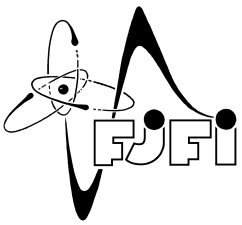
CZECH TECHNICAL UNIVERSITY IN PRAGUE
FACULTY OF NUCLEAR SCIENCES AND PHYSICAL ENGINEERING

DOCTORAL THESIS

INVERSE PROBLEMS IN IMAGE RESTORATION



PRAGUE 2025 TOMÁŠ KEREPECKÝ

I declare that I have carried out this doctoral thesis independently, using only the cited sources, literature, and other professional references. This thesis has not been used to obtain any other degree, whether the same or different. To enhance clarity and structure, I used tools such as Grammarly, ChatGPT, and Gemini during the preparation of this work. However, I thoroughly reviewed and refined the text manually, taking full responsibility for the final version.

In Prague, March 25, 2025 .....

Author's signature

Bibliographic entry

Author	Ing. Tomáš Kerepecký Czech Technical University in Prague Faculty of Nuclear Sciences and Physical Engineering Department of Mathematics
Title of Dissertation	Inverse Problems in Image Restoration
Field of Study	Mathematical Engineering
Supervisor	Doc. Ing. Filip Šroubek, Ph.D. DSc. Institute of Information Theory and Automation Czech Academy of Sciences
Academic Year	2025
Number of Pages	116
Keywords	inverse problems, image restoration, demosaicking, deconvolution, deep learning, implicit neural networks, image processing

Bibliografický záznam

Autor	Ing. Tomáš Kerepecký České vysoké učení technické v Praze Fakulta jaderná a fyzikálně inženýrská Katedra matematiky
Název práce	Řešení inverzních problémů rekonstrukce obrazu
Studijní obor	Matematické inženýrství
Školitel	Doc. Ing. Filip Šroubek, Ph.D. DSc. Ústav teorie informace a automatizace Akademie věd České republiky, v.v.i.
Akademický rok	2025
Počet stran	116
Klíčová slova	inverzní problémy, rekonstrukce obrazu, demosaicking, dekonvoluce, hluboké učení, implicitní neuronové reprezentace, digitální zpracování obrazu

Abstract

This doctoral thesis presents a comprehensive study of inverse problems in image restoration, focusing on recovering high-quality images from a variety of degraded inputs. It introduces five novel image reconstruction methods that collectively bridge classical model-based algorithms and modern deep learning approaches. First, an iterative Wiener filtering and thresholding technique (IWFT) is developed to perform image deblurring while suppressing ringing artifacts, addressing limitations of traditional deconvolutional methods. Second, a deep unrolled neural network (D3Net) is designed to jointly solve demosaicking, deblurring, and deringing in a unified optimization-inspired framework, blending the interpretability of classical model-based methods with the flexibility of learned models. Third, a dual-view self-supervised approach (Dual-Cycle) leverages cycle-consistent generative modeling to fuse two orthogonal light-sheet microscopy images, producing high-resolution 3D reconstructions without the need for any paired training data. The fourth contribution pioneers the use of implicit neural representations for image restoration: a neural field-based demosaicking method (NeRD) that represents images as continuous functions and achieves reconstruction quality on par with state-of-the-art supervised methods. Finally, a self-adaptive implicit framework (INRID) is proposed for image demosaicking, which optimizes a coordinate-based network per image and robustly handles additional degradations such as blur and noise without requiring retraining.

Abstrakt

Tato disertační práce představuje ucelenou studii inverzních problémů ve zpracování obrazu se zaměřením na rekonstrukci vysoce kvalitních obrazů z různě degradovaných vstupů. Přináší pět nových metod rekonstrukce obrazu, jež dohromady propojují klasické algoritmy s moderními technikami hlubokého učení. Nejprve je vyvinuta metoda iterativní Wienerovy filtrace a prahování (IWFT) k odstranění rozmazání obrazu, která zároveň potlačuje prstencové artefakty kolem hran a překonává omezení tradičních metod dekonvoluce. Dále je navržena neuronová síť D3Net, která v jednotném frameworku společně řeší problém demosaickingu, dekonvoluce a potlačení prstencových artefaktů, přičemž kombinuje interpretovatelnost klasických postupů s flexibilitou naučených modelů pomocí techniky “deep unrolling". Třetím přístupem je metoda Dual-Cycle určená pro rekonstrukci dat ve fluorescenční mikroskopii. Využívá cyklicky konzistentní generativní model k fúzi dvou ortogonálních snímků a dosahuje vysoce kvalitní 3D rekonstrukce bez potřeby párových trénovacích dat. Čtvrtým přínosem je průkopnické využití implicitních neuronových reprezentací pro demosaicking. Metoda NeRD reprezentuje obraz jako spojitou funkci definovanou neuronovou sítí a dosahuje kvality rekonstrukce srovnatelné se současnými pokročilými metodami. Nakonec je představena samoadaptivní metoda demosaickingu nazvaná INRID, která pro každý jednotlivý snímek optimalizuje vlastní implicitní neuronovou reprezentaci a dokáže robustně zvládat degradace, jako je rozostření a šum, aniž by vyžadovala další přetrénování.

Acknowledgments

First and foremost, I would like to thank two remarkable women in my life: my mother, ***Hana***, whose unwavering love and support carried me through the first half of my Ph.D., and my wife, ***Zuzana***, whose strength, grace, and encouragement sustained me in the second half. Thank you both for your belief in me.

I am deeply grateful to my supervisor, ***Filip Šroubek***, for his friendly guidance, patience, and invaluable insights. Despite thousands of responsibilities, he always made time for me. I was exceptionally privileged to be his Ph.D. student.

I am thankful to the entire **Department of Image Processing** for providing a supportive and welcoming environment, especially its leaders, ***Barbara Zitová*** and ***Jan Flusser*** for their warm and personal leadership and generous funding. Special thanks to ***Adam Novozámský***, whose help and friendship were essential at both the beginning and throughout my Ph.D. journey.

I would also like to express my heartfelt gratitude to the **Institute of Information Theory and Automation** and the **Czech Technical University**. These institutions provided not only a supportive environment and resources for my research and teaching but also the freedom to pursue my Ph.D. while actively contributing to society through educational and community-focused initiatives.

Special thanks to the **Fulbright Commission**, which acknowledged my commitment to both research and societal engagement by awarding me the Fulbright-Masaryk Award. This incredible opportunity allowed me to work as a visiting scholar in the United States.

Lastly, my deepest thanks to **God**, for I can truly relate to these words: *"I have fought the good fight, I have finished the race, and I have kept the faith."* (Bible, 2 Tim. 4:7)

My Ph.D. studies were supported by these grants:

- GAČR - GA25-15933S - *Dynamic Inverse Problems in Time-Lapse Microscopy*
- GAČR - GA24-10069S - *Hybrid Neural Network Architectures for Image Recognition*
- 2022 Fulbright-Masaryk Award - *Deep Learning-Based Next Generation Biomedical Imaging Technology*
- GAČR - GA21-03921S - *Inverse Problems in Image Processing*
- GAČR - GA18-05360S - *Solving Inverse Problems for the Analysis of Fast Moving Objects*
- 2017 Praemium Academicae - *Akademická prémie AV ČR (Jan Flusser)*

Contents

Part I

INTRODUCTION

Inverse Problems in Image Restoration

1

Inverse problems, in general, involve recovering complete information from incomplete or noisy data, much like deducing the "reality" from the shadows in Plato's allegory of the cave [1]. In *image restoration* [2], inverse problems arise when we try to reconstruct a high-quality image from degraded or partial measurements. While this task may seem like an abstract mathematical challenge [3, 4], it is central to a wide range of real-world imaging applications, including digital photography, astronomical imaging, remote sensing, industrial inspection, and biomedical imaging (see [5–8] for an in-depth overview).

Figure 1.1 illustrates the forward and inverse imaging processes in modalities such as digital photography, although the fundamental principles of inverse problems remain consistent across all other imaging techniques. In a forward model, the underlying scene $\boldsymbol{u}$ is mapped through a physical or computational degradation operator $\boldsymbol{\mathcal{D}}$ (representing the imaging system, e.g., a digital camera), to produce measurements $\boldsymbol{g}$ (often corrupted by noise $\boldsymbol{n}$), according to

$$\boldsymbol{g} = \boldsymbol{\mathcal{D}}(\boldsymbol{u}) + \boldsymbol{n}. \tag{1.1}$$

By contrast, an inverse model $\boldsymbol{\mathcal{R}}$ seeks to restore the original $\boldsymbol{u}$ from the observed $\boldsymbol{g}$, where the estimate is given by $\hat{\boldsymbol{u}} = \boldsymbol{\mathcal{R}}(\boldsymbol{g})$. This reversal of the forward process tends to be mathematically and computationally challenging, as many potential images $\boldsymbol{u}$ can yield the same input data, and measurement noise $\boldsymbol{n}$ further complicates the recovery process. Such problems, where a unique and stable solution may not exist, are commonly referred to as *ill-posed* [9].

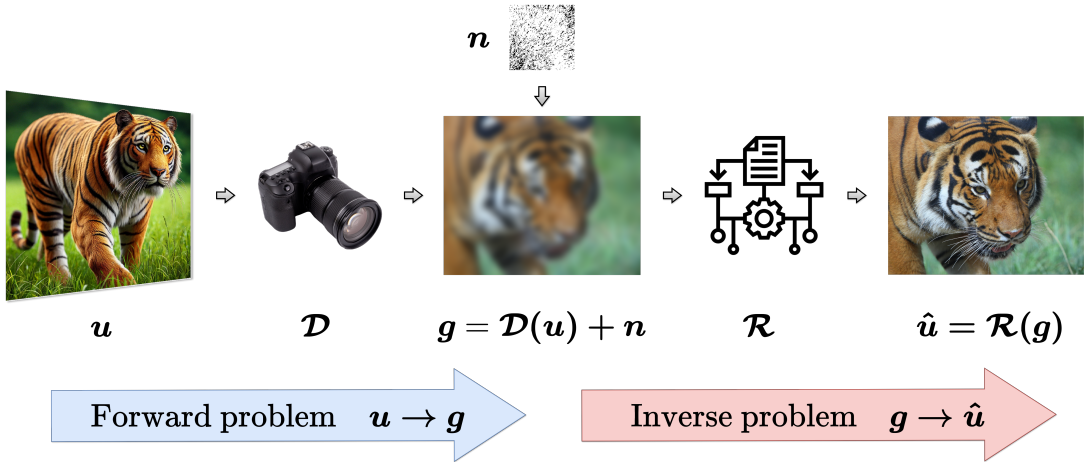


Figure 1.1: Schematic illustration of forward and inverse imaging problems in digital photography. In the forward problem (first arrow), a known object $\boldsymbol{u}$ is transformed into measured data $\boldsymbol{g}$ via an imaging system $\boldsymbol{\mathcal{D}}$, subject to degradation such as blur and noise $\boldsymbol{n}$. In the inverse problem (second arrow), the goal is to recover the estimate $\hat{\boldsymbol{u}}$ from $\boldsymbol{g}$. The task is typically ill-posed and very sensitive to measurement quality.

REAL-WORLD examples clearly illustrate the importance of inverse problems. In image restoration, problems are typically categorized as single-channel or multi-channel. To exemplify these paradigms, we focus on single-channel digital photography [10–14] and multi-channel light-sheet fluorescence microscopy (LSFM) [15–17]. Both cases also reflect the central themes of this thesis.

Digital Photography

When capturing an image with a digital camera, one might assume that the sensor records a full-color, sharp picture. In reality, almost all cameras use a single sensor with a color filter array (CFA), such as the Bayer pattern, where each pixel records only one of the three RGB colors. Moreover, physical limitations such as out-of-focus blur, motion blur, lens imperfections and sensor noise further degrade the captured data. As a result, the camera must not only reconstruct the missing colors (*demosaicking* [18–20]) but also correct for the blur (*deblurring* [21–23]). This leads to a joint demosaicking-deblurring problem that may also incorporate *denoising* [24–26] (see Figure 1.2).

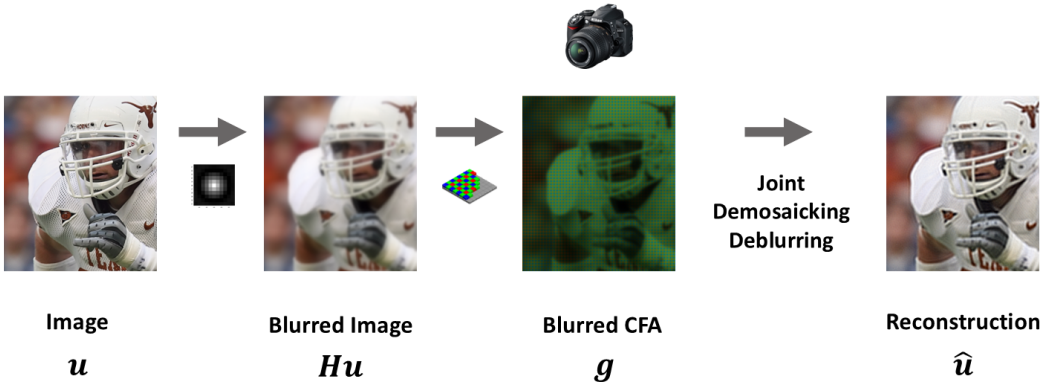


Figure 1.2: Illustration of the joint demosaicking and deblurring inverse problem in digital photography. A real-world scene, mathematically represented as an image $\boldsymbol{u}$, is formed on the camera sensor by the lens, which may introduce optical blur, resulting in a blurred image $\boldsymbol{Hu}$. Simultaneously, a color filter array imposes color sampling, producing the blurred raw data $\boldsymbol{g}$. The inverse problem then reconstructs the image $\hat{\boldsymbol{u}}$, restoring both colors and details.

The general forward model (1.1) with the degradation operator $\mathcal{D}$ is now expressed as the matrix composition $\mathcal{D}(\boldsymbol{u}) = \boldsymbol{S}\,\boldsymbol{H}\boldsymbol{u}$, where $\boldsymbol{u}$ is in vectorized form, $\boldsymbol{H}$ denotes a blur operator, typically modeled as a convolution with a known point-spread function (PSF), and $\boldsymbol{S}$ represents the sampling operator corresponding to the given CFA. With this decomposition, the observation model becomes

$$\boldsymbol{g} = \boldsymbol{S}\,\boldsymbol{H}\,\boldsymbol{u} + \boldsymbol{n}\,. \tag{1.2}$$

Poor reconstruction $\boldsymbol{u}$ from measurement $\boldsymbol{g}$ can lead to visible artifacts such as color Moiré, zippering, or ringing, while effective reconstruction preserves fine details and faithfully reproduces the original scene. In everyday photography, therefore, addressing the inverse problem of joint demosaicking-deblurring is indispensable for achieving images of high perceptual quality. The challenges associated with this problem are investigated in Chapter 5 and further detailed in Part II of this thesis.

Fluorescence Microscopy

Beyond consumer photography, inverse problems are crucial in advanced scientific imaging, particularly in fluorescence microscopy. LSFM is a powerful technique for capturing 3D images of biological specimens with minimal photodamage. Instead of illuminating the entire sample, LSFM selectively excites the fluorophores in a thin optical plane, reducing out-of-focus blur and allowing for extended imaging of living specimens. However, because each view is captured plane by plane, the resulting image stack may suffer from incomplete structural information and anisotropic resolution (with lower detail along the optical axis). To mitigate these issues, a dual-view Selective Plane Illumination Microscope (diSPIM) [27] captures images from two perpendicular directions, each providing complementary information (see Figure 1.3).

In this case, the two degradation processes are modeled as

$$\boldsymbol{g}_1 = \boldsymbol{A}_1\boldsymbol{H}_1\boldsymbol{u} + \boldsymbol{n}_1, \qquad \boldsymbol{g}_2 = \boldsymbol{A}_2\boldsymbol{H}_2\boldsymbol{u} + \boldsymbol{n}_2, \tag{1.3}$$

where $\boldsymbol{H}_i$ represents optical blur for view i and $\boldsymbol{A}_i$ is the affine transform (rotation and misalignment) for each camera.

The challenge becomes merging these two incomplete and noisy 3D measurements, $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$, into a single high-quality 3D reconstruction $\boldsymbol{u}$. This is a typical example of two joint inverse problems, *image fusion* [28, 29] and *super-resolution* [30, 31], where multiple low resolution inputs are combined to recover a high-resolution image. A more in-depth exploration of this problem is presented in Section 5.3.

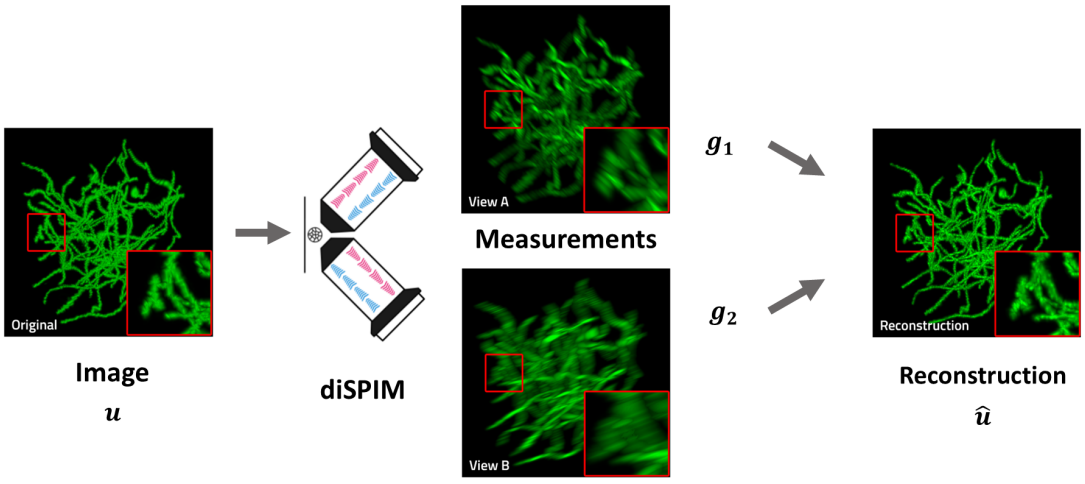


Figure 1.3: Illustration of the inverse problem in dual-view light-sheet fluorescence microscopy (diSPIM). A 3D biological sample, represented as $\boldsymbol{u}$, is imaged from two perpendicular views, producing 3D measurements $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$ with anisotropic resolution and noise. The inverse problem fuses these views to reconstruct $\hat{\boldsymbol{u}}$, enhancing structural details and isotropy.

From digital photography to cutting-edge biological imaging, this illustrates the central theme of this thesis: imaging systems rarely capture perfect data, and restoring a high-quality image requires solving an inverse problem.

CLASSICAL restoration methods include techniques that rely on mathematical models of the imaging processes and incorporate prior knowledge through handcrafted regularizers. In this context, a regularizer is a function designed to impose constraints on the solution, promoting features like smooth transitions or sharp edges, which are characteristic of natural images. Figure 1.4 demonstrates the effect of regularization.

For some inverse problems, a solution can be found in an explicit form. For example, when the only degradation is blur combined with additive noise, the Wiener filter is a well-established method that provides a closed-form solution in the frequency (Fourier) domain [32]. Although elegant, the Wiener filter often produces *ringing artifacts*, which are a manifestation of the Gibbs phenomenon (see [33]). As shown in Figure 1.4b, these artifacts are most prominent near edges. In Chapter 5 we have proposed an effective solution for suppressing these artifacts.

(a) Blurred (b) Wiener (c) Regularized (d) Ground-truth

Figure 1.4: Deblurring results using classical model-based approaches. (a) The blurred image; (b) the output from a Wiener filter – an optimal linear filter that can produce ringing artifacts near strong edges; (c) a regularized solution [11] that imposes constraints to suppress these artifacts while preserving natural features such as smooth transitions and sharp edges; and (d) the ground truth.

Variational Approach and Explicit Regularization

Rather than seeking a closed-form solution, classical *model-based* approaches reformulate the inverse problem as an optimization task that balances consistency to the observed data with prior assumptions about the underlying image. The goal is to find an estimate $\hat{\boldsymbol{u}}$ that minimizes

$$\hat{\boldsymbol{u}} = \arg\min_{\boldsymbol{u}} \{\boldsymbol{\mathcal{L}}(\boldsymbol{u}) + \lambda \boldsymbol{\Phi}(\boldsymbol{u})\}, \tag{1.4}$$

where the first term $\boldsymbol{\mathcal{L}}(\boldsymbol{u})$ measures the fidelity of the estimated $\boldsymbol{u}$ to the image formation model (1.1) and $\boldsymbol{\Phi}(\boldsymbol{u})$ encodes *explicit regularization* (e.g., smoothness, sparsity, or edge preservation). The parameter λ controls the trade-off between these components. When we assume the model error follows a Gaussian distribution, the data term becomes the ℓ_2 norm: $\boldsymbol{\mathcal{L}}(\boldsymbol{u}) = \frac{1}{2}\|\boldsymbol{\mathcal{D}}(\boldsymbol{u}) - \boldsymbol{g}\|_2^2$. When assuming the model error follows a Laplace

distribution, the ℓ_1 norm is used. A variety of regularization strategies have been proposed, such as Total Variation (TV) [34], wavelet-domain sparsity [35], and non-local self-similarity. For instance, the Non-Local Total Variation [36] integrates self-similarity into the regularization term, while effectively capturing long-range dependencies.

A classical example of early regularization methods is Tikhonov regularization [9], which employs a quadratic penalty $\boldsymbol{\Phi}(\boldsymbol{u}) = \|\boldsymbol{u}\|_2^2$. Although elementary, it often serves as a starting point for variational methods in inverse problems. In this formulation, the optimization problem becomes

$$\hat{\boldsymbol{u}} = \arg\min_{\boldsymbol{u}} \left\{ \frac{1}{2} \|\boldsymbol{\mathcal{D}}(\boldsymbol{u}) - \boldsymbol{g}\|_2^2 + \lambda \|\boldsymbol{u}\|_2^2 \right\}. \tag{1.5}$$

When considering a linear degradation operator in matrix form, $\boldsymbol{\mathcal{D}}(\boldsymbol{u}) = \boldsymbol{D}\,\boldsymbol{u}$, the formulation reduces to classical ridge regression [37]. Differentiating with respect to $\boldsymbol{u}$ and setting the gradient to zero leads to the normal equations:

$$\boldsymbol{D}^T(\boldsymbol{D}\,\boldsymbol{u} - \boldsymbol{g}) + 2\lambda\,\boldsymbol{u} = 0 \quad \Longrightarrow \quad (\boldsymbol{D}^T\boldsymbol{D} + 2\lambda\boldsymbol{I})\boldsymbol{u} = \boldsymbol{D}^T\boldsymbol{g}. \tag{1.6}$$

Assuming $(\boldsymbol{D}^T\boldsymbol{D} + 2\lambda\boldsymbol{I})$ is invertible, we arrive at a closed-form solution:

$$\hat{\boldsymbol{u}} = \boldsymbol{\mathcal{R}}(\boldsymbol{g}) = (\boldsymbol{D}^T\boldsymbol{D} + 2\lambda\boldsymbol{I})^{-1}\boldsymbol{D}^T\boldsymbol{g}. \tag{1.7}$$

This explicit solution is especially attractive for problems such as denoising (with $\boldsymbol{D} = \boldsymbol{I}$) or when the degradation operator is well-conditioned. However, for applications like deblurring ($\boldsymbol{D} = \boldsymbol{H}$) or joint demosaicking and deblurring ($\boldsymbol{D} = \boldsymbol{S}\boldsymbol{H}$), where $\boldsymbol{D}$ is often ill-conditioned or non-invertible, direct inversion is problematic. This issue persists even in a regularized framework, as a small λ leads to instability and noise amplification, while a large λ ensures stability but over-smooths the result, making it closely resemble the degraded input.

Iterative Optimization Techniques

In practice, solving the general variational formulation of inverse problems (1.4) often requires iterative numerical methods. Early methods employed gradient descent or Gauss-Seidel iterations [38, 39]; however, more advanced algorithms have been developed to handle the non-smooth optimization problems. Notably, the Iterative Shrinkage-Thresholding Algorithm (ISTA) [40] and its accelerated variant FISTA [41] leverage proximal operators to efficiently handle ℓ_1-regularized terms.

For example, TV regularization replaces the quadratic penalty with the ℓ_1-norm of the image gradient $\nabla\boldsymbol{u}$, thereby promoting piecewise-smooth solutions while preserving edges:

$$\hat{\boldsymbol{u}} = \arg\min_{\boldsymbol{u}} \left\{ \frac{1}{2} \|\boldsymbol{\mathcal{D}}(\boldsymbol{u}) - \boldsymbol{g}\|_2^2 + \lambda \|\nabla\boldsymbol{u}\|_1 \right\}. \tag{1.8}$$

Very popular choice for solving such non-smooth convex problems is the *Alternating Direction Method of Multipliers* (ADMM) [42] due to its ability to decouple the data fidelity and regularization terms.

By introducing an auxiliary variable $\boldsymbol{z}$, one can rewrite the minimization problem as

$$\min_{\boldsymbol{u},\boldsymbol{z}} \left\{ \frac{1}{2}\|\boldsymbol{\mathcal{D}}(\boldsymbol{u}) - \boldsymbol{g}\|_2^2 + \lambda\|\nabla \boldsymbol{z}\|_1 \right\} \quad \text{subject to} \quad \boldsymbol{u} = \boldsymbol{z}, \tag{1.9}$$

and then solve the resulting augmented Lagrangian

$$L_\rho(\boldsymbol{u}, \boldsymbol{z}, \boldsymbol{v}) = \frac{1}{2}\|\boldsymbol{\mathcal{D}}(\boldsymbol{u}) - \boldsymbol{g}\|_2^2 + \lambda\|\nabla \boldsymbol{z}\|_1 + \frac{\rho}{2}\|\boldsymbol{u} - \boldsymbol{z} + \boldsymbol{v}\|_2^2 - \frac{\rho}{2}\|\boldsymbol{v}\|_2^2, \tag{1.10}$$

iteratively. Here $\boldsymbol{v}$ is the scaled dual variable and $\rho > 0$ is a penalty parameter.

The ADMM algorithm proceeds by updating the variables in an alternating fashion. Specifically, at iteration k, the updates are as follows:

1. $\boldsymbol{u}$**-update:**

$$\boldsymbol{u}^{k+1} = \arg\min_{\boldsymbol{u}} \left\{ \frac{1}{2}\|\boldsymbol{\mathcal{D}}(\boldsymbol{u}) - \boldsymbol{g}\|_2^2 + \frac{\rho}{2}\|\boldsymbol{u} - \boldsymbol{z}^k + \boldsymbol{v}^k\|_2^2 \right\}. \tag{1.11}$$

2. $\boldsymbol{z}$**-update:**

$$\boldsymbol{z}^{k+1} = \arg\min_{\boldsymbol{z}} \left\{ \lambda\|\nabla \boldsymbol{z}\|_1 + \frac{\rho}{2}\|\boldsymbol{u}^{k+1} - \boldsymbol{z} + \boldsymbol{v}^k\|_2^2 \right\}. \tag{1.12}$$

3. **Dual variable update:**

$$\boldsymbol{v}^{k+1} = \boldsymbol{v}^k + \boldsymbol{u}^{k+1} - \boldsymbol{z}^{k+1}. \tag{1.13}$$

Here, the whole iterative algorithm functions as the reconstruction operator $\boldsymbol{\mathcal{R}}$, progressively refining the estimate of $\hat{\boldsymbol{u}}$.

In the context of digital photography, applying ADMM to solve the inverse problem (1.2) involves a $\boldsymbol{u}$-update step that minimizes a quadratic objective, accounting for both the blurring introduced by the operator $\boldsymbol{H}$ and the subsampling from the operator $\boldsymbol{S}$. Leveraging the convolutional structure of $\boldsymbol{H}$, one can use fast Fourier transform (FFT), while the sparse nature of $\boldsymbol{S}$ is efficiently handled with iterative solvers like conjugate gradients [39]. Meanwhile, the $\boldsymbol{z}$-update is addressed via an appropriate proximal operator, such as soft-thresholding for an ℓ_1-based regularizer. This decoupling in the ADMM framework enables effective reconstruction for joint demosaicking and deblurring, even in the presence of noise and other ill-conditioning effects.

Similarly, the primal-dual method of Chambolle and Pock [43] offers an efficient framework for handling non-smooth convex objectives by updating the primal and dual variables simultaneously.

Although classical model-based methods offer clear interpretability and, in some cases, closed-form solutions, they often struggle with the ill-posedness of inverse problems and the limitations of handcrafted priors. These challenges, along with the computational demands of iterative solvers, have spurred interest in alternative approaches. This naturally leads us to consider data-driven *deep learning* methods [6, 7], which learn the inverse mapping or the prior directly from data.

LEARNING-BASED reconstruction methods adopt a fundamentally different paradigm from traditional, model-based optimization strategies. Recall that classical approaches explicitly design a reconstruction operator

$$\boldsymbol{\mathcal{R}} : \boldsymbol{g} \to \boldsymbol{u}, \tag{1.14}$$

which inverts the forward degradation process (1.1) to recover the underlying image $\boldsymbol{u}$ from its noisy or incomplete measurement $\boldsymbol{g}$.

In contrast, *supervised* deep learning methods replace this manually designed operator with a data-driven learning mechanism

$$\boldsymbol{\mathcal{L}} : \{(\boldsymbol{g}^{(i)}, \boldsymbol{u}^{(i)})\}_{i=1}^{N} \to \boldsymbol{\mathcal{R}}_{\boldsymbol{\theta}}. \tag{1.15}$$

Here, a training set comprising N paired examples, $\{(\boldsymbol{g}^{(i)}, \boldsymbol{u}^{(i)})\}_{i=1}^{N}$, is used to learn an inverse operator $\boldsymbol{\mathcal{R}}_{\boldsymbol{\theta}}$, where $\boldsymbol{\theta}$ represents the trainable parameters. In practice, this operator is implemented as a deep neural network $\boldsymbol{\mathcal{N}}_{\boldsymbol{\theta}}$, and once trained, the reconstructed image is obtained by

$$\hat{\boldsymbol{u}} = \boldsymbol{\mathcal{N}}_{\hat{\boldsymbol{\theta}}}(\boldsymbol{g}). \tag{1.16}$$

During training, algorithms such as stochastic gradient descent (SGD) [45] are employed to minimize a loss function, typically the mean squared error (MSE), across the training samples:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \|\boldsymbol{\mathcal{N}}_{\boldsymbol{\theta}}(\boldsymbol{g}^{(i)}) - \boldsymbol{u}^{(i)}\|_2^2. \tag{1.17}$$

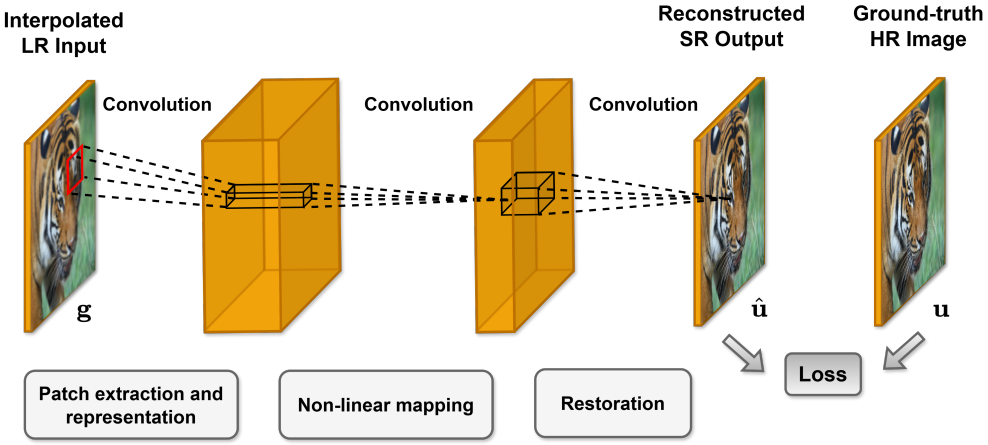


Figure 1.5: An illustration of SRCNN [44], one of the earliest deep learning CNNs that replaced traditional image restoration methods. Rather than explicitly defining an inverse operator, SRCNN learns it from low-resolution (LR) and high-resolution (HR) image pairs by minimizing a pixel-wise loss (e.g., mean squared error). The network has three stages: (1) a convolutional layer that extracts patch-based features from the interpolated LR input, (2) a layer that maps these features into a higher-dimensional space, and (3) a final layer that reconstructs the super-resolved (SR) image by combining the mapped features. SRCNN leverages large datasets to learn complex implicit priors, achieving high-quality restoration in a single forward pass. It outperforms many classical methods based on handcrafted priors and iterative optimization.

By learning the inverse mapping directly from data, deep learning approaches can *implicitly* capture complex image priors and degradation models that are difficult to characterize analytically. As a result, these methods often outperform traditional model-based techniques, achieving significant advancements in inverse problems such as denoising, deblurring, and super-resolution [46–48].

Convolutional Neural Networks

Convolutional neural networks (CNNs) [49] have long served as the backbone of deep image restoration. Early successes, such as the Super-Resolution CNN (SRCNN) proposed by Dong et al. [44], demonstrated that even a modest three-layer CNN (Figure 1.5) could learn to upsample low-resolution images, outperforming then state-of-the-art sparse-coding methods. This breakthrough paved the way for deeper architectures; for example, Very Deep Super-Resolution (VDSR) [50] leveraged a 20-layer network with residual learning (ResNet) [51] to further enhance super-resolution performance. DnCNN [52] employed residual CNNs to achieve superior denoising compared to classical filtering or advanced model-based approaches such as BM3D [24]. We used CNNs in Sections 5.2-5.4 as core building blocks to tackle joint demosaicking-deblurring problem and fluorescence imaging restoration.

A key advantage of CNN-based approaches is their computational efficiency. Once trained, restoration is accomplished in a single forward pass, unlike the iterative solvers commonly used in classical methods. Moreover, CNNs, which incorporate activation functions, inherently act as nonlinear operators that adapt to input features, enabling them to handle structured noise and artifacts that are challenging to model analytically. This efficiency, however, comes at the expense of requiring large, representative training datasets and a careful tuning to avoid overfitting. In practice, techniques such as data augmentation, early stopping, and weight regularization are employed to enhance generalization [53].

Vision Transformers and Generative Networks

More recently, attention-based models have emerged as a powerful alternative to CNNs. Vision Transformers (ViTs) [54] utilize self-attention mechanisms [55] to capture long-range dependencies that are challenging for convolutional filters with limited receptive fields. Models such as the Image Processing Transformer (IPT) [56] and SwinIR [57] have demonstrated state-of-the-art performance in a variety of inverse problems by leveraging global context.

Deep learning models for image restoration are often optimized using pixel-wise losses (e.g., MSE), which can achieve high peak signal-to-noise ratio (PSNR) [32] but tend to produce overly smooth images that miss fine textures and details. This limitation has motivated the exploration of perceptual losses like Structural Similarity Index Measure (SSIM) [58] or Learned Perceptual Image Patch Similarity (LPIPS) [59] and, ultimately, adversarial training.

Generative adversarial networks (GANs), introduced by Goodfellow et al. [60], have become a major milestone in the evolution of deep image restoration. In the GAN framework, two neural networks, a generator $\mathcal{N}_G$ and a discriminator $\mathcal{N}_D$, are trained simultaneously in an adversarial setting: while $\mathcal{N}_G$ strives to produce restored images that are indistinguishable from clean images, $\mathcal{N}_D$ learns to differentiate between

genuine and generated outputs. This adversarial interplay encourages the generator to produce images with high perceptual quality, complementing the fidelity term (e.g., L2) in the loss function. For instance, SRGAN [61] demonstrated that incorporating an adversarial loss into the training process yields super-resolved images with significantly enhanced sharpness and fine details compared to those generated by models optimized solely with MSE. An illustration of a conditional GAN framework for super-resolution is shown in Figure 1.6. Similarly, DeblurGAN [62] employed a conditional GAN to map blurry inputs to sharp outputs, effectively harnessing the discriminator as a learned *implicit regularizer*.

Despite their success, GANs can be challenging to train, motivating researchers to explore alternative generative paradigms such as diffusion models, which offer greater stability and improved image fidelity (see Section 2.1).

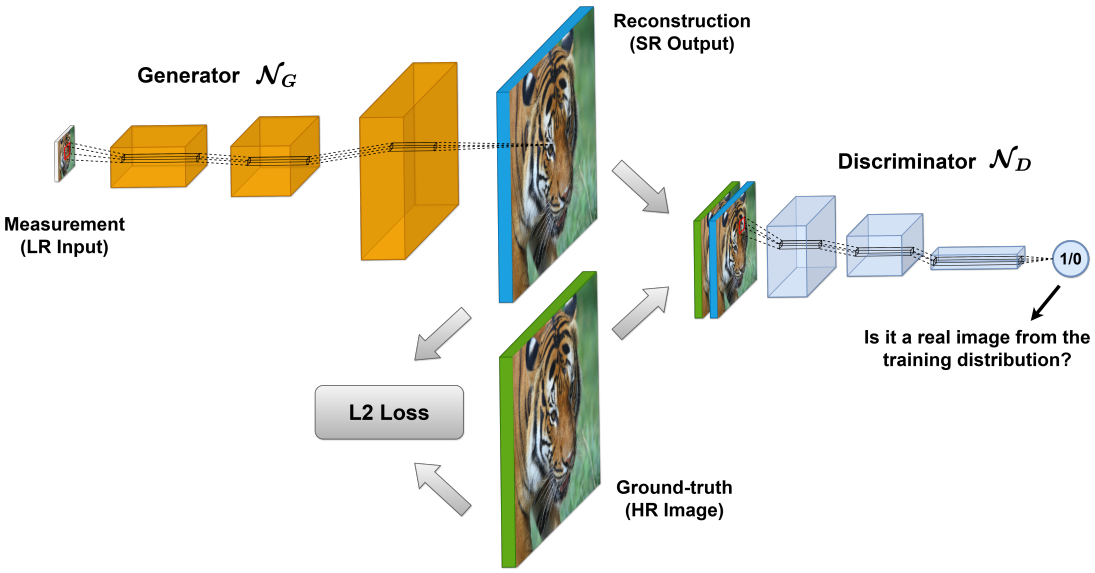


Figure 1.6: An illustration of a conditional GAN framework for image super-resolution. The generator $\boldsymbol{\mathcal{N}_G}$ receives a low-resolution (LR) input and produces a super-resolved (SR) output, while a pixel-wise (e.g., L2) loss ensures fidelity to the ground-truth high-resolution (HR) image. Concurrently, the discriminator $\boldsymbol{\mathcal{N}_D}$ distinguishes between real HR images and generated outputs, guiding the generator to synthesize sharper, more realistic details. By placing the generator and discriminator in a competitive setting, adversarial training encourages the production of highly realistic outputs that are difficult to achieve with purely pixel-wise objectives. Although illustrated here for super-resolution, the same principle can be applied to other inverse problems such as deblurring or denoising.

Self-supervised Learning

Generative frameworks provide substantial benefits in scenarios where aligned training image pairs are missing. For instance, Cycle-Consistent Generative Adversarial Network (CycleGAN) [63], as depicted in Figure 1.7, introduces a cycle-consistency loss that facilitates the learning of mappings between degraded and clean image domains using unpaired data. This approach is particularly valuable in real-world restoration tasks, such as in computed tomography or fluorescence microscopy [64, 65], where acquiring perfectly aligned ground-truth images is often impractical or even infeasible.

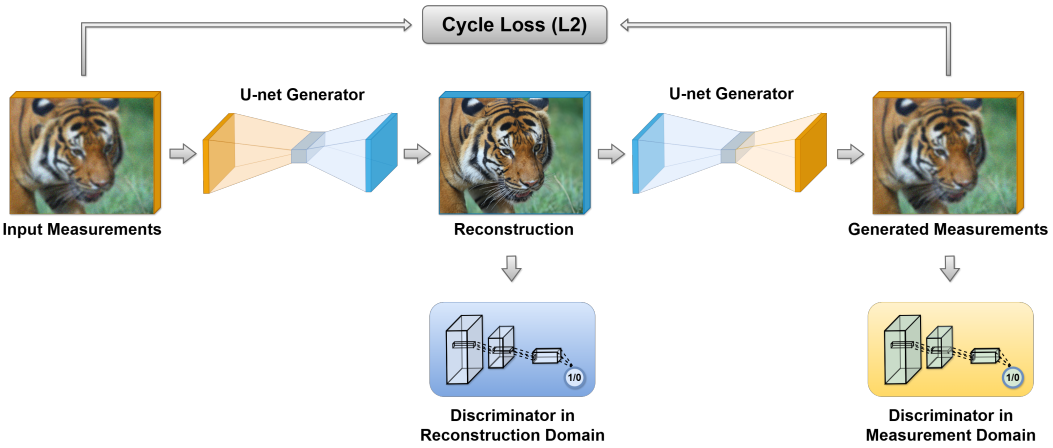


Figure 1.7: A schematic illustration of a cycle-consistent generative adversarial framework (CycleGAN) [63], which enables learning mappings between a "measurement" domain (e.g., blurred) and a "reconstruction" domain using unpaired data. Each domain has its own generator-discriminator pair, commonly a U-Net [66] generator and a PatchGAN [67] discriminator to assess local realism. A cycle-consistency loss (e.g., L_2) enforces that an image translated from one domain to the other and back remains close to its original form. This approach is particularly advantageous in real-world image restoration in computed tomography or fluorescence microscopy [64, 65], where perfectly aligned ground-truth data are unavailable. By leveraging adversarial training and cycle-consistency, the model can capture complex transformations and produce perceptually convincing outputs.

Moreover, when ground-truth data is completely unavailable, CycleGAN can operate in a fully *self-supervised* mode by relying on internal consistency signals derived directly from the measurement data. For instance, consider the diSPIM problem defined in Equation (1.3). In Section 5.3, an approach is presented in which a U-net generator [66] produces a 3D reconstruction that is then degraded to mimic the measurement process, ensuring consistency with the acquired data. A PatchGAN-based discriminator [67] further reinforces uniform detail by mapping high-resolution lateral details onto the axial view. This cycle consistency creates an intrinsic self-supervisory signal, enabling the recovery of fine details without the need of ground-truth data.

Alternative self-supervised paradigms have emerged that eliminate the need for external training datasets. One seminal example is the *deep image prior* (DIP) approach proposed by Ulyanov et al. [68]. DIP is based on the observation that the architecture of a randomly initialized CNN can serve as an effective image prior, as illustrated in Figure 1.8. In this approach, a network $\boldsymbol{\mathcal{N}}_{\boldsymbol{\theta}}$ is fitted to a single degraded image $\boldsymbol{g}$ by minimizing the reconstruction loss:

$$\min_{\boldsymbol{\theta}} \ \|\boldsymbol{\mathcal{D}}(\boldsymbol{\mathcal{N}}_{\boldsymbol{\theta}}(\boldsymbol{n})) - \boldsymbol{g}\|_2^2, \tag{1.18}$$

where $\boldsymbol{\mathcal{D}}$ denotes the degradation operator and $\boldsymbol{n}$ is a fixed random input (e.g., uniform noise). By optimizing this loss, the network recovers a clean estimate $\hat{\boldsymbol{u}} = \boldsymbol{\mathcal{N}}_{\boldsymbol{\theta}}(\boldsymbol{n})$ that explains the observed data $\boldsymbol{g}$. A crucial aspect of DIP is that the network's architectural bias acts as an *implicit regularization*, favoring natural image statistics and thereby preventing overfitting to noise when early stopping is applied. SelfDeblur [69] extends this concept to blind deblurring by jointly optimizing for both the latent sharp image and the unknown blur kernel (PSF), thereby leveraging the network's inherent architectural bias as an implicit regularizer to enforce natural image statistics. In Section 5.3, we combine DIP with CycleGAN to deblur 3D fluorescence microscopy images.

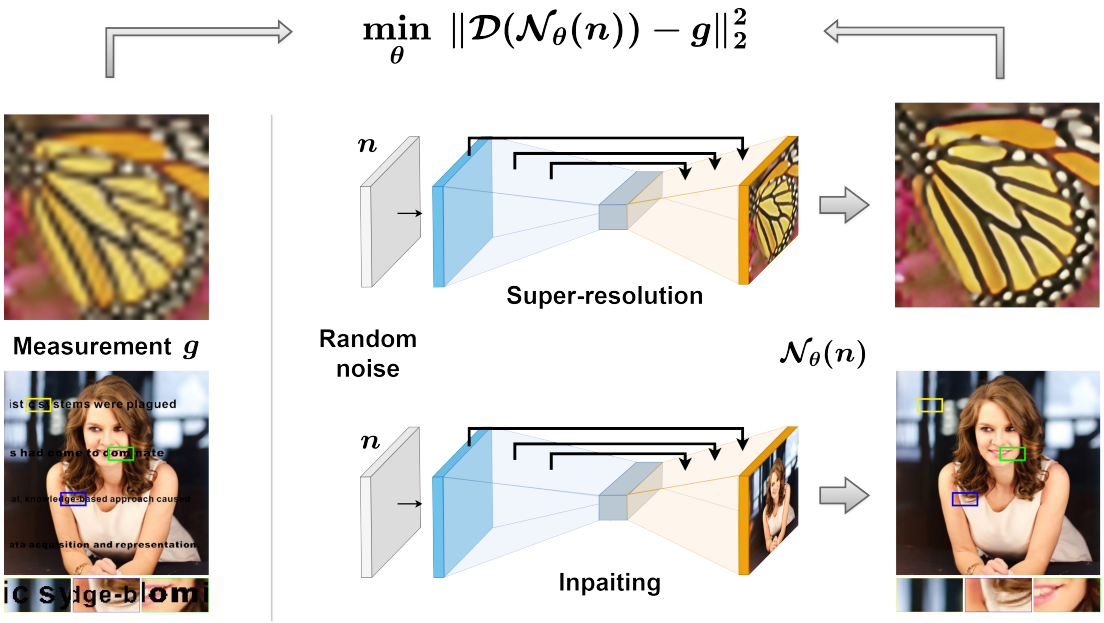


Figure 1.8: Illustration of the deep image prior (DIP) concept [68], where a randomly initialized network receives random noise as input and is optimized to reconstruct a single degraded measurement $\boldsymbol{g}$. Unlike traditional supervised approaches, DIP does not require external training data; instead, it exploits the inherent biases of the CNN architecture as an implicit prior for natural images. By merely fitting the network to the observed data, DIP provides high-quality restorations in various inverse problems (e.g., super-resolution in the top row, inpainting in the bottom row). This self-supervised paradigm has been adapted to tasks such as blind deblurring [69] and combined with adversarial frameworks [17], demonstrating its flexibility and effectiveness in image restoration.

Similarly to DIP, *implicit neural representations* [70], explained in Section 2.2, also leverage the inherent biases of network architectures as priors. Their built-in regularization restricts the solution space to natural, smooth functions, much like the effect seen in DIP, thereby helping to prevent overfitting to noise.

Other self-supervised methods like Noise2Noise [71] have demonstrated that denoising networks can be trained solely on pairs of independently corrupted images. Under the assumption of zero-mean noise, the optimal mapping learned from noisy inputs to noisy targets converges to the mapping from noisy inputs to the clean image. Extensions such as Noise2Void [72] and Noise2Self [73] further enable effective training from a single noisy image by employing masking strategies, thus eliminating the need for paired data altogether.

While data-driven deep learning methods have revolutionized image restoration by learning complex inverse mappings directly from data, they are not without limitations. Supervised approaches typically rely on large training datasets and face interpretability challenges, as their internal mechanisms can be difficult to analyze and explain. Similarly, self-supervised techniques, although they eliminate the need for ground-truth data, can be computationally intensive and time-consuming, since they often require optimizing the network separately for each new image, and they too present interpretability issues. To address these challenges, *hybrid* and model-based learning strategies have been developed that incorporate explicit forward models with flexible learned priors, thereby enhancing both robustness and efficiency.

HYBRID (or physics-informed) methods in image restoration build upon the classical framework detailed in Section 1.2. In the traditional approach, one formulates the problem as the optimization task (1.4), where the data fidelity term ensures consistency with the observed measurements and the regularization term $\boldsymbol{\Phi}(\boldsymbol{u})$ incorporates prior knowledge about the image (e.g., smoothness or sparsity). Hybrid approaches [74–76] enhance these classical models by integrating deep neural networks, thereby embedding data-driven priors and fine-tuned models directly into the established optimization framework.

Plug-and-Play Priors

We revisit the ADMM, Equations (1.11-1.13), where an auxiliary variable $\boldsymbol{z}$ is introduced to decouple the fidelity and regularization terms. A key observation is that the $\boldsymbol{z}$-update step (1.12) can be interpreted as a denoising operation. To understand this, we can rewrite the $\boldsymbol{z}$-update in a more suggestive form:

$$\boldsymbol{z}^{k+1} = \arg\min_{\boldsymbol{z}} \left\{ \frac{1}{2}\|\boldsymbol{z} - (\boldsymbol{u}^{k+1} + \boldsymbol{v}^k)\|_2^2 + \frac{\lambda}{\rho}\boldsymbol{\Phi}(\boldsymbol{z}) \right\}. \tag{1.19}$$

In this formulation, the term $\|\boldsymbol{z} - (\boldsymbol{u}^{k+1} + \boldsymbol{v}^k)\|_2^2$ forces the variable $\boldsymbol{z}$ to be close to the current estimate $\boldsymbol{u}^{k+1} + \boldsymbol{v}^k$, which can be thought of as a noisy version of the true image. The regularization term $\boldsymbol{\Phi}(\boldsymbol{z})$ acts to suppress the noise and enforce the desired image properties. By the definition of the proximal operator, Equation (1.19) is equivalent to:

$$\boldsymbol{z}^{k+1} = \operatorname{prox}_{\frac{\lambda}{\rho}\boldsymbol{\Phi}}\left(\boldsymbol{u}^{k+1} + \boldsymbol{v}^k\right). \tag{1.20}$$

This operator acts as a denoising function, mapping a noisy input to a cleaner version that adheres to the prior encoded by $\boldsymbol{\Phi}$.

Plug-and-play priors (PnP) introduced by Venkatakrishnan et al. [77] exploit this interpretation by replacing the proximal operator with an off-the-shelf denoiser (e.g., BM3D). Subsequent developments [78] extended this idea by using learned denoisers, denoted by $\boldsymbol{\mathfrak{D}}_\sigma$, typically deep CNNs trained to remove additive noise at specific levels σ. In the PnP framework, the $\boldsymbol{z}$-update is modified as follows:

$$\boldsymbol{z}^{k+1} = \boldsymbol{\mathfrak{D}}_\sigma\left(\boldsymbol{u}^{k+1} + \boldsymbol{v}^k\right). \tag{1.21}$$

This replacement bypasses the need to explicitly define or solve for $\boldsymbol{\Phi}(\boldsymbol{u})$; instead, the denoiser implicitly enforces a learned image prior based on large-scale training data. The success of this approach stems from the denoiser's ability to remove noise while preserving essential structures and textures, thereby effectively replacing the proximal operator of a designed regularizer.

Regularization by Denoising

Regularization by Denoising (RED) proposed by Romano et al. [79] builds on a similar idea but explicitly incorporates the denoiser into the regularization term. In RED, the regularizer is defined as

$$\boldsymbol{\Phi}(\boldsymbol{u}) = \frac{1}{2}\boldsymbol{u}^T\Big(\boldsymbol{u} - \boldsymbol{\mathfrak{D}}_\sigma(\boldsymbol{u})\Big). \tag{1.22}$$

This formulation directly ties the regularization penalty to the performance of the denoiser. By minimizing this term, the optimization process encourages the reconstructed image $\boldsymbol{u}$ to be close to its denoised version, effectively leveraging the power of the denoiser both for regularization and to improve convergence properties.

By extending the denoiser-centric framework of RED, Liu et al. introduced Regularization by Artifact-Removal (RARE) [80], which leverages deep networks trained to suppress a broader range of artifacts beyond mere noise. This approach enables the reconstruction algorithm to incorporate richer prior information, thereby enhancing robustness and image quality in more challenging imaging scenarios.

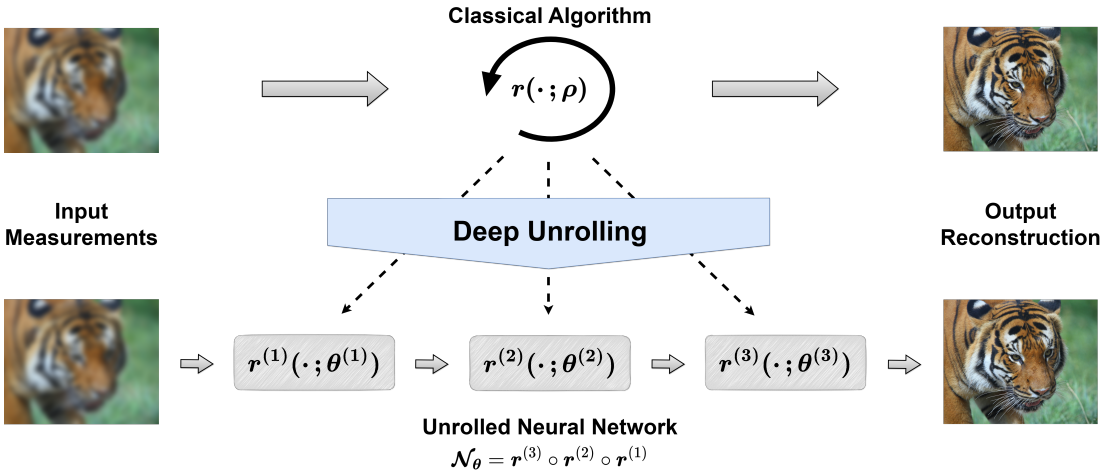


Figure 1.9: A conceptual illustration of deep unrolling for image deblurring. A classical iterative algorithm (e.g., ADMM) denoted by $\boldsymbol{r}(\cdot;\boldsymbol{\rho})$ with parameters $\boldsymbol{\rho}$ is "unrolled" into a fixed number of layers $\boldsymbol{r}^{(k)}$, each corresponding to one iteration of the original method. The input measurements (left) and the output reconstruction (right) are exemplified by a degraded and a restored tiger image, respectively. The unrolled layers incorporate learnable parameters $\boldsymbol{\theta} = \{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \boldsymbol{\theta}^{(3)}\}$, enabling end-to-end training for improved restoration performance.

Deep Unrolling

Deep algorithm unrolling [46, 81–83], also known as deep unfolding, bridges classical iterative algorithms and deep learning by "unrolling" each iteration of an optimization method (e.g., ADMM) into a corresponding layer in a neural network (see Figure 1.9). This transforms a classical update rule $\boldsymbol{r}(\cdot;\boldsymbol{\rho})$, parametrized by fixed $\boldsymbol{\rho}$, into a structured, multi-layer model:

$$\boldsymbol{\mathcal{N}}_{\boldsymbol{\theta}} = \boldsymbol{r}^{(K)} \circ \cdots \circ \boldsymbol{r}^{(1)}, \tag{1.23}$$

where each $\boldsymbol{r}^{(k)}$ has learnable parameters $\boldsymbol{\theta}^{(k)}$, and $\boldsymbol{\theta} = \{\boldsymbol{\theta}^{(1)}, \cdots, \boldsymbol{\theta}^{(K)}\}$.

Basic unrolled models keep the algorithmic structure largely intact, only learning certain parameters such as step sizes or penalty weights. More advanced approaches go further by replacing key steps (e.g., proximal operators) with learned neural modules. For instance, in unrolled ADMM, instead of using a fixed denoising prior in the $\boldsymbol{z}$-update, one can learn a denoising network $\mathfrak{D}_{\boldsymbol{\theta}_z}^{(k)}$, similar to PnP methods. The $\boldsymbol{u}$-update step can also be implemented as a trainable neural block $\boldsymbol{\mathcal{N}}_{\boldsymbol{\theta}_u}^{(k)}$ that directly embeds the degradation operator $\boldsymbol{\mathcal{D}}$ to enforce data fidelity. This formulation ensures that the physical model is incorporated at each iteration, guiding the network to respect forward-model constraints. The dual variable update in unrolled ADMM often remains unchanged.

As illustrated in Figure 1.9, deep unrolling takes an initial degraded input and applies a sequence of learned updates $\boldsymbol{r}^{(k)}$, mimicking classical iterations, to reconstruct a clean output image. Each layer's parameters, $\boldsymbol{\theta}^{(k)} = \{\boldsymbol{\theta}_{\boldsymbol{u}}^{(k)}, \boldsymbol{\theta}_{\boldsymbol{z}}^{(k)}\}$ in the ADMM-based example, are learned end-to-end using supervised learning. This hybrid formulation maintains the interpretability and convergence behavior of classical algorithms, while improving performance through data-driven learning. The result is a task-specific, trainable network $\boldsymbol{\mathcal{N}}_{\boldsymbol{\theta}}$ that often outperforms traditional iterative approaches in both speed and quality.

Deep Equilibrium Models

Deep equilibrium models (DEQ) [84] build on deep unrolling and go even further by defining the network implicitly through its fixed point rather than by stacking a predetermined number of layers. In DEQ, the solution $\hat{\boldsymbol{u}}$ is defined by the equilibrium condition

$$\hat{\boldsymbol{u}} = \boldsymbol{\mathcal{N}}_{\boldsymbol{\theta}}(\hat{\boldsymbol{u}}; \boldsymbol{g}), \tag{1.24}$$

where $\boldsymbol{\mathcal{N}}_{\boldsymbol{\theta}}$ is a learnable update function designed to mimic a single iteration of a traditional optimization algorithm by taking as input an image estimate along with the measurement data $\boldsymbol{g}$ and producing an updated estimate that better adheres to the forward model and desired image characteristics. DEQ seeks a fixed point through iterative refinement (using fixed-point iterations or implicit differentiation), thereby decoupling the number of model parameters from the number of iterations required for convergence. This yields a compact yet powerful representation while maintaining efficiency during both training and inference.

The methods discussed above illustrate the trend toward hybrid strategies that merge the structure of classical model-based formulations with the learning capacity of deep networks. By integrating explicit knowledge of the forward operator with data-driven priors (either through advanced denoisers or via learned iterative schemes) these hybrid approaches achieve both interpretability and state-of-the-art restoration performance.

Modern Trends in Image Restoration 2

Recent advances in image restoration have embraced the integration of deep generative models into classical inversion frameworks. Diffusion models, for example, have emerged as powerful learned priors operating through the iterative refinement of noise. Originally devised to generate new images from learned distributions, this process effectively regularizes the inversion task. We close this chapter with a brief introduction to implicit neural representations, a fundamentally different approach to addressing inverse problems that has become the primary focus of our recent research (Sections 5.4-5.5).

2.1 Bayesian Approach and Diffusion Models

In Section 1.2, we formulated classical restoration methods as variational optimization problems. An alternative yet closely related viewpoint is provided by the Bayesian paradigm [85], which provides a principled framework to tackle ill-posed problems by introducing prior knowledge as a probability distribution. Bayes' theorem allows us to compute the posterior distribution of the unknown image $\boldsymbol{u}$ given the observation $\boldsymbol{g}$:

$$p(\boldsymbol{u}\,|\,\boldsymbol{g}) \propto p(\boldsymbol{g}\,|\,\boldsymbol{u})\,p(\boldsymbol{u})\,, \tag{2.1}$$

where $p(\boldsymbol{g}\,|\,\boldsymbol{u})$ is the likelihood, defined by the forward model and noise statistics (e.g. Equation (1.1)) and $p(\boldsymbol{u})$ is the prior distribution reflecting our knowledge of plausible images. In other words, among all images that could explain $\boldsymbol{g}$, we prefer those that are *a priori* more likely. This Bayesian formulation is very general and modular, and can be leveraged in different ways.

A common practical approach is maximum a posteriori (MAP) estimation, which finds the single most probable $\boldsymbol{u}$ given $\boldsymbol{g}$. Maximizing the posterior probability in Equation (2.1) is equivalent to minimizing the negative log-posterior:

$$\hat{\boldsymbol{u}}_{MAP} = \arg\min_{\boldsymbol{u}} \Big\{ -\log p(\boldsymbol{g}|\boldsymbol{u}) \;-\; \log p(\boldsymbol{u}) \Big\}\,. \tag{2.2}$$

Here, we can define the data fidelity term as $\boldsymbol{\mathcal{L}}(\boldsymbol{u}) = -\log p(\boldsymbol{g}|\boldsymbol{u})$ and the regularization term as $\lambda\boldsymbol{\Phi}(\boldsymbol{u}) = -\log p(\boldsymbol{u})$. Thus, MAP estimation naturally leads to the classical variational formulation of the inverse problem in Equation (1.4). For example, assuming a smoothness prior for $\boldsymbol{u}$ might lead to using TV regularization, while employing a quadratic penalty $\boldsymbol{\Phi}(\boldsymbol{u}) = \|\boldsymbol{u}\|_2^2$ results in the well-known Tikhonov regularization.

Furthermore, if we assume that the noise $\boldsymbol{n}$ in Equation (1.1) is independent and Gaussian – that is, $\boldsymbol{n} \sim \mathcal{N}(0, \sigma^2\boldsymbol{I})$, where σ^2 is the noise variance – the likelihood can be written as

$$p(\boldsymbol{g}|\boldsymbol{u}) \propto \exp\!\left(-\frac{1}{2\sigma^2}\|\boldsymbol{\mathcal{D}}(\boldsymbol{u}) - \boldsymbol{g}\|_2^2\right). \tag{2.3}$$

Thus, minimizing the negative log-likelihood $\boldsymbol{\mathcal{L}}(\boldsymbol{u})$ recovers the familiar least-squares error, as in Equation (1.5).

Classical methods use handcrafted priors $\boldsymbol{\Phi}(\boldsymbol{u})$ to encode image knowledge, though these are often too simplistic. Standard deep learning approaches learn implicit priors from $(\boldsymbol{g}, \boldsymbol{u})$ pairs, but they are typically task-specific and require a separate model for each problem. More recently, generative models (such as diffusion models) have emerged.

Although they also demand substantial data, they learn the complete distribution $p(\boldsymbol{u})$ directly. This learned prior can then be combined with the known likelihood $p(\boldsymbol{g}|\boldsymbol{u})$ via Bayes' rule (2.1), offering a universal framework for addressing various inverse problems in imaging.

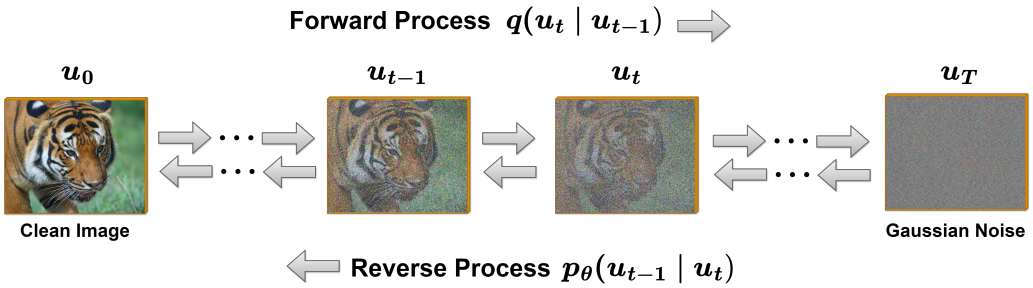


Figure 2.1: Schematic illustration of the forward (and reverse) diffusion process used in diffusion models [86]. An original image is gradually corrupted by Gaussian noise across multiple steps until it becomes nearly random. By learning to reverse this noising process, diffusion models capture rich image priors that can be integrated with known forward operators (likelihoods) to solve ill-posed inverse problems. Unlike classical model-based approaches that rely on handcrafted priors (or supervised deep learning methods that often require large paired datasets) diffusion models provide a robust generative framework that can address a wide range of restoration tasks through iterative refinement.

Diffusion Models

Diffusion models, also known as denoising diffusion probabilistic models or score-based generative models [86–88], learn the full image distribution $p(\boldsymbol{u})$ by reversing a gradual noising process. In the forward process (Figure 2.1), a clean image $\boldsymbol{u}_0$ is progressively perturbed by Gaussian noise over T steps, such that at each step

$$q(\boldsymbol{u}_t \,|\, \boldsymbol{u}_{t-1}) = \mathcal{N}\Big(\boldsymbol{u}_t; \sqrt{1-\beta_t}\,\boldsymbol{u}_{t-1},\ \beta_t \boldsymbol{I}\Big), \tag{2.4}$$

for $t = 1, ..., T$ with $0 < \beta_t < 1$. Composing these transitions yields

$$q(\boldsymbol{u}_t \,|\, \boldsymbol{u}_0) = \mathcal{N}\Big(\boldsymbol{u}_t; \sqrt{\bar{\alpha}_t}\,\boldsymbol{u}_0,\ (1-\bar{\alpha}_t)\boldsymbol{I}\Big), \tag{2.5}$$

where $\bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s)$. For a well-chosen schedule, $\bar{\alpha}_t$ diminishes with t so that $\boldsymbol{u}_T$ approaches a standard normal distribution.

The reverse process (Figure 2.1) can be modeled by a parameterized Markov chain $p_\theta(\boldsymbol{u}_{t-1} \,|\, \boldsymbol{u}_t)$ that approximates the true reverse conditionals $q(\boldsymbol{u}_{t-1} \,|\, \boldsymbol{u}_t, \boldsymbol{u}_0)$ [86]. This is typically defined as

$$p_\theta(\boldsymbol{u}_{t-1} \,|\, \boldsymbol{u}_t) = \mathcal{N}\Big(\boldsymbol{u}_{t-1};\ \boldsymbol{\mu}_\theta(\boldsymbol{u}_t, t),\ \beta_t \boldsymbol{I}\Big), \tag{2.6}$$

with the mean computed from the predicted noise $\boldsymbol{\epsilon}_\theta(\boldsymbol{u}_t, t)$ as

$$\boldsymbol{\mu}_\theta(\boldsymbol{u}_t, t) = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\boldsymbol{u}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\,\boldsymbol{\epsilon}_\theta(\boldsymbol{u}_t, t)\right). \tag{2.7}$$

The noise prediction function, $\boldsymbol{\epsilon}_\theta(\boldsymbol{u}_t, t)$, is typically a deep neural network with a U-net backbone [66] often enhanced with attention mechanisms [55]. The training

objective is to minimize the error between the true noise and the network's prediction. Specifically, the loss is defined as

$$L_\theta = \mathbb{E}_{\boldsymbol{u}_0,\,\boldsymbol{\epsilon},\,t}\left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta\left(\sqrt{\bar{\alpha}_t}\,\boldsymbol{u}_0 + \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\epsilon},\; t\right)\right\|^2, \tag{2.8}$$

where $\boldsymbol{\epsilon}$ is sampled from $\mathcal{N}(0, \boldsymbol{I})$. After training, generation is performed by initializing $\boldsymbol{u}_T$ from a standard normal distribution $\mathcal{N}(0, \boldsymbol{I})$ and iteratively applying the reverse transitions $p_\theta(\boldsymbol{u}_{t-1} \,|\, \boldsymbol{u}_t)$ to obtain $\boldsymbol{u}_0$.

Unlike GANs, which rely on unstable adversarial training, diffusion models generate images through a sequence of T iterative refinement steps. By learning a comprehensive generative prior directly from data and effectively integrating measurement fidelity, these models offer a flexible and robust approach to solving a wide range of inverse imaging problems.

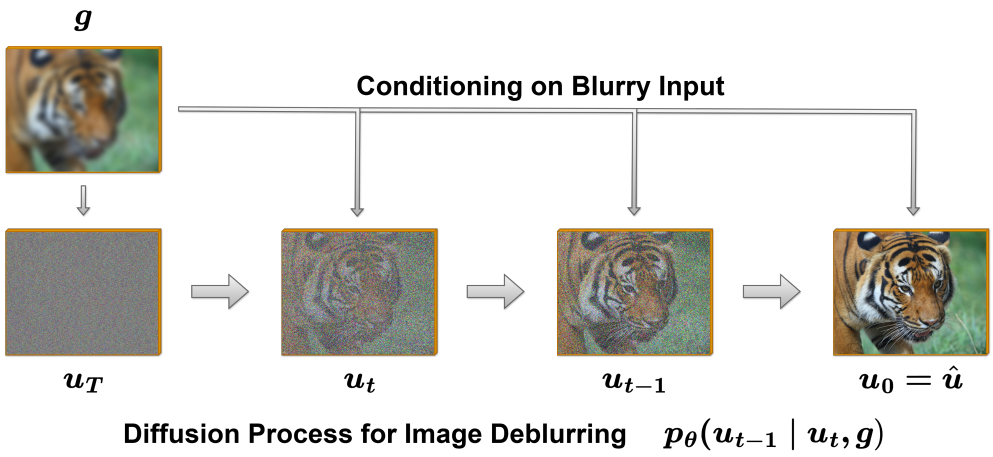


Figure 2.2: Schematic illustration of a diffusion-based image restoration process for deblurring. A measured image $\boldsymbol{g}$ serves as a conditioning signal, which can be incorporated either through an end-to-end learned model (e.g., SR3 [89]) or in a plug-and-play fashion (e.g., Denoising Diffusion Restoration Models [90]). Starting from random noise $\boldsymbol{u}_T$, the model iteratively refines its estimate $\boldsymbol{u}_t$ until it converges to a deblurred output $\boldsymbol{u}_0$.

Diffusion Models in Inverse Problems

In the context of inverse problems, the diffusion model acts as a learned prior $p_\theta(\boldsymbol{u}_0)$. By incorporating the likelihood $p(\boldsymbol{g} \,|\, \boldsymbol{u})$ derived from the forward model, the restoration process can be guided toward solutions that are both plausible and consistent with the observed data (see Figure 2.2).

Several strategies have been proposed to incorporate the likelihood into diffusion-based inverse problems [91, 92]. One approach trains an end-to-end conditional model that learns the reverse process $p_\theta(\boldsymbol{u}_{t-1} \,|\, \boldsymbol{u}_t, \boldsymbol{g})$ by directly conditioning on the observed data (e.g., via concatenation or cross-attention) to sample from posterior $p(\boldsymbol{u} \,|\, \boldsymbol{g})$. For example, Saharia et al. [89] trained a diffusion model (SR3) conditioned on low-resolution images to perform super-resolution. While this approach demands a large paired dataset and custom training per problem, it enables fast inference.

Alternatively, plug-and-play style methods [93–95] integrate measurement consistency directly into the reverse process. In its simplest form, one applies a gradient-based

update after each step, typically by minimizing a term such as $\|\boldsymbol{\mathcal{D}}(\boldsymbol{u}) - \boldsymbol{g}\|_2^2$. This is analogous to iterative techniques like PnP (explained in the previous chapter), where the degradation operator $\boldsymbol{\mathcal{D}}$ guides the reconstruction toward solutions consistent with the forward model. Another strategy uses a reference image either to align low-frequency components during diffusion [96] or to perform exact conditional sampling from a Gaussian posterior [90].

To address general and even nonlinear inverse problems, Chung et al. introduced Diffusion Posterior Sampling [97], which alternates between unconditional generation and measurement-guided gradient descent. Beyond classical inverse problems, diffusion models extend to domain-specific tasks. In medical imaging, they facilitate compressed sensing MRI, producing reconstructions that closely match acquired k-space measurements [98]. They also enable blind restoration, where the degradation model is partially or entirely unknown [99].

A key strength of diffusion-based inverse problem solvers is their ability to produce realistic, high-detail outputs that respect the measurements. On the downside, diffusion methods typically involve many iterations and thus can be computationally intensive. Ongoing research efforts such as one-step distillation aim to accelerate the sampling process without sacrificing reconstruction quality [100].

Latent Diffusion and Text-Driven Image Restoration

Another active direction formulates inversion in the latent space of generative models. Latent diffusion models [101] compress images into lower-dimensional representations, enabling more efficient restoration. However, working in the latent space poses unique challenges: the measurement must be accurately mapped into the latent domain, and the decoded image must remain consistent with the original observation. Recent studies have provided rigorous analyses and convergence guarantees in this context, with algorithms such as [102, 103] ensuring measurement consistency during latent sampling. Notably, "Gaussian is All You Need" [104] introduced a covariance-corrected likelihood objective that improves convergence to the true posterior without the need to backpropagate through the entire diffusion chain.

A particularly promising avenues are text-guided diffusion models [105]. Methods such as prompt tuning [106], regularization by text [107], and the SPIRE framework [108] condition the restoration process on natural language descriptions, thereby enabling user-driven semantic guidance alongside the physics-based measurements. These techniques help resolve ambiguities that cannot be addressed by purely pixel-wise or even perceptual fidelity terms, thereby opening up new possibilities for interactive inverse problem solving.

In conclusion, diffusion models offer a promising framework for inverse problems by integrating measurement consistency with learned priors that capture the complete image distribution. Recent surveys [92, 109–111] highlight their wide applicability. Future work should focus on efficiency [112] and exploring novel approaches like interactive restoration.

Real-world scenes are inherently continuous, yet measurement devices, such as digital camera sensors, capture only discrete samples, potentially missing fine details. To address this challenge, researchers have developed a promising approach that represents images as continuous functions. Implicit neural representations (INRs), also known as Neural Fields [113–115], embody this concept by encoding an image $\boldsymbol{u}$ as a coordinate-based neural network (Figure 2.3). In these models, spatial coordinates are fed into a multilayer perceptron (MLP) that generates the corresponding intensity values. This method bypasses the constraints of fixed pixel grids by allowing arbitrary sampling resolutions, and its storage requirements depend on scene complexity rather than pixel count.

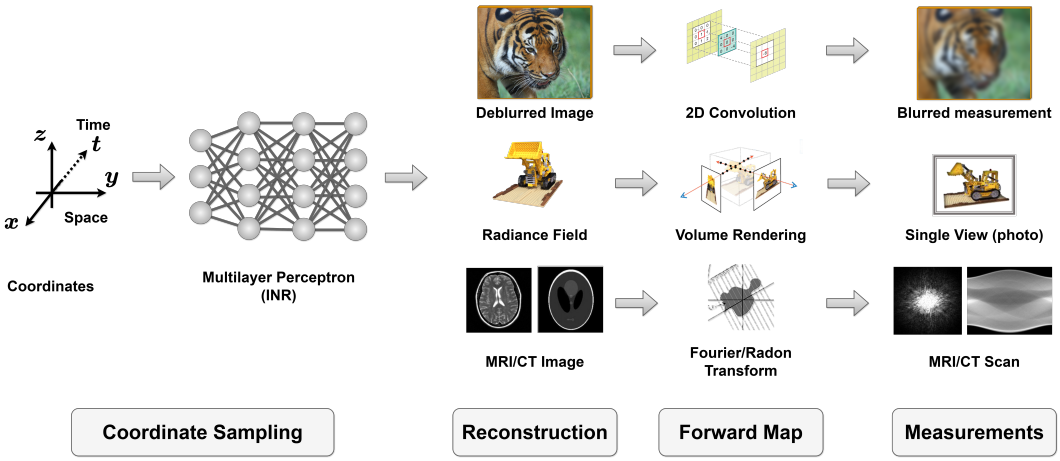


Figure 2.3: A high-level illustration of implicit neural representations (INRs), where a coordinate-based neural network (for example, a multilayer perceptron) replaces the traditional pixel grid. Given spatial (and potentially temporal and other) coordinates, the INR outputs the corresponding intensity or color values, allowing flexible sampling at arbitrary resolutions. By integrating physics-based forward operators (for example, convolution for deblurring, volume rendering for 3D scenes, or Radon transforms for tomography) into this framework, a wide range of inverse problems (such as image deblurring, multi-view 3D reconstruction, or CT imaging) can be tackled in a single end-to-end optimization. Because the model is fully differentiable, it is directly trained on measured data, ensuring high fidelity to the underlying measurements while leveraging the network's architecture as an implicit prior for natural image features.

Moreover, the differentiable nature of INRs facilitates their integration into end-to-end optimization frameworks, including inverse problem solvers that use physics-based forward models. In this process, the reconstructed image is implicitly encoded in a weight space that relies on the architecture as a prior for natural image features. Therefore, the network's inherent *implicit regularization* is used to improve reconstruction quality [116].

Mathematical Formulation

Formally, let $\boldsymbol{u} : \Omega \to \mathbb{R}^C$ be the unknown continuous image (with C channels) defined on a domain Ω. An INR models the image $\boldsymbol{u}$ as a coordinate-based MLP $\boldsymbol{u}_\theta(\mathbf{r})$, with parameters $\boldsymbol{\theta}$, and coordinates $\mathbf{r} \in \Omega$. Solving the inverse problem with an INR entails finding network parameters $\boldsymbol{\theta}$ such that $\boldsymbol{u}_{\boldsymbol{\theta}}$ explains the observations $\boldsymbol{g} = \boldsymbol{\mathcal{D}}(\boldsymbol{u}) + \boldsymbol{n}$.

This is typically done by minimizing a suitable loss function enforcing data fidelity, possibly with additional regularization $\boldsymbol{\Phi}(\boldsymbol{u_\theta})$:

$$\hat{\boldsymbol{\theta}} \;=\; \arg\min_{\boldsymbol{\theta}} \left\{ \| \boldsymbol{\mathcal{D}}(\boldsymbol{u_\theta}) - \boldsymbol{g} \|_2^2 + \lambda\, \boldsymbol{\Phi}(\boldsymbol{u_\theta}) \right\}. \tag{2.9}$$

In practice, the loss can be computed either by sampling the continuous network $\boldsymbol{u}_\theta$ at the pixel locations of $\boldsymbol{g}$ or by defining $\boldsymbol{\mathcal{D}}$ (and possibly the regularizer $\boldsymbol{\Phi}$) continuously and then discretizing their outputs for loss computation. In many INR approaches, no explicit regularizer is used ($\lambda = 0$); instead, the network's architecture serves as an implicit prior that discourages fitting spurious high-noise solutions. The optimization in Equation (2.9) is typically performed via gradient descent on $\boldsymbol{\theta}$, treating the problem analogously to a MAP estimation: the likelihood $p(\boldsymbol{g} \,|\, \boldsymbol{u})$ (from the forward model) is encoded by the data fidelity term, and any prior belief $p(\boldsymbol{u})$ can be encoded implicitly by the network $\boldsymbol{u_\theta}$, explicitly by a regularizer $\boldsymbol{\Phi}$, or both.

In practice, one might initialize $\boldsymbol{\theta}$ randomly and optimize until convergence or until the $\boldsymbol{\mathcal{D}}(\boldsymbol{u_\theta})$ fits $\boldsymbol{g}$ to an acceptable degree without overfitting noise (early stopping therefore acts as another form of regularization [117]). The result is a set of learned weights $\hat{\boldsymbol{\theta}}$, from which the restored image is obtained as $\hat{\boldsymbol{u}}(\boldsymbol{r}) = \boldsymbol{u}_{\hat{\boldsymbol{\theta}}}(\boldsymbol{r})$. Because $\boldsymbol{u_\theta}$ is continuous, the reconstruction $\hat{\boldsymbol{u}}$ can be queried at any resolution or coordinate, providing a super-resolved or geometry-adapted solution as needed.

Comparison with Other Methods

The INR approach stands apart from other inverse problem solvers in several respects. In contrast to methods that require external training data, such as supervised CNNs or transformers, the INR framework typically optimizes solely on the given measurement $\boldsymbol{g}$ (see Equation (2.9)), much like the DIP approach. However, while DIP employs a CNN that outputs an image on a fixed grid, the INR's coordinate-based MLP provides a truly continuous representation. This allows reconstructions to be sampled at arbitrary resolutions, offering increased flexibility.

Modern diffusion models, by comparison, leverage large-scale datasets to learn expressive priors that can generate high-frequency details and photo-realistic textures. Yet, their effectiveness comes with the cost of extensive pre-training and often task-specific conditioning, making them less adaptable in scenarios where data are scarce or the degradation $\boldsymbol{\mathcal{D}}$ is unique.

Similarly, supervised deep learning methods learn a mapping from $\boldsymbol{g}$ to $\boldsymbol{u}$ using paired data and achieve faster inference through a single forward pass. However, these models may not enforce strict measurement fidelity in the inverse problem setting and can be prone to errors when the test input deviates from the training distribution.

Standard INR optimization typically yields a single estimate, whereas generative models (such as diffusion models and GANs) can sample multiple plausible solutions to capture posterior uncertainty. Recent research, including the Implicit Diffusion Model (IDM) [118] for high-fidelity continuous image super-resolution, has integrated INR into the decoding phase of diffusion denoising frameworks. Although these approaches show considerable promise, fully harnessing INR within generative models remains an open research challenge.

Recent Developments in INRs

Recent advances in INRs have enabled precise signal representation and reconstruction using neural networks. Early breakthroughs, such as NeRF [70] for mapping 3D coordinates to radiance and density, and DeepSDF [119] for representing 3D shapes via continuous signed distance functions, demonstrated that MLPs can encode complex scenes. The Local Implicit Image Function (LIIF) [120] confirmed that INRs can effectively represent 2D images, enabling extrapolation to resolutions up to 30 times higher than the original. NeRV [121] even showed the capability to encode entire videos in neural networks. Furthermore, recent research has extended the INR framework to the realm of image compression [122, 123].

Early INR models, however, suffered severely from spectral bias [124–126]. Spectral bias is a phenomenon where MLPs exhibit a natural tendency to learn low-frequency functions more easily than high-frequency details. This issue was particularly pronounced in early models due to the use of ReLU activations, which often led to oversmoothing of fine details. One effective strategy to address spectral bias is positional encoding. Initially introduced in Fourier feature mappings [127] and later adopted by NeRF, positional encoding helps the network learn higher-frequency components more effectively. SIREN [116] further advanced this approach by substituting ReLU activations with sine functions, significantly enhancing the network's ability to capture fine details. Recent advancements in the field have introduced new techniques and architectures that further expand the INR toolkit. Gaussian INRs [128] provide improved frequency localization, while wavelet-based INRs (WIRE) [117] leverage multi-scale analysis to enhance robustness against noise.

Additionally, novel activation functions such as the hyperbolic oscillation function (HOSC) [129] and the sinus cardinal function (SINC) [130], inspired by the Shannon sampling theorem, extend the network's capacity to capture high-frequency details. Finally, architectures like High-Order Implicit Neural Representation (HOIN) [131] and FINER [132], which utilize variable-periodic activation functions, have emerged as strong contenders for establishing a new state-of-the-art INR backbone.

Conditioning has emerged as another critical factor in enhancing INR performance, as it integrates contextual or domain-specific information (often via latent embeddings) into the network. This additional guidance helps the network achieve higher-fidelity reconstructions and improved generalization. The most common conditioning approaches are modulation of activation functions [133–135], latent embedding concatenation [119], and hypernetwork frameworks [136].

Scalability and efficiency remain central challenges. Recent adaptive strategies like ACORN [137], multiresolution hash encoding [138], meta-learning [139], and hierarchical frameworks such as KiloNeRF [140] and MINER [141], have shown promise in accelerating training and inference. Furthermore, dictionary-based methods such as Neural Implicit Dictionary [142] offer compact representations.

Applications in Inverse Problems

INRs excel in many ill-posed inverse imaging scenarios. In super-resolution, models like LIIF [120] and IDM [118] learn continuous mappings from low-resolution images, enabling arbitrary resolution queries and high-quality reconstructions. For image denoising, wavelet-based INRs effectively filter noise while preserving structural details [117]. In medical imaging, methods such as CoIL [143] directly map spatial coordinates to inten-

sities for CT reconstruction, while techniques inspired by NeRP [144] extend these ideas to MRI. INRs also excel in image inpainting and interpolation, seamlessly filling missing regions by learning continuous functions [120]. For video interpolation, spatial-temporal consistency is achieved in models like CURE [145] and VideoINR [146]. Additional inverse problem applications include fundamental tasks such as deblurring [147, 148]. Interestingly, Xu et al. introduced an implicit neural signal processing network [149] to perform inverse problem tasks like image deblurring directly on INRs, without explicit decoding.

Collectively, these advancements demonstrate that INR-based models offer robust and efficient alternatives to traditional discrete methods across a wide range of inverse imaging challenges. Continued research is expanding the practical applications and enhancing the theoretical understanding of INRs, as highlighted in recent surveys [113–115].

Goals of the Thesis

3

Advancing the field of image restoration, particularly in addressing inverse problems, requires a careful integration of classical optimization methods and modern deep learning approaches. The challenges posed by real-world degradations such as blur, noise, and subsampling necessitate robust solutions that can generalize across diverse imaging conditions. This thesis explores these challenges through three key objectives:

- Enhance traditional deblurring techniques by refining Wiener filtering, addressing its limitations with an iterative approach that mitigates ringing artifacts while maintaining computational efficiency.
- Develop deep learning-based image restoration frameworks leveraging deep unrolling and self-supervised learning techniques to jointly tackle multiple restoration tasks (e.g., deblurring, demosaicking, denoising, super-resolution, and image fusion). Emphasis will be placed on effectively reducing restoration-induced artifacts while eliminating dependence on extensive paired training datasets.
- Establish implicit neural representations as a novel tool for image restoration tasks that have not been previously explored with INR (e.g., demosaicking); evaluate their ability to generalize across diverse degradations; and design a self-adaptive INR-based framework for robust image reconstruction under varying real-world conditions.

Structure of the Thesis 4

This thesis addresses the challenge of inverse problems in image restoration by developing innovative methods to reconstruct high-quality images from degraded observations across diverse applications. Two tables guide the reader through the contributions of this work. Table 4.1 outlines the restoration tasks addressed by each method, ranging from denoising to super-resolution. Table 4.2 illustrates the evolution of the techniques, from classical model-based methods to advanced hybrid and self-supervised strategies with implicit and explicit regularization. Organized as five interconnected studies, the thesis presents a cohesive body of work that bridges traditional image restoration approaches with modern, data-driven solutions.

	IWFT (Sec. 5.1)	**D3Net** (Sec. 5.2)	**Dual-Cycle** (Sec. 5.3)	**NeRD** (Sec. 5.4)	**INRID** (Sec. 5.5)
Denoising	✓	✓	✓	✓	✓
Deblurring	✓	✓	✓		✓
Demosaicking		✓		✓	✓
Deringing	✓	✓			
Image fusion			✓		
Super-resolution			✓	(✓)	(✓)

Table 4.1: Overview of inverse problem applications addressed in the thesis. This table summarizes the capabilities of each proposed method – IWFT, D3Net, Dual-Cycle, NeRD, and INRID – in tackling specific restoration tasks such as deblurring, demosaicking, deringing, image fusion, and super-resolution. Note that denoising is inherently incorporated to varying extents across all approaches through the application of appropriate regularization strategies, whether implicit or explicit. (✓) indicates that the method was not explicitly evaluated for super-resolution but may be capable of it due to the resolution-agnostic nature of implicit neural representations.

	IWFT (Sec. 5.1)	**D3Net** (Sec. 5.2)	**Dual-Cycle** (Sec. 5.3)	**NeRD** (Sec. 5.4)	**INRID** (Sec. 5.5)
Model-based	✓				
Hybrid		✓	✓		✓
Supervised		✓		✓	
Self-supervised			✓		✓
Explicit reg.	✓	✓			✓
Implicit reg.		✓	✓	✓	✓

Table 4.2: Summary of methodological attributes of the proposed methods. The table highlights the evolution from classical model-based techniques to modern deep learning frameworks, detailing aspects such as hybrid learning strategies, supervised versus self-supervised training, and the integration of explicit and implicit regularization, as discussed across the dissertation.

The first study tackles image deblurring using an iterative Wiener filtering approach named IWFT (Sec. 5.1). While traditional Wiener filtering is analytically elegant, it suffers from ringing artifacts near high-contrast regions. IWFT integrates a thresholding step within an ADMM-inspired framework to progressively refine the reconstruction, focusing on deblurring and suppressing ringing artifacts (i.e., deringing). As shown in the tables, this approach is rooted in a model-based framework with explicit regularization.

The second study introduces D3Net (Sec. 5.2), a deep unrolling method that reformulates iterative optimization as a structured neural network. By replacing handcrafted iterative updates with learnable convolutional filters, D3Net jointly handles demosaicking, deblurring, and deringing. It leverages supervised learning along with both implicit and explicit regularization, embodying a hybrid framework that combines classical interpretability with modern deep learning adaptability.

The third study presents Dual-Cycle (Sec. 5.3), a self-supervised framework tailored for multi-view fluorescence microscopy. Exploiting cycle-consistency in a CycleGAN-based architecture, Dual-Cycle fuses two perpendicular measurements to reconstruct high-resolution 3D images without the need for paired training data. This method extends its application to super-resolution and image fusion, as highlighted in the tables, and relies on implicit regularization through the DIP (Deep Image Prior).

In the fourth study, the focus shifts to Neural Field-Based Demosaicking (NeRD) (Sec. 5.4). By representing images as continuous functions via a coordinate-based multilayer perceptron with sinusoidal activations, NeRD facilitates smooth interpolation and better adaptation to image structures. Supported by a supervised CNN-based encoder and implicit regularization through INRs, NeRD distinguishes itself as the first method to leverage implicit neural representations for image demosaicking, achieving reconstruction quality on par with state-of-the-art approaches.

While supervised learning has driven significant advancements in image demosaicking, its reliance on fixed training distributions often leads to reduced performance when confronted with out-of-distribution examples. Addressing these limitations, the final study introduces INRID (Sec. 5.5), a fully self-supervised framework for image demosaicking. INRID dynamically optimizes network parameters on a per-image basis while incorporating both a Bayer-pattern consistency loss and an initial estimation-based regularization. This robust framework effectively handles real-world degradations such as blur and noise by merging self-supervision with hybrid learning and additional implicit regularization through INRs.

Collectively, these studies illustrate a progression from classical model-based methods toward hybrid approaches (e.g., deep unrolling), advancing further into self-supervised learning leveraging deep image priors, and ultimately reaching continuous implicit representations, culminating in a novel paradigm for image demosaicking. The following chapter details the specific contributions of each publication.

Publications 5

5.1 Paper 1 - IWFT

Filip Šroubek, Tomáš Kerepecký, and Jan Kamenický, "Iterative Wiener filtering for deconvolution with ringing artifact suppression," in *2019 27th European Signal Processing Conference (EUSIPCO)*. September 2019, pp. 1–5, IEEE

Image deblurring addresses the problem of recovering images degraded by convolution with a blur kernel, a fundamental challenge in computational imaging; leading to loss of fine details. While Wiener filtering is a well-established method for deconvolution, it is prone to ringing artifacts near strong edges due to the ill-posed nature of the problem. This work introduces *IWFT* (Iterative Wiener Filtering with Thresholding), an approach that refines the classical Wiener filter by incorporating an iterative optimization process inspired by the ADMM. The key innovation of the method lies in decomposing the deblurring process into a sequence of restoration and update steps, where each iteration progressively refines the image to suppress artifacts while preserving detail. The method operates entirely in the spatial domain, avoiding frequency-domain boundary artifacts, and introduces learned restoration and update filters that adapt to different types of blur and noise. The framework is flexible and can be extended to other restoration tasks, such as demosaicking and super-resolution. Experimental results (visually presented in Figure 5.1) confirm that the proposed iterative Wiener filtering method substantially reduces ringing artifacts compared to traditional Wiener filtering. For additional quantitative results and analysis of computational efficiency, refer to the reprinted paper.

(a) Blurred (b) Wiener (c) Proposed

Figure 5.1: Deconvolution without ringing artifacts: (a) the blurred input image and PSF (inset), (b) restoration using the Wiener filter, (c) restoration using the proposed iterative Wiener filtering and thresholding. Notice ringing artifacts in the Wiener solution (b) and their absence in the proposed method (c).

Contribution of the Paper

This paper introduced an iterative extension of Wiener filtering to address ringing artifacts in image deblurring. By integrating restoration and update filters within an optimization-inspired framework, the approach significantly enhances deconvolution quality (reduces ringing while preserving detail). The method avoids frequency-domain constraints, enabling a flexible and generalizable solution applicable to various inverse problems, including demosaicking and super-resolution.

Contribution of the Author

The main contribution of Tomáš Kerepecký in this work was to perform experiments and to provide final feedback on the manuscript. These efforts helped refine the overall study and supported the validation of its methodologies.

Tomáš Kerepecký and Filip Šroubek, "D3net: Joint demosaicking, deblurring and deringing," in *2020 25th International Conference on Pattern Recognition (ICPR)*. January 2021, pp. 1–8, IEEE

MODERN camera sensors capture images using a color filter array, with only one color channel recorded per pixel. To reconstruct a full RGB image, demosaicking estimates the missing color information. However, raw sensor data is further degraded by issues such as lens blur and sensor noise, which require multiple restoration steps. Traditional pipelines typically address these tasks independently, leading to suboptimal results as errors can propagate from one stage to the next. This paper presents a hybrid approach that leverages deep learning to jointly perform demosaicking, deblurring, and deringing, three critical processes for effective digital camera image restoration.

This work proposes *D3Net*, a CNN designed to perform joint restoration in an end-to-end manner. Inspired by model-based method IWFT (Section 5.1), the network structure mimics the iterative steps of classical deconvolution methods while replacing fixed filters with learnable parameters (see Figure 5.2). The unrolled architecture consists of a restoration layer "`rConv`" responsible for demosaicking and initial deblurring, followed by three update layers that iteratively refine the image and suppress ringing artifacts. D3Net retains interpretability and efficiency while significantly improving restoration quality over conventional approaches.

A key advantage of D3Net is its ability to learn effective restoration filters from a minimal amount of training data. Unlike conventional deep learning models that require large datasets, D3Net can be trained using only a single pair of degraded and ground-truth images. This is made possible by the structured nature of the network, which follows the principles of deep unrolling.

The method is evaluated against state-of-the-art demosaicking and deblurring techniques, demonstrating superior performance in both objective image-quality metrics (PSNR/SSIM) and visual clarity. The results further confirm that joint restoration leads to significantly better reconstructions compared to sequential approaches.

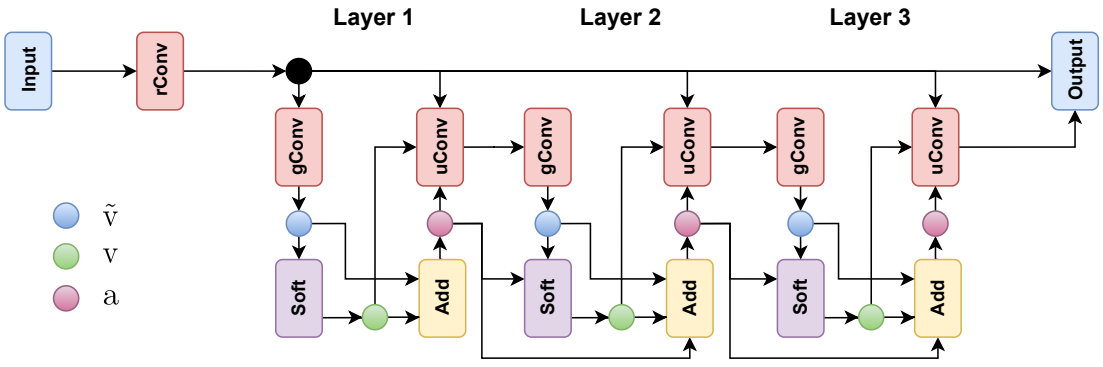


Figure 5.2: Architecture of the proposed D3Net based on the unrolled IWFT algorithm. Here, `rConv`, `gConv`, and `uConv` denote convolutional filters, `Soft` represents soft thresholding, and `Add` indicates a simple addition operation.

Contribution of the Paper

This paper introduces *D3Net*, a joint demosaicking, deblurring, and deringing network inspired by the deep unrolling of optimization algorithms. The network architecture mimics an ADMM-based iterative restoration process, ensuring both interpretability and high performance. By integrating all three restoration tasks into a single framework, the method avoids error accumulation and significantly outperforms sequential pipelines. The structured approach enables training on a single image pair while maintaining the ability to generalize, making it computationally efficient and practical for real-world imaging applications.

Contribution of the Author

The main contribution of Tomáš Kerepecký in this paper involved extensive research on deep unrolling, formulating the core idea, and developing the methods described. Tomáš Kerepecký was responsible for preparing data from the relevant databases, thoroughly testing all mentioned algorithms and drafting the initial version of the manuscript. The final version was then refined collaboratively by both authors. Tomáš Kerepecký presented the paper as an oral contribution at the international conference.

Tomáš Kerepecký, Jiaming Liu, Xuan Wei Ng, David W. Piston, and Ulugbek S. Kamilov, "Dual-cycle: Self-supervised dual-view fluorescence microscopy image reconstruction using cyclegan," in *2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. June 2023, pp. 1–5, IEEE

This paper extends the research into self-supervised learning for image restoration, with a particular focus on fluorescence microscopy. Three-dimensional fluorescence microscopy often suffers from spatial-resolution anisotropy, where the resolution in the axial direction is significantly lower than in the lateral plane due to optical diffraction limits and system aberrations. Traditional computational approaches attempt to mitigate this issue using model-based multi-view deconvolution, where images from multiple viewpoints are fused to reconstruct a higher-resolution image. However, these methods rely on accurately designed fusion strategies and precise knowledge of the imaging system's PSF. On the other hand, using supervised deep learning techniques for this task is limited by the lack of sufficient ground truth data.

This work introduces *Dual-Cycle*, a self-supervised deep learning framework designed to perform joint deconvolution and fusion of dual-view fluorescence microscopy images. The method is inspired by the CycleGAN approach, leveraging a cycle-consistency loss to enable high-resolution image reconstruction without requiring paired training data. Dual-Cycle comprises two main components: a dual-view generator that fuses perpendicular views of the sample to enhance spatial resolution, and a degradation model that incorporates estimated PSFs to guide the learning process. By embedding physics-based priors into the learning framework, the network effectively reconstructs isotropic-resolution 3D images while adapting to variations in imaging conditions.

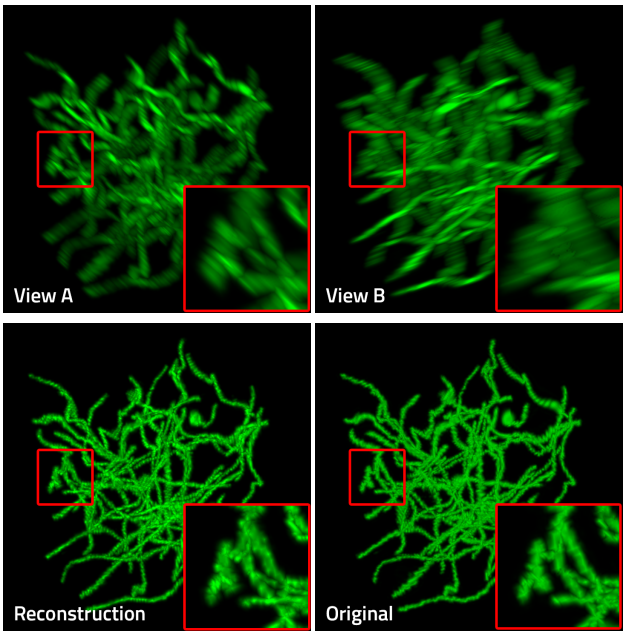


Figure 5.3: Given two views with anisotropic resolution, the Dual-Cycle reconstructs a 3D image with isotropic resolution given two views. A and B are two perpendicular views of the same sample.

Unlike traditional deconvolution techniques, which require manually tuned priors and are sensitive to system-specific parameters, Dual-Cycle learns an optimal restoration function directly from measurements. The model is trained solely on dual-view measurements, enabling it to generalize to various imaging conditions.

Experimental validation was conducted on two types of data: synthetic microscopy datasets (see Figure 5.3) and real fluorescence microscopy acquisitions (details in the paper). The experiments demonstrate that Dual-Cycle significantly improves axial resolution while preserving fine structural details. Furthermore, the results indicate that the method outperforms traditional multi-view deconvolution techniques by achieving higher PSNR and superior visual quality.

Contribution of the Paper

This paper presents *Dual-Cycle*, a self-supervised framework for dual-view fluorescence microscopy image reconstruction, based on cycle-consistent generative networks. The proposed method eliminates the need for external ground-truth supervision by leveraging the inherent structural consistency in multi-view imaging. The framework integrates a physics-guided degradation model, enabling the network to learn an optimal deconvolution strategy. Experimental results on both synthetic and real microscopy datasets confirm that Dual-Cycle significantly improves axial resolution and outperforms existing multi-view fusion methods.

Contribution of the Author

The main contribution of Tomáš Kerepecký in this work consisted of outlining the core conceptual framework, conducting and analyzing the principal experiments, and preparing the final draft of the manuscript. This involved designing the study methodology, gathering and evaluating the data, and ensuring the coherence of the paper's overall structure and presentation. Tomáš Kerepecký presented the work at the international conference.

Tomáš Kerepecký, Filip Šroubek, Adam Novozámský, and Jan Flusser, "Nerd: Neural field-based demosaicking," in *2023 IEEE International Conference on Image Processing (ICIP)*. October 2023, pp. 1735–1739, IEEE

DEMOSAICKING, a fundamental step in image processing that reconstructs full-color images from raw sensor data, has traditionally relied on computationally efficient model-based methods, such as bilinear or edge-directed interpolation, which often struggle with fine textures and color artifacts. More recent deep learning-based approaches, particularly CNNs and transformers, have significantly improved demosaicking performance, but they are often constrained by their reliance on pixel-grid-based processing, limiting their adaptability across different resolutions.

This work presents *NeRD* (Neural Field-Based Demosaicking), the first application of implicit neural representations to the problem of demosaicking (depicted in Figure 5.4). Instead of treating images as discrete pixel arrays, NeRD represents images as continuous functions parameterized by a neural network. The core of NeRD is a coordinate-based multilayer perceptron with sine activation functions, which maps spatial coordinates to corresponding RGB values. The demosaicking process is guided by a hybrid architecture combining a residual CNN and a U-Net-based encoder, which extracts local image priors from the Bayer pattern and conditions the implicit neural representation. This allows the model to leverage both local and global features, improving the consistency and accuracy of color reconstruction.

The method is evaluated against state-of-the-art demosaicking techniques, including CNN-based and transformer-based models. Experimental results demonstrate that NeRD outperforms traditional and CNN-based methods while significantly reducing the performance gap between INR-based and transformer-based demosaicking models. The ablation study highlights the importance of local feature encoding and skip connections in the implicit representation framework, showing that these architectural choices significantly enhance reconstruction quality.

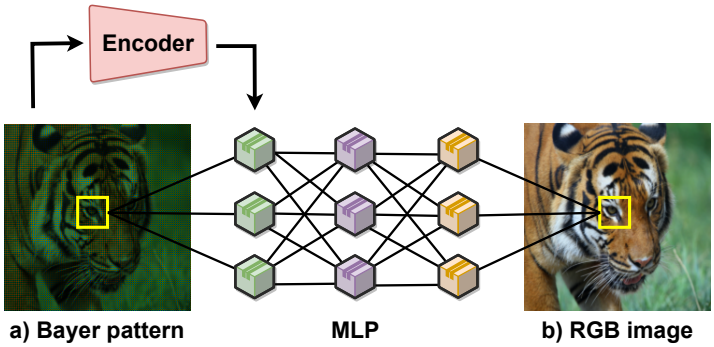


Figure 5.4: Simplified illustration of NeRD, the first demosaicking method using a coordinate-based implicit neural representation and a local encoding technique (ResNet + U-Net). For exact architecture details, refer to the reprinted paper below.

Contribution of the Paper

This paper introduces *NeRD*, the first neural field-based approach to demosaicking, leveraging implicit neural representations to reconstruct full-color images from Bayer patterns. By integrating a hybrid ResNet and U-Net encoder with a coordinate-based MLP, the method preserves spatial consistency and enhances color reconstruction quality. The resolution-agnostic nature of NeRD allows it to generalize across different image sizes, outperforming traditional and CNN-based demosaicking methods. The experimental evaluation confirms its effectiveness, demonstrating competitive performance against state-of-the-art transformer-based models.

Contribution of the Author

The main contribution of Tomáš Kerepecký in this paper involved extensive research on implicit neural representations, formulating the core idea, developing the methods described, and thorough testing of the proposed algorithms. Tomáš Kerepecký was responsible for preparing relevant data, conducting experiments, and drafting the initial version of the manuscript. The final paper was then refined collaboratively with other co-authors and presented at the international conference.

Tomáš Kerepecký, Filip Šroubek, and Jan Flusser, "Implicit neural representation for image demosaicking," *Digital Signal Processing*, p. 105022, 2025, Elsevier

---

RECONSTRUCTION of full-color images from raw sensor data remains a challenging problem, especially when the input data is degraded by noise, blur, or other real-world imperfections. Traditional demosaicking methods, both interpolation-based and deep learning-based, often struggle when confronted with out-of-distribution conditions (i.e., degradations not seen during training or not accounted for by the algorithm).

In response, this paper introduces *INRID* (Implicit Neural Representation for Image Demosaicking), a fully self-supervised framework that adapts to each individual image. By leveraging implicit neural representations, INRID is able to enhance the reconstruction quality of raw sensor data, achieving superior generalization compared to conventional CNN and transformer models trained solely on clean datasets.

The core idea of INRID is to encode each image as a coordinate-based multilayer perceptron that maps spatial coordinates to RGB values. Unlike prior INR-based demosaicking approaches, such as NeRD, which required supervised training on large datasets to condition INR, INRID operates in a self-supervised manner. Figure 5.5 illustrates the conceptual diagram of the framework. The approach integrates two key loss functions: a Bayer loss, which ensures fidelity to the raw sensor data by minimizing the reconstruction error in the observed CFA measurements, and a complementary loss, which regularizes the reconstruction using an initial estimate from traditional or state-of-the-art demosaicking methods. This hybrid formulation enables INRID to effectively utilize prior knowledge from classical algorithms while dynamically refining the reconstruction for each image.

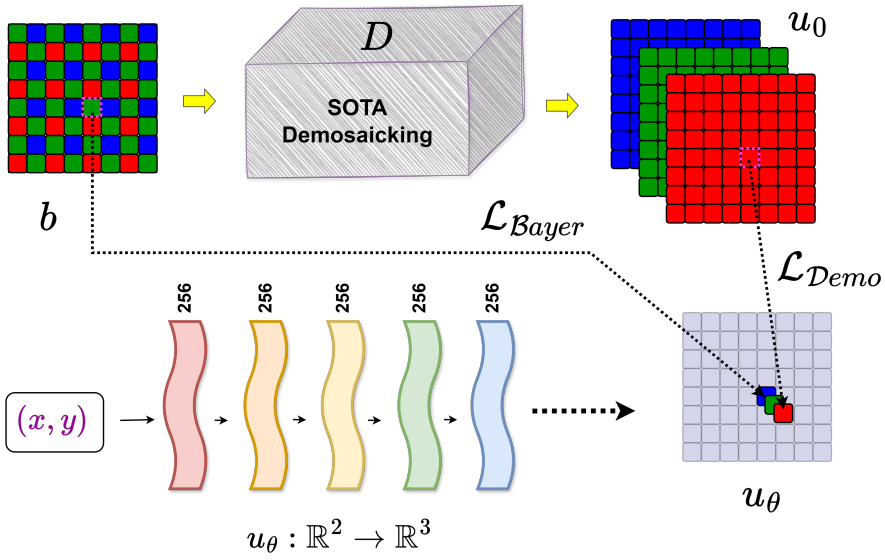


Figure 5.5: Conceptual diagram of INRID. The proposed approach performs demosaicking using an implicit neural representation $u_\theta : \mathbb{R}^2 \rightarrow \mathbb{R}^3$, optimized by minimizing the mean squared error $\mathcal{L}_{Bayer}$ between the reconstruction u_θ and the Bayer measurement b, as well as between the reconstruction u_θ and the initial demosaicked image u_0 ($\mathcal{L}_{Demo}$).

One of the major advantages of INRID is its ability to adapt to challenging degradations, such as blur and noise, without requiring retraining on new datasets. The method employs an optimization process that iteratively updates the INR parameters to fit each image, allowing it to recover high-frequency details lost in traditional demosaicking methods. Experimental evaluations on standard benchmarks and real-world raw sensor images demonstrate that INRID outperforms both traditional and deep learning-based demosaicking approaches, particularly in cases where the input data deviates from standard training distributions. The study further investigates the impact of various INR architectures, including sinusoidal representations and Fourier feature mappings, highlighting the benefits of different signal representations for inverse problems in imaging.

Contribution of the Paper

This paper introduces *INRID*, a fully self-supervised INR-based technique for image demosaicking that dynamically adapts to individual images, improving reconstruction quality for out-of-distribution inputs. By combining Bayer loss and complementary loss, the framework effectively integrates classical demosaicking priors with modern INR-based reconstruction, demonstrating significant improvements over both conventional and deep learning-based methods. INRID handles real-world degradations such as blur and noise, showcasing its robustness across diverse imaging conditions.

Contribution of the Author

The main contribution of Tomáš Kerepecký in this paper involved formulating the core idea, conducting extensive research on implicit neural representations, and developing the methods described. Tomáš Kerepecký was also responsible for preparing relevant data, carrying out experiments, and drafting the initial manuscript. The final version was then refined collaboratively with other co-authors, incorporating additional insights to produce the published work.

Contribution of the Thesis 6

Research presented in this thesis introduces novel methodologies that enhance existing image restoration techniques and propose innovative approaches for solving inverse problems in digital photography and fluorescence microscopy.

The main contributions of this thesis include:

- We introduced an iterative Wiener filtering framework (IWFT) for deblurring, incorporating restoration and update filters to effectively suppress ringing artifacts, while maintaining computational efficiency.
- We designed a deep unrolled neural network (D3Net) for joint image restoration, leveraging optimization-inspired updates to enable demosaicking, deblurring, and deringing in a unified learnable framework.
- We proposed a self-supervised learning approach (Dual-Cycle) for reconstructing dual-view light-sheet fluorescence microscopy images. By incorporating a cycle-consistency loss, this method eliminates the need for paired training data.
- We introduced the first implicit neural representation technique for demosaicking (NeRD), achieving reconstruction quality on par with state-of-the-art deep learning methods and showcasing the advantages of continuous function-based representations over pixel-based approaches.
- We introduced INRID, a self-adaptive INR framework for image demosaicking, designed to dynamically adapt to varying degradation types, ensuring robust generalization across distortions such as blur and noise without requiring dataset-specific retraining.

The work presented in this thesis bridges the gap between model-based restoration, deep learning, self-supervised and self-adaptive representations and offers new insights into the potential of implicit neural networks for solving inverse problems in imaging.

Bibliography

[1] Charles W. Groetsch, *Inverse problems in the mathematical sciences*, vol. 52, Springer, 1993.

[2] Aggelos K. Katsaggelos, *Digital image restoration*, Springer-Verlag, 1991.

[3] Andreas Kirsch, *An introduction to the mathematical theory of inverse problems*, vol. 120, Springer, 2011.

[4] Martin Hanke, *A taste of inverse problems: basic theory and examples*, SIAM, 2017.

[5] Mario Bertero, Patrizia Boccacci, and Christine De Mol, *Introduction to inverse problems in imaging*, CRC Press, 2021.

[6] Alice Lucas, Michael Iliadis, Rafael Molina, and Aggelos K. Katsaggelos, "Using deep neural networks for inverse problems in imaging: beyond analytical methods," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 20–36, 2018, IEEE.

[7] Gregory Ongie, Ajil Jalal, Christopher A. Metzler, Richard G. Baraniuk, Alexandros G. Dimakis, and Rebecca Willett, "Deep learning techniques for inverse problems in imaging," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 1, pp. 39–56, 2020, IEEE.

[8] Michael T. McCann and Michael Unser, "Biomedical image reconstruction: From the foundations to deep neural networks," *Foundations and Trends in Signal Processing*, vol. 13, no. 3, pp. 283–359, 2019, Now Publishers, Inc.

[9] Andrey Nikolayevich Tikhonov, "Solutions of ill-posed problems," *VH Winston and Sons*, 1977.

[10] Felix Heide, Markus Steinberger, Yun-Ta Tsai, Mushfiqur Rouf, Dawid Pająk, Dikpal Reddy, Orazio Gallo, Jing Liu, Wolfgang Heidrich, Karen Egiazarian, et al., "Flexisp: A flexible camera image processing framework," *ACM Transactions on Graphics (ToG)*, vol. 33, no. 6, pp. 1–13, 2014, ACM New York, NY, USA.

[11] Filip Šroubek, Tomáš Kerepecký, and Jan Kamenický, "Iterative Wiener filtering for deconvolution with ringing artifact suppression," in *2019 27th European Signal Processing Conference (EUSIPCO)*. September 2019, pp. 1–5, IEEE.

[12] Tomáš Kerepecký and Filip Šroubek, "D3net: Joint demosaicking, deblurring and deringing," in *2020 25th International Conference on Pattern Recognition (ICPR)*. January 2021, pp. 1–8, IEEE.

[13] Tomáš Kerepecký, Filip Šroubek, Adam Novozámský, and Jan Flusser, "Nerd: Neural field-based demosaicking," in *2023 IEEE International Conference on Image Processing (ICIP)*. October 2023, pp. 1735–1739, IEEE.

[14] Tomáš Kerepecký, Filip Šroubek, and Jan Flusser, "Implicit neural representation for image demosaicking," *Digital Signal Processing*, p. 105022, 2025, Elsevier.

[15] Ernst HK Stelzer, Frederic Strobl, Bo-Jui Chang, Friedrich Preusser, Stephan Preibisch, Katie McDole, and Reto Fiolka, "Light sheet fluorescence microscopy," *Nature Reviews Methods Primers*, vol. 1, no. 1, pp. 73, 2021, Nature Publishing Group, UK, London.

[16] Rory M. Power and Jan Huisken, "A guide to light-sheet fluorescence microscopy for multiscale imaging," *Nature Methods*, vol. 14, no. 4, pp. 360–373, 2017, Nature Publishing Group US New York.

[17] Tomáš Kerepecký, Jiaming Liu, Xuan Wei Ng, David W. Piston, and Ulugbek S. Kamilov, "Dual-cycle: Self-supervised dual-view fluorescence microscopy image reconstruction using cyclegan," in *2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. June 2023, pp. 1–5, IEEE.

[18] Daniele Menon and Giancarlo Calvagno, "Color image demosaicking: An overview," *Signal Processing: Image Communication*, vol. 26, no. 8-9, pp. 518–533, 2011, Elsevier.

[19] Michaël Gharbi, Gaurav Chaurasia, Sylvain Paris, and Frédo Durand, "Deep joint demosaicking and denoising," *ACM Transactions on Graphics (ToG)*, vol. 35, no. 6, pp. 1–12, 2016, ACM New York, NY, USA.

[20] Filippos Kokkinos and Stamatios Lefkimmiatis, "Deep image demosaicking using a cascade of convolutional residual denoising networks," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 303–319.

[21] William H. Richardson, "Bayesian-based iterative method of image restoration," *Journal of the Optical Society of America*, vol. 62, no. 1, pp. 55–59, 1972.

[22] Deepa Kundur and Dimitrios Hatzinakos, "Blind image deconvolution," *IEEE signal processing magazine*, vol. 13, no. 3, pp. 43–64, 2002, IEEE.

[23] Kaihao Zhang, Wenqi Ren, Wenhan Luo, Wei-Sheng Lai, Björn Stenger, Ming-Hsuan Yang, and Hongdong Li, "Deep image deblurring: A survey," *International Journal of Computer Vision (IJCV)*, vol. 130, no. 9, pp. 2103–2130, 2022, Springer.

[24] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on image processing (TIP)*, vol. 16, no. 8, pp. 2080–2095, 2007, IEEE.

[25] Chunwei Tian, Lunke Fei, Wenxian Zheng, Yong Xu, Wangmeng Zuo, and Chia-Wen Lin, "Deep learning on image denoising: An overview," *Neural Networks*, vol. 131, pp. 251–275, 2020, Elsevier.

[26] Michael Elad, Bahjat Kawar, and Gregory Vaksman, "Image denoising: The deep learning revolution and beyond—a survey paper," *SIAM Journal on Imaging Sciences*, vol. 16, no. 3, pp. 1594–1654, 2023, SIAM.

[27] Yicong Wu, Peter Wawrzusin, Justin Senseney, Robert S. Fischer, Ryan Christensen, Anthony Santella, Andrew G. York, Peter W. Winter, Clare M. Waterman, Zhirong Bao, et al., "Spatially isotropic four-dimensional imaging with dual-view plane illumination microscopy," *Nature biotechnology*, vol. 31, no. 11, pp. 1032–1038, 2013, Nature Publishing Group UK London.

[28] Hao Zhang, Han Xu, Xin Tian, Junjun Jiang, and Jiayi Ma, "Image fusion meets deep learning: A survey and perspective," *Information Fusion*, vol. 76, pp. 323–336, 2021, Elsevier.

[29] Harpreet Kaur, Deepika Koundal, and Virender Kadyan, "Image fusion techniques: a survey," *Archives of computational methods in Engineering*, vol. 28, no. 7, pp. 4425–4447, 2021, Springer.

[30] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang, "Super-resolution image reconstruction: a technical overview," *IEEE signal processing magazine*, vol. 20, no. 3, pp. 21–36, 2003, IEEE.

[31] Kamal Nasrollahi and Thomas B. Moeslund, "Super-resolution: a comprehensive survey," *Machine vision and applications*, vol. 25, pp. 1423–1468, 2014, Springer.

[32] Rafael C. Gonzalez, *Digital image processing*, Pearson education india, 2009.

[33] Lu Yuan, Jian Sun, Long Quan, and Heung-Yeung Shum, "Image deblurring with blurred/noisy image pairs," *ACM transactions on graphics (TOG)*, vol. 26, no. 3, 2007.

[34] Leonid I. Rudin, Stanley Osher, and Emad Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992, Elsevier.

[35] David L. Donoho, "De-noising by soft-thresholding," *IEEE transactions on information theory*, vol. 41, no. 3, pp. 613–627, 2002, IEEE.

[36] Guy Gilboa and Stanley Osher, "Nonlocal operators with applications to image processing," *Multiscale Modeling & Simulation*, vol. 7, no. 3, pp. 1005–1028, 2009, SIAM.

[37] Trevor Hastie, Robert Tibshirani, and Jerome H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, vol. 2, Springer, 2009.

[38] Curtis R Vogel, *Computational methods for inverse problems*, SIAM, 2002.

[39] Gene H. Golub and Charles F. Van Loan, *Matrix computations*, JHU press, 2013.

[40] Ingrid Daubechies, Michel Defrise, and Christine De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 57, no. 11, pp. 1413–1457, 2004, Wiley Online Library.

[41] Amir Beck and Marc Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009, SIAM.

[42] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan 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, Now Publishers, Inc.

[43] Antonin Chambolle and Thomas Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of mathematical imaging and vision*, vol. 40, pp. 120–145, 2011, Springer.

[44] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Learning a deep convolutional network for image super-resolution," in *13th European Conference on Computer Vision (ECCV), pages=184–199, year=2014, organization=Springer.*

[45] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning*, MIT Press, 2016.

[46] Michael T. McCann, Kyong Hwan Jin, and Michael Unser, "Convolutional neural networks for inverse problems in imaging: A review," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 85–95, 2017, IEEE.

[47] Antonio Torralba, Phillip Isola, and William T. Freeman, *Foundations of computer vision*, MIT Press, 2024.

[48] Jingwen Su, Boyan Xu, and Hujun Yin, "A survey of deep learning approaches to image restoration," *Neurocomputing*, vol. 487, pp. 46–65, 2022, Elsevier.

[49] Yann LeCun, Bernhard Boser, John S. Denker, Donnie Henderson, Richard E. Howard, Wayne Hubbard, and Lawrence D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989, MIT Press.

[50] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2016, pp. 1646–1654.

[51] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2016, pp. 770–778.

[52] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE transactions on image processing (TIP)*, vol. 26, no. 7, pp. 3142–3155, 2017, IEEE.

[53] Simon J.D. Prince, *Understanding Deep Learning*, The MIT Press, 2023.

[54] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations (ICLR)*, 2021.

[55] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin, "Attention is all you need," *Advances in neural information processing systems (NeurIPS)*, vol. 30, 2017.

[56] Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao, "Pre-trained image processing transformer," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, 2021, pp. 12299–12310.

[57] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte, "Swinir: Image restoration using swin transformer," in *Proceedings of the IEEE/CVF international conference on computer vision (CVPR)*, 2021, pp. 1833–1844.

[58] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing (TIP)*, vol. 13, no. 4, pp. 600–612, 2004, IEEE.

[59] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2018, pp. 586–595.

[60] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2014, vol. 27, pp. 2672–2680.

[61] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2017, pp. 4681–4690.

[62] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas, "Deblurgan: Blind motion deblurring using conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2018, pp. 8183–8192.

[63] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision (ICCV)*, 2017, pp. 2223–2232.

[64] Xiao Liang, Liyuan Chen, Dan Nguyen, Zhiguo Zhou, Xuejun Gu, Ming Yang, Jing Wang, and Steve Jiang, "Generating synthesized computed tomography (ct) from cone-beam computed tomography (cbct) using cyclegan for adaptive radiation therapy," *Physics in Medicine & Biology*, vol. 64, no. 12, pp. 125002, 2019, IOP Publishing.

[65] Hyoungjun Park, Myeongsu Na, Bumju Kim, Soohyun Park, Ki Hean Kim, Sunghoe Chang, and Jong Chul Ye, "Deep learning enables reference-free isotropic super-resolution for volumetric fluorescence microscopy," *Nature Communications*, vol. 13, no. 1, pp. 1–12, 2022, Nature Publishing Group.

[66] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *18th international conference of Medical image computing and computer-assisted intervention (MICCAI)*. Springer, 2015, pp. 234–241.

[67] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE*

conference on computer vision and pattern recognition (CVPR), 2017, pp. 1125–1134.

[68] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky, "Deep image prior," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2018, pp. 9446–9454.

[69] Dongwei Ren, Kai Zhang, Qilong Wang, Qinghua Hu, and Wangmeng Zuo, "Neural blind deconvolution using deep priors," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, 2020, pp. 3341–3350.

[70] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," in *European Conference on Computer Vision (ECCV)*, 2020, pp. 405–421.

[71] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila, "Noise2Noise: Learning image restoration without clean data," in *Proceedings of the 35th International Conference on Machine Learning (ICML)*, Jennifer Dy and Andreas Krause, Eds., 10–15 Jul 2018, vol. 80 of *Proceedings of Machine Learning Research*, pp. 2965–2974.

[72] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug, "Noise2void-learning denoising from single noisy images," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, 2019, pp. 2129–2137.

[73] Joshua Batson and Loic Royer, "Noise2self: Blind denoising by self-supervision," in *Proceedings of the 36th International Conference on Machine Learning (ICML)*, 2019, pp. 524–533.

[74] Ulugbek S. Kamilov, Charles A. Bouman, Gregery T. Buzzard, and Brendt Wohlberg, "Plug-and-play methods for integrating physical and learned models in computational imaging: Theory, algorithms, and applications," *IEEE Signal Processing Magazine*, vol. 40, no. 1, pp. 85–97, 2023, IEEE.

[75] Nir Shlezinger, Jay Whang, Yonina C. Eldar, and Alexandros G. Dimakis, "Model-based deep learning: Key approaches and design guidelines," in *2021 IEEE Data Science and Learning Workshop (DSLW)*. 2021, pp. 1–6, IEEE.

[76] Hemant K. Aggarwal, Merry P. Mani, and Mathews Jacob, "Modl: Model-based deep learning architecture for inverse problems," *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 394–405, 2018, IEEE.

[77] Singanallur V. Venkatakrishnan, Charles A. Bouman, and Brendt Wohlberg, "Plug-and-play priors for model based reconstruction," in *2013 IEEE global conference on signal and information processing*. IEEE, 2013, pp. 945–948.

[78] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte, "Plug-and-play image restoration with deep denoiser prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 44, no. 10, pp. 6360–6376, 2021, IEEE.

[79] Yaniv Romano, Michael Elad, and Peyman Milanfar, "The little engine that could: Regularization by denoising (red)," *SIAM Journal on Imaging Sciences*, vol. 10, no. 4, pp. 1804–1844, 2017, SIAM.

[80] Jiaming Liu, Yu Sun, Cihat Eldeniz, Weijie Gan, Hongyu An, and Ulugbek S. Kamilov, "Rare: Image reconstruction using deep priors learned without groundtruth," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 6, pp. 1088–1099, 2020, IEEE.

[81] Yan Yang, Jian Sun, Huibin Li, and Zongben Xu, "Deep admm-net for compressive sensing mri," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2016, vol. 29.

[82] Vishal Monga, Yuelong Li, and Yonina C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 18–44, 2021, IEEE.

[83] Aayush Karan, Kulin Shah, Sitan Chen, and Yonina Eldar, "Unrolled denoising networks provably learn to perform optimal bayesian inference," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 37, pp. 135264–135298, 2024.

[84] Davis Gilton, Gregory Ongie, and Rebecca Willett, "Deep equilibrium architectures for inverse problems in imaging," *IEEE Transactions on Computational Imaging (TCI)*, vol. 7, pp. 1123–1133, 2021, IEEE.

[85] Albert Tarantola, *Inverse problem theory and methods for model parameter estimation*, SIAM, 2005.

[86] Jonathan Ho, Ajay Jain, and Pieter Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems (NeurIPS)*, vol. 33, pp. 6840–6851, 2020.

[87] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *Proceedings of the 32th International conference on machine learning (ICML)*, 2015, pp. 2256–2265.

[88] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole, "Score-based generative modeling through stochastic differential equations," in *International Conference on Learning Representations (ICLR)*, 2020.

[89] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J. Fleet, and Mohammad Norouzi, "Image super-resolution via iterative refinement," *IEEE transactions on pattern analysis and machine intelligence (TPAMI)*, vol. 45, no. 4, pp. 4713–4726, 2022, IEEE.

[90] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song, "Denoising diffusion restoration models," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 35, pp. 23593–23606, 2022.

[91] Berthy T. Feng, Jamie Smith, Michael Rubinstein, Huiwen Chang, Katherine L Bouman, and William T. Freeman, "Score-based diffusion models as principled

priors for inverse imaging," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 10520–10531.

[92] Giannis Daras, Hyungjin Chung, Chieh-Hsin Lai, Yuki Mitsufuji, Jong Chul Ye, Peyman Milanfar, Alexandros G Dimakis, and Mauricio Delbracio, "A survey on diffusion models for inverse problems," *arXiv preprint arXiv:2410.00083*, 2024.

[93] Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool, "Denoising diffusion models for plug-and-play image restoration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 1219–1229.

[94] Alexandros Graikos, Nikolay Malkin, Nebojsa Jojic, and Dimitris Samaras, "Diffusion models as plug-and-play priors," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 35, pp. 14715–14728, 2022.

[95] Hengkang Wang, Xu Zhang, Taihui Li, Yuxiang Wan, Tiancong Chen, and Ju Sun, "Dmplug: A plug-in method for solving inverse problems with diffusion models," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 37, pp. 117881–117916, 2024.

[96] Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon, "Ilvr: Conditioning method for denoising diffusion probabilistic models," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, 2021, pp. 14347–14356.

[97] Hyungjin Chung, Jeongsol Kim, Michael T. McCann, Marc L. Klasky, and Jong Chul Ye, "Diffusion posterior sampling for general noisy inverse problems," in *11th International Conference on Learning Representations (ICLR)*, 2023.

[98] Hyungjin Chung and Jong Chul Ye, "Score-based diffusion models for accelerated mri," *Medical image analysis*, vol. 80, pp. 102479, 2022, Elsevier.

[99] Hyungjin Chung, Jeongsol Kim, Sehui Kim, and Jong Chul Ye, "Parallel diffusion models of operator and image for blind inverse problems," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 6059–6069.

[100] Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever, "Consistency models," in *Proceedings of the 40th International Conference on Machine Learning (ICML)*, 2023, pp. 32211–32252.

[101] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, 2022, pp. 10684–10695.

[102] Kushagra Pandey and Stephan Mandt, "A complete recipe for diffusion generative models," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 4261–4272.

[103] Litu Rout, Yujia Chen, Abhishek Kumar, Constantine Caramanis, Sanjay Shakkottai, and Wen-Sheng Chu, "Beyond first-order tweedie: Solving inverse problems using latent diffusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 9472–9481.

[104] Nebiyou Yismaw, Ulugbek S. Kamilov, and M. Salman Asif, "Gaussian is all you need: A unified framework for solving inverse problems via diffusion posterior sampling," *arXiv preprint arXiv:2409.08906*, 2024.

[105] Ron Mokady, Amir Hertz, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or, "Null-text inversion for editing real images using guided diffusion models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, 2023, pp. 6038–6047.

[106] Hyungjin Chung, Jong Chul Ye, Peyman Milanfar, and Mauricio Delbracio, "Prompt-tuning latent diffusion models for inverse problems," in *Proceedings of the 41st International Conference on Machine Learning (ICML)*, 2024, pp. 8941–8967.

[107] Jeongsol Kim, Geon Yeong Park, Hyungjin Chung, and Jong Chul Ye, "Regularization by texts for latent diffusion inverse solvers," in *The 13th International Conference on Learning Representations (ICLR)*, 2025.

[108] Chenyang Qi, Zhengzhong Tu, Keren Ye, Mauricio Delbracio, Peyman Milanfar, Qifeng Chen, and Hossein Talebi, "Spire: Semantic prompt-driven image restoration," in *European Conference on Computer Vision (ECCV)*. Springer, 2024, pp. 446–464.

[109] Florinel-Alin Croitoru, Vlad Hondru, Radu Tudor Ionescu, and Mubarak Shah, "Diffusion models in vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 45, no. 9, pp. 10850–10869, 2023, IEEE.

[110] Xin Li, Yulin Ren, Xin Jin, Cuiling Lan, Xingrui Wang, Wenjun Zeng, Xinchao Wang, and Zhibo Chen, "Diffusion models for image restoration and enhancement–a comprehensive survey," *arXiv preprint arXiv:2308.09388*, 2023.

[111] Brian B. Moser, Arundhati S. Shanbhag, Federico Raue, Stanislav Frolov, Sebastian Palacio, and Andreas Dengel, "Diffusion models, image super-resolution, and everything: A survey," *IEEE Transactions on Neural Networks and Learning Systems*, 2024, IEEE.

[112] Hui Shen, Jingxuan Zhang, Boning Xiong, Rui Hu, Shoufa Chen, Zhongwei Wan, Xin Wang, Yu Zhang, Zixuan Gong, Guangyin Bao, et al., "Efficient diffusion models: A survey," *arXiv preprint arXiv:2502.06805*, 2025.

[113] Yiheng Xie, Towaki Takikawa, Shunsuke Saito, Or Litany, Shiqin Yan, Numair Khan, Federico Tombari, James Tompkin, Vincent Sitzmann, and Srinath Sridhar, "Neural fields in visual computing and beyond," *Computer Graphics Forum*, 2022, The Eurographics Association and John Wiley Sons Ltd.

[114] Amirali Molaei, Amirhossein Aminimehr, Armin Tavakoli, Amirhossein Kazerouni, Bobby Azad, Reza Azad, and Dorit Merhof, "Implicit neural representation in medical imaging: A comparative survey," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 2381–2391.

[115] Amer Essakine, Yanqi Cheng, Chun-Wun Cheng, Lipei Zhang, Zhongying Deng, Lei Zhu, Carola-Bibiane Schönlieb, and Angelica I Aviles-Rivero, "Where do we stand with implicit neural representations? a technical and performance survey," *arXiv preprint arXiv:2411.03688*, 2024.

[116] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein, "Implicit neural representations with periodic activation functions," *Advances in neural information processing systems (NeurIPS)*, vol. 33, pp. 7462–7473, 2020.

[117] Vishwanath Saragadam, Daniel LeJeune, Jasper Tan, Guha Balakrishnan, Ashok Veeraraghavan, and Richard G Baraniuk, "Wire: Wavelet implicit neural representations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 18507–18516.

[118] Sicheng Gao, Xuhui Liu, Bohan Zeng, Sheng Xu, Yanjing Li, Xiaoyan Luo, Jianzhuang Liu, Xiantong Zhen, and Baochang Zhang, "Implicit diffusion models for continuous super-resolution," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, 2023, pp. 10021–10030.

[119] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove, "Deepsdf: Learning continuous signed distance functions for shape representation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 165–174.

[120] Yinbo Chen, Sifei Liu, and Xiaolong Wang, "Learning continuous image representation with local implicit image function," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 8628–8638.

[121] Hao Chen, Bo He, Hanyu Wang, Yixuan Ren, Ser Nam Lim, and Abhinav Shrivastava, "Nerv: Neural representations for videos," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 34, pp. 21557–21568, 2021.

[122] Yannick Strümpler, Janis Postels, Ren Yang, Luc Van Gool, and Federico Tombari, "Implicit neural representations for image compression," in *European Conference on Computer Vision (ECCV)*. Springer, 2022, pp. 74–91.

[123] Emilien Dupont, Hrushikesh Loya, Milad Alizadeh, Adam Golinski, Yee Whye Teh, and Arnaud Doucet, "Coin++: neural compression across modalities," *Transactions on Machine Learning Research*, vol. 2022, no. 11, 2022.

[124] Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville, "On the spectral bias of neural networks," in *Proceedings of the 36th International Conference on Machine Learning (ICML)*, 2019, pp. 5301–5310.

[125] Gizem Yüce, Guillermo Ortiz-Jiménez, Beril Besbinar, and Pascal Frossard, "A structured dictionary perspective on implicit neural representations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 19228–19238.

[126] Shaowen Xie, Hao Zhu, Zhen Liu, Qi Zhang, You Zhou, Xun Cao, and Zhan Ma, "Diner: Disorder-invariant implicit neural representation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 6143–6152.

[127] Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng, "Fourier features let networks learn high frequency functions in low dimensional domains," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 7537–7547, 2020.

[128] Sameera Ramasinghe and Simon Lucey, "Beyond periodicity: Towards a unifying framework for activations in coordinate-mlps," in *European Conference on Computer Vision (ECCV)*. Springer, 2022, pp. 142–158.

[129] Danzel Serrano, Jakub Szymkowiak, and Przemyslaw Musialski, "Hosc: A periodic activation function for preserving sharp features in implicit neural representations," *arXiv preprint arXiv:2401.10967*, 2024.

[130] Hemanth Saratchandran, Sameera Ramasinghe, Violetta Shevchenko, Alexander Long, and Simon Lucey, "A sampling theory perspective on activations for implicit neural representations," in *Proceedings of the 41st International Conference on Machine Learning (ICML)*, 2024, pp. 43422–43444.

[131] Yang Chen, Ruituo Wu, Yipeng Liu, and Ce Zhu, "Hoin: High-order implicit neural representations," *arXiv preprint arXiv:2404.14674*, 2024.

[132] Zhen Liu, Hao Zhu, Qi Zhang, Jingde Fu, Weibing Deng, Zhan Ma, Yanwen Guo, and Xun Cao, "Finer: Flexible spectral-bias tuning in implicit neural representation by variable-periodic activation functions," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 2713–2722.

[133] Ishit Mehta, Michaël Gharbi, Connelly Barnes, Eli Shechtman, Ravi Ramamoorthi, and Manmohan Chandraker, "Modulated periodic activations for generalizable local functional representations," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 14214–14223.

[134] Emilien Dupont, Hyunjik Kim, SM Ali Eslami, Danilo Jimenez Rezende, and Dan Rosenbaum, "From data to functa: Your data point is a function and you can treat it like one," in *International Conference on Machine Learning (ICML)*, 2022, pp. 5694–5725.

[135] Amirhossein Kazerouni, Reza Azad, Alireza Hosseini, Dorit Merhof, and Ulas Bagci, "Incode: Implicit neural conditioning with prior knowledge embeddings," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2024, pp. 1298–1307.

[136] Ivan Skorokhodov, Savva Ignatyev, and Mohamed Elhoseiny, "Adversarial generation of continuous images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 10753–10764.

[137] Julien Martel, David B. Lindell, Connor Z. Lin, Eric R. Chan, Marco Monteiro, and Gordon Wetzstein, "Acorn: adaptive coordinate networks for neural scene representation," *ACM Transactions on Graphics (TOG)*, vol. 40, no. 4, pp. 1–13, 2021, ACM New York, NY, USA.

[138] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *ACM Transactions on Graphics (TOG)*, vol. 41, no. 4, pp. 1–15, 2022, ACM New York, NY, USA.

[139] Matthew Tancik, Ben Mildenhall, Terrance Wang, Divi Schmidt, Pratul P. Srinivasan, Jonathan T. Barron, and Ren Ng, "Learned initializations for optimizing coordinate-based neural representations," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, 2021, pp. 2846–2855.

[140] Christian Reiser, Songyou Peng, Yiyi Liao, and Andreas Geiger, "Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (CVPR)*, 2021, pp. 14335–14345.

[141] Vishwanath Saragadam, Jasper Tan, Guha Balakrishnan, Richard G. Baraniuk, and Ashok Veeraraghavan, "Miner: Multiscale implicit neural representation," in *European Conference on Computer Vision (ECCV)*. Springer, 2022, pp. 318–333.

[142] Peihao Wang, Zhiwen Fan, Tianlong Chen, and Zhangyang Wang, "Neural implicit dictionary learning via mixture-of-expert training," in *Proceedings of the 39th International Conference on Machine Learning (ICML)*, 2022, pp. 22613–22624.

[143] Yu Sun, Jiaming Liu, Mingyang Xie, Brendt Egon Wohlberg, and Ulugbek S. Kamilov, "Coil: Coordinate-based internal learning for imaging inverse problems," *IEEE Transactions on Computational Imaging (TCI)*, vol. 7, no. LA-UR-21-21526, 2021.

[144] Liyue Shen, John Pauly, and Lei Xing, "Nerp: implicit neural representation learning with prior embedding for sparsely sampled image reconstruction," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 1, pp. 770–782, 2022, IEEE.

[145] Wentao Shangguan, Yu Sun, Weijie Gan, and Ulugbek S. Kamilov, "Learning cross-video neural representations for high-quality frame interpolation," in *European Conference on Computer Vision (ECCV)*. Springer, 2022, pp. 511–528.

[146] Zeyuan Chen, Yinbo Chen, Jingwen Liu, Xingqian Xu, Vidit Goel, Zhangyang Wang, Humphrey Shi, and Xiaolong Wang, "Videoinr: Learning video implicit neural representation for continuous space-time super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 2047–2057.

[147] Li Ma, Xiaoyu Li, Jing Liao, Qi Zhang, Xuan Wang, Jue Wang, and Pedro V. Sander, "Deblur-nerf: Neural radiance fields from blurry images," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, 2022, pp. 12861–12870.

[148] Yuhui Quan, Xin Yao, and Hui Ji, "Single image defocus deblurring via implicit neural inverse kernels," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 12600–12610.

[149] Dejia Xu, Peihao Wang, Yifan Jiang, Zhiwen Fan, and Zhangyang Wang, "Signal processing for implicit neural representations," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 35, pp. 13404–13418, 2022.

List of Author's Publications

1. Tomáš Kerepecký, Filip Šroubek, and Jan Flusser, "Implicit neural representation for image demosaicking," *Digital Signal Processing*, p. 105022, 2025, Elsevier

2. Tomáš Kerepecký, Filip Šroubek, Barbara Zitová, and Jan Flusser, "Automated actor recognition in video content," in *Data Science in Applications: Towards AI Driven Approaches*, G. Dzemyda, J. Bernatavičienė, and J. Kacprzyk, Eds., vol. 1206. Springer Nature, June 2025. (accepted)

3. Tomáš Kerepecký, Filip Šroubek, and Jan Flusser, "Inverse problems in image restoration," in *Inverse Problems: Modeling and Simulation*, A. Hasanoğlu, R. Novikov, and K. Bockstal, Eds., vol. 11. Springer Nature, May 2025. (accepted)

4. Tomáš Kerepecký, Filip Šroubek, Barbara Zitová, and Jan Flusser, "Star: Screen time and actor recognition in video content," in *2024 46th DAGM German Conference on Pattern Recognition (GCPR)*. vol. 15298. Springer Nature, May 2025, (accepted)

5. Tomáš Karella, Tomáš Suk, Václav Košík, Leonid Bedratyuk, Tomáš Kerepecký, and Jan Flusser, "3d non-separable moment invariants and their use in neural networks," *SN Computer Science*, vol. 5, no. 8, pp. 1–16, 2024, Springer.

6. Tomáš Kerepecký, Filip Šroubek, Adam Novozámský, and Jan Flusser, "Nerd: Neural field-based demosaicking," in *2023 IEEE International Conference on Image Processing (ICIP)*. October 2023, pp. 1735–1739, IEEE

7. Roman Staněk, Tomáš Kerepecký, Adam Novozámský, Filip Šroubek, Barbara Zitová, and Jan Flusser, "Real-time wheel detection and rim classification in automotive production," in *2023 IEEE International Conference on Image Processing (ICIP)*. October 2023, pp. 1410–1414, IEEE.

8. Tomáš Kerepecký, Jiaming Liu, Xuan Wei Ng, David W. Piston, and Ulugbek S. Kamilov, "Dual-cycle: Self-supervised dual-view fluorescence microscopy image reconstruction using cyclegan," in *2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. June 2023, pp. 1–5, IEEE

9. Tomáš Kerepecký and Filip Šroubek, "D3net: Joint demosaicking, deblurring and deringing," in *2020 25th International Conference on Pattern Recognition (ICPR)*. January 2021, pp. 1–8, IEEE

10. Filip Šroubek, Tomáš Kerepecký, and Jan Kamenický, "Iterative Wiener filtering for deconvolution with ringing artifact suppression," in *2019 27th European Signal Processing Conference (EUSIPCO)*. September 2019, pp. 1–5, IEEE

Part II

PAPERS

Iterative Wiener Filtering for Deconvolution with Ringing Artifact Suppression

Filip Šroubek, Tomáš Kerepecký and Jan Kamenický
Institute of Information Theory and Automation
Czech Academy of Sciences
Prague, Czech Republic
sroubekf@utia.cas.cz

***Abstract*—Sensor and lens blur degrade images acquired by digital cameras. Simple and fast removal of blur using linear filtering, such as Wiener filter, produces results that are not acceptable in most of the cases due to ringing artifacts close to image borders and around edges in the image. More elaborate deconvolution methods with non-smooth regularization, such as total variation, provide superior performance with less artifacts, however at a price of increased computational cost. We consider the alternating directions method of multipliers, which is a popular choice to solve such non-smooth convex problems, and show that individual steps of the method can be decomposed to simple filtering and element-wise operations. Filtering is performed with two sets of filters, called restoration and update filters, which are learned for the given type of blur and noise level with two different learning methods. The proposed deconvolution algorithm is implemented in the spatial domain and can be easily extended to include other restoration tasks such as demosaicing and super-resolution. Experiments demonstrate performance of the algorithm with respect to the size of learned filters, number of iterations, noise level and type of blur.**

***Index Terms*—Wiener filter, LMMSE, deconvolution, total variation, ADMM, non-smooth optimization**

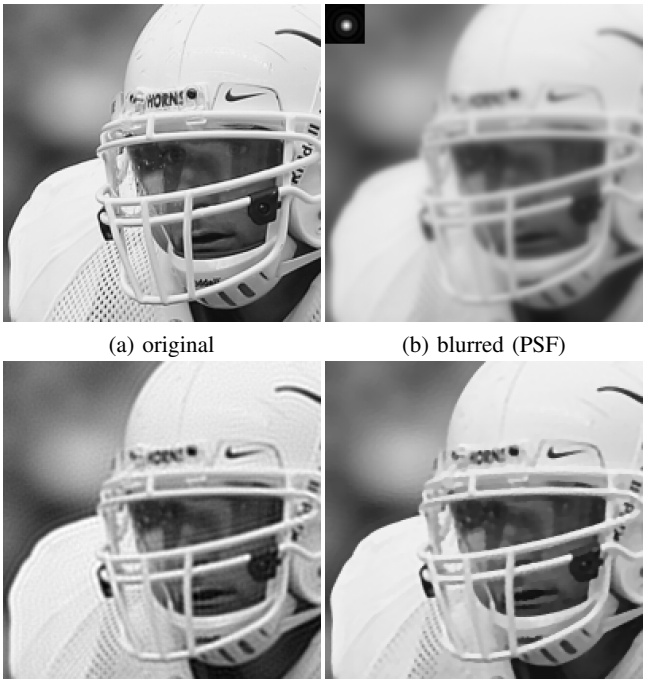
I. Introduction

Digital cameras are present in various measuring systems including microscopes, telescopes, and also small embedded systems like smartphones. Data acquired by camera sensors are subject to various types of signal degradation, for example lens and sensor blur, aberrations, color filter array (CFA) and noise. To obtain true images of the measured scene, it is necessary to correctly process the acquired data. The processing step is designed to run in the camera with limited computational capacity that allows only pixel-wise operations and some basic filtering.

Blur degradation often remains unattended in cameras as the cost to remove it is very high. Blur is modeled by convolution and even if the convolution kernel – called point spread function (PSF) – is known, the inverse problem of deconvolution is ill-posed due to values close to zero in spectra of common PSFs. For this reason, deconvolution methods based solely on linear operators, such as filtering, produce poor results.

From all linear filters, the optimal is the well-known Wiener filter, which is popular for having an explicit form in the frequency (Fourier) domain (FD) and estimates a sharp image in one step. However, since it is a linear operator, the estimated image exhibits ringing artifacts around edges; see an example of Wiener output in Fig. 1(c) obtained by filtering the blurred image in Fig. 1(b). Another disadvantage is that implementation in the FD implicitly assumes circular convolution, which in real scenarios is violated and the so-called problem of boundary conditions results in disturbing artifacts close to image borders. Proposed remedies either solve the boundary pixels separately in the spatial (image) domain (SD) [1], [2], or modify the boundary pixels in the blurred image to better comply with circular convolution [3], [4].

Equivalently, the problem of boundary conditions can be



(a) original (b) blurred (PSF)

(c) Wiener (d) proposed iterative Wiener

Fig. 1. Deconvolution without ringing artifacts: (a) the original image, (b) the blurred input image and PSF (inset), (c) restoration using the Wiener filter with power spectrum of the original image, (d) restoration using the proposed iterative Wiener filtering after 10 iterations. Notice ringing artifacts in the Wiener solution (c) and their absence in the proposed method (d).

This work was supported by Czech Science Foundation grant GA18-05360S and by the Praemium Academiae awarded by the Czech Academy of Sciences.

solved if only a 'valid' part of convolution is considered. Deconvolution formulated as a least squares optimization problem with Tikhonov quadratic regularization has a closed-form linear solution of type $Ax = b$; review classical restoration methods in [5]. However, when the 'valid' part of convolution is used, the inversion of A is typically not feasible and iterative numerical methods, such as Conjugate Gradient, must be used.

To achieve deconvolution results without artifacts, we have to leave the space of linear operators and allow non-linear ones. This is done by introducing non-smooth regularization terms, such as total variation [6], in the optimization problem; see for example [7]. Solving non-smooth convex problems requires specialized techniques, of which saddle-point methods [8] are probably the most popular. Generally, the solution is not in a closed form anymore and instead an iterative procedure is applied, which consists of multiple update equations and some of them are non-linear.

In this work, we propose to solve the deconvolution problem by combining the computational efficiency of Wiener filtering and the superior restoration quality of non-smooth optimization methods. We show that in the saddle-point methods linear update equations can be interpreted as Wiener-like filters and the non-linear update equations as soft thresholding. All steps are implemented in the SD using only filtering and element-wise operations, which naturally solves the problem of boundary artifacts. The filters are learned by solving a separate optimization problem on training data, which are specifically generated for the learning procedure. We foresee that the proposed algorithm easily extends to space-variant deconvolution, demosaicing or super-resolution.

The paper is organized as follows. In Section II, learning filters in the SD and the proposed algorithm is introduced. Section III experimentally validates the algorithm performance with respect to various conditions and Section IV concludes the paper with a short discussion of possible extensions of the algorithm.

II. Methodology

The discrete formation model considered in this work is a standard convolution process

$$g = Hu + n\,, \tag{1}$$

where g is the blurred and noisy image, u is the unknown sharp image, $H(\cdot) \equiv h * \cdot$ denotes a degradation operator (matrix) performing convolution with some known PSF h, and $n \approx \mathcal{N}(0, \sigma^2)$ is additive white Gaussian noise (AWGN) with zero mean and variance $\mathrm{Var}(n) = \sigma^2$. We consider scalar-valued digital images represented as column vectors $u \in \mathbb{R}^m$ and $g \in \mathbb{R}^p$. Pixels are indexed as $(u)_i$. In practice, H models 'valid' convolution and thus $m \geq p$. We define a discrete gradient operator $\mathsf{D} : \mathbb{R}^m \to \mathbb{R}^{m\times 2}$, which in its simplest form returns horizontal and vertical differences of pixels. It is a multidimensional array (tensor) consisting of two matrix components $[D_x, D_y]$ that perform convolution with $[1, -1]$ and $[1; -1]$ filters. The operator D can be more general and have more components, e.g. diagonal differences for isotropic behavior, or differences of pixels in a larger neighborhood to better capture correlation of pixels. On the vector-valued output of D, we define following norms:

$$\|\mathsf{D}u\|_{2,1}\colon \mathbb{R}^{m\times 2} \to \mathbb{R} \equiv \sum_i \left((\mathsf{D}u)_{i,1}^2 + (\mathsf{D}u)_{i,2}^2\right)^{1/2}\,,$$

$$\|\mathsf{D}u\|_{2,1}^2\colon \mathbb{R}^{m\times 2} \to \mathbb{R} \equiv \sum_i \left((\mathsf{D}u)_{i,1}^2 + (\mathsf{D}u)_{i,2}^2\right)\,.$$

If 'valid' convolution is replaced with circular convolution then (1) rewrites in the FD as

$$\mathcal{G} = \mathcal{H}\mathcal{U} + \mathcal{N}\,, \tag{2}$$

where capital calligraphic letters denote the Fourier transform $F(\cdot)$ of the corresponding function, e.g. $\mathcal{G} = Fg$, $\mathcal{H} = Fh$.

First let us formulate deconvolution as an optimization problem with Tikhonov regularization

$$\hat{u} = \arg\min_u \frac{\gamma}{2}\|Hu - g\|_2^2 + \|\mathsf{D}u\|_{2,1}^2\,, \tag{3}$$

where the norm of the first term is the classical ℓ_2-norm. A closed-form solution exists in this case and if circular convolution is assumed, the result in the FD has an explicit form of linear filtering

$$\hat{\mathcal{U}} = \hat{\mathcal{W}}\mathcal{G} = \frac{\mathcal{H}^*}{|\mathcal{H}|^2 + \frac{1}{\gamma}|\mathcal{D}|^2}\mathcal{G}\,, \tag{4}$$

where $\hat{\mathcal{W}}$ is a restoration filter in the FD. Note that since D is a tensor then $|\mathcal{D}|^2 = |\mathcal{D}_x|^2 + |\mathcal{D}_y|^2$, where $\mathcal{D}_{(\cdot)}$'s are Fourier transforms of gradient operator components. The power spectrum of D is identical to the spectrum of the Laplacian operator, which is f^2 with f being the spatial frequency.

The solution $\hat{\mathcal{W}}$ resembles a standard Wiener filter with a modified power spectrum of the original image. Recall that the Wiener filter is a linear minimum mean square error (LMMSE) estimator defined as

$$\hat{\mathcal{W}} = \arg\min_{\mathcal{W}} \mathbb{E}_{u,n}\{\|\mathcal{W}\mathcal{G} - \mathcal{U}\|_2^2\}\,, \tag{5}$$

where $\mathbb{E}_{u,n}\{\cdot\}$ denotes the expectation with respect to the distribution of images and noise. The solution is the well-known formula $\hat{\mathcal{W}} = \mathcal{H}^*/(|\mathcal{H}|^2 + \mathcal{S}_{nn}/\mathcal{S}_{uu})$ where $\mathcal{S}_{uu}$ and $\mathcal{S}_{nn}$ are power spectra of the original image and noise, respectively, and are assumed to be known. The filter $\hat{\mathcal{W}}$ in (4), which is the solution of Tikhonov regularization, is thus the Wiener filter for noise $n \approx \mathcal{N}(0, 1/\gamma)$ and images with the power spectrum $\mathcal{S}_{uu} = 1/|\mathcal{D}|^2$.

A. Learning restoration filters

To avoid the problem of boundary conditions in convolution, it is preferable to have restoration filters in the SD. We discuss two approaches. A straightforward one is to use the closed form solution in (4) and estimate the corresponding SD filter $\hat{w} \in \mathbb{R}^s$ for some given size s by solving

$$\hat{w} = \arg\min_w \|\hat{\mathcal{W}} - Fw\|^2 \quad \text{s.t.} \quad w \in \mathbb{R}^s,\ \sum_i (w)_i = (\hat{\mathcal{W}})_0\,, \tag{6}$$

where the second equality constraint guarantees that the filter mean is preserved. The above constrained optimization problem has a simple solution using the method of Lagrange multipliers: transform $\hat{\mathcal{W}}$ to the SD, crop it to size s, and add an appropriate constant to preserve the original mean value. However, there are two disadvantages of this approach. First, the filter is optimal in the sense of ℓ_2-norm calculated in the PSF domain, which has limited relation to the quality of the restoration. Second, it can be used only if the explicit form (4) in the FD exists. For example, if downsampling is present in the formation model, such as in super-resolution or demosaicing, the inversion must be done numerically and the FD explicit form is not viable.

A remedy to the above problems is the second approach that solves LMMSE (5) directly in the SD, see this approach applied to demosaicing in [9]. We take an arbitrary image $\tilde{u}$ and perform type of spectral whitening by modifying the image power spectrum to $\mathcal{S}_{\tilde{u}\tilde{u}} = 1/|\mathcal{D}|^2$. Then we generate a blurred image $\tilde{g}$ following the formation model (1) with $n \approx \mathcal{N}(0, 1/\gamma)$. The pair $(\tilde{u}, \tilde{g})$ is a training set, which we then use in optimization (5) by replacing the expectation with a sample mean. The complete algorithm is summarized in Alg. 1.

Algorithm 1 Learning restoration filters

Input: h – PSF, s – filter size, σ^2 – noise variance, $\mathcal{S}$ – power spectrum

Output: $\hat{w}$ – filter of size s

1: **Generate a training pair** $(\tilde{u}, \tilde{g})$ **:**
2: Take some image $\tilde{u}$, modify its spectrum ($\mathcal{S}_{\tilde{u}\tilde{u}} = \mathcal{S}$), and generate a blurred image $\tilde{g} = h * \tilde{u} + n$ with $n \approx \mathcal{N}(0, \sigma^2)$.
3: **Solve for** $w \in \mathbb{R}^s$**:**
4: $\hat{w} = \arg\min_w \|w * \tilde{g} - \tilde{u}\|_2^2$

B. Proposed iterative algorithm

Let us now reformulate deconvolution as an optimization problem with total variation regularization [6]

$$\hat{u} = \arg\min_u \frac{\gamma}{2}\|Hu - g\|_2^2 + \|\mathsf{D}u\|_{2,1}\,. \tag{7}$$

Saddle-point methods are frequently used for solving such non-smooth convex problems. A popular choice is the 'alternating directions method of multipliers' (ADMM) [10], which is also considered here, however similar results are obtained also for 'primal-dual' methods of Chambolle and Pock [11]. The ADMM introduces an auxiliary variable $\mathsf{v} \in \mathbb{R}^{m\times 2}$ and an equality constraint $\mathsf{v} = \mathsf{D}u$, and rewrites (7) as a saddle-point problem for an 'augmented Lagrangian':

$$\min_{u,\mathsf{v}} \frac{\gamma}{2}\|Hu - g\|_2^2 + \|\mathsf{v}\|_{2,1} + \frac{\beta}{2}\|\mathsf{D}u - \mathsf{v} - \mathsf{a}\|_{2,1}^2\,, \tag{8}$$

where $\mathsf{a} \in \mathbb{R}^{m\times 2}$ is the Lagrange multiplier. Minimization with respect to the image u leads to a linear problem and if circular convolution is assumed, the result can be written in the FD as

$$\mathcal{U} = \frac{\mathcal{H}^*}{|\mathcal{H}|^2 + \frac{\beta}{\gamma}|\mathcal{D}|^2}\mathcal{G} + \frac{\mathcal{D}^*}{|\mathcal{D}|^2 + \frac{\gamma}{\beta}|\mathcal{H}|^2}(\mathcal{V} + \mathcal{A}) \tag{9}$$

The first term is the solution of Tikhonov regularization (4) and is equivalent to Wiener filtering with PSF h, image power spectrum $1/|\mathcal{D}|^2$ and noise $n \approx \mathcal{N}(0, \beta/\gamma)$. Alg. 1 is applied with parameters $\sigma^2 = \beta/\gamma$ and $\mathcal{S} = 1/|\mathcal{D}|^2$ to learn the corresponding filter in the SD. We refer to this filter as '*restoration filter*' $w_1 \in \mathbb{R}^s$. The second term can be considered as another Wiener filtering of $(\mathsf{v} + \mathsf{a})$ with vertical and horizontal filters $[1, -1]$, image power spectrum $1/|\mathcal{H}|^2$ and noise $n \approx \mathcal{N}(0, \gamma/\beta)$. In this case, Alg. 1 is unstable since the PSF power spectrum $|\mathcal{H}|^2$ typically contains values close to zero in higher frequencies and setting the image power spectrum to $1/|\mathcal{H}|^2$ is not feasible. Instead, we apply the approach in (6) and refer to the estimated filters as '*update filters*' $\mathsf{w}_2 \in \mathbb{R}^{s\times 2}$. Note that w_2 is a vector-valued function and it consists of two filters one for each component of the gradient operator D (or more if D is more complex). Examples of restoration and update filters for two different PSFs are shown in Fig. 2. The remaining update equations for the auxiliary variable v and Lagrange multiplier a are in accordance with the ADMM and consist of simple element-wise operations. In the thresholding step, the norm on the vector-valued image is calculated per pixel as $\|\mathsf{D}u - \mathsf{a}\|_2 : \mathbb{R}^{m\times 2} \to \mathbb{R}^m \equiv \left((\mathsf{D}u - \mathsf{a})_{i,1}^2 + (\mathsf{D}u - \mathsf{a})_{i,2}^2\right)^{1/2}$ and the multiplication of the vector-valued image $(\mathsf{D}u - \mathsf{a})$ with the scalar-valued image $\max(\cdot)/\|\cdot\|_2$ is done element-wise by replicating the scalar-valued image.

Algorithm 2 Iterative Wiener filtering and thresholding (IWFT)

Input: g – blurred image, (w_1, w_2) – restoration and update filters, and N – number of iterations

Output: u – sharp image

1: **Initial estimation with restoration filter:**
2: $u_1 \leftarrow w_1 * g$
3: $k \leftarrow N$, $\mathsf{a} \leftarrow 0$, $\beta \leftarrow 10\max(g)$, $u \leftarrow u_1$
4: **repeat**
5: **Element-wise soft thresholding:**
6: $\mathsf{v} \leftarrow (\mathsf{D}u - \mathsf{a}) \cdot \dfrac{\max\left(\|\mathsf{D}u - \mathsf{a}\|_2 - \frac{1}{\beta}, 0\right)}{\|\mathsf{D}u - \mathsf{a}\|_2}$
7: $\mathsf{a} \leftarrow \mathsf{a} - \mathsf{D}u + \mathsf{v}$
8: **Improve the image with update filter:**
9: $u \leftarrow u_1 + \mathsf{w}_2 * (\mathsf{v} + \mathsf{a})$
10: $k \leftarrow k - 1$
11: **until** $k = 0$ **or** relative tolerance $< 10^{-4}$

The whole algorithm, which we call the iterative Wiener filtering and thresholding (IWFT) is summarized in Alg. 2. The filters w_1 and w_2 are inputs to the algorithm and they are precomputed for the given blur and noise level in the degraded image g. The algorithm consists of three main steps: initial filtering (line 2), element-wise computation (lines 6 and 7)

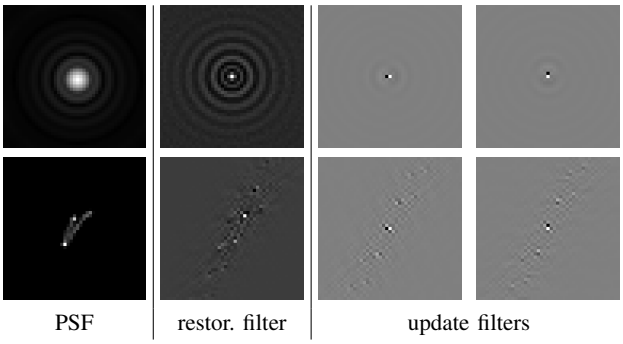


Fig. 2. Learned restoration and update filters for the Airy disk (top row) and motion blur (bottom row). From left to right: blur (h), initial restoration filter (w_1), two update filters (w_2) for horizontal and vertical differences. The filter size (s) is 45×45.

and update filtering (line 9). The first filtering step provides the initial estimator of u, which corresponds to the first term in (9). The remaining two steps are run iteratively. Update of v is element-wise soft thresholding with the threshold $1/\beta$. The parameter β is the same used in the construction of filters w_1 and w_2. Experimentally we have validated that the best results are achieved for β 10-times the range of intensity values, i.e. image gradients below $1/10$th of the intensity range are zeroed out in v. The update of the sharp image u is performed by filtering each component of the vector-valued image $(\mathsf{v} + \mathsf{a})$ with the corresponding update filter from w_2 and summing the results over the components. The algorithm finishes after satisfying one the convergence criteria: number of iterations or relative tolerance between the new estimation and the old one.

III. Experiments

The proposed IWFT algorithm solves the deconvolution problem (7) using the ADMM, which is guaranteed to converge. The appealing property of the proposed method is that all steps are implemented either by linear filters or simple element-wise operations, and thus the problem of boundary conditions in convolution is not present. The practical usage of the method is however determined by several other factors: by what margin the method outperforms the classical Wiener filter, how many iterations are generally required, and what is the minimum filter size to achieve these results. The following experiment addresses these issues in question.

The method performance was evaluated with respect to the filter size, number of iterations and noise level. The standard peak signal-to-noise (PSNR) ratio in dB was used as a performance measure. We also evaluated SSIM [12] and the results were equivalent. We took sharp images, blurred them with two types of blur – Airy disk modeling sensor blur and motion blur modeling camera shake – and added noise with SNR $= 50$, 30, 20dB. Figs. 1(a) and (b) illustrate an example of the original image and the corresponding one blurred by Airy disk, respectively. The restoration and update filters of different sizes were learned for each PSF and noise level with parameters $\gamma = 10^5$ (50dB), $\gamma = 10^3$ (30dB) and $\gamma = 10^2$ (20dB). Fig. 3 summarizes PSNR of the IWFT algorithm after $N = 0$, 1, 5 and 15 iterations for all generated images. $N = 0$ means that only the initial filtering with the restoration filter w_1 in step 2 is performed and the result is equivalent to the standard Wiener filter for the image power spectrum $1/|\mathcal{D}|^2$. In this case, strong ringing artifacts are present in the restored images as illustrated in the first column of Fig. 4. A noticeable improvement of the restored image both in the PSNR sense and visually is achieved after one application of the update filters w_2 (see the 1st iteration in the second column of Fig. 4). Additional iterations further improve the image, yet after 15 iterations (the last column) improvements are negligible.

The quality of restoration improves with the increasing filter size as expected. When the Airy disk is used, PSNR saturates for filter sizes of 45×45 in the case of 50dB. When noise increases in the image, the restoration and update filters perform more denoising than deconvolution and filters of smaller size become sufficient. So in the case of 30dB (20dB), PSNR saturates already for filter sizes of around 15×15 (10×10), however due to increased noise the achieved PSNR is lower. When the motion blur of similar effective size as the Airy disk is used, we notice that the maximum achievable PSNR is much higher. This is in accordance with the fact that Gaussian blurs (including Airy disk and out-of-focus) are more destructive than motion blurs. Examples of restored images for two noise levels and both blur types using filter size 45×45 are summarized in Fig. 5

IV. Conclusions

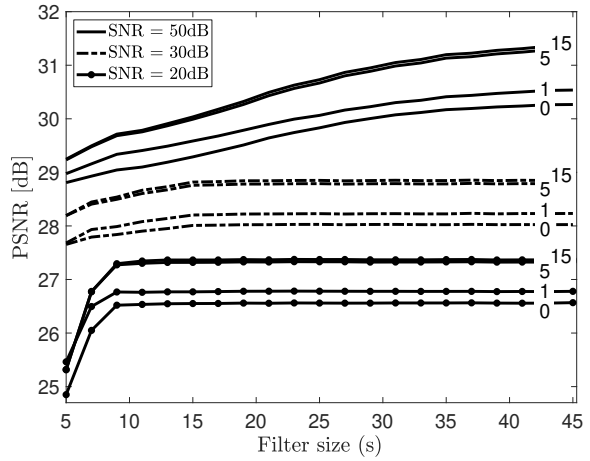
We have proposed a computationally efficient image restoration algorithm IWFT consisting of only filtering and element-wise operations, which makes it particularly suitable for implementation in digital cameras. The algorithm is based on the alternating directions method of multipliers and iteratively solves a non-smooth convex problem of deconvolution with total variation regularization using two linear filters. One filter is for initial restoration and another for updating the current estimate. Filters are implemented in the spatial domain and learned by two proposed learning methods. Experiments illustrate that the IWFT algorithm performs well for moderate filter sizes and removes ringing artifacts after only a few iterations in the case of realistic sensor and lens blurs.

A promising feature of the algorithm, which we plan to investigate in the near future, is the capability to seamlessly incorporate other restoration tasks. The algorithm for learning restoration filters is sufficiently general to estimate filters that in addition to deconvolution perform, e.g, demosaicing and super-resolution. In this case, we can replace the restoration filter for initial estimation with the newly learned filters and the rest of the algorithm remains the same.

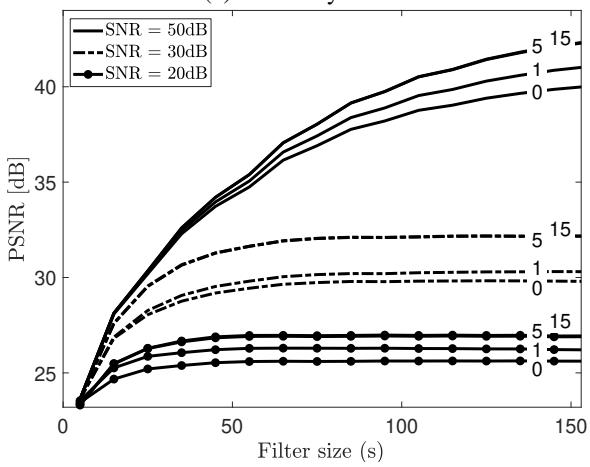
References

[1] S. J. Reeves, "Fast image restoration without boundary artifacts," *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1448–1453, 2005.

[2] M. Šorel, "Removing boundary artifacts for real-time iterated shrinkage deconvolution," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2329–2334, Apr 2012.



(a) PSF Airy disk



(b) PSF motion

Fig. 3. PSNR performance of the proposed iterative Wiener filtering with respect to the size (s) of the restoration and update filters (w_1, w_2): (a) deconvolution of the Airy disk (top row in Fig. 2), (b) deconvolution of the motion blur (bottom row in Fig. 2). Performance curves are for 0 (only initial restoration filtering), 1, 5 and 15 iterations with noise levels 50dB (solid), 30dB (dashed) and 20dB (solid&marker). PSNR improves by a wide margin after the first 5 iterations and additional iterations are unnecessary. With increasing noise the minimum optimal filter size decreases. For the motion blur, achieved PSNRs are much higher yet the method is less practical as the performance plateau is reached for larger filter sizes.

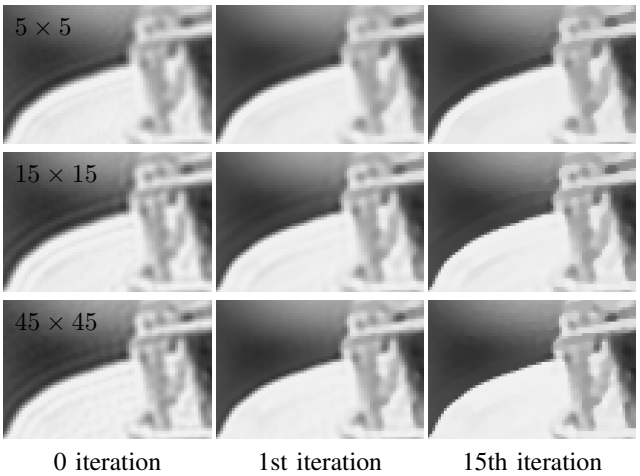


Fig. 4. Close-ups of restored images for different filter sizes and number of iterations: from top to bottom row – filter sizes of 5×5, 15×15 and 45×45; from left to right – 0 (only initial restoration filtering), 1 and 15 iterations. Images correspond to results of Airy disk deconvolution for SNR $=$ 50dB with PSNR summarized in Fig. 3(a). Parts of the original and blurred image are in Fig. 1(a)-(b).

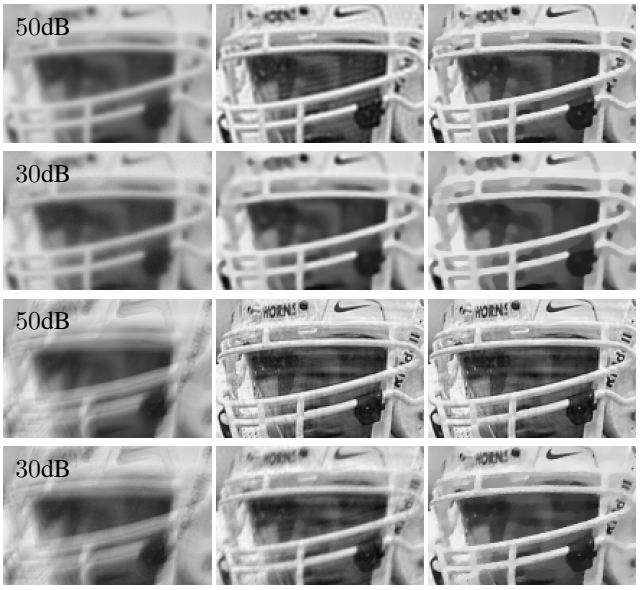


Fig. 5. Close-ups of restored images for different noise levels and number of iterations: from top to bottom row – noise levels of 50dB and 30dB; from left to right – input blurred image, 0 (only initial restoration filtering), 15 iterations. Images correspond to results of Airy disk (top two rows) and motion blur (bottom two rows) deconvolution for filter size 45×45.

[3] R. Liu and J. Jia, "Reducing boundary artifacts in image deconvolution," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE, 2008, pp. 505–508.

[4] J. Portilla, "Maximum likelihood extension for non-circulant deconvolution," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Oct. 2014, pp. 4276–4279.

[5] M. R. Banham and A. K. Katsaggelos, "Digital image restoration," *IEEE Signal Process. Mag.*, vol. 14, no. 2, pp. 24–41, Mar. 1997.

[6] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

[7] M. Almeida and M. Figueiredo, "Deconvolving images with unknown boundaries using the alternating direction method of multipliers," *Image Processing, IEEE Transactions on*, vol. 22, no. 8, pp. 3074–3086, Aug 2013.

[8] A. Chambolle and T. Pock, "An introduction to continuous optimization for imaging," *Acta Numerica*, vol. 25, pp. 161–319, 2016.

[9] J. Portilla, D. Otaduy, and C. Dorronsoro, "Low-complexity linear demosaicing using joint spatial-chromatic image statistics," in *Proc. IEEE Int. Conf. Image Processing 2005*, vol. 1, Sep. 2005, pp. I–61.

[10] 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.

[11] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of mathematical imaging and vision*, vol. 40, no. 1, pp. 120–145, 2011.

[12] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

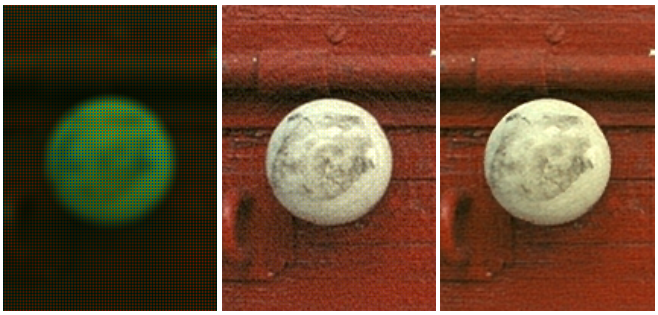
D^{3}Net: Joint Demosaicking, Deblurring and Deringing

Tomáš Kerepecký
The Czech Academy of Sciences,
Institute of Information Theory and Automation
Pod Vodárenskou věží 4, Prague, Czechia
Email: http://www.utia.cas.cz/people/kerepeck

Filip Šroubek
The Czech Academy of Sciences,
Institute of Information Theory and Automation
Pod Vodárenskou věží 4, Prague, Czechia
Email: http://www.utia.cas.cz/people/sroubek

***Abstract*—Images acquired with standard digital cameras have Bayer patterns and suffer from lens blur. A demosaicking step is implemented in every digital camera, yet blur often remains unattended due to computational cost and instability of deblurring algorithms. Linear methods, which are computationally less demanding, produce ringing artifacts in deblurred images. Complex non-linear deblurring methods avoid artifacts, however their complexity imply offline application after camera demosaicking, which leads to sub-optimal performance. In this work, we propose a joint demosaicking deblurring and deringing network with a light-weight architecture inspired by the alternating direction method of multipliers. The proposed network has a transparent and clear interpretation compared to other black-box data driven approaches. We experimentally validate its superiority over state-of-the-art demosaicking methods with offline deblurring.**

***Index Terms*—Demosaicking, deblurring, deringing, ADMM, CNN**



(a) Raw data (b) Demosaicking & Deblurring (c) Deringing

Fig. 1. The proposed convolutional neural network joints three restoration tasks: demosaicking, deblurring and deringing.

I. Introduction

Data acquired by modern digital camera sensors are subject to various types of signal degradation, such as lens and sensor blur, aberrations, color filter array (CFA) and noise. To convert the raw data from the imaging sensor into an image suitable for the human visual system, it is necessary to correctly process the acquired data, particularly by applying demosaicking and deblurring procedures. Sequential demosaicking and deblurring provides sub-optimal solutions [1], yet they are still used for their simplicity. Joint demosaicking and deblurring was studied earlier [1]–[4] using traditional model-based optimization approaches. More recent learning-based methods focus only on joint demosaicking and denoising [5]–[8] and disregard blur, which is present in DSLR and mobile phone cameras even if the lens is in focus, see Fig. 2.

An important, yet often neglected, property of restoration algorithms is their ability to run in the camera with limited computational capacity, such as pixel-wise operations and basic filtering. In this regard, a computationally efficient algorithm for deconvolution was proposed in [9]. The algorithm is based on the alternating direction method of multipliers (ADMM) [10] and performs deblurring by iterative Wiener filtering and thresholding (IWFT). Removing blur with the Wiener filter (ideal linear filter) produces mediocre results in most of the cases due to ringing artifacts around edges in the image. The IWFT algorithm instead uses two sets of filters, one for the initial restoration (deblurring) and another for the ringing artifact suppression (deringing). These filters are precomputed offline for the given type of degradation, i.e. blur and noise level.

Recent works have revealed that, with the aid of model-based optimization methods, such as Primal-Dual or ADMM, it is possible to design convolutional neural networks (CNN) with clear interpretation [8], [11]. Inspired by these studies, we design a light-weight CNN imitating the IWFT concept, which is directly applicable to raw camera data (Fig. 1). The proposed network – called D^{3}Net – performs joint demosaicking, deblurring and deringing. Network filters have clear interpretation and they become learnable parameters, which is an important advantage over the IWFT algorithm. A relatively small number of training parameters allows us to efficiently train the network by only a single pair of degraded and ground-truth images. We perform quantitative and qualitative evaluation of D^{3}Net and compare it with state-of-the-art demosaicking methods with and without offline deblurring.

This work was supported by Czech Science Foundation grant GA18-05360S and by the Praemium Academiae awarded by the Czech Academy of Sciences.

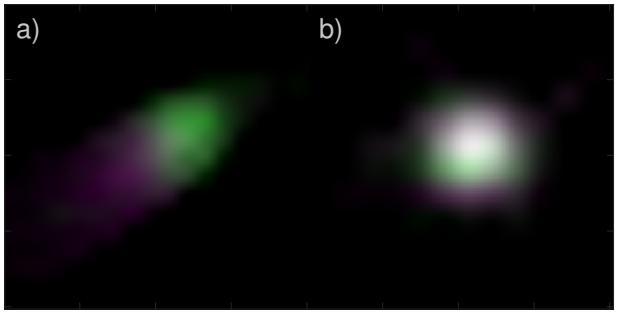


Fig. 2. Intrinsic camera blur (combination of sensor blur and lens aberrations): a) DSLR, b) mobile phone.

II. PROBLEM FORMULATION

To solve the joint demosaicking-deblurring problem, one of the most frequently used approaches in the literature relies on the following linear observation model

$$g = SHu + n\,, \tag{1}$$

where $g \in \mathbb{R}^p$ is the blurred, noisy raw image and $u \in \mathbb{R}^m$ is the unknown high-resolution sharp image. Both u and g correspond to the vectorized forms of the images. $H(\cdot) \equiv h * \cdot$ denotes a degradation operator (matrix) performing convolution with some known point spread function (PSF) h. For simplicity, we employ a stationary blur model. S represents the down-sampling operator, which models the particular CFA pattern. It corresponds to a binary matrix which excludes the spatial and channel location in the image where color information is missing. We consider additive white Gaussian noise $n \approx \mathcal{N}(0, \sigma^2)$ with zero mean and variance σ^2.

To solve the ill-posed inverse problem, we adopt the optimization problem with total variation regularization [12]:

$$\hat{u} = \arg\min_u \frac{\gamma}{2}\|SHu - g\|_2^2 + \phi^1\big(\{D_j u\}\big)\,, \tag{2}$$

where the norm of the first term is the classical ℓ_2-norm. $\phi^s\left(\{D_j u\}\right) = \sum_i (\sum_j [D_j u]_i^2)^{s/2}$ represents the regularization function. $D_j(\cdot) \equiv d_j * \cdot$ denotes the j-th feature extraction operator implemented as a convolution with the filter d_j. For example, if the set $\{d_j\}$ comprises only vertical and horizontal differences, $\{D_j u\}$ corresponds to the discrete image gradient and $\phi^1(\cdot)$ is the sum of gradient magnitudes. Pixels are indexed as $[u]_i$. Parameter γ is is the weight between the data term and regularization.

A popular choice for solving such non-smooth convex problems is ADMM [10]. The method introduces an auxiliary variables $v_j \in \mathbb{R}^m$ and an equality constraints $v_j = D_j u$, and rewrites (2) as a saddle-point problem for an 'augmented Lagrangian':

$$\min_{u,\{v_j\}} \frac{\gamma}{2}\|SHu - g\|_2^2 + \phi^1\big(\{v_j\}\big) + \frac{\beta}{2}\phi^2\big(\{D_j u - v_j - a_j\}\big)\,, \tag{3}$$

where $a_j \in \mathbb{R}^m$ represents the Lagrange multiplier.

In order to solve joint demosaicking-deblurring minimization problem (3) as well as to deal with ringing artifacts after deconvolution, we imitate IWFT concept and propose computationally efficient algorithm summarized in Alg. 1. See Algorithm 2 in [9] for more details.

Algorithm 1 Joint demosaicking, deblurring and deringing

Input: g – blurred image, N – number of iterations, $\{r_k\}$ – set of restoration filters, $\{w_j\}$ – set of update filters
Output: u – sharp image

1: **Initial estimation with restoration filter:**
2: $u_r \leftarrow P(\{r_k * g\})$ `[rConv]`
3: $k \leftarrow N$, $\{a_j\} \leftarrow 0$, $\beta \leftarrow 10\max(g)$, $u \leftarrow u_r$
4: **repeat**
5: $\tilde{v}_j \leftarrow d_j * u \quad \forall j$ `[gConv]`
6: **Soft thresholding:**
7: $v_j \leftarrow \texttt{SoftThr}\left(\tilde{v}_j - a_j, \frac{1}{\beta}\right) \quad \forall j$ `[Soft]`
8: **Update the Lagrange multiplier:**
9: $a_j \leftarrow a_j + (v_j - \tilde{v}_j) \quad \forall j$ `[Add]`
10: **Improve the image with update filter:**
11: $u \leftarrow u_r + \sum_j w_j * (v_j + a_j)$ `[uConv]`
12: $k \leftarrow k - 1$
13: **until** $k = 0$

ADMM sequentially performs alternating minimization with respect to u and $\{v_j\}$. Minimization over v_j leads to soft thresholding with the threshold $1/\beta$ (line 7). Parameter β is set to 10-times the range of intensity values of the blurred image g. In the case of minimization over u (line 11), the update step can be written as

$$u = P\left(\{r_k * g\}\right) + \sum_j w_j * (v_j + a_j)\,, \tag{4}$$

where $\{r_k\}$ and $\{w_j\}$ are the sets of restoration and update filters, respectively. Operator P performs pixel shuffling to assemble the final RGB image. These filters are inputs to the algorithm and they are precomputed offline, similarly as in [9], for the given type of degradation, i.e. blur, CFA pattern and noise level. The Lagrange multiplier a_j is updated by the term $(v_j - d_j * u)$ (line 9).

III. PROPOSED NETWORK ARCHITECTURE

Alg. 1 consists of only filtering and element-wise operations and therefore can be used to design the architecture of the light-weight convolutional neural network. As a result, all convolutional filters in the algorithm becomes learnable parameters.

The architecture of our proposed network D^{3}Net is shown in Fig. 3. The first filtering step (Alg. 1 - line 2) provides the initial estimator of reconstructed image u, which corresponds to the demosaicking and deblurring tasks. The remaining steps in Alg. 1 (line 5 - 12) are run iteratively and perform the ringing artifact suppression (deringing task). Experimentally we have validated that three iterations provide balanced results between ringing artifact suppression and over-regularization.

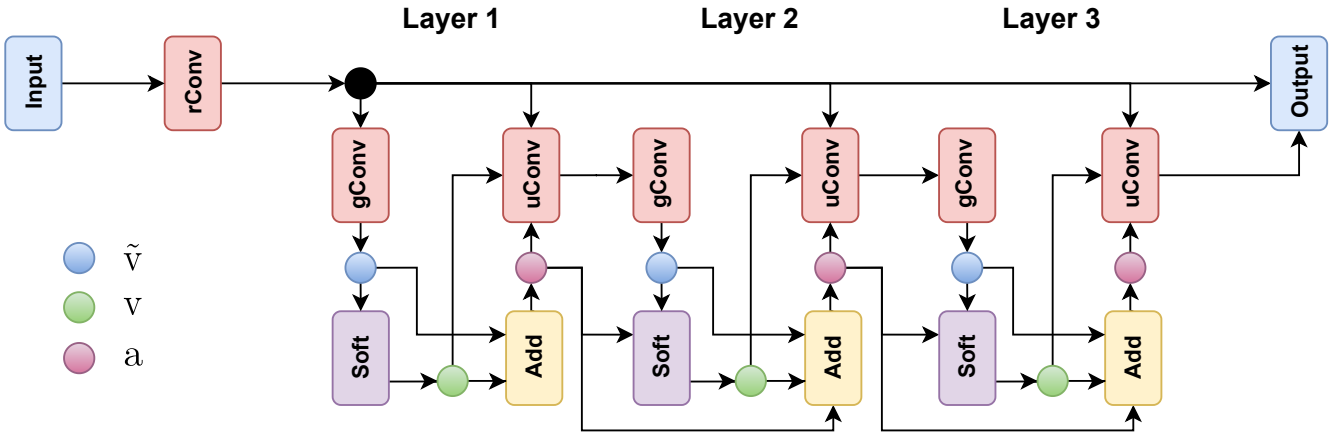


Fig. 3. Proposed network architecture of the D^{3}Net. See Alg. 1 for the interpretation of individual blocks.

The restoration layer (rConv) is therefore followed by three update layers consisting of four operations: gradient filtering (gConv), soft thresholding with the threshold $1/\beta$ (Soft), element-wise operation (Add) and update convolutional layer (uConv).

The input to our network is the degraded blurred image g (Fig. 1a) separated into four channels according to the Bayer CFA pattern. The network can be modified to any other CFA (e.g. X-Trans). The restoration layer (rConv) consists of convolution with 12 four-channel filters $\{r_k\}$ followed by pixel shuffling P which results in the restored RGB image u_r (Fig. 1b) with three times more data than g. An example of one out of 48 restoration filters for demosaicking and deconvolution is in Fig. 4. Filters are initialized using the result of IWFT algorithm (Fig. 4a) and further improved (Fig. 4b) by training with the traditional back-propagation process. The difference between the restoration filter before and after the training is demonstrated in Fig. 4c. Filter size is the hyper parameter of the network and is set, in our example, to 25×25. The human visual system is more sensitive to high frequencies in the luminance channel, therefore restored RGB image is transformed to a YCbCr color space and all update layers are applied only on luminance channel of the restored image u_r. The complexity of update layers depends on the number of gradient filters. In our implementation, we use 4 filters in the convolutional layers (gConv) which correspond to horizontal, vertical and two diagonal difference operators (Fig. 5a). Consequently, there are four update filters for each update layer (Fig. 6a). Gradient and update filters are also initialized by IWFT. Their modification through the learning procedure is more significant than in the case of restoration filters. The final output of the network (Fig. 1c) is the reconstructed image u, which is the restored luminance channel with chrominance channels transformed back to RGB color space. In total, the proposed network contains 36 convolutions in seven convolutional blocks.

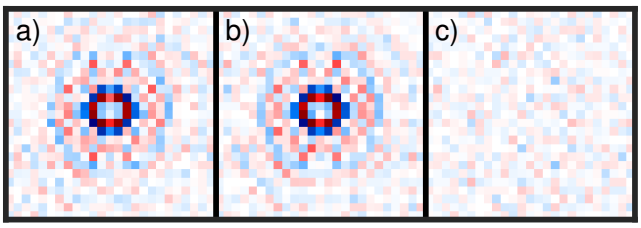


Fig. 4. An example of restoration filter (25x25) for demosaicking and deblurring Bayer data distorted by out-of-focus blur. There are 48 restoration filters in layer rConv. a) initialization from IWFT, b) learned by D^{3}Net, c) difference. Blue-white-red colormap represents numbers from -1 to 1.

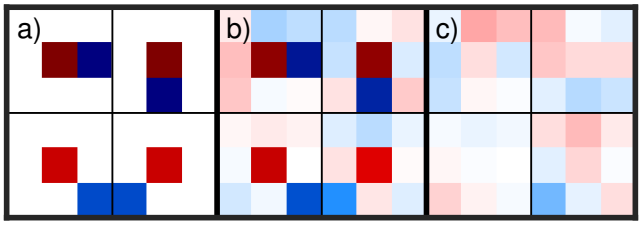


Fig. 5. An example of horizontal, vertical and two diagonal gradient filters (3x3) in layer gConv for deringing images distorted by out-of-focus blur. a) initialization from IWFT, b) learned by D^{3}Net, c) difference. Blue-white-red colormap represents numbers from -1 to 1.

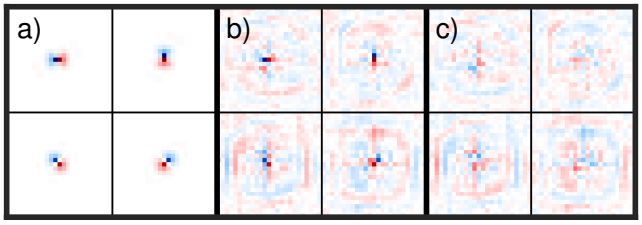


Fig. 6. An example of 4 update filters (25x25) in layer uConv for deringing images distorted by out-of-focus blur. The update filters are related to gradient filters. a) initialization from IWFT, b) learned by D^{3}Net, c) difference. Blue-white-red colormap represents numbers from -1 to 1.

IV. EXPERIMENTS

We use Pytorch as our framework for implementing D^{3}Net. First, we demonstrate the improvement of results from our network over augmented IWFT. In addition, we focus on deringing effect of the proposed network as well as to show the influence of different blurs on final reconstruction. Then we compare our results with sequential demosaicking and deconvolution methods. Finally, we test our network on real images. Throughout the experiments, two objective quality measures were used: Peak Signal to Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM).

A. IWFT vs. D^{3}Net

For training and evaluation of the proposed network, we used publicly available Kodak PhotoCD image dataset. One image from the set was used as a training set and the remaining 23 images formed a validation set. In this experiment, input images were randomly cropped into patches of size 200×200 pixels. We converted the Kodak images into blurred Bayer images by performing an image degradation process (1). Blurred Kodak images were down-sampled with the Bayer pattern GRBG and finally Gaussian noise was added.

Batch size was set to 4. The network was optimized with the mean-squared-error loss. All weights were initialized by filters of the IWFT algorithm, computed similarly as in [9]. Optimization was carried out using the stochastic gradient descent algorithm with learning rate 0.01 and momentum 0.9. Training was super fast with only one epoch, which corresponds to approximately 3.5 minutes on a GeForce RTX 2080 Ti.

We tested both methods, IWFT and D^{3}Net, on out-of-focus blur represented by circular PSF with radius 5 and noise levels 30dB and 40dB. Size of the restoration and update filters were the same and ranged from 5×5 to 35×35. Size of the gradient filters was 3×3. Fig. 7 demonstrates the improvement of the results using the learning-based approach. It can be concluded that proposed network gives significantly better PSNR results than IWFT for all filter sizes and noise levels. For the given out-of-focus blur with radius 5, the performance of both methods flattens out for filter sizes of 25×25 and more.

B. Deringing effect

As discussed in Sec. III, update filters change through training more than restoration filters. Therefore we analyzed the performance of update layers, mainly their deringing effect. To train our network, images from Kodak dataset were degraded in the same way as in the Sec. IV-A. This time we used two types of blur: out-of-focus blur with radius 5 and Gaussian blur with variance 3. To form training set with 582930 degraded and ground-truth image pairs, we used 18 images from Kodak dataset and cropped them into patches of size 100×100. The remaining six Kodak images composed validation set. We considered Gaussian noise 40dB and filter sizes 25×25. Other parameters remained the same as in the previous case. Training of our network lasted approximately 9 hours on a GeForce RTX 2080 Ti.

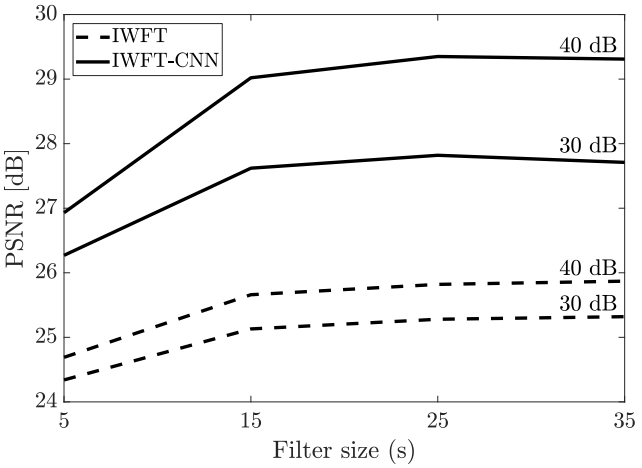


Fig. 7. Average PSNR performance of the proposed D^{3}Net (solid) and IWFT (dashed) [9] with respect to the size (s) of the restoration and update filters ($\{r_k\}, \{w_j\}$). Out-of-focus blur and noise levels with 30dB and 40dB are considered. Proposed network outperforms IWFT for all filter sizes.

Tabs. I and II) compares average PSNR and SSIM of the reconstructed images for D^{3}Net, standard demosaicking method [13], Wiener filter and IWFT algorithm in the case of out-of-focus blur and Gaussian blur, respectively.

TABLE I
OUT-OF-FOCUS BLUR: AVERAGE PSNR AND SSIM RESULTS FOR DIFFERENT RECONSTRUCTION METHODS.

Method	PSNR [dB]	SSIM
Demosaicked [13]	24.69	0.757
Wiener	26.10	0.874
IWFT	25.82	0.823
D^{3}Net (proposed)	**29.94**	**0.926**

TABLE II
GAUSSIAN BLUR: AVERAGE PSNR AND SSIM RESULTS FOR DIFFERENT RECONSTRUCTION METHODS.

Method	PSNR [dB]	SSIM
Demosaicked [13]	24.53	0.752
Wiener	24.34	0.780
IWFT	25.27	0.856
D^{3}Net (proposed)	**26.89**	**0.870**

(a) Raw data (b) Demosaicked [13] (c) Wiener (d) IWFT (e) D^{3}Net (proposed) (f) Ground-truth

Fig. 8. Visual results from the Kodak dataset. (a) degraded data by out-of-focus blur, CFA Bayer pattern and Gaussian noise with 40dB, (b) applying only demosaicking [13], (c) Wiener filtering, (d) joint demosaicking and deblurring using IWFT method [9], (e) joint demosaicking and deblurring using our proposed network, (f) original sharp image. Our proposed method D^{3}Net retain fine details as opposed to IWFT method that over-smooths highly textured areas while suppresses ringing artifacts when only Wiener filtering is considered.

Wiener filter is equivalent to initial restoration in the IWFT algorithm (i.e. layer rConv in D^{3}Net without learning). It is a popular deconvolution method, however, as a linear filter, the estimated image exhibits ringing artifacts around edges (Fig 8c). Non-linear update steps of IWFT algorithm performed well in suppressing ringing artifacts, hence results were visually better. However, such reconstructed images were over-smoothed (Fig. 8d). This eventually led to lower PSNR and SSIM values for IWFT than for Wiener filter when out-of-focus blur was considered (Tab. I). It was not the case for images blurred by Gaussian PSF, although images remained too smoothed as can be seen in Fig. 9d. Images corrected by D^{3}Net did not suffer from these problems. In Fig. 8e details of the wall are still recognizable as opposed to IWFT. Overall, the proposed network was able to recover more realistic details than the optimization-based IWFT as well as produce images with less visually disturbing artifacts than Wiener-like filters.

C. Joint vs. sequential approach

This experiment presents the comparison of the proposed joint approach with sequential demosaicking and deblurring procedures and evaluates the effect of training the proposed network on more than one image pair. The network trained in Sec. IV-B using 18 Kodak images is denoted D^{3}Net v2 and the network trained on a single pair of degraded and ground-truth image is D^{3}Net v1. We evaluated our networks on the McMaster dataset [14] and compared them with recent demosaicking methods FlexISP [3], DeepJoint [15] and JointADMM [5] followed by robust non-blind deconvolution method (non-blind deconvolution step in [16]). The kernel part of those algorithms, including all parameters, remains the same as their authors provided. Applying offline deblurring is identified by the asterisk symbol *.

An interesting result is that the reconstructed image provided by standard IWFT as well as Wiener filter (Fig. 10e-f) looks visually better and retain relatively fine details as opposed to the other sequential demosaicking and deblurring

(a) Raw data (b) Demosaicked [13] (c) Wiener (d) IWFT (e) D^{3}Net (proposed) (f) Ground-truth

Fig. 9. Visual results from the Kodak dataset. (a) degraded data by Gaussian blur, CFA Bayer pattern and Gaussian noise with 40dB, (b) applying only demosaicking [13], (c) Wiener filtering, (d) joint demosaicking and deblurring using IWFT method [9], (e) joint demosaicking and deblurring using our proposed network, (f) original sharp image.

methods, yet they received worse PSNR values (Tab. III).

From Tab. III we observe that D^{3}Net yields substantially better results than all other tested methods. Surprisingly, even network trained on a single image pair (D^{3}Net v1) outperforms sequential demosaicking and deblurring. Our methods leads to better and more visually pleasing results, as it can be seen in Fig. 10l.

TABLE III
OUT-OF-FOCUS BLUR: AVERAGE PSNR AND SSIM RESULTS FOR THE DIFFERENT RECONSTRUCTION METHODS.

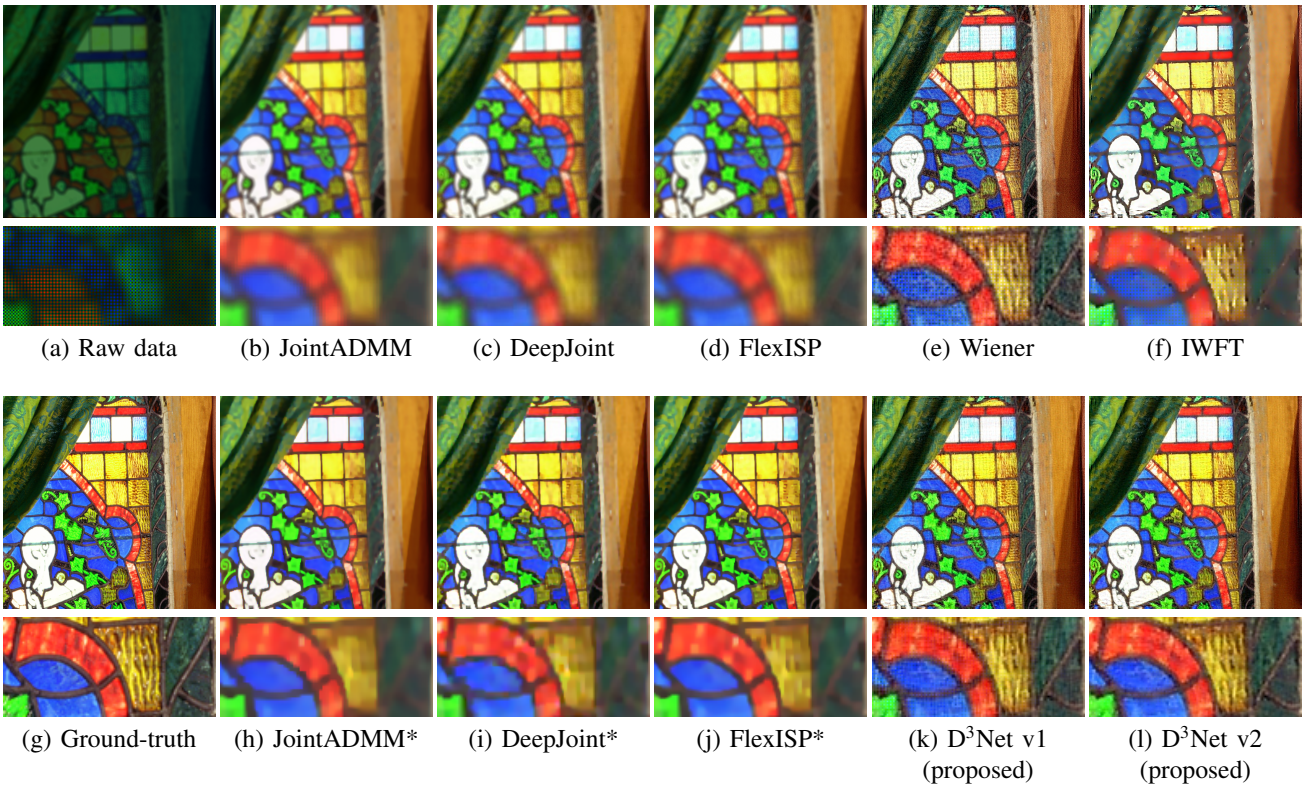
Method	PSNR [dB]	SSIM
JointADMM	23.06	0.742
DeepJoint	23.40	0.751
FlexISP	23.37	0.763
Wiener	23.57	0.826
IWFT	24.07	0.843
JointADMM*	25.44	0.839
DeepJoint*	25.62	0.846
FlexISP*	26.48	0.882
D^{3}Net v1	27.61	0.887
D^{3}Net v2	**28.91**	**0.912**

D. Results on real image

We tested D^{3}Net on real data captured by LG Nexus 5 mobile phone camera (8 MP, f/2.4, 4 mm, 1/6 sec, RGGB). The mobile phone processed the raw data and stored the image as JPEG. We analyzed cropped patch with the size of 449×433 which is shown in Fig 11a. In the inset of the figure, zoomed minipatch of size 46×46 is presented. Notice the demosaicking artifacts in the image.

Intrinsic camera blur kernels for different regions of the input image can be estimated in advance according to [1]. To train our network we artificially blurred test images from Kodak dataset with PSF (Fig. 2b) corresponding to the selected patch of the captured image. We considered additive Gaussian noise 35 dB. Eventually, blurred Kodak images were down-sampled with the Bayer pattern RGGB. In order to prevent over-fitting of the network, size of the restoration and update filters were set to 3×3. To form training and validation set, Kodak images were cropped into patches of size 210×310. Other parameters were set as in IV-B. Training of our network lasted approximately 5 minutes on a GeForce RTX 2080 Ti.

The raw image as returned by the camera API was processed through D^{3}Net and the output is seen in Fig 11b. By comparison, the result of our method reveals greater detail, looks visually more pleasing and does not suffer from disturbing demosaicking artifacts. Small size (3×3) of restoration and update filters makes it particularly suitable for implementation in small embedded system like digital cameras.



(a) Raw data (b) JointADMM (c) DeepJoint (d) FlexISP (e) Wiener (f) IWFT

(g) Ground-truth (h) JointADMM* (i) DeepJoint* (j) FlexISP* (k) D^{3}Net v1 (proposed) (l) D^{3}Net v2 (proposed)

Fig. 10. Comparison of our joint demosaicking deblurring and deringing network D^{3}Net with sequential demosaicking and deblurring methods (FlexISP [3], DeepJoint [15] and JointADMM [5] followed by robust non-blind deconvolution method [16]). Evaluated on McMaster dataset [14]. Symbol * means that offline deblurring was applied. D^{3}Net v1 was trained using a single pair and D^{3}Net v2 was trained on 18 pairs of degraded and ground-truth images from Kodak dataset.

Fig. 11. Image reconstruction of real data captured by LG Nexus 5 phone camera. a) Demosaicking processed by phone, b) D^{3}Net.

V. CONCLUSIONS

In this work, we presented a novel portable CNN for joint demosaicking, deblurring and deringing of raw image data. The light-weight structure of the network makes it particularly suitable for implementation in digital cameras. Architecture of the proposed network is inspired by the model-based optimization algorithm IWFT. We adopted the IWFT idea, extended it to perform also demosaicking, and designed it as a CNN. Results demonstrate that filters used for image reconstruction can be further improved by adopting the learning-based approach. We have shown, that our joint approach outperforms state-of-the-art demosaicking methods with offline deblurring.

REFERENCES

[1] C. J. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf, "Non-stationary correction of optical aberrations," in *2011 International Conference on Computer Vision*. IEEE, 2011, pp. 659–666.

[2] H. Q. Luong, B. Goossens, J. Aelterman, A. Pižurica, and W. Philips, "A primal-dual algorithm for joint demosaicking and deconvolution," in *2012 19th IEEE International Conference on Image Processing*. IEEE, 2012, pp. 2801–2804.

[3] F. Heide, M. Steinberger, Y.-T. Tsai, M. Rouf, D. Pajak, D. Reddy, O. Gallo, J. Liu, W. Heidrich, K. Egiazarian *et al.*, "Flexisp: A flexible camera image processing framework," *ACM Transactions on Graphics (TOG)*, vol. 33, no. 6, pp. 1–13, 2014.

[4] D. S. Yoo, M. K. Park, and M. G. Kang, "Joint deblurring and demosaicing using edge information from bayer images," *IEICE TRANSACTIONS on Information and Systems*, vol. 97, no. 7, pp. 1872–1884, 2014.

[5] H. Tan, X. Zeng, S. Lai, Y. Liu, and M. Zhang, "Joint demosaicing and denoising of noisy bayer images with admm," in *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017, pp. 2951–2955.

[6] D. S. Tan, W.-Y. Chen, and K.-L. Hua, "Deepdemosaicking: Adaptive image demosaicking via multiple deep fully convolutional networks," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2408–2419, 2018.

[7] E. Schwartz, R. Giryes, and A. M. Bronstein, "Deepisp: Toward learning an end-to-end image processing pipeline," *IEEE Transactions on Image Processing*, vol. 28, no. 2, pp. 912–923, 2018.

[8] F. Kokkinos and S. Lefkimmiatis, "Iterative joint image demosaicking and denoising using a residual denoising network," *IEEE Transactions on Image Processing*, vol. 28, no. 8, pp. 4177–4188, 2019.

[9] F. Šroubek, T. Kerepeckỳ, and J. Kamenickỳ, "Iterative wiener filtering for deconvolution with ringing artifact suppression," in *2019 27th European Signal Processing Conference (EUSIPCO)*. IEEE, 2019, pp. 1–5.

[10] 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.

[11] S. Wang, S. Fidler, and R. Urtasun, "Proximal deep structured models," in *Advances in Neural Information Processing Systems*, 2016, pp. 865–873.

[12] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

[13] H. S. Malvar, L.-w. He, and R. Cutler, "High-quality linear interpolation for demosaicing of bayer-patterned color images," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3. IEEE, 2004, pp. iii–485.

[14] L. Zhang, X. Wu, A. Buades, and X. Li, "Color demosaicking by local directional interpolation and nonlocal adaptive thresholding," *Journal of Electronic imaging*, vol. 20, no. 2, p. 023016, 2011.

[15] M. Gharbi, G. Chaurasia, S. Paris, and F. Durand, "Deep joint demosaicking and denoising," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 6, pp. 1–12, 2016.

[16] J. Pan, Z. Lin, Z. Su, and M.-H. Yang, "Robust kernel estimation with outliers handling for image deblurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2800–2808.

DUAL-CYCLE: SELF-SUPERVISED DUAL-VIEW FLUORESCENCE MICROSCOPY IMAGE RECONSTRUCTION USING CYCLEGAN

Tomas Kerepecky[1,2], *Jiaming Liu*[2], *Xue Wen Ng*[3], *David W. Piston*[3], *and Ulugbek S. Kamilov*[2]

[1]The Czech Academy of Sciences, Institute of Information Theory and Automation, Prague, Czechia
[2]Computational Imaging Group (CIG), Washington University in St. Louis, St. Louis, MO 63130
[3]Department of Cell Biology and Physiology, Washington University School of Medicine, St. Louis, MO 63110

ABSTRACT

Three-dimensional fluorescence microscopy often suffers from anisotropy, where the resolution along the axial direction is lower than that within the lateral imaging plane. We address this issue by presenting Dual-Cycle, a new framework for joint deconvolution and fusion of dual-view fluorescence images. Inspired by the recent Neuroclear method, Dual-Cycle is designed as a cycle-consistent generative network trained in a self-supervised fashion by combining a dual-view generator and prior-guided degradation model. We validate Dual-Cycle on both synthetic and real data showing its state-of-the-art performance without any external training data.

Index Terms— Light-sheet fluorescence microscopy, Dual-view imaging, deep learning, image deconvolution.

1. INTRODUCTION

Three-dimensional fluorescence imaging, such as light-sheet fluorescence microscopy (LSFM) [1,2] is an essential tool for revealing important structural information in biological samples. However, it is common for 3D fluorescence microscopy to suffer from spatial-resolution anisotropy, where the axial direction is more blurry than the lateral imaging plane. Such anisotropy is due to several factors, including the diffraction of light and axial undersampling.

The spatial-resolution anisotropy is often addressed using image deconvolution methods, such as Richardson-Lucy algorithm [3, 4]. However, achieving isotropic resolution from a single 3D volume is an ill-posed inverse problem. The problem can be simplified by using multiview microscopy systems, such as dual-view inverted selective plane illumination microscope (diSPIM) [5,6], equipped with classical joint multi-view deconvolution and fusion methods [5, 7, 8].

Deep learning (DL) has emerged as an alternative to the classical deconvolution algorithms [9–11]. Neuroclear [10] is a recent self-supervised DL framework that uses cycle-consistent generative adversarial network (CycleGAN) [12] to improve the axial resolution from a single 3D input image without any knowledge of the point spread function (PSF). However, in many cases, the experimental PSF can be readily measured using either fluorescent beads [8, 13] or small structures within samples [14], or derived theoretically [15].

In this paper, we present Dual-Cycle as an improvement to Neuroclear that extends it into a dual-view self-supervised model-based framework. The inclusion of an additional view as input improves the reconstruction capability, while the additional prior on estimated PSFs allows our model to account for the expected degradation process. We experimentally validate Dual-Cycle on synthetic and real data showing that it can outperform Neuroclear as well as traditional dual view reconstruction algorithms.

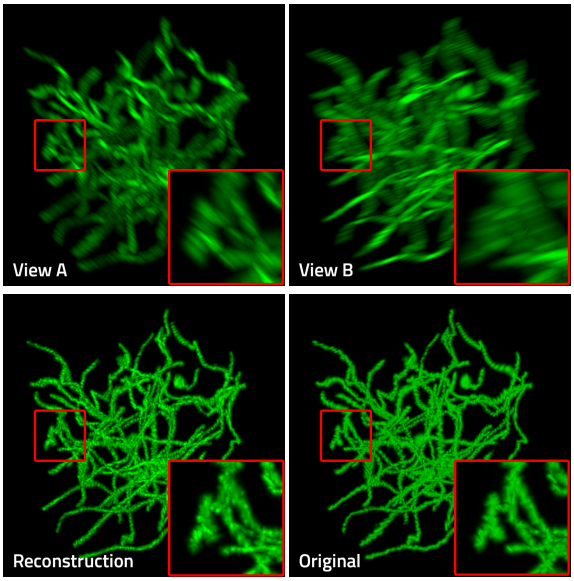


Fig. 1. Dual-Cycle reconstructs a 3D image with isotropic resolution given two views, A and B, of the same sample.

This work was supported in part by the Czech Science Foundation grant GA21-03921S, the NSF CAREER award CCF-2043134, the Fulbright commission under the Fulbright-Masaryk award, and by the Beckman Center for Advanced Light-Sheet Microscopy at Washington University in St. Louis.

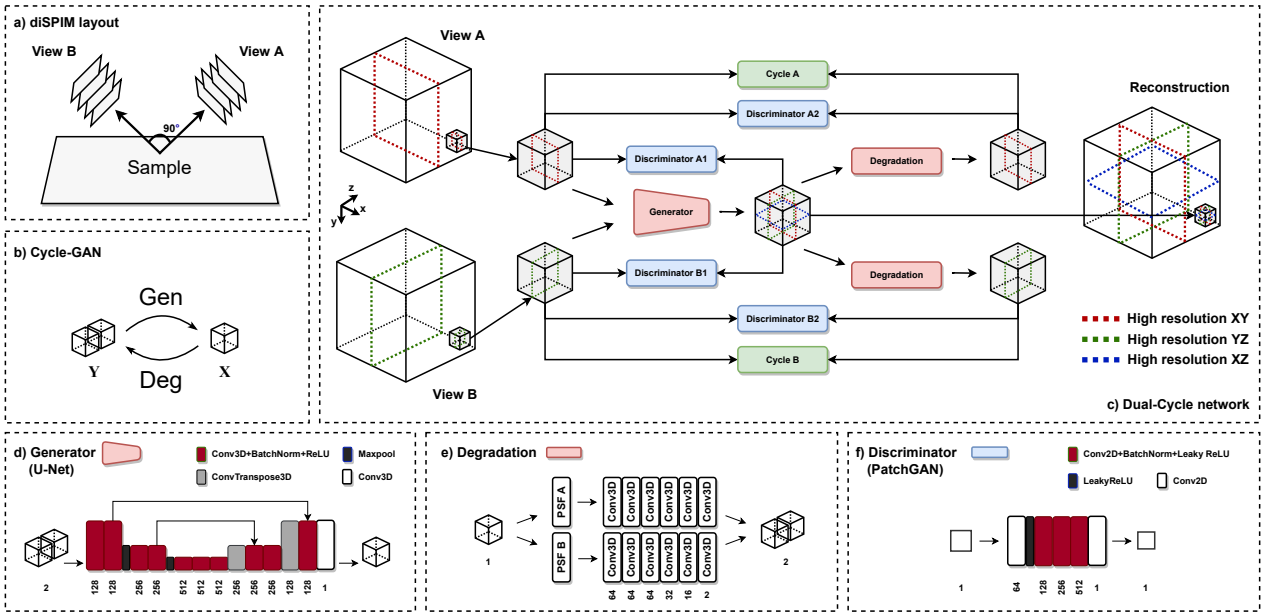


Fig. 2. Schematic illustration of the Dual-Cycle framework. a) Scheme of dual-view inverted selective plane illumination microscope (diSPIM). b) CycleGAN approach: for two domains Y and X, CycleGAN learns two mutually-inverse generator mappings Gen and Deg with the assistance of corresponding discriminators. c) Dual-Cycle network architecture. d) Schematic of the generator based on U-Net. e) Degradation forms two paths each consisting of blurring with known PSF followed by the deep linear generator. f) PatchGAN-based [16] discriminators work on 2D slices of input 3D volumes.

2. FORWARD PROBLEM

We focus on images recorded with single-plane illumination microscopes (SPIMs) [17] in a dual-view setup (diSPIM, Fig. 2a). Data is acquired by two cameras, A and B, with an ideal relative rotation of 90 degrees. The image formation process (forward model) can be represented as the following linear observation model:

$$\begin{aligned} \boldsymbol{g}_A &= \boldsymbol{\mathcal{A}}_A \boldsymbol{\mathcal{H}}_A \boldsymbol{u} + \boldsymbol{n}, \\ \boldsymbol{g}_B &= \boldsymbol{\mathcal{R}}_\perp \boldsymbol{\mathcal{A}}_B \boldsymbol{\mathcal{H}}_B \boldsymbol{u} + \boldsymbol{n}. \end{aligned} \tag{1}$$

where $\boldsymbol{g}_A$, $\boldsymbol{g}_B$, and $\boldsymbol{u}$ correspond to the vectorized forms of deskewed 3D volumes, measured by camera A (View A), camera B (View B), and the original high-resolution 3D volume (Fig. 1). $\boldsymbol{\mathcal{H}}_A$ (resp. $\boldsymbol{\mathcal{H}}_B$) denote 3D convolution along the axial direction z (resp. x) with some known PSF h_A (resp. h_B). To model the mismatch from an ideal dual view setup, we include operators $\boldsymbol{\mathcal{A}}_{A/B}$, representing 3D affine transformation. We assume a coordinate system of unknown image $\boldsymbol{u}$ to be the same as $\boldsymbol{g}_A$ and that the ideal rotation of View B with respect to View A is 90 degrees around axes y, denoted as $\boldsymbol{\mathcal{R}}_\perp$. We omit subsampling in the axial directions by interpolating measurements to have voxels of equal size. In the general case, we consider additive noise $\boldsymbol{n}$.

Problem 1 leads to an inherently ill-posed inverse problem. To solve it, we adopt and extend the approach in [10].

3. INVERSE PROBLEM

Our proposed framework is illustrated in Fig. 2c. In our setup, View A has a higher resolution in the xy plane and is blurred in the axial direction z, while View B has a higher resolution in the yz plane and is blurred in the axial direction x. Our goal is to reconstruct the original 3D volume with an isotropic resolution. We focus mainly on joint deconvolution and fusion with additional fine registration. Our framework is based on a CycleGAN approach illustrated in Fig. 2b and consists of two cycle-consistency paths, hence the name Dual-Cycle. It is worth mentioning that Dual-Cycle does not require any external training data beside the test object to be reconstructed.

The two views of the 3D volume are used as input for the 3D U-net-based generator (Fig. 2d). The result of the generator is one 3D image representing the original 3D volume with isotropic resolution. To achieve this, we employ two sets of discriminators A1 and B1 (Fig. 2f). Discriminators A1 distinguish between xy planes of View A and xy and xz planes of the reconstructed volume. Discriminators B1 distinguish between yz planes of View B and yz and xz planes of the reconstructed volume. To regularize and stabilize learning, the dual-cycle consistency is imposed. Therefore, the reconstructed image is degraded along two paths to imitate the forward problem (1). Consequently, *Degradation* A and B, Fig. 2e, consist of 3D convolution with given PSFs h_A

and h_B followed by a deep linear generator (DLG) to address ideal model mismatch caused by affine operators $\mathcal{A}$. For the blind case, when PSFs are unknown, degradation can be performed by DLGs only. Eventually, two other sets of discriminators A2 and B2 are added to map the distribution of corresponding planes of input View A/B onto generated View A/B. All discriminators are PatchGAN-based [16] and work on 2D slices of analyzed 3D volumes (Fig. 2f). Pixel-wise L1 loss between View A/B and generated View A/B is added to the GAN objective function to enforce cycle consistency.

4. EXPERIMENTAL VALIDATION

We now present the numerical evaluation of Dual-Cycle on synthetic and real light-sheet data.

4.1. Synthetic data

We first illustrate possible improvements due to our dual-view framework over the single-view Neuroclear [10]. Additionally, we compare our network with other commonly used multi-view reconstruction techniques diSPIMFusion [9] and MIPAV-generatefusion [6]. The performance was measured using the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM).

We consider a dataset of six generated 3D volumes (120 × 120 × 120 voxels), shown in Fig. 3. We drew 30-50 lines randomly in space and applied 3D elastic grid-based deformation. These volumes were treated as original ground truth volumes. All images were scaled to have values in the range 0-1. To obtain degraded volumes View A/B, we used the degradation process (1), without noise and 90-degree rotation. The original volume was blurred in the z direction for View A and in the x direction for View B (blurring by Gaussian kernel with a standard deviation in range 2-4). Further, we applied random affine transformations to simulate the imperfection of the registration method. Relatively small mismatch (representing by $\mathcal{A}$ in eq. (1)) is implemented as transformation of 3D points $\boldsymbol{p}$ as follows: $\boldsymbol{p}' = (\mathbb{I} + \mathbb{N})\boldsymbol{p} + \boldsymbol{t}$, where $\mathbb{I}$ is identity matrix and $\mathbb{N}$ is random matrix with elements from a uniform distribution over $[-0.0025, 0.0025]$, and $\boldsymbol{t}$ is random translation vector sampled from a uniform distribution over $[-0.05, 0.05]^3$.

Except for Neuroclear, all methods use prior knowledge about the PSFs and both views as input. Visual comparison of reconstructed volumes corresponding to the first 3D volume of the synthetic dataset is in Fig. 4. All methods can effectively perform the reconstruction, yet the improvement of Dual-Cycle compared to single view baseline is visually noticeable and corroborated by an increase in SSIM. Table 1 summarizes the average PSNR/SSIM results of the tested methods. Overall, Dual-Cycle improves over the second best methods by 1.49 db (PSNR) and 0.017 (SSIM).

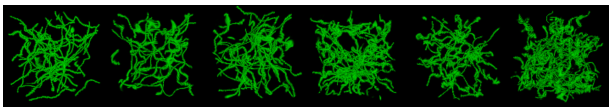


Fig. 3. The set of six generated 3D volumes used in experiments.

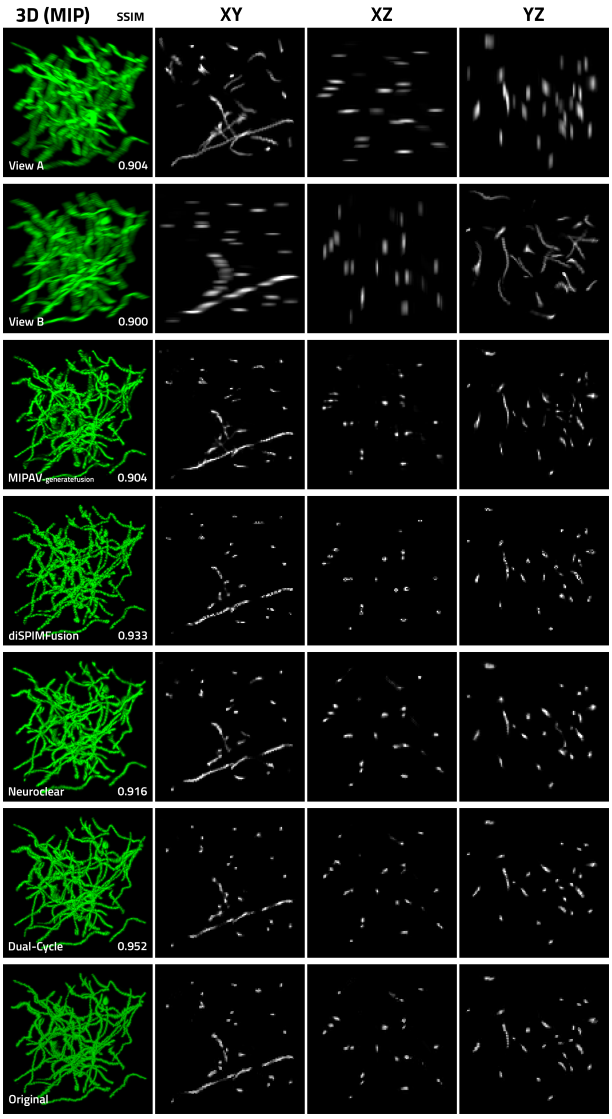


Fig. 4. Comparison of MIPAV-generatefusion [6], diSPIM-Fusion [9], Neuroclear [10], and Dual-Cycle applied on the views A and B generated from the first 3D volume in the synthetic dataset in Fig. 3. Visualized XY, XZ, and YZ images represent central cross-sections of the corresponding cubes in xy, xz, and yz planes. Each reconstruction is labeled with its SSIM value with respect to the original volume.

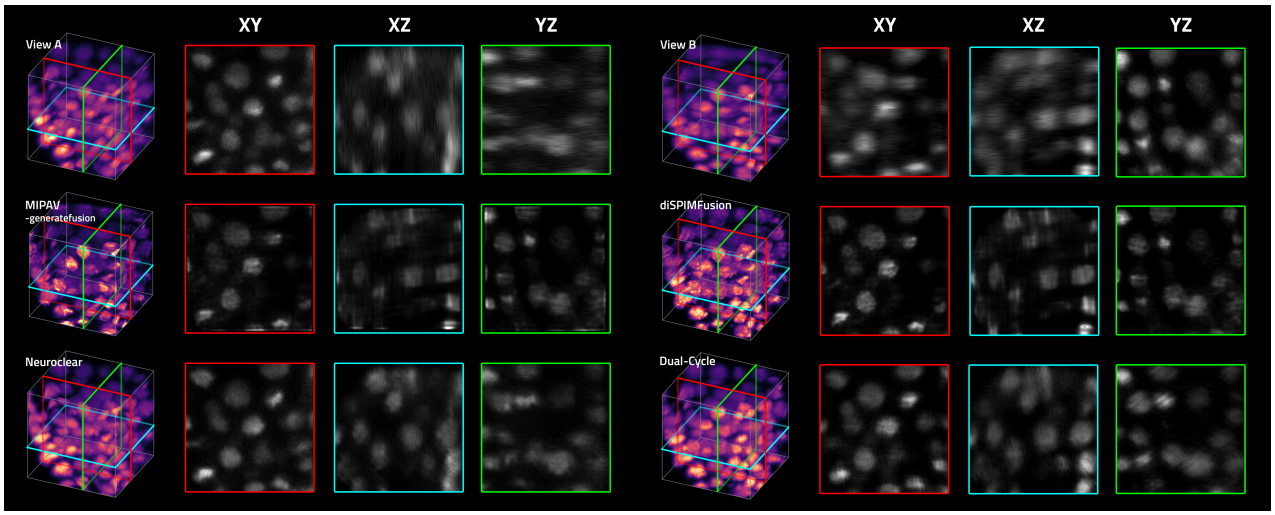


Fig. 5. Image reconstruction from real diSPIM data from [9] using reconstruction methods MIPAV-generatefusion [6], diSPIM-Fusion [9], Neuroclear [10], and the proposed Dual-Cycle framework.

Implementation of Dual-Cycle was based on the Neuroclear and CycleGAN PyTorch framework; we used Adam optimizer and learning rate set to 0.0001. The network was initialized with weights pre-trained on the first volume. The training of the first (resp. following volumes) lasted approximately 12 hours (resp. 3-6 hours) using NVIDIA RTX A5000.

4.2. Real data

We also tested reconstruction on diSPIM data from [9]. Data was preprocessed using the Fiji software [18]. The preprocessing involved denoising of both views and performing initial coarse registration of View B on View A. For both views: the minimum brightness value was truncated at value 78, volumes were normalized to 0-1 range, and were interpolated to have voxel sizes equal to $(0.1625\ \mu\text{m})^3$. For the registration of view B on view A, we used Fiji plugin Fijiyama [19]. Images were cropped to $120 \times 120 \times 120$ voxels and tested with the same methods as in Sec. 4.1. Visual comparison of reconstructed volumes is presented in Fig. 5. The improvement of Dual-Cycle reconstruction over the Neuroclear is indicated via cross sections. Overall, Dual-Cycle achieves comparable or better performance relative to the state-of-the-art methods.

Table 1. The average PSNR/SSIM results of the blurred view A/B, MIPAV-generatefusion, diSPIMFusion, Neuroclear and Dual-Cycle on the testing 3D volumes.

Method	PSNR [dB]	SSIM
View A	29.32	0.929
View B	29.13	0.927
MIPAV-generatefusion [6]	29.13	0.931
diSPIMFusion [9]	28.55	0.943
Neuroclear [10]	29.79	0.942
Dual-Cycle (our)	**31.28**	**0.960**

5. CONCLUSION

We presented Dual-Cycle, a self-supervised framework for dual-view fluorescence image reconstruction. The proposed method extends the recent Neuroclear method based on the CycleGAN framework. Compared to Neuroclear, Dual-Cycle includes two perpendicular views of the sample as input and uses prior knowledge on the estimated PSFs as a part of the degradation process within the framework. We have experimentally shown that Dual-Cycle achieves the state-of-the-art performance on synthetic and real data. While we only explored the dual-view setup in this work, our framework can be readily expanded into the multiple-view regime.

6. REFERENCES

[1] Ernst HK Stelzer, Frederic Strobl, Bo-Jui Chang, Friedrich Preusser, Stephan Preibisch, Katie McDole, and Reto Fiolka, "Light sheet fluorescence microscopy," *Nature Reviews Methods Primers*, vol. 1, no. 1, pp. 1–25, 2021.

[2] Renhao Liu, Yu Sun, Jiabei Zhu, Lei Tian, and Ulugbek S Kamilov, "Recovery of continuous 3d refractive index maps from discrete intensity-only measurements using neural fields," *Nature Machine Intelligence*, pp. 1–11, 2022.

[3] William Hadley Richardson, "Bayesian-based iterative method of image restoration," *JoSA*, vol. 62, no. 1, pp. 55–59, 1972.

[4] Leon B Lucy, "An iterative technique for the rectification of observed distributions," *The astronomical journal*, vol. 79, pp. 745, 1974.

[5] Yicong Wu, Peter Wawrzusin, Justin Senseney, Robert S Fischer, Ryan Christensen, Anthony Santella, Andrew G York, Peter W Winter, Clare M Waterman, Zhirong Bao, et al., "Spatially isotropic four-dimensional imaging with dual-view plane illumination microscopy," *Nature biotechnology*, vol. 31, no. 11, pp. 1032–1038, 2013.

[6] Abhishek Kumar, Yicong Wu, Ryan Christensen, Panagiotis Chandris, William Gandler, Evan McCreedy, Alexandra Bokinsky, Daniel A Colón-Ramos, Zhirong Bao, Matthew McAuliffe, et al., "Dual-view plane illumination microscopy for rapid and spatially isotropic imaging," *Nature protocols*, vol. 9, no. 11, pp. 2555–2573, 2014.

[7] Stephan Preibisch, Fernando Amat, Evangelia Stamataki, Mihail Sarov, Robert H Singer, Eugene Myers, and Pavel Tomancak, "Efficient bayesian-based multiview deconvolution," *Nature methods*, vol. 11, no. 6, pp. 645–648, 2014.

[8] Maja Temerinac-Ott, Olaf Ronneberger, Peter Ochs, Wolfgang Driever, Thomas Brox, and Hans Burkhardt, "Multiview deblurring for 3-d images from light-sheet-based fluorescence microscopy," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1863–1873, 2011.

[9] Min Guo, Yue Li, Yijun Su, Talley Lambert, Damian Dalle Nogare, Mark W Moyle, Leighton H Duncan, Richard Ikegami, Anthony Santella, Ivan Rey-Suarez, et al., "Rapid image deconvolution and multiview fusion for optical microscopy," *Nature biotechnology*, vol. 38, no. 11, pp. 1337–1346, 2020.

[10] Hyoungjun Park, Myeongsu Na, Bumju Kim, Soohyun Park, Ki Hean Kim, Sunghoe Chang, and Jong Chul Ye, "Deep learning enables reference-free isotropic super-resolution for volumetric fluorescence microscopy," *Nature Communications*, vol. 13, no. 1, pp. 1–12, 2022.

[11] Yicong Wu, Xiaofei Han, Yijun Su, Melissa Glidewell, Jonathan S Daniels, Jiamin Liu, Titas Sengupta, Ivan Rey-Suarez, Robert Fischer, Akshay Patel, et al., "Multiview confocal super-resolution microscopy," *Nature*, vol. 600, no. 7888, pp. 279–284, 2021.

[12] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[13] "Huygens psf distiller," `https://svi.nl/Huygens-PSF-Distiller`, Accessed: 2022-08-24.

[14] Jacques Boutet de Monvel, Eric Scarfone, Sophie Le Calvez, and Mats Ulfendahl, "Image-adaptive deconvolution for three-dimensional deep biological imaging," *Biophysical journal*, vol. 85, no. 6, pp. 3991–4001, 2003.

[15] Klaus Becker, Saiedeh Saghafi, Marko Pende, Inna Sabdyusheva-Litschauer, Christian M Hahn, Massih Foroughipour, Nina Jährling, and Hans-Ulrich Dodt, "Deconvolution of light sheet microscopy recordings," *Scientific reports*, vol. 9, no. 1, pp. 1–14, 2019.

[16] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[17] Jan Huisken, Jim Swoger, Filippo Del Bene, Joachim Wittbrodt, and Ernst HK Stelzer, "Optical sectioning deep inside live embryos by selective plane illumination microscopy," *Science*, vol. 305, no. 5686, pp. 1007–1009, 2004.

[18] Johannes Schindelin, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, Curtis Rueden, Stephan Saalfeld, Benjamin Schmid, et al., "Fiji: an open-source platform for biological-image analysis," *Nature methods*, vol. 9, no. 7, pp. 676–682, 2012.

[19] Romain Fernandez and Cédric Moisy, "Fijiyama: a registration tool for 3d multimodal time-lapse imaging," *Bioinformatics*, vol. 37, no. 10, pp. 1482–1484, 2021.

NERD: NEURAL FIELD-BASED DEMOSAICKING

Tomáš Kerepecký [1,2] , *Filip Šroubek* [1] , *Adam Novozámský* [1] , *Jan Flusser* [1]

[1] Institute of Information Theory and Automation, The Czech Academy of Sciences, Czechia
[2]Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Czechia

ABSTRACT

We introduce NeRD, a new demosaicking method for generating full-color images from Bayer patterns. Our approach leverages advancements in neural fields to perform demosaicking by representing an image as a coordinate-based neural network with sine activation functions. The inputs to the network are spatial coordinates and a low-resolution Bayer pattern, while the outputs are the corresponding RGB values. An encoder network, which is a blend of ResNet and U-net, enhances the implicit neural representation of the image to improve its quality and ensure spatial consistency through prior learning. Our experimental results demonstrate that NeRD outperforms traditional and state-of-the-art CNN-based methods and significantly closes the gap to transformer-based methods.

Index Terms— Demosaicking, neural field, implicit neural representation.

1. INTRODUCTION

Raw data acquired by modern digital camera sensors is subject to various types of signal degradation, one of the most severe being the color filter array. To convert the raw data (Fig. 1a) into an image suitable for human visual perception (Fig. 1b), a demosaicking procedure is necessary [1].

Two main categories of image demosaicking exist: model-based and learning-based methods. Model-based methods, such as bilinear interpolation, Malvar [2], or Menon [3], are still widely used, but they fail to match the performance of recent deep learning-based approaches using deep convolutional networks (CNN) [4, 5, 6] or Swin Transformers [7].

Recently, Transformer networks have seen remarkable success in computer vision tasks and have become a state-of-the-art approach in demosaicking. However, a new paradigm in deep learning, Neural Fields (NF) [8], is gaining attention due to its comparable or superior performance in several computer vision tasks [8, 9, 10, 11, 12, 13, 14]. The basic idea behind NF is to represent data as the weights of a Multilayer Perceptron (MLP), known as implicit neural representation.

This work was supported in part by the Czech Science Foundation grant GA21-03921S, the *Praemium Academiae* awarded by the Czech Academy of Sciences, and the Fulbright commission under the Fulbright-Masaryk award.

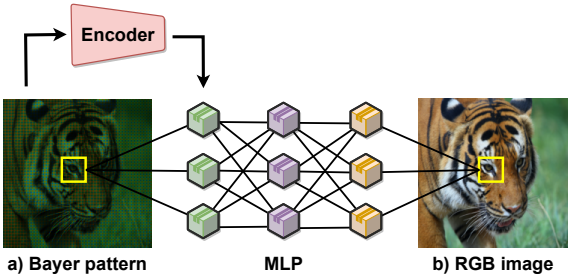


Fig. 1. An illustration of demosaicking using coordinate-based Multilayer Perceptron and local encoding technique.

NF has been applied in various domains and applications including Neural Radiance Fields (NeRF) [9] which achieved state-of-the-art results in representing complex 3D scenes. NeRV [11] encodes entire videos in neural networks. The Local Implicit Image Function (LIIF) [12] represents an image as a neural field capable of extrapolating to 30 times higher resolution. SIREN [13] uses a sinusoidal neural representation and demonstrates superiority over classical ReLU MLP in representing complex natural signals such as images.

Prior information from training data can be encoded into neural representation through conditioning (local or global) using methods such as concatenation, modulation of activation functions [15], or hypernetworks [14]. For example, CURE [10], a state-of-the-art method for video interpolation based on NF, uses an encoder to impose space-time consistency using local feature codes.

NF has also been used in image-to-image translation tasks such as superresolution, denoising, inpainting, and generative modeling [8]. However, to the best of our knowledge, no NF method has been proposed for demosaicking.

In this paper, we present NeRD, a novel approach for image demosaicking based on NF. The proposed method employs a joint ResNet and U-Net architecture to extract prior information from high-resolution ground-truth images and their corresponding Bayer patterns. This information is then used to condition the MLP using local feature encodings. The proposed approach offers a unique and innovative solution for image demosaicking.

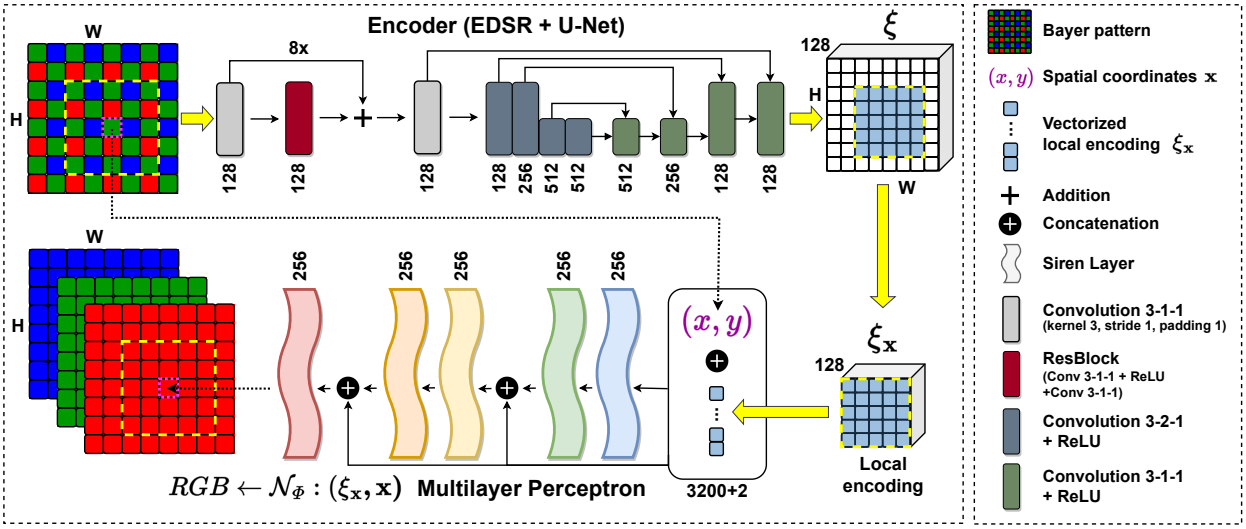


Fig. 2. The overall architecture of NeRD. Encoder consisting of 8 residual blocks and U-net architecture generates encoding ξ for the input Bayer pattern. Numbers below each layer in the encoder represent the number of output channels. Spatial coordinates $\mathbf{x} = (x, y)$ concatenated with the corresponding local encoding vector $\xi_{\mathbf{x}}$ are transformed into RGB value using a multilayer perceptron with 5 hidden layers each with 256 output channels, siren activation functions, and two skip connections.

2. PROPOSED METHOD

NeRD converts spatial coordinates and local encodings into RGB values. The local encodings are generated by an encoder that integrates consistency priors in NeRD. The overall architecture of NeRD is depicted in Fig. 2.

The core of NeRD is a fully connected feedforward network $\mathcal{N}_\Phi : (\xi_{\mathbf{x}}, \mathbf{x}) \rightarrow \mathbf{n}$ with 5 hidden layers, each with 256 output channels and sine activation functions. Φ denotes the network weights. The input is a spatial coordinate $\mathbf{x} = (x, y) \in \mathbb{R}^2$ and local encoding vector $\xi_{\mathbf{x}}$. The output is a single RGB value $\mathbf{n} = (r, g, b) \in \mathbb{R}^3$. The SIREN architecture [13] was chosen for its ability to model signals with greater precision compared to MLPs with ReLU. There are two skip connections that concatenate the input vector with the output of the second and fourth hidden layers.

Using the MLP without local encoding $\xi_{\mathbf{x}}$ leads to suboptimal demosaicking results due to the insufficient information contained in the training image. This is demonstrated by the result in Fig. 3-NeRD.0, where the reconstructed image is the output of the SIREN model trained only on original input Bayer pattern in self-supervised manner. The lack of spatial consistency in these results highlights the need for additional prior information in the form of spatial encoding, which is why we utilize an encoder.

The encoder provides local feature codes $\xi_{\mathbf{x}}$ for a given coordinate $\mathbf{x}$ and its architecture is shown in the first row of Fig. 2. The Bayer pattern is processed through a combined network that incorporates 8 residual blocks (using the EDSR architecture [16]) and 4 downsampling and 4 upsampling layers (U-Net architecture [17]) connected by multiple skip connections. The result is a global feature encoding $H \times W \times 128$, where H and W denote the height and width of the initial Bayer pattern in pixels. The local encoding $\xi_{\mathbf{x}}$ is extracted from the global encoding as a 5×5 region centered at $\mathbf{x}$, which is then flattened into a 3200-dimensional feature vector. The architecture of the encoder is adopted from [10].

The final RGB image is produced by independently retrieving the RGB pixel values from NeRD at the coordinates specified by the input Bayer pattern.

3. EXPERIMENT

We numerically validated NeRD on standard image datasets. Experiments also include an ablation study highlighting the key components of the proposed architecture and comparisons with state-of-the-art methods.

3.1. Dataset and Evaluation Metrics

A training set was created by combining multiple high-resolution datasets, such as DIV2K [18], Flickr2K [16], and OST [19], resulting in a total of 12 000 images. During each epoch, 10 000 randomly cropped patches of size 200×200 and corresponding Bayer patterns (GBRG) were generated. The Kodak and McM [20] datasets were used for testing.

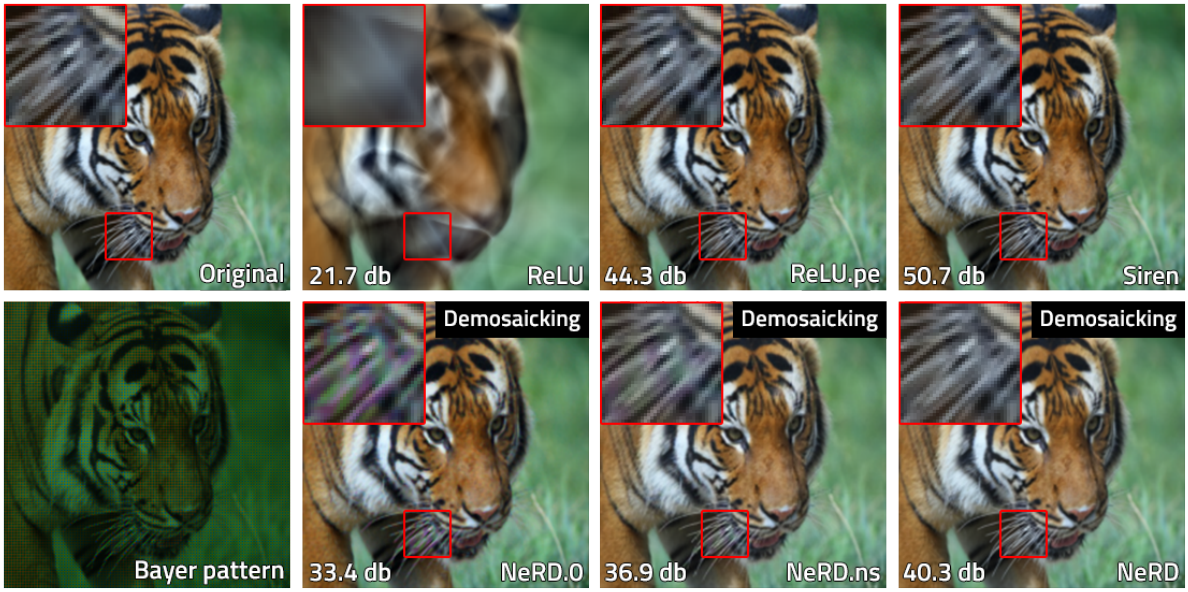


Fig. 3. The ablation study of NeRD. The original image is from DIV2K dataset. "ReLU" and "Siren" models show the implicit neural representation of the original image using MLP with ReLU and sine activation functions, respectively. These models were trained in a self-supervised manner to fit the original image. "ReLU.pe" stands for "ReLU" model with additional positional encoding in the form of Fourier feature mapping. "NeRD.0" model is identical to "Siren" model but is only trained using the input Bayer pattern. "NeRD" is the proposed demosaicking method, while "NeRD.ns" represents the proposed architecture without skip connections in the MLP. Each image is labeled with its PSNR value with respect to the original image.

The evaluation was performed using Peak Signal to Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM).

3.2. Training Configuration

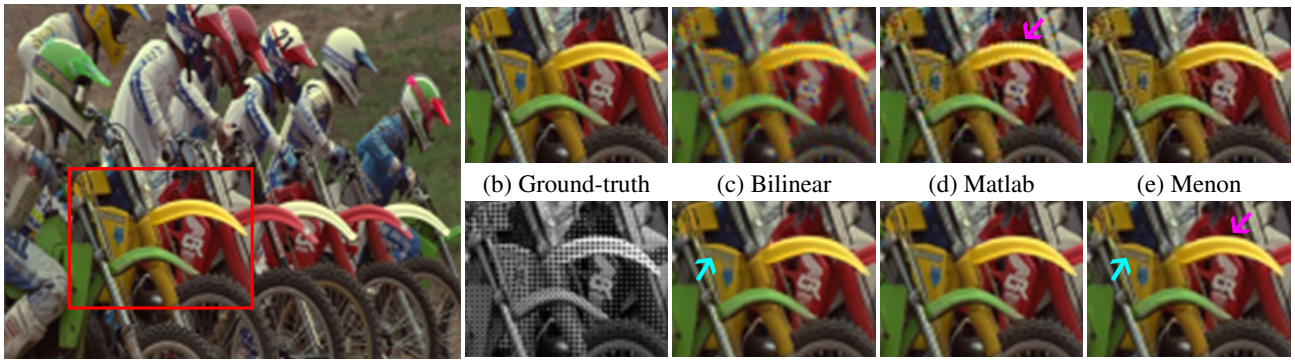
The training was conducted using an Nvidia L40s GPU. All INR models were optimized using the Mean Squared Error loss function, and the Adam optimizer was used with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The initial learning rate was set to 0.0009, and a step decay was applied, reducing the learning rate by 0.95 every epoch consisting of 10 000 iterations. The patch size was set to 200×200 and the batch size was 5.

3.3. Ablation Study

MLP and activation functions. RGB images can be represented as the weights of a fully connected feedforward neural network. This representation is achieved by training an MLP in a self-supervised manner to fit the original image. However, the usage of standard ReLU activation functions in MLPs produces unsatisfactory results, as shown in Fig. 3-ReLU. To significantly improve reconstruction, Fourier feature mapping of input spatial coordinates can be used (see Fig. 3-ReLU.pe). This technique is referred to as "positional encoding". Nonetheless, an even better outcome can be achieved by replacing ReLU with sine functions, also known as SIRENs. They demonstrate the capability of MLPs as image decoders and hold promise for demosaicking applications. SIREN architecture has the capacity to model RGB images with great precision. As demonstrated in Fig. 3-Siren, the SIREN with 5 hidden layers, each with 256 neurons, achieved a PSNR of 50.7 dB when trained for just 1000 iterations to fit the original image.

Encoder. The naive approach of decoding RGB images from Bayer patterns using SIREN architecture fails as it loses two-thirds of the original information, as shown in Fig. 3-NeRD.0. To improve the demosaicking capability of the MLP, prior information must be incorporated through an encoder. This encoder learns prior information across various training image pairs and conditions the MLP with local encodings. The effectiveness of the encoder is demonstrated in Fig. 3-NeRD, which shows the results of demosaicking using the NeRD architecture described in Sec. 2.

Skip Connections. The integration of encoding into the MLP can be achieved through various methods. However, methods such as modulation of activation functions or the use of hypernetworks present challenges in terms of parallelization. Hence, we utilized a method of concatenation, where the



(a) Kodak:kodim05.png (cropped) (b) Ground-truth (c) Bilinear (d) Matlab (e) Menon (f) Raw (g) DeepDemosaick (h) RSTCANet (i) NeRD

Fig. 4. A visual comparison of NeRD and the current state-of-the-art methods on an example from the Kodak dataset. The visual differences are highlighted by close-ups, which correspond to the red box in the original image. Although NeRD exhibits slightly inferior visual performance compared to RSTCANet, it outperforms traditional methods in terms of reconstruction accuracy (indicated by the magenta arrow) and avoids over-smoothing details, as seen with the DeepDemosaick method (indicated by the cyan arrow).

coordinates and feature vectors are combined at the input and later concatenation of the input with the second and fourth hidden layers is performed using skip connections. The significance of incorporating skip connections into the MLP is illustrated in Fig. 3-NeRD.ns (no-skip). This figure demonstrates a degradation in both the quality of the reconstruction and the PSNR value when these connections are omitted.

3.4. Comparison With Existing Methods

The evaluation of the proposed NeRD demosaicking algorithm was performed on the McM and Kodak datasets, which were resized and cropped to 200×200 px. A comparison of NeRD with traditional demosaicking algorithms and state-of-the-art methods is presented in Table 1 in terms of average PSNR and SSIM values calculated from the demosaicked images. The results show that NeRD outperforms traditional methods and the CNN-based DeepDemosaick [4], but falls slightly behind the transformer-based RSTCANet [7].

Table 1. Average PSNR/SSIM obtained by NeRD and the current state-of-the-art methods on the McM* and Kodak* datasets (*resized and cropped to 200×200 px). **Bold** and underline highlights the highest and second highest values, respectively. Note the superior results of NeRD over the CNN-based and traditional methods. Only RSTCANet, which is based on transformers, has slightly higher scores.

Method	McM* [20] PSNR/SSIM	Kodak* PSNR/SSIM
Bilinear	27.15/0.912	28.01/0.894
Matlab (Malvar) [2]	30.54/0.923	33.52/0.957
Menon [3]	31.40/0.918	35.20/0.968
DeepDemosaick [4]	33.31/0.942	37.76/0.976
RSTCANet [7]	**37.77/0.978**	**40.84/0.988**
NeRD	36.18/0.969	39.07/0.984

A visual comparison of the demosaicked images is presented in Fig. 4. The figure highlights differences between NeRD and the other methods and provides insights into their performance. One notable characteristic of NeRD is that it avoids over-smoothing details, unlike the DeepDemosaick [4] method, as indicated by the cyan arrow in the Fig. 4g. Furthermore, NeRD outperforms traditional methods in terms of preserving fine details and avoiding unpleasant artifacts, as indicated by the magenta arrow in the Fig. 4d.

4. CONCLUSION

This paper presents a novel demosaicking algorithm, NeRD, that leverages the recent class of techniques known as Neural Fields. The ablation study results emphasize the significance of incorporating an encoder and skip connections within the MLP, which results in significant improvement over traditional techniques and outperforms the CNN-based DeepDemosaick method in preserving fine details while avoiding undesirable artifacts. Although NeRD shows slightly lower visual performance compared to the transformer-based RSTCANet, it still demonstrates remarkable accuracy in terms of reconstruction. Future research can focus on enhancing NeRD through fine-tuning using input Bayer pattern-specific loss functions and integrating Transformer networks or ConvNeXt into the encoder. In addition, expanding the training set by more diverse datasets can improve the prior. Albeit NeRD may not attain the performance level of Transformer-based demosaicking, our contribution broadens the range of domains where Neural Fields can be applied.

5. REFERENCES

[1] Daniele Menon and Giancarlo Calvagno, "Color image demosaicking: An overview," *Signal Processing: Image Communication*, vol. 26, no. 8-9, pp. 518–533, 2011.

[2] Henrique S Malvar, Li-wei He, and Ross Cutler, "High-quality linear interpolation for demosaicing of bayer-patterned color images," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 2004, vol. 3, pp. iii–485.

[3] Daniele Menon, Stefano Andriani, and Giancarlo Calvagno, "Demosaicing with directional filtering and a posteriori decision," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 132–141, 2006.

[4] Filippos Kokkinos and Stamatios Lefkimmiatis, "Iterative joint image demosaicking and denoising using a residual denoising network," *IEEE Transactions on Image Processing*, vol. 28, no. 8, pp. 4177–4188, 2019.

[5] Michaël Gharbi, Gaurav Chaurasia, Sylvain Paris, and Frédo Durand, "Deep joint demosaicking and denoising," *ACM Transactions on Graphics (ToG)*, vol. 35, no. 6, pp. 1–12, 2016.

[6] Tomáš Kerepecky and Filip Šroubek, "D3net: Joint demosaicking, deblurring and deringing," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 1–8.

[7] Wenzhu Xing and Karen Egiazarian, "Residual swin transformer channel attention network for image demosaicing," in *2022 10th European Workshop on Visual Information Processing (EUVIP)*. IEEE, 2022, pp. 1–6.

[8] Yiheng Xie, Towaki Takikawa, Shunsuke Saito, Or Litany, Shiqin Yan, Numair Khan, Federico Tombari, James Tompkin, Vincent Sitzmann, and Srinath Sridhar, "Neural fields in visual computing and beyond," in *Computer Graphics Forum*. Wiley Online Library, 2022, vol. 41, pp. 641–676.

[9] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.

[10] Wentao Shangguan, Yu Sun, Weijie Gan, and Ulugbek S Kamilov, "Learning cross-video neural representations for high-quality frame interpolation," in *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XV*. Springer, 2022, pp. 511–528.

[11] Hao Chen, Bo He, Hanyu Wang, Yixuan Ren, Ser Nam Lim, and Abhinav Shrivastava, "Nerv: Neural representations for videos," *Advances in Neural Information Processing Systems*, vol. 34, pp. 21557–21568, 2021.

[12] Yinbo Chen, Sifei Liu, and Xiaolong Wang, "Learning continuous image representation with local implicit image function," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 8628–8638.

[13] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein, "Implicit neural representations with periodic activation functions," *Advances in Neural Information Processing Systems*, vol. 33, pp. 7462–7473, 2020.

[14] Ivan Skorokhodov, Savva Ignatyev, and Mohamed Elhoseiny, "Adversarial generation of continuous images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 10753–10764.

[15] Emilien Dupont, Hyunjik Kim, SM Eslami, Danilo Rezende, and Dan Rosenbaum, "From data to functa: Your data point is a function and you should treat it like one," *arXiv preprint arXiv:2201.12204*, 2022.

[16] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 136–144.

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[18] Eirikur Agustsson and Radu Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.

[19] Chao Dong Xintao Wang, Ke Yu and Chen Change Loy, "Recovering realistic texture in image super-resolution by deep spatial feature transform," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

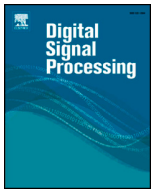
[20] Lei Zhang, Xiaolin Wu, Antoni Buades, and Xin Li, "Color demosaicking by local directional interpolation and nonlocal adaptive thresholding," *Journal of Electronic imaging*, vol. 20, no. 2, pp. 023016–023016, 2011.

Digit. Signal Process. 159 (2025) 105022

Contents lists available at ScienceDirect

Digital Signal Processing

journal homepage: www.elsevier.com/locate/dsp



Implicit neural representation for image demosaicking

Tomáš Kerepecký [a,b,*], Filip Šroubek [a], Jan Flusser [a]

[a] Institute of Information Theory and Automation, The Czech Academy of Sciences, Pod Vodárenskou věží 4, Prague, CZ-18200, Czechia
[b] Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Břehová 78/7, Prague, CZ-11519, Czechia

ARTICLE INFO

Keywords:
Demosaicking
Implicit neural representation
Inverse problems

ABSTRACT

We propose a novel approach to enhance image demosaicking algorithms using implicit neural representations (INR). Our method employs a multi-layer perceptron to encode RGB images, combining original Bayer measurements with an initial estimate from existing demosaicking methods to achieve superior reconstructions. A key innovation is the integration of two loss functions: a Bayer loss for fidelity to sensor data and a complementary loss that regularizes reconstruction using interpolated data from the initial estimate. This combination, along with INR's inherent ability to capture fine details, enables high-fidelity reconstructions that incorporate information from both sources. Furthermore, we demonstrate that INR can effectively correct artifacts in state-of-the-art demosaicking methods when input data diverge from the training distribution, such as in cases of noise or blur. This adaptability highlights the transformative potential of INR-based demosaicking, offering a robust solution to this challenging problem.

1. Introduction

Digital camera sensors typically capture raw image data through a Color Filter Array (CFA), resulting in sub-sampled color information that requires reconstruction through a process known as demosaicking. Traditional demosaicking algorithms, such as bilinear interpolation, Malvar [1], and Menon [2], offer computational efficiency but are prone to artifacts like color Moiré, zippering, and false color patterns. These artifacts degrade image quality by introducing undesirable visual effects. Moiré patterns appear as repetitive interference patterns in areas with high-frequency textures. Zippering manifests as jagged edges along sharp transitions. False color patterns distort natural color representation, often due to processing errors during demosaicking.

More advanced approaches have aimed to mitigate these issues by integrating demosaicking with other image processing tasks. For instance, joint demosaicking and denoising or deblurring methods [3–8] employ model-based optimization techniques to achieve better reconstruction quality.

Recent advancements in deep learning have significantly enhanced the performance of demosaicking algorithms [9–15]. These techniques have set new benchmarks by leveraging Convolutional Neural Networks (CNNs) or Transformers to reduce artifacts and improve the fidelity of reconstructed images. However, these methods often struggle when faced with input data that diverge from their training distribution, such as images affected by blur, common in both DSLR and mobile phone cameras (Fig. 1), even when the lens is in focus.

In response to these challenges, we propose a novel deep learning-based approach named INRID (Implicit Neural Representation for Image Demosaicking), which leverages Implicit Neural Representations (INR) [16] to enhance image reconstruction in both traditional and state-of-the-art demosaicking methods. By representing each individual image through the weights of a Multilayer Perceptron (MLP), our approach provides a more flexible and powerful reconstruction.

INRID reconstructs the image by adapting to the specific characteristics of two key inputs: the raw Bayer data and the initial demosaicked image from methods such as Malvar or Menon. A Bayer loss function enforces fidelity to the original raw sensor data, minimizing the mean squared error (MSE) between the reconstructed Bayer pattern and the raw measurements. Simultaneously, the complementary pixel values — those missing in the Bayer pattern — are reconstructed by aligning them with the initial estimate while ensuring consistency with the raw Bayer data. This combined process enables INRID to capture fine image details and correct residual artifacts, that traditional methods often leave unaddressed.

For state-of-the-art deep learning methods, INRID extends beyond refinement to address out-of-distribution scenarios, such as blurred or noisy inputs. By incorporating the forward degradation process—e.g., simulating blur or noise—directly into the optimization, INRID aligns

* Corresponding author.
E-mail address: kerepecky@utia.cas.cz (T. Kerepecký).

https://doi.org/10.1016/j.dsp.2025.105022

Available online 23 January 2025

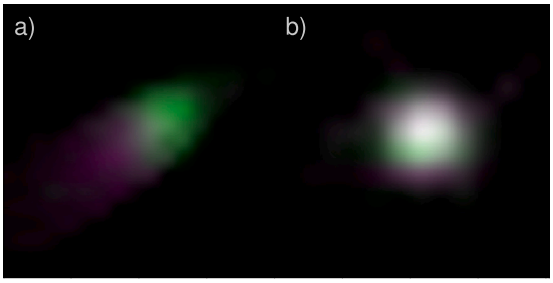


Fig. 1. Intrinsic camera blur (a combination of sensor blur and lens aberrations, present even when the lens is in focus): a) DSLR, b) mobile phone. These intrinsic blur kernels are about 7x7 pixels in size for 16 MPx images. Interpolation was used to magnify the blur kernels for visual presentation.

the reconstruction with both the degraded Bayer data and the initial estimate. This approach ensures robust adaptation to challenging conditions, recovering high-frequency details and reducing artifacts. As a result, INRID significantly enhances demosaicking performance, even when the input data diverge from the training distribution.

The rest of the paper is organized as follows: Section 2 reviews related work. Section 3 describes the proposed methodology, including the inverse problem and definition of loss functions used for training. Section 4 presents experimental results that demonstrate our approach in enhancing existing methods. Section 5 discusses the limitations of our work and potential future directions for improvement. Finally, Section 6 offers concluding remarks.

2. Related work

2.1. Image demosaicking

Traditional demosaicking methods have predominantly relied on interpolation techniques, which, despite their computational efficiency, are prone to introducing artifacts, especially in regions with high-frequency content. Early methods, such as bilinear interpolation, provided a simple yet effective approach for reconstructing missing color information [17]. The work by Malvar et al. [1] improved upon these techniques by introducing a gradient-corrected bilinear interpolation method, optimized using a Wiener filtering approach, which aimed to reduce the visibility of common artifacts. Menon et al. [2] further advanced the field by incorporating directional filtering and a posteriori decision-making, which improved edge preservation and reduced color artifacts. However, these methods struggled with handling complex textures and often produced noticeable artifacts, such as color Moiré patterns and zipper effects.

Optimization-based methods tackle demosaicking by formulating it as an inverse problem and integrating regularization terms to enhance reconstruction quality. For instance, the multiframe demosaicking and super-resolution method by Farsiu et al. [18] applies a maximum a posteriori (MAP) estimation framework. This approach effectively reduces artifacts and addresses degradations such as noise and blur, while requiring increased computational complexity compared to interpolation-based methods.

The advent of deep learning has led to significant advancements in demosaicking. One notable approach, commonly named DeepDemosaick, is the method proposed by Kokkinos and Lefkimmiatis [10], which introduces a deep convolutional residual network designed to jointly perform demosaicking and denoising. This approach leverages a cascade of convolutional layers to model the underlying patterns in raw sensor data and predict a high-quality full-resolution RGB image. The network is inspired by optimization strategies from classical image regularization methods and is trained end-to-end on a dataset of mosaicked and ground-truth images. This design enables the model to capture complex pixel-level dependencies, resulting in superior color reconstruction and reduced artifacts compared to previous methods.

Another state-of-the-art method, RSTCANet [11], currently a leading method in the field, builds upon the Swin Transformer framework with the introduction of Residual Swin Transformer Channel Attention Blocks. This advanced design captures both spatial and channel-wise dependencies more effectively, thanks to its hierarchical structure and shifted windows, while the residual connections allow for deeper network architectures by mitigating the vanishing gradient problem. RSTCANet excels in preserving fine details and handling complex textures, delivering high-quality demosaicking results across various datasets.

These deep learning-based methods, including RSTCANet and DeepDemosaick, are pre-trained on large datasets to learn a mapping from mosaiced inputs to full-color images. While effective on images similar to the training data, their reliance on pre-training limits their ability to generalize to out-of-distribution data, such as images with blur or noise not represented in the training set. Pre-trained models cannot easily adapt to variations not seen during training, which can lead to suboptimal performance in challenging scenarios.

In contrast, our hybrid approach employs an INR that is optimized individually for each input image. Instead of relying on pre-trained weights, we solve an optimization problem over the network parameters specific to each image, rather than over pixel values as in traditional methods. This per-image optimization allows our model to adapt to the unique characteristics of each image, providing robustness to out-of-distribution data such as noisy or blurred images. By optimizing over network parameters, our method can capture fine image details and correct artifacts more effectively.

2.2. Implicit neural representation

INRs have emerged as a powerful tool in computer vision, representing images and 3D shapes continuously through fully connected feed-forward networks. Early work, such as DeepSDF [19], showcased the effectiveness of ReLU-based MLPs for shape representation.

For images, INR maps spatial coordinates to RGB values using an MLP, enabling continuous image representation, unlike conventional pixel grids. This approach allows high-quality reconstructions, even from sparse or incomplete data.

However, ReLU-based networks, while foundational, struggle to capture fine details, particularly high-frequency information, due to their piecewise linear structure.

To address this limitation, Fourier feature mapping, also known as positional encoding, was introduced [20]. This technique involves mapping the input spatial coordinates into a higher-dimensional space using sinusoidal functions, which helps the MLP capture finer details and improves the reconstruction quality. This approach was popularized by works such as NeRF (Neural Radiance Fields) [21], where it was used to represent 3D scenes with high fidelity.

Building on these advancements, SIREN (Sinusoidal Representation Networks) [22] was introduced, which replaced ReLU with sine activation functions. SIRENs demonstrated the ability to model high-frequency details with greater precision, as sine functions naturally encode oscillatory patterns that are prevalent in image data. This architecture significantly improved the performance of MLPs as image decoders, enabling them to achieve state-of-the-art results in various tasks, including image superresolution and inpainting.

Recently, WIRE (Wavelet Implicit Representations) [23] has pushed the boundaries of INR even further by introducing wavelet-based activation functions. WIRE leverages the multi-resolution properties of wavelets, allowing the MLP to model both coarse and fine details simultaneously.

INCODE [24] further advances INR by introducing a harmonizer network that dynamically adjusts the activation functions based on prior knowledge. This innovation allows INCODE to adaptively fine-tune key parameters like amplitude and frequency of sinusoidal activation func-

tions, enabling the MLP to better capture details and broader signal patterns.

Ramasinghe and Lucey [25] proposed additional activation functions such as Gaussian, Laplacian, and so-called Quadratic to broaden the family of INRs, offering alternatives for capturing fine details without relying on periodic functions.

However, for our demosaicking approach, SIREN and INCODE remain particularly promising due to their sinusoidal activation functions, which are well-suited for interpolating missing data and capturing the complex signal patterns required in this problem.

2.3. Implicit neural representation for image demosaicking

In our previous work, Neural field-based Demosaicking (NERD) [26], we extended the application of INR to the domain of demosaicking. NERD introduced a method that combined ResNet [27] and U-Net [28] architectures to condition the MLP using high-resolution image features extracted from ground-truth images and their corresponding Bayer patterns. This approach demonstrated the potential of INR in handling the challenging task of demosaicking by leveraging the strengths of coordinate based neural networks.

Compared to NERD, the approach presented in this paper significantly reduces computational complexity by eliminating the encoder component while also leveraging the strengths of existing demosaicking methods. Rather than merely introducing a new demosaicking technique, the proposed hybrid framework is designed to substantially enhance reconstruction capabilities and improve the robustness of both traditional and state-of-the-art methods.

3. Problem formulation

In the context of digital image processing, the forward problem involves modeling the degradation process that occurs during image acquisition with a digital camera. This process encompasses blurring due to the camera optical system, subsampling caused by the CFA, commonly implemented as a Bayer pattern, and noise introduced by the sensor. The forward model for Bayer measurement b is expressed as:

$$b = S_B H u + n_B \tag{1}$$

where $u \in \mathbb{R}^M$ represents the vectorized form of the unknown high-resolution sharp image, $H(\cdot) \equiv h * \cdot$ denotes the channel-dependent blurring operator, where h is the Point Spread Function (PSF) estimated from calibration data and $*$ indicates convolution. $n_B \approx \mathcal{N}(0, \sigma_B^2)$ represents additive white Gaussian noise with zero mean and variance σ_B^2, and S_B is the down-sampling operator corresponding to the Bayer pattern (e.g. RGGB), resulting in the observed mosaiced image $b \in \mathbb{R}^P$, where $M = 3P$.

Additionally, for the complementary pixel values, we can hypothesize a forward model:

$$c = S_C H u + n_C \tag{2}$$

where S_C is the down-sampling operator corresponding to the remaining 2/3 of the original pixel values that are complementary to the Bayer pattern (therefore $c \in \mathbb{R}^{2P}$). The term $n_C \approx \mathcal{N}(0, \sigma_C^2)$, with variance σ_C^2, represents additive noise associated with these complementary pixels.

3.1. Inverse problem

The inverse problem seeks to reconstruct the high-resolution image u from a degraded observation b. Our approach incorporates not only the forward model for the Bayer measurement b (Equation (1)) but also a second forward model for the complementary pixel values c (Equation (2)). Since c is not directly available, we estimate a rough reconstruction $u_0 = D(b)$ using an initial demosaicking method D. From this reconstruction, the complementary pixel values are approximated as $c \approx S_C u_0$. The inverse problem is inherently ill-posed due to the combined effects of blur, noise, and incomplete color information, requiring a robust optimization strategy.

In our framework, the inverse problem is formulated as training an INR, u_ψ, to reconstruct the high-resolution image u by parameterizing it as a continuous function modeled by the weights ψ of an MLP. Optimization of parameters ψ ensures that the outputs of u_ψ, when passed through the degradation models, match both the observed Bayer measurement b and the complementary pixel estimates c. Furthermore, added regularization promotes smoothness and edge preservation. This optimization is carried out for each individual image using stochastic gradient descent or its variants, with backpropagation applied to minimize loss functions derived from the forward models of b and c.

Formally, the optimization problem is expressed as:

$$\hat{\psi} = \arg\min_{\psi} \Big\{ \alpha \mathcal{L}_{\text{Bayer}}\Big(\hat{b}, b\Big) + \beta \mathcal{L}_{\text{Demo}}\Big(\hat{c}, S_C u_0\Big) + \gamma \mathcal{R}\Big(u_\psi\Big) \Big\}, \tag{3}$$

where $u_{\hat{\psi}}$ is the final reconstruction. $\hat{b} = S_B H u_\psi$ represents the predicted INR that is subject to the given degradation and corresponds to the Bayer pattern. To perform the degradation we sample u_ψ at all pixel locations and consider its vectorized form. The Bayer loss $\mathcal{L}_{\text{Bayer}}(\hat{b}, b)$ ensures fidelity to the original sensor data. Additionally, $\hat{c} = S_C H u_\psi$ denotes the degraded INR at complementary pixel locations, which lack direct Bayer measurements. The complementary loss, $\mathcal{L}_{\text{Demo}}$, minimizes the error between $\hat{c}$ and the corresponding values in the initial demosaiced image u_0. The overall optimization is balanced by the weighting factors α, β, and γ, which control the contributions of the Bayer loss, complementary loss, and the Total Variation (TV) regularization $\mathcal{R}(u_\psi)$.

In our ablation study for selecting optimal weighting factors (Section 4.4), β is fixed at 1 while α is varied to balance the Bayer and complementary losses. The parameter γ, when set to values between 10^{-6} and 10^{-5}, is used specifically for joint demosaicking, and deblurring tasks, as described in Section 4.6. The specific values of these weighting factors are further detailed in the experimental section.

3.2. Bayer loss

The Bayer loss $\mathcal{L}_{\text{Bayer}}$ is defined as the MSE between the predicted Bayer image and the observed (inherently blurred) mosaiced image b:

$$\mathcal{L}_{\text{Bayer}}\Big(\hat{b}, b\Big) = \frac{1}{P} \| S_B H u_\psi - b \|_2^2. \tag{4}$$

3.3. Complementary loss

The complementary loss $\mathcal{L}_{\text{Demo}}$ is calculated as the MSE between the predicted complementary pixel values and the corresponding values in the initial demosaiced (inherently blurred) image u_0:

$$\mathcal{L}_{\text{Demo}}\Big(\hat{c}, S_C u_0\Big) = \frac{1}{2P} \| S_C H u_\psi - S_C u_0 \|_2^2. \tag{5}$$

3.4. Total variation regularization

We apply Color TV regularization $\mathcal{R}(u_\psi)$ to ensure smoothness while preserving edges [29]. Total variation measures the gradient magnitude across the image, penalizing rapid intensity changes to reduce noise and retain key features. In INR models, the continuous image representation allows gradient computation at any point using automatic differentiation, enabling efficient total variation minimization. In our framework, TV regularization proves especially beneficial for tasks such as joint demosaicking and deblurring, where it not only stabilizes the reconstruction process but also helps preserve important image details, making it particularly impactful for processing real-world images.

Table 1
Configurations for INR Models.

Parameter	Gauss	ReLU	FFN	SIREN	WIRE	INCODE
Activation	Gaussian	ReLU	ReLU	Sine	Wavelet	Sine
Modulation	—	—	Positional Encoding (Gaussian)	—	—	Harmonizer (ResNet34)
Hidden Layers	5	5	5	5	5	5
Neurons per Layer	256	256	256	256	256	256
Learning Rate	1×10^{-4}	1×10^{-4}	1×10^{-4}	1×10^{-4}	7×10^{-4}	1×10^{-4}
Batch Size	128×128	128×128	128×128	128×128	128×128	128×128
Trainable Parameters	330499	330499	461059	330499	330499	568359
Special Parameters	—	—	—	$\omega_{\text{first}} = 30$, $\omega_{\text{hidden}} = 30$	$\omega = 30$, $\sigma = 10$	$a = 0.1993$, $b = 0.0196$, $c = 0.0588$, $d = 0.0269$

Table 2
Image Reconstruction With INR: Average PSNR values for image representation using different INR models on the Kodak dataset, across various image sizes and training iterations. **Bold** and underline highlight the highest and second highest values, respectively.

INR Model	Original size (768×512)			1/2 Resize (384×256)			1/4 Resize (192×128)		
	500 Iter	1000 Iter	2000 Iter	500 Iter	1000 Iter	2000 Iter	500 Iter	1000 Iter	2000 Iter
Gauss	31.01	33.68	35.99	34.26	38.97	43.25	60.24	**90.23**	80.24
ReLU	21.92	22.52	23.07	21.79	22.82	23.61	21.60	23.22	24.60
SIREN	37.89	40.29	41.92	39.62	46.46	49.88	44.50	51.73	60.52
WIRE	37.32	40.51	42.65	41.71	43.04	46.79	56.73	66.70	74.74
FFN	32.83	34.98	36.88	35.15	40.05	43.76	37.60	45.29	52.35
Incode	**39.65**	**41.35**	**42.87**	**52.78**	**50.21**	**51.93**	**72.23**	79.72	**90.94**

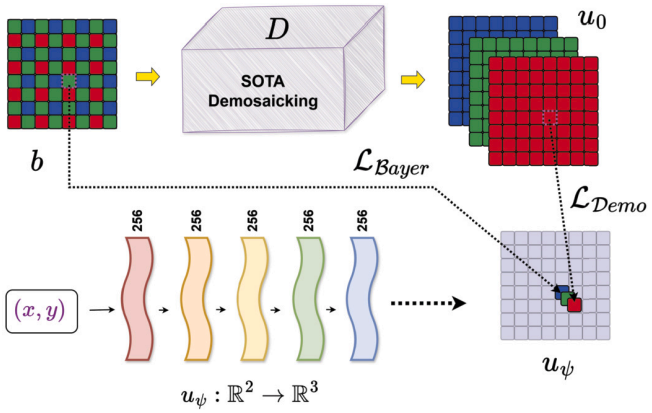


Fig. 2. Illustration of INRID: The proposed approach performs demosaicking using an implicit neural representation $u_\psi : \mathbb{R}^2 \to \mathbb{R}^3$, optimized by minimizing the mean squared error $\mathcal{L}_{Bayer}$ between the reconstruction u_ψ and the Bayer measurement b, as well as between the reconstruction u_ψ and the initial demosaicked image u_0 ($\mathcal{L}_{Demo}$). The INR consists of five layers, each with 256 neurons, and employs sinusoidal activations to effectively capture high-frequency image details. Unlike traditional activation functions such as ReLU or sigmoid, sinusoidal activations enable a more expressive representation, improving the reconstruction of fine structures and textures critical for accurate demosaicking (see ablation study in Section 4.3).

We call the algorithm that solves (3) INRID, standing for Implicit Neural Representation for Image Demosaicking. Fig. 2 provides a conceptual overview of the INRID framework. It highlights the key components: the raw Bayer measurement b, the initial demosaicked image u_0, and the learned implicit representation u_ψ. This high-level visualization is intended to help readers grasp the primary relationships and flow of the optimization process.

4. Experimental results

To solve the minimization in (3), we employ a self-supervised approach where the INR is trained directly on the degraded image data without requiring ground truth high-resolution images. This enables the INR model to reconstruct the high-resolution image solely based on the observed mosaiced image and complementary pixel information.

We begin by demonstrating image representation using INR and comparing various architectures. Next, we show that using the Bayer loss only for image representation exceeds basic demosaicking approaches such as nearest neighbor and bilinear interpolation, and in some cases outperforms traditional methods like Malvar and Menon. We then illustrate how the combination of Bayer and complementary loss within the INRID framework significantly improves reconstruction performance and exceeds all traditional methods. Furthermore, we showcase the joint demosaicking, denoising, or deblurring capabilities of INRID, enhancing state-of-the-art demosaicking methods such as DeepDemosaick and RSTCANet. Finally, we demonstrate the effectiveness of our approach on real-world data from mobile phone cameras.

4.1. Experimental setup

Table 1 summarizes the hyperparameter configurations for all INR models used in our experiments. Each model consists of five hidden layers with 256 neurons per layer. The Gauss and ReLU models employ Gaussian and ReLU activation functions, respectively, while the Fourier Feature Networks (FFN) utilize Gaussian positional encoding with ReLU activations. The SIREN model uses sine activation functions, parameterized by frequency terms ω for the first and hidden layers. The WIRE model incorporates a wavelet activation function, characterized by a frequency term ω and a scale term σ, which enable the balance of global and local signal representation. The INCODE model builds on a modified SIREN architecture, augmented with a harmonizer network based on the ResNet34 [27] backbone. The specific parameters, including frequency and scale terms for SIREN, WIRE, and INCODE, are summarized in Table 1 and detailed in their respective original works [22–24].

The training was conducted using an Nvidia L40s GPU. All models were optimized using the MSE loss function, and the Adam optimizer, with decay rates for gradient and squared gradient averages set to 0.9 and 0.999, respectively. A learning rate scheduler was applied to gradually reduce the learning rate during training. The initial learning rate was set to 0.0001 for most models, except for WIRE, which used a

Table 3
Image Reconstruction With INR: Average PSNR values for image representation using different INR models on the McM dataset (500×500 version), across various image sizes and training iterations. **Bold** and underline highlight the highest and second highest values, respectively.

INR Model	Original size (500×500)			1/2 Resize (250×250)			1/4 Resize (125×125)		
	500 Iter	1000 Iter	2000 Iter	500 Iter	1000 Iter	2000 Iter	500 Iter	1000 Iter	2000 Iter
Gauss	29.65	32.75	36.14	36.50	45.51	51.32	86.48	113.20	**117.79**
ReLU	20.81	21.67	22.47	19.82	21.14	22.31	18.61	20.49	22.34
SIREN	39.10	41.79	43.61	39.19	46.08	51.71	44.07	52.38	62.06
WIRE	37.28	41.05	44.17	46.10	49.58	47.83	53.00	58.83	65.72
FFN	35.16	37.67	39.47	35.89	41.24	46.02	32.70	43.02	52.30
Incode	**41.56**	**43.17**	**44.64**	**64.79**	**67.54**	**53.96**	**107.24**	**116.99**	117.33

learning rate of 0.0007. Batch sizes were fixed at 128×128 for all experiments.

To evaluate the performance of the INR models, we tested on the Kodak [30] and McMaster [31] datasets. The evaluation metrics included Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM).

The source code used in this study is publicly available at https://github.com/kereptom/inrid2024.

4.2. Image reconstruction with INR

In the image reconstruction experiment, we evaluated the performance of different INR architectures in representing images. In other words, we trained INR models in a self-supervised manner to fit the original image. This corresponds to setting $\alpha = 1$, $\beta = 2$, and $\gamma = 0$ in Equation (3), with the initial reconstruction u_0 replaced by the original ground truth pixel values u. We set $\beta = 2$ because the complementary loss involves twice as many pixels as the Bayer loss.

Specifically, we tested six different INR architectures across three image sizes and three different numbers of iterations, calculating average PSNR results for both the Kodak and McM datasets. The results, as shown in Tables 2 and 3, indicate that the INR model with ReLU activation consistently performed the worst across all conditions. On average, INCODE delivered the best results in nearly all scenarios. SIREN and WIRE were strong contenders, especially at the original size and half size. While both Gauss and FFN showed moderate results overall, FFN performed slightly better at the original size, whereas Gauss was particularly effective for smaller images. SIREN, although not always achieving the highest scores, produced stable and reliable results across different image sizes and iteration counts, making it a strong performer in a wide range of conditions.

The visual demonstration in Fig. 3 supports these findings, showing the reconstruction of Kodak image #23 at its original size after 2000 iterations. The ReLU INR model shows significant blurring, particularly in areas with fine textures. In contrast, the other methods produce visually pleasing and accurate reconstructions, with INCODE, WIRE, and SIREN standing out for their near-perfect results (see Fig. 3, especially in the close-ups).

We also analyzed the progression of PSNR values with extended training on the McM dataset (500 × 500 version) beyond 2000 iterations, as shown in Fig. 4. The results reveal continued improvement across all models, but with a diminishing rate of gain after 2000 iterations. Models such as INCODE, WIRE, and SIREN exhibit high performance and retain their advantage. Given this diminishing improvement, it becomes important to consider the trade-off between further enhancing reconstruction quality and the associated computational cost, which will be discussed further in Section 5.

Although WIRE showed competitive performance, we encountered instability with the learning rate, making its training less reliable compared to other models. Based on these results, we chose to proceed with two INR architectures for further experiments: INCODE, which dominated in most scenarios, and SIREN, which consistently performed well and demonstrated stability across various conditions; and also included FFN and Gauss for reference.

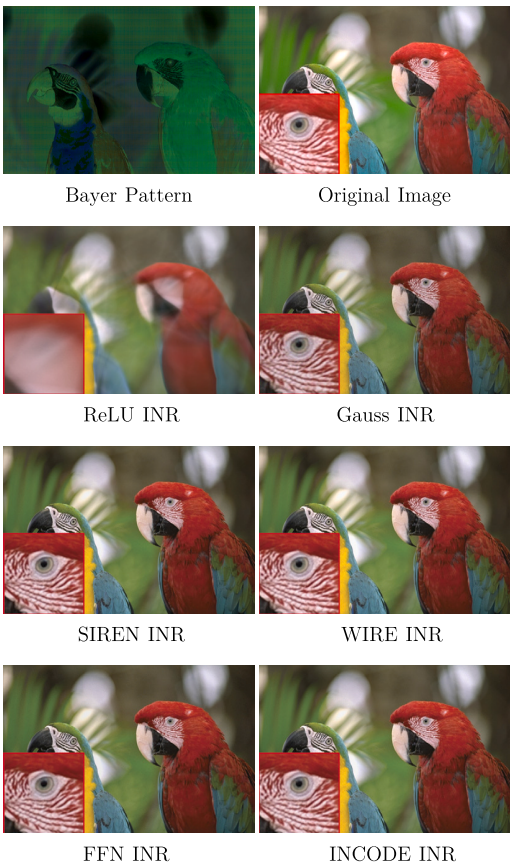


Fig. 3. Image reconstruction using different INR architectures on an example from the Kodak dataset. The Bayer Pattern (top left) shows the raw subsampled data for demonstration purposes. All INR models were trained in a self-supervised manner to fit the original image (top right). ReLU INR struggles to model high-frequency details, resulting in noticeable blurring, especially in regions with fine textures, such as the bird's feathers. In contrast, INCODE, SIREN and WIRE architectures provide the most visually pleasing reconstructions, capturing details with higher fidelity. This example illustrates the results after 2000 training iterations for each INR architecture. The corresponding average PSNR values for the entire dataset are reported in Table 2, 4th column.

4.3. Image demosaicking with INR

Following our image representation study, we extended our experiments to image demosaicking using INR architectures, focusing solely on Bayer measurements, which is equivalent to setting $\alpha = 1$, $\beta = 0$, and $\gamma = 0$ in Equation (3). The results, shown in Tables 4 and 5, indicate a decline in PSNR values as image size decreases, contrasting with the full

Table 4
Image Demosaicking With INR: Average PSNR values for image demosaicking using various INR models on the Kodak dataset. The models were overfitted on Bayer measurements across different image sizes and training iterations, as opposed to Table 2, where all image pixels were taken into account. In this setup, minimization was performed using the objective in (3), where the complementary loss was neglected ($\beta = 0$) and $\gamma = 0$. **Bold** and underline highlight the highest and second highest values, respectively.

INR Model	Original size (768 × 512)			1/2 Resize (384 × 256)			1/4 Resize (192 × 128)		
	500 Iter	1000 Iter	2000 Iter	500 Iter	1000 Iter	2000 Iter	500 Iter	1000 Iter	2000 Iter
Gauss	25.66	26.76	27.75	16.23	16.04	17.91	14.23	14.43	14.29
SIREN	**34.21**	**34.11**	33.96	**31.25**	**31.19**	**30.93**	**31.11**	**31.38**	**31.33**
FFN	31.36	33.13	**34.41**	28.75	29.93	30.22	25.34	25.16	25.40
Incode	33.38	33.64	33.95	26.54	28.23	30.10	20.28	20.37	20.47

Table 5
Average PSNR values for image demosaicking using various INR models on the MCM dataset, following the same setup as described for Table 4. **Bold** and underline highlight the highest and second highest values, respectively.

INR Model	Original size (500 × 500)			1/2 Resize (250 × 250)			1/4 Resize (125 × 125)		
	500 Iter	1000 Iter	2000 Iter	500 Iter	1000 Iter	2000 Iter	500 Iter	1000 Iter	2000 Iter
Gauss	21.83	22.48	23.57	13.71	13.82	14.13	12.30	12.27	12.23
SIREN	**34.87**	**34.75**	34.71	**31.49**	**31.63**	**31.48**	**29.48**	**29.88**	**29.88**
FFN	32.26	33.15	33.75	25.34	26.29	26.64	19.09	20.19	20.33
Incode	33.99	34.65	**35.16**	21.99	22.89	25.49	15.45	15.58	15.50

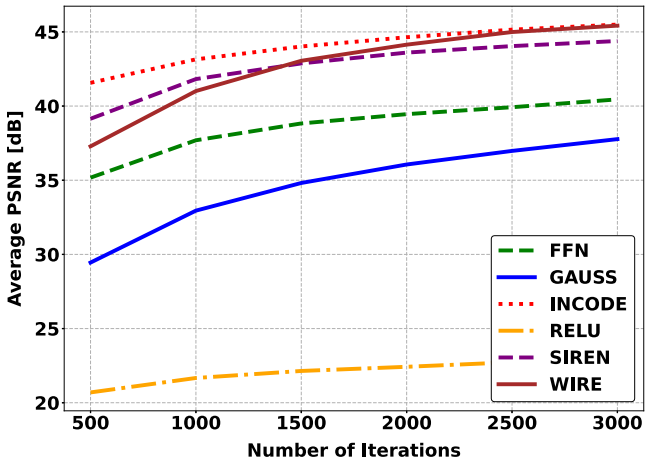


Fig. 4. Image Reconstruction With INR: Average PSNR values with increasing training iterations for various INR models on the McM dataset (500 × 500 version). The plot demonstrates continued improvement in PSNR with additional iterations, though the rate of gain decreases over time for all models.

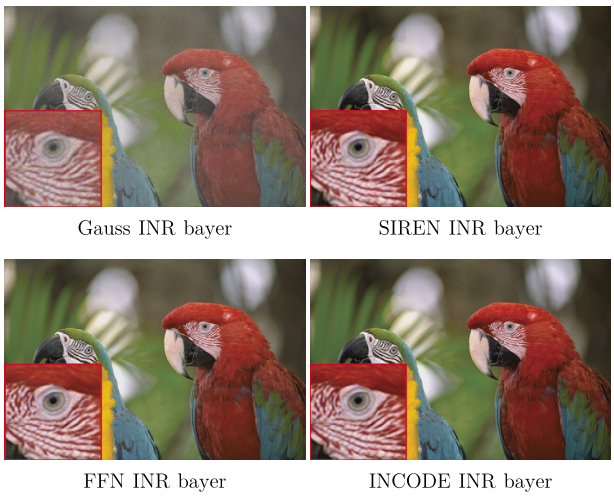


Fig. 5. Image Demosaicking With INR: Demosaicking results for Kodak image #23 (original size) using various INR architectures trained on Bayer measurements (Fig. 3, top-left). INCODE, SIREN, and FFN outperform the Gaussian model after 2000 iterations. PSNR values are listed in Table 4, 4th column.

image representation results in the previous section. This decline is due to the reduced availability of ground truth pixels and increased impact of CFA degradation in smaller images.

Interestingly, while INCODE excelled in full image representation, the SIREN architecture outperforms it in the demosaicking task, particularly with smaller images. SIREN's superior PSNR values highlight its robustness in scenarios requiring significant interpolation.

The visual demonstration is presented in Figs. 5 and 6. In the original size (Fig. 5), only Gauss exhibits improper reconstruction. When resized to half the original size (Fig. 6), SIREN begins to handle the reconstruction more effectively, producing a more colorful image. As the image size is reduced further, SIREN becomes the only model capable of adequately managing the interpolation.

The naive approach to INR-based demosaicking explained in this section, especially when using the SIREN architecture, surpasses basic algorithms like nearest neighbor and bilinear interpolation (see Table 6). As will be seen in the next section, it also highlights the potential of SIREN for boosting traditional demosaicking methods when initial information about missing pixels is provided.

4.4. Enhancing image demosaicking with INR

We now take full advantage of the INRID framework by incorporating both Bayer and complementary loss functions. This experiment corresponds to setting $\beta = 1$ in Equation (3), while varying α to balance the contributions of the Bayer and complementary losses. We keep the TV regularization turned off.

The inclusion of complementary loss leverages initial demosaicking reconstructions, regularizing the problem and, with the aid of Bayer loss, ultimately boosting the demosaicking capabilities of the original methods. This approach helps the INR model to learn from not only the available Bayer data but also the estimated values from the initial demosaicking process, thus improving the overall reconstruction quality.

Optimal Alpha Selection: To determine the optimal value for α, we conducted experiments using various demosaicking methods. Fig. 7 shows the average PSNR and SSIM from the McM dataset as a function of α for traditional demosaicking methods such as Malvar and Menon. The results indicate that, while the complementary loss plays a crucial role

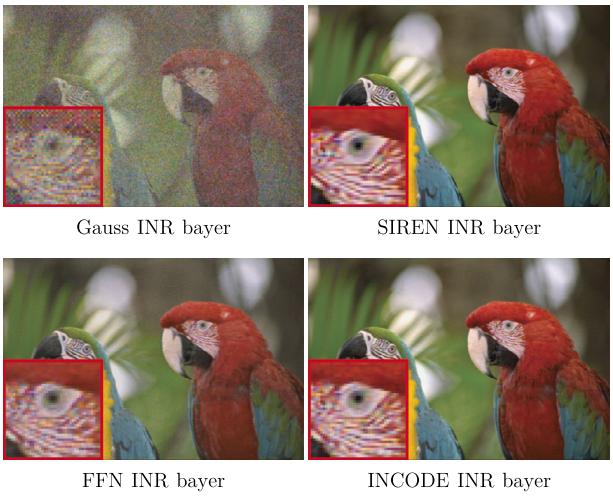


Fig. 6. Image Demosaicking With INR: Kodak image #23 resized to 384×256. SIREN INR produces the best visual outcome after 2000 iterations, while Gaussian INR shows significant artifacts. PSNR values are in Table 4, 7th column.

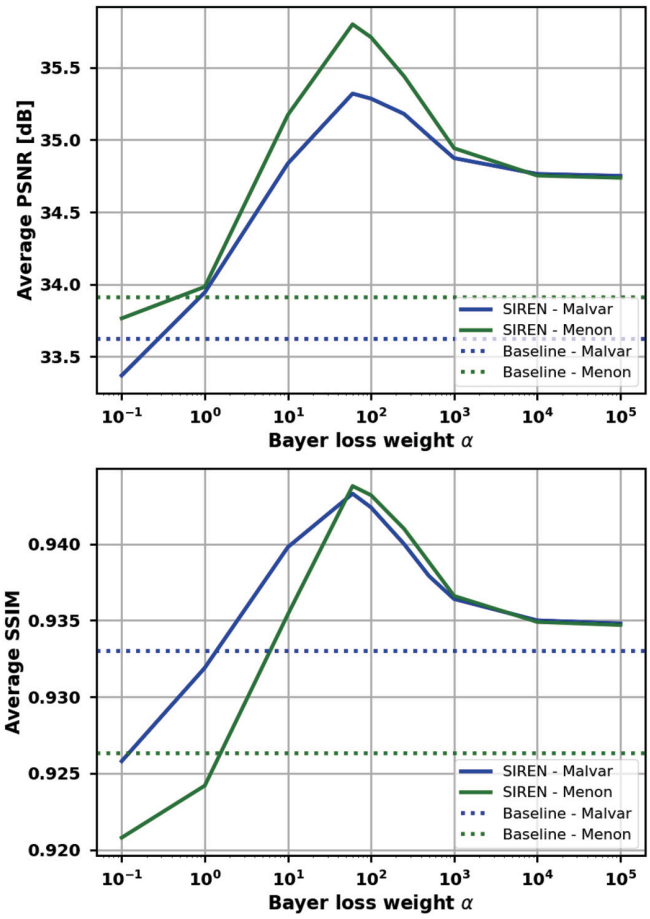


Fig. 7. Optimal Alpha Selection: Average PSNR and SSIM vs Alpha for SIREN-based demosaicking over the McM dataset (500×500). The plots show the performance of SIREN models to improve Malvar and Menon demosaicking methods as a function of α. The solid lines represent SIREN results, while the dotted lines show the baseline performance. The results indicate that for $\alpha > 1$, our INRID framework enhances demosaicking quality, with peak improvements around $\alpha = 60$, as seen in both PSNR and SSIM. As α approaches infinity, performance approaches the naive INR-based demosaicking.

Table 6
Enhancing Image Demosaicking with INR: Average PSNR and SSIM values for different demosaicking methods on the McM (500×500) dataset after 2000 iterations and the Kodak (192×128) dataset after 10000 iterations with $\alpha = 60$. **Bold** indicates the highest values.

Method	McM (500×500) PSNR/SSIM	Kodak (192×128) PSNR/SSIM
Nearest Neighbor	27.54/0.8594	25.11/0.7973
Bilinear	30.41/0.9276	26.61/0.8685
Bayer INRID	**34.71/0.9348**	**31.33/0.9283**
Malvar	33.62/0.9330	31.68/0.9420
Malvar INRID	**35.31/0.9433**	**32.58/0.9449**
Menon	33.91/0.9263	33.29/0.9571
Menon INRID	**35.80/0.9438**	**33.74/0.9586**
RSTCANet	**40.06/0.9739**	**38.40/0.9839**
RSTCANet INRID	36.95/0.9501	38.38/0.9836

in guiding the training process, the Bayer loss remains more dominant. When α is below 1, which emphasizes the complementary loss more than the Bayer loss, performance degrades compared to the baseline (dotted line), providing no enhancement at all. However, when α is greater than 1, INRID begins to enhance the original demosaicking methods, with the most significant improvements occurring when α is within the range of $(10, 200)$. As α increases further towards infinity, the complementary loss influence diminishes, and the model essentially reverts to the naive INR demosaicking approach discussed in the previous section. Based on these findings, we selected $\alpha = 60$, which yielded the best average improvements in both PSNR and SSIM.

Boosting Demosaicking Performance: Table 6 highlights the impact of the INRID approach in improving traditional demosaicking techniques, specifically Malvar and Menon, using the SIREN architecture. The results are consistent across both the Kodak and McM datasets, where the INRID framework significantly boosts the performance, leading to noticeable improvements in both PSNR and SSIM. For both traditional demosaicking methods, Malvar and Menon, integrating the INRID approach results in visibly better reconstruction quality, particularly in challenging areas with fine details or high-frequency content, as illustrated in Fig. 8.

Basic demosaicking algorithms like nearest neighbor and bilinear interpolation are surpassed by even the naive INR demosaicking (Bayer INRID) introduced in the previous section. For these methods, incorporating initial reconstruction degrades the enhancement.

It is worth noting, however, that INRID has limitations. Once the initial demosaicking reconstruction reaches a certain level of accuracy, further improvement of a given method is limited. This is evident in the PSNR values for the Transformer-based demosaicking method RSTCANet, as seen in Table 6. However, for state-of-the-art methods, the INRID framework can still be valuable when addressing joint problems such as demosaicking combined with denoising or deblurring.

4.5. Joint demosaicking and denoising

While INRID may not directly enhance state-of-the-art demosaicking methods, like RSTCANet, it shows significant improvements when dealing with out-of-distribution data, such as images corrupted by noise. To demonstrate this robustness, we conducted experiments on the joint demosaicking and denoising task using the Kodak dataset resized to 192×128. We introduced varying levels of Gaussian noise, with signal-to-noise ratios (SNRs) ranging from 10 dB to 40 dB, and then applied our INRID framework with RSTCANet as the initial demosaicking reconstruction.

We compared our approach against a baseline and two specialized methods for joint demosaicking and denoising. The first is a classical method that builds upon the demosaicking technique of Farsiu et

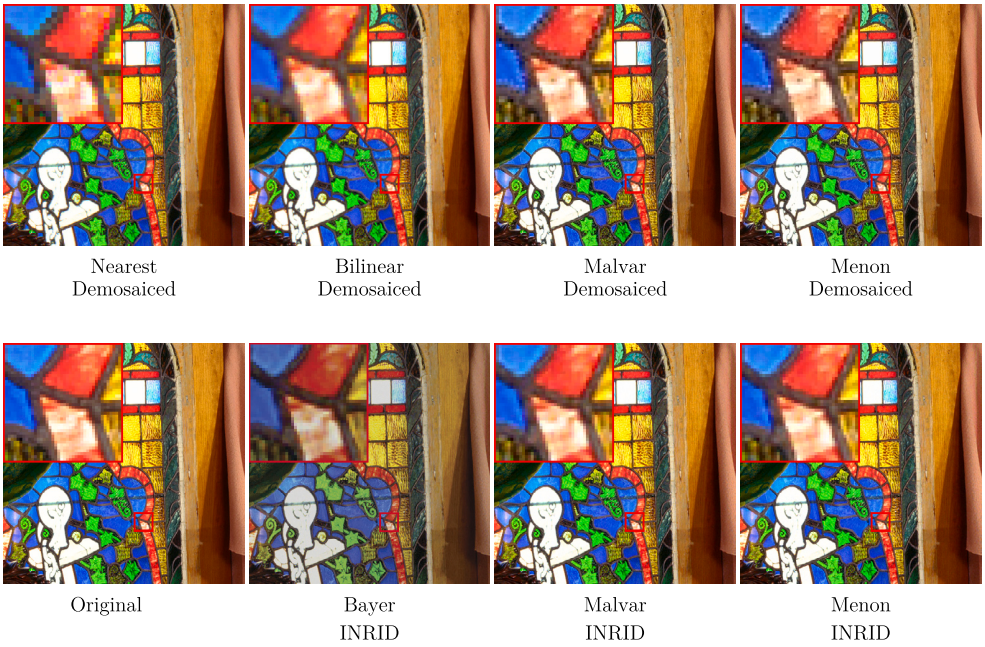


Fig. 8. Enhancing Image Demosaicking with INR: Visual Comparison of Demosaicking Methods with and without INRID Enhancement (applied to McM image #1, 500×500). The top row showcases traditional demosaicking results from Nearest Neighbor, Bilinear, Malvar, and Menon methods. The bottom row starts with the original image. The Bayer INRID approach overfits directly to the Bayer measurement without initial demosaicking, demonstrating a higher reconstruction quality compared to Nearest Neighbor and Bilinear methods. INRID significantly boosts the performance of the Malvar and Menon methods, particularly in high-frequency regions, such as along the stained glass edges (see red close-ups).

Table 7
Joint Demosaicking and Denoising: Average PSNR and SSIM values for different iterations and SNR levels on the Kodak dataset (192×128). The parameters are: $\alpha = 1$, $\beta = 1$, and $\gamma = 0$, with noise levels in SNR (dB). INRID uses initial reconstruction from RSTCANet.

Model	Iterations	PSNR / SSIM			
		10 dB	20 dB	30 dB	40 dB
Classical (HQ)	-	23.68/0.5826	27.75/0.7934	30.20/0.9063	30.78/0.9348
DeepDemosaick	-	17.80/0.2985	28.83/0.8076	34.27/0.9489	35.70/0.9666
RSTCANet	-	17.51/0.2878	26.57/0.6620	34.31/0.9225	**37.74/0.9777**
INRID (RSTCANet init)	500 Iter	**23.75/0.6352**	**29.19**/0.8238	33.20/0.9247	33.94/0.9385
	1000 Iter	23.73/0.6324	29.17/**0.8250**	34.67/0.9415	36.48/0.9661
	2000 Iter	23.69/0.6309	29.11/0.8238	**34.77/0.9425**	37.63/0.9770

al. [18], formulated via half-quadratic (HQ) approximation in a multiplicative form [32] and solved by alternating minimization. While this approach can also integrate deblurring using a suitable kernel, we used a delta kernel here to focus solely on denoising and demosaicking. The second method is DeepDemosaick (introduced in Section 2), a deep convolutional residual network designed for joint demosaicking and denoising.

Since RSTCANet already provides high-quality initial reconstructions, we placed equal emphasis on the Bayer and complementary losses, setting $\alpha = 1$ and $\beta = 1$. To showcase the denoising capabilities of INR, TV regularization was disabled ($\gamma = 0$). To mitigate overfitting to noisy measurements, an early-stopping mechanism is employed for INRID, with training concluding after 500 to 2000 iterations.

Table 7 presents a comparative analysis of the four methods across various noise levels. For heavy to moderate noise conditions (10–30 dB SNR), INRID consistently surpasses RSTCANet and outperforms both the classical and deep-learning-based approaches. While the classical method delivers competitive results under severe noise conditions (10–20 dB SNR), its performance diminishes as noise levels decrease. Notably, DeepDemosaick closely matches INRID's performance at 20 dB SNR.

At 20 dB SNR, INRID achieves a PSNR of 29.19 dB after 500 iterations, outperforming RSTCANet's 26.57 dB and the classical method's 27.75 dB. DeepDemosaick achieves 28.83 dB at this noise level and its visual quality is comparable to INRID (see Fig. 9). Visually, RSTCANet and the classical method exhibit noticeable artifacts, whereas INRID effectively removes noise, especially around detailed regions such as window shutters.

Under extreme noise conditions (10 dB SNR), the advantage of INRID becomes more pronounced, with a PSNR of 23.75 dB compared to RSTCANet's 17.51 dB, representing a significant improvement. At 30 dB SNR, where noise levels are lower, the performance gap between INRID and its initial RSTCANet reconstruction narrows. Finally, at 40 dB SNR, where noise is minimal, RSTCANet achieves the highest PSNR (37.74 dB), and further refinement by INRID does not yield additional improvements. These results align with the conclusions drawn in the preceding section.

4.6. Joint demosaicking and deblurring

This experiment evaluates INRID's impact on traditional and advanced demosaicking methods when integrated with deblurring. Image

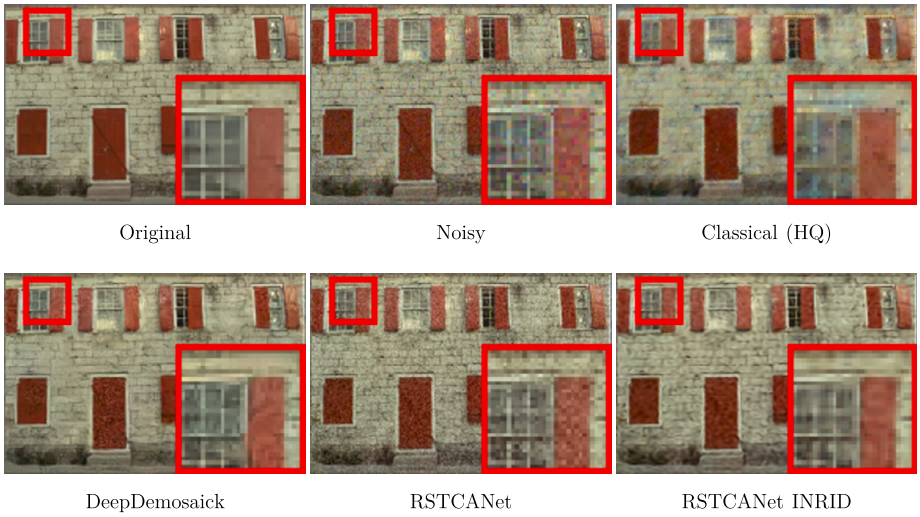


Fig. 9. Joint Demosaicking and Denoising: The first row presents the original image, its noisy counterpart (input SNR = 20 dB), and the output from the Classical method (HQ minimization). The second row showcases results from DeepDemosaick, RSTCANet, and RSTCANet INRID. The corresponding PSNR and SSIM results are shown in Table 7, 4th column. The image is from the Kodak dataset (resized to 192 × 128), with INRID parameters set to $\alpha = \beta = 1$ and $\gamma = 0$.

Table 8
Joint Demosaicking and Deblurring with Uniform Kernel: Average PSNR and SSIM for joint demosaicking and deblurring on the Kodak dataset (resized to 192 × 128) with a uniform kernel (3 × 3), 50 dB noise after 10,000 iterations. +TV indicates that total variation with $\gamma = 10^{-6}$ was added in the minimization problem (3). Significant improvements with INRID are particularly evident in methods like RSTCANet and DeepDemosaick, with the best results highlighted in **bold**. *Italicized* entries indicate demosaicking results without deblurring.

Method	Original	INRID Enhancement			
		$\alpha = 1$	$\alpha = 1$, +TV	$\alpha = 60$	$\alpha = 60$, +TV
Nearest	*25.71/0.7701*	27.51/0.8232	27.56/0.8236	31.57/0.9113	**31.58/0.9115**
Bilinear	*25.06/0.7999*	26.52/0.8420	26.52/0.8413	31.32/0.9120	**31.36/0.9124**
Malvar	*27.34/0.8293*	30.84/0.9071	30.96/0.9120	32.40/0.9252	**32.44/0.9257**
Menon	*27.03/0.8159*	31.19/0.9093	31.34/0.9144	32.67/0.9272	**32.74/0.9285**
DeepDemosaick	*27.29/0.8232*	32.17/0.9198	32.27/0.9211	32.95/0.9289	**33.01/0.9302**
RSTCANet	*27.31/0.8251*	33.38/0.9374	**33.60/0.9409**	33.24/0.9324	33.35/0.9344
Wiener Filtering	26.53/0.7959	—	—	—	—
IWFT	26.48/0.8330	—	—	—	—
D3Net	29.86/0.8736	—	—	—	—
HQ	31.66/0.9280	—	—	—	—
Bayer INRID	32.59/0.9241	—	—	—	—

quality degradation in such scenarios primarily arises from convolution operations, which introduce blur during image acquisition. Since initial demosaicking guides INRID in interpolating missing data, the effectiveness of deconvolution critically depends on the quality of this preliminary interpolation.

Experiments were conducted on the Kodak dataset using Gaussian and uniform blur kernels with added noise at 50 dB to simulate the forward problem (1). Results are summarized in Tables 8 and 9 for 3 × 3 kernels and in Tables 10 and 11 for 7 × 7 kernels. First, it shows that INRID enhancement of traditional techniques, such as Nearest neighbor, Bilinear interpolation and Malvar's method, is suboptimal and outperformed by INRID deblurring with Bayer measurements alone (Bayer INRID). The traditional approaches produce initial reconstructions that fail to adequately match the original image distribution, leading to insufficient deconvolution performance.

In contrast, state-of-the-art methods, such as DeepDemosaick and RSTCANet, deliver more accurate initial demosaicking results, which, when enhanced with INRID, yield significantly improved deblurring performance. For instance, under Uniform 3 × 3 blur (Table 9), DeepDemosaick improves from 27.29 dB to 33.01 dB, and RSTCANet improves from 27.31 dB to 33.60 dB.

In most scenarios, choosing $\alpha = 60$ yields higher PSNR and SSIM values, as discussed in Section 4.4. However, when using RSTCANet initialization for images blurred with a smaller 3 × 3 kernel, the best results occur at $\alpha = 1$.

The addition of TV regularization with $\gamma = 10^{-6}$ provides a modest improvement in PSNR for all methods (except RSTCANet with Gaussian 3 × 3 blur). Notably, for all tested scenarios, DeepDemosaick and RSTCANet paired with INRID outperform Bayer INRID, underscoring the importance of accurate initial demosaicking. For state-of-the-art methods, the complementary loss in INRID plays a significant role in enhancing reconstruction quality. Fig. 10 demonstrates how INRID integration performs demosaicking and deblurring effectively. Nevertheless, with larger blur kernels, the INRID enhancement of state-of-the-art demosaicked methods is less significant compared to the Bayer INRID (see Table 10).

We further compared INRID with other joint demosaicking and deblurring methods. Specifically, we extended Iterative Wiener Filtering and Thresholding (IWFT) [33] to handle demosaicking and deblurring by incorporating formation model (1) into equation (1) in [33]. After integrating the new degradation model, the algorithm was modified accordingly, and the rest follows the original IWFT pipeline. The first

Table 9
Joint Demosaicking and Deblurring with Gaussian Kernel: Average PSNR and SSIM for joint demosaicking and deblurring on the Kodak dataset (resized to 192 × 128) with a Gaussian kernel (3 × 3), 50 dB noise after 10,000 iterations. +TV indicates that total variation with $\gamma = 10^{-6}$ was added in the minimization problem (3). Significant improvements with INRID are particularly evident in methods like RSTCANet and DeepDemosaick, with the best results highlighted in **bold**. *Italicized* entries indicate demosaicking results without deblurring.

Method	Original	INRID Enhancement			
		$\alpha = 1$	$\alpha = 1$, +TV	$\alpha = 60$	$\alpha = 60$, +TV
Nearest	*25.98/0.7877*	27.57/0.8314	27.56/0.8311	31.54/0.9142	**31.54/0.9145**
Bilinear	*25.54/0.8218*	26.77/0.8552	26.77/0.8548	31.31/0.9152	**31.34/0.9155**
Malvar	*28.48/0.8666*	31.02/0.9125	31.03/0.9145	32.80/0.9319	**32.85/0.9320**
Menon	*28.24/0.8600*	31.32/0.9141	31.33/0.9158	32.89/0.9331	**32.90/0.9336**
DeepDemosaick	*28.39/0.8567*	32.22/0.9249	32.15/0.9234	33.15/0.9354	**33.17/0.9359**
RSTCANet	*28.64/0.8693*	**33.84/0.9440**	33.66/0.9415	33.74/0.9401	33.71/0.9398
Wiener Filtering	28.13/0.8336	—	—	—	—
IWFT	28.18/0.8540	—	—	—	—
D3Net	31.06/0.8988	—	—	—	—
HQ	32.45/0.9399	—	—	—	—
Bayer INRID	33.07/0.9329	—	—	—	—

Table 10
Joint Demosaicking and Deblurring with Uniform Kernel: Average PSNR and SSIM on the Kodak dataset (resized to 192 × 128) with a uniform kernel (7 × 7), 50 dB noise after 10,000 iterations. + TV indicates that total variation with $\gamma = 10^{-6}$ was added in the minimization problem. Each row's best PSNR/SSIM is in **bold**. *Italicized* entries indicate demosaicking results without deblurring.

Method	Original	INRID Enhancement			
		$\alpha = 1$	$\alpha = 1$, +TV	$\alpha = 60$	$\alpha = 60$, +TV
Nearest	*23.32/0.5986*	26.57/0.7627	26.61/0.7666	29.60/0.8450	**29.72/0.8499**
Bilinear	*22.60/0.6078*	24.47/0.6804	24.47/0.6812	28.94/0.8274	**28.96/0.8291**
Malvar	*23.55/0.6084*	27.87/0.8071	27.97/0.8125	29.94/0.8514	**30.10/0.8574**
Menon	*23.50/0.6065*	29.25/0.8422	29.38/0.8474	30.12/0.8559	**30.36/0.8637**
DeepDemosaick	*23.55/0.6105*	29.06/0.8329	28.87/0.8273	30.11/0.8545	**30.31/0.8615**
RSTCANet	*23.55/0.6096*	30.16/0.8587	29.97/0.8548	30.14/0.8549	**30.39/0.8623**
Wiener Filtering	25.28/0.7129	—	—	—	—
IWFT	25.21/0.7114	—	—	—	—
D3Net	26.16/0.7266	—	—	—	—
HQ	30.01/0.8773	—	—	—	—
Bayer INRID	30.11/0.8540	—	—	—	—

Table 11
Joint Demosaicking and Deblurring with Gaussian Kernel: Average PSNR and SSIM on the Kodak dataset (resized to 192 × 128) with a Gaussian kernel (7 × 7), 50 dB noise after 10,000 iterations. + TV indicates that total variation with $\gamma = 10^{-6}$ was added in the minimization problem. Each row's best PSNR/SSIM is in **bold**. *Italicized* entries indicate demosaicking results without deblurring.

Method	Original	INRID Enhancement			
		$\alpha = 1$	$\alpha = 1$, +TV	$\alpha = 60$	$\alpha = 60$, +TV
Nearest	*24.03/0.6509*	26.58/0.7752	26.43/0.7718	29.22/0.8420	**29.29/0.8448**
Bilinear	*23.24/0.6619*	24.87/0.7180	24.87/0.7190	28.58/0.8275	**28.64/0.8305**
Malvar	*24.47/0.6734*	27.58/0.8092	27.41/0.8048	29.35/0.8441	**29.37/0.8465**
Menon	*24.37/0.6681*	28.79/0.8360	28.28/0.8227	29.46/0.8451	**29.55/0.8489**
DeepDemosaick	*24.41/0.6685*	28.45/0.8248	27.84/0.8087	29.44/0.8456	**29.49/0.8482**
RSTCANet	*24.44/0.6716*	29.50/0.8480	28.55/0.8249	29.49/0.8459	**29.55/0.8488**
Wiener Filtering	25.34/0.7487	—	—	—	—
IWFT	25.78/0.7909	—	—	—	—
D3Net	26.58/0.7334	—	—	—	—
HQ	29.06/0.8543	—	—	—	—
Bayer INRID	29.46/0.8445	—	—	—	—

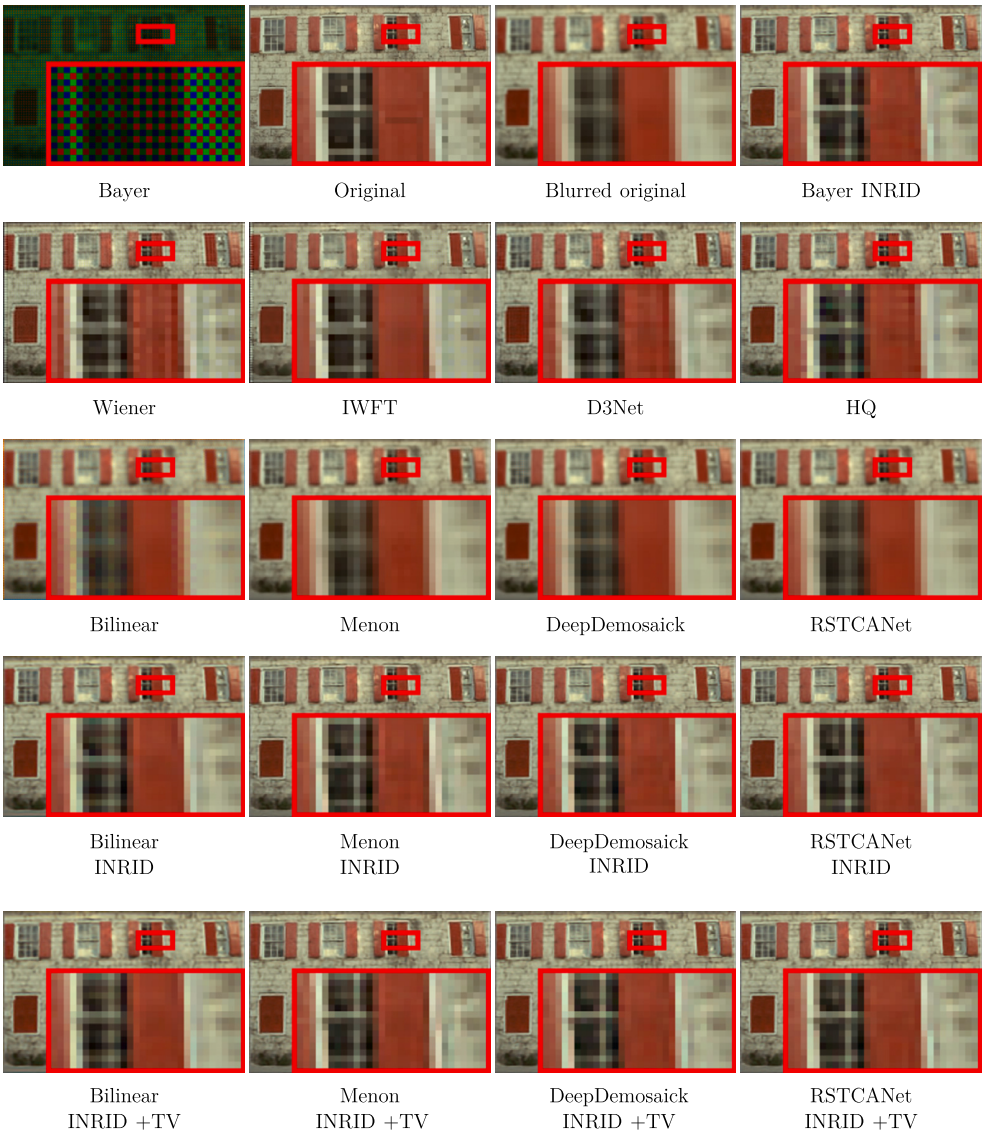


Fig. 10. Joint Demosaicking and Deblurring: Visual example of INRID's performance on Kodak image #1 (192 × 128), degraded with 3 × 3 Gaussian blur and 50 dB noise. Corresponding PSNR, SSIM, and other metrics are in Table 9.

reconstruction step in IWFT corresponds to Wiener Filtering, a popular deconvolution technique. When applied to our test set, this method introduced ringing artifacts around edges due to its linear nature (clearly visible around the window shutters in Fig. 10).

The complete IWFT algorithm then uses non-linear update steps to refine the Wiener-based reconstruction, effectively suppressing these artifacts and producing smoother images. However, IWFT tends to oversmooth images, leading to a slight reduction in PSNR (specifically in the case of uniform blur) despite the noticeable visual improvements.

D3Net [9] is an end-to-end CNN developed for joint demosaicking, deblurring, and deringing. It surpasses IWFT in terms of PSNR but remains constrained by its lightweight architecture, which targets embedded devices with limited computational resources. HQ is a robust optimization framework for joint demosaicking and deblurring method introduced in Section 4.5. While HQ outperforms D3Net, it still does not reach the reconstruction quality offered by INRID methods.

While INRID significantly enhances demosaicking and achieves superior reconstruction quality compared to baseline and joint techniques, its computational cost remains a notable drawback. Fig. 11 illustrates the average PSNR and processing time for RSTCANet INRID over the Kodak dataset. The PSNR improves steadily up to approximately 11,000 iterations, after which it saturates, whereas the runtime continues to increase linearly, exceeding 100 seconds for 11,000 iterations on an NVIDIA L40S GPU. In contrast, traditional methods such as Wiener filtering, IWFT, and HQ are significantly faster, completing reconstruction in just a few seconds. D3Net achieves similar inference times to these methods but requires several minutes of pretraining for each specific blur and noise level. Despite the superior reconstruction quality of INRID-enhanced methods, their computational cost limits their practicality for scenarios demanding fast or resource-efficient processing, a limitation that will be further addressed in Section 5.

4.7. Real data

To further validate the reconstruction effectiveness of our INRID approach, we tested it on real raw data captured from an LG Nexus 5 camera. The blur kernels (displayed in Fig. 1b) were estimated from calibration data.

This experiment demonstrates the practical benefits of applying joint demosaicking and deblurring on actual image data. As shown in Fig. 12,

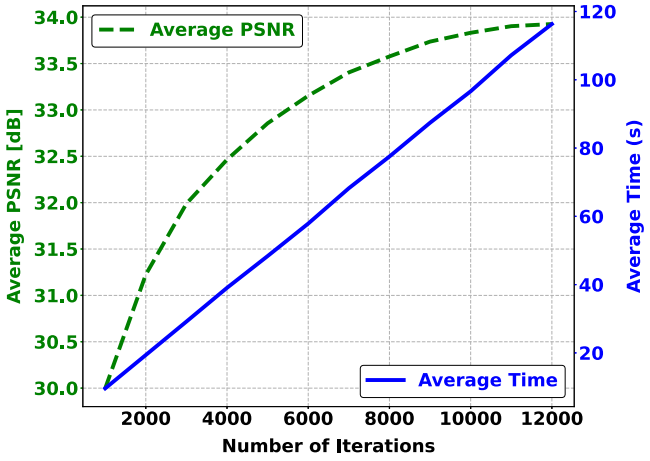


Fig. 11. Computational cost of RSTCANet-INRID with $\alpha = \beta = 1$ and $\gamma = 0$ on the Kodak dataset (resized to 192×128). The figure shows the average PSNR (green dashed line) and runtime (blue solid line) as a function of the number of iterations, using the parameters described in Table 9. The PSNR improves steadily until saturating around 11,000 iterations, while the runtime increases linearly, reaching 2 minutes for 12,000 iterations on an NVIDIA L40S GPU.

the INRID framework, combined with prior information about the camera's PSFs, yields a significant improvement over both the standard JPEG output and the advanced RSTCANet model.

The raw Bayer data are seen in the second column of Fig. 12. The JPEG output (third column) exhibits considerable compression artifacts and blurring, especially in magnified areas. RSTCANet, shown in the fourth column, improves the reconstruction quality but still leaves some residual blurring and noise.

In contrast, the final column illustrates the result of applying RSTCANet in conjunction with the INRID framework, leveraging PSF priors for all four RGGB channels in the raw Bayer data. By setting $\alpha = 1$ and $\beta = 1$, and enabling TV regularization with a value of 10^{-5}, our approach effectively removes noise and reduces blurring. This leads to a visually sharper and more accurate reconstruction, as highlighted by the red close-up in Fig. 12. Zoomed-in views of the green and blue bordered regions show further evidence of INRID's ability in enhancing the baseline RSTCANet method.

5. Discussion and future work

The results demonstrate the significant potential of INRID in enhancing traditional and state-of-the-art demosaicking methods. However, the computational cost associated with per-image training remains a notable limitation, particularly for large datasets or high-resolution images. While this study focuses on reconstruction quality, addressing efficiency is a crucial challenge for expanding the practical utility of INRID, especially in real-time applications.

Training INRID for each image is time-intensive. For example, processing a 192×128 image required approximately 96 seconds for 10,000 iterations on an NVIDIA L40s GPU, using around 1 GB of memory. This is orders of magnitude slower than traditional methods like Malvar or Menon, which complete within milliseconds for similar resolutions. Similarly, pre-trained models like RSTCANet offer real-time inference but fail to handle scenarios involving corrupted inputs, such as blurred or noisy data. In contrast, iterative joint demosaicking and deblurring methods process images within seconds but may not match INRID's reconstruction fidelity.

To make INRID more practical, future efforts should aim to reduce its computational overhead. Promising approaches include multiresolution hash encoding [34], which could cut training times to seconds, and dictionary-based representations like Neural Implicit Dictionary (NID) [35], which leverage pre-learned basis functions for efficient reconstructions without a long per-image training.

Scaling INRID to gigapixel-resolution images also presents challenges due to the extensive training times required for basic architectures. Techniques such as tiling, which processes overlapping sections of an image with smaller MLPs, can enable parallel computation but may introduce stitching artifacts at boundaries. Refinements like KiloNeRF [36], which divides scenes into thousands of compact neural networks, and Multiscale Implicit Neural Representation (MINER) [37], which processes images hierarchically, offer promising solutions. Additionally, hybrid frameworks like ACORN [38] dynamically allocate resources based on local signal complexity, optimizing both memory usage and training time for high-resolution applications.

Beyond computational improvements, extending INRID to related tasks such as super-resolution and inpainting is a natural progression, given the similar challenge of reconstructing missing data. Integrating conditioning mechanisms, such as activation function modulations [39] or meta-learning paradigms [40], could further enhance generalization across diverse images while reducing per-image training requirements. This is particularly relevant for refining state-of-the-art methods in scenarios with out-of-distribution data, such as blur or noise.

Theoretical advancements addressing spectral bias [41]—a tendency of MLPs to prioritize low-frequency components over high-frequency details—are also essential. A structured dictionary perspective [40], where MLPs learn representations from a set of predefined basis functions, offers a promising direction to improve high-frequency detail reconstruction and overall image fidelity.

Ultimately, while INRID achieves superior reconstruction quality, its computational demands highlight clear challenges and opportunities for future work. Advances in training efficiency, scalability, and generalization will be crucial in realizing the broader applicability of INRID across diverse image reconstruction tasks while preserving its fidelity.

6. Conclusion

This paper introduced INRID, a novel framework leveraging Implicit Neural Representations for image demosaicking. By integrating Bayer loss to enforce fidelity to sensor data and complementary loss to utilize initial reconstructions, INRID significantly enhances traditional methods like Malvar and Menon, achieving PSNR improvements of up to 2 dB. The framework also addresses limitations in deep learning-based methods, effectively correcting artifacts and demonstrating resilience in challenging scenarios, including blur and noise. Real-world validation on raw sensor data from mobile cameras further underscored INRID's capability to produce sharper and more accurate reconstructions compