Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

REVIEW ARTICLE

Robustness and Exploration of Variational and Machine Learning Approaches to Inverse Problems: An Overview

Alexander Auras*[†1] | Kanchana Vaishnavi Gandikota[†1] | Hannah Droege[2] | Michael Moeller[1]

[†] These authors contributed equally

[1]Institute for Vision and Graphics, University of Siegen, NRW, Germany
[2]Institute of Computer Science, Rheinische Friedrich-Wilhelms-Universität Bonn, NRW, Germany

Correspondence
*Alexander Auras,
Institute for Vision and Graphics,
University of Siegen,
Adolf-Reichwein-Straße 2a,
57076 Siegen,
Germany.
Email: alexander.auras@uni-siegen.de

Abstract

This paper attempts to provide an overview of current approaches for solving inverse problems in imaging using variational methods and machine learning. A special focus lies on point estimators and their robustness against adversarial perturbations. In this context results of numerical experiments for a one-dimensional toy problem are provided, showing the robustness of different approaches and empirically verifying theoretical guarantees. Another focus of this review is the exploration of the subspace of data consistent solutions through explicit guidance to satisfy specific semantic or textural properties.

KEYWORDS:
Inverse Problems; Machine Learning; Robustness; Explorability;

1 | INTRODUCTION

The goal of image reconstruction is to recover an unknown image from indirect or distorted measurements, i.e., to recover a ground truth image u from measurements

$$f = A(u) + n. \tag{1}$$

for a forward operator f and (additive) noise n. When the forward measurement process is linear, recovering u becomes a linear inverse problem, which is what we focus on in this paper. Simple approaches compute reconstructions u for (1) linearly via least-squares estimates, possibly including an additional smoothing or regularization. Examples of this approach include filtered back projection [50] for tomographic image reconstruction, and Wiener filtering for image deconvolution, which incorporates regularization through linear filtering in Fourier space. Variational approaches (c.f. [16]) to such problems find the minimizer of a suitable cost function, typically consisting of a data fidelity term $E(A, u, f)$ that measures the discrepancy from the observation model (1) and a regularizer $R(u)$ that incorporates prior knowledge about the image to be recovered,

$$\hat{u} = \arg\min_u E(A, u, f) + R(u). \tag{2}$$

At least in finite dimensions, the above perspective relates to the Bayesian approach to inverse problems (see e.g. [44]), via the concept of *maximum a-posteriori probability* (MAP) estimates By modeling the unknown image u and the measurements f as realizations of random variables with respective distributions $p(u)$ and $p(f)$, and computing the MAP estimate as the argument that maximizes the posterior distribution $p(u|f)$, the application of Bayes law yields (2) with $E(A, u, f) = -\log p(f|u)$ and $R(u) = -\log(p(u))$. While (2) is a point estimate, the Bayesian approach to inverse problems inverse models or learns the entire posterior probability $p(u|f)$ and/or sampling schemes for it. Though their notion of solutions is different, both approaches consider the inverse problem to be well-posed if a unique solution exists and depends on the data continuously.

Classical approaches have thoroughly analyzed ill-posed problems and a large body of works provide stability and convergence guarantees, e.g. by selecting (noise-level-dependent) regularizers with suitable properties in (2). Yet, these regularizers are typically not expressive enough to model the distribution of natural/realistic images faithfully. In the past decade, deep learning has achieved remarkable success in image reconstruction through the ability to capture data-dependent structures, showing great improvements in reconstruction quality over earlier classical approaches. This survey attempts to present an overview of recent deep learning-based image reconstruction methods and discuss issues related to the stability, robustness, and explorability of solutions.

2 | OVERVIEW OF DEEP LEARNING FOR INVERSE PROBLEMS IN IMAGING

In this section, we give a brief overview of different strategies to solve inverse problems by involving concepts (i.e. learning or parametrization strategies) from deep learning. We distinguish two subcategories, i.e., point estimators (Sec. 2.1) that deterministically predict a single solution to an inverse problem (similar to the MAP estimate (2)), and methods that allow stochastic sampling from the posterior $p(u|f)$ (Sec. 2.2).

2.1 | Deep Learning for Point Estimates to Inverse Problems

Direct Approaches

In our summary below, we distinguish two types of approaches: Direct approaches (this paragraph) that aim at directly predicting a solution of an inverse problem and are therefore specific to a (class of) forward operators A, and network priors that learn (encodings of) the prior probability $p(u)$ only to keep the versatility of variational approaches (2).

Fully Learned Approaches. A straightforward approach is to train a deep network $\mathcal{G}_\theta$, i.e., a function parameterized by θ, in a supervised manner to directly invert the measurement process:

$$\hat{u} = \mathcal{G}_\theta(f), \tag{3}$$

where $\mathcal{G}_\theta$ is trained by minimizing the expectation of some distortion measure $\mathcal{L}$ between the network output and the ground truth over a set of training examples. Common networks are trained using simple pixel-wise ℓ_2 or ℓ_1 reconstruction losses [173, 21, 91, 180], losses based on structural similarity metrics [179], perceptual losses [87] or adversarial losses by discriminating between the network output and the clean data distribution [166, 97] in addition to pixel-wise losses. Networks trained using such an approach have achieved better reconstruction quality in many imaging tasks than classical approaches, as learning on specific datasets allows the networks to capture a data-dependent context. As such approaches typically do not take the forward operator A into account in the reconstruction process, they can also be applied when the forward model is not known or cannot be modeled accurately, for example, in blind image restoration tasks [127, 175]. Yet, such fully learned approaches are specific to the task they have been trained on, which contrasts the rather versatile variational approaches (2), in which the forward operator, discrepancy measure, and regularization term can be exchanged flexibly. Moreover, using deep, rather general parametrizations without encoding knowledge of the measurement process (1) explicitly commonly leads to networks whose properties remain poorly understood and are frequently referred to as "black-box" approaches. The latter has led to significant research on combinations of interpretable classical and powerful learning-based approaches.

Learned Post-processors. The simplest approach to incorporate model knowledge into a learning-based approach is to train a post-processor network on removing artifacts from an initial reconstruction obtained from an analytical linear reconstruction operator $A^\dagger_{\text{reg}}$ such as the adjoint, the pseudo-inverse, or a regularized version thereof (c.f. [122, 101, 27]):

$$\hat{u} = \mathcal{G}_\theta(A^\dagger_{\text{reg}}(f)). \tag{4}$$

A common approach for such post-processing networks is to use residual learning [73] to recover the difference between initial reconstruction and ground truth. To ensure measurement consistency of solutions obtained by such a post-processing scheme, [151] explicitly constrain the learned residual to be in the null space of the forward operator.

Unrolled Optimization. Unrolled optimization [66] uses the knowledge of the forward model to alternate between measurement and image spaces in an iterative algorithm with a fixed number of iterations, where some of the intermediate operations are learned using parameterized deep network modules. Starting from [66], several works proposed unfolding different iterative model-based algorithms, for instance, learned ISTA [66, 176], learned ADMM [159], learned gradient descent [1, 58], learned

primal-dual [2], or proximal gradient algorithms [111, 136]. Instead of learning a different set of network parameters for the proximal step in each iteration, a few works [72, 68, 3] share the same network for the proximal steps across iterations. In comparison with the fully learned approaches, unrolled networks tend to require less training data, and allow for more interpretable, and parameter-efficient learning [119]. As the unrolled networks are trained typically using a small number of unrolled steps, the inference is also faster in comparison to classical variational approaches, which may need more iterations to converge. On the other hand, testing unrolled networks using more inference steps than used in training typically results in severe artifacts. Recent work [56] addresses this shortcoming through deep equilibrium models [13] which incorporate fixed-point convergence by construction. Such models automatically share the same set of parameters across any number of iterations for solving the fixed-point equation.

Neural Network Priors

An alternate approach for incorporating model knowledge into learning-based techniques is to use neural networks as priors in variational inference. In contrast to the dedicatedly trained networks, this approach endows the algorithm with the flexibility to handle different measurement models, while improving the performance of handcrafted priors. This class of methods includes learning regularizers, using trained networks such as denoisers, generative models, and even untrained neural networks as priors in variational image recovery.

Learned Regularizers: Classical approaches to learning regularizers include the Field of Experts [143], dictionary learning [5, 110], learning regularizer via bi-level optimization, e.g. [96, 28], or learning a (componentwise) proximal operator [150]. Learning deep network regularizers often involves explicitly parameterizing the regularization functional $R(\cdot)$ in (2) using a neural network R_θ [102, 108, 124, 93, 64] which may be trained based on different objectives. While [102] uses a neural network trained to penalize artifacts in the recovered solution, [93] trains a neural network regularizer motivated by sparsity penalties. [108, 134, 124] learn regularizers which are trained adversarially to distinguish between samples from the training data distribution and degraded samples. Instead of directly parameterizing the regularizers, [26] learns a proximal operator of a regularizer, and [74] learns the projection operators onto clean data manifolds. While learned regularizers improve reconstruction performance over handcrafted priors, they may not always guarantee stability or convergence, which requires imposing additional constraints on the regularizer. [118] instead train networks to output descent direction with a provable convergence to a minimizer of the (differentiable) energy, while [48] expanded this approach for non-smooth energies. Some works constrain the regularizer to ensure a convergent iterative scheme. [108, 124] impose Lipschitz-continuity on the regularizer via a soft-penalty, and [123] enforce convexity of regularizer using input convex networks [7] for convergence. We refer to [125] for a review of learned reconstruction methods with convergence guarantees.

Denoisers as Priors: Pretrained denoisers have been employed as priors in image recovery- as proximal operators, or in a functional representing the gradient of regularizer. Plug-and-Play (PnP) methods [164, 24] replace proximal operators of a regularizer by generic denoisers such as non-local means [20] or BM3D [40] in proximal splitting algorithms. Subsequently [177, 115, 178] proposed the use of pretrained neural network denoisers as proximal operators with good empirical results. An alternate approach is regularization by denoising (RED) using denoisers D_θ in a regularization functional of the form $\langle u, u - D_\theta(u)\rangle$ [17, 141] in a gradient descent based scheme. While both PnP and RED approaches empirically provide very good reconstructions, they require strong conditions on the denoiser to have convergence guarantees. The denoiser replacing the proximal operator should be non-expansive, or in the RED framework, the denoiser should additionally have a symmetric Jacobian. These restrictive conditions are not satisfied by arbitrary denoising networks [140]. A few approaches constrain the denoiser to satisfy properties required for convergence, for instance, [146, 161] train denoisers with constrained Lipschitz constants, [38] derive image denoisers with symmetric Jacobians, and [71] parameterize 1-Lipschitz operators for denoising. Instead of constraining the denoisers, [156] project the outputs of arbitrary denoisers onto the cone of descent directions to a given energy in a (proximal) gradient descent algorithm for provable convergence.

Untrained Neural Network Prior: In [162] Ulyanov et al. proposed to use the structure of a randomly initialized convolutional generator to capture natural image statistics, referred to as 'Deep Image Prior' (DIP), and used this to solve inverse problems by optimizing untrained network weights to minimize reconstruction error:

$$\hat{u} = \mathcal{G}(z_0; \hat{\theta}) \;\text{ s.t. }\; \hat{\theta} = \arg\min_\theta \|f - A\mathcal{G}(z_0; \theta)\|^2. \tag{5}$$

Their work used an over-parameterized UNet [142] for $\mathcal{G}$ and suggested early stopping of the optimization in (5) to prevent overfitting. [75] instead use an under-parameterized non-convolutional generator which prevents overfitting. More recent works [29, 77, 105] even search for neural architectures to be used as deep image priors. [81] present a projected gradient descent scheme

for solving (5) using under-parameterized networks [75] and provide convergence guarantees for their scheme. A few works [31, 104, 114] have also combined deep image priors with additional regularization. [31] considers a Bayesian perspective of the deep image prior as a Gaussian process and computes MMSE estimate $\hat{u}$ by optimizing both θ and z using ℓ_2 regularization on both. [104] employ TV regularization on the DIP network output. [114] use a combination of deep image prior and regularization by denoising, and [163] combine DIP with learned regularization.

2.2 | Deep Learning for Bayesian Approaches to Inverse Problems

While point estimators are certainly interesting, they do not describe all possible solutions given the observation f, which are commonly represented by the *posterior* $p(u|f)$. Stochastic approaches to inverse problems allow to sample solutions from $p(u|f)$, which is particularly desirable for under-determined inverse problems. This is useful for uncertainty quantification [126, 154] as it not only allows computing point estimates such as the minimum mean squared error (MMSE) estimator but also deriving higher-order statistics such as the variance of each pixel of the reconstructions. Existing deep learning approaches to stochastic image recovery frequently utilize conditional or unconditional generative models. The former approach involves training a conditional generative model, i.e., a model that gets additional inputs such as the measured data, for a specific recovery task, and the latter approach uses a pre-trained generative model as a prior for the image recovery. We will briefly summarize works in both directions below.

Conditional Generative Models: Conditional generative models train generative approaches with the data of an inverse problem as an additional input, where the training phase ensures, or at least encourages data-consistency of the predicted solutions following (1). Consequently, they can sample multiple solutions $\hat{u}$ for a given observation f [9, 130, 103, 106, 132]. Most commonly this is achieved by supervised training of conditional generative models such as conditional GANs [86, 11, 129], conditional flow models [172, 106, 154, 85], or conditional diffusion models [148, 103, 147, 171]. A few of these methods also guarantee consistency of the reconstruction with input either by an explicit projection operation [11], or by choosing inherently invertible generative models [106, 85, 9, 130]. Only a few works [155, 145] use unsupervised or unpaired learning to learn conditional generative models. Any of the above approaches represents the posterior as a transformation of a simple distribution (e.g. a Gaussian) via a parameterized mapping, the conditional generative model, such that samples from the posterior can be drawn by feeding different samples from the initial distribution into the trained network, or, more formally, considering the push-forward of the latent distribution under the conditional generative model.

Generative Priors: Generative models such as generative adversarial networks (GANs) [60], variational autoencoders (VAEs) [92], normalizing flows [47], and diffusion-based or score-based models [158, 76] are trained to produce new samples from the underlying distribution of the training data, and therefore can serve as useful priors when the image to be recovered belongs to this distribution. These models learn a generator $\mathcal{G}_\theta$ to transform a simple distribution $p(z)$ on a latent space (e.g. a Gaussian) to the image distribution $p(u)$ (as opposed to the previous paragraph, where models try to directly predict the posterior). [19] proposed the use of deep generative model priors for image recovery by optimizing for a vector in the smaller dimensional latent space of a trained GAN or a VAE to minimize the reconstruction error:

$$\hat{u} = \mathcal{G}_\theta(\hat{z}) \;\text{ s.t. }\; \hat{z} = \arg\min_z \|f - A\mathcal{G}_\theta(z)\|^2, \tag{6}$$

with an ℓ_2 regularization on z using simple gradient descent-based methods, and demonstrated significant improvements over the classical priors for compressive sensing with a small number of measurements. For compressed sensing using random Gaussian matrices, they show that (6) results in solutions close to the ground truth with high probability under certain conditions. Their work was later extended to non-linear inverse problems in [70, 18]. [174] proposed image inpainting by using Poisson blending using the image that is closest in the latent space of the generator to the input corrupted image. [25] proposed latent space optimization of a generative autoencoder for light field recovery. [152, 138] investigated the use of projected gradient descent, and [98] proposed the use of the alternating direction method of multipliers (ADMM) for image recovery using GAN priors. [135] utilize hierarchical VAEs in an efficient Plug-and-Play algorithm for general inverse problems. An advantage of latent space optimization is the ability to obtain multiple solutions by using different initial latent codes [116, 113, 131], which can be accelerated by finding latent space directions in the null space of the forward operator [120]. Yet, such strategies rather sample different local maxima of the posterior $p(u|f)$ than reflecting the posterior itself, see Fig. 1 for an illustration. A major limitation of the latent space optimization (6) is that samples outside the range manifold of the generator cannot be reconstructed accurately resulting in a non-trivial representation error. Subsequent works attempt to reduce this representation error using different approaches. [46] allows a small deviation of the recovered image from the range of a generator with sparsity prior on

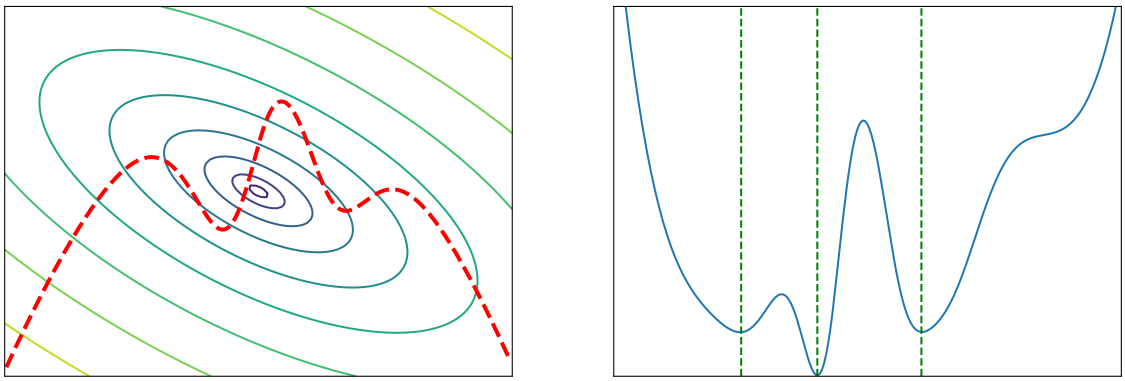


Figure 1 The approach (6) corresponds to a posterior that is the product of the likelihood $p(f|u)$ and a prior $p(u)$ that restricts u to the range of a generator $\mathcal{G}_\theta(\hat{z})$. The left plot illustrates level lines of $-\log(p(f|u))$ in 2d along with a lower dimensional manifold that is the range of $\mathcal{G}_\theta(\hat{z})$ as a dashed red line. The resulting costs are shown on the right. Running gradient descent from different starting points can merely sample local minima (dashed green lines) that correspond to local maxima of the posterior.

their difference, which is extended to optimizing intermediate layer representations in [41]. [80, 131] adopt a two-step approach of latent space optimization followed by fine-tuning both the latent vector and generator parameters. [59] proposes a framework for inverse problems using VAEs by considering a joint posterior distribution of latent z and image space u which guarantees converges to a stationary point. [10] replace the GAN prior in (6) with a flow-based generator, with an ℓ_2 regularization on z. [169, 94] generalize this to arbitrary differentiable measurement operators and measurement noises using a maximum a-posteriori framework, with [94] using a generalized version of flow models which progressively increase dimension from a low-dimensional latent space.

The works [83, 90, 100, 82] adopt Langevin dynamics for linear inverse problems and incorporate the guidance from measurement through the gradient of the log-posterior into their iterative process, or via a projection operation [90]. Several recent works incorporate the knowledge of the forward operator to modify the reverse sampling process in denoising diffusion models. This can be done by alternating between a standard reverse diffusion step and a projection operation for promoting measurement consistency [32, 35, 167, 107]. An alternate approach is to e.g. use a gradient descent step on the data fidelity term [37, 36] or the pseudo-inverse of the forward operator [157] using a clean estimate at each step of the reverse diffusion process, where [36] additionally include a correction step through projection. While approaches that employ only one projection operation per denoising step (such as [167, 107]) are faster, they are restricted to linear inverse problems. On the other hand, guidance using the gradient of a data fidelity term [37] can be applied even to non-linear inverse problems, yet it is more expensive as it requires back-propagation through the diffusion model weights at each iteration. [170, 51] train conditional flow based models to parameterize the posterior $p(u|f)$, given a pretrained generative model representing the prior. More recently, [112] adopt diffusion models in a regularization-by-denoising framework, and [181] demonstrate their utility for plug-and-play image restoration as an effective alternative to the standard Gaussian denoisers.

3 | STABILITY AND ROBUSTNESS

As deep learning approaches are increasingly adopted in image recovery tasks, characterizing the vulnerabilities and instabilities of neural networks for image recovery is important, especially in safety-critical applications like medical imaging. While adversarial robustness is extensively studied for image classification, see e.g. [160, 61, 109], it is less studied in the context of image recovery. The notion of robustness itself is very different for classification and reconstruction problems. For a classifier, robustness can be characterized by the minimal perturbation which can cause a sample to cross the decision boundary, leading to a change in classification outcome. For reconstruction tasks, the outputs are not discrete labels, and there is no notion of a decision boundary. Instead, the output of the reconstruction algorithm should vary continuously/smoothly with changing inputs. The latter links to the mathematical study of inverse problems in infinite-dimensional spaces, where the pseudo-inverse of the

linear operator A in (1) is neither continuous nor defined everywhere as soon as A is compact and has an infinitely dimensional range. Consequently, one has to balance the desire to have a suitable type of continuity in the reconstruction with the faithful approximation of the pseudo-inverse depending on the expected noise. Furthermore, the ill-posedness of the inverse problem might arise from a lack of uniqueness, e.g., due to a forward operator with a non-trivial null-space.

For finite dimensional linear inverse problems, the degree of "ill-posedness" can be quantified by the condition number $\kappa(A) = \frac{\sigma_{max(A)}}{\sigma_{min(A)}}$ of the given operator A, with $\sigma_{\{max,min\}}(A)$ denoting the largest/smallest singular value. In the infinite-dimensional setting, 0 is an accumulation point of the singular values and the severity of the ill-posedness is characterized by how fast the singular values converge to zero.

The aforementioned notion of ill-posedness motivates the (classical) linear regularization of the singular values (spectral regularization): Consider a forward operator A with singular value decomposition

$$Au = \sum_{i=1}^{\infty} \sigma_i \langle u, \mu_i \rangle \nu_i, \tag{7}$$

where σ_i again denote the singular values, while μ_i and ν_i are the i-th left/right singular vectors. One replaces the (unbounded) $1/\sigma_i$ that arise in the pseudoinverse by a suitable bounded approximation $g_{\alpha_i}(\sigma_i)$, i.e.,

$$R_\alpha(f) = \sum_{i=1}^{\infty} g_{\alpha_i}(\sigma_i) \langle f, \nu_i \rangle \mu_i \tag{8}$$

with regularization functions $g_{\alpha_i}(\sigma)$ parameterized by one or multiple α_i that determine a (noise-dependent) balance between the boundedness of the reconstruction operator and the faithfulness of approximating $A^\dagger$. Although being limited to specific linear regularizations (and therefore typically being suboptimal in imaging applications) the analysis of such approaches is well-established and conditions for choosing α such that $R_\alpha(f)$ converges to the true solution are well-established. Recent work has extended such analysis to learned regularization functions g_{α_i}. It demonstrated that an analytical solution for the optimal regularization can be computed and yields stability guarantees, see e.g. [14, 89].

A common way to characterize stability in finite dimensions, particularly in the neural network community, has been to use the notion of Lipschitz continuity. If a reconstruction algorithm $\mathcal{G}$ satisfies

$$\|\mathcal{G}(f + \delta) - \mathcal{G}(f)\| \leq L\|\delta\|, \tag{9}$$

then $\mathcal{G}$ is a Lipschitz continuous mapping with Lipschitz constant L, where $\|\cdot\|$ on both sides of (9) is commonly chosen to be the ℓ_2 norm. From the point of view of stability, a small value of L is desirable to ensure that the maximal change in the reconstruction caused by a small change in measurements remains small. While Lipschitz continuity provides a useful notion of stability, analyzing the stability of common neural networks in terms of Lipschitz constants is difficult, owing to the high complexity involved in its exact computation, even for moderately sized neural networks [88]. In particular, computing the smallest Lipschitz constant was shown to be NP-hard even for a 2-layer fully connected network in [165]. As a result most works [168, 165, 39] only compute approximations and upper bounds for the smallest Lipschitz constant of neural networks. On the other hand, stability in terms of a Lipschitz continuous network seems to come at the cost of the reconstruction performance. For instance, [156] observed that enforcing non-expansiveness ($L \leq 1$) drastically decreased the denoising performance of neural network denoisers. Moreover, even analytically, for an ill-posed problem with a forward operator A that yields a one-to-one correspondence between ground truth and measurements, the Lipschitz constant has to be noise-level-dependent and has to tend to infinity as the reconstruction operator approximates $A^\dagger$. A noise-level independent, fixed (small) value of L implies the inability of $\mathcal{G}$ to accurately reconstruct the ground truth. See [63] for a discussion.

[45] show that variational energy minimization approaches of the specific form $\hat{u} = \arg\min_u \|Au - f\|^2 + \lambda\|u\|_p^p$ for $p \in (1, \infty)$ show good stability properties, with Tikhonov ℓ_2 regularized reconstruction map being globally Lipschitz continuous. The reconstruction map is locally Lipschitz continuous in the measurement space for $p \in (1, 2)$, and globally $\frac{1}{p-1}$ Hölder continuous[1] for $p \in (2, \infty)$. In general, all variational energy minimization approaches permit a stability estimate in the case of linear inverse problems using convex regularizers, as shown in [22]. The optimality condition for (2) with $E(A, u, f) = \frac{1}{2}\|Au - f\|^2$ is

$$0 \in A^*(Au - f) + \partial R(u). \tag{10}$$

[1] $\mathcal{G}$ is α Hölder continuous if $\|\mathcal{G}(f + \delta) - \mathcal{G}(f)\|_p \leq K\|\delta\|_2^\alpha$

(a) Mean of the images in (b)

(b) Various natural images

Figure 2 Convex regularizers rate convex combinations (a) of images (b) as at least as "natural" as one of the images itself.

Taking the difference between the two optimality conditions arising from two different measurements f_1 and f_2 with their corresponding reconstructions u_1 and u_2 and subsequently taking the inner product with $u_1 - u_2$ yields

$$\begin{aligned} 0 &\in \langle A^*(Au_1 - f_1) - A^*(Au_2 - f_2) + \partial R(u_1) - \partial R(u_2), u_1 - u_2 \rangle, \\ &\in \|Au_1 - Au_2\|^2 - \langle f_1 - f_2, Au_1 - Au_2 \rangle + \langle \partial R(u_1) - \partial R(u_2), u_1 - u_2 \rangle. \end{aligned} \tag{11}$$

The second term can now be bounded from above by applying $\langle a, b \rangle \leq \|a\|\|b\|$ and $ab \leq \frac{a^2}{2} + \frac{b^2}{2}$:

$$\langle f_1 - f_2, Au_1 - Au_2 \rangle \leq \frac{1}{2}\|f_1 - f_2\|^2 + \frac{1}{2}\|Au_1 - Au_2\|^2.$$

We find that there have to exist subgradients $p_1 \in \partial R(u_1)$ and $p_2 \in \partial R(u_2)$ such that

$$\frac{1}{2}\|f_1 - f_2\|^2 \geq \frac{1}{2}\|Au_1 - Au_2\|^2 + \langle p_1 - p_2, u_1 - u_2 \rangle$$

or alternatively that

$$\frac{1}{2}\|f_1 - f_2\|^2 \geq \frac{1}{2}\|Au_1 - Au_2\|^2 + \mathcal{D}_R(u_1, u_2) \tag{12}$$

with $\mathcal{D}_R(a, b)$ denoting the symmetric Bregman distance with respect to the convex regularizer R. Note that – as opposed to (9) – the natural stability of a variational approach is the sum of a data consistency and a regularization-specific measure of distance. The latter can be a rather weak notion of a difference as only the properties $\mathcal{D}_R(u_1, u_2) \geq 0$ and $\mathcal{D}_R(u, u) = 0$ can be guaranteed, unless R is m-strongly convex, in which case $\mathcal{D}_R(u_1, u_2) \geq m\|u_1 - u_2\|^2$. Therefore, even though learned convex regularizers provably satisfy (12), their underlying symmetric Bregman distance might remain difficult to interpret. Interesting future research could therefore involve architectures or additional loss functions for encouraging a particularly meaningful Bregman distance. We further refer the reader to [15] for error estimates with non-quadratic data fidelity terms.

Yet, although the analysis of convex variational methods, their ability to determine global minimizers and stability properties like (12) are highly appealing, they are systematically suboptimal to act as a prior for natural images as illustrated in Fig. 2: According to any convex regularizer R, any convex combination of natural images is "at least as likely" to be a natural image as one of its constituting images. This property makes a strong case for nonconvex regularizers, which, even though being more mathematically challenging to analyze, are slowly coming into the focus of the research community (see e.g. [133] for invex regularizers). Since nonconvex or general learning-based approaches, however, remain very challenging to analyze mathematically, a large body of work on the empirical analysis of stability exists for different notions of robustness including robustness to adversarial perturbations, robustness in recovering fine details, robustness to changes in forward measurement operator and robustness to distribution shifts in data. We will discuss these notions of robustness along with recent findings in the following subsections.

3.1 | Adversarial Robustness

Adversarial robustness of a learned reconstruction operator $\mathcal{G}_\theta$ can be quantified by the maximum deviation caused in the reconstruction by a small perturbation in the measurement. The adversarial perturbation causing the maximum reconstruction can be

obtained as

$$\delta_{\text{adv}} = \underset{\delta \in B_\epsilon}{\arg\max}\ d(\mathcal{G}_\theta(f+\delta), \mathcal{G}_\theta(f)) \tag{13}$$

for a suitable measure of distance d between two reconstructions, most commonly $d(\mathcal{G}_\theta(f+\delta), \mathcal{G}_\theta(f)) = \left\|\mathcal{G}_\theta(f+\delta) - \mathcal{G}_\theta(f)\right\|_2$, and a suitable set B_ϵ of perturbations, e.g. $B = \{\delta \mid \|\delta\|_\infty \leq \epsilon\}$. When the maximum deviation $d(\mathcal{G}_\theta(f+\delta), \mathcal{G}_\theta(f))$ is small with respect to δ_{adv}, the reconstruction operator $\mathcal{G}_\theta$ can be considered robust. This notion of robustness is closely related to Lipschitz continuity. In practice the optimization problem in (13) is seldom solved exactly as this is prohibitively complex, and is approximated using a small number of projected gradient ascent steps.

Recent works starting from [8, 33] have characterized the instabilities of deep learning-based image recovery methods. The authors of [33] demonstrate the susceptibility of deep networks for image super-resolution to adversarial perturbations, with a focus on untargeted attacks. [8] show that end-to-end trained deep networks for image recovery are susceptible to adversarial perturbations. They find that perturbations optimized for the networks do not transfer to classical approaches like ℓ_1 minimization with sparsity constraints, and conclude that these classical methods are more robust than learned approaches. On the other hand, [55, 42, 54] analyze the stability of both classical and deep learning approaches to image recovery, and show that even classical approaches are susceptible to adversarial perturbations optimized for these methods, considering $d = \|\cdot\|_2$ and B_ϵ being the ℓ_∞ ball. Further, the work [54] also shows that adversarial perturbations optimized for classical approaches transfer well to learned approaches, indicating that these methods do share some common directions of vulnerability. One would anticipate such directions of vulnerability to lie in a subspace spanned by singular vectors of the forward operator corresponding to small singular values, which remains an interesting investigation for future research, particularly considering that classification networks remain susceptible to low-frequency attacks, c.f. [67], which are not part of the aforementioned subspace for many prominent inverse imaging problems.

Because most works analyze robustness by considering deviations in the reconstructed image in terms of the ℓ_2 metric, the conclusion that classical (convex, variational) methods are also susceptible to adversarial perturbations in general, would be too broad, particularly considering that they are *provably robust* to ℓ_2 perturbations in the data space when robustness is measured in terms of data consistency and the symmetric Bregman distance of the regularizer, see (3). We will illustrate the differences between these perspectives in a toy problem of recovering one-dimensional signals following the empirical evaluation of Genzel *et al.*[55] below.

Different Notions of Robustness - Numerical Experiments on a Toy Problem.

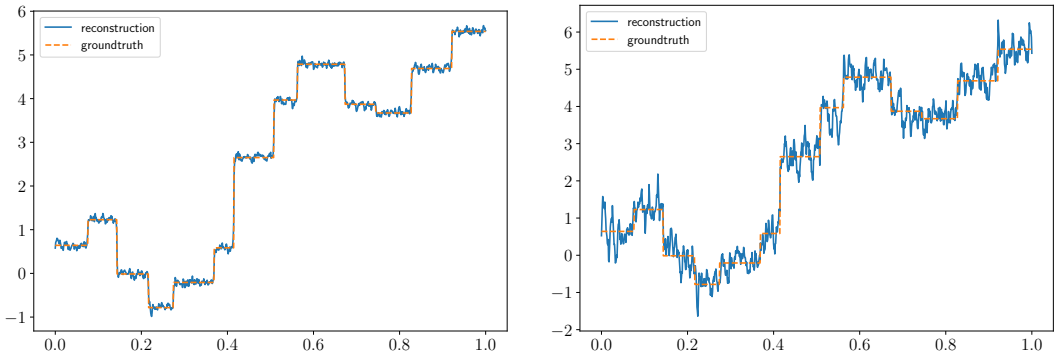
We generate signals u as discretizations of piece-wise constant functions with a random number of jumps, each varying in height. The forward operator $A \in \mathbb{R}^{\frac{N}{2} \times N}$ is chosen to be a compressed sensing operator with $N = 1024$, containing random numbers drawn from a Gaussian distribution with mean 0 and variance 0.05. We compare the model-based approaches of Tikhonov (ℓ_2-squared) and total variation (TV) reconstruction to learning-based approaches with a post-processing U-Net [142] architecture, an end-to-end learned Tiramisu architecture [84] without any model based components and a model-motivated (plug-and-play) architecture (denoted as *ItNet*), incorporating a U-Net based proximal step. We measure their robustness against adversarial attacks by projected gradient descent with a projection to an ℓ_∞-ball with radius ϵ. We chose $\epsilon = 0.2$ in our experiments. Details about the optimization process, hyperparameter choices, and further information can be found in the codebase, which will be made publically available[2]. In our first experiment, we directly reconstruct the signal by employing a Tikhonov-like regularized inversion method

$$\hat{u}_{Tik} = (A^T A + \alpha D^T D)^{-1} A^T f \tag{14}$$

where the matrix D serves as a finite difference matrix, $\hat{u}_{Tik}$ is the estimated solution, and α is a parameter that balances the fidelity to the data with the smoothness of the reconstruction.

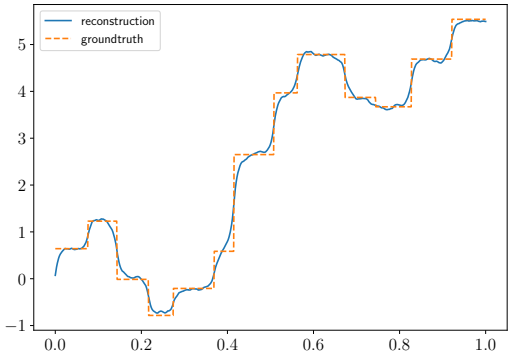
As shown in figure 3 (a), this reconstruction, while being able to capture the basic structure of the solution, suffers due to the noisy nature of the measurements, leaving a lot of room for improvement. Additionally, due to the low regularization strength, the approach is very susceptible to adversarial noise, as is clearly visible in figure 3 (b). While $\alpha = 10^{-7}$ is a good value for minimizing the ℓ_2-error in comparison to the ground truth, we also show the results for a much stronger regularization (figure 3 (c) and (d)): While the resulting reconstruction exhibits a clear oversmoothing, it is significantly less affected by the adversarial attack, showing a tradeoff between robustness and fidelity.

