Image-Adaptive GAN Based Reconstruction

Shady Abu Hussein,* Tom Tirer,* Raja Giryes

School of Electrical Engineering
Tel Aviv University, Tel Aviv, Israel

Abstract

In the recent years, there has been a significant improvement in the quality of samples produced by (deep) generative models such as variational auto-encoders and generative adversarial networks. However, the representation capabilities of these methods still do not capture the full distribution for complex classes of images, such as human faces. This deficiency has been clearly observed in previous works that use pre-trained generative models to solve imaging inverse problems. In this paper, we suggest to mitigate the limited representation capabilities of generators by making them image-adaptive and enforcing compliance of the restoration with the observations via back-projections. We empirically demonstrate the advantages of our proposed approach for image super-resolution and compressed sensing.

Introduction

The developments in deep learning (Goodfellow, Bengio, and Courville 2016) in the recent years have led to significant improvement in learning generative models. Methods like variational auto-encoders (VAEs) (Kingma and Welling 2013), generative adversarial networks (GANs) (Goodfellow et al. 2014) and latent space optimizations (GLOs) (Bojanowski et al. 2018) have found success at modeling data distributions. However, for complex classes of images, such as human faces, while these methods can generate nice examples, their representation capabilities do not capture the full distribution. This phenomenon is sometimes referred to in the literature, especially in the context of GANs, as *mode collapse* (Arjovsky, Chintala, and Bottou 2017; Karras et al. 2017). Yet, as demonstrated in (Richardson and Weiss 2018), it is common to other recent learning approaches as well.

Another line of works that has gained a lot from the developments in deep learning is imaging inverse problems, where the goal is to recover an image $\mathbf{x}$ from its degraded or compressed observations $\mathbf{y}$ (Bertero and Boccacci 1998). Most of these works have been focused on training a convolutional neural network (CNN) to learn the inverse mapping from $\mathbf{y}$ to $\mathbf{x}$ for a *specific* observation model (e.g. super-resolution with certain scale factor and bicubic anti-aliasing kernel (Dong et al. 2014)). Yet, recent works have suggested to use neural networks for handling only the image prior in a way that does not require exhaustive offline training for each different observation model. This can be done by using CNN denoisers (Zhang et al. 2017; Meinhardt et al. 2017; Rick Chang et al. 2017) plugged into iterative optimization schemes (Venkatakrishnan, Bouman, and Wohlberg 2013; Metzler, Maleki, and Baraniuk 2016; Tirer and Giryes 2018), training a neural network from scratch for the imaging task directly on the test image (based on internal recurrence of information inside a single image) (Shocher, Cohen, and Irani 2018; Ulyanov, Vedaldi, and Lempitsky 2018), or using generative models (Bora et al. 2017; Yeh et al. 2017; Hand, Leong, and Voroninski 2018).

Methods that use generative models as priors can only handle images that belong to the class or classes on which the model was trained. However, the generative learning equips them with valuable semantic information that other strategies lack. For example, a method which is not based on a generative model cannot produce a perceptually pleasing image of human face if the eyes are completely missing in an inpainting task (Yeh et al. 2017). The main drawback in restoring complex images using generative models is the limited representation capabilities of the generators. Even when one searches over the range of a pre-trained generator for an image which is closest to the original $\mathbf{x}$, he is expected to get a significant mismatch (Bora et al. 2017).

In this work, we propose a strategy to mitigate the limited representation capabilities of generators when solving inverse problems. The strategy is based on a gentle internal learning phase at test time, which essentially makes the generator image-adaptive while maintaining the useful information obtained in the offline training. In addition, in scenarios with low noise level, we propose to further improve the reconstruction by a back-projection step that strictly enforces compliance of the restoration with the observations $\mathbf{y}$. We empirically demonstrate the advantages of our proposed approach for image super-resolution and compressed sensing.

[*] The authors have contributed equally to this work.

Related Work

Our work is mostly related to the work by Bora et al. (2017), which have suggested to use pre-trained generative models for the compressive sensing (CS) task (Donoho 2006; Candes, Romberg, and Tao 2006): reconstructing an unknown signal $\mathbf{x} \in \mathbb{R}^n$ from observations $\mathbf{y} \in \mathbb{R}^m$ of the form

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}, \tag{1}$$

where $\mathbf{A}$ is an $m \times n$ measurement matrix, $\mathbf{e} \in \mathbb{R}^m$ represents the noise, and the number of measurements is much smaller than the ambient dimension of the signal, i.e. $m \ll n$. Following the fact that in highly popular generative models (e.g. GANs, VAEs and GLOs) a generator $\mathbf{G}(\cdot)$ learns a mapping from a low dimensional space $\mathbf{z} \in \mathbb{R}^k$ to the signal space $\mathbf{G}(\mathbf{z}) \subset \mathbb{R}^n$, the authors (Bora et al. 2017) have proposed a method, termed CSGM, that estimates the signal as $\hat{\mathbf{x}} = \mathbf{G}(\hat{\mathbf{z}})$, where $\hat{\mathbf{z}}$ is obtained by minimizing the non-convex[1] cost function

$$f(\mathbf{z}) = \|\mathbf{y} - \mathbf{A}\mathbf{G}(\mathbf{z})\|_2^2, \tag{2}$$

using backpropagation and standard gradient based optimizers.

For specific classes of images, such as handwritten digits and human faces, the experiments in (Bora et al. 2017; Hand, Leong, and Voroninski 2018) have shown that using learned generative models enables to reconstruct nice looking images with much fewer measurements than methods that use non-generative (e.g. model-based) priors. However, unlike the latter, it has been also shown that CSGM and its variants cannot provide accurate recovery even when there is no noise and the number of observations is very large. This shortcoming is mainly due to the limited representation capabilities of the generative models (see Section 6.3 in (Bora et al. 2017)), and is common to related recent works (Hand, Leong, and Voroninski 2018; Bora, Price, and Dimakis 2018; Dhar, Grover, and Ermon 2018; Shah and Hegde 2018).

Note that using specific structures of $\mathbf{A}$, the model (1) can be used for different imaging inverse problems, making the CSGM method applicable for these problems as well. For example, it can be used for denoising task when $\mathbf{A}$ is the $n \times n$ identity matrix $\mathbf{I}_n$, inpainting task when $\mathbf{A}$ is an $m \times n$ sampling matrix (i.e. a selection of m rows of $\mathbf{I}_n$), deblurring task when $\mathbf{A}$ is a blurring operator, and super-resolution task if $\mathbf{A}$ is a composite operator of blurring (i.e. anti-aliasing filtering) and down-sampling.

Our image-adaptive approach is inspired by (Tirer and Giryes 2019), which is influenced itself by (Shocher, Cohen, and Irani 2018; Ulyanov, Vedaldi, and Lempitsky 2018). These works follow the idea of internal recurrence of information inside a single image within and across scales (Glasner, Bagon, and Irani 2009). However, while the two methods (Shocher, Cohen, and Irani 2018; Ulyanov, Vedaldi, and Lempitsky 2018) completely avoid an offline training phase and optimize the weights of a deep neural network only in the test phase, the work in (Tirer and Giryes 2019) incorporates external and internal learning by taking offline trained CNN denoisers, fine-tuning them in test time and then plugging them into a model-based optimization scheme (Tirer and Giryes 2018). Note, though, that the internal learning phase in (Tirer and Giryes 2019) uses patches from $\mathbf{y}$ as the ground truth for a denoising loss function ($f(\tilde{\mathbf{x}}) = \|\mathbf{y} - \tilde{\mathbf{x}}\|_2^2$), building on the assumption that $\mathbf{y}$ directly includes patterns which recur also in $\mathbf{x}$. Therefore, this method requires that $\mathbf{y}$ is not very degraded, which makes it suitable perhaps only for the super-resolution task, similarly to (Shocher, Cohen, and Irani 2018), which is also restricted to this problem.

Note that the method in (Ulyanov, Vedaldi, and Lempitsky 2018), termed as deep image prior (DIP), can be applied to different observation models. However, the advantage of our approach stems from the offline generative learning that captures valuable semantic information that DIP lacks. As mentioned above, a method like DIP, which is not based on a generative model, cannot produce a perceptually pleasing image of human face if the eyes are completely missing in an inpainting task (Yeh et al. 2017). In this paper, we demonstrate that this advantage holds also for highly ill-posed scenarios in image super-resolution and compressed sensing. In addition, note that the DIP approach typically works only with huge U-Nets like architectures that need to be modified for each inverse problem and require much more memory than common generators. Indeed, we struggled (GPU memory overflow, long run-time) to apply DIP to the 1024×1024 images of CelebA-HQ dataset (Karras et al. 2017).

The Proposed Method

In this work, our goal is to make the solutions of inverse problems using generative models more faithful to the observations and more accurate, despite the limited representation capabilities of the pre-trained generators. To this end, we propose an image-adaptive approach, whose motivation is explained both verbally and mathematically (building on theoretical results from (Bora et al. 2017)). We also discuss a back-projection post-processing step that can further improve the results for scenarios with low noise level. While this post-processing, typically, only moderately improves the results of model-based super-resolution algorithms (Glasner, Bagon, and Irani 2009; Yang et al. 2010), we will show that it is highly effective for generative priors. To the best of our knowledge, we are the first to use it in reconstructions based on generative priors.

An Image-Adaptive Approach

We propose to handle the limited representation capabilities of the generators by making them image-adaptive (IA) using internal learning in test-time. In details, instead of recovering the latent signal $\mathbf{x}$ as $\hat{\mathbf{x}} = \mathbf{G}(\hat{\mathbf{z}})$, where $\mathbf{G}(\cdot)$ is a pre-trained generator and $\hat{\mathbf{z}}$ is a minimizer of (2), we suggest to simultaneously optimize $\mathbf{z}$ and the parameters of the generator, denoted as $\boldsymbol{\theta}$, by minimizing the cost function

$$f_{IA}(\boldsymbol{\theta}, \mathbf{z}) = \|\mathbf{y} - \mathbf{A}\mathbf{G}_{\boldsymbol{\theta}}(\mathbf{z})\|_2^2. \tag{3}$$

[1] The function $f(\mathbf{z})$ is non-convex due to the non-convexity of $\mathbf{G}(\mathbf{z})$.

The optimization is done using backpropagation and standard gradient based optimizers. The initial value of $\boldsymbol{\theta}$ is the pre-trained weights, and the initial value of $\mathbf{z}$ is $\hat{\mathbf{z}}$, obtained by minimization with respect to $\mathbf{z}$ alone, as done in CSGM. Then, we perform joint-minimization to obtain $\hat{\boldsymbol{\theta}}_{IA}$ and $\hat{\mathbf{z}}_{IA}$, and recover the signal using $\hat{\mathbf{x}}_{IA} = \mathbf{G}_{\hat{\boldsymbol{\theta}}_{IA}}(\hat{\mathbf{z}}_{IA})$.

The rationale behind our approach can be explained as follows. Current leading learning strategies cannot train a generator whose representation range covers *every* sample of a complex distribution, thus, optimizing only $\mathbf{z}$ is not enough. However, the expressive power of deep neural networks (given by optimizing the weights $\boldsymbol{\theta}$ as well) allows to create a *single* specific sample that agrees with the observations $\mathbf{y}$. Yet, contrary to prior works that optimize the weights of neural networks only by internal learning (Shocher, Cohen, and Irani 2018; Ulyanov, Vedaldi, and Lempitsky 2018), here we incorporate information captured in the test-time with the valuable semantic knowledge obtained by the offline generative learning.

To make sure that the information captured in test-time does not come at the expense of offline information which is useful for the *test image at hand*, we start with optimizing $\mathbf{z}$ alone, as mentioned above, and then apply the joint minimization with a small learning rate and early stopping (details in the experiments section below).

Mathematical Motivation for Image-Adaptation

To motivate the image-adaptive approach, let us consider an L-layer neural network generator

$$\mathbf{G}(\mathbf{z}; \{\mathbf{W}_\ell\}_{\ell=1}^L) = \mathbf{W}_L \sigma(\mathbf{W}_{L-1}\sigma(\ldots\sigma(\mathbf{W}_1\mathbf{z})\ldots)), \tag{4}$$

where $\sigma(\cdot)$ denotes element-wise ReLU activation, and $\mathbf{W}_\ell \in \mathbb{R}^{k_\ell \times k_{\ell-1}}$ such that $k_L = n$. Recall that typically $k_0 < k_1 < \ldots < k_L$ (as $k_0 \ll n$). The following theorem, which has been proven in (Bora et al. 2017) (Theorem 1.1 there), provides an upper bound on the reconstruction error.

Theorem 1. *Let* $\mathbf{G}(\mathbf{z}) : \mathbb{R}^k \to \mathbb{R}^n$ *as given in* (4)*,* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *with* $A_{ij} \sim \mathcal{N}(0, 1/m)$*,* $m = \Omega\left(kL\log n\right)$*, and* $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$*. Let* $\hat{\mathbf{z}}$ *minimize* $\|\mathbf{y} - \mathbf{A}\mathbf{G}(\mathbf{z})\|_2$ *to within additive* ϵ *of the optimum. Then, with probability* $1 - \mathrm{e}^{-\Omega(m)}$ *we have*

$$\|\mathbf{G}(\hat{\mathbf{z}}) - \mathbf{x}\|_2 \le 6E_{rep}(\mathbf{G}(\cdot), \mathbf{x}) + 3\|\mathbf{e}\|_2 + 2\epsilon, \tag{5}$$

where $E_{rep}(\mathbf{G}(\cdot), \mathbf{x}) := \underset{\mathbf{z}}{min}\|\mathbf{G}(\mathbf{z}) - \mathbf{x}\|_2$.

Note that $E_{rep}(\mathbf{G}(\cdot), \mathbf{x})$ is in fact the representation error of the generator for the specific image $\mathbf{x}$. This term has been empirically observed in (Bora et al. 2017) to dominate the overall error, e.g. more than the error of the optimization algorithm (represented by ϵ). The following proposition builds on Theorem 1 and motivates the joint optimization of $\mathbf{z}$ and $\mathbf{W}_1$ by guaranteeing a decreased representation error term.

Proposition 2. *Consider the generator defined in* (4) *with* $k_0 < k_1$*,* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *with* $A_{ij} \sim \mathcal{N}(0, 1/m)$*,* $m = \Omega\left(k_1 L\log n\right)$*, and* $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$*. Let* $\hat{\mathbf{z}}$ *and* $\hat{\mathbf{W}}_1$ *minimize* $\tilde{f}(\mathbf{z}, \mathbf{W}_1) = \|\mathbf{y} - \mathbf{A}\mathbf{G}(\mathbf{z}; \{\mathbf{W}_\ell\}_{\ell=1}^L)\|_2$ *to within additive* ϵ *of the optimum. Then, with probability* $1 - \mathrm{e}^{-\Omega(m)}$ *we have*

$$\|\mathbf{G}(\hat{\mathbf{z}}; \hat{\mathbf{W}}_1, \{\mathbf{W}_\ell\}_{\ell=2}^L) - \mathbf{x}\|_2 \le 6\tilde{E}_{rep} + 3\|\mathbf{e}\|_2 + 2\epsilon, \tag{6}$$

where $\tilde{E}_{rep} \le E_{rep}(\mathbf{G}(\cdot), \mathbf{x})$.

Proof. Define $\tilde{\hat{\mathbf{z}}} := \hat{\mathbf{W}}_1\hat{\mathbf{z}}$ and $\tilde{\mathbf{G}}(\tilde{\mathbf{z}}) := \mathbf{W}_L\sigma(\mathbf{W}_{L-1}\sigma(\ldots\sigma(\mathbf{W}_2\sigma(\mathbf{I}_{k_1}\tilde{\mathbf{z}}))\ldots))$. Note that $\mathbf{G}(\hat{\mathbf{z}}; \hat{\mathbf{W}}_1, \{\mathbf{W}_\ell\}_{\ell=2}^L) = \tilde{\mathbf{G}}(\tilde{\hat{\mathbf{z}}})$, therefore $\tilde{\hat{\mathbf{z}}}$ minimize $\|\mathbf{y} - \mathbf{A}\tilde{\mathbf{G}}(\tilde{\mathbf{z}})\|_2$ to within additive ϵ of the optimum. Applying Theorem 1 on $\tilde{\mathbf{G}}(\tilde{\mathbf{z}})$ and $\tilde{\hat{\mathbf{z}}}$ we have

$$\|\tilde{\mathbf{G}}(\tilde{\hat{\mathbf{z}}}) - \mathbf{x}\|_2 \le 6E_{rep}(\tilde{\mathbf{G}}(\cdot), \mathbf{x}) + 3\|\mathbf{e}\|_2 + 2\epsilon, \tag{7}$$

with the advertised probability. Now, note that

$$\begin{aligned} &E_{rep}(\tilde{\mathbf{G}}(\cdot), \mathbf{x}) \\ &= \min_{\tilde{\mathbf{z}}\in\mathbb{R}^{k_1}} \|\mathbf{W}_L\sigma(\mathbf{W}_{L-1}\sigma(\ldots\sigma(\mathbf{W}_2\sigma(\mathbf{I}_{k_1}\tilde{\mathbf{z}}))\ldots)) - \mathbf{x}\|_2 \\ &\le \min_{\mathbf{z}\in\mathbb{R}^{k_0}} \|\mathbf{W}_L\sigma(\mathbf{W}_{L-1}\sigma(\ldots\sigma(\mathbf{W}_2\sigma(\mathbf{W}_1\mathbf{z}))\ldots)) - \mathbf{x}\|_2 \\ &= E_{rep}(\mathbf{G}(\cdot), \mathbf{x}), \end{aligned} \tag{8}$$

where the inequality follows from $\mathbf{W}_1\mathbb{R}^{k_0} \subset \mathbb{R}^{k_1}$. We finish with substituting $\tilde{\mathbf{G}}(\tilde{\hat{\mathbf{z}}}) = \mathbf{G}(\hat{\mathbf{z}}; \hat{\mathbf{W}}_1, \{\mathbf{W}_\ell\}_{\ell=2}^L)$ in (7) and defining $\tilde{E}_{rep} := E_{rep}(\tilde{\mathbf{G}}(\cdot), \mathbf{x})$.

□

Proposition 2 shows that under the mathematical framework of Theorem 1, and under the (reasonable) assumption that the output dimension of the first layer is larger than its input, it is possible to *further* reduce the representation error of the generator for $\mathbf{x}$ (the term that empirically dominates the overall error) by optimizing the weights of the first layer as well. The result follows from obtaining an increased set in which the nearest neighbor of $\mathbf{x}$ is searched.

Note that if $k_0 < k_1 < \ldots < k_L$, then the procedure which is described in Proposition 2 can be repeated sequentially layer after layer to further reduce the representation error. However, note that this theory loses its meaningfulness at high layers because $m = \Omega\left(k_\ell L\log n\right)$ approaches $\Omega\left(n\right)$ (so no prior is necessary). Yet, it presents a motivation to optimize all the weights, as we suggest to do in practice.

"Hard" vs. "Soft" Compliance to Observations

The image-adaptive approach improves the agreement between the recovery and the observations. We turn now to describe another complementary way to achieve this goal.

Denote by $\hat{\mathbf{x}}$ an estimation of $\mathbf{x}$, e.g. using CSGM method or our IA approach. Assuming that there is no noise, i.e. $\mathbf{e} = 0$, a simple post-processing to strictly enforce compliance of the restoration with the observations $\mathbf{y}$ is back-projecting (BP) the estimator $\hat{\mathbf{x}}$ onto the affine subspace $\{\mathbf{A}\mathbb{R}^n = \mathbf{y}\}$

$$\hat{\mathbf{x}}_{bp} = \underset{\tilde{\mathbf{x}}}{\text{argmin}}\ \|\tilde{\mathbf{x}} - \hat{\mathbf{x}}\|_2^2 \quad \text{s.t.} \quad \mathbf{A}\tilde{\mathbf{x}} = \mathbf{y}. \tag{9}$$

Note that this problem has a closed-form solution

$$\begin{aligned} \hat{\mathbf{x}}_{bp} &= \mathbf{A}^\dagger\mathbf{y} + (\mathbf{I}_n - \mathbf{A}^\dagger\mathbf{A})\hat{\mathbf{x}} \\ &= \mathbf{A}^\dagger(\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}) + \hat{\mathbf{x}}, \end{aligned} \tag{10}$$

where $\mathbf{A}^\dagger := \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}$ is the pseudoinverse of $\mathbf{A}$ (assuming that $m < n$, which is the common case, e.g. in super-resolution and compressed sensing tasks). In practical cases,

