al. 2022). For a given degraded function $D(\cdot)$ and degraded image $I_d$, the latent vector $w_{re}$ is iterated through back-propagation by minimizing the reconstruction loss of $I_d$ and the degraded synthesis image $D(\phi_{Synthesis}(w_{re}))$, where $\phi_{Synthesis}$ is the synthesis layers of GAN. For example, PTI (Roich et al. 2021) inverts high-quality face images to the latent space and uses the result for image attributes manipulation. Nevertheless, PTI cannot be generalized to degraded input images from natural scenes. On the other hand, DGP (Pan et al. 2021) provides compelling results to generate missing semantics, e.g., color, patch, resolution, for various images. But the scheme is limited to vanilla GAN models (i.e., BigGAN (Brock, Donahue, and Simonyan 2019)) which only generate results with relatively low resolution (e.g., $128^2$, $256^2$).

In this paper, we present an effective framework to generate high-resolution natural images from degraded inputs. We utilize the powerful capability of StyleGAN-XL (Sauer, Schwarz, and Geiger 2022), the state-of-the-art model for large-scale image synthesis, to generate high-quality images from degraded images. However, simply applying existing inversion methods (Pan et al. 2021; Roich et al. 2021; Menon et al. 2020) on StyleGAN-XL pre-trained on ImageNet (Krizhevsky, Sutskever, and Hinton 2012) obtains unsatisfactory results and faces two challenges:

1. The latent space of StyleGAN-XL pre-trained on ImageNet is much larger and more complex compared to other popular GAN models like BigGAN and StyleGAN (Karras, Laine, and Aila 2019). Due to the diversity of data, the scale of the generator increases. For instance, for face images (i.e., FFHQ (Karras, Laine, and Aila 2019)) StyleGAN2 (Karras et al. 2020) at resolution $1024^2$ only uses 18 latent vectors, while StyleGAN-XL trained on ImageNet at resolution $512^2$ needs 37 latent vectors since the data domain is more complicated. In the meantime, StyleGAN-XL employs an intermediate latent space compared to BigGAN which is the essence of its flexible editability but also adds to the inversion difficulty. For such a complex latent space, finding a good initialization of latent vector is quite important.

2. The visual quality of the result is not explicitly imposed by the existing inversion methods. For example, images $I_{re}$ and $I_t$ generated by $w_{re}$ and $w_t$ in Figure 1(a) result in the same degraded images ($D(I_{re})$ and $D(I_t)$). Minimizing the traditional pixel or feature loss as existing meth-
We fully explore the potential of GAN inversion for degraded natural images in diverse scenes. To our best knowledge, we are the first to adopt StyleGAN-XL to generate high-quality images from degraded inputs.

• We propose a simple yet effective GAN inversion method to address the challenge of inversion with StyleGAN-XL. The key idea is to find a better starting latent vector and introduce a constraint on the image quality for optimization, summarized as Clustering & Regularize.

• We conduct extensive experiments on restoration tasks with different datasets and GAN models. The experimental results demonstrate the superiority of our method against other inversion methods.

Related Works

Style-based Generator

The Style-based generator is first proposed by (Karras, Laine, and Aila 2019). A mapping module is introduced to map the random values (denoted by \( z \in \mathcal{Z} \)) to an intermediate latent space named \( \mathcal{W} \) space. By feeding latent vectors \( w \in \mathcal{W} \) to each synthesis layer, StyleGAN can control the attributes of the generated image. It quickly surpassed previous image generative models (Brock, Donahue, and Simonyan 2019; Karras et al. 2017), showing superior perceptual quality and variety. StyleGAN2 (Karras et al. 2020) introduced path length regularization and weight demodulation, which further improves the image quality. Recently, StyleGAN3 (Karras et al. 2021) addressed the aliasing problem and proposed a new architecture that boosts the generator to be fully equivariant to translation and rotation.

