Wasserstein Patch Prior for Image Superresolution

Johannes Hertrich, Antoine Houdard, and Claudia Redenbach

Abstract—Many recent superresolution methods are based on supervised learning. That means, that they require a large database of pairs of high- and low-resolution images as training data. However, for many applications, acquiring pairs of high and low resolution data or even imaging a large area with a high resolution is unrealistic. To overcome this problem, we introduce a Wasserstein patch prior for unsupervised superresolution of two- and three-dimensional images. In addition to the low-resolution observation, our method only requires one, possibly small, reference image which has a similar patch distribution as the high resolution ground truth. This assumption can e.g. be fulfilled when working with texture images or images of homogeneous material microstructures. The proposed regularizer penalizes the Wasserstein-2-distance of the patch distributions within the reconstruction and the reference image at different scales. We demonstrate the performance of the proposed method by applying it to two- and three-dimensional images of materials’ microstructures.

Index Terms—Superresolution, wasserstein distances, patch priors, computed tomography, FIB-SEM imaging.

I. INTRODUCTION

SUPERRESOLUTION is the task to reconstruct a high-resolution image based on a low-resolution observation.

From a mathematical point of view, image superresolution belongs to the class of inverse problems. More precisely, we assume that the high-resolution image \( x \in \mathbb{R}^n \) and the low-resolution observation \( y \in \mathbb{R}^d \) are related by

\[ y = f(x) + \eta, \]

where \( f: \mathbb{R}^n \to \mathbb{R}^d \) is a (not necessarily linear) operator and \( \eta \) is some noise. For inverse problems (including superresolution), the operator \( f \) is usually differentiable but ill-posed or not injective. Thus, reconstructing the true image \( x \) from a given observation \( y \) is difficult and we have to use some prior information. This is classically done by minimizing a functional

\[ \mathcal{J}(x) = d(x, y) + \lambda R(x), \]

where \( d \) is a data fidelity term, which ensures that the distance of \( f(x) \) and \( y \) is small, and \( R \) is a regularizer, which incorporates the prior information. The hyper parameter \( \lambda \) serves as a weighting between both terms. If \( \eta \) is Gaussian noise, then the data fidelity term is usually set to \( d(x, y) = \frac{1}{2} \| f(x) - y \|^2 \). During the last decades a wide range of regularizers was proposed for different inverse problems.

Simple regularizers, as the total variation (TV) \([39]\), can be used very generally for a large number of problems. However, they typically show a weaker performance than reconstruction methods which are adapted to specific problems and the special structure of the images under consideration. In image processing, powerful methods for denoising \([4], [25], [29]\), noise level estimation \([22], [49]\) or superresolution \([24], [41]\) are based on similarities of small \( p \times p \) patches of the images under consideration.

Here, we particularly focus on so-called patch priors \([34], [63]\). These are regularizers which depend on the patch distribution of \( x \), i.e. they are of the form

\[ R(x) = \tilde{R}(P^N_i x_{i=1}), \]

where the values \( P^N_i x \) are \( p \times p \) patches from \( x \) and the operator \( P^N \) extracts the \( i \)-th patch of \( x \).

Using the idea that a texture or texture-like image can be represented by its patch distribution, the authors of \([15], [21], [26], [30]\) proposed to synthesize textures by minimizing the Wasserstein distance of the patch distribution of the synthesized image to the patch distribution of some reference image. Here, we transfer these ideas to superresolution.

We assume that we have access to a high resolution reference image \( \hat{x} \) which has the same patch distribution as the (unknown) true high resolution image \( x \). The actual content of \( \hat{x} \) and \( x \) need not coincide. This assumption is motivated by the typical situation when imaging materials’ microstructures. The imaging setup, e.g. the detector size in micro computed tomography, poses restrictions on the size of the acquired image. Hence, choosing the voxel size or magnification requires a trade-off between resolution and size of the imaged volume. While it is possible to scan a large sample with a low resolution, only a very small sample can be imaged with a higher resolution. Generating registered pairs of high and low resolution images as training data is either practically infeasible or impossible, e.g., when using destructive imaging techniques. In contrast, scanning one small additional sample of the material at higher resolution—Many recent superresolution methods are based on supervised learning. That means, that they require a large database of pairs of high- and low-resolution images as training data. However, for many applications, acquiring pairs of high and low resolution data or even imaging a large area with a high resolution is unrealistic. To overcome this problem, we introduce a Wasserstein patch prior for unsupervised superresolution of two- and three-dimensional images. In addition to the low-resolution observation, our method only requires one, possibly small, reference image which has a similar patch distribution as the high resolution ground truth. This assumption can e.g. be fulfilled when working with texture images or images of homogeneous material microstructures. The proposed regularizer penalizes the Wasserstein-2-distance of the patch distributions within the reconstruction and the reference image at different scales. We demonstrate the performance of the proposed method by applying it to two- and three-dimensional images of materials’ microstructures.

Index Terms—Superresolution, wasserstein distances, patch priors, computed tomography, FIB-SEM imaging.

I. INTRODUCTION

SUPERRESOLUTION is the task to reconstruct a high-resolution image based on a low-resolution observation.

From a mathematical point of view, image superresolution belongs to the class of inverse problems. More precisely, we assume that the high-resolution image \( x \in \mathbb{R}^n \) and the low-resolution observation \( y \in \mathbb{R}^d \) are related by

\[ y = f(x) + \eta, \]

where \( f: \mathbb{R}^n \to \mathbb{R}^d \) is a (not necessarily linear) operator and \( \eta \) is some noise. For inverse problems (including superresolution), the operator \( f \) is usually differentiable but ill-posed or not injective. Thus, reconstructing the true image \( x \) from a given observation \( y \) is difficult and we have to use some prior information. This is classically done by minimizing a functional

\[ \mathcal{J}(x) = d(x, y) + \lambda R(x), \]

where \( d \) is a data fidelity term, which ensures that the distance of \( f(x) \) and \( y \) is small, and \( R \) is a regularizer, which incorporates the prior information. The hyper parameter \( \lambda \) serves as a weighting between both terms. If \( \eta \) is Gaussian noise, then the data fidelity term is usually set to \( d(x, y) = \frac{1}{2} \| f(x) - y \|^2 \). During the last decades a wide range of regularizers was proposed for different inverse problems.

Simple regularizers, as the total variation (TV) \([39]\), can be used very generally for a large number of problems. However, they typically show a weaker performance than reconstruction methods which are adapted to specific problems and the special structure of the images under consideration. In image processing, powerful methods for denoising \([4], [25], [29]\), noise level estimation \([22], [49]\) or superresolution \([24], [41]\) are based on similarities of small \( p \times p \) patches of the images under consideration.

Here, we particularly focus on so-called patch priors \([34], [63]\). These are regularizers which depend on the patch distribution of \( x \), i.e. they are of the form

\[ R(x) = \tilde{R}(P^N_i x_{i=1}), \]

where the values \( P^N_i x \) are \( p \times p \) patches from \( x \) and the operator \( P^N \) extracts the \( i \)-th patch of \( x \).

