

# Adaptive Quantile Sparse Image (AQuaSI) Prior for Inverse Imaging Problems

Franziska Schirmacher<sup>1</sup>, Christian Riess<sup>2</sup>, and Thomas Köhler<sup>1</sup>

<sup>1</sup>Pattern Recognition Lab, Friedrich-Alexander-University (FAU)  
Erlangen-Nürnberg, Erlangen 91058, Germany. E-mail: {franziska.schirmacher,  
thomas.koehler}@fau.de

<sup>2</sup>IT Security Infrastructures Lab, FAU Erlangen-Nürnberg, Erlangen 91058,  
Germany.

## Abstract

Inverse problems play a central role for many classical computer vision and image processing tasks. Many inverse problems are ill-posed, and hence require a prior to regularize the solution space. However, many of the existing priors, like total variation, are based on ad-hoc assumptions that have difficulties to represent the actual distribution of natural images. Thus, a key challenge in research on image processing is to find better suited priors to represent natural images.

In this work, we propose the Adaptive Quantile Sparse Image (AQuaSI) prior. It is based on a quantile filter, can be used as a joint filter on guidance data, and be readily plugged into a wide range of numerical optimization algorithms. We demonstrate the efficacy of the proposed prior in joint RGB/depth upsampling, on RGB/NIR image restoration, and in a comparison with related regularization by denoising approaches.

## 1 Introduction

In computer vision and image processing, many tasks are modeled as inverse problems ranging from low-level vision to image analysis and scene understanding. Inverse problems refer to the inference of latent information from noisy observations, whereas both quantities are linked by a generative model. For instance, in image denoising, we are interested in obtaining clean images from noisy ones under a certain noise model. Despite widespread application domains, such formulations can be tackled from two complementary perspectives: *Model-based* approaches aim at inverting the generative model using iterative optimization. This includes classical Bayesian statistics [1] or the most recent compressive sensing methodologies [2]. *Learning-based* approaches infer mappings in forward direction by learning from exemplars. Here, dictionary-based methods [3] constitute the classical approach, while more recently deep neural network pushed the frontiers towards end-to-end learning, for example in super-resolution [4] or motion deblurring [5].

An inverse problem is oftentimes highly *ill-posed*. This implies that its solution is not necessarily unique and continuously related to the given observations. Turning ill-posed problems to well-posed ones requires *priors* on the desired solution spaces. Such prior knowledge can either be incorporated implicitly via learning from example data, or explicitly by enforcing properties of a solution. The latter has the merit that it is exclusively driven by available input data without relying on external training data. It is achieved by a proper regularization of the problem, which can be steered by general assumptions like smoothness or sparsity of the desired solution. For instance, traditional models like total variation (TV) or related methods regularize with *linear* filter operations (e.g., gradients or Wavelet transforms) using sparsity-inducing loss functions in that filter domain. Such priors are well understood and can be modeled in Bayesian form but, despite their flexibility, they are

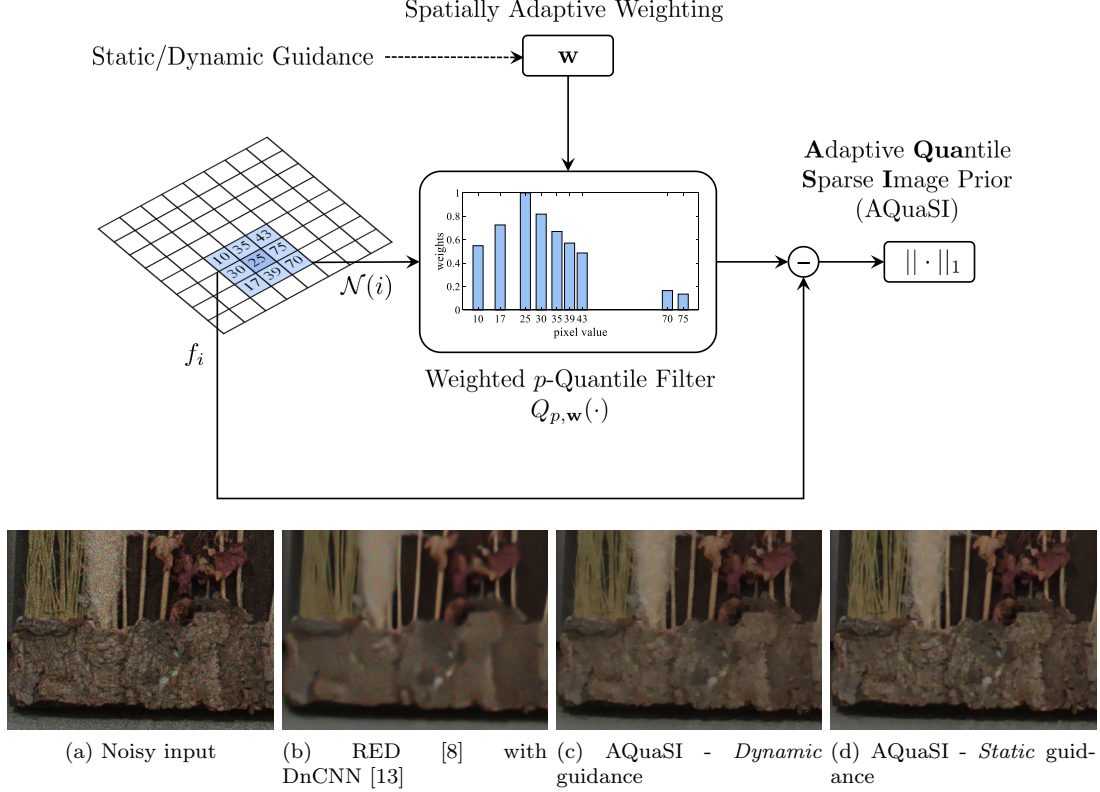


Figure 1: Top: Schematic representation of the AquaSI prior. Bottom: Denoising results on an image corrupted by multiplicative speckle noise. The weighted  $p$ -quantile filter is applied to an input image  $\mathbf{f}_i$ . Then, the  $L_1$  norm is applied to the difference of the filtered output and the input image. The filter sorts the pixel values within a predefined neighborhood  $\mathcal{N}(i)$  in ascending order. Weights are assigned based on the similarity to the center pixel and the intensity distribution in a guidance image. The guidance can be gained by external data (*static* guidance) or from the input data itself (*dynamic* guidance).

based on too simplistic design choices to reasonably constraining the solution space. For example, TV priors in image restoration oftentimes cause staircasing artifacts that are typically not observed in natural images.

Recently, powerful priors that build on image denoising methods have been demonstrated for their performance in inverse imaging. Plug-and-play (PnP) [6] formulations, autoencoding priors [7], as well as regularization by denoising (RED) [8] and variants thereof [9, 10] are some manifestations of this approach. Specifically, RED allows to relate *non-linear* operators with image priors in analytical form. However, despite its practical success, Reehorst and Schniter [11] showed that RED algorithms can only be related to specific priors under denoisers with symmetric Jacobian. Unfortunately, this property does not hold true for popular denoisers like BM3D or deep neural nets. Moreover, RED is based on a cross correlation between an image and its residual noise estimated by a denoiser, which is assumed to be white noise. This model can be translated into Tikhonov-like regularization also known as graph Laplacian in case of symmetric smoothing [12]. This is a delicate property as it would require a perfect denoiser to infer the residual noise. Current denoising engines can barely infer the residual in form of pure white noise, especially if observations are degraded by challenging noise types like Poisson or multiplicative models. This can limit the power of the prior in many applications.

In this paper, we pursue the goal of designing a more appropriate model and propose the **A**daptive

**Quantile Sparse Image (AQuaSI) prior** – a novel prior for ill-posed inverse imaging problems. The proposed AQuaSI prior falls into the broad class of RED formulations with some crucial design choices: 1) We exploit the general class of *weighted quantile filters* as a specific denoising type to build our model. Contrary to standard denoisers, such filters are particularly effective for heavy-tailed noise distributions differing from Gaussian models, e. g., impulsive noise [14]. In our framework, they can also be linearized, and hence very efficiently deployed for numerical optimization. 2) AQuaSI assumes images to be fixed points under weighted quantile filtering similar to autoencoding priors, but with *sparsity-promoting regularization* contributing to its robustness. In particular, we use  $L_1$  norm regularization for fixed points and can relate this model to a prior distribution in the sense of Bayesian statistics. In contrast, the graph Laplacian in RED is sensitive to denoising inaccuracy of the used filter. While both priors are expected to perform equally under ideal conditions like additive Gaussian noise, our sparse model gains in importance under challenging noise distributions. The advantages of these properties can be seen in Fig. 1 (left) for multiplicative speckle noise removal. Here, RED with DnCNN [13] as denoiser fails to handle this noise distribution and overpenalizes the denoising inaccuracy in its regularization term, resulting in oversmoothing. In contrast, AQuaSI with its weighted quantile filter and sparse regularization achieves superior structure preservation.

The proposed model generalizes our previous paper on the QuaSI prior [15]. This earlier work proposes a quantile filter as prior. In this work, we make it spatially adaptive by introducing filter weights. This filter class also contains weighted median filtering [16, 17] as a special instance. As such, AQuaSI enables regularization via *joint filtering* by calculating the filter weights from *guidance data*. In this case, the pixel values within the filter kernel are sorted and weighted by values derived from the guidance data to calculate quantiles. The spatial weighting can be either derived from an external image, termed *static* guidance, or from the given input image, termed *dynamic* guidance. In contrast, QuaSI exploits uniform weights. A schematic representation of our prior is shown in Fig. 1 (right).

We further show that AQuaSI is universally usable and how it can be deployed in different model-based optimization approaches to solve typical inverse imaging problems<sup>1</sup>. In different applications, including non-blind image deblurring, joint upsampling of RGB/depth images, and restoration of RGB/near infrared images, we demonstrate that our prior improves upon hand-crafted priors that are driven by ad-hoc assumptions.

This paper is organized as follows. We first mathematically define the inverse problem and list related work on it in Sec. 2. The AQuaSI prior and its numerical optimization are presented in Sec. 3 and Sec. 4, respectively. Quantitative and qualitative experiments on three applications of AQuaSI are presented in Sec. 5. Finally, Sec. 6 concludes this work.

## 2 Problem Statement and Related Work

The focus of this work is on inverse problems in computer vision. An inverse problem aims at recovering a latent image  $\mathbf{f} \in \mathbb{R}^n$  from measurements  $\mathbf{g} \in \mathbb{R}^m$ . In a maximum a-posteriori (MAP) formulation, the latent image maximizes the posterior distribution  $p(\mathbf{f}|\mathbf{g})$  and can be inferred as the solution of an optimization problem using Bayes rule

$$\hat{\mathbf{f}} = \underset{\mathbf{f}}{\operatorname{argmax}} p(\mathbf{f}|\mathbf{g}) = \underset{\mathbf{f}}{\operatorname{argmax}} p(\mathbf{g}|\mathbf{f})p(\mathbf{f}) . \quad (1)$$

The prior  $p(\mathbf{f})$  models the distribution of the latent image and the observation model  $p(\mathbf{g}|\mathbf{f})$  denotes the probability of observing the measurement  $\mathbf{g}$  from the latent image  $\mathbf{f}$ . Using the negative log-likelihood  $-\log(p(\mathbf{f}|\mathbf{g}))$ , Eq. 1 can be written as the energy minimization task

$$\hat{\mathbf{f}} = \underset{\mathbf{f}}{\operatorname{argmin}} \{ \mathcal{L}(\mathbf{f}, \mathbf{g}) + \lambda R(\mathbf{f}) \} , \quad (2)$$

---

<sup>1</sup>The source code to our methods and supplementary material can be downloaded at: <https://github.com/franziska-schirmacher/AQuaSI>

where  $\mathcal{L}(\mathbf{f}, \mathbf{g}) = -\log(p(\mathbf{g}|\mathbf{f}))$  is a *data term* to measure the fidelity of an estimate  $\mathbf{f}$  w.r.t.  $\mathbf{g}$  under a given image formation model, and  $\lambda R(\mathbf{f}) = -\log(p(\mathbf{f}))$  is a *regularization term* with weight  $\lambda \geq 0$  derived from prior knowledge on the solution  $\hat{\mathbf{f}}$ . In common frameworks for denoising, deblurring, or super-resolution, the data term is  $\mathcal{L}(\mathbf{f}, \mathbf{g}) = \|\mathbf{g} - \mathbf{W}\mathbf{f}\|_2^2$ , where  $\mathbf{W} \in \mathbb{R}^{n \times m}$  is a task-dependent operator that models an underlying linear image formation model. While the choice of the data term is highly dependent on the application, it is oftentimes possible to define more generic priors for the regularization term  $R(\mathbf{f})$ . The large body of existing works on image priors can roughly be organized into priors on sparsity, priors from deep learning, and priors from denoising.