[2]https://github.com/AlexanderAuras/GAMM-Overview-23/

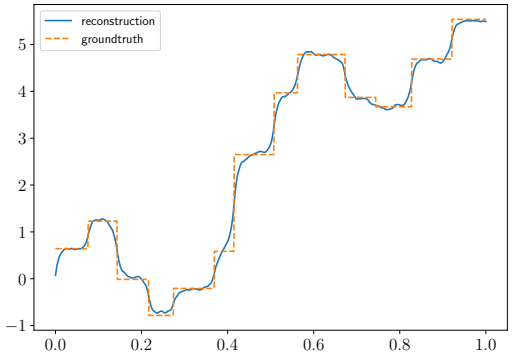


(a) Tikhonov reconstruction with $\alpha = 10^{-7}$

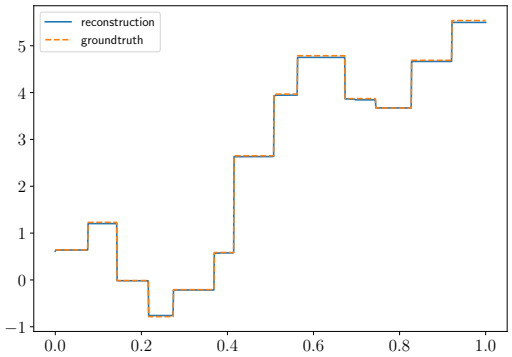
(b) Tikhonov adversarial reconstruction with $\alpha = 10^{-7}$



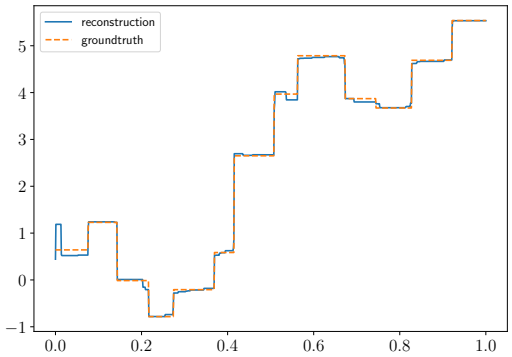
(c) Tikhonov reconstruction with $\alpha = 10^{2}$



(d) Tikhonov adversarial reconstruction with $\alpha = 10^{-7}$



(e) TV reconstruction



(f) TV adversarial reconstruction

Figure 3 Results of the total-variation-based reconstruction of 1D-signals in a compressed sensing setting.

We continue by considering the regularization via total variation [144],

$$\hat{u}_{TV} = \arg\min_{u} \frac{1}{2}\|Au - f\|_2^2 + \alpha \mathrm{TV}(u), \tag{15}$$

where $\mathrm{TV}(u)$ represents the total variation of u, and α is a regularization parameter. We employ the alternating directions method of multipliers (ADMM) for solving (15). Figure 3 (e) shows that this approach can reconstruct the ground truth nearly perfectly, exhibiting only minor deviations, possibly due to its bias or the mean-seeking behavior of the total variation regularization. While the reconstruction after an adversarial attack (Fig.3 (f)), is not altered too severely, there are some clear deviations visible. While these deviations can cause a noticeable change in the ℓ_2-norm in comparison to the reconstruction without adversarial attack, one can see that all jumps (and jump directions) of the ground truth solution are preserved, which is what one would expect from a small symmetric Bregman distance with respect to the total variation, c.f. (3).

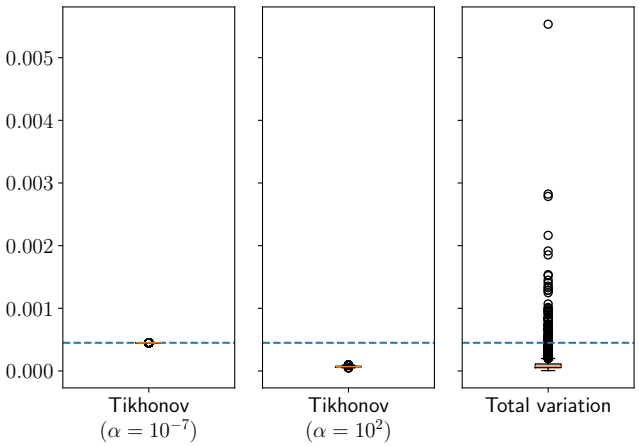


Figure 4 Results of the empirical evaluation of the bound in equation 12 for different variational reconstruction methods using Gaussian noise. Violations are artifacts of limited precision calculations.

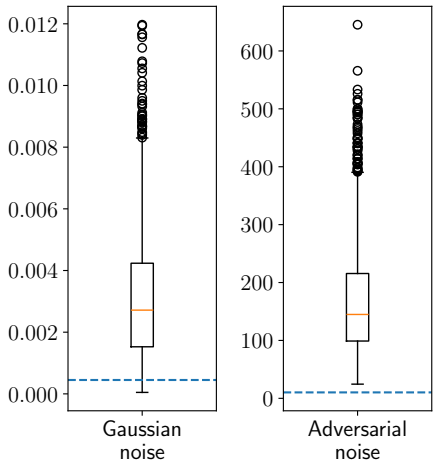


Figure 5 Results of the empirical evaluation of the bound in equation 12 for Tikhonov regularized reconstruction ($\alpha = 10^2$) applied to total variation reconstruction, exhibiting obvious violations.

We further verify empirically that the bound in equation 12 holds by calculating it over a test dataset, and show the results in figures 4 and 6, for both adversarial and white Gaussian noise and Tikhonov and total variation regularized reconstructions. In all cases, the calculated deviation lies close to or below the bound (shown here as a dotted line). The bound is dictated by the size ϵ of the projection step used during the adversarial attack, describing the maximal distance between the original and the adversarial sample. We emphasize that the results, while seemingly violating the bounds, all lie close to or below the bound, while the remaining deviations can be attributed to the limits of the numerical precision available. The dependence of equation 12 on the subgradient of the regularizer leads to different notions of stability for different reconstruction approaches. We show a concrete example in figure 5, visualizing that a total variation reconstruction tends to violate the bounds for a Tikhonov reconstruction.

While the bound, especially the Bregman distance, is not applicable for neural networks trained for image recovery, measuring robustness in terms of measurement consistency $\left\| A\mathcal{G}_\theta(f + \delta_{\text{adv}}) - f \right\|_2$ is interesting as there can be multiple solutions to ill-posed problems which satisfy a similar level of consistency, even when there is a large discrepancy between them in terms of ℓ_2 error in image space. [54] find that adversarial reconstructions are remarkably stable in terms of measurement consistency, even when there is a significant degradation in the quality of reconstructions. This, again, hints at the fact that adversarial attacks happen in subspaces corresponding to small singular values of the forward operator, such that attacks can exploit any

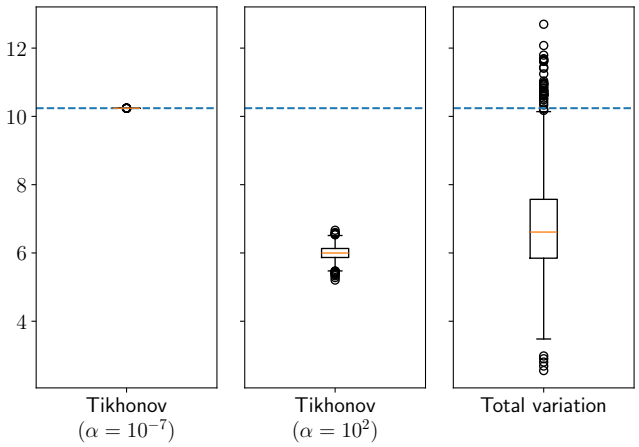


Figure 6 Results of the empirical evaluation of the bound in equation 12 for different variational reconstruction methods using adversarial noise. Violations are artifacts of limited precision calculations.

under-regularization. The work further demonstrates universal attacks are feasible and also transferable across different recovery networks showing the potential of black box attacks on image recovery.

For completeness, we also show the performance of the neural network-based methods from Genzel *et al.*[55] in terms of the aforementioned measurement consistency. All learned approaches are trained on a dataset of 8192 samples with AWGN with mean 0 and a standard deviation of 0.03 ($\mathbb{E}(\|n\|^2) = 0.6785$, with n denoting the noise). First, we show the reconstruction capability of a U-Net architecture, where the U-Net serves as a post-processor, refining the image obtained from an initial Tikhonov regularized solution, transformation,

$$\hat{u} = \mathcal{G}_{\hat{\theta}}(\hat{u}_{Tik}), \tag{16}$$

where $\mathcal{G}_\theta$ represents a U-Net model. In Figure 7 (a) and (b) the resulting reconstruction as well as the reconstruction of an adversarial example are shown. The U-Net reconstruction is comparable in quality to the total variation, while the adversarial example (obtained using the same attack and hyperparameters as in the total variation case), has a somewhat stronger influence on the network performance.

In the next approach, the reconstruction is predicted by a Tiramisu model, where the network is responsible for improving the results of a learned linear forward operator:

$$\hat{u} = \mathcal{G}_{\hat{\theta}_1}(L_{\hat{\theta}_2} f) \tag{17}$$

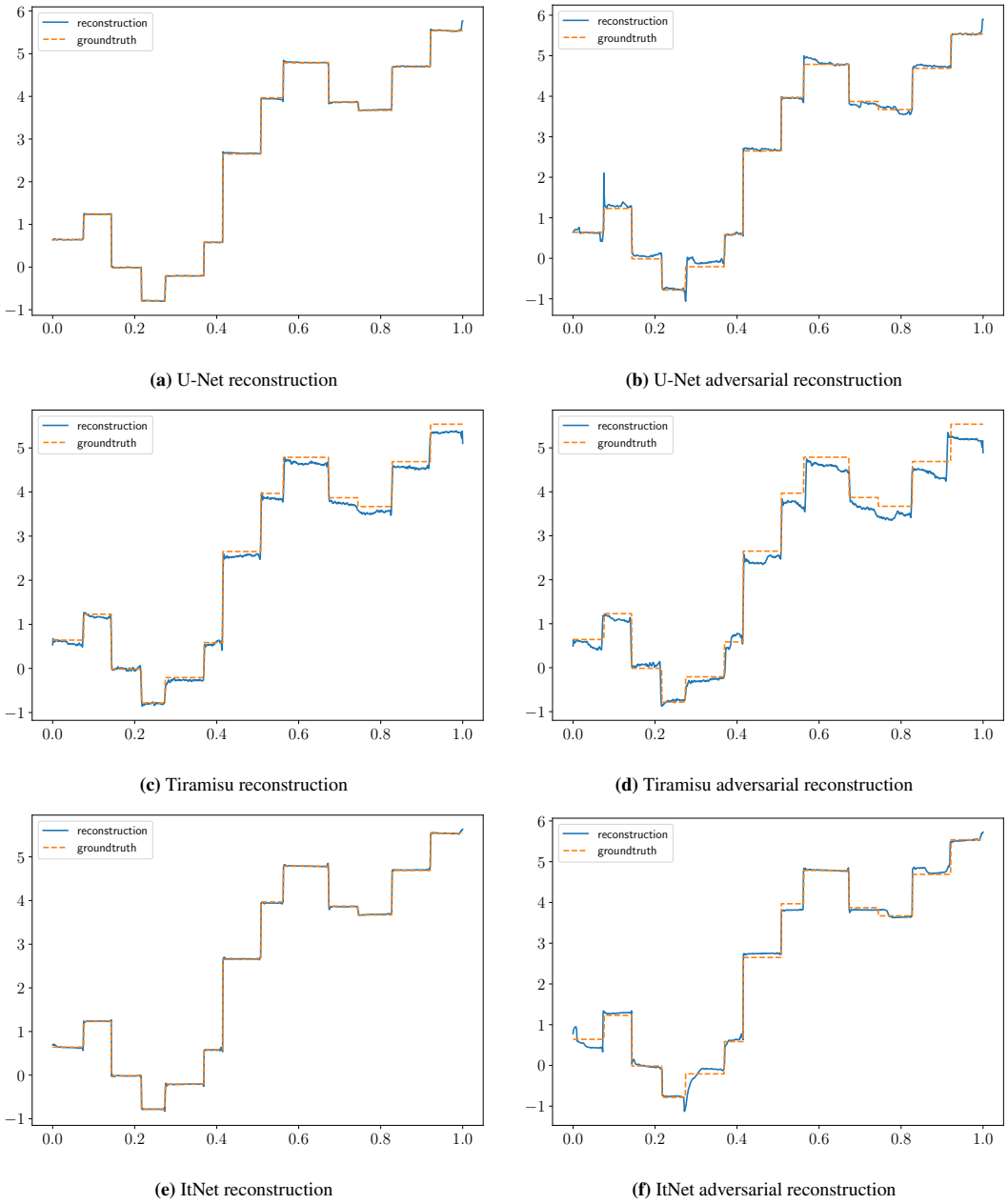
Here $\mathcal{G}_\theta$ denotes a neural network based on the Tiramisu architecture [84], while L_θ represents a learned linear transformation which is intended to substitute the Tikhonov reconstruction operator used in the U-Net approach. The achieved reconstruction quality (figure 7 (c) and (d)) in the normal and the attacked case, after comparable training efforts, is visibly worse than in any other approach. This demonstrates the influence of model information on the reconstruction process as well as the difficulty of training modern neural network architectures, due to the amount of resources required for training and fine-tuning hyperparameters.