Using the idea that a texture or texture-like image can be represented by its patch distribution, the authors of \([15], [21], [26], [30]\) proposed to synthesize textures by minimizing the Wasserstein distance of the patch distribution of the synthesized image to the patch distribution of some reference image. Here, we transfer these ideas to superresolution.

We assume that we have access to a high resolution reference image \( \hat{x} \) which has the same patch distribution as the (unknown) true high resolution image \( x \). The actual content of \( \hat{x} \) and \( x \) need not coincide. This assumption is motivated by the typical situation when imaging materials’ microstructures. The imaging setup, e.g. the detector size in micro computed tomography, poses restrictions on the size of the acquired image. Hence, choosing the voxel size or magnification requires a trade-off between resolution and size of the imaged volume. While it is possible to scan a large sample with a low resolution, only a very small sample can be imaged with a higher resolution. Generating registered pairs of high and low resolution images as training data is either practically infeasible or impossible, e.g., when using destructive imaging techniques. In contrast, scanning one small additional sample of the material at higher resolution...
resolution is possible. For spatially homogeneous materials, the patch distribution of this additional sample can be expected to be similar to that of the desired high resolution image.

Nowadays, many state-of-the-art superresolution methods are based on machine learning with deep neural networks, see e.g. [14], [38], [40], [55], [59], [62]. However, most of these methods are trained on registered pairs of high- and low-resolution images. As outlined above, such data is often not available in practice.

There are only very few neural network based approaches working with similar assumptions as our approach. Examples are the deep image prior [50] and Plug-and-Play (PnP) methods [47], [54]. We compare the proposed method with both in Section III. One possible extension of the Plug and Play framework is the MACE algorithm [5], which is closely related to the parallel Douglas-Rachford splitting [3], [10]. The MACE algorithm in connection with a multi-resolution data fusion was further investigated and applied for superresolution in [36]. Finally, the authors of [46] propose an unsupervised superresolution method based on internal learning. However, their main assumption is that the images admit a self-similarity across different resolutions. While this assumption is often met by two-dimensional natural images, it is typically violated when considering three-dimensional materials’ microstructures.

In summary, our paper provides the following contributions.

– Building upon [26], we propose to use a Wasserstein patch prior for unsupervised reconstructions in inverse problems. This prior penalizes the Wasserstein-2 distance of the patch distributions of the reconstruction and the reference image.
– We apply the Wasserstein patch prior for unsupervised superresolution of real-world material images.
– In particular, we extend the Wasserstein patch prior to three-dimensional images, where not many competing methods exist.

An implementation of the two-dimensional examples is available online.¹

The paper is structured as follows: In Section II, we introduce the proposed regularizer and describe how the functional consisting of data fidelity term and regularizer can be minimized. We provide numerical examples for real and synthetic data in 2D and 3D in Section III. Conclusions are drawn in Section IV.

II. A WASSERSTEIN PATCH PRIOR

In this section, we describe the proposed superresolution method. First, in Subsection II-A, we introduce the Wasserstein patch prior as a regularizer and derive our objective functional. In Subsection II-B we focus on the minimization of the objective functional. Finally, in Subsection II-C we discuss an approach for making the method robust to small deviations between the patch distributions of the reference and the ground truth image.

A. Proposed Regularizer

Let \( P_i : \mathbb{R}^d \rightarrow \mathbb{R}^p \) be the operator which extracts the \( i \)-th patch of size \( p \) from an image of size \( d_1 \times d_2 \) with \( d = d_1 \times d_2 \).

Then, we define the empirical patch distribution \( \mu_x \) of an image \( x \) by the discrete measure

\[
\mu_x = \frac{1}{N} \sum_{i=1}^{N} \delta_{P_i x}.
\]

Here, the set \( \{ P_i x : i = 1, \ldots, N \} \) is either the whole set of all overlapping patches within the image \( x \) or a random subset thereof. The main assumption of the following proposed regularizer is that the images \( x \) and \( \tilde{x} \) are similar in the sense that they have a similar empirical patch distribution.

To measure the distance of the patch distribution within an image \( x \) and its reference image \( \tilde{x} \), we use the Wasserstein-2 distance. For two discrete measures

\[
\mu = \frac{1}{N} \sum_{i=1}^{N} \delta_{p_i}, \quad \tilde{\mu} = \frac{1}{M} \sum_{j=1}^{M} \delta_{\tilde{p}_j}, \quad p_i, \tilde{p}_j \in \mathbb{R}^p,
\]

the Wasserstein-2 distance is defined by

\[
W_2^2(\mu, \tilde{\mu}) = \min_{\Pi(\mu, \tilde{\mu})} \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{1}{2} \| p_i - \tilde{p}_j \|^2 \pi_{i,j},
\]

where

\[
\Pi(\mu, \tilde{\mu}) = \left\{ (\pi_{i,j})_{i,j} : \sum_{i=1}^{N} \pi_{i,j} = \frac{1}{M}, \quad \sum_{j=1}^{M} \pi_{i,j} = \frac{1}{N} \right\}.
\]

Engineering materials frequently show a multiscale microstructure with components of significantly different size. To incorporate image information across these scales in the regularizer, we consider several downscaled versions of the original image \( x \). More precisely, let \( A \) be some downsampling operator. Then, we define sequences of downsampled versions of \( x \) and \( \tilde{x} \) via \( x_1 = x, x_l = A x_{l-1}, \tilde{x}_1 = \tilde{x} \) and \( \tilde{x}_l = A \tilde{x}_{l-1} \) for \( l = 2, \ldots, L \). Our regularizer measures the Wasserstein distance between patches of \( x_l \) and \( \tilde{x}_l \) at each layer \( l \), i.e., we consider

\[
R(x) = \sum_{l=1}^{L} W_2^2(\mu_{x_l}, \mu_{\tilde{x}_l}), \quad (1)
\]

Fig. 1 illustrates this decomposition and the considered Wasserstein patch prior.

Finally, we propose to minimize the functional

\[
\mathcal{J}(x) = d(x, y) + \lambda R(x).
\]

Within our numerical examples, we focus on the data fidelity term \( d(x, y) = \frac{1}{2} \| f(x) - y \|^2 \), such that \( \mathcal{J} \) can be rewritten as

\[
\mathcal{J}(x) = \frac{1}{2} \| f(x) - y \|^2 + \lambda \sum_{l=1}^{L} W_2^2(\mu_{x_l}, \mu_{\tilde{x}_l}). \quad (2)
\]

For the rest of the paper, we will refer to this regularizer as the Wasserstein Patch Prior (WPP).

Note that the authors of [26] propose to find a local minimum of the regularizer \( R \) to generate textures based on a reference texture \( \tilde{x} \) for a random initialization \( x \). Using the same definitions, the idea of Wasserstein patch priors could also be transferred to other optimal transport costs.