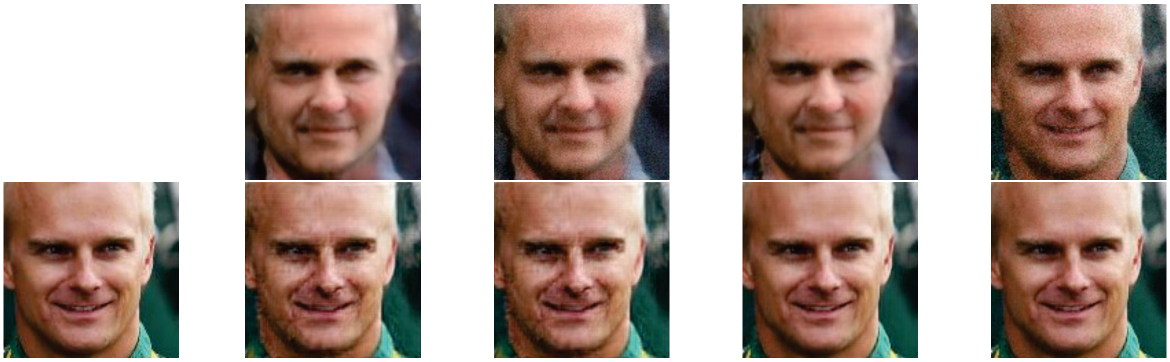

Figure 1: Compressed sensing with Gaussian measurement matrix using BEGAN. From left to right and top to bottom: original image, CSGM for $m/n = 0.122$, CSGM-BP for $m/n = 0.122$, CSGM for $m/n = 0.61$, CSGM-BP for $m/n = 0.61$, IAGAN for $m/n = 0.122$, IAGAN-BP for $m/n = 0.122$, IAGAN for $m/n = 0.61$, IAGAN-BP for $m/n = 0.61$.

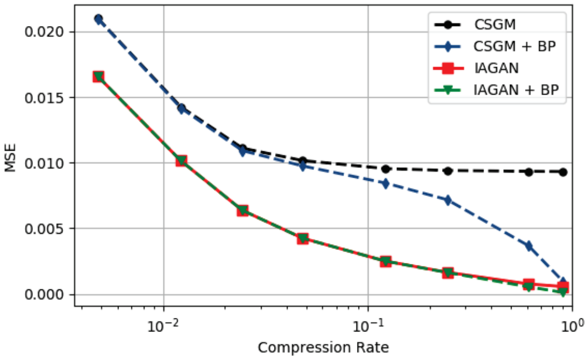

Figure 2: Compressed sensing with Gaussian measurement matrix using BEGAN. Reconstruction MSE (averaged over 100 images from CelebA) vs. the compression ratio m/n.

where the problem dimensions are high, the matrix inversion in $\mathbf{A}^\dagger$ can be avoided by using the conjugate gradient method (Hestenes and Stiefel 1952). Note that when $\mathbf{y}$ is noisy, the operation $\mathbf{A}^\dagger\mathbf{y}$ in (10) is expected to amplify the noise. Therefore, the BP post-processing is useful as long as the noise level is low.

Let $\mathbf{P}_A := \mathbf{A}^\dagger\mathbf{A}$ denote the orthogonal projection onto the row space of $\mathbf{A}$, and $\mathbf{Q}_A := \mathbf{I}_n - \mathbf{A}^\dagger\mathbf{A}$ denote its orthogonal complement. Substituting (1) into (10) gives

$$\hat{\mathbf{x}}_{bp} = \mathbf{P}_A\mathbf{x} + \mathbf{Q}_A\hat{\mathbf{x}} + \mathbf{A}^\dagger\mathbf{e}, \qquad (11)$$

which shows that $\hat{\mathbf{x}}_{bp}$ is consistent with $\mathbf{y}$ on $\mathbf{P}_A\mathbf{x}$ (i.e. displays *"hard" compliance*), and considers only the projection of $\hat{\mathbf{x}}$ onto the null space of $\mathbf{A}$. Therefore, for an estimate $\hat{\mathbf{x}}$ obtained via a generative model, the BP technique essentially *eliminates* the component of the generator's representation error that resides in the row space of $\mathbf{A}$, but does not change at all the component in the null space of $\mathbf{A}$. Still, from the (Euclidean) accuracy point of view, this strategy is very effective at low noise levels, as demonstrated in the experiments section.

Interestingly, note that our image-adaptive strategy enforces only a *"soft" compliance* of the restoration with the observations $\mathbf{y}$, because our gentle joint optimization (which prevents overriding the offline semantic information) may not completely diminish the component of the generator's representation error that resides in the row space of $\mathbf{A}$, as done by BP. On the other hand, intuitively, the strong prior (imposed by the offline training and by the generator's structure) is expected to improve the restoration also in the null space of $\mathbf{A}$ (unlike BP). Indeed, as shown below, by combining the two approaches, i.e. applying the IA phase and then the BP on $\hat{\mathbf{x}}_{IA}$, we obtain better results than only applying BP on CSGM. This obviously implies decreasing the component of reconstruction error in the null space of $\mathbf{A}$.

Experiments

In our experiments we use two recently proposed GAN models, which are known to generate very high quality samples of human faces. The first is BEGAN (Berthelot, Schumm, and Metz 2017), trained on CelebA dataset (Liu et al. 2015), which generates a 128×128 image from a uniform random vector $\mathbf{z} \in \mathbb{R}^{64}$. The second is PGGAN (Karras et al. 2017), trained on CelebA-HQ dataset (Karras et al. 2017) that generates a 1024×1024 image from a Gaussian random vector $\mathbf{z} \in \mathbb{R}^{512}$. We use the official pre-trained models, and for details on the models and their training procedures we refer the reader to the original publications. Note that previous works, which use generative models for solving inverse problems, have considered much simpler datasets, such as MNIST (LeCun et al. 1998) or a small version of CelebA (downscaled to size 64×64), which perhaps do not demonstrate how severe the effect of mode collapse is.

The test-time procedure is done as follows, and is almost the same for the two models. For CSGM we follow (Bora et al. 2017) and optimize (2) using ADAM optimizer (Kingma and Ba 2014) with learning rate (LR) of 0.1. We use 1600 iterations for BEGAN and 1800 iterations for PGGAN. The final $\mathbf{z}$, i.e. $\hat{\mathbf{z}}$, is chosen to be the one with minimal objective value $f(\mathbf{z})$ along the iterations, and the CSGM recovery is $\hat{\mathbf{x}} = \mathbf{G}(\hat{\mathbf{z}})$. Performing a post-processing BP step (10) gives us also a reconstruction that we denote by CSGM-BP.

In the reconstruction based on image-adaptive GANs, which we denote by IAGAN, we initialize $\mathbf{z}$ with $\hat{\mathbf{z}}$, and then

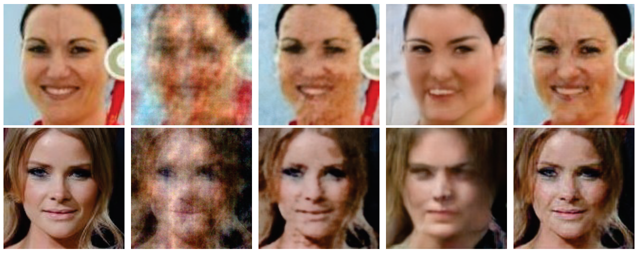

Figure 3: Compressed sensing with 30% (top) and 50% (bottom) subsampled Fourier measurements and noise level of 10/255, for CelebA images. Left to right: original image, naive reconstruction (zero padding and IFFT), DIP, CSGM, and IAGAN. CSGM and IAGAN use the BEGAN prior.

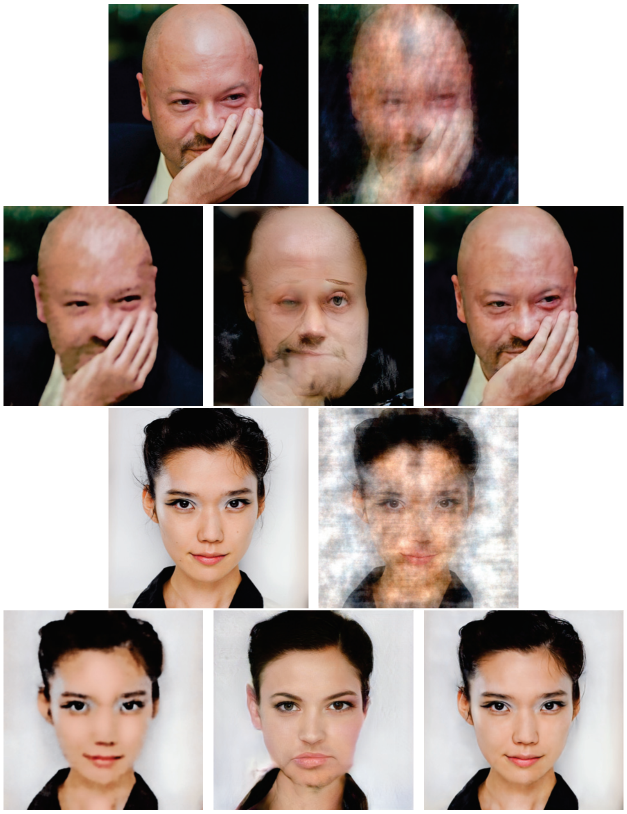

Figure 4: Compressed sensing with 30% (top group) and 50% (bottom group) subsampled Fourier measurements and noise level of 10/255, for CelebA-HQ images. From left to right and top to bottom: original image, naive reconstruction (zero padding and IFFT), DIP, CSGM, and IAGAN. Note that CSGM and IAGAN use the PGGAN prior.

optimize (3) jointly for $\mathbf{z}$ and $\boldsymbol{\theta}$ (the generator parameters). For BEGAN we use LR of 10^{-4} for both $\mathbf{z}$ and $\boldsymbol{\theta}$ in all scenarios, and for PGGAN we use LR of 10^{-4} and 10^{-3} for $\mathbf{z}$ and $\boldsymbol{\theta}$, respectively. For BEGAN, we use 600 iterations for compressed sensing and 500 for super-resolution. For PGGAN we use 500 and 300 iterations for compressed sensing and super-resolution, respectively. In the examined noisy scenarios we use only half of the amount of iterations, to avoid overfitting the noise. The final minimizers $\hat{\boldsymbol{\theta}}_{IA}$ and $\hat{\mathbf{z}}_{IA}$ are chosen according to the minimal objective value, and the IAGAN result is obtained by $\hat{\mathbf{x}}_{IA} = \mathbf{G}_{\hat{\boldsymbol{\theta}}_{IA}}(\hat{\mathbf{z}}_{IA})$. Another recovery, which uses also the post-processing BP step (10) on $\hat{\mathbf{x}}_{IA}$, is denoted by IAGAN-BP.

We also compare the methods with DIP (Ulyanov, Vedaldi, and Lempitsky 2018). We use DIP official implementation for the noiseless scenarios, and for the examined noisy scenarios we reduce the number of iterations by a factor of 4 (tuned for best average performance) to prevent the network from overfitting the noise.

Apart from presenting visual results[2], we compare the performance of the different methods using two quantitative measures. The first one is the widely-used mean squared error (MSE) (sometimes in its PSNR form[3]). The second is a distance between images that focuses on perceptual similarity (PS), which has been proposed in (Zhang et al. 2018) (we use the official implementation). Displaying the PS is important since it is well known that PSNR/MSE may not correlate with the visual/perceptual quality of the reconstruction. Note that in the PS score — lower is better.

Compressed Sensing

In the first experiment we demonstrate how the proposed IA and BP techniques significantly outperform or improve upon CSGM for a large range of compression ratios. We consider noiseless compressed sensing using an $m \times n$ Gaussian matrix $\mathbf{A}$ with i.i.d. entries drawn from $A_{ij} \sim \mathcal{N}(0, 1/m)$, similar to the experiments in (Bora et al. 2017). In this case, there is no efficient way to implement the operators $\mathbf{A}$ and $\mathbf{A}^T$. Therefore, we consider only the BEGAN that generates 128×128 images (i.e. $n = 3 \times 128^2 = 49,152$), which are much smaller than those generated by PGGAN.

Figure 1 shows several visual results and Figure 2 presents the reconstruction MSE of the different methods as we change the number of measurements m (i.e. we change the compression ratio m/n). The results are averages over 20 images from CelebA dataset. It is clearly seen that IAGAN outperforms CSGM for all the values of m. Note that due to the limited representation capabilities of BEGAN (equivalently – its mode collapse), CSGM performance reaches a plateau in a quite small value of m, contrary to IAGAN error that continues to decrease. The back-projection strategy

Figure 5: Binary masks for compressed sensing with 30% (left) and 50% (right) subsampled Fourier measurements.

[2] More examples are presented in the companion technical report (Abu Hussein, Tirer, and Giryes 2019).

[3] **We compute the average PSNR as $10\log_{10}(255^2/\overline{\mathrm{MSE}})$, where $\overline{\mathrm{MSE}}$ is averaged over the test images.**

Table 1: Compressed sensing with subsampled Fourier measurements. Reconstruction PSNR [dB] (left) and PS (Zhang et al. 2018) (right), averaged over 100 images from CelebA and CelebA-HQ, for compression ratios 0.3 and 0.5, with noise level of 10/255.