Lastly, we reconstruct the signal using a Plug-and-play approach (ItNet), where the U-Net architecture replaces the proximal step in a proximal gradient descent approach

$$\hat{u} = u^I, \quad u^{k+1} = \mathcal{G}_{\hat{\theta}}\left(u^k - \tau A^T\left(Au^k - f\right)\right), \quad u^0 = u_{Tik} \tag{18}$$

where $\mathcal{G}_\theta$ again represents a U-Net model, applied as prox-operator in I many proximal gradient descent steps. The reconstruction quality achieved is similar to that of the post-processing U-Net (see figure 7 (e) and (f)). The attacked reconstruction exhibits artifacts comparable to a weak form of the artifacts in the U-Net case, demonstrating in another way the effects of the amount of incorporated model information.

Table 1 shows the quantitative results of our evaluation in a variety of relevant metrics for each approach with noisy measurements as well as adversarial examples. We consider the following metrics- reconstruction quality in terms of proximity to ground



(a) U-Net reconstruction

(b) U-Net adversarial reconstruction

(c) Tiramisu reconstruction

(d) Tiramisu adversarial reconstruction

(e) ItNet reconstruction

(f) ItNet adversarial reconstruction

Figure 7 Results of the learned reconstructions of 1D-signals in a compressed sensing setting.

truth for both noisy inputs $\|\hat{u} - u_{gt}\|^2$, and adversarial examples $\|\hat{u}_{adv} - u_{gt}\|^2$, measurement consistency of the reconstructions for noisy inputs $\|A\hat{u} - f\|^2$ and adversarial examples $\|A\hat{u}_{adv} - f_{adv}\|^2$. Obtaining both good reconstruction quality and measurement consistency are crucial for reconstruction algorithm. In addition to these metrics, we also evaluate $\|D\hat{u} - D\hat{u}_{adv}\|^2$ for characterizing the theoretical bound in the Tikhonov case, which also empirically shows the trade-off between robustness and reconstruction quality. Finally, we measure the deviation between the noisy measurements and adversarial inputs $\|f - f_{adv}\|^2$, and corresponding deviations in the reconstructions $\|\hat{u} - \hat{u}_{adv}\|^2$. The ratio of these measures $\|\hat{u} - \hat{u}_{adv}\|^2$ and $\|f - f_{adv}\|^2$ describes the Lipschitz-constant of reconstruction algorithm.

The best performance in each of the metrics described is marked in bold in Table 1. In terms of reconstruction quality with noisy inputs, we observe the best results with ItNet, followed by the postprocessing U-Net approach and the TV reconstruction. Tikhonov regularized reconstructions are acceptable for good choices of α, while large values of α lead to over-regularization and bad reconstruction quality. This highlights the superiority of learned approaches regarding reconstruction quality.

Table 1 Quantitative results of the different reconstruction methods. In the rows below, $\hat{u}$ denotes the reconstruction on noisy data f (which is the same for all methods). $\hat{u}_{adv}$ denotes the reconstruction on (method-specific) data f_{adv} computed via an adversarial attack on each method by trying to maximize $\|\hat{u}_{adv} - u_{gt}\|$ with one FGSM step. We report the reconstruction quality and data fidelity for both, the normal and adversarially attacked solutions, as well as the difference between the two solutions in an ℓ_2 sense, in the symmetric Bregman distances of the Tikhonov regularization, and in terms of their difference in data space, i.e., after applying the forward operator.

	Tikhonov ($\alpha = 10^{-7}$)	**Tikhonov** ($\alpha = 10^{2}$)	**TV**	**U-Net**	**Tiramisu**	**ItNet**
$\|\hat{u} - u_{gt}\|^2$	4.39	30.85	0.66	0.38	8.68	**0.34**
$\|\hat{u}_{adv} - u_{gt}\|^2$	86.37	51.45	**11.54**	25.51	46.97	23.57
$\|A\hat{u} - f\|^2$	**0**	31.44	0.84	0.46	10.68	0.38
$\|A\hat{u}_{adv} - f_{adv}\|^2$	**0**	27.73	7.67	26.29	41.42	24.9
$\|A\hat{u} - A\hat{u}_{adv}\|^2$	20.48	**6.91**	10.34	21.43	28.45	20.17
$\|D\hat{u} - D\hat{u}_{adv}\|^2$	30.41	0.03	1.62	3.85	5.47	4.9
$\|\hat{u} - \hat{u}_{adv}\|^2$	66.74	**6.52**	9.28	22.26	25.75	20.74
$\|f - f_{adv}\|^2$	20.48	20.48	20.48	20.48	20.48	20.48

To summarize: In our experiments, we analyzed common variational, model-based approaches such as TV and Tikhonov regularized reconstruction on the one hand, and on the other hand, learned approaches incorporating varying degrees of model information (following the experiments in [55]). We also found a tradeoff between reconstruction quality and the robustness of the approach to adversarial attacks. Additionally, we showed that it is possible to empirically verify the bounds given by equation 12. We would like to note, however, that this bound is regularizer-dependent, resulting in varying notions of robustness. These robustness measures are not consistent (e.g. TV solutions might violate the robustness measure of the Tikhonov regularization). Learned approaches are also harder to quantify in terms of bounds, leading to again further robustness measures.

Targeted changes

In addition to untargeted attacks which aim to degrade the quality of reconstructions, it is also interesting to evaluate the robustness to targeted attacks which trigger the reconstruction method to produce a specific, (realistic) reconstruction. [30] perform adversarial attacks to generate tiny features, which cannot be recovered well by MRI reconstruction networks, and propose adversarial training to improve the network's sensitivity to such features. [42, 121, 79] show that adversarial perturbations can alter diagnostically relevant regions. In [54] the authors demonstrate that localized adversarial attacks targeting diagnostically relevant regions can recover diagnostically different images even with extremely small perturbations such that resulting solutions still maintain a high degree of measurement consistency. While this appears to be contradictory to [53], where drastic changes that are highly inconsistent with the measurements were obtained after small adversarial changes of the input data, differences in the knowledge of the forward operator could be an explanation for the vastly different behaviors: While [54] considered the same forward operator for all instance of the data set, [53] considered variable and even unknown forward operators.

Defense

The simplest defense to deal with additive perturbations is training with additive noise. [95, 55] show that training with noise improves the adversarial robustness of reconstruction networks, with [95] showing this to be the optimal strategy for training robust denoisers. Prior works [137, 4, 34, 23] also perform adversarial training [109] or regularization to improve robustness. While adversarial training can improve robustness when the attack is (roughly) known, yet even this does not "guarantee" robustness. Further, improved robustness through adversarial training leads to reduced quality reconstruction. On the other hand, high reconstruction quality invariably comes at a cost of reduced robustness, we refer to [128, 63] for a discussion on this trade-off.

3.2 | Robustness to distribution shifts

The work [42] studies the effect of distribution shifts due to different acquisition techniques, different anatomies, and difficult-to-reconstruct samples as evaluated by a state-of-the-art method. The authors find that the performance drop on distribution shifts is

similar for trained and untrained methods (e.g. model-based approaches or untrained neural network priors). Untrained methods using hyperparameters tuned for a particular distribution do not perform as well with distribution shifts. Further theoretical analysis is performed in [153], where explicit error bounds for mismatched CNN-priors for steepest descent RED are derived. [43] propose to fix the effects of distribution shifts through a self-supervised domain adaptation method paired with inference-time training to improve the robustness to distribution shifts.

3.3 | Robustness to changes in forward measurement operator

Another desirable property of an image reconstruction algorithm is the robustness to changes in the measurement model. Classical variational approaches allow modifications, for example, changes in the noise model or modifications of the forward operator A, as they can easily be incorporated into the energy minimization by appropriate changes of the energy function. While this also holds for learned regularizers, denoising priors, or generative priors, end-to-end trained neural networks, including the model-based unrolled networks suffer from a lack of adaptivity. This means that a network trained for a specific forward operator A and noise model suffers from a significant performance drop if these are modified, and therefore have to be retrained for the new measurement model [8]. To address this, [57] propose a fine-tuning-based as well as a training-free approach to adapt trained models to variations in forward operator, whereas [62] propose training with different forward operators. [78] show that unrolled networks based on deep equilibrium models [56] are robust to changes in the measurement model.

3.4 | Robustness in recovering fine details

The authors of [8] find that different trained networks have different degrees of robustness in recovering fine details not seen in training data, ranging from the complete removal of such details to their faithful recovery. [42] observe that this ability to recover fine details is directly correlated with the overall reconstruction performance, and improving it also improves the ability to recover fine details. [8] consider fine details not belonging to the null space of the forward operator. As network hallucinations, changes and removal of details are common problems encountered in learning-based approaches, research focussing on the enforcement of data-consistent solutions has emerged, e.g. [118], where a gradient descent algorithm utilizing network-predicted descent directions is modified to converge globally to the minimizer of the data fidelity. Yet, when certain details belong to the null space of the forward operator, the problem of recovering them is rather a generative task and leads to the desire to be able to draw realistic possible sample reconstructions or to actively *exploit* solutions with certain properties. We will briefly summarize the former before providing some more details on the latter.

3.5 | Robustness for Bayesian Methods

Instead of the recovery of a single solution and the investigation of how the single predicted solution changes as the measurements change, the perspective of *Baysian inverse problems* is that the prediction of the posterior should exist, be unique, and be locally Lipschitz continuous for changing data (c.f. [99]. Consequently, the terms continuity and stability depend on a suitable choice of distance between probability measures and can yield well-posed problems far more often than in the variational setting (see [99]). In finite dimensions, this effect can be understood by relating energy minimization methods to the Bayesian setting via maximum a-posterior probability estimates. Naturally, one has to expect that $\arg\min_u \log p(u|f)$ can be discontinuous even if $p(u|f)$ depends on f continuously, and we refer to [6] for a nice example. We also refer to [6] for proving that the Lipschitz continuity of the conditional generative model transfers to a stability estimate for the posterior, and to the references therein (e.g. [65, 117, 69, 149]) for further discussions on the trade-off between the regularity and the expressivity of (conditional) generators.