¹[Online]. Available: https://github.com/joherwich/Wasserstein_Patch_Prior
Remark 1: To achieve computational efficiency, we do not use all patches within the measure \( \mu_x \) arising from the reference image but only a random subset of 10000 patches. For the measure \( \mu_x \) arising from our reconstruction, we use all patches.

B. Minimization of the Objective Function

For the minimization of \( J \), we assume that \( f \) is differentiable and use a gradient descent scheme based on the minimization of \( R \). Here, the main challenge is to compute the derivative of \( W_2^2 \left( \frac{1}{N} \sum_{i=1}^{N} \delta_{P_i, x}, \frac{1}{N} \sum_{i=1}^{N} \delta_{P_i, \tilde{x}} \right) \) with respect to \( x \). For this, we follow the lines of [26].

Using the dual formulation of the Wasserstein-2 distance, we obtain that

\[
\nabla_x W_2^2 \left( \frac{1}{N} \sum_{i=1}^{N} \delta_{P_i, x}, \frac{1}{N} \sum_{i=1}^{N} \delta_{P_i, \tilde{x}} \right) = \arg \max_{\psi \in \mathbb{R}^N} F(\psi, x),
\]

where

\[
F(\psi, x) = \frac{1}{N} \sum_{i=1}^{N} \psi(x^i) + \frac{1}{N} \sum_{i=1}^{N} \psi_i,
\]

and \( \psi^c(x) = \min_j \{ x_j - P_j \tilde{x} \} \) is the c-transform of \( \psi \). Now, the following theorem was shown in [26], Theorem 2.

**Theorem 2**: Consider \( x_0 \) such that \( x \mapsto W_2^2 \left( \frac{1}{N} \sum_{i=1}^{N} \delta_{P_i, x}, \frac{1}{N} \sum_{i=1}^{N} \delta_{P_i, \tilde{x}} \right) \) and \( x \mapsto F(\psi^*, x) \) are differentiable at \( x_0 \) with \( \psi^* \in \arg \max_{\psi} F(\psi, x_0) \). Then it holds

\[
\nabla_x W_2^2 \left( \frac{1}{N} \sum_{i=1}^{N} \delta_{P_i, x_0}, \frac{1}{N} \sum_{i=1}^{N} \delta_{P_i, \tilde{x}} \right) = \nabla_x F(\psi^*, x_0).
\]

Note that for almost every \( x \) the set of minimizers

\[
\sigma_\psi(i) = \arg \min_{j} (\frac{1}{2} \| P_i x - P_j \tilde{x} \|)^2 - \psi_j
\]

is single-valued. Hence,

\[
\nabla_x \psi^c(P_i x) = \frac{1}{2} \| P_i x - P_j \tilde{x} \|^2
\]

for almost every \( x \). Note that for computing this gradient it is required to compute \( \psi^* \). For doing so, the authors of [26] suggest to use a (stochastic) gradient ascent as proposed in [18].

Summarized, the gradient of \( J(x) \) in (2) can be computed by Algorithm 1.

**Remark 3 (Relation to Wasserstein GANs)**: At first glance, the minimization of the dual form (3) of the Wasserstein distance looks similar to the loss function of (Wasserstein) GANs [2],...
Algorithm 1: Gradient Computation.

for $l = 1, \ldots, L$ do

Set

$$F_l(\psi, x_l) := \frac{1}{N_l} \sum_{i=1}^{N_l} \psi^c(P, x_l) + \frac{1}{N_l} \sum_{j=1}^{N_l} \psi_j$$

Use stochastic gradient ascent to compute $\psi^l \in \arg \max_{\psi} F_1(\psi, x_l)$.

end for

Compute $\nabla J(x)$ as

$$\nabla J(x) = \nabla f(x)(f(x) - y) + \lambda \sum_{l=1}^{L} (A^l)^T \left( \frac{1}{N_l} \sum_{i=1}^{N_l} P_i^T P_{x_l} - P_i^T P_{\tilde{x}_l(i)}\tilde{x}_l \right),$$

where $\nabla f(x)$ is the Jacobian of $f$ at $x$.

[19]. In particular, the minimization of the WPP shares the minimax structure with GAN-type loss functions via the right side of (3). However, there are some fundamental methodological differences:

- A Wasserstein GAN aims at minimizing the Wasserstein-1 distance of the probability distribution induced by some generator to the probability distribution defined by some training data. In particular, a Wasserstein GAN compares probability distributions on the space of $n \times m$ images, which might be very high-dimensional. Furthermore, Wasserstein GANs for inverse problems are often trained in a supervised way [1], [45], [56], such that the access to a large database of ground truth images is required. In contrast, the proposed regularizer (1) considers the Wasserstein distance of the patch distributions of the reconstruction and the reference image. These are measures on the space of $p \times p$ patches, which has a much smaller dimension (in Section III we use $p = 6$). These patches are extracted from the reconstruction and one single reference image. Consequently, our proposed method is unsupervised and only needs one reference image.

- The optimization with respect to the discriminator of a Wasserstein GAN is a highly non-concave maximization problem with often millions of parameters. Thus, there is no chance to end up even close to the global optimum. Consequently, it is not clear how well a Wasserstein GAN really approximates the Wasserstein distance. Indeed, the authors of a recent paper [48] found that Wasserstein GANs fail to approximate the Wasserstein distance and that the use of a better approximation yields much weaker results. In contrast, the maximization with respect to $\psi$ in (3) is a concave optimization problem with much fewer parameters. Thus, any local maximizer is a global maximizer such that Algorithm 1 is really a gradient descent on the Wasserstein distance rather than a training routine of a GAN.

Consequently, Wasserstein GANs are only very loosely related to the WPP.

C. Compensation of Slightly Different Patch Distributions

So far, we assumed that the patch distributions in the (unknown) ground truth $x$ and the reference image $\tilde{x}$ are identical. This assumption is not realistic in practice. To compensate for slight differences of $\mu_x$ and $\mu_{\tilde{x}}$, we introduce the new operator $g : \mathbb{R}^{m+2q+n+2q} \rightarrow \mathbb{R}^d$ defined by $g(x) = f(Cx)$, where $C : \mathbb{R}^{m+2q+n+2q} \rightarrow \mathbb{R}^{m+n}$ crops the middle $m \times n$ pixels from $x \in \mathbb{R}^{m+2q+n+2d}$. Then, we minimize

$$\mathcal{I}(x) = \frac{1}{2} \|g(x) - y\|^2 + \lambda \sum_{l=1}^{L} W_2^2(\mu_{x_l}, \mu_{\tilde{x}_l}).$$

Finally, our reconstruction is given by

$$\hat{x} = Cz, \quad \text{with} \quad z \in \arg \min_{z} \mathcal{I}(x).$$

Note that the data fidelity term $\frac{1}{2} \|g(x) - y\|^2$ is by definition not affected by the boundary of size $q$ in $x$. On the other hand, the boundary of $x$ influences the patch distribution of $x$ such that it can compensate small differences between the patch distributions of the reference image $\tilde{x}$ and the unknown ground truth $x$.