### 2.0.1 Natural Scene Statistics and Sparsity Priors

The vast majority of classical image priors  $p(\mathbf{f})$  for Eq. 1 are derived from statistical models of natural scenes. Some common assumptions are to consider natural images as smooth, piecewise smooth, or piecewise constant signals under certain transforms. This includes the well known TV prior [18] that has been widely adopted for image restoration. Many elaborated optimizers exist to solve inverse problems with a TV prior, but the underlying statistical assumption of a piecewise constancy of image intensities is in many applications too simplistic to appropriately model the complex nature of natural scenes.

Further advances in the field of natural scene statistics led to Hyper-Laplacian priors [19–21] or normalized scale-invariant priors [22]. Specifically for deblurring, non-linear extreme value priors like dark [23] or bright channels [24] have been proposed. Such approaches exploit sparsity of natural images as a stronger prior for image restoration. Spectral models [25, 26] are likewise heading towards this direction but exploit low-rank properties of non-local similar patches. These are more suitable to model natural scene statistics but require specific optimization strategies like re-weighted minimization [20] or variable splitting [21], which limits their flexibility. In contrast, we aim in this work at designing an *out-of-the box* image prior that is widely applicable within different optimization schemes.

### 2.0.2 Deep Learning-Based Priors

The use of deep learning architectures like convolutional neural networks (CNNs) is a different track to design the prior  $p(\mathbf{f})$  in Eq. 1 and has become popular to solve low-level vision problems. In early works, classical approaches to sparse coding like [27] have been newly interpreted via CNNs and learned in an end-to-end fashion from exemplars [28]. Other approaches are to integrate denoising networks to regularize model-based optimization in variable splitting schemes [29, 30]. However, these methods require learning from large-scale example datasets, which are not readily available in many real-world applications. In contrast, the proposed prior in this work is *unsupervised*.

Deep image priors [10, 31] are one alternative, as prior knowledge on natural images is extracted from network architectures instead of example data. In principle, this avoids the use of simplistic hand-crafted features for regularization, but such priors are — other than the proposed prior — difficult to deploy in existing model-based image restoration frameworks.

### 2.0.3 Transforming a Denoiser into a Regularizer

Most closely related to our proposed approach are strategies to form the prior  $p(\mathbf{f})$  in Eq. 1 via image denoising. PnP priors are one of these attempts and replace proximal operators in variable splitting algorithms like the alternating direction method of multiplier (ADMM) [6, 32] or primal dual optimization [33] by invoking denoising algorithms. Moreover, variants of the iterative shrinkage/thresholding algorithm (ISTA) [34] have been proposed to extend the PnP formulation towards non-linear [35] and online optimization problems [36]. This implicitly models a prior on the solution space and have gained much popularity in signal recovery and image restoration. Chan [12] provided theoretical insights to the actual performance of PnP by deriving relations to classical MAP inference and consensus equilibrium [37] for the class of symmetric smoothing priors.

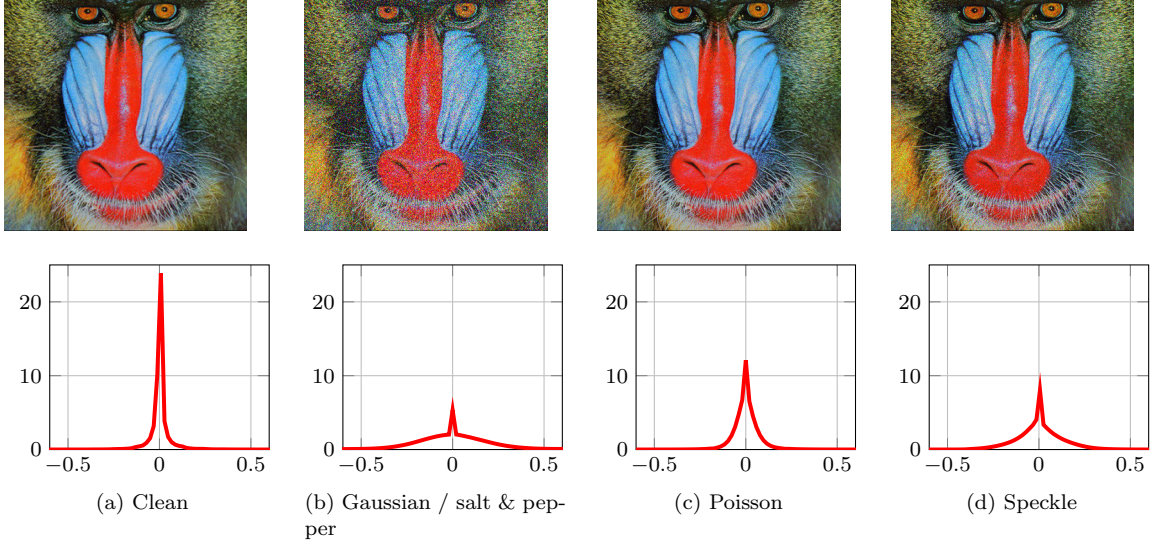


Figure 2: Influence of different images degradations to the behavior of the AquaSI prior. Top row: clean and degraded *baboon* image using Gaussian, mixed Gaussian / salt & pepper, Poisson, and speckle noise, respectively. Bottom row: empirical distribution of the quantile residual  $\mathbf{r} = \mathbf{f} - Q_{p,\mathbf{w}}(\mathbf{f})$  for the different degradations. Notice that the distributions on clean images are strongly concentrated around 0, while the distribution of noisy images are very broad.

In that sense, an alternative approach includes regularization by denoising (RED) as proposed by Romano *et al.* [8]. Their model considers clean images as a fixed point of image denoising. In its core, RED extends the well known fixed point property of natural images as raised in related works on autoencoding [7] or mean-shift priors [38] to general denoisers. This allows the integration of many advanced filters to solve inverse imaging problems under different optimization schemes and broadens the scope compared to PnP. In principle, RED can also be combined with other complementary models like deep image priors [10]. The used denoiser can be related to an analytical prior in a MAP framework provided that it has symmetric Jacobian [11] – a property that is, however, not fulfilled by many elaborated filters including those covered in this work. From a practical perspective, RED is based on Tikhonov regularization similar to related autoencoding priors as proposed in the work of Bigdeli *et al.* [7]. In [39], the authors showed that PnP can also be related to Tikhonov regularization. A direct comparison in [12] shows that RED is less sensitive to parameters compared to PnP, while PnP yields better performance in terms its mean squared error. However, overall Tikhonov regularization can be expected to be less robust to non-Gaussian noise, e. g., than  $L_1$ -based regularizer. The proposed AquaSI prior aims at increasing the robustness using a sparsity-promoting model without relying on strong assumptions like a symmetric Jacobian for its interpretation.

### 3 Adaptive Quantile Sparse Image (AquaSI) Prior

The proposed AquaSI prior is a spatially adaptive model for natural images. It is sensitive to common image degradations, and can be efficiently implemented in various optimization frameworks. In this section, we first introduce the prior itself and its behavior under common image degradations. Then, we show its applicability in joint filtering, and finally present a linearization of the filter for insertion into an optimizer.

### 3.1 Definition of the AQuaSI Prior

Turning denoising algorithms into an image prior for regularization is an emerging topic in image restoration. Our approach follows a similar train of thought and is based on *weighted quantile filtering* as a particular denoiser. We denote a weighted  $p$ -quantile filter with  $0 \leq p \leq 1$  for an image  $\mathbf{f}$  as  $Q_{p,\mathbf{w}}(\mathbf{f})$ , where  $\mathbf{w} \in \mathbb{R}^N$  are spatially adaptive weights. The weighted quantile filter selects a pixel from a neighborhood based on two considerations: a) the intensity distribution in the neighborhood, b) the distribution of similarities of each pixel to the center pixel. That way, pixels that are very dissimilar to the center can be weighed down, and effectively be ignored by the quantile filter.

Let  $\mathbf{w}_i \in \mathbb{R}^n$  with  $n = |\mathcal{N}(i)|$  be the weights within the kernel  $\mathcal{N}(i)$  of this filter centered at the  $i$ -th pixel. Then, we define a permutation of these weights,  $\tilde{\mathbf{w}}_i = m_i(\mathbf{w})$ , such that the associated pixel intensities  $m_i(f_{ij}) \in \mathcal{N}(i)$  are sorted in ascending order. The weighted  $p$ -quantile is the element  $[Q_{p,\mathbf{w}}(\mathbf{f})]_i = f_l$  with

$$w_l = m_i^{-1}(\tilde{w}_{ik^*}) \quad (3)$$

where  $m_i^{-1}(\tilde{\mathbf{w}}_i)$  maps the elements of the permutation  $\tilde{\mathbf{w}}_i$  back to weights in the filter kernel  $\mathbf{w}_i$  and the index  $k^*$  is given by:

$$k^* = \min k \quad \text{s.t.} \quad \sum_{j=1}^k \tilde{w}_{ij} \geq p \sum_{j=1}^n \tilde{w}_{ij} . \quad (4)$$

The general model in Eq. 3 and Eq. 4 depends on the percentage parameter  $p$ , which is a hyper-parameter to adjust the filter for specific applications. If  $p = 0.5$ , we obtain the weighted median filter, which has been shown to be particularly suitable for denoising under zero-mean noise [16, 17].

Conceptionally, the proposed AQuaSI prior follows the idea that a natural image should be a fixed point under the weighted quantile filter. That is, the *quantile residual* of a clean image  $\mathbf{f}$  is zero, i. e.  $\mathbf{f} - Q_{p,\mathbf{w}}(\mathbf{f}) = \mathbf{0}$ . In order to apply this assumption for the regularization of inverse problems, we define the prior with slight relaxation according to the sparsity-inducing term

$$R_{\text{AQuaSI}}(\mathbf{f}) = \|\mathbf{f} - Q_{p,\mathbf{w}}(\mathbf{f})\|_1 . \quad (5)$$

Under the general MAP framework in Eq. 1, we can relate this regularization term to the prior distribution  $p(\mathbf{f}) = \exp(-\lambda R_{\text{AQuaSI}}(\mathbf{f}))$ .

### 3.2 Impact of Image Degradations on the AQuaSI Prior

We validate the key assumptions behind our AQuaSI prior on natural images that have been subject to different types of degradations. Specifically, we present the behavior of AQuaSI for several noise distributions. We use the images from Set14 [40] as *clean* images, and simulate *noisy* counterparts using mixed Gaussian and salt & pepper noise, Poisson noise, and speckle noise. Sample images are shown in the top row of Fig. 2. For each noise model, we compute the quantile residual

$$\mathbf{r} = \mathbf{f} - Q_{p,\mathbf{w}}(\mathbf{f}) , \quad (6)$$

with  $p = 0.5$  and  $\mathbf{w} = \mathbf{1}$  on a neighborhood of  $n = 5^2$  pixels, which corresponds to the usual median operator on a window of  $5 \times 5$  pixels. In the bottom row of Fig. 2, histograms of the residuals are shown.

Two observations can be made from these distributions. First, the residual distributions of noisy images are very broad. In contrast to that, residual distributions on clean images are strongly concentrated around zero. Second, also for clean images, there are a few notable outliers that largely differ from zero. Both observations together indicate that the proposed quantile residual is an adequate model for natural images: the overall concentration around zero indicates that the quantile is a good predictor for most pixels, and the use of the sparsity-promoting  $L_1$  norm ensures that the influence of the few outliers in the clean image does not dominate the overall result.



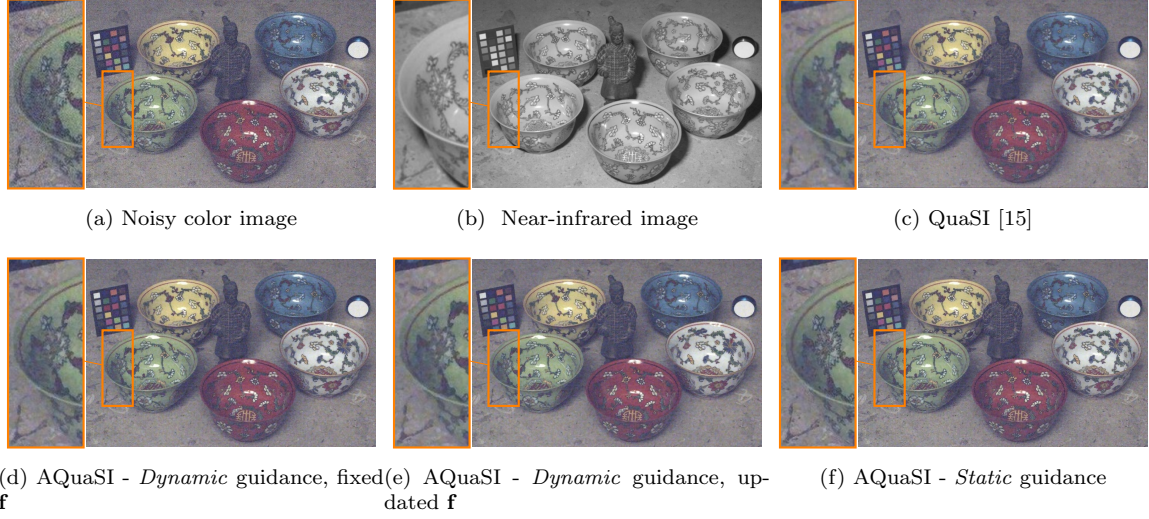


Figure 3: Denoising of the *bowls* color image [41] using the ADMM optimization in combination with different variants of our prior. QuaSI [15] is spatially non-adaptive and can only exploit color data for regularization. AQuaSI is spatially adaptive and can be steered by color data (either the fixed input or an intermediate ADMM result) in a *dynamic* guidance mode or by external near-infrared data in a *static* guidance mode. Notice the superior recovery of small image structures achieved by the spatially adaptive variants.