GAN Inversion

GAN inversion aims at mapping a given image back into the latent space of a pre-trained GAN model, which can be...
viewed as the inverse problem of image synthesis. It was first introduced by (Zhu et al. 2016). Generally, inversion methods can be divided into optimization-based (Creswell and Bharath 2018; Abdal, Qin, and Wonka 2020; Roich et al. 2021) and learning-based (Richardson et al. 2021; Dinh et al. 2022; Tov et al. 2021; Dong et al. 2021). The optimization-based approaches directly update the latent vector by minimizing the reconstruction loss. Learning-based methods employ an encoder to learn the mapping from the given image to its corresponding latent. Recently, (Dong et al. 2021) studies the invertibility of an arbitrary pre-trained DNN using learning-based inversion. However, despite the gain in inference speed, the reconstruction quality of learning-based methods is often worse than that of optimization-based approaches. Also, the encoder cannot be adapted to out-of-domain data. Our proposed CRI is based on optimization as we find that the relatively easy optimization-based approaches still do not work well for complex scenes.

Due to the many desirable properties of StyleGANs, recent works focus on conducting inversion with StyleGANs. For StyleGANs, inversion is usually conducted in the $W$ space. It has been shown that the extended $W+$ where different latent vectors are fed into each of the generator’s layers is much more expressive and enables better image preservation (Abdal, Qin, and Wonka 2019). As shown in (Tov et al. 2021), there exist trade-offs between distortion, perceptual quality, and editability within the $W+$ space. Recently, PTI (Roich et al. 2021) proposed pivotal tuning to mitigate the distortion-editability trade-off for out-of-distribution images. In comparison, we consider GAN inversion as a tool for exploiting the GAN priors for restoration tasks. Finding a “sweet spot” in the distortion-perception trade-off rather than the distortion-editability trade-off is the primary purpose and challenge of our work.

Deep Priors in Image Restoration

Image restoration is the task of recovering a clean image given its degraded version. Due to the powerful ability to generate photo-realistic images, many works exploit pre-trained GANs as priors to improve the quality (Geonung Kim 2022; Wei et al. 2022; Bartz et al. 2020). Leveraging a pre-trained StyleGAN is very common in face restoration (Menon et al. 2020; Yang et al. 2021). As one representative work, PULSE (Menon et al. 2020) propose to solve face image super-resolution task with inversion on StyleGAN. As for more complex real-world images, attempts are still limited to vanilla GANs (e.g., BigGAN). DGP (Pan et al. 2021) presents a relaxed formulation for mining the priors in BigGAN. The generator is jointly finetuned with the latent vectors. (Wu et al. 2021) ’retrieves’ color information encapsulated in BigGAN via an encoder and then incorporates these features with feature modulations. In this paper, we explore the potential of StyleGAN-XL for degraded images from natural complex scenes, achieving better results than existing GAN inversion methods.

Method

Suppose the given input image $I_d$ is obtained via $I_d = D(I)$ where $I$ is the original high-quality image, $D(\cdot)$ is the degradation function. Our aim is to generate high-quality images. By utilizing StyleGAN-XL, the problem can be translated as finding a latent vector $w$ that satisfies:

$$w = \arg \min_w L(D(\phi_{\text{Synthesis}}(w)), I_d),$$

where $\phi_{\text{Synthesis}}(\cdot)$ is the synthesis network of the pre-trained generator, $L(\cdot)$ is a distance metric in the image or feature space.

Although various attempt has been made (Menon et al. 2020; Pan et al. 2021) to seek generative priors, generating high-resolution natural images remains challenging. Since the latent space is much more complex for StyleGAN-XL pre-trained on ImageNet, it is essential to provide additional help for the optimization process. We proposed a Clustering & Regularize strategy to break through this problem. Clustering offers a better initial starting point for the optimization, while Regularize constrains the scope of $w$ during optimization. In this section, we will first give details about this strategy and then describe the overall pipeline.

Start from a Centroid

One main challenge for restoring natural images with StyleGAN-XL lies in the complexity of the latent space. The latent vectors for StyleGAN-ImageNet present a multi-modal characteristic. Simply starting from an average latent vector becomes sub-optimal. Thus, we begin the optimization by finding a proper centroid as shown in Figure 1(b).