III. NUMERICAL RESULTS FOR SUPERRESOLUTION

In this section, we apply the Wasserstein patch regularization for the problem of superresolution of two- and three-dimensional images.

Many superresolution approaches found in the literature focus on natural 2D images. However, problems due to low resolution may also arise in processing and analysis of (3D) images of materials’ microstructures. For instance, selecting the voxel size in micro computed tomography requires a trade-off between representativity of the imaged volume and ability to reconstruct fine structure details such as thin fibres or cracks [43]. In serial sectioning imaging techniques such as scanning electron microscopy coupled with focused ion beam milling (FIB-SEM), limiting the processing time may require to choose the voxel size larger than desired, in particular in the slicing direction. The resulting anisotropic voxels induce an artificial anisotropy in the data which influences the results of a quantitative image analysis [37]. In these cases, application of a superresolution approach may help to overcome limitations due to the imaging setup. However, the reconstruction quality must be validated critically to ensure that geometric or material characteristics obtained from the reconstructed image are reliable.
For superresolution, the forward operator \( f \) usually is a composition of a blur operator and a downsampling operator. We demonstrate the performance of our approach by using images of materials’ microstructures obtained by synchrotron micro-computed tomography (\( \mu \text{CT} \)). Additionally, we consider synthetic images that were obtained by simulating the process of FIB-SEM imaging. Finally, the approach is applied to a real FIB-SEM image stack.

For evaluating the quality of our results we use the following quality measures:

- **Peak-signal to noise ratio (PSNR):** For two images \( x \) and \( y \) on \([0, 1]^{m \times n}\) it is defined as

\[
\text{PSNR}(x, y) = -10 \log_{10}(\frac{1}{mn} \|x - y\|^2).
\]

- **Blur effect [11]:** This metric measure the sharpness of our results and is based on comparing the variation within an input image \( x \) with a blurred version \( x_{\text{blur}} \). Sharp images \( x \) will lose much variation through the blur, while the loss will be small for blurred \( x \). The blur effect describes the percentage of variation which is retained after the blurring process. It takes values in \([0, 1]\), where a small blur effect indicates that \( x \) is very sharp while a large blur effect means that \( x \) is very blurry.

- **Feature-based similarity index (FSIM) [60]:** The basic idea of the FSIM is not to compare the raw pixel values, but to compare the similarity of certain feature maps extracted from the images. We use the same feature maps as the in the original paper [60].

- **Learned perceptual image patch similarity (LPIPS) [61]:** LPIPS measures the perceptual similarity of our results and the ground truth. The basic idea is to compare the feature maps extracted from some deep neural network that is trained for some classical imaging task which is not necessarily related to our original problem. In our numerical examples, we use the AlexNet [28] trained for classification of the ImageNet data set [13]. A small value of LPIPS indicates a high perceptual similarity. Note that LPIPS was originally proposed and implemented for 2D images. Although LPIPS could probably be extended to 3D images (see [44] for some work in this direction), this is not within the scope of our paper. For the 3D data, we will use the mean of the LPIPS values over all slices of the image.

We compare our results with the bi- and tricubic interpolation, \( L^2 - TV \), the expected patch log likelihood (EPLL) using a domain-specific patch distribution, the deep image prior (DIP) and plug-and-play FBS (PnP-DRUNet). Within the PnP-DRUNet we use once a denoiser trained on natural images and once a denoiser trained on 500 domain-specific images (PnP-DS), i.e., high-resolution images from the corresponding material. Note that the input assumptions of the used methods differ slightly. While the interpolation, \( L^2 - TV \), DIP and PnP-DRUNet require only the knowledge of the operator, the WPP and EPLL additionally need one high-resolution reference image. Finally,

\[\text{PnP-DS requires additionally extensive knowledge about the considered image domain represented by a large data base of high-resolution images of the corresponding material, which is usually not available in practice. A short description of these methods can be found in Appendix A.}

All numerical examples are implemented in Python and PyTorch and are based on the code of [26].\[^{3}\] An implementation of the 2D examples is provided online.\[^{4}\]

### A. Synchrotron Computed Tomography Data

First, we consider material data which was also used in [24]. Here, a series of multi-scale 3D images has been acquired by synchrotron micro-computed tomography at the SLS beamline TOMCAT. Samples of two materials were selected to provide 3D images of size \( 2560 \times 2560 \times 2160 \) having different levels of complexity:

- The first one is a composite ("SiC Diamonds") obtained by microwave sintering of silicon and diamonds, see [53].
- The second one is a sample of Fontainebleau sandstone ("FS"), a rather homogeneous natural rock that is commonly used in the oil industry for flow experiments.

In the first step, we consider only 2D sectional images of the CT scans. They have the size \( 600 \times 600 \) pixels with a pixel spacing of \( 1.625 \mu m \). Some ground truth and the reference images in our two-dimensional experiments are given in Fig. 2. Since we require that the forward operator \( f \) is known, we use the real material images as high resolution ground-truth and as reference

\[\text{[Online]. Available: https://github.com/ahoudard/wgenpatex}\]

\[\text{[Online]. Available: https://github.com/johertrich/Wasserstein_Patch_Prior}\]
TABLE I

| Method | PSNR | FSIM | L2-TV | EPLL | DIP | PnP-DRUnet | PnP-DS | WPP |
|--------|------|------|-------|------|-----|------------|--------|-----|
| SiC    | 25.52| 27.76| 27.66 | 27.36| 27.75| 28.24      | 27.72  |     |
|        | 0.8836 | 0.9142 | 0.9213 | 0.9145 | 0.8753 | 0.9125 | 0.9236 |     |
|        | 0.5445 | 0.3938 | 0.4109 | 0.3805 | 0.4425 | 0.3941 | 0.3590 |     |
|        | 0.4126 | 0.2008 | 0.2102 | 0.2013 | 0.3239 | 0.2380 | 0.1595 |     |
|        | 0s    | 2s   | 10s*  | 120s | 100s* | 100s* | 600s   |     |
| FS     | 29.78 | 31.42 | 31.44 | 30.95 | 31.21 | 32.18 | 31.18 |     |
|        | 0.8577 | 0.8576 | 0.9093 | 0.8614 | 0.7842 | 0.9029 | 0.9472 |     |
|        | 0.1903 | 0.4010 | 0.3671 | 0.3815 | 0.4811 | 0.3641 | 0.3200 |     |
|        | 0.3521 | 0.2604 | 0.1969 | 0.2537 | 0.3601 | 0.2595 | 0.1542 |     |
|        | 0s    | 2s   | 10s*  | 120s | 100s* | 100s* | 100s* |     |

Required reference images | 0 | 0 | 1 | 0 | 0 | 500 | 1 |

Fig. 3. Reconstruction of the high resolution image “SiC Diamonds” using different methods. Top: Full image, middle and bottom: zoomed-in parts. We marked the zoomed-in parts by a white box in the full images. WPP and PnP-DS yield much sharper and visually better results than the comparison methods. However, PnP-DS requires the knowledge of a large data base of domain-specific images, while the WPP only needs a single reference image.
HERTRICH et al.: WASSERSTEIN PATCH PRIOR FOR IMAGE SUPERRESOLUTION

Fig. 4. Reconstruction of the high resolution image “Fontainebleau Sandstone” using different methods. Top: Full image, middle and bottom: zoomed-in parts. From the visual impression, the reconstruction with the WPP and PnP-DS are much more realistic than the reconstructions of the comparison methods. However, PnP-DS requires the knowledge of a large data base of domain-specific images, while the WPP only needs a single reference image.

Fig. 5. FRC to measure the effective resolution of the different reconstruction of the high resolution images. It turns out that WPP and PnP-DS have a significantly higher effective resolution than the comparisons. Here, the effective resolution for PnP-DS is slightly higher than for WPP. However, this is probably to be explained by the more extensive knowledge about the image domain, which is required for PnP-DS. We used the code from [52] to generate this figure.

measures the effective resolution of an image by computing the correlation of the Fourier spectra of two images along concentric rings. Then, the FRC represents the function which maps the radius of the ring to the corresponding correlation. Now, the effective image resolution is given by the intersection of this function with a given threshold function. The FRC curves for our reconstructions are given in Fig. 5. To keep the figure clear, we restrict ourselves to the reconstructions of bicubic interpolation, WPP, DIP, PnP-DRUNet and PnP-DS. We observe that WPP and PnP-DS have a significantly higher effective resolution than the comparison methods. The PnP-DS has a slightly higher effective resolution than WPP which can be explained by the more extensive knowledge of the image domain.

To assess the performance of the method in 3D, we redo the experiments with 3D images of size $176 \times 176 \times 176$. The forward operator is the same as for the 2D images, which results in low resolution observations of size $41 \times 41 \times 41$. Due to the computational effort, we reduce the patch size to $p = 4$ and do not add a boundary as in Section II-C. The parameter $\lambda$ is set to $\lambda = 10000/p^3$ for the SiC image and to $\lambda = 12000/p^3$ for the FS image. The resulting error measures for the 3D images are given in Table II. Similar to the two dimensional case, the Wasserstein patch prior reconstruction is significantly better than $L^2$-TV and the tricubic interpolation.

Table II: PSNR, BLUR EFFECT AND LPIPS VALUE OF THE HIGH-RESOLUTION RECONSTRUCTION OF THE 3D-MATERIAL DATA USING DIFFERENT METHODS. WPP IS THE BEST METHOD OVER ALL QUALITY MEASURES.

| Material | PSNR | Blur effect | LPIPS | LPIPS | WPP |
|----------|------|-------------|-------|-------|-----|
| SiC      | 23.06| 0.6156      | 0.4756| 0.3840|     |
| FS       | 27.16| 0.6545      | 0.4649| 0.3452|     |

FS and SiC Diamonds are given in Table II. Similar to the two dimensional case, the Wasserstein patch prior reconstruction is significantly better than $L^2$-TV and the tricubic interpolation.

B. Synthetic FIB-SEM Images

As a second imaging technique, we consider focused ion beam scanning electron microscopy imaging (FIB-SEM). We first use synthetic image data which are obtained by simulating...
the FIB-SEM imaging process by using the protocol described in [35]. The imaged geometry is a realization of a Boolean model of spheres of identical radius [9]. Images at several resolutions ranging from (cubical) voxel sizes of 3 nm to 18 nm were generated. For more details on the image data we refer to [37]. As ground truth image, we will use the images with 6 nm voxel size. The low resolution images are given by the versions with spacings 9 nm, 12 nm, 15 nm and 18 nm. To reduce the noise on the images, we preprocess each image by a $3 \times 3 \times 3$ median filter. Images at the different resolutions are registered, but the forward operator $f$ is unknown and its formulation is not straightforward. Therefore, we propose to estimate $f$ in a first step based on one registered pair of a high-resolution and a low-resolution image. Afterwards, we minimize the functional (2) using the estimation of $f$. The detailed description of the estimation process for the forward operator $f$ can be found in Appendix B.

We set the number of scales to $L = 2$ and use a patch size of $p = 4$. The parameter $\lambda$ is set to $\lambda = 9000$ for low-resolution voxel spacing of 9 nm, $\lambda = 3000$ for 12 nm, $\lambda = 2000$ for 15 nm and $\lambda = 1000$ for 18 nm. As a comparison, we use $L^2$-TV using the operator $\hat{f}$ as forward model. The resulting error measures are given in Table III. Slices of the reconstructions using the WPP and $L^2$-TV are shown in Fig. 6. Again, the WPP yields the best reconstruction both visually and in terms of the quality measures. In particular for larger magnification factors, the WPP recovers the structure of the high-resolution image much better than the other methods.

C. Real FIB-SEM Images

As a real data example we consider a pair of FIB-SEM images of a porous zirconium dioxide filtration membrane produced by spin coating [6]. The sample is imaged using a Carl Zeiss
Hertrich et al.: Wasserstein Patch Prior for Image Superresolution

Fig. 7. Slices of the reconstructions of the high-resolution 3D-FIB-SEM-image (spacing 10 nm) for the real data. Top: xy section (SEM imaging plane), bottom: yz section. We can see that the WPP significantly reduces the blur within the image.

Crossbeam NVision 40 Field Emission Scanning Electron Microscope (FE-SEM) with integrated advanced tomography package Atlas 5 3-D-Tomography. To reduce charging during SEM imaging of the poorly electrically conductive ZrO$_2$, the sample is sputtered with gold. An additional 1-2 μm platinum layer is applied locally by FIB deposition to smooth the rough porous surface.

The high resolution image contains $649 \times 452 \times 161$ voxels with cubic voxels of edge length 10 nm. The size of the low resolution image is $355 \times 272 \times 52$ voxels with a voxel edge length of 20 nm.

a) Forward Operator: As in the previous subsection, the forward operator is unknown. As the synthetic data from Subsection III-B simulates the FIB-SEM imaging process, we will use the forward operator from Subsection III-B for superresolution from the 12 nm image to the 6 nm image.

b) Results: We set the number of scales to $L = 2$, the patch-size to $p = 4$ and the parameter $\lambda$ to $\lambda = 3000$. Single slices of the resulting reconstruction using the Wasserstein patch prior and the tricubic interpolation are shown in Fig. 7. The WPP reconstruction is sharper than the interpolation and enhances the structures from the observation. Indeed, the blur effect of the interpolation is $0.5107$, while the blur effect of the Wasserstein patch prior reconstruction is $0.4344$. As no ground truth image is available, the other quality measures cannot be applied in this example.

IV. Conclusion

In this paper, we introduced a Wasserstein-2 patch prior for image superresolution, which penalizes the Wasserstein-2 distance of the patch distribution in the reconstruction to the patch distribution in some reference image. The minimization of the arising objective functional can be done via a gradient descent scheme based on [26]. We demonstrated the performance of the new prior on 2D and 3D material images. In particular, we have shown that the method is applicable to real-world superresolution problems. In case of the FIB-SEM images, the forward operator estimated from the synthetic data also yields reasonable results for the real data. An investigation of the generality and robustness of this approach will be subject of future research.