4 | EXPLORABILITY

In the case where many of the singular values of the forward operator are either zero or very small in comparison to the expected noise level of the measurements, any reconstruction method has to select a solution from many possible choices. For instance, MAP estimates try to select the most probable one, and Bayesian methods allow picking multiple ones by sampling from the estimated posterior. Yet, considering the high dimensionality of the underlying space as well as the risk of complex (not well-localized) posteriors, a very large number of samples could be necessary to get a good impression of the variety of possible

(a) Low-res **(b)** **(c)** **(d)** **(e)**

Figure 8 Exploring solutions to 16× super-resolution through text using method from [52]. The text prompts used are "a high resolution photograph of a face of b) a man c) a child d) a smiling child with curly hair e) a smiling woman".

reconstructions. Among such samples, many will be similar, and - depending on the application - only a few of them might be relevant to answer an underlying question of interest. Therefore, some researchers have started focussing on the explorability of inverse reconstruction problems: To provide more control during the reconstruction process, a guiding mechanism can provide solutions that are not only data consistent but also fulfill additional criteria, such as specific semantic interpretations or particular texture properties.

For instance, Bahat *et al.* [11] address a limitation in existing super-resolution methods, which typically produce a single high-resolution output from a low-resolution input. The authors propose the composition of a Generative Adversarial Network (GAN) with a function that provably enforces the consistency with the measurements. To control the reconstruction, they introduce a control signal z, which is induced into each layer of the neural network and is designed to manipulate image gradients, thereby enabling texture modification. In [12] this work is extended to JPEG image decompression including an option to generate the control z via the optimization over an image classifier to guide the reconstruction towards a specific classification.

Following the approach of learning a classifier to guide the reconstruction, Droege *et al.* [49] proposed to explore the space of possible computed tomography reconstruction via an energy minimization technique considering

$$\min_{u\in[0,1]^N} \|Au - f\|_2^2 + \lambda H(C_\theta(u) - d) \tag{19}$$

with H denoting a suitable loss function, C_θ denoting a robust (fixed) classifier, and d being a guidance class parameter. As an application, the reconstruction of computerized tomography images of the human lung with different levels d of predicted malignancy of (localized) nodules is presented.

Recently, Gandikota and Chandramouli [52] introduced an approach for text-driven exploration of solutions to super-resolution using a pre-trained text-to-image diffusion model [139]. They ensure analytical data consistency through projection at every step in the reverse diffusion process following [167]:

$$\hat{u}_{0|t} := A^\dagger f + (I - A^\dagger A)u_{0|t}. \tag{20}$$

where $u_{0|t}$ is the MMSE estimate of the clean image at step t of the reverse diffusion process. [52] adapt this to the cascaded diffusion process at different resolutions in [139] by appropriately modifying the forward operator at each resolution. Figure 8 exemplifies a result of this approach for the task of 16× super-resolution.

5 | CONCLUSIONS

We have provided an overview of model-based, learning-based, and hybrid techniques for linear inverse problems with a focus on the robustness of point-based predictors. Our goal was to show that - although the notion of ℓ_2 stability is dominant in the machine learning literature - at least convex variational methods give rise to provable stability but in a different metric, i.e., the sum of consistency in the data space (after applying the forward operator to the reconstruction) and the symmetric Bregman distance with respect to the used regularizer. To what extent different neural network architectures and training schemes could also lead to different notions of stability, remains an interesting direction of future research. Furthermore, a clear bias-variance (or expressiveness-robustness) trade-off seems to persist. Beyond point-based estimates of solutions, the entire posterior might be difficult to sample from, such that we advertised research in the active (application-specific) exploration of different meaningful

and realistic solutions. The latter can include a control for specific classification problems in medical as well as different forms of guidance, including text, for the reconstruction of natural RGB images in challenging situations, with diffusion models being a promising recent technique for representing strong generative priors.

6 | ACKNOWLEDGEMENTS

Alexander Auras acknowledges the support of the German Research Foundation, Priority Program 2298, Project 464101190.

References

[1] J. Adler and O. Öktem, *Solving ill-posed inverse problems using iterative deep neural networks*, Inverse Problems **33** (2017), no. 12, 124007.

[2] J. Adler and O. Öktem, *Learned primal-dual reconstruction*, IEEE Transactions on Medical Imaging **37** (2018), no. 6, 1322–1332.

[3] H. K. Aggarwal, M. P. Mani, and M. Jacob, *Modl: Model-based deep learning architecture for inverse problems*, IEEE Transactions on Medical Imaging **38** (2018), no. 2, 394–405.

[4] S. Agnihotri et al., *On the unreasonable vulnerability of transformers for image restoration-and an easy fix*, *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 3707–3717.

[5] M. Aharon, M. Elad, and A. Bruckstein, *K-svd: An algorithm for designing overcomplete dictionaries for sparse representation*, IEEE Transactions on Signal Processing **54** (2006), no. 11, 4311–4322.

[6] F. Altekrüger, P. Hagemann, and G. Steidl, *Conditional generative models are provably robust: Pointwise guarantees for bayesian inverse problems*, Transactions on Machine Learning Research (2023). URL https://openreview.net/forum?id=Wcui061fxr.

[7] B. Amos, L. Xu, and J. Z. Kolter, *Input convex neural networks*, *International Conference on Machine Learning*, PMLR, 146–155.

[8] V. Antun et al., *On instabilities of deep learning in image reconstruction and the potential costs of ai*, National Academy of Sciences **117** (2020), no. 48, 30088–30095. URL https://www.pnas.org/doi/abs/10.1073/pnas.1907377117.

[9] L. Ardizzone et al., *Analyzing inverse problems with invertible neural networks*, *International Conference on Learning Representations*. URL https://openreview.net/forum?id=rJed6j0cKX.

[10] M. Asim et al., *Invertible generative models for inverse problems: mitigating representation error and dataset bias*, *International Conference on Machine Learning*, PMLR, 399–409.

[11] Y. Bahat and T. Michaeli, *Explorable super resolution*, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2716–2725.

[12] Y. Bahat and T. Michaeli, *What's in the image? explorable decoding of compressed images*, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2908–2917.

[13] S. Bai, J. Z. Kolter, and V. Koltun, *Deep equilibrium models*, Advances in Neural Information Processing Systems **32** (2019).

[14] H. Bauermeister, M. Burger, and M. Moeller, *Learning spectral regularizations for linear inverse problems*, *NeurIPS 2020 Workshop on Deep Learning and Inverse Problems*. URL https://openreview.net/forum?id=lUgF584nOSY.

[15] M. Benning and M. Burger, *Error estimates for general fidelities*, Electronic Transactions on Numerical Analysis **38** (2011), no. 44-68, 77.

[16] M. Benning and M. Burger, *Modern regularization methods for inverse problems*, Acta Numerica **27** (2018), 1–111.

[17] S. A. Bigdeli et al., *Deep mean-shift priors for image restoration*, Advances in Neural Information Processing Systems **30** (2017).

[18] P. Bohra et al., *Bayesian inversion for nonlinear imaging models using deep generative priors*, IEEE Transactions on Computational Imaging **8** (2022), 1237–1249.

[19] A. Bora et al., *Compressed sensing using generative models*, *34th International Conference on Machine Learning-Volume 70*, JMLR. org, 537–546.

[20] A. Buades, B. Coll, and J.-M. Morel, *A non-local algorithm for image denoising*, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 60–65 vol. 2, .

[21] H. C. Burger, C. J. Schuler, and S. Harmeling, *Image denoising: Can plain neural networks compete with BM3D?*, *IEEE International Conference Computer Vision and Pattern Recognition*, 2392–2399.

[22] M. Burger, E. Resmerita, and L. He, *Error estimation for bregman iterations and inverse scale space methods in image restoration*, Computing **81** (2007), no. 2, 109–135.

[23] A. Castillo et al., *Generalized real-world super-resolution through adversarial robustness*, *IEEE/CVF International Conference on Computer Vision*, 1855–1865.

[24] S. H. Chan, X. Wang, and O. A. Elgendy, *Plug-and-play admm for image restoration: Fixed-point convergence and applications*, IEEE Transactions on Computational Imaging **3** (2016), no. 1, 84–98.

[25] P. Chandramouli et al., *A generative model for generic light field reconstruction*, IEEE Transactions on Pattern Analysis & Machine Intelligence **44** (2022), no. 04, 1712–1724.

[26] R. J. Chang et al., *One network to solve them all–solving linear inverse problems using deep projection models*, *IEEE International Conference on Computer Vision*, 5888–5897.

[27] H. Chen et al., *Low-dose ct with a residual encoder-decoder convolutional neural network*, IEEE Transactions on Medical Imaging **36** (2017), no. 12, 2524–2535.

[28] Y. Chen, R. Ranftl, and T. Pock, *Insights into analysis operator learning: From patch-based sparse models to higher order mrfs*, IEEE Transactions on Image Processing **23** (2014), no. 3, 1060–1072.

[29] Y.-C. Chen et al., *Nas-dip: Learning deep image prior with neural architecture search*, *16th European Conference on Computer Vision*, Springer, 442–459.

[30] K. Cheng et al., *Addressing the false negative problem of deep learning mri reconstruction models by adversarial attacks and robust training*, *3rd Conference on Medical Imaging with Deep Learning*, PMLR. URL http://proceedings.mlr.press/v121/cheng20a.html.

[31] Z. Cheng et al., *A bayesian perspective on the deep image prior*, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5443–5451.

[32] J. Choi et al., *Ilvr: Conditioning method for denoising diffusion probabilistic models*, arXiv preprint arXiv:2108.02938 (2021).

[33] J.-H. Choi et al., *Evaluating robustness of deep image super-resolution against adversarial attacks*, *IEEE/CVF International Conference on Computer Vision*.

[34] J.-H. Choi et al., *Adversarially robust deep image super-resolution using entropy regularization*, *Asian Conference on Computer Vision*.

[35] H. Chung, B. Sim, and J. C. Ye, *Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction*, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12413–12422.

[36] H. Chung et al., *Improving diffusion models for inverse problems using manifold constraints*, A. H. Oh et al., (eds.), *Advances in Neural Information Processing Systems*. URL https://openreview.net/forum?id=nJJjv0JDJju.

[37] H. Chung et al., *Diffusion posterior sampling for general noisy inverse problems*, *International Conference on Learning Representations*. URL https://openreview.net/forum?id=OnD9zGAGT0k.

[38] R. Cohen et al., *It has potential: Gradient-driven denoisers for convergent solutions to inverse problems*, Advances in Neural Information Processing Systems **34** (2021), 18152–18164.

[39] P. L. Combettes and J.-C. Pesquet, *Lipschitz certificates for layered network structures driven by averaged activation operators*, SIAM Journal on Mathematics of Data Science **2** (2020), no. 2, 529–557.

[40] K. Dabov et al., *Image denoising by sparse 3-d transform-domain collaborative filtering*, IEEE Transactions on image processing **16** (2007), no. 8, 2080–2095.

[41] G. Daras et al., *Intermediate layer optimization for inverse problems using deep generative models*, *International Conference on Machine Learning (ICML)*.

[42] M. Z. Darestani, A. S. Chaudhari, and R. Heckel, *Measuring robustness in deep learning based compressive sensing*, *International Conference on Machine Learning*, PMLR, 2433–2444.

[43] M. Z. Darestani, J. Liu, and R. Heckel, *Test-time training can close the natural distribution shift performance gap in deep learning based compressed sensing*, K. Chaudhuri et al., (eds.), *Proceedings of the 39th International Conference on Machine Learning*, *Proceedings of Machine Learning Research*, vol. 162, PMLR, 4754–4776. URL https://proceedings.mlr.press/v162/darestani22a.html.