For a given class, we first randomly sample $M$ latent vectors $\{w_j\}_{j=1}^M$ using the pre-trained mapping module $\phi_{\text{Mapping}}$ of StyleGAN-XL:

$$\{w_j\}_{j=1}^M = \phi_{\text{Mapping}}(\{z_j\}_{j=1}^M, c),$$

where $\{z_j\}_{j=1}^M \sim \mathcal{N}(0, 1)$ and $c$ is a one-hot vector of the given class. $\{w_j\}_{j=1}^M$ are clustered into $N$ clusters, obtaining $N$ centroids $\{cen_i\}_{i=1}^N$ corresponding to $N$ center images:

$$\{cen_i\}_{i=1}^N = \phi_{\text{Synthesis}}(\{w_k\}_{k=1}^N_{cen}).$$

Given degraded input image $I_d$, we measure the distance between $I_d$ and $\{cen_i\}_{i=1}^N$ in the feature space of a pre-trained model (e.g. VGG (Simonyan and Zisserman 2014)). The corresponding $w_k$ of the “nearest” center image is selected as the starting point for the subsequent optimization.

Regularized Offset

Due to the ill-posedness of image restoration, we may find multiple latent vectors that satisfy Eq. 1. However, some of the generated images may be beyond the manifold of natural images. Applying the traditional loss on $I_d$ and $D(\phi_{\text{Synthesis}}(w))$ does not explicitly guarantee the perception for inversion. We propose to improve the perceptual quality of the result images by constraining the scope of the latent vectors during optimization. Instead of optimizing $w$ directly, we introduce an offset term $w_{\text{off}}$ and keep the centroid $w_{cen}^k$ frozen. The output image is defined as:

$$I_{\text{syn}} = \phi_{\text{Synthesis}}(w_{cen}^k + w_{\text{off}}).$$
Then we add a regularization term for $w^o_{off}$

$$reg = ||w^o_{off}||_2.$$  \tag{5}

This regularization term limits the range of the output latent $w = w^c_{cen} + w^o_{off}$ to a certain extent so that the output image does not exceed the natural manifold. By doing so, the latent is allowed to be gradually rectified towards the target while remaining high perceptual quality.

**Loss Function**

The overall pipeline of CRI can be divided into two stages: the optimization stage and the finetune stage. As noted in DGP (Pan et al. 2021), a fixed generator is perhaps a crucial limitation for faithfully reconstructing unseen and complex images. However, we found that finetuning the generator together with the latent vector at the same time like DGP is sub-optimal for more sophisticated GANs. For StyleGANGs with disentangled latent vectors, dividing optimization and finetune into two stages helps reduce the inversion difficulty.

In the first stage, only the $w^o_{off}$ is optimized and the weights of the generator are frozen. The overall loss for the optimization stage can be defined as:

$$L_{op} = L_{LPIPS}(I_d, D(I_{syn})) + \lambda_1 L_2(I_d, D(I_{syn})) + \lambda_2 ||w^o_{off}||_2,$$  \tag{6}

where $D(\cdot)$ is the degradation function, $\lambda_1$ and $\lambda_2$ are hyper-parameters. As the noise input is removed in StyleGAN3, we do not optimize this term. However, the noise vector can be considered to improve the visual details.

We use the pivotal tuning technique (Roich et al. 2021) in the second stage, where the pivotal latent $w = w^c_{cen} + w^o_{off}$ is frozen. The generator is finetuned with the following loss:

$$L_{ft} = L_{LPIPS}(I_d, D(I_{syn})) + \lambda_{L2}L_2(I_d, D(I_{syn})) + \lambda_{R}L_R,$$  \tag{7}

where $\lambda_{L2}$ and $\lambda_{R}$ are hyperparameters, $L_R$ is the locality regularization term defined as:

$$L_R = L_{LPIPS}(x_r, x^*_r) + \lambda^R_{L2}L_2(x_r, x^*_r),$$  \tag{8}

where $\lambda^R_{L2}$ is a hyperparameter, $x_r = \phi_{synthesis}(w_r; \theta)$ is generated with the original weights of the generator and $x^*_r = \phi_{synthesis}(w_r; \theta^*)$ is generated using the currently tuned ones, $w_r$ is the interpolated code between a random latent vector and the pivotal latent vector. Pseudocode of CRI is summarized in Algorithm 1.