Appendix

A. Comparison to Established Methods

We compare our results with the following methods:

- Bi- and tricubic interpolation [27], [31]: These interpolations are based on the local approximation of the image by polynomials of degree 3.

- $L^2$-TV reconstruction [39]: It is defined as a solution $\hat{x}$ of

$$
\hat{x} \in \arg \min_x \frac{1}{2} \| f(x) - y \|^2 + \lambda \text{TV}(x),
$$

where TV is the total variation. It is defined by $\text{TV}(x) = \| Dx \|_1$, where $D$ is the matrix which maps $x$ onto the vector $Dx$ which contains all differences of neighboured pixels. The scalar $\lambda > 0$ serves as weighting between data fidelity term and regularizer. Note that the objective (5) is convex, such that the solution can be easily computed by classical methods like the alternating direction method of multipliers [16], [17] or a Chambolle-Pock algorithm [7].
For the 2D images, we additionally consider the following methods:

- **Expected patch log likelihood (EPLL) prior [34], [63]:** The idea of EPLL is to assume that the patch distribution can be approximated by a Gaussian mixture model which is fitted to the patch distribution of the reference image. Once the parameters of the GMM \( p \) are estimated, the minimization problem

\[
\hat{x} \in \arg \min_x \frac{1}{2} \| f(x) - y \|^2 - \lambda \sum_{i=1}^{N} \log p(P_i x)
\]

is solved approximately. For denoising tasks, EPLL has shown great performance and beats several classical methods as e.g. BM3D [12], see [63].

- **Deep image prior (DIP) [50]:** The idea of DIP is to solve the optimization problem

\[
\hat{\theta} \in \arg \min_{\theta} \| f(G_\theta(z)) - y \|^2,
\]

where \( G_\theta \) is a UNet with parameters \( \theta \) and \( z \) is a randomly chosen input. We use the same architecture as [50]. Then, the reconstruction \( \hat{x} \) is given by \( \hat{x} = G_\theta(z) \). It was shown in [50], that DIP admits competitive results for many inverse problems.

- **Plug and Play forward backward splitting (PnP-DRUNet and PnP-DS):** Plug and Play methods were first introduced in [54]. The main idea is to consider a classical algorithm from convex optimization and replace the proximal operator with a more general denoiser. Here, we consider the forward backward splitting algorithm for minimizing the function \( F(x) = d(x) + R(x) \) with \( d(x) = \frac{1}{2} \| f(x) - y \|^2 \) given by

\[
x_{r+1} = \text{prox}_R(x_r - \eta \nabla d(x_r)).
\]

Then, by replacing the proximal operator by some denoiser \( D \), we obtain the plug-and-play iteration

\[
x_{r+1} = D(x_r - \eta \nabla d(x_r)).
\]

In our case, we consider two kinds of neural networks as denoiser \( D \). The first one is the DRUNet from [57] which is trained for denoising natural images. We abbreviate the numerical examples with this network by PnP-DRUNet. The second kind is a DnCNN [58] trained on a large training set of the high resolution images of the corresponding material. We abbreviate the corresponding examples by PnP-DS. Note that the use of PnP-DS requires that, we are able to train the denoiser. In particular, we need to have access to a large data-base of domain-specific images which is a much stronger assumption than for any of the other methods. Finally, we set the reconstruction to be \( \hat{x} = x_{100} \). It was shown in [8], [20], [23], [32], [33], [57], [59] that Plug-and-Play methods can achieve state-of-the-art performance for several applications.

---

5We use the code of [34] available at https://github.com/pshibby/fepll_public.

6We use the original implementation from [50] available at https://github.com/DmitryUlyanov/deep-image-prior

B. **Estimation of the Forward Operator**

In Subsection III-A, we used a strided convolution as down-sampling operator because of its simplicity. However, if we consider also non-integral magnification factors, this appears to be infeasible. In this case, we assume that our forward operator is given by \( f(x) = S(k \ast x + b) \) for a \( 15 \times 15 \times 15 \) blur kernel \( k \), a bias \( b \in \mathbb{R} \) and a downsampling operator \( S \). To ensure that \( f \) describes our data as well as possible, we will adapt the blur kernel \( k \) and the bias \( b \) to the registered pair \((\tilde{x}, \tilde{y})\) as described in the paragraph below.

Further, for the downsampling operator \( S \), we make use of Fourier transforms. Given a 3D-image \( x \in \mathbb{R}^{n_z \times n_y \times n_m} \) the three-dimensional discrete Fourier transform (DFT) is defined by \( F_n := F_{n_x} \otimes F_{n_y} \otimes F_{n_m} \), where \( F_n = (\exp(-2\pi i k l/n))^{n-1}_{k,l,n} \). Now, the downsampling operator \( S : \mathbb{R}^{m_z \times m_y \times m_m} \to \mathbb{R}^{n_z \times n_y \times n_m} \) is given by

\[
S = n_z n_y n_m \sum_{m_z}^{m_z} F_{n_z}^{-1} F_{n_y} F_{n_m} D F_{m_z} F_{m_y} F_{m_m},
\]

where for \( x \in \mathbb{C}^{m_z \times m_y \times m_m} \) the \((i,j,k)\)-th entry of \( D(x) \) is given by \( x_{i,j,k'} \), where

\[
j' = \begin{cases} i, & \text{if } i \leq \frac{n_z}{2}, \\ i + m_z - n_z, & \text{otherwise}. \end{cases}
\]

and \( j' \) and \( k' \) are defined analogously. Thus, the operator \( S \) generates a downsampled version \( S(x) \) of an image \( x \) by removing the high-frequency part from \( x \). Note that even if the Fourier matrix \( F_{n_x} \) is complex valued, the range of \( S \) is real-valued, as \( D \) preserves Hermitian-symmetric spectra.

**Estimation of Blur Kernel and Bias:** We assume that we have given images \( \tilde{x} \in \mathbb{R}^{m_z \times m_y \times m_m} \) and \( \tilde{y} \in \mathbb{R}^{n_z \times n_y \times n_m} \) related by \( \tilde{y} \approx S(\hat{k} \ast \tilde{x} + b) \), where the blur kernel \( k \in \mathbb{R}^{15 \times 15 \times 15} \) and the bias \( b \in \mathbb{R} \) are unknown. In the following, we aim at reconstructing \( k \) and \( b \) from \( \tilde{x} \) and \( \tilde{y} \). Here, we use the notations \( N = n_z n_y n_m \) and \( M = m_z m_y m_m \). Further let \( k \in \mathbb{R}^{m_z \times m_y \times m_m} \) be the kernel \( k \) padded with zeros such that it still corresponds to the same convolution as \( k \), but has size \( m_z \times m_y \times m_m \).