### 3.3 Joint Filtering

Similar to existing local [17, 42] or global [43] optimization-based methods, the proposed prior facilitates *joint filtering* by steering the regularizer in Eq. 5 with either internal or external image features. This is achieved by controlling the weights  $\mathbf{w}$  of the underlying weighted quantile operator. At the same time, the non-uniform choice of weights make the proposed technique spatially adaptive.

More specifically, for regularization of  $\mathbf{f}$ , the weight  $w_{ij}$  for the pixel  $f_j$  w.r.t. the pixel  $f_i$  is determined according to

$$w_{ij} = G_{\sigma_w}(z_i, z_j) \quad , \quad (7)$$

where  $z_i$  and  $z_j$  are the respective pixels in a *guidance image*  $\mathbf{z}$ , and  $G_{\sigma_w}(z_i, z_j)$  is a measure for their spatial neighborhood. For instance, we can use an isotropic Gaussian kernel

$$G_{\sigma_w}(z_i, z_j) = \exp\left(-\frac{(z_i - z_j)^2}{2\sigma_w^2}\right) \quad , \quad (8)$$

where  $\sigma_w$  denotes the standard deviation. Joint filtering can be done in two modes: In *static* guidance mode [17, 42], spatially adaptive weights  $\mathbf{w}$  are derived from an external guidance image  $\mathbf{z}$ . In *dynamic* guidance mode [43], spatially adaptive weights  $\mathbf{w}$  are internally derived in the domain of the regularized image without considering additional guidance data. This can be done using either the fixed input  $\mathbf{g}$  to obtain spatially adaptive weights (i.e.,  $\mathbf{z} = \mathbf{g}$ ) or using an indeterminate result  $\mathbf{f}$  (i.e.  $\mathbf{z} = \mathbf{f}$ ) to gradually update the weights within iterative optimization as introduced in Sec. 4.

Spatial adaptivity has a large impact on image structure preservation, and as such greatly improves upon our earlier, spatially non-adaptive prior, which we then called “QuaSI” [15]. In order to show the impact of spatial adaptivity, we compare AQuaSI with *static* and *dynamic* guidance to the non-adaptive QuaSI prior. This is done for the task of RGB and near-infrared (NIR) image restoration (see Sec. 5.4) [44], where the filtering of noisy color images is guided by NIR data. The parameters of the quantile filter are  $p = 0.5$  for a median filter on a  $n = 5^2$  neighborhood and  $\sigma_w = 0.1$  for the Gaussian kernel  $G_{\sigma_w}$ .

Figure 3 shows the behavior of the different priors on the *bowls* scene. Fine details, such as the border of the flower printing, are better preserved in the filtered color image using AQuaSI with *static* guidance obtained from NIR data: the edges in Fig. 3f are sharper than the edges in Fig. 3c. Also between the spatially non-adaptive QuaSI and AQuaSI with *dynamic* guidance, AQuaSI performs better in preserving more details of the flower printing while smoothing homogeneous regions. We also observe superior denoising performance for *dynamic* mode when using intermediate results as guidance (Fig. 3e) than when using the noisy input image as guidance (Fig. 3d). Also the flower printings are better preserved.

Thus, the spatial adaptivity and the possibility of using additional guidance data broadens the applicability and enhances the performance of the proposed AQuaSI prior compared to the QuaSI prior.

### 3.4 Pseudo-Linear Filtering

The proposed AQuaSI prior is both non-convex and highly non-linear due to the underlying weighted quantile filtering. However, this flexibility somewhat complicates the numerical optimization when using this model for an image restoration task. To mitigate this issue, a proper linearization can be used to deploy our prior for the solution of inverse imaging problems. We propose to formulate  $Q_{p,\mathbf{w}}(\mathbf{f})$  in the pseudo-linear form [8]:

$$Q_{p,\mathbf{w}}(\mathbf{f}) = \mathbf{Q}\mathbf{f} , \quad (9)$$

where  $\mathbf{Q}$  denotes the weighted quantile filter in matrix notation constructed from the image  $\mathbf{f}$ . This decomposes the filter into the construction of its pseudo-linear form  $\mathbf{Q}$  adapted to the image  $\mathbf{f}$  and the actual filtering. While this formulation applies in principle to many popular non-linear models, like non-local means or bilateral filtering, the proposed weighted quantile filtering allows for the efficient linearization

$$Q_{ij} = \begin{cases} 1 & \text{if } j = z \text{ in the sense of Eq. 3} \\ 0 & \text{otherwise} \end{cases} , \quad (10)$$

where  $z$  denotes the position of the  $p$ -weighted quantile in the neighborhood  $\mathcal{N}(i)$  centered at the  $i$ -th pixel.

The pseudo-linear form in Eq. 9 is accurate for a fixed image  $\mathbf{f}$ . In case of small perturbations in the image that do not affect the ordering of the pixel values in the neighborhood  $\mathcal{N}(i)$ , this form is still accurate and  $\mathbf{Q}$  remains constant.

## 4 Numerical Optimization with AQuaSI Prior

AQuaSI can be deployed within several widely used iterative optimization schemes for inverse imaging problems. These schemes follow the quite general energy minimization framework of Eq. 2.

### 4.1 Gradient Descent

Gradient descent is among the most frequently used optimization methods to solve inverse imaging problems. We study steepest descent as a conceptional simple, yet flexible gradient descent iteration scheme. Given an estimate  $\mathbf{f}^t$  for the solution of the inverse problem in Eq. 2 with the problem-specific data fidelity term  $\mathcal{L}(\mathbf{f}, \mathbf{g})$  and our proposed regularization term  $R_{\text{AQuaSI}}(\mathbf{f})$ , we refine  $\mathbf{f}^t$  from iteration  $t$  to  $t + 1$  according to the update equation

$$\mathbf{f}^{t+1} = \mathbf{f}^t - \mu \left( \nabla_{\mathbf{f}} \mathcal{L}(\mathbf{f}^t, \mathbf{g}) + \lambda \nabla_{\mathbf{f}} R_{\text{AQuaSI}}(\mathbf{f}^t) \right) , \quad (11)$$

where  $\nabla_{\mathbf{f}}$  denotes the gradient calculated at  $\mathbf{f} = \mathbf{f}^t$  and  $\mu$  is the iteration step size parameter. To calculate the gradient of the AQuaSI regularization term for this update rule,  $Q_{p,\mathbf{w}}(\mathbf{f})$  is constructed



with the pseudo-linear form in Eq. 9 around the current estimate  $\mathbf{f}^t$  such that  $Q_{p,\mathbf{w}}(\mathbf{f}^t) = \mathbf{Q}\mathbf{f}^t$ . Based on this linearization, the gradient of the AquaSI term is

$$\begin{aligned}\nabla_{\mathbf{f}} R_{\text{AquaSI}}(\mathbf{f}^t) &= \nabla_{\mathbf{f}} \|\mathbf{f} - Q_{p,\mathbf{w}}(\mathbf{f})\|_1 \big|_{\mathbf{f}=\mathbf{f}^t} \\ &= (\mathbf{I} - \mathbf{Q})^\top \text{sign}(\mathbf{f} - \mathbf{Q}\mathbf{f}^t) .\end{aligned}\quad (12)$$

To handle the non-differentiability of the  $L_1$  norm at zero, we use an  $\epsilon$ -smoothed version with  $\text{sign}(u) \approx u(\sqrt{u^2 + \epsilon})^{-1}$  for small  $\epsilon$ .

The above iteration scheme can be employed with fixed step size parameter  $\mu$  or line searches to determine  $\mu$  dynamically per iteration. Moreover, we can revise the selection of the search direction from the steepest descent approach to related techniques like conjugate gradient (CG) iterations.

## 4.2 Variable Splitting

AquaSI can also be used in variable splitting techniques for minimization of Eq. 2. As one of the most popular techniques within this class of optimizers, we study the alternating direction method of multipliers (ADMM) [45], where Eq. 2 is rewritten as the constrained optimization problem

$$\begin{aligned}(\hat{\mathbf{f}}, \hat{\mathbf{u}}) &= \underset{\mathbf{f}, \mathbf{u}}{\text{argmin}} \{ \mathcal{L}(\mathbf{f}, \mathbf{g}) + \lambda \|\mathbf{u}\|_1 \} \\ \text{s.t. } \mathbf{u} &= \mathbf{f} - Q_{p,\mathbf{w}}(\mathbf{f})\end{aligned}\quad (13)$$

with auxiliary variable  $\mathbf{u}$ . This is equivalent to enforcing the non-linear constraint  $\mathbf{u} = \mathbf{f} - Q_{p,\mathbf{w}}(\mathbf{f})$  by minimizing the Augmented Lagrangian

$$\begin{aligned}(\hat{\mathbf{f}}, \hat{\mathbf{u}}) &= \underset{\mathbf{f}, \mathbf{u}}{\text{argmin}} \left\{ \mathcal{L}(\mathbf{f}, \mathbf{g}) + \lambda \|\mathbf{u}\|_1 \right. \\ &\quad \left. + \frac{\alpha}{2} \|\mathbf{u} - \mathbf{f} + Q_{p,\mathbf{w}}(\mathbf{f}) - \mathbf{b}_u\|_2^2 \right\} ,\end{aligned}\quad (14)$$

where  $\alpha$  and  $\mathbf{b}_u$  denote the corresponding Lagrangian multiplier and a Bregman variable, respectively.

The minimization of Eq. 14 is accomplished in an alternating manner

$$\begin{aligned}\mathbf{f}^{t+1} &= \underset{\mathbf{f}}{\text{argmin}} \left\{ \mathcal{L}(\mathbf{f}, \mathbf{g}) \right. \\ &\quad \left. + \frac{\alpha}{2} \|\mathbf{u}^t - \mathbf{f} + Q_{p,\mathbf{w}}(\mathbf{f}) - \mathbf{b}_u^t\|_2^2 \right\} ,\end{aligned}\quad (15)$$

$$\begin{aligned}\mathbf{u}^{t+1} &= \underset{\mathbf{u}}{\text{argmin}} \left\{ \lambda \|\mathbf{u}\|_1 \right. \\ &\quad \left. + \frac{\alpha}{2} \|\mathbf{u} - \mathbf{f}^{t+1} + Q_{p,\mathbf{w}}(\mathbf{f}^{t+1}) - \mathbf{b}_u^t\|_2^2 \right\} ,\end{aligned}\quad (16)$$

$$= \text{shrink} \left( \mathbf{f}^{t+1} + Q_{p,\mathbf{w}}(\mathbf{f}^{t+1}) + \mathbf{b}_u^t, \frac{\lambda}{\alpha} \right) ,\quad (17)$$

$$\mathbf{b}_u^{t+1} = \mathbf{b}_u^t + (\mathbf{f}^{t+1} - Q_{p,\mathbf{w}}(\mathbf{f}^{t+1}) - \mathbf{u}^{t+1}) ,\quad (18)$$

where  $\text{shrink}(u, \gamma) = \text{sign}(u) \max(u - \gamma, 0)$  is the shrinkage operator associated with the  $L_1$  norm  $\|\mathbf{u}\|_1$ . To reduce the influence of the selection of the Lagrangian multiplier, we adopt varying penalty parameters [46]. The weighted quantile filter  $Q_{p,\mathbf{w}}(\mathbf{f})$  is again approximated by the pseudo-linear form in Eq. 9.

## 4.3 Multi-Channel Inverse Problems

The AquaSI prior can be easily generalized to handle multi-channel data like RGB color or multi-spectral images. Applying this model to each channel separately is straightforward but would ignore

inter-channel correlations like structures present in all channels. Multi-channel images can be jointly optimized in order to allow for simultaneous processing of all channels. To this end, we couple all channels in the AQuaSI prior.