**Experiments**

In this section, we present comprehensive experiments to evaluate our method. We experiment with image inpainting, image colorization and super-resolution. We start by comparing with current inversion methods on StyleGAN-XL (Sauer, Schwarz, and Geiger 2022) pre-trained on ImageNet (Krizhevsky, Sutskever, and Hinton 2012). We also compare the generating results of our method on StyleGAN-XL and DGP on BigGAN. For face images, we conduct experiments with StyleGAN2 (Karras et al. 2020) pre-trained on FFHQ (Karras, Laine, and Aila 2019), which shows the generalizability of our method. Then, we extend our method to out-of-domain images. Finally, we conduct an ablation study to justify the effectiveness of our inversion method. All the methods are inverted to $W^+$ space except PTI-colorization as we found $W$ space generates better results.

**Restoration on ImageNet**

We first compare our method with other GAN inversion methods (Menon et al. 2020; Pan et al. 2021; Roich et al. 2021) on ImageNet. We use StyleGAN-XL (Sauer, Schwarz, and Geiger 2022) as it is state-of-the-art on large-scale image synthesis. The resolution of the generated images is $512^2$. We invert 1000 images from the validation set of ImageNet, each from different classes, to quantitatively evaluate the methods. The scale factor of the SR task is set to 4.

**Qualitative Results**

Figure 2 presents a qualitative comparison against other GAN-inversion methods (Menon et al. 2020; Pan et al. 2021; Roich et al. 2021) on ImageNet dataset. Our method achieves superior results for all tasks, i.e. image inpainting, image colorization and image super-resolution. Due to the lack of finetuning stage, PULSE fails to generalize to complex real-world images. It can only produce the general shape, but could not fill in the missing information. At the same time, since the latent vector and the generator are simultaneously optimized and finetuned, it is
difficult for DGP to find a faithful output. Finetuning with incorrect latent vectors disrupts the generative capacity of the pre-trained generator which result in poor visual quality. Compared to PTI, our method is capable of recovering realistic fine details (e.g. smoother contours for inpainting, truer colors for colorization and sharper details for SR).

**Quantitative Results** Following (Wu et al. 2021; Zeng et al. 2021), we use LPIPS (Zhang et al. 2018), FID (Heusel et al. 2017) for image inpainting and image colorization. Following (Pan et al. 2021), LPIPS and NIQE (Mittal, Soundararajan, and Bovik 2012) are used for SR. Also, we find these metrics are close to visual perception. As shown in Table 1, the quantitative results align with our qualitative results as we achieve the best score for all metrics.

**Comparison with DGP-BigGAN** We compare the results of DGP with BigGAN-ImageNet and the results of our methods. Quantitative and Qualitative results are shown in Table 2 and Figure 3. Since BigGAN only supports the resolution of $256^2$, we resize the output of DGP+BigGAN to $512^2$ for comparison. For SR, the input of both DGP and CRI are set to $128^2$. Our CRI+StyleGAN-XL outperforms DGP+BigGAN on image inpainting, colorization and SR.

**Restoration on Face** To illustrate our method is not limited by the GAN model or data, we conduct experiments on StyleGAN2 (Karras et al. 2020) pre-trained on FFHQ (Karras, Laine, and Aila 2019). The resolution of the generated images is $1024^2$. We invert 500 images from CelebA-HQ (Karras et al. 2017). The scale factor of SR is set to 16.

---

![Figure 2: Image Restoration with images from the ImageNet valid set. StyleGAN-XL pre-trained on ImageNet is used. Our method achieves more natural transitions, more realistic colors and sharper details. Zoom in for better visualization.](image)

<table>
<thead>
<tr>
<th>Task</th>
<th>Measure</th>
<th>DGP</th>
<th>PTI</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inpainting</td>
<td>LPIPS ↓</td>
<td>0.2612</td>
<td>0.1219</td>
<td><strong>0.1164</strong></td>
</tr>
<tr>
<td></td>
<td>FID ↓</td>
<td>166.24</td>
<td>54.62</td>
<td><strong>44.73</strong></td>
</tr>
<tr>
<td>Colorization</td>
<td>LPIPS ↓</td>
<td>0.3105</td>
<td>0.1994</td>
<td><strong>0.1914</strong></td>
</tr>
<tr>
<td></td>
<td>FID ↓</td>
<td>78.32</td>
<td>47.51</td>
<td><strong>45.90</strong></td>
</tr>
<tr>
<td>SR</td>
<td>LPIPS ↓</td>
<td>0.2726</td>
<td>0.2429</td>
<td><strong>0.2390</strong></td>
</tr>
<tr>
<td></td>
<td>NIQE ↓</td>
<td>8.4044</td>
<td>5.5659</td>
<td><strong>5.3065</strong></td>
</tr>
</tbody>
</table>

Table 1: Quantitative comparison against DGP and PTI on image inpainting, image colorization and image super-resolution. We use StyleGAN-XL pre-trained on ImageNet, and invert images from the ImageNet valid set.

<table>
<thead>
<tr>
<th>Task</th>
<th>Measure</th>
<th>DGP</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inpainting</td>
<td>LPIPS ↓</td>
<td>0.1316</td>
<td>0.1211</td>
</tr>
<tr>
<td></td>
<td>FID ↓</td>
<td>116.12</td>
<td>80.20</td>
</tr>
<tr>
<td>Colorization</td>
<td>LPIPS ↓</td>
<td>0.2256</td>
<td>0.1842</td>
</tr>
<tr>
<td></td>
<td>FID ↓</td>
<td>189.42</td>
<td>92.48</td>
</tr>
<tr>
<td>SR</td>
<td>LPIPS ↓</td>
<td>0.2226</td>
<td>0.2224</td>
</tr>
<tr>
<td></td>
<td>NIQE ↓</td>
<td>9.2102</td>
<td>5.5686</td>
</tr>
</tbody>
</table>

Table 2: Quantitative comparison between DGP with BigGAN pre-trained on ImageNet and CRI with StyleGAN-XL pre-trained on ImageNet.
Figure 3: Qualitative comparison between DGP with BigGAN pre-trained on ImageNet and CRI with StyleGAN-XL pre-trained on ImageNet. Zoom in for better visualization.

(a) BSD100  (b) DIV2K

Figure 4: Evaluation of CRI on out-of-domain datasets, including (a) image from BSD100, (b) image from DIV2K.

Quantitative Results For face restoration, we employ LPIPS and FID as evaluation metrics shown in Table 3. The FID score is calculated between the inverted 500 images and the CelebA-HQ dataset (30000 images in total). Our method obtains the lowest LPIPS and FID for all tasks, indicating that our results are perceptually close to the ground truth and have a close distance to the real face distribution.

Qualitative Results Qualitative results of each task are shown in Figure 6. Our method outperforms other methods in terms of image quality and identity preservation. PULSE and DGP can produce results of good visual quality, but the similarity is relatively low. For instance, DGP-SR fails to preserve the hairstyle. Images generated by PTI are similar to the input in terms of known information. For example, the background of the image in the inpainting task and the structure in the colorization task are quite authentic. However, the overall view of the generated images is not natural since the inverted latent vectors beyond the manifold of natural images. Compared to PTI, our results are more natural and generate sharper details.

Quantitative Results Table 3: Quantitative comparison against PULSE, DGP and PTI on image inpainting, image colorization and image super-resolution. We use StyleGANv2 pre-trained on FFHQ, and invert images from the CelebA-HQ.

Robustness on Out-of-domain Images

We also experiments with images from BSD100 (Martin et al. 2001) dataset and DIV2K (Agustsson and Timofte 2017) dataset. Deit-M (Touvron et al. 2021) is used as a classifier to provide the class information \( i.e., c \) in Eq. 2) for the input. As shown in Figure 4, our CRI can reconstruct the missing part of out-of-domain images.

Ablation Study

Effect of Centroids We show the effectiveness of the clustering centroids by adopting different cluster numbers. To exclude the effect of the random sampling of \( \{z_j\}_{j=1}^M \), we cluster the same \( \{z_j\}_{j=1}^M \) into \( N = \{1, 5, 10, 15\} \) clusters respectively using KMeans. The results of colorization are presented in Figure 5. The performance becomes progressively better as the number of centroids increases. However, the clustering time also becomes longer. Considering the performance and the clustering time trade-off, we choose \( N = 10 \) as the default.

Effect of Regularized Offset To validate the effectiveness of our regularized offset, we conduct an ablation study on face colorization with StyleGAN2. 60 images from CelebA-HQ are used. We first remove the regularization term and then replace the \( L_2 \) regularization with \( L_1 \). As shown in Table 4, using \( L_2 \) regularized offset is beneficial for both LPIPS and the ID Similarity compared to \( L_1 \) regularization. Optimizing without a regularization term (w/o Reg) fails to generate authentic colors (see Figure 7) but achieves high ID similarity, which also demonstrates the inherent trade-off between distortion and perception. In the meanwhile, our

Table 3: Quantitative comparison against PULSE, DGP and PTI on image inpainting, image colorization and image super-resolution. We use StyleGANv2 pre-trained on FFHQ, and invert images from the CelebA-HQ.

Table 4: Ablation study on face colorization with StyleGAN2. 60 images from CelebA-HQ are used. We first remove the regularization term and then replace the \( L_2 \) regularization with \( L_1 \). As shown in Table 4, using \( L_2 \) regularized offset is beneficial for both LPIPS and the ID Similarity compared to \( L_1 \) regularization. Optimizing without a regularization term (w/o Reg) fails to generate authentic colors (see Figure 7) but achieves high ID similarity, which also demonstrates the inherent trade-off between distortion and perception. In the meanwhile, our

Figure 5: Comparison of numbers of the cluster.
proposed CRI can achieve a “sweet spot” in the distortion-perception trade-off by introducing the regularized offset.

**Implementation Details**

For StyleGAN-XL with ImageNet, we use 1000 iterations for the optimization stage and use 350 iterations for the fine-tune stage as in (Sauer, Schwarz, and Geiger 2022). For StyleGAN2 with FFHQ, we use 500 iterations for the first stage and 20 iterations for the second stage as inversion with a face is simpler. For more detailed implementations, please refer to the supplementary materials.

**Conclusion**

We have presented a general framework for generating high-resolution images from degraded images in diverse natural scenes. This is achieved by leveraging the generative power of StyleGAN-XL. Furthermore, we propose a simple yet effective technique, term as Clustering & Regularize, to ease the inversion difficulty of StyleGAN-XL pre-trained on massive natural images. Clustering is introduced to solve the difficulty caused by the large and complex latent space of StyleGAN-XL. It divides the latent space into sub-spaces and provides the inversion with a better initialization. While Regularize is derived from the perceptual aspect. By introducing an offset term and constraining it with regularization we can bind inversion for better visual quality. CRI allows us to achieve good perception and can effectively provide the results with rich image semantics. Extensive experiments on image restoration tasks illustrate the effectiveness of CRI. By designing the degradation function \( D(\cdot) \), we believe our framework can be applied to other degradations, e.g., noise and blur, which will be left for future work.
Acknowledgments

This work was supported by the National Key Research and Development Program of China (2021ZD0111000); National Natural Science Foundation of China No. 62222602, 62176092, 62106075, Shanghai Science and Technology Commission No.2151100700, Natural Science Foundation of Shanghai (202R1417700).

References


Brock, A.; Donahue, J.; and Simonyan, K. 2019. Large scale GAN training for high fidelity natural image synthesis. ICLR.