Applying the DFT on both sides of \( \tilde{y} \approx S(\tilde{x} + b) \) and using the definition of \( S \), we obtain that

\[
\tilde{y} = \frac{N}{M} D(\tilde{k} \circ \tilde{x} + M b e) = \frac{N}{M} D(\tilde{k}) \circ D(\tilde{x}) + N b e,
\]

where \( \circ \) is the elementwise product and \( e \) denotes the first unit vector (i.e. \( e_{0,0,0} = 1 \) and all other entries are zero). Now, we can conclude that

\[
D(\tilde{k}) = \frac{M}{N} \tilde{y} \circ D(\tilde{x}) - \frac{M b}{x_{0,0,0}} e,
\]

where \( \circ \) is the elementwise quotient. In practice, we stabilize this quotient by increasing the absolute value of \( D(\tilde{x}) \) by \( 10^{-5} \) while retaining the phase. Thus, assuming that the high-frequency part of \( k \) is negligible (i.e., that \( D^2 Dk = k \)), we can approximate \( \tilde{k} \) by

\[
\tilde{k} \approx \frac{M}{N} D^{-1} \tilde{y} \circ D(\tilde{x}) - \frac{M b}{x_{0,0,0}} e.
\]
Applying the inverse DFT this becomes
\[
\hat{k} \approx F^{-1}\left(\hat{m}_{x,m_y,m_z} \cdot \frac{M}{N} D_T(\hat{y} \otimes D(\hat{x})) \right) - \frac{b}{\hat{x}_{0,0,0}}.
\]
Using the assumption that \( \hat{k} \) is zero outside of the \( 15 \times 15 \times 15 \) patch, where \( k \) is located, we can estimate \( b \) by taking the mean over all pixels of \( F^{-1}\left(\hat{m}_{x,m_y,m_z} \cdot \frac{M}{N} D_T(\hat{y} \otimes D(\hat{x})) \right) \) outside of this \( 15 \times 15 \times 15 \) patch. Afterwards, we estimate \( k \) by reprojecting
\[
F^{-1}\left(\hat{m}_{x,m_y,m_z} \cdot \frac{M}{N} D_T(\hat{y} \otimes D(\hat{x})) \right) - \frac{b}{\hat{x}_{0,0,0}}
\]
to the set of all real \( 15 \times 15 \times 15 \) kernels.

ACKNOWLEDGMENT

The authors would like to thank Paul Scherrer Institut, Villigen, Switzerland for provision of synchrotron radiation beamtime at the TOMCAT beamline X02DA of the SLS. Furthermore, we thank Diego Roldán from the TU Kaiserslautern for the generation of the synthetic FIB-SEM images from Section III-B and Sören Höhn (Fraunhofer IKTS) for providing the real FIB-SEM data.

REFERENCES

[1] J. Adler and O. Öktem, “Deep Bayesian Inversion,” 2018, arXiv: 1811.05910.
[2] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein generative adversarial networks,” in Proc. Int. Conf. Mach. Learn., 2017, pp. 214–223.
[3] R. Bergmann, J. Persch, and G. Steidl, “A parallel Douglas-Rachford algorithm for minimizing ROF-like functionals on images with values in symmetric Hadamard manifolds,” SIAM J. Imag. Sci., vol. 9, no. 3, pp. 901–937, 2016.
[4] A. Buades, B. Coll, and J.-M. Morel, “A non-local algorithm for image denoising,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2005, vol. 2, pp. 60–65.
[5] G. T. Buzzard, S. H. Chan, S. Sreehari, and C. A. Bouman, “Plug-and-play upblurred: Optimization-free reconstruction using consensus equilibrium,” SIAM J. Imag. Sci., vol. 11, no. 3, pp. 2001–2020, 2018.
[6] C. B. Carter and M. G. Norton. Ceramic Materials: Science and Engineering—New York, NY, USA: Springer-Verlag, 2007.
[7] A. Chambolle and T. Pock, “A first-order primal-dual algorithm for convex problems with applications to imaging,” J. Math. Imag. Vis., vol. 40, no. 1, pp. 120–145, 2011.
[8] S. H. Chan, X. Wang, and O. A. Elgendy, “Plug-and-play ADMM for image restoration: Fixed-point convergence and applications,” IEEE Trans. Comput. Imag., vol. 3, no. 1, pp. 84–98, Mar. 2017.
[9] S. Chiu, D. Stoyan, W. Kendall, and J. Mecke. Stochastic Geometry and its Applications, Hoboken, NJ, USA: Wiley, 2013.
[10] P. L. Combettes and J.-C. Pesquet, “A proximal decomposition method for solving convex variational inverse problems,” Inverse problems,” vol. 24, no. 6, 2008, Art. no. 065014.
[11] F. Crete, T. Dolmiere, P. Ladret, and M. Nicolas, “The blur effect: Perception and estimation with a new no-reference perceptual blur metric,” in Hum. Vis. Electron. Imag., vol. 6492, pp. 196–206, 2007.
[12] A. Danielyan, V. Katkovnik, and K. Egiazarian, “BM3D frames and variational image deblurring,” IEEE Trans. Image Process., vol. 21, no. 4, pp. 1715–1728, Apr. 2012.
[13] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2009, pp. 248–255.
[14] C. Dong, C. C. Loy, K. He, and X. Tang, “Image super-resolution using deep convolutional networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 2, pp. 295–307, Feb. 2016.
[15] A. A. Efros and T. K. Leung, “Texture synthesis by non-parametric sampling,” in Proc. 7th IEEE Int. Conf. Comput. Vis., 1999, vol. 2, pp. 1033–1038.
[16] D. Gabay, “Applications of the Method of Multipliers to Variational Inequalities,” in Studies in Mathematics and Its Applications, Amsterdam, The Netherlands: Elsevier, 1983, vol. 15, pp. 299–331.
[17] D. Gabay and B. Mercier, “A dual algorithm for the solution of nonlinear variational problems via finite element approximation,” Comput. Math. Appl., vol. 2, no. 1, pp. 17–40, 1976.
[18] A. Genevay, M. Cuturi, G. Peyre, and F. R. Bach, “Stochastic optimization for large-scale optimal transport,” in Proc. Int. Conf. Neural Inf. Process. Syst., 2016, pp. 3440–3448.
[19] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, “Improved training of wasserstein GANS. Advances in neural information processing systems,” in Proc. 31st Int. Conf. Neural Inf. Process. Syst., 2017, pp. 5769–5779.
[20] H. Gupta, K. H. Jin, H. Q. Nguyen, M. T. McCann, and M. Unser, “CNN-based projected gradient descent for consistent CT image reconstruction,” IEEE Trans. Med. Imag., vol. 37, no. 6, pp. 1440–1453, Jun. 2018.
[21] J. Gutierrez, J. Rabin, B. Galeme, and T. Hurtat, “Optimal patch assignment for statistically constrained texture synthesis,” in Proc. Int. Conf. Scale Space Variational Methods Comput. Vis., 2017, pp. 172–183.
[22] M. Hasanasab, J. Hertrich, F. Laus, and G. Steidl, “Alternatives to the EM algorithm for ML estimation of location, scatter matrix, and degree of freedom of the student-t distribution,” Numer. Algorithms, vol. 87, no. 1, pp. 91–118, 2021.
[23] J. Hertrich, S. Neumayer, and G. Steidl, “Convolutional proximal neural networks and plug-and-play algorithms,” Linear Algebra Appl., vol. 631, pp. 203–234, 2021.
[24] J. Hertrich et al., “PCA reduced Gaussian mixture models with applications in superresolution,” Inverse Problems Imag., vol. 16, no. 2, pp. 341–366, 2021.
[25] A. Houdard, C. Bouveyron, and J. Delon, “High-dimensional mixture models for unsupervised image denoising (HDMI),” SIAM J. Imag. Sci., vol. 11, no. 4, pp. 2815–2846, 2018.
[26] A. A. Houdard, N. Leclaire Papadakis, and J. Rabin, “Wasserstein generative models for patch-based texture synthesis,” in Proc. Int. Conf. Scale Space Variational Methods Comput. Vis., 2021, pp. 269–280.
[27] R. Keys, “Cubic convolution interpolation for digital image processing,” IEEE Trans. Acoust., Speech, Signal Process., vol. 29, no. 6, pp. 1153–1160, Dec. 1981.
[28] T. K. Leung, K. P. S. Tam, and M. Unser, “A new multi-scale approach to image deblurring,” IEEE Trans. Med. Imag., vol. 23, no. 3, pp. 214–223, 2004.
[29] A. Leclaire and J. Rabin, “A fast multi-layer approximation to semi-variational problems via finite element approximation,” Appl. Numer. Math., vol. 2, no. 1, pp. 17–40, 1976.
[30] S. Ono, “Primal-dual plug-and-play image restoration,” in Proc. Int. Conf. Scale Space Variational Methods Comput. Vis., 2017, pp. 1781–1790.
[31] S. Ono, “Primal-dual plug-and-play image restoration,” IEEE Signal Process. Lett., vol. 24, no. 8, pp. 1108–1112, Aug. 2017.
[32] S. Parameswaran, C. Deledalle, L. Denis, and T. Q. Nguyen, “Accelerating GMM-based patch priors for image restoration: Three ingredients for a 100x speed-up,” IEEE Trans. Image Process., vol. 28, no. 2, pp. 687–698, Feb. 2019.
[33] Y. Romano, J. Isidoro, and P. Milanfar, “RAISR: Rapid and accurate image super resolution,” IEEE Trans. Comput. Imag., vol. 3, no. 1, pp. 110–125, 2017.
[34] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” Physica D: Nonlinear Phenomena, vol. 60, no. 1–4, pp. 259–268, 1992.
[35] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, “Image super-resolution via iterative refinement,” 2021, arXiv:2104.07636.
[41] P. Sandeep and T. Jacob, “Single image super-resolution using a joint GMM method,” IEEE Trans. Image Process., vol. 25, no. 9, pp. 4233–4244, Sep. 2016.