[44] M. Dashti and A. M. Stuart, *The bayesian approach to inverse problems*, Handbook of Uncertainty Quantification (2017), 311–428.

[45] P. del Aguila Pla, S. Neumayer, and M. Unser, *Stability of image-reconstruction algorithms*, IEEE Transactions on Computational Imaging **9** (2023), 1–12.

[46] M. Dhar, A. Grover, and S. Ermon, *Modeling sparse deviations for compressed sensing using generative models*, *International Conference on Machine Learning*, PMLR, 1214–1223.

[47] L. Dinh, J. Sohl-Dickstein, and S. Bengio, *Density estimation using real NVP*, *International Conference on Learning Representations*. URL https://openreview.net/forum?id=HkpbnH9lx.

[48] H. Dröge, T. Möllenhoff, and M. Möller, *Non-smooth energy dissipating networks*, *2022 IEEE International Conference on Image Processing (ICIP)*, 3281–3285, .

[49] H. Dröge et al., *Explorable data consistent ct reconstruction*, *33rd British Machine Vision Conference 2022, BMVC 2022, London, UK, November 21-24, 2022*, BMVA Press. URL https://bmvc2022.mpi-inf.mpg.de/0746.pdf.

[50] L. A. Feldkamp, L. C. Davis, and J. W. Kress, *Practical cone-beam algorithm*, Journal of the Optical Society of America A, Optics and image science **1** (1984), no. 6, 612–619. URL https://opg.optica.org/josaa/abstract.cfm?URI=josaa-1-6-612.

[51] B. T. Feng et al., *Score-based diffusion models as principled priors for inverse imaging*, *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 10520–10531.

[52] K. V. Gandikota and P. Chandramouli, *Exploring open domain image super-resolution through text*, 2023.

[53] K. V. Gandikota, P. Chandramouli, and M. Moeller, *On adversarial robustness of deep image deblurring*, *IEEE International Conference on Image Processing*, 3161–3165.

[54] K. V. Gandikota et al., *Evaluating adversarial robustness of low dose CT recovery*, *Medical Imaging with Deep Learning*. URL https://openreview.net/forum?id=L-N1uAxfQk1.

[55] M. Genzel, J. Macdonald, and M. März, *Solving inverse problems with deep neural networks-robustness included*, IEEE Transactions on Pattern Analysis and Machine Intelligence (2022).

[56] D. Gilton, G. Ongie, and R. Willett, *Deep equilibrium architectures for inverse problems in imaging*, IEEE Transactions on Computational Imaging **7** (2021), 1123–1133.

[57] D. Gilton, G. Ongie, and R. Willett, *Model adaptation for inverse problems in imaging*, IEEE Transactions on Computational Imaging **7** (2021), 661–674.

[58] D. Gong et al., *Learning deep gradient descent optimization for image deconvolution*, IEEE Transactions on Neural Networks and Learning Systems **31** (2020), no. 12, 5468–5482.

[59] M. González, A. Almansa, and P. Tan, *Solving inverse problems by joint posterior maximization with autoencoding prior*, SIAM Journal on Imaging Sciences **15** (2022), no. 2, 822–859. URL https://doi.org/10.1137/21M140225X.

[60] I. Goodfellow et al., *Generative adversarial nets*, *Advances in neural information processing systems*, 2672–2680.

[61] I. J. Goodfellow, J. Shlens, and C. Szegedy, *Explaining and harnessing adversarial examples*, *International Conference on Learning Representations*.

[62] A. Gossard and P. Weiss, *Training adaptive reconstruction networks for inverse problems*, arXiv preprint arXiv:2202.11342 (2022).

[63] N. M. Gottschling et al., *The troublesome kernel: why deep learning for inverse problems is typically unstable*, arXiv preprint arXiv:2001.01258 (2020).

[64] A. Goujon et al., *A neural-network-based convex regularizer for inverse problems*, IEEE Transactions on Computational Imaging (2023), 1–15.

[65] H. Gouk et al., *Regularisation of neural networks by enforcing lipschitz continuity*, Mach. Learn. **110** (2021), no. 2, 393–416. URL https://doi.org/10.1007/s10994-020-05929-w.

[66] K. Gregor and Y. LeCun, *Learning fast approximations of sparse coding*, *27th international conference on international conference on machine learning*, 399–406.

[67] C. Guo, J. S. Frank, and K. Q. Weinberger, *Low frequency adversarial perturbation*, R. P. Adams and V. Gogate, (eds.), *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, *Proceedings of Machine Learning Research*, vol. 115, PMLR, 1127–1137.

[68] H. Gupta et al., *Cnn-based projected gradient descent for consistent ct image reconstruction*, IEEE Transactions on Medical Imaging **37** (2018), no. 6, 1440–1453.

[69] P. Hagemann and S. Neumayer, *Stabilizing invertible neural networks using mixture models*, Inverse Problems **37** (2021), no. 8, 085002. URL https://dx.doi.org/10.1088/1361-6420/abe928.

[70] P. Hand, O. Leong, and V. Voroninski, *Phase retrieval under a generative prior*, Advances in Neural Information Processing Systems **31** (2018).

[71] M. Hasannasab et al., *Parseval proximal neural networks*, Journal of Fourier Analysis and Applications **26** (2020), 1–31.

[72] J. He et al., *Optimizing a parameterized plug-and-play admm for iterative low-dose ct reconstruction*, IEEE Transactions on Medical Imaging **38** (2018), no. 2, 371–382.

[73] K. He et al., *Deep residual learning for image recognition*, *IEEE conference on computer vision and pattern recognition*, 770–778.

[74] H. Heaton et al., *Wasserstein-based projections with applications to inverse problems*, SIAM Journal on Mathematics of Data Science **4** (2022), no. 2, 581–603.

[75] R. Heckel et al., *Deep decoder: Concise image representations from untrained non-convolutional networks*, *International Conference on Learning Representations*.

[76] J. Ho, A. Jain, and P. Abbeel, *Denoising diffusion probabilistic models*, *Advances in Neural Information Processing Systems*, vol. 33, Curran Associates, Inc., 6840–6851.

[77] K. Ho et al., *Neural architecture search for deep image prior*, Computers & graphics **98** (2021), 188–196.

[78] J. Hu et al., *Robustness of deep equilibrium architectures to changes in the measurement model*, *IEEEInternational Conference on Acoustics, Speech and Signal Processing*, 1–5, .

[79] Y. Huang et al., *Some investigations on robustness of deep learning in limited angle tomography*, *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 145–153.

[80] S. A. Hussein, T. Tirer, and R. Giryes, *Image-adaptive gan based reconstruction*, *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, 3121–3129.

[81] G. Jagatap and C. Hegde, *Algorithmic guarantees for inverse imaging with untrained network priors*, H. Wallach et al., (eds.), *Advances in Neural Information Processing Systems*, vol. 32, Curran Associates, Inc. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/831b342d8a83408e5960e9b0c5f31f0c-Paper.pdf.

[82] A. Jalal et al., *Instance-optimal compressed sensing via posterior sampling*, *International Conference on Machine Learning*, PMLR, 4709–4720.

[83] A. Jalal et al., *Robust compressed sensing mri with deep generative priors*, M. Ranzato et al., (eds.), *Advances in Neural Information Processing Systems*, vol. 34, Curran Associates, Inc., 14938–14954. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/7d6044e95a16761171b130dcb476a43e-Paper.pdf.

[84] S. Jégou et al., *The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation*, *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 11–19.

[85] Y. Jo, S. Yang, and S. J. Kim, *Srflow-da: Super-resolution using normalizing flow with deep convolutional block*, *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 364–372.

[86] Y. Jo et al., *Tackling the ill-posedness of super-resolution through adaptive target generation*, *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16236–16245.

[87] J. Johnson, A. Alahi, and L. Fei-Fei, *Perceptual losses for real-time style transfer and super-resolution*, *European Conference on Computer Vision*, Springer, 694–711.

[88] M. Jordan and A. G. Dimakis, *Exactly computing the local lipschitz constant of relu networks*, *Advances in Neural Information Processing Systems*, vol. 33, Curran Associates, Inc., 7344–7353. URL https://proceedings.neurips.cc/paper/2020/file/5227fa9a19dce7ba113f50a405dcaf09-Paper.pdf.

[89] S. Kabri et al., *Convergent data-driven regularizations for ct reconstruction*, arXiv preprint arXiv:2212.07786 (2022).

[90] Z. Kadkhodaie and E. P. Simoncelli, *Stochastic solutions for linear inverse problems using the prior implicit in a denoiser*, A. Beygelzimer et al., (eds.), *Advances in Neural Information Processing Systems*. URL https://openreview.net/forum?id=x5hh6N9bUUb.

[91] J. Kim, K. Lee, and K. M. Lee, *Accurate image super-resolution using very deep convolutional networks*, *IEEE Conference on Computer Vision and Pattern Recognition*, 1646–1654.

[92] D. P. Kingma and M. Welling, *Auto-encoding variational bayes*, arXiv preprint arXiv:1312.6114 (2013).

[93] E. Kobler et al., *Total deep variation for linear inverse problems*, *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7549–7558.

[94] K. Kothari et al., *Trumpets: Injective flows for inference and inverse problems*, *Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, vol. 161, PMLR, 1269–1278. URL https://proceedings.mlr.press/v161/kothari21a.html.

[95] A. Krainovic, M. Soltanolkotabi, and R. Heckel, *Learning provably robust estimators for inverse problems via jittering*, arXiv preprint arXiv:2307.12822 (2023).

[96] K. Kunisch and T. Pock, *A bilevel optimization approach for parameter learning in variational models*, SIAM Journal on Imaging Sciences **6** (2013), no. 2, 938–983. URL https://doi.org/10.1137/120882706.

[97] O. Kupyn et al., *Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better*, *IEEE/CVF International Conference on Computer Vision*.

[98] F. Latorre, V. Cevher et al., *Fast and provable admm for learning with generative priors*, *Advances in Neural Information Processing Systems*.

[99] J. Latz, *Bayesian inverse problems are usually well-posed*, SIAM Review **65** (2023), no. 3, 831–865. URL https://doi.org/10.1137/23M1556435.

[100] R. Laumont et al., *Bayesian imaging using plug & play priors: when langevin meets tweedie*, SIAM Journal on Imaging Sciences **15** (2022), no. 2, 701–737.

[101] D. Lee, J. Yoo, and J. C. Ye, *Deep residual learning for compressed sensing mri*, *2017IEEE14th International Symposium on Biomedical Imaging (ISBI 2017)*, 15–18, .

[102] H. Li et al., *Nett: Solving inverse problems with deep neural networks*, Inverse Problems **36** (2020), no. 6, 065005.

[103] H. Li et al., *Srdiff: Single image super-resolution with diffusion probabilistic models*, Neurocomputing **479** (2022), 47–59.

[104] J. Liu et al., *Image restoration using total variation regularized deep image prior*, *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Ieee, 7715–7719.

[105] Y. Liu et al., *The devil is in the upsampling: Architectural decisions made simpler for denoising with deep image prior*, *International Conference on Computer Vision*.

[106] A. Lugmayr et al., *Srflow: Learning the super-resolution space with normalizing flow*, *European Conference on Computer Vision*, Springer, 715–732.

[107] A. Lugmayr et al., *Repaint: Inpainting using denoising diffusion probabilistic models*, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11461–11471.

[108] S. Lunz, O. Öktem, and C.-B. Schönlieb, *Adversarial regularizers in inverse problems*, *32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada*, 8507–8516.

[109] A. Madry et al., *Towards deep learning models resistant to adversarial attacks*, *International Conference on Learning Representations*.

[110] J. Mairal et al., *Supervised dictionary learning*, D. Koller et al., (eds.), *Advances in Neural Information Processing Systems*, vol. 21, Curran Associates, Inc. URL https://proceedings.neurips.cc/paper_files/paper/2008/file/c0f168ce8900fa56e57789e2a2f2c9d0-Paper.pdf.