Let  $\mathbf{f}_c \in \mathbb{R}^N$ ,  $c = 1, \dots, C$  be the individual channels of a multi-channel image denoted in vector notation. Based on Eq. 2, we plug the AQuaSI prior into a multi-channel inverse problem according to:

$$\underset{\mathbf{f}_1, \dots, \mathbf{f}_C}{\operatorname{argmin}} \left\{ \sum_{c=1}^C \mathcal{L}(\mathbf{f}_c, \mathbf{g}_c) + \lambda \|\bar{\mathbf{f}} - Q_{p, \mathbf{w}}(\bar{\mathbf{f}})\|_1 \right\} , \quad (19)$$

where  $\bar{\mathbf{f}}$  denotes the weighted average:

$$\bar{\mathbf{f}} = \sum_{c=1}^C m_c \mathbf{f}_c . \quad (20)$$

For instance, in case of RGB color images, using the weights  $\mathbf{m} = (0.299, 0.587, 0.114)^\top$  is equivalent to a standard grayscale conversion. We replace the non-linear filter  $Q_{p, \mathbf{w}}(\cdot)$  with its pseudo-linear form  $\bar{\mathbf{Q}}$  computed from  $\bar{\mathbf{f}}$  to exploit inter-channel correlations and to allow for better structure preservation. Then, we can resort the AQuaSI prior on the average image:

$$\begin{aligned} \|\bar{\mathbf{f}} - Q_{p, \mathbf{w}}(\bar{\mathbf{f}})\|_1 &= \left\| \sum_{c=1}^C m_c \mathbf{f}_c - Q_{p, \mathbf{w}} \left( \sum_{c=1}^C m_c \mathbf{f}_c \right) \right\|_1 \\ &\approx \left\| \sum_{c=1}^C m_c \mathbf{f}_c - \sum_{c=1}^C m_c \bar{\mathbf{Q}} \mathbf{f}_c \right\|_1 \\ &= \left\| \sum_{c=1}^C m_c (\mathbf{f}_c - \bar{\mathbf{Q}} \mathbf{f}_c) \right\|_1 . \end{aligned} \quad (21)$$

Thus, we apply a single linearization  $\bar{\mathbf{Q}}$  to the individual channels of the underlying inverse problem. All channels are stacked in the variables  $\mathbf{f}$  and  $\mathbf{g}$ :

$$\begin{aligned} \mathbf{f} &= (\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_C)^\top , \\ \mathbf{g} &= (\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_C)^\top . \end{aligned} \quad (22)$$

Then, the objective function in Eq. 19 can be rewritten as

$$\hat{\mathbf{f}} = \underset{\mathbf{f}}{\operatorname{argmin}} \{ \mathcal{L}(\mathbf{f}, \mathbf{g}) + \lambda \|\mathbf{M}(\mathbf{f} - \mathbf{Q}\mathbf{f})\|_1 \} , \quad (23)$$

where  $\mathbf{Q} = \operatorname{diag}(\bar{\mathbf{Q}}, \dots, \bar{\mathbf{Q}})$ ,  $\mathbf{M} = \operatorname{diag}(\mathbf{M}_1, \dots, \mathbf{M}_C)$ , and  $\mathbf{M}_c = \operatorname{diag}(m_c, \dots, m_c) \in \mathbb{R}^{CN \times CN}$ . The optimization problem in Eq. 23 can be solved using the algorithms described in Sec. 4.2.

We compare the channel-wise and multi-channel (MC) approach for RGB/NIR image restoration (see Sec. 5.4) on an example of the ARRI dataset [47] in Fig. 4. In general, the advantage of the multi-channel approach compared to a channel-wise optimization is two-fold: 1) AQuaSI MC yields better structure preservation than AQuaSI channel-wise. For example, the wooden sticks at the back of the image are less blurred and the gaps between the sticks are better visible. 2) AQuaSI MC requires less computation time than simple channel-wise processing. This is explained by the fact that only a single linearization  $\bar{\mathbf{Q}}$  is required per iteration. In contrast, channel-wise processing needs to construct one linearization per image channel. In Fig. 4, AQuaSI MC converged after 388 seconds, whereas the channel-wise approach converged after 6678 seconds on an image of size  $1575 \times 2800$  pixels.

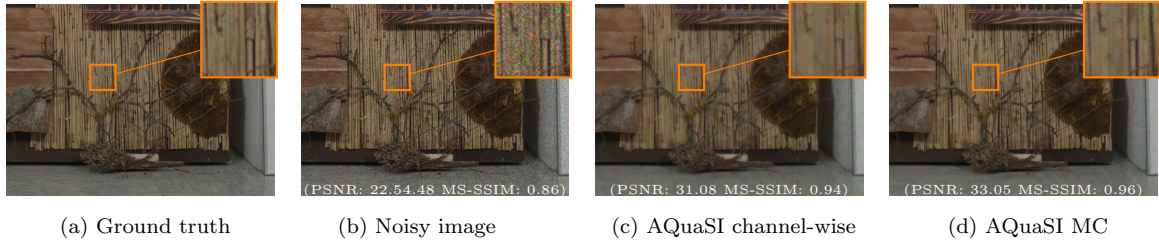


Figure 4: RGB/NIR restoration on the ARRI dataset [47]. (a) Ground truth color image, (b) color image corrupted by speckle noise, (c) AQuaSI channel-wise with *static* guidance, and (d) AQuaSI multi-channel with *static* guidance.

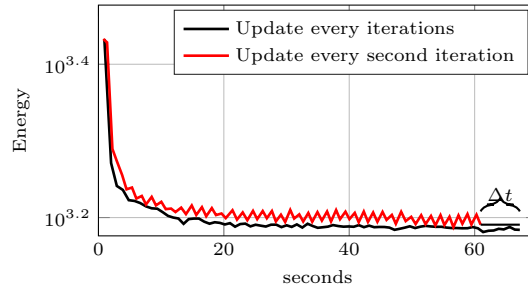


Figure 5: Convergence analysis for iterative optimization with the AQuaSI prior. We evaluated the overall energy function optimized by ADMM using an update of the AQuaSI pseudo-linear form after every iteration and after every second iteration. The time gap  $\Delta t$  indicates the performance improvement when updating every second iteration for images of size  $256 \times 256$  pixels. This improvement increases with the image resolution.

#### 4.4 Implementation Remarks

The iterative algorithms with our AQuaSI prior are easy to implement and can be seamlessly integrated into existing inverse imaging frameworks. There are two main steps at each iteration: 1) estimation of the image  $\mathbf{f}^{t+1}$ , and 2) the construction of the linearization  $\mathbf{Q}$  for weighted quantile filtering  $Q_{p,\mathbf{w}}(\mathbf{f})$ . While the first step involves efficient matrix/vector operations, the computational complexity of an iteration is mainly related to the second step.

To decrease the computation time, we found that it is sufficient to re-use the same  $\mathbf{Q}$  for several iterations instead of updating  $\mathbf{Q}$  in every iteration. In this case, the pseudo-linear form is an approximation of the actual filtering. We validate this approximation experimentally by examining the convergence of our iteration scheme. Figure 5 depicts the convergence of ADMM in terms of the optimized energy for RGB/NIR restoration (see Sec. 5.4) versus the computation time of the algorithm. We evaluated the convergence on image patches of size  $256 \times 256$  pixels. Here, we compare an update of  $\mathbf{Q}$  at every iteration against an update after every second iteration. It is worth noting that both variants convergences to comparable local minima, while the approximation is considerably faster in computation. Especially for large-scale images, significantly lower processing time are possible.

## 5 Experimental Evaluation and Applications

In order to emphasize the effectiveness of the AQuaSI prior, we relate it to TV regularization and RED algorithms. Additionally, we evaluate the performance of the proposed method on two common

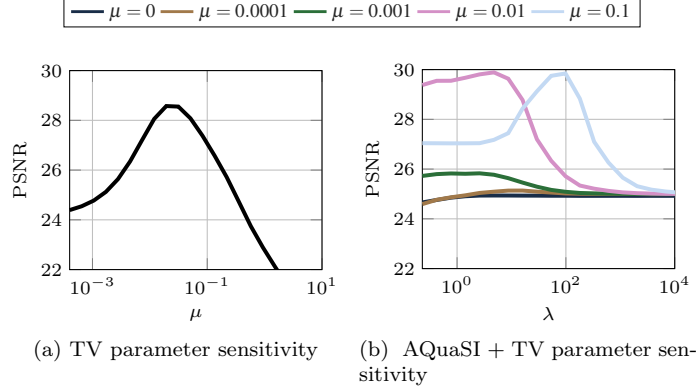


Figure 6: Parameter sensitivity analysis for the TV prior with regularization weight  $\mu$  (a) and the combination of AquaSI with regularization weight  $\lambda$  and TV (b) on the RGB/NIR image restoration result.

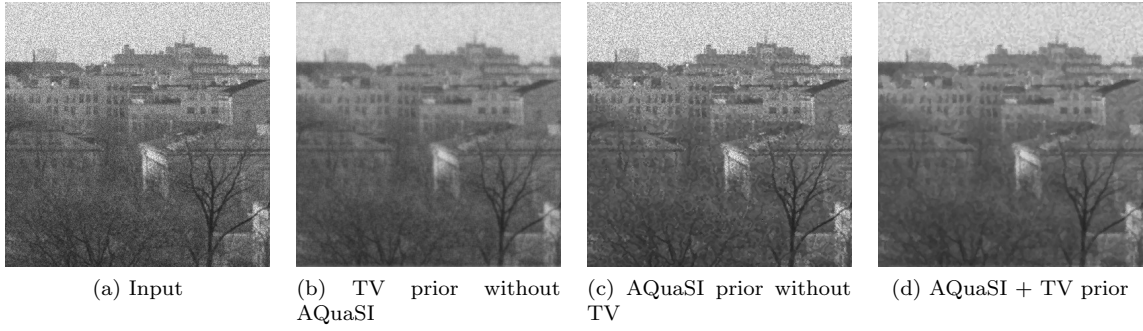


Figure 7: Influence of the TV prior and the AquaSI prior to image denoising. We compare a noisy input (a) to denoised images using only the TV prior (b), only AquaSI (c), as well as the combination TV + AquaSI (d). For each configuration, we report denoised images with the highest PSNR.

inverse problems in computer vision, namely joint RGB/depth map upsampling and RGB/NIR image restoration.

## 5.1 Comparison to Total Variation Regularization

We investigate the influence of the AquaSI prior with *dynamic* guidance for RGB/NIR image restoration (see Sec. 5.4) and compare it to the influence of the TV prior. We varied the regularization weights of the TV and AquaSI prior in Eq. 27 separately, while setting the other one to zero. We additionally combined AquaSI with TV for varying regularization weight of the TV prior.

Figure 6 shows the peak signal-to-noise ratio (PSNR) for TV and AquaSI with different regularization weights. A small regularization weight  $\mu$  for the TV prior effectively removes the term from the optimization. Increasing  $\mu$  for the TV prior beyond the optimum point drastically decreases the performance again, since TV starts to oversmooth the results. Changing the weight for AquaSI from  $\lambda = 0$  to a larger value corresponds to activating it. The jump in the activation can be steered by the choice of  $\sigma_w$ , the standard deviation of the Gaussian weighting function. AquaSI smooths the image with large  $\sigma_w$  and allows for better structure preservation with smaller  $\sigma_w$ . For this experiment, we choose  $n = 3^2$  and  $\sigma_w = 0.1$ , which is rather low. Thus, compared to TV, AquaSI does not achieve a competitive denoising performance. The advantage of this configuration of AquaSI can be seen in combination with the TV prior in Fig. 7. The smoothing property of TV is retained and AquaSI helps to preserve the original image content. This property of AquaSI is underlined in the following

Table 1: Averaged peak signal to noise ratio (PSNR) and no-reference deblurring metric (DM) [49] for non-blind deblurring using RED [8] with WMF [17] or BM3D denoising [51], and AQuaSI on the Levin dataset [48]. Results are reported for different noise levels of additive white Gaussian noise (AWGN) and speckle noise (SN).

	Level $\sigma_{\text{noise}}$	RED with WMF		RED with BM3D		AQuaSI	
		PSNR	DM	PSNR	DM	PSNR	DM
AWGN	0.0001	29.40	-13.96	<b>30.62</b>	-14.11	30.47	<b>-12.94</b>
	0.0009	27.12	-14.02	<b>28.41</b>	-14.60	27.51	<b>-12.77</b>
	0.0025	25.87	<b>-13.10</b>	<b>27.32</b>	-15.46	26.34	-13.36
SN	0.0001	33.69	-13.13	<b>35.04</b>	-13.48	33.92	<b>-12.47</b>
	0.0009	30.04	-13.05	<b>30.87</b>	-14.23	30.25	<b>-12.30</b>
	0.0025	28.41	-12.89	28.58	-14.93	<b>28.76</b>	<b>-12.10</b>

sections, where AQuaSI drastically improves the performance of common image priors. We provide additional qualitative results for comparison to TV regularization in the supplementary material.

## 5.2 Comparison to Regularization by Denoising (RED)

AQuaSI is similar in spirit to the RED framework proposed by Romano *et al.* [8]. Using a weighted median filter  $Q_{p,\mathbf{w}}(\cdot)$ ,  $p = 0.5$  as denoising engine in the same variant as in the proposed prior, RED is modeled by the regularization term

$$R_{\text{RED}}(\mathbf{f}) = \mathbf{f}^\top (\mathbf{f} - Q_{p,\mathbf{w}}(\mathbf{f})) . \quad (24)$$

That is, instead of regularizing with the  $L_1$  norm, RED uses the square of the image  $\mathbf{f}$ . For a comparison between these models, we apply both on denoising and deblurring tasks.