CelebA	naive IFFT	DIP	CSGM	IAGAN
CS ratio 0.3	19.23 / 0.540	**25.96** / 0.139	20.12 / 0.246	25.50 / **0.092**
CS ratio 0.5	20.53 / 0.495	27.21 / 0.125	20.32 / 0.241	**27.59 / 0.066**
CelebA-HQ	naive IFFT	DIP	CSGM	IAGAN
CS ratio 0.3	19.65 / 0.625	24.97 / 0.566	21.38 / 0.520	**25.80 / 0.429**
CS ratio 0.5	20.45 / 0.597	26.29 / 0.535	21.82 / 0.514	**28.26 / 0.378**

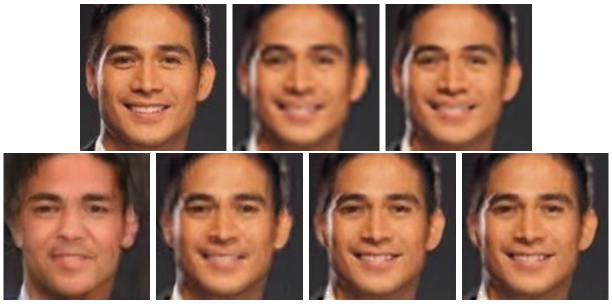

Figure 6: Super-resolution with scale factor 4, bicubic kernel, and no noise, for CelebA images. From left to right and top to bottom: original image, bicubic upsampling, DIP, CSGM, CSGM-BP, IAGAN, and IAGAN-BP. Note that CSGM and IAGAN use the BEGAN prior.

is shown to be very effective, as it makes sure that CSGM-BP is rescued from the plateau of CSGM. The fact that IAGAN still has a very small error when the compression ratio is almost 1 follows from our small learning rates and early stopping, which have been found necessary for small values of m, where the null space of $\mathbf{A}$ is very large and it is important to avoid overriding the offline semantic information. However, this small error is often barely visible, as demonstrated by visual results in Figure 1, and further decreases by the BP step of IAGAN-BP.

In order to examine our proposed IA strategy for the larger model PGGAN as well, we turn to use a different measurement operator $\mathbf{A}$ which can be applied efficiently – the subsampled Fourier transform. This acquisition model is also more common in practice, e.g. in sparse MRI (Lustig, Donoho, and Pauly 2007). We consider scenarios with compression ratios of 0.3 and 0.5, and noise level of 10/255 (due to the noise we do not apply the BP post-processing). The PSNR and PS results (averaged on 100 images from each dataset) are given in Table 1, and several visual examples are shown in Figures 3 and 4. In Figure 5 we present the binary masks used for 30% and 50% Fourier domain sampling of 128×128 images in CelebA. The binary masks that have been used for CelebA-HQ have similar forms.

The unsatisfactory results obtained by CSGM clearly demonstrate the limited capabilities of both BEGAN and PGGAN for reconstruction: Despite the fact that both of them can generate very nice samples (Berthelot, Schumm, and Metz 2017; Karras et al. 2017), they typically can-

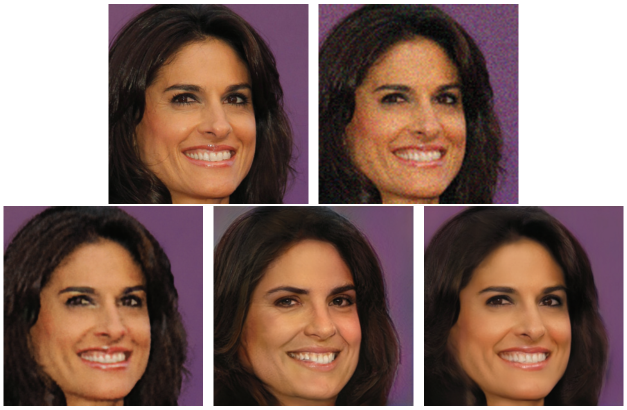

Figure 7: Super-resolution with bicubic kernel, scale factor 8, and noise level of 10/255, for CelebA-HQ images. From left to right and top to bottom: original image, bicubic upsampling, DIP, CSGM, and IAGAN. Note that CSGM and IAGAN use the PGGAN prior.

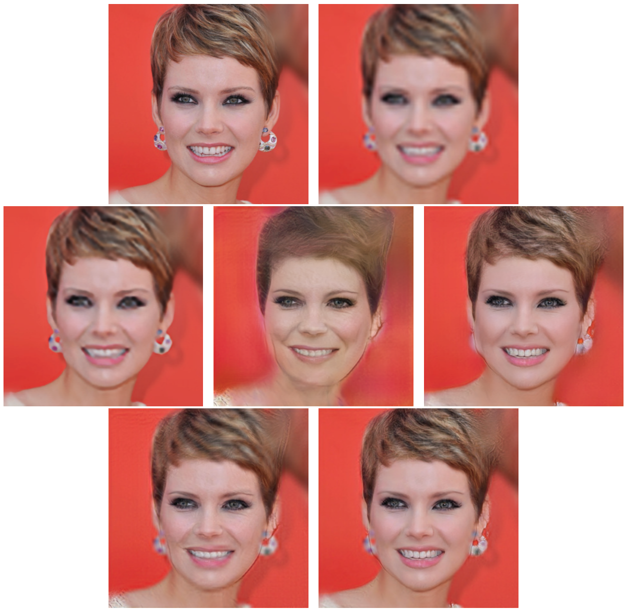

Figure 8: Super-resolution with bicubic kernel, scale factor 16, and no noise, for CelebA-HQ images. From left to right and top to bottom: original image, bicubic upsampling, DIP, CSGM, IAGAN, CSGM-BP, and IAGAN-BP. Note that CSGM and IAGAN use the PGGAN prior.

Table 2: Super-resolution with bicubic downscaling kernel. Reconstruction PSNR [dB] (left) and PS (Zhang et al. 2018) (right), averaged over 100 images from CelebA and CelebA-HQ, for scale factors 4, 8 and 16, with no noise.

CelebA	Bicubic	DIP	CSGM	CSGM-BP	IAGAN	IAGAN-BP
SR x4	26.50 / 0.165	**27.35** / 0.159	20.51 / 0.235	26.44 / 0.165	27.16 / **0.092**	27.14 / **0.092**
SR x8	22.39 / 0.212	23.45 / 0.339	20.23 / 0.240	22.71 / 0.212	23.49 / 0.158	**23.53** / **0.157**

CelebA-HQ	Bicubic	DIP	CSGM	CSGM-BP	IAGAN	IAGAN-BP
SR x8	29.94 / 0.398	**30.01** / 0.400	22.62 / 0.505	28.54 / 0.398	28.76 / 0.387	28.76 / **0.360**
SR x16	27.43 / 0.437	**27.51** / 0.480	22.34 / 0.506	26.20 / 0.437	26.28 / 0.421	25.86 / **0.411**

Table 3: Super-resolution with bicubic downscaling kernel. Reconstruction PSNR [dB] (left) and PS (Zhang et al. 2018) (right), averaged over 100 images from CelebA and CelebA-HQ, for scale factors 4, 8 and 16, with noise level of 10/255.

CelebA	Bicubic	DIP	CSGM	IAGAN
SR x4	24.72 / 0.432	24.19 / 0.280	20.57 / 0.238	**25.54 / 0.133**
SR x8	21.65 / 0.660	21.22 / 0.513	20.22 / **0.243**	**21.72 / 0.243**

CelebA-HQ	Bicubic	DIP	CSGM	IAGAN
SR x8	26.31 / 0.801	**27.61** / 0.430	21.60 / 0.519	26.30 / **0.421**
SR x16	**25.02** / 0.781	24.20 / 0.669	21.31 / 0.516	24.73 / **0.455**

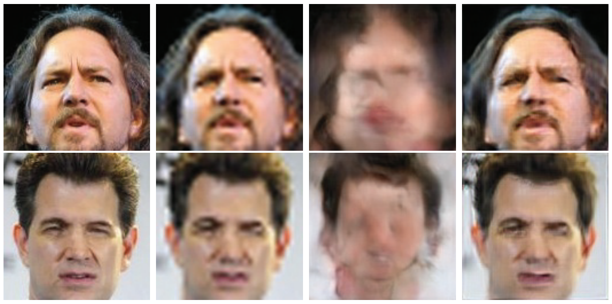

Figure 9: Super-resolution of misaligned images with bicubic kernel and scale factor 4 using BEGAN. Left to right: original image, bicubic upsampling, CSGM, and IAGAN.

not represent well an image that fits the given observations $\mathbf{y}$. This is resolved by our image-adaptive approach. For CelebA dataset DIP has competitive PSNR with our IAGAN. However, both the qualitative examples and the PS (perceptual similarity) measure agree that IAGAN results are much more pleasing. For CelebA-HQ dataset our IAGAN clearly outperforms the other methods.

Inference run-time. Since IAGAN performs a quite small number of ADAM iterations to jointly optimize $\mathbf{z}$ and $\boldsymbol{\theta}$ (the generator's parameters), it requires only a small additional time compared to CSGM. Yet, both methods are much faster than DIP, which trains from scratch a large CNN at test-time. For example, for compression ratio of 0.5, using NVIDIA RTX 2080ti GPU we got the following per image run-time: for CelebA: DIP ~100s, CSGM ~30s, and IAGAN ~35s; and for CelebA-HQ: DIP ~1400s, CSGM ~120s, and IAGAN ~140s. The same behavior, i.e. CSGM and IAGAN are much faster than DIP, holds throughout the experiments in the paper (e.g. also for the super-resolution task).

Super-Resolution

We turn to examine the super-resolution task, for $\mathbf{A}$ which is a composite operator of blurring with a bicubic anti-aliasing kernel followed by down-sampling. For BEGAN we use super-resolution scale factors of 4 and 8, and for PGGAN we use scale factors of 8 and 16. We check a noiseless scenario and a scenario with noise level of 10/255. For the noiseless scenario we also examine the GAN-based recovery after a BP post-processing, which can be computed efficiently, because $\mathbf{A}^{\dagger}$ can be implemented by bicubic upsampling. The PSNR and PS results (averaged on 100 images from each dataset) of the different methods are given in Tables 2 and 3, and several visual examples are shown in Figures 6 - 8.

Once again, the results of the plain CSGM are not satisfying. Due to the limited representation capabilities of BEGAN and PGGAN, the recovered faces look very different than the ones in the test images. The BP post-processing is very effective in reducing CSGM representation error when the noise level is minimal. For our IAGAN approach, the BP step is less effective (i.e. IAGAN and IAGAN-BP have similar recoveries), which implies that the ”soft-compliance” of IAGAN obtains similar results as the ”hard-compliance” of the BP in the row space of $\mathbf{A}$. In the noiseless case, DIP often obtains better PSNR than IAGAN. However, as observed in the compressed sensing experiments, both the visual examples and the PS (perceptual similarity) measure agree that IAGAN results are much more pleasing and sharper, in both noisy and noiseless scenarios. A similar tradeoff between distortion and perception has been recently investigated by Blau and Michaeli (2018). Their work supports the observation that the balance between fitting the measurements and preserving the generative prior, which is the core of our IAGAN approach, may limit the achievable PSNR in some cases but significantly improves the perceptual quality.

We finish this section with an extreme demonstration of mode collapse. In this scenario we use the BEGAN model to super-resolve images with scale factor of 4. Yet, this time the images are slightly misaligned — they are vertically translated by a few pixels. The PSNR[dB] / PS results (averaged on 100 CelebA images) are 19.18 / 0.374 for CSGM and 26.73 / 0.127 for IAGAN. Several visual results are shown in Figure 9. The CSGM is highly susceptible to the poor capabilities of BEGAN in this case, while our IAGAN strategy is quite robust.

Table 4: Deblurring with 9×9 uniform filter and noise level of 10/255. Reconstruction PSNR [dB] (left) and PS (Zhang et al. 2018) (right), averaged over 100 images from CelebA and CelebA-HQ.

CelebA	Blurred	DIP	CSGM	IAGAN
Deb U(9x9)	22.21 / 0.490	25.63 / 0.203	20.37 / 0.241	**26.15 / 0.110**
CelebA-HQ	Blurred	DIP	CSGM	IAGAN
Deb U(9x9)	25.80 / 0.622	**28.28** / 0.458	21.62 / 0.507	28.25 / **0.388**

Deblurring

We briefly demonstrate that the advantage of IAGAN carries to more inverse problems by examining a deblurring scenario, where the operator $\mathbf{A}$ represents blurring with a 9×9 uniform filter, and the noise level is 10/255 (so we do not apply the BP post-processing). The PSNR and PS results (averaged on 100 images from each dataset) of DIP, CSGM, and IAGAN are given in Table 4, and several visual examples are presented in Figure 10.

Similarly to the previous experiments, the proposed IAGAN often exhibits the best PSNR and consistently exhibits the best perceptual quality.

Conclusion

In this work we considered the usage of generative models for solving imaging inverse problems. The main deficiency in such applications is the limited representation capabilities of the generators, which unfortunately do not capture the full distribution for complex classes of images. We suggested two strategies for mitigating this problem. One technique is a post-processing back-projection step, which is applicable at low noise level, that essentially eliminates the component of the generator's representation error that resides in the row space of the measurement matrix. The second technique, which is our main contribution, is an image adaptive approach, termed IAGAN, that improves the generator capability to represent the *specific* test image. This method can improve also the restoration in the null space of the measurement matrix. One can also use the two strategies together. Experiments on compressed sensing and super-resolution tasks demonstrated that our strategies, especially the image-adaptive approach, yield significantly improved reconstructions, which are both more accurate and perceptually pleasing than other alternatives.

Acknowledgments SAH is supported by the NSF-BSF grant (No. 2017729). TT is supported by the European research council (ERC StG 757497 PI Giryes).

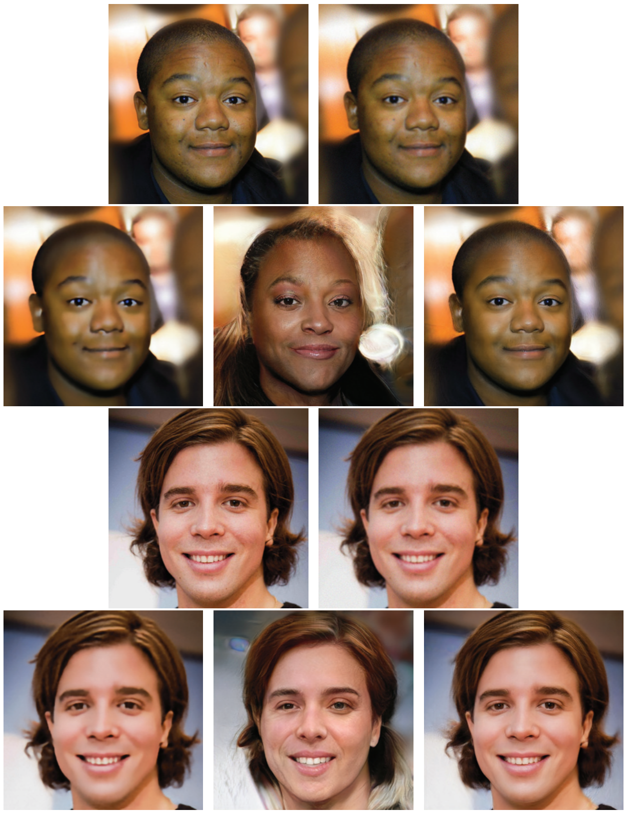

Figure 10: Deblurring with 9×9 uniform filter and noise level of 10/255, for CelebA-HQ images. From left to right and top to bottom (in each group): original image, blurred and noisy image, DIP, CSGM, and IAGAN. Note that CSGM and IAGAN use the PGGAN prior.

References

Abu Hussein, S.; Tirer, T.; and Giryes, R. 2019. Image-adaptive GAN based reconstruction. *arXiv preprint arXiv:1906.05284*.

Arjovsky, M.; Chintala, S.; and Bottou, L. 2017. Wasserstein generative adversarial networks. In *International Conference on Machine Learning*, 214–223.

Bertero, M., and Boccacci, P. 1998. *Introduction to inverse problems in imaging*. CRC press.

Berthelot, D.; Schumm, T.; and Metz, L. 2017. Began: Boundary equilibrium generative adversarial networks. *arXiv preprint arXiv:1703.10717*.

Blau, Y., and Michaeli, T. 2018. The perception-distortion tradeoff. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 6228–6237.

Bojanowski, P.; Joulin, A.; Lopez-Pas, D.; and Szlam, A. 2018. Optimizing the latent space of generative networks. In *International Conference on Machine Learning*, 599–608.

Bora, A.; Jalal, A.; Price, E.; and Dimakis, A. G. 2017. Compressed sensing using generative models. In *International Conference on Machine Learning*, 537–546.

Bora, A.; Price, E.; and Dimakis, A. G. 2018. Ambientgan: Generative models from lossy measurements. In *International Conference on Learning Representations (ICLR)*.

Candes, E.; Romberg, J.; and Tao, T. 2006. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory* 52(2):489–509.

Dhar, M.; Grover, A.; and Ermon, S. 2018. Modeling sparse deviations for compressed sensing using generative models. *arXiv preprint arXiv:1807.01442*.

Dong, C.; Loy, C. C.; He, K.; and Tang, X. 2014. Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*, 184–199. Springer.

Donoho, D. 2006. Compressed sensing. *IEEE Transactions on information theory* 52(4):1289–1306.

Glasner, D.; Bagon, S.; and Irani, M. 2009. Super-resolution from a single image. In *Computer Vision, 2009 IEEE 12th International Conference on*, 349–356. IEEE.

Goodfellow, I.; Bengio, Y.; and Courville, A. 2016. *Deep learning*.

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.

Hand, P.; Leong, O.; and Voroninski, V. 2018. Phase retrieval under a generative prior. In *Advances in Neural Information Processing Systems*, 9154–9164.

Hestenes, M. R., and Stiefel, E. 1952. *Methods of conjugate gradients for solving linear systems*, volume 49.

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*.

Kingma, D. P., and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kingma, D. P., and Welling, M. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P.; et al. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86(11):2278–2324.

Liu, Z.; Luo, P.; Wang, X.; and Tang, X. 2015. Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, 3730–3738.

Lustig, M.; Donoho, D.; and Pauly, J. M. 2007. Sparse MRI: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 58(6):1182–1195.

Meinhardt, T.; Moller, M.; Hazirbas, C.; and Cremers, D. 2017. Learning proximal operators: Using denoising networks for regularizing inverse imaging problems. In *ICCV*, 1781–1790.

Metzler, C. A.; Maleki, A.; and Baraniuk, R. G. 2016. From denoising to compressed sensing. *IEEE Transactions on Information Theory* 62(9):5117–5144.

Richardson, E., and Weiss, Y. 2018. On GANs and GMMs. In *Advances in Neural Information Processing Systems*, 5852–5863.

Rick Chang, J.; Li, C.-L.; Poczos, B.; Vijaya Kumar, B.; and Sankaranarayanan, A. C. 2017. One network to solve them all–solving linear inverse problems using deep projection models. In *ICCV*, 5888–5897.

Shah, V., and Hegde, C. 2018. Solving linear inverse problems using gan priors: An algorithm with provable guarantees. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4609–4613. IEEE.

Shocher, A.; Cohen, N.; and Irani, M. 2018. ”zero-shot” super-resolution using deep internal learning. In *CVPR*.

Tirer, T., and Giryes, R. 2018. Image restoration by iterative denoising and backward projections. *IEEE Transactions on Image Processing* 28(3):1220–1234.

Tirer, T., and Giryes, R. 2019. Super-resolution via image-adapted denoising CNNs: Incorporating external and internal learning. *IEEE Signal Processing Letters*.

Ulyanov, D.; Vedaldi, A.; and Lempitsky, V. 2018. Deep image prior. In *CVPR*.

Venkatakrishnan, S. V.; Bouman, C. A.; and Wohlberg, B. 2013. Plug-and-play priors for model based reconstruction. In *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*, 945–948. IEEE.

Yang, J.; Wright, J.; Huang, T. S.; and Ma, Y. 2010. Image super-resolution via sparse representation. *IEEE transactions on image processing* 19(11):2861–2873.

Yeh, R. A.; Chen, C.; Yian Lim, T.; Schwing, A. G.; Hasegawa-Johnson, M.; and Do, M. N. 2017. Semantic image inpainting with deep generative models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5485–5493.

Zhang, K.; Zuo, W.; Gu, S.; and Zhang, L. 2017. Learning deep cnn denoiser prior for image restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, 3929–3938.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 586–595.