[111] M. Mardani et al., *Neural proximal gradient descent for compressive imaging*, Advances in Neural Information Processing Systems **31** (2018).

[112] M. Mardani et al., *A variational perspective on solving inverse problems with diffusion models*, arXiv preprint arXiv:2305.04391 (2023).

[113] R. Marinescu, D. Moyer, and P. Golland, *Bayesian image reconstruction using deep generative models*, *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*.

[114] G. Mataev, P. Milanfar, and M. Elad, *Deepred: Deep image prior powered by red*, *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 0–0.

[115] T. Meinhardt et al., *Learning proximal operators: Using denoising networks for regularizing inverse imaging problems*, *IEEE International Conference on Computer Vision*, 1781–1790.

[116] S. Menon et al., *Pulse: Self-supervised photo upsampling via latent space exploration of generative models*, *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2437–2445.

[117] T. Miyato et al., *Spectral normalization for generative adversarial networks*, *International Conference on Learning Representations*. URL https://openreview.net/forum?id=B1QRgziT-.

[118] M. Moeller, T. Mollenhoff, and D. Cremers, *Controlling neural networks via energy dissipation*, *IEEE/CVF International Conference on Computer Vision*, 3256–3265.

[119] V. Monga, Y. Li, and Y. C. Eldar, *Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing*, IEEE Signal Processing Magazine **38** (2021), no. 2, 18–44.

[120] A. Montanaro, D. Valsesia, and E. Magli, *Exploring the solution space of linear inverse problems with gan latent geometry*, *IEEE International Conference on Image Processing*, 1381–1385, .

[121] J. N. Morshuis et al., *Adversarial robustness of mr image reconstruction under realistic perturbations*, *International Workshop on Machine Learning for Medical Image Reconstruction*, Springer, 24–33.

[122] A. Mousavi, A. B. Patel, and R. G. Baraniuk, *A deep learning approach to structured signal recovery*, *2015 53rd annual allerton conference on communication, control, and computing (Allerton)*, IEEE, 1336–1343.

[123] S. Mukherjee et al., *Learned convex regularizers for inverse problems*, arXiv:2008.02839v2 (2020).

[124] S. Mukherjee et al., *End-to-end reconstruction meets data-driven regularization for inverse problems*, *Advances in Neural Information Processing Systems*, vol. 34, 21413–21425.

[125] S. Mukherjee et al., *Learned reconstruction methods with convergence guarantees: A survey of concepts and applications*, IEEE Signal Processing Magazine **40** (2023), no. 1, 164–182.

[126] D. Narnhofer et al., *Posterior-variance-based error quantification for inverse problems in imaging*, arXiv preprint arXiv:2212.12499 (2022).

[127] M. Noroozi, P. Chandramouli, and P. Favaro, *Motion deblurring in the wild*, *39th German Conference on Pattern Recognition*, Springer, 65–77.

[128] G. Ohayon, T. Michaeli, and M. Elad, *The perception-robustness tradeoff in deterministic image restoration*, arXiv preprint arXiv:2311.09253 (2023).

[129] G. Ohayon et al., *High perceptual quality image denoising with a posterior sampling cgan*, *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1805–1813.

[130] G. A. Padmanabha and N. Zabaras, *Solving inverse problems using conditional invertible neural networks*, Journal of Computational Physics **433** (2021), 110194.

[131] X. Pan et al., *Exploiting deep generative prior for versatile image restoration and manipulation*, IEEE Transactions on Pattern Analysis and Machine Intelligence **44** (2021), no. 11, 7474–7489.

[132] S. Peng and K. Li, *Generating unobserved alternatives: A case study through super-resolution and decompression*, OpenReview (2020). URL https://openreview.net/forum?id=_EQxgdRFUHG.

[133] S. Pinilla et al., *Improved imaging by invex regularizers with global optima guarantees*, A. H. Oh et al., (eds.), *Advances in Neural Information Processing Systems*.

[134] J. Prost et al., *Learning local regularization for variational image restoration*, *Scale Space and Variational Methods in Computer Vision: 8th International Conference, SSVM 2021, Virtual Event, May 16–20, 2021, Proceedings*, Springer, 358–370.

[135] J. Prost et al., *Inverse problem regularization with hierarchical variational autoencoders*, arXiv preprint arXiv:2303.11217 (2023).

[136] P. Putzky and M. Welling, *Recurrent inference machines for solving inverse problems*, 2017. URL https://openreview.net/forum?id=HkSOlP9lg.

[137] A. Raj, Y. Bresler, and B. Li, *Improving robustness of deep-learning-based image reconstruction*, *International Conference on Machine Learning*, *PMLR*, vol. 119, 7932–7942. URL http://proceedings.mlr.press/v119/raj20a.html.

[138] A. Raj, Y. Li, and Y. Bresler, *Gan-based projector for faster recovery with convergence guarantees in linear inverse problems*, *IEEE/CVF International Conference on Computer Vision (ICCV)*.

[139] A. Ramesh et al., *Hierarchical text-conditional image generation with clip latents*, arXiv preprint arXiv:2204.06125 (2022).

[140] E. T. Reehorst and P. Schniter, *Regularization by denoising: Clarifications and new interpretations*, IEEE Transactions on Computational Imaging **5** (2018), no. 1, 52–67.

[141] Y. Romano, M. Elad, and P. Milanfar, *The little engine that could: Regularization by denoising (red)*, SIAM Journal on Imaging Sciences **10** (2017), no. 4, 1804–1844.

[142] O. Ronneberger, P. Fischer, and T. Brox, *U-net: Convolutional networks for biomedical image segmentation*, *International Conference on Medical image computing and computer-assisted intervention*, Springer, 234–241.

[143] S. Roth and M. Black, *Fields of experts: a framework for learning image priors*, *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, 860–867 vol. 2, .

[144] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D: nonlinear phenomena **60** (1992), no. 1-4, 259–268.

[145] C. Runkel et al., *Learning posterior distributions in underdetermined inverse problems*, *International Conference on Scale Space and Variational Methods in Computer Vision*, Springer, 187–209.

[146] E. Ryu et al., *Plug-and-play methods provably converge with properly trained denoisers*, *International Conference on Machine Learning*, PMLR, 5546–5557.

[147] C. Saharia et al., *Palette: Image-to-image diffusion models*, *ACM SIGGRAPH 2022 Conference Proceedings*, 1–10.

[148] C. Saharia et al., *Image super-resolution via iterative refinement*, IEEE Transactions on Pattern Analysis and Machine Intelligence **45** (2023), no. 4, 4713–4726.

[149] A. Salmona et al., *Can push-forward generative models fit multimodal distributions?*, A. H. Oh et al., (eds.), *Advances in Neural Information Processing Systems*. URL https://openreview.net/forum?id=Tsy9WCO_fK1.

[150] U. Schmidt and S. Roth, *Shrinkage fields for effective image restoration*, *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2774–2781, .

[151] J. Schwab, S. Antholzer, and M. Haltmeier, *Deep null space learning for inverse problems: convergence analysis and rates*, Inverse Problems **35** (2019), no. 2, 025008.

[152] V. Shah and C. Hegde, *Solving linear inverse problems using gan priors: An algorithm with provable guarantees*, *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, IEEE, 4609–4613.

[153] S. Shoushtari et al., *Deep model-based architectures for inverse problems under mismatched priors*, IEEE Journal on Selected Areas in Information Theory **3** (2022), no. 3, 468–480.

[154] A. Siahkoohi et al., *Faster uncertainty quantification for inverse problems with conditional normalizing flows*, arXiv preprint arXiv:2007.07985 (2020).

[155] B. Sim et al., *Optimal transport driven cyclegan for unsupervised learning in inverse problems*, SIAM Journal on Imaging Sciences **13** (2020), no. 4, 2281–2306.

[156] H. Sommerhoff, A. Kolb, and M. Moeller, *Energy dissipation with plug-and-play priors*, *NeurIPS 2019 Workshop on Solving Inverse Problems with Deep Networks*. URL https://openreview.net/forum?id=SJxRjQncLH.

[157] J. Song et al., *Pseudoinverse-guided diffusion models for inverse problems*, *International Conference on Learning Representations*. URL https://openreview.net/forum?id=9_gsMA8MRKQ.

[158] Y. Song et al., *Score-based generative modeling through stochastic differential equations*, *International Conference on Learning Representations*.

[159] J. Sun et al., *Deep admm-net for compressive sensing mri*, Advances in neural information processing systems **29** (2016).

[160] C. Szegedy et al., *Intriguing properties of neural networks*, *International Conference on Learning Representations*.

[161] M. Terris et al., *Enhanced convergent pnp algorithms for image restoration*, *IEEE International Conference on Image Processing (ICIP)*, 1684–1688.

[162] D. Ulyanov, A. Vedaldi, and V. Lempitsky, *Deep image prior*, *IEEE Conference on Computer Vision and Pattern Recognition*, 9446–9454.

[163] D. Van Veen et al., *Compressed sensing with deep image prior and learned regularization*, arXiv preprint arXiv:1806.06438 (2018).

[164] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, *Plug-and-play priors for model based reconstruction*, *2013IEEEGlobal Conference on Signal and Information Processing*, IEEE, 945–948.

[165] A. Virmaux and K. Scaman, *Lipschitz regularity of deep neural networks: analysis and efficient estimation*, Advances in Neural Information Processing Systems **31** (2018).

[166] X. Wang et al., *Esrgan: Enhanced super-resolution generative adversarial networks*, *European Conference on Computer Vision Workshops*.

[167] Y. Wang, J. Yu, and J. Zhang, *Zero-shot image restoration using denoising diffusion null-space model*, *International Conference on Learning Representations*.

[168] L. Weng et al., *Towards fast computation of certified robustness for relu networks*, *International Conference on Machine Learning*, PMLR, 5276–5285.

[169] J. Whang, Q. Lei, and A. Dimakis, *Solving inverse problems with a flow-based noise model*, *38th International Conference on Machine Learning*, vol. 139, PMLR, 11146–11157.

[170] J. Whang, E. Lindgren, and A. Dimakis, *Composing normalizing flows for inverse problems*, *International Conference on Machine Learning*, PMLR, 11158–11169.

[171] J. Whang et al., *Deblurring via stochastic refinement*, *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16293–16303.

[172] C. Winkler et al., *Learning likelihoods with conditional normalizing flows*, 2020. URL https://openreview.net/forum?id=rJg3zxBYwH.

[173] J. Xie, L. Xu, and E. Chen, *Image denoising and inpainting with deep neural networks*, *Advances in Neural Information Processing Systems*, 341–349.

[174] R. A. Yeh et al., *Semantic image inpainting with deep generative models*, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[175] S. W. Zamir et al., *Restormer: Efficient transformer for high-resolution image restoration*, *IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

[176] J. Zhang and B. Ghanem, *Ista-net: Interpretable optimization-inspired deep network for image compressive sensing*, *IEEE Conference on Computer Vision and Pattern Recognition*.

[177] K. Zhang et al., *Learning deep CNN denoiser prior for image restoration*, *IEEE Conference on Computer Vision and Pattern Recognition*.

[178] K. Zhang et al., *Plug-and-play image restoration with deep denoiser prior*, IEEE Transactions on Pattern Analysis and Machine Intelligence (2021).

[179] H. Zhao et al., *Loss functions for image restoration with neural networks*, IEEE Transactions on Computational Imaging **3** (2016), no. 1, 47–57.

[180] B. Zhu et al., *Image reconstruction by domain-transform manifold learning*, Nature **555** (2018), no. 7697, 487–492.

[181] Y. Zhu et al., *Denoising diffusion models for plug-and-play image restoration*, *IEEE Conference on Computer Vision and Pattern Recognition Workshops (NTIRE)*.