We evaluated AQuaSI and RED for non-blind deblurring using the dataset of Levin *et al.* [48] comprising eight blur kernels and four ground truth grayscale images. We degraded the ground truths by convolutions with the known blur kernels. In addition, each blurred image is corrupted by additive white Gaussian noise and speckle noise ( $\sigma_{\text{noise}} \in \{0.0001, 0.009, 0.0025\}$  for each model and images in  $[0, 1]$ ). For deblurring, we applied AQuaSI with *dynamic* guidance and RED with the weighted median filter ( $n = 5^2$ ,  $\sigma_w = 0.6$ ) and BM3D in the framework provided by Romano *et al.* [8]. The deblurring quality was assessed by the PSNR and the no-reference motion deblurring metric (DM) [49]. While PSNR measures distortions relative to the ground truth, DM is specifically designed to assess the perceptual quality of the deblurring methods and lower values express better performance. Quantitative results for both noise models are reported in Tab. 1. RED with BM3D outperforms AQuaSI w. r. t. PSNR. However AQuaSI excels in the deblurring metric DM. Here, AQuaSI overall outperforms RED with WMF and RED with BM3D especially under heavy speckle noise that is difficult to handle by of-the-shelf denoisers. Analogously to the other image restoration problems (cf. Sec. 5.4), there is an inherent tradeoff between signal distortion and perceptual quality [50]. In our experiments, AQuaSI performs reasonably in terms of distortions (second best in PSNR) while providing excellent perceptual quality (best in DM) leading to good perception-distortion tradeoffs. This is also confirmed in the qualitative results in Fig. 8.

Figure 9 depicts RED versus the proposed AQuaSI prior for RGB/NIR image restoration (see Sec. 5.4). Here, NIR data guides the restoration of noisy RGB images that are degraded by speckle noise ( $\sigma_{\text{noise}} = 0.2$ ). For fair comparison we set  $\mu = 0$ , thus eliminate the influence of the TV prior. We couple both priors with an  $L_2$  norm data term and use the weighted median filter ( $n = 9^2$ ,  $\sigma_w = 0.006$ ) in *static* guidance mode within ADMM optimization. It is worth noting that AQuaSI yields considerably better perceptual quality in terms of noise reduction as well as a higher PSNR. We explain this behavior by the different models underlying the priors: RED with its quadratic model has difficulties in handling non-Gaussian noise on its own, while AQuaSI with the  $L_1$  norm



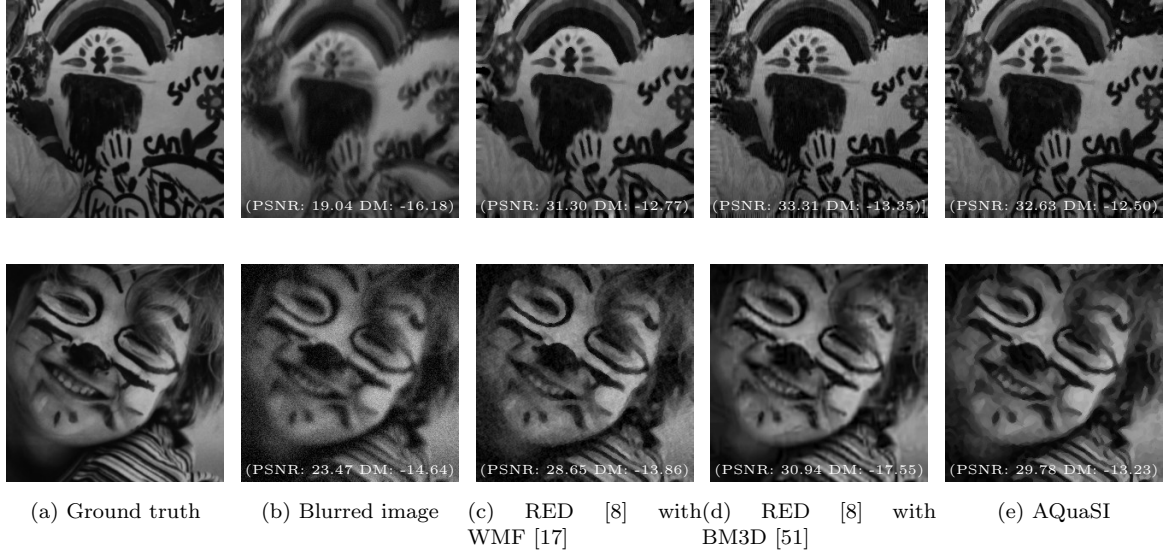


Figure 8: Qualitative deblurring results for two images of the Levin dataset [48]. The top row shows ground truth data, blurry input data corrupted by additive white Gaussian noise with  $\sigma_{\text{noise}} = 0.0001$  and deblurring results of different algorithms. The bottom row depicts qualitative results for  $\sigma_{\text{noise}} = 0.025$ .

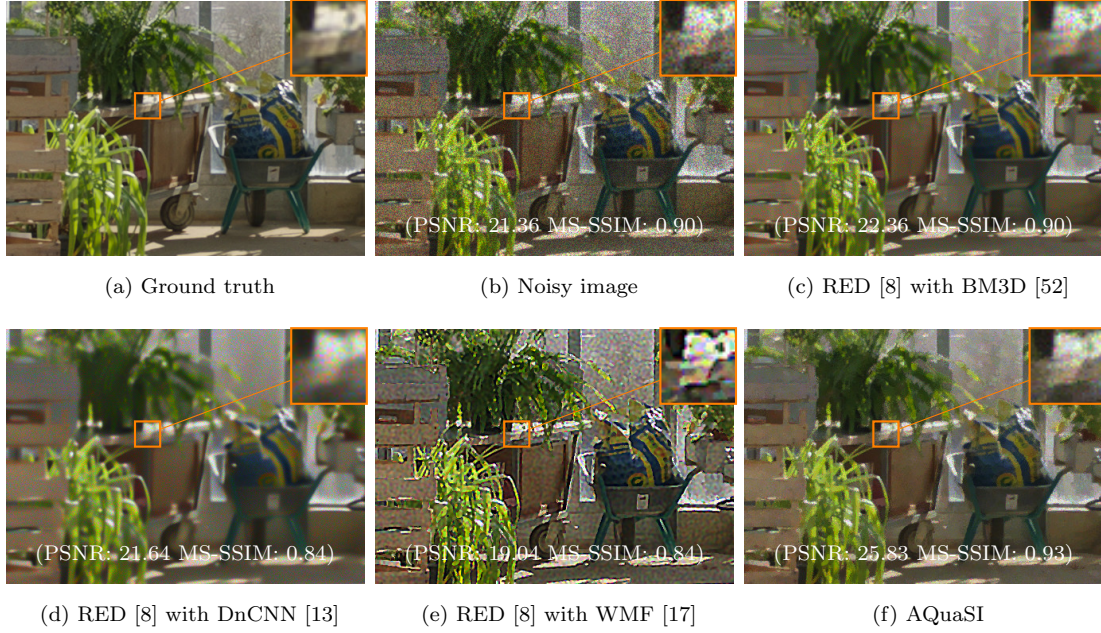


Figure 9: RGB/NIR restoration using our AQUaSI prior under non-Gaussian noise in comparison to regularization by denoising (RED) [8]. (a) Ground truth color image, (b) color image corrupted by speckle noise, (c) RED with weighted median filter (WMF), and (f) AQUaSI prior. The priors are steered by NIR data in a *static* guidance mode. The reported PSNR and MS-SSIM values are calculated from the sections to be seen.

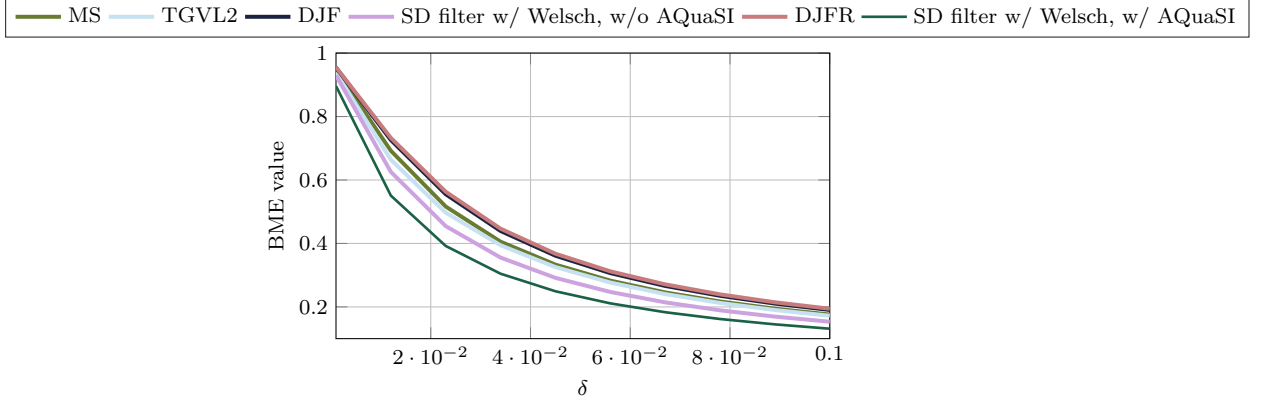


Figure 10: Bad matching error (BME) of various joint RGB/depth map upsampling methods at different error thresholds  $\delta$  on the Middlebury data.

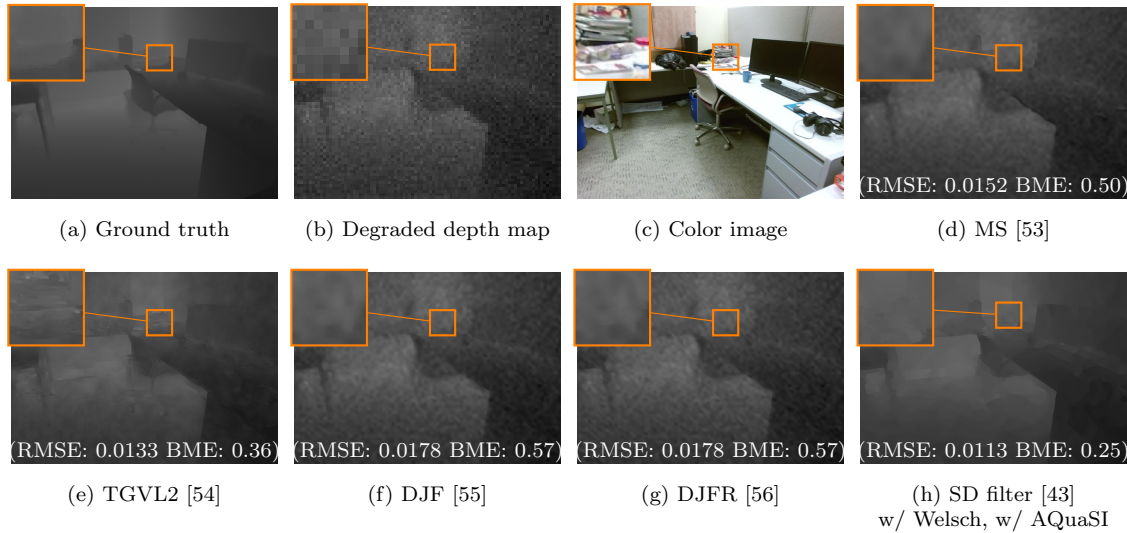


Figure 11: Joint RGB/depth map upsampling on a SUN image pair. (a) Ground truth depth map, (b) degraded depth map, (c) color image, (d) - (h) results of mutual structure (MS) filter, TGVL2 upsampling, deep joint filter (DJF), residual-based deep joint filter (DJFR), and SD filter w/ Welsch, w/ AQUaSI. RMSE and BME, with  $\delta = 0.01$ , are reported for each result.

model is less sensitive in that case. This also confirms the properties of our sparsity-promoting model shown in Sec. 3.2. We additionally report the results of RED without guidance data using BM3D [52], DnCNN [13] as denoiser. While DnCNN tends to oversmooth the result, BM3D provides a better tradeoff between noise reduction and structure preservation.

### 5.3 Joint Upsampling

In joint upsampling, we aim at upsampling a noisy depth map under the guidance of an RGB image. To demonstrate the applicability of the proposed method for this task, we augment the popular joint static and dynamic (SD) filter [43] by our AQUaSI prior. Given a noisy depth map  $\mathbf{g}$  and a complementary RGB color image  $\mathbf{z}$ , we obtain an upsampled depth map  $\mathbf{f}$  according to the inverse



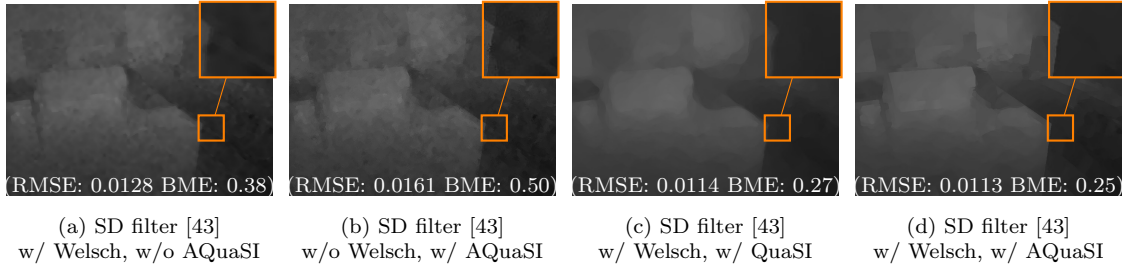


Figure 12: Comparison of the different variants of the SD filter [43] on a SUN image pair. (a) SD filter w/ Welsch, w/o AQuaSI, (b) SD filter w/o Welsch, w/ AQuaSI, (c) SD filter w/ Welsch, w/ QuaSI, (d) SD filter w/ Welsch, w/ AQuaSI. RMSE and BME with  $\delta = 0.01$  are reported for each result.

Table 2: Averaged root mean square error (RMSE) for RGB/depth map upsampling using different general-purpose (MS [53], DJF [55], and DJFR [56]) and customized methods for depth maps (TGVL2 [54]) on the Middlebury, NYU, and SUN benchmark datasets.

Method	RMSE Middlebury	RMSE NYU	RMSE SUN
MS [53]	0.0245	0.0167	0.0190
TGVL2 [54]	0.0245	0.0169	0.0213
DJF [55]	0.0266	-	0.0218
DJFR [56]	0.0273	-	0.0218
SD filter [43] w/ Welsch, w/o AQuaSI	0.0220	0.0143	0.0178
SD filter [43] w/o Welsch, w/ AQuaSI	0.0253	0.0176	0.0205
SD filter [43] w/ Welsch, w/ QuaSI	0.0211	0.0136	0.0177
SD filter [43] w/ Welsch, w/ AQuaSI	<b>0.0197</b>	<b>0.0124</b>	<b>0.0165</b>

problem

$$\hat{\mathbf{f}} = \underset{\mathbf{f}}{\operatorname{argmin}} \mathcal{L}(\mathbf{f}, \mathbf{g}) + \underbrace{\mu R(\mathbf{f}, \mathbf{z})}_{\text{SD prior}} + \underbrace{\lambda \|\mathbf{f} - Q_{p, \mathbf{w}}(\mathbf{f})\|_1}_{\text{AQuaSI prior}}, \quad (25)$$

where  $R(\mathbf{f}, \mathbf{z})$  denotes the Welsch function as proposed in [43] with regularization weight  $\mu \geq 0$ . The data fidelity term is defined analogously to [43] as  $\mathcal{L}(\mathbf{f}, \mathbf{g}) = \|\mathbf{c} \odot (\mathbf{f} - \mathbf{g})\|_2^2$ , where  $\mathbf{c}$  denotes a confidence map to mask out outliers, e.g., invalid depth measurements, and  $\odot$  is the Hadamard (element-wise) product. In order to solve Eq. 25, we perform iteratively reweighted least squares as in [43], which essentially is a variant of gradient descent as introduced in Sec. 4.1.

We study joint upsampling on the Middlebury dataset (using the *Art*, *Books*, *Dolls*, *Laundry*, *Moebius* and *Reindeer* depth maps), the NYU dataset [57], and the SUN dataset [58–60]. We generated synthetic depth maps normalized in  $[0, 1]$  from their ground truth by Gaussian blurring with standard deviation  $\sigma_{\text{blur}} = 4$  followed by nearest neighbor downsampling with a factor of 8 and by adding Gaussian noise with standard deviation  $\sigma_{\text{noise}} = 0.0005$ . Joint upsampling is assessed quantitatively with the Root Mean Square Error (RMSE) between the ground truth depth map  $\tilde{\mathbf{f}}$  and a recovered depth map  $\mathbf{f}$ . Additionally, we report the Bad Matching Error (BME) [61]

$$E_{\text{BME}} = \frac{1}{N} \sum (|f_i - \tilde{f}_i| > \delta), \quad (26)$$

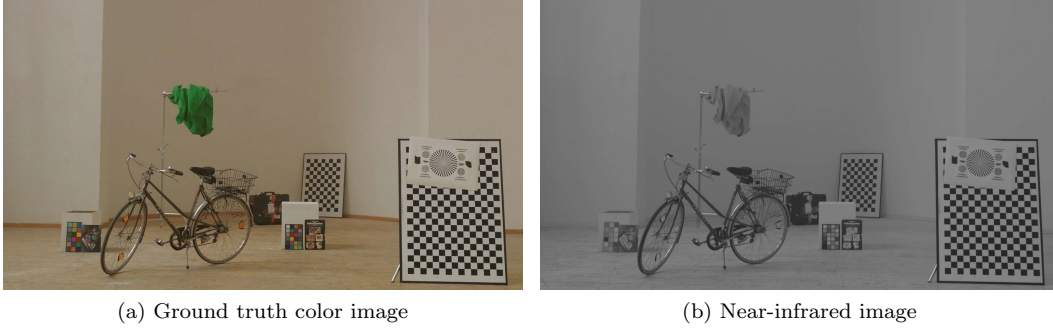


Figure 13: Example color and near-infrared image from the ARRI dataset [47].

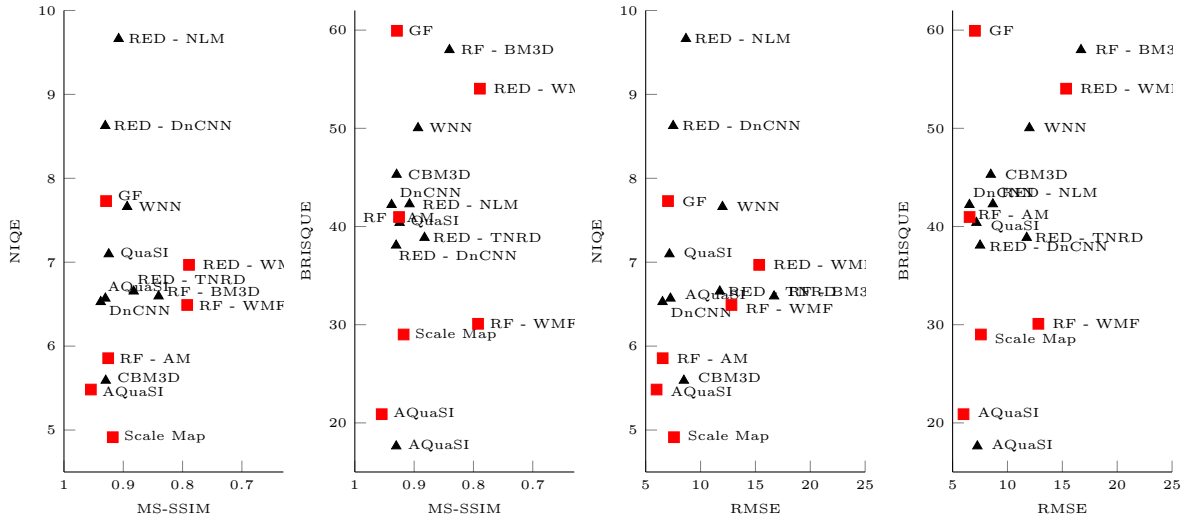


Figure 14: Quantitative results of all competing methods on the ARRI dataset using the perception-distortion plane to visualize BRISQUE and NIQE over RMSE and MS-SSIM. The red squares indicate methods with near-infrared guidance and the black triangles indicate methods without near-infrared guidance. All variants of QuaSI and AquaSI use the multi-channel mode.

where  $N$  are the number of pixels in a depth map and  $\delta$  denotes a scalar that can be set accordingly.

The SD filter with the AquaSI prior according to Eq. 25 (SD filter, w/ AquaSI) is compared to TGVL2 upsampling [54], the mutual structure filter (MS) [53], the deep joint filter (DJF) [55], and the residual-based deep joint filter (DJFR) [56]. We also conducted an ablation study using Eq. 25 and evaluated different variants of the proposed joint upsampling by omitting individual regularization terms (SD filter w/ Welsch and w/o AquaSI, SD filter w/o Welsch and w/ AquaSI) and by replacing AquaSI with its non-adaptive counterpart (SD filter w/ Welsch and w/ QuaSI). The parameters of AquaSI were set to  $p = 0.5$ ,  $n = 9^2$ ,  $\sigma_w = 0.1$ , and  $\lambda = 0.1$  and the parameters of the competing methods were chosen according to suggestions of the authors.

Table 2 shows the average RMSEs on the three datasets achieved by the different methods<sup>2</sup>. Notice that the combination of the SD filter with the AquaSI prior yields the lowest errors in all benchmarks. In comparison to the baseline SD filter (w/ Welsch and w/o AquaSI), SD w/ Welsch, w/ AquaSI reduces the RMSE by 10.5%, 13.3%, and 7.3% in the Middlebury, NYU, and SUN

<sup>2</sup>DJF and DJFR are not evaluated on the NYU dataset, since they used the NYU dataset for training.

benchmark, respectively. Figure 10 shows the mean BME relative to the ground truth depth maps for different  $\delta$  values on the Middlebury data. The AQuaSI prior further improves the results of the SD filter and outperforms all competing methods.

Figure 11 shows qualitative results of the competing upsampling approaches. The MS and DJFR filters were sensitive to noise. In contrast, TGVL2 erroneously transfers textures from the color image to the depth map. An example can be seen in Fig. 11e, where the layers of the books cause artifacts in the depth map. The SD filter w/ Welsch and w/ AQuaSI achieves the best upsampling performance, which is consistent with the quantitative results. The SD filter w/ Welsch and w/ QuaSI leads to a good noise reduction but causes to oversmoothing as shown in Fig. 12c. While the SD filter w/o Welsch and w/ AQuaSI does not remove noise as effectively as SD filter w/ Welsch and w/o AQuaSI, edge preservation is greatly improved. This observation underlines the reported structure preservation property of the AQuaSI prior.

## 5.4 RGB/NIR Cross Field Image Restoration

Another popular application is color (RGB) and near-infrared (NIR) image restoration. The scene is captured using a RGB and a NIR sensor. While NIR images in such setups are oftentimes flashed and feature high signal-to-noise ratios, the RGB data can be noisy. Here, we aim at denoising RGB images. AQuaSI is plugged into classical TV denoising [45]. That is, given a noisy RGB image  $\mathbf{g}$ , we recover its clean counterpart  $\mathbf{f}$  according to

$$\begin{aligned} \hat{\mathbf{f}} = \underset{\mathbf{f}}{\operatorname{argmin}} \mathcal{L}(\mathbf{f}, \mathbf{g}) + \underbrace{\mu (\|\nabla_x \mathbf{f}\|_1 + \|\nabla_y \mathbf{f}\|_1)}_{\text{TV prior}} \\ + \underbrace{\lambda \|\mathbf{f} - Q_{p,\mathbf{w}}(\mathbf{f})\|_1}_{\text{AQuaSI prior}}, \end{aligned} \quad (27)$$

where  $\nabla_x \mathbf{f}$  and  $\nabla_y \mathbf{f}$  denote the derivatives of  $\mathbf{f}$  in  $x$ - and  $y$ -direction to define the TV prior with regularization weight  $\mu \geq 0$  and  $\mathcal{L}(\mathbf{f}, \mathbf{g}) = \|\mathbf{f} - \mathbf{g}\|_2^2$  models the data fidelity. The optimization of Eq. 27 is performed via ADMM (see Sec. 4.2). We apply AQuaSI with  $p = 0.5$ ,  $n = 9^2$ , and  $\sigma_w = 0.006$  and the TV prior with  $\mu = 0.05$  and  $\beta = 7$  for the corresponding auxiliary variable. We compare the channel-wise approach of AQuaSI with  $\lambda = 13$  and  $\alpha = 1100$  to the multi-channel (MC) mode to simultaneously denoise three color channels with  $\lambda = 13 \cdot 3$ ,  $\alpha = 1100 \cdot 3$ .

### 5.4.1 With Near-Infrared Guidance

First, we study the case, where NIR images are used to restore clean RGB images using the *static* guidance mode of our prior. The evaluation is performed on eleven registered ground truth RGB/NIR image pairs taken from the publicly available ARRI dataset [47] as shown in Fig. 13. We simulate noisy color images by adding speckle noise (variance  $\sigma_{\text{noise}} = 0.2$ ) to the respective ground truth images. We compare AQuaSI against two state-of-the-art static guidance based methods, namely guided filtering [42] and cross-field restoration via scale map [44]. In addition, we adopt regularization by denoising (RED) [8] and zero-order reverse filtering (RF) [62]. Specifically, we equipped RED with the weighted median filter [17] (RED with WMF) and the RF framework with the weighted median filter (RF with WMF) as well as adaptive manifold filtering [63] (RF with AM).

For a quantitative comparison, the methods are plotted on a perception-distortion plane using the evaluation framework by Blau and Michaeli [50]. The distortion is quantified by the RMSE and multiscale structural similarity (MS-SSIM) [64]. The perceptual quality is quantified by the natural image quality evaluator (NIQE) [65] and the blind/referenceless image spatial quality evaluator (BRISQUE) [66]. These metrics are arranged in the perception-distortion plane in a way that better methods are closer to the bottom left of the plane. For BRISQUE and NIQE, low values imply high image quality in terms of observed naturalness. The closer MS-SSIM is to the value 1 the higher is the structural similarity between the restored image and the ground truth.