[42] W. Saxton and W. Baumeister, “The correlation averaging of a regularly arranged bacterial cell envelope protein,” J. Microsc., vol. 127, no. 2, pp. 127–138, 1982.

[43] K. Schladitz, “Quantitative micro-CT,” J. Microsc., vol. 243, no. 2, pp. 111–117, 2011.

[44] M. Segu, M. Grinvald, R. Siegwart, and F. Tombari, “3DSNet: Unsupervised shape-to-shape 3D style transfer,” 2020, arXiv:2011.13388.

[45] J. Shao, L. Chen, and Y. Wu, “SRWGANTV: Image super-resolution through wasserstein generative adversarial networks with total variational regularization,” in Proc. Int. Conf. Comput. Res. Develop., 2021, pp. 21–26.

[46] A. Shocher, N. Cohen, and M. Irani, “Zero-shot,” super-resolution using deep internal learning in Proc. IEEE Conf. Comput. Vis. pattern Recognit., 2018, pp. 3118–3126.

[47] S. Sreehari et al., “Plug-and-play priors for bright field electron tomography and sparse interpolation,” IEEE Trans. Comput. Imag., vol. 2, no. 4, pp. 408–423, Dec. 2016.

[48] J. Stanczuk, C. Etmann, L. M. Kreusser, and C.-B. Schönlieb, “Wasserstein GANs work because they fail (to approximate the wasserstein distance),” 2021, arXiv:2103.01678.

[49] C. Sutour, C.-A. Deledalle, and J.-F. Aujol, “Estimation of the noise level function based on a nonparametric detection of homogeneous image regions,” SIAM J. Imag. Sci., vol. 8, no. 4, pp. 2622–2661, 2015.

[50] D. Ulyanov, A. Vedaldi, and V. Lempitsky, “Deep image prior,” in Proc. IEEE Conf. Comput. Vis. pattern Recognit., 2018, pp. 9446–9454.

[51] M. Van Heel, W. Keegstra, W. Schutter, and E. V. Bruggen, “Arthropod hemocyanin structures studied by image analysis,” Life Chem. Rep. Suppl., vol. 1, pp. 69–73, 1982.

[52] M. van Heel and M. Schatz, “Fourier shell correlation threshold criteria,” J. Struct. Biol., vol. 151, no. 3, pp. 250–262, 2005.

[53] S. Vaucher et al., “On-line tools for microscopic and macroscopic monitoring of microwave processing,” Physica B: Condens. Matter, vol. 398, no. 2, pp. 191–195, 2007.

[54] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, “Plug-and-play priors for model based reconstruction,” in Proc. IEEE Glob. Conf. Signal Inf. Process., 2013, pp. 945–948.

[55] Z. Wang, J. Chen, and S. C. Hoi, “Deep learning for image super-resolution: A survey,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 10, pp. 3365–3387, Oct. 2021.

[56] F. Wu, B. Wang, D. Cui, and L. Li, “Single image super-resolution based on wasserstein guns,” in Proc. 37th Chin. Control Conf., 2018, pp. 9649–9653.

[57] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte, “Plug-and-play image restoration with deep denoiser prior,” IEEE Trans. Pattern Anal. Mach. Intell., early access, doi: 10.1109/TPAMI.2021.3088914.

[58] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, “Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising,” IEEE Trans. Image Process., vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[59] K. Zhang, W. Zuo, and L. Zhang, “Deep plug-and-play super-resolution for arbitrary blur kernels,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2019, pp. 1671–1681.

[60] L. Zhang, L. Zhang, X. Mou, and D. Zhang, “FSIM: A feature similarity index for image quality assessment,” IEEE Trans. Image Process., vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[61] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” in Proc. IEEE Conf. Comput. Vis. pattern Recognit., 2018, pp. 586–595.

[62] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, “Residual dense network for image super-resolution,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 2472–2481.

[63] D. Zoran and Y. Weiss, “From learning models of natural image patches to whole image restoration,” in Proc. IEEE Int. Conf. Comput. Vis., 2011, pp. 479–486.