Chapter 1
Inverse Problems in Image Restoration

Tomáš Kerepecký and Filip Šroubek

Abstract This work addresses inverse problems in image restoration, focusing on recovering high-quality images from degraded observations, a critical task in fields like microscopy and digital photography. We examine both traditional variational methods and modern deep learning techniques, highlighting hybrid approaches that merge mathematical modeling with data-driven learning. Classical model-based methods use explicit regularization, like total variation, to incorporate prior knowledge and stabilize the inversion process. Meanwhile, deep learning approaches, both supervised and self-supervised, leverage implicit regularization, where network architectures capture and learn prior information from data. We present our recent advancements in this field and discuss the effectiveness of these complementary approaches in solving complex image restoration problems in theory and practice.

Key words: inverse problems; image restoration; deep learning; regularization.

Inverse problems in image restoration focus on recovering a sharp, high-quality image from its degraded version, a task essential in applications like microscopy and digital photography, where images are often compromised by noise, blur, or other distortions during acquisition. These challenges arise due to the inherent limitations of real-world imaging systems, such as sensor imperfections, optical distortions, sampling issues, and other hardware constraints which degrade image quality. The objective is to model the degradation process and develop methods that can reverse it, reconstructing the original latent image from incomplete or corrupted data. Image restoration methods often leverage mathematical tools provided by variational

Tomáš Kerepecký
Institute of Information Theory and Automation, Czech Academy of Sciences, Czechia
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague, Czechia
e-mail: kerepecky@utia.cas.cz

Filip Šroubek
Institute of Information Theory and Automation, Czech Academy of Sciences, Czechia

calculus, numerical mathematics, and machine learning, allowing for sophisticated solutions to these inverse problems.

Considering a general multi-view scenario, the image formation model relates the latent image u to the acquired measurements g_k according to the following formula:

$$g_k = D_k u + n_k, \tag{1.1}$$

where D_k is a degradation operator for the k-th view, and n_k is the additive noise specific to that view. The degradation operator D_k accounts for both geometric and radiometric degradations. The geometric part affects the spatial domain of u, modeling transformations such as those caused by camera movements between acquisitions, which are often modeled by affine transformations. The radiometric part modifies the intensity values of u. Common forms of radiometric degradation include subsampling on camera sensors, which reduces resolution, and blurring, which occurs due to factors like improper focus or motion during acquisition. For example, in digital cameras, out-of-focus lenses produce blur shaped by the aperture, while in fluorescence microscopy, 3D blur can be introduced by anisotropies in the scanning process.

Multi-view image acquisition is common in applications like video processing, where adjacent frames provide slightly different perspectives, or in microscopy, where multi-view setups provide complementary information. In single-channel scenarios, such as traditional photography, only one observation g is available, and the degradation is modeled by a single operator D.

Image restoration aims to invert the image formation model (1.1) and recover the latent image u from either a single observation g or multiple images g_k. If D_k is the identity operator and only noise is present, the inverse problem is called denoising. If D_k involves geometric degradation, the task becomes image registration. If D_k is a blurring operator, the inverse problem is deconvolution. When the downsampling operator is present, the addressed task is called super-resolution or demosaicking. The challenge with D_k is its often ill-conditioned nature, which means that noise n_k gets amplified during inversion, leading to unstable solutions.

The classical **model-based** approach for solving (1.1) adopts variational calculus and formulates the problem as a regularized minimization that aims to find a unique solution. The minimization problem is formulated as:

$$\min_u \mathcal{L}(u) + \mathcal{R}(u). \tag{1.2}$$

The data term $\mathcal{L}(u)$ measures the fidelity of the estimated u to the image formation model (1.1). When we assume the model error follows a Gaussian distribution, the appropriate measure is the L_2 norm: $\mathcal{L}(u) = \sum_k \|D_k u - g_k\|_2^2$. When assuming the model error follows a Laplace distribution, the L_1 norm is used.

The **explicit regularization** term $\mathcal{R}(u)$ imposes prior knowledge on u and reduces the set of admissible solutions, thereby stabilizing the ill-posed problem. Regularization terms often involve non-smooth functions and frequently depend on the image gradient. A classic example is the total variation (TV) [1], but there are other

examples, such as promoting sparsity of image features using wavelets. The minimization of (1.2) is done iteratively by proximal algorithms such as Primal-Dual [3] or Alternating Direction Method of Multipliers (ADMM) [4]. These algorithms split the original problem into several easier-to-solve subproblems by replacing non-smooth terms with proximal operators.

While model-based methods rely on explicit regularization, modern deep learning approaches introduce **implicit regularization**. These methods treat the inverse problem as a learning task, where a Neural Network (NN) is trained to minimize a loss function representing the difference between the output of NN and reference images (either high-quality ground truth images or original measurements).

In **supervised learning** and a single-view scenario, the training data consists of pairs $\{g^{(i)}, u^{(i)}\}$, with the degraded images $g^{(i)}$ generated from the acquisition process (1.1) applied to a high-quality ground truth images $u^{(i)}$. A typical setup for a supervised learning process is given by the following Mean Squared Error (MSE) minimization problem:

$$\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{N} \|\mathcal{N}_\theta\left(g^{(i)}\right) - u^{(i)}\|_2^2, \tag{1.3}$$

where $\mathcal{N}_\theta$ is a neural network parameterized by θ, mapping the degraded images $g^{(i)}$ to the target images $u^{(i)}$. Here, $\hat{\theta}$ represents the optimal set of network parameters that minimize the MSE loss function over the N training data pairs. The final reconstruction procedure is carried out by the trained network $\mathcal{N}_{\hat{\theta}}$. For more details on supervised learning for inverse problems, the reader is referred to [2].

In contrast, **self-supervised** multi-view approaches focus on minimizing the difference between the network's output, after applying the degradation process, and the original K degraded measurements g_k. These methods do not require ground truth images. Instead, the network is trained to reconstruct an image that, when passed through the degradation operator D_k, closely resembles the original degraded observation g_k. A typical self-supervised setup can be formulated as:

$$\hat{\theta} = \arg\min_{\theta} \sum_{k=1}^{K} \|D_k \mathcal{N}_\theta(g_k) - g_k\|_2^2, \tag{1.4}$$

where D_k is either known a priori or learned as part of the training process. This approach is especially useful in scenarios where ground truth data is unavailable, such as in computed tomography or fluorescence microscopy, allowing the network to learn directly from the degraded observations without external supervision. A common example is Deep Image Prior (DIP) [7].

Both supervised and self-supervised deep learning methods represent a significant shift in how we approach inverse problems in image restoration. Traditional model-based methods depend on explicit regularization, where a specific term, $\mathcal{R}(u)$, is defined to stabilize the inversion. In contrast, deep learning methods utilize implicit regularization. They learn prior knowledge from extensive datasets during training, or, in the case of self-supervised approaches, directly from the measure-

ments, effectively extracting regularization from the data. The network architecture itself acts as a regularizer, imposing structure and constraints on the restored image.

Recent developments combine learning-based and model-based techniques. Notable trends in this integration include methods like Deep Unrolling [8] or Deep Equilibrium Models [9], which reformulate traditional iterative algorithms as neural networks.

There are other hybrid approaches that have shown significant potential in addressing image restoration tasks by leveraging both physical model-driven and data-driven methodologies. Plug-and-Play Priors [5] and Regularization by Denoising [6] incorporate pre-trained denoisers as regularizers within traditional optimization frameworks. Similarly, the DIP incorporates the forward model from the acquisition process (1.1) while solving inverse problems in a self-supervised manner, relying on the inherent structure learned by neural networks.

We applied these approaches in our recent works, as outlined in Table 1.1, and discuss them in detail in the subsequent sections.

	Model based	**Supervised learning**	**Self-supervised learning**	**Explicit reg.**	**Implicit reg.**
IWFT [10]	✓			✓	
NeRD [11]		✓			✓
D3Net [12]	✓	✓		✓	✓
Dual-Cycle [13]	✓		✓		✓

Table 1.1 Comparison of different approaches to inverse problems in image restoration, highlighting the use of model-based methods, supervised and self-supervised deep learning strategies, as well as the integration of explicit and implicit regularization.

Example of a model-based technique with explicit regularization is our Iterative Wiener Filtering and Thresholding (IWFT) [10] for solving the deconvolution problem in embedded devices with limited computational power. Here, the degradation operator D is modeled as a convolution, representing the camera blur. The objective is to minimize the L_2 norm of the error in the blurred observation while introducing an explicit TV regularization term. IWFT employs ADMM as the optimization framework to alternate between solving the fidelity term and regularization in (1.2). A key innovation in IWFT is that optimization steps are translated into simple filtering operations. Specifically, the method introduces two filters: a restoration filter applied during the initial Wiener filtering step to handle the deblurring, and an update filter used iteratively to refine the image by addressing the TV regularization. These filters are pre-computed for different blur types and noise levels, making the final restoration process computationally efficient and capable of significantly reducing artifacts, such as ringing, commonly seen in traditional deconvolution approaches.

An example of the supervised deep learning approach with implicit regularization is Neural Field-Based Demosaicking (NeRD) [11], where we address the problem of reconstructing full-color images from raw camera sensor data. The degradation process is represented as sub-sampling introduced by the Bayer pattern in digital

cameras. In our method we impose implicit regularization through the use of Implicit Neural Representations (INRs) [14], which model continuous signals, such as images, as functions parameterized by neural networks rather than as discrete pixel values. NeRD utilizes INR $\mathcal{N}_\theta$ to represent the latent image u. The neural network, parameterized by θ, maps spatial coordinates x and learned feature encodings $\xi_x(g)$ to RGB values: $\mathcal{N}_\theta(\xi_x, x) \to u(x)$. This approach allows NeRD to bypass classical regularization by embedding prior knowledge in the network architecture and achieving reconstruction through its implicit structure.

A typical example of a hybrid approach is our recent D3Net [12], which incorporates both explicit and implicit regularization using a deep unrolling technique. D3Net tackles the joint problem of demosaicking, deblurring, and deringing by unrolling traditional ADMM iterations into network layers. The degradation operator models both camera blur through convolution and downsampling via the Bayer pattern, similar to NeRD. The method starts with a modified version of IWFT for demosaicking, which is later fine-tuned using a supervised deep neural network.

D3Net applies an initial restoration filter to address demosaicking and deblurring, followed by iterative update layers that suppress ringing artifacts. These filters are further optimized through back-propagation during training. D3Net demonstrates significant improvements over purely model-based methods, combining the strengths of classical optimization with the flexibility of neural networks.

An example of a model-based, self-supervised approach with implicit regularization is Dual-Cycle [13], which addresses the reconstruction of isotropic 3D volumes from dual-view fluorescence microscopy images. In this case, the degradation processes D involves blurring and affine transformations, modeled as:

$$g_1 = D_1 u + n_1\,, \quad g_2 = D_2 u + n_2, \tag{1.5}$$

where g_1 and g_2 are two observed degraded volumes, and u represents the latent 3D volume.

For reconstruction, Dual-Cycle employs a U-Net [16], leveraging implicit regularization through DIP. The network is trained in a self-supervised manner using a cycle consistency loss to simulate the degradation and enforce consistency with the observations. Additionally, GAN discriminators [15] are integrated to guide the training, further improving the reconstruction. The loss function combines L_1 cycle consistency and adversarial GAN losses $\mathcal{L}_{GAN}$:

$$\hat{\theta}, \hat{\phi} = \arg\min_{\theta,\phi} \sum_{i\in\{1,2\}} \|D_{\phi,i}\mathcal{N}_\theta(g_1, g_2) - g_i\|_1 + \mathcal{L}_{GAN}. \tag{1.6}$$

Here, $\mathcal{N}_\theta$ represents the reconstruction network responsible for solving the inverse problem and estimating the latent 3D volume. The learned operator $D_{\phi,i}$ models the degradation, utilizing known blur kernels. Both the U-Net parameters θ and the parameters of the degradation operator ϕ are optimized jointly during the training process.

We developed various methods to address inverse problems in image restoration, focusing on model-based and deep learning approaches, supervised and self-supervised learning, and explicit versus implicit regularization. Our work covers techniques that handle various degradation models, such as blur, subsampling, and affine transformations, applied in both single-view and multi-view scenarios. We tested classical regularization like total variation in IWFT and implicit regularization in NeRD. Hybrid approaches, such as D3Net and Dual-Cycle, demonstrate the effectiveness of combining model-based and learning-based paradigms to solve complex inverse problems in image restoration.

Acknowledgements This work was supported by the Czech Science Foundation GA21-03921S.

References

1. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**(1–4), 259–268 (1992)
2. McCann, M.T., Jin, K.H., Unser, M.: Convolutional neural networks for inverse problems in imaging: A review. IEEE Signal Processing Magazine **34**(6), 85–95 (2017)
3. Chambolle, A., Pock, T.: An introduction to continuous optimization for imaging. Acta Numerica **25**, 161–319 (2016)
4. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. Now Publishers Inc. (2011)
5. Kamilov, U.S., Bouman, C.A., Buzzard, G.T., Wohlberg, B.: Plug-and-play methods for integrating physical and learned models in computational imaging: Theory, algorithms, and applications. IEEE Signal Processing Magazine **40**(1), 85–97 (2023)
6. Romano, Y., Elad, M., Milanfar, P.: The little engine that could: Regularization by denoising (RED). SIAM Journal on Imaging Sciences **10**(4), 1804–1844 (2017)
7. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 9446–9454 (2018)
8. Monga, V., Li, Y., Eldar, Y.C.: Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. IEEE Signal Processing Magazine **38**(2), 18–44 (2021)
9. Gilton, D., Ongie, G., Willett, R.: Deep equilibrium architectures for inverse problems in imaging. IEEE Transactions on Computational Imaging **7**, 1123–1133 (2021)
10. Šroubek, F., Kerepecký, T., Kamenický, J.: Iterative Wiener filtering for deconvolution with ringing artifact suppression. EUSIPCO 2019, 1–5 (2019)
11. Kerepecký, T., Šroubek, F., Novozámský, A., Flusser, J.: Nerd: Neural field-based demosaicking. ICIP 2023, 1735–1739 (2023)
12. Kerepecký, T., Šroubek, F.: D3net: Joint demosaicking, deblurring and deringing. ICPR 2020, 1–8 (2021)
13. Kerepecky, T., Liu, J., Piston, D.W., Kamilov, U.S.: Dual-cycle: Self-supervised dual-view fluorescence microscopy image reconstruction using cyclegan. ICASSP 2023, 1–5 (2023)
14. Sitzmann, V., Martel, J. N. P., Bergman, A. W., Lindell, D. B., Wetzstein, G.: Implicit Neural Representations with Structured Latent Spaces. Advances in Neural Information Processing Systems **33**, 5521–5532 (2020)
15. Zhu, J.-Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of the IEEE International Conference on Computer Vision 2017, 2223–2232 (2017)
16. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention 9351, 234–241 (2015)