Figure 14 shows the averaged perception-distortion planes. The distortion measure MS-SSIM is used in the first two plots, where AQuaSI performs best. For the associated perceptual metrics NIQE and BRISQUE, the best performer is in one case Scale Map and in the other case AQuaSI. The third and fourth plots evaluate RMSE as a distortion measure and draw a similar picture. Again, AQuaSI performs best w.r.t. RMSE, again only challenged by scale map for the perceptual metric NIQE. Under BRISQUE, AQuaSI performs best. Overall, AQuaSI shows a highly competitive performance, and is in all cases closely located to the bottom left corner.

Figure 15 depicts qualitative results for several methods. The guided filter slightly oversmooths the image. For example, the contours on the bike are barely visible. Conversely, scale map produces a crisp output at the expense of noise and color artifacts. Especially homogeneous areas, such as the checkerboard fields, remain noisy. RED with WMF removes noisy color pixels but cause blob-like artifacts. RF with AM produces color block artifacts and blurs small structures. In contrast, RF with WMF causes grid-like artifacts. Overall, AQuaSI with *static* guidance yields high perceptual image quality conforming the perception-distortion tradeoff analysis.

#### 5.4.2 Without Near-Infrared Guidance

We also examine the more challenging problem of restoring a clean RGB image without guidance by NIR data. For that use case, we compare AQuaSI in its *dynamic* guidance mode to different well known stand-alone image denoising methods, namely weighted nuclear norm minimization (WNN) [26], color block-matching and 3D filtering (CBM3D) [51], and feed-forward denoising convolutional neural networks (DnCNN) [13]. We also adopted the RF approach in combination with BM3D [52] (RF with BM3D) as well as RED [8] with non-local means (NLM) [67], BM3D [52], DnCNN [13], and trainable nonlinear reaction-diffusion (TNRD) [68].

In Fig. 14 the quantitative results are shown. In terms of distortion, AQuaSI with *dynamic* guidance and CBM3D yield similar results. In terms of perceptual quality, AQuaSI with *dynamic* guidance has the best BRISQUE value whereas CBM3D performs better w.r.t. NIQE.

Figure 16 shows the qualitative results of all competing methods that do not use additional guidance data for denoising. WNN produces a sharp output at the expense of colored noise, especially in homogeneous areas. Compared to WNN, CBM3D suppresses more noise, but produces small streaking artifacts in homogeneous areas. While CBM3D does contain more residual noise than DnCNN, small details such as the spokes of the wheel are preserved. QuaSI - MC provides competitive denoising performance w.r.t. distortion, but produces slightly oversmoothed results. AQuaSI with *dynamic* guidance has a similar performance to QuaSI, but avoids color artifacts. To enable a fair comparison, AQuaSI’s smoothing parameter  $\sigma_w$  was optimized for *static* guidance mode. Hence, the performance of the *dynamic* mode can be further improved by tuning  $\sigma_w$  either towards low distortions or good perceptual image quality in the perception-distortion plane.

#### 5.4.3 Comparison of Different AQuaSI Variants

In Table 3, the average RMSE and MS-SSIM values of all four variants of AQuaSI on the ARRI dataset are shown. In terms of distortion measures, the multi-channel approach slightly outperforms channel-wise processings in both guidance modes. We explain this behavior by the avoidance of color artifacts under a simultaneous processing of the RGB channels. As shown in Sec. 4.3, another major advantage of the multi-channel over the channel-wise approach is the lower computation time. Thus, the multi-channel approach is preferable to simple channel-wise processing. Moreover, AQuaSI with *dynamic* guidance is outperformed by the *static* guidance mode. Notice that the former achieves a better recovery of small structures, as shown in Fig. 15 (g). This reveals the benefit of exploiting NIR data for regularizing RGB denoising.

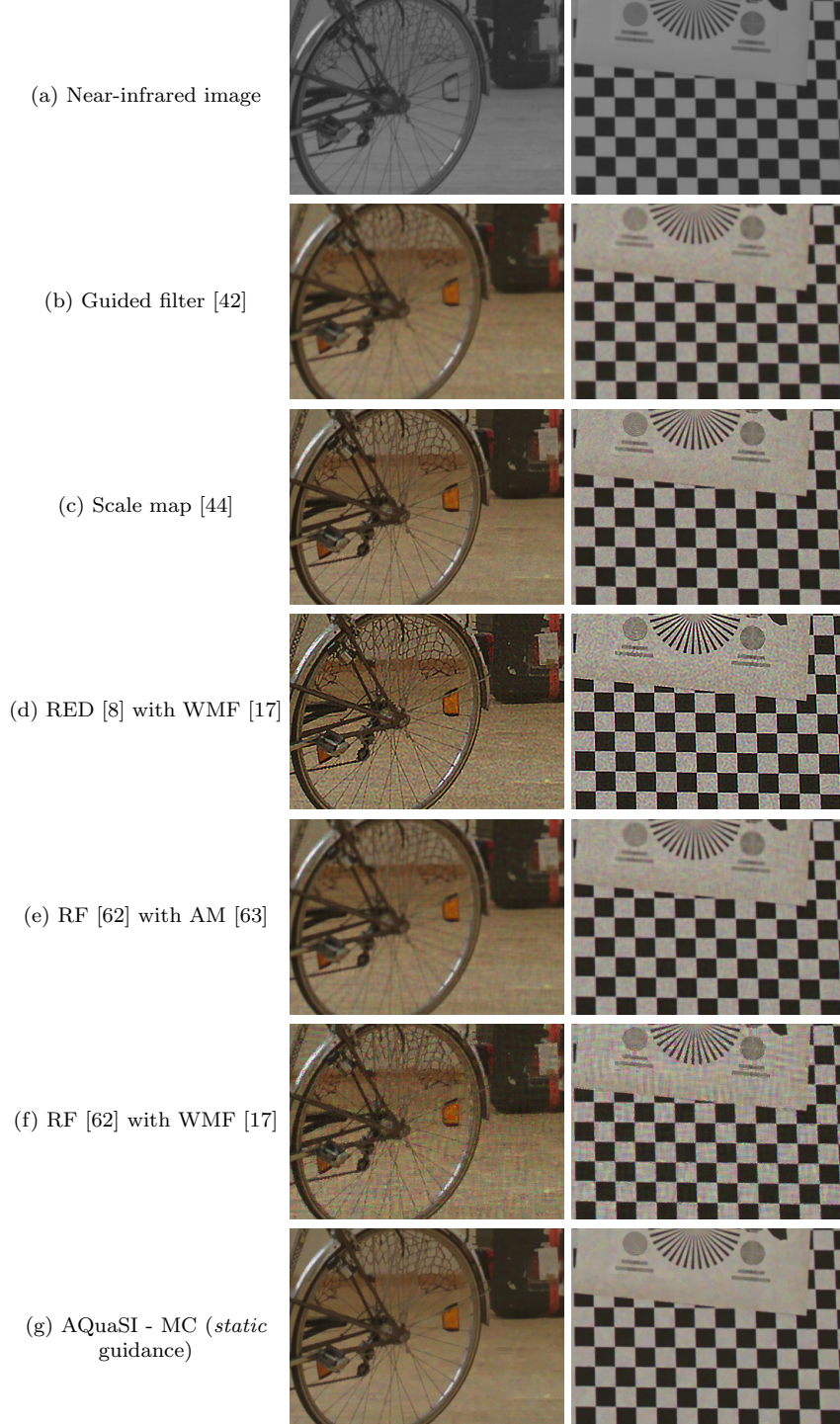


Figure 15: RGB/NIR image restoration on the ARRI dataset [47]. All competing methods enable guided filtering to denoise the degraded color image using the near-infrared image for guidance.



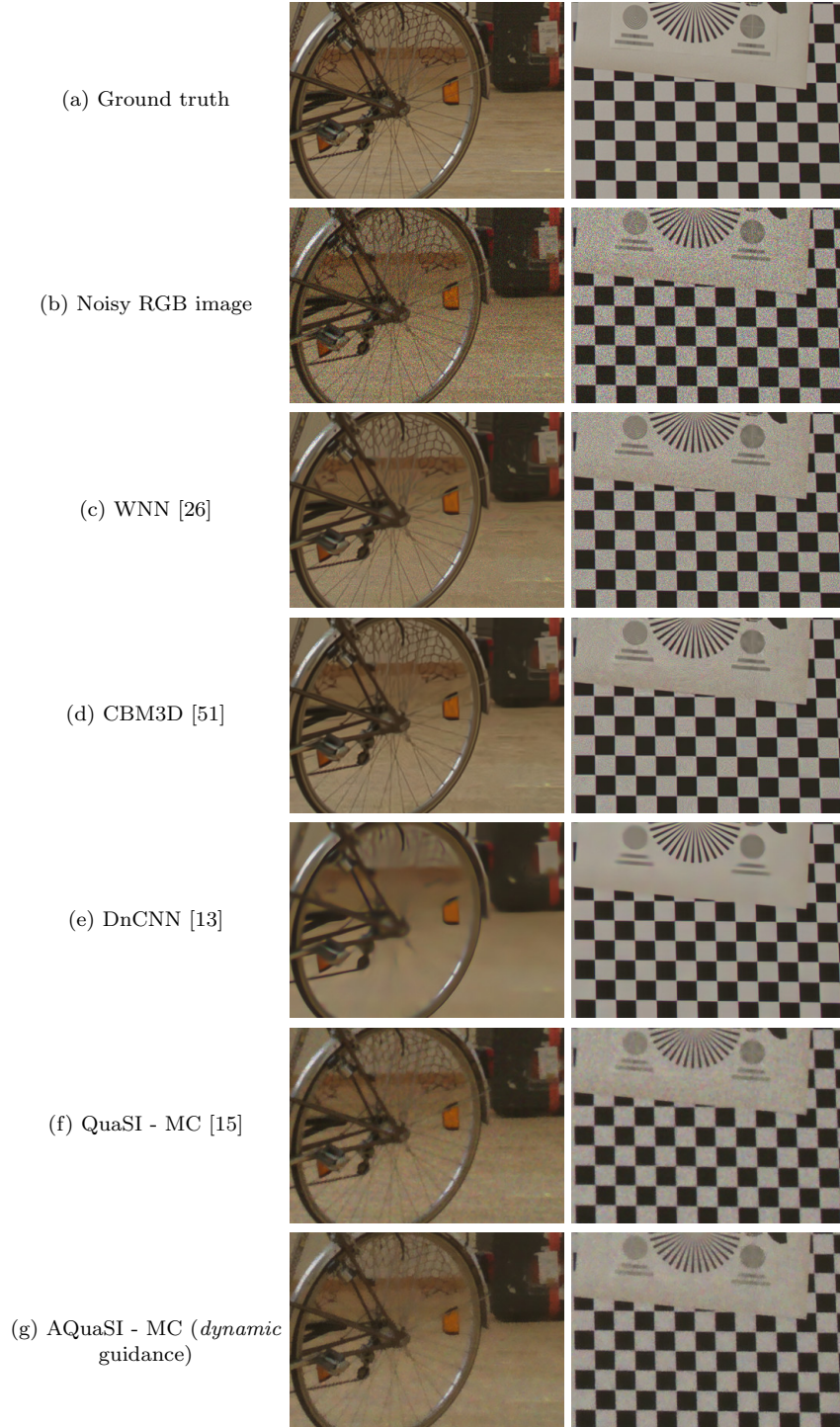


Figure 16: RGB/NIR image restoration on the ARRI dataset [47]. All competing methods perform denoising on the degraded color image only.

Table 3: Averaged RMSE and MS-SSIM for RGB/NIR image restoration using all AQuaSI variants on the ARRI dataset.

Method	RMSE	MS-SSIM
AQuaSI - channel-wise, <i>dynamic</i> guidance	14.89	0.84
AQuaSI - channel-wise, <i>static</i> guidance	6.62	<b>0.95</b>
AQuaSI - multi-channel, <i>dynamic</i> guidance	7.26	0.93
AQuaSI - multi-channel, <i>static</i> guidance	<b>5.99</b>	<b>0.95</b>

## 6 Conclusions

We propose AQuaSI, a novel prior for solving inverse problems. The key assumption of this prior is that good natural images are fixed points under a quantile filter with a spatially adaptive weighting scheme. The flexibility of AQuaSI can be attributed to the optional use of a guidance image, which can be either used in the direction of joint filtering based on a *static* guidance, or in *dynamic* guidance, in which case it is similar in spirit to filtering. We present AQuaSI within a full mathematical framework to seamlessly integrate it into popular optimization schemes such as gradient descent and ADMM.

In our experimental results, we demonstrate the strength of AQuaSI for the application of RGB/depth map upsampling, RGB/NIR image restoration, and image deblurring. Our prior is universally applicable and compares favorably against the state-of-the-art in the investigated applications. Its importance over related RED or PnP approaches is especially evident for inverse problems with unknown or non-Gaussian noise, e. g., RGB/NIR restoration. In analyzing perception-distortion planes, we found that AQuaSI shows decent tradeoffs between low signal distortions and high perceptual quality. In the simpler cases of noise-free problems or degradations by moderate levels of Gaussian noise, RED algorithms perform equally well as shown for non-blind deblurring. We also found that for certain problems like RGB/depth map upsampling, task-specific priors can outperform the solely use of AQuaSI due to their superior denoising capabilities. However, thanks to its structure preservation, state-of-the-art performance is achieved when combining AQuaSI with task-specific regularization.

Due to the ever-increasing number of learning-based image restoration techniques, we believe that transferring our universal prior to the field of deep learning, e. g., as operator in neural networks, is an interesting area for future research and can help to boost the performance of these methods.

## Acknowledgment

This work was supported in part by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project number 146371743 TRR 89 Invasive Computing

## References

- [1] S. D. Babacan, R. Molina, and A. K. Katsaggelos, “Variational Bayesian Super Resolution,” *IEEE Trans. Image Process.*, vol. 20, no. 4, pp. 984–999, 2011.
- [2] P. Ochs, A. Dosovitskiy, T. Brox, and T. Pock, “An iterated l1 algorithm for non-smooth non-convex optimization in computer vision,” in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1759–1766.



- [3] M. Elad and M. Aharon, “Image denoising via sparse and redundant representation over learned dictionaries,” *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.
- [4] J. Kim, J. K. Lee, and K. M. Lee, “Accurate Image Super-Resolution Using Very Deep Convolutional Networks,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1646–1654.
- [5] P. Wieschollek, M. Hirsch, B. Scholkopf, and H. P. Lensch, “Learning Blind Motion Deblurring,” in *2017 IEEE International Conference on Computer Vision*, 2017, pp. 231–240.
- [6] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, “Plug-and-Play priors for model based reconstruction,” in *2013 IEEE Global Conference on Signal and Information Processing*, 2013, pp. 945–948.
- [7] S. A. Bigdeli and M. Zwicker, “Image restoration using autoencoding priors,” *arXiv preprint arXiv:1703.09964*, 2017.
- [8] Y. Romano, M. Elad, and P. Milanfar, “The Little Engine that Could: Regularization by Denoising (RED),” *SIAM Journal on Imaging Sciences*, vol. 10, no. 4, pp. 1804–1844, 2017.
- [9] Y. Sun, J. Liu, and U. S. Kamilov, “Block coordinate regularization by denoising,” *arXiv preprint arXiv:1905.05113*, 2019.
- [10] G. Mataev, M. Elad, and P. Milanfar, “Deepred: Deep image prior powered by RED,” *CoRR*, vol. abs/1903.10176, 2019.
- [11] E. T. Reehorst and P. Schniter, “Regularization by denoising: Clarifications and new interpretations,” *IEEE Trans. Comput. Imag.*, vol. 5, no. 1, pp. 52–67, 2018.
- [12] S. H. Chan, “Performance analysis of plug-and-play admm: A graph signal processing perspective,” *IEEE Trans. Comput. Imag.*, vol. 5, no. 2, pp. 274–286, 2019.
- [13] 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, 2017.
- [14] A. C. Bovik, *Handbook of image and video processing*. Academic press, 2010.
- [15] F. Schirmacher, T. Köhler, J. Endres, T. Lindenberger, L. Husvagt, J. G. Fujimoto, J. Hornegger, A. Dörfler, P. Hoelter, and A. K. Maier, “Temporal and volumetric denoising via quantile sparse image prior,” *Medical image analysis*, vol. 48, pp. 131–146, 2018.
- [16] Z. Ma, K. He, Y. Wei, J. Sun, and E. Wu, “Constant Time Weighted Median Filtering for Stereo Matching and Beyond,” in *2013 IEEE International Conference on Computer Vision*, 2013, pp. 49–56.
- [17] Q. Zhang, L. Xu, and J. Jia, “100+ Times Faster Weighted Median Filter (WMF),” in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2830–2837.
- [18] D. Perrone and P. Favaro, “A Clearer Picture of Total Variation Blind Deconvolution,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 6, pp. 1041–1055, 2016.
- [19] D. Krishnan and R. Fergus, “Fast Image Deconvolution using Hyper-Laplacian Priors,” in *Advances in Neural Information Processing Systems*, 2009, pp. 1033–1041.
- [20] T. Köhler, X. Huang, F. Schebesch, A. Aichert, A. Maier, and J. Hornegger, “Robust Multiframe Super-Resolution Employing Iteratively Re-Weighted Minimization,” *IEEE Trans. Comput. Imag.*, vol. 2, no. 1, pp. 42 – 58, 2016.

- [21] J. Pan, Z. Hu, Z. Su, and M.-H. Yang, “L<sub>0</sub>-Regularized Intensity and Gradient Prior for Deblurring Text Images and Beyond,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8828, pp. 1–14, 2016.
- [22] D. Krishnan, T. Tay, and R. Fergus, “Blind deconvolution using a normalized sparsity measure,” in *2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2011, pp. 233–240.
- [23] J. Pan, D. Sun, H. Pfister, and M.-H. Yang, “Blind Image Deblurring Using Dark Channel Prior,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1628–1638.
- [24] Y. Yan, W. Ren, Y. Guo, R. Wang, and X. Cao, “Image Deblurring via Extreme Channels Prior,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6978–6986.
- [25] S. Wang, L. Zhang, and Y. Liang, “Nonlocal spectral prior model for low-level vision,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7726 LNCS, 2013, pp. 231–244.
- [26] S. Gu, L. Zhang, W. Zuo, and X. Feng, “Weighted nuclear norm minimization with application to image denoising,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2862–2869.
- [27] J. Yang, J. Wright, T. S. Huang, and Y. Ma, “Image Super-Resolution via Sparse Representation,” *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, 2010.
- [28] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang, “Deep networks for image super-resolution with sparse prior,” in *2015 IEEE International Conference on Computer Vision*, 2015, pp. 370–378.
- [29] K. Zhang, W. Zuo, S. Gu, and L. Zhang, “Learning deep cnn denoiser prior for image restoration,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3929–3938.
- [30] T. Meinhardt, M. Moeller, C. Hazirbas, and D. Cremers, “Learning Proximal Operators: Using Denoising Networks for Regularizing Inverse Imaging Problems,” in *2017 IEEE International Conference on Computer Vision*, 2017, pp. 1799–1808.
- [31] D. Ulyanov, A. Vedaldi, and V. Lempitsky, “Deep image prior,” in *2018 IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9446–9454.
- [32] 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, 2017.
- [33] S. Ono, “Primal-dual plug-and-play image restoration,” *IEEE Signal Processing Letters*, vol. 24, no. 8, pp. 1108–1112, 2017.
- [34] I. Daubechies, M. Defrise, and C. De Mol, “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.
- [35] U. S. Kamilov, H. Mansour, and B. Wohlberg, “A plug-and-play priors approach for solving nonlinear imaging inverse problems,” *IEEE Signal Processing Letters*, vol. 24, no. 12, pp. 1872–1876, 2017.
- [36] Y. Sun, B. Wohlberg, and U. S. Kamilov, “An online plug-and-play algorithm for regularized image reconstruction,” *IEEE Trans. Comput. Imag.*, 2019.

- [37] G. T. Buzzard, S. H. Chan, S. Sreehari, and C. A. Bouman, “Plug-and-play unplugged: Optimization-free reconstruction using consensus equilibrium,” *SIAM Journal on Imaging Sciences*, vol. 11, no. 3, pp. 2001–2020, 2018.
- [38] S. A. Bigdeli, M. Zwicker, P. Favaro, and M. Jin, “Deep mean-shift priors for image restoration,” in *Advances in Neural Information Processing Systems*, 2017, pp. 763–772.
- [39] R. Ahmad, C. A. Bouman, G. T. Buzzard, S. Chan, E. T. Reehorst, and P. Schniter, “Plug and play methods for magnetic resonance imaging,” *arXiv preprint arXiv:1903.08616*, 2019.
- [40] R. Zeyde, M. Elad, and M. Protter, “On single image scale-up using sparse-representations,” in *Curves and Surfaces*. Springer, 2012, pp. 711–730.
- [41] D. Krishnan and R. Fergus, “Dark flash photography,” in *ACM SIGGRAPH 2009 Papers*, 2009, pp. 96:1–96:11.
- [42] K. He, J. Sun, and X. Tang, “Guided Image Filtering,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–409, 2013.
- [43] B. Ham, M. Cho, and J. Ponce, “Robust Image Filtering using Joint Static and Dynamic Guidance,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4823–4831.
- [44] Q. Yan, X. Shen, L. Xu, S. Zhuo, X. Zhang, L. Shen, and J. Jia, “Cross-field joint image restoration via scale map,” in *The IEEE International Conference on Computer Vision*, 2013.
- [45] T. Goldstein and S. Osher, “The Split Bregman Method for L1-Regularized Problems,” *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323–343, 2009.
- [46] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [47] J. Lüthen, J. Wörmann, M. Kleinsteuber, and J. Steurer, “A rgb/nir data set for evaluating dehazing algorithms,” *Electronic Imaging*, vol. 2017, no. 12, pp. 79–87, 2017.
- [48] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, “Understanding and evaluating blind deconvolution algorithms,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1964–1971.
- [49] Y. Liu, J. Wang, S. Cho, A. Finkelstein, and S. Rusinkiewicz, “A no-reference metric for evaluating the quality of motion deblurring,” *ACM Trans. Graph.*, vol. 32, no. 6, pp. 175–1, 2013.
- [50] Y. Blau and T. Michaeli, “The perception-distortion tradeoff,” in *2018 IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6228–6237.
- [51] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Color image denoising via sparse 3d collaborative filtering with grouping constraint in luminance-chrominance space,” in *2007 IEEE International Conference on Image Processing*, vol. 1, 2007, pp. I–313.
- [52] —, “Image denoising by sparse 3-d transform-domain collaborative filtering,” *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 145–149, 2007.
- [53] X. Shen, C. Zhou, L. Xu, and J. Jia, “Mutual-structure for joint filtering,” in *2015 IEEE International Conference on Computer Vision*, 2015, pp. 3406–3414.
- [54] D. Ferstl, C. Reinbacher, R. Ranftl, M. Ruether, and H. Bischof, “Image guided depth upsampling using anisotropic total generalized variation,” in *2013 IEEE International Conference on Computer Vision*, 2013, pp. 993–1000.

- [55] Y. Li, J.-B. Huang, N. Ahuja, and M.-H. Yang, “Deep joint image filtering,” in *European Conference on Computer Vision*. Springer, 2016, pp. 154–169.
- [56] —, “Joint image filtering with deep convolutional networks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 1909–1923, 2019.
- [57] P. K. Nathan Silberman, Derek Hoiem and R. Fergus, “Indoor segmentation and support inference from rgb-d images,” in *2012 IEEE European Conference on Computer Vision*, 2012, pp. 746–760.
- [58] S. Song, S. P. Lichtenberg, and J. Xiao, “Sun rgb-d: A rgb-d scene understanding benchmark suite,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 567–576.
- [59] A. Janoch, S. Karayev, , J. T. Barron, M. Fritz, K. Saenko, and T. Darrell, “A category-level 3-d object dataset: Putting the kinect to work,” in *2011 IEEE International Conference on Computer Vision Workshops*, 2011, pp. 1168–1174.
- [60] J. Xiao, A. Owens, and A. Torralba, “Sun3d: A database of big spaces reconstructed using sfm and object labels,” in *2013 IEEE International Conference on Computer Vision*, 2013, pp. 1625–1632.
- [61] D. Scharstein, R. Szeliski, and R. Zabih, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” in *IEEE Workshop on Stereo and Multi-Baseline Vision*, 2001, pp. 131–140.
- [62] X. Tao, C. Zhou, X. Shen, J. Wang, and J. Jia, “Zero-order reverse filtering,” in *2017 IEEE International Conference on Computer Vision*, 2017, pp. 222–230.
- [63] E. S. Gastal and M. M. Oliveira, “Adaptive manifolds for real-time high-dimensional filtering,” *ACM Trans. Graph.*, vol. 31, no. 4, p. 33, 2012.
- [64] Z. Wang, E. P. Simoncelli, and A. C. Bovik, “Multiscale structural similarity for image quality assessment,” in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers*, 2003, vol. 2. IEEE, 2003, pp. 1398–1402.
- [65] A. Mittal, R. Soundararajan, and A. C. Bovik, “Making a completely blind image quality analyzer,” *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.
- [66] A. Mittal, A. K. Moorthy, and A. C. Bovik, “No-reference image quality assessment in the spatial domain,” *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, 2012.
- [67] A. Buades, B. Coll, and J. . Morel, “A non-local algorithm for image denoising,” in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 2005, pp. 60–65.
- [68] Y. Chen and T. Pock, “Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1256–1272, 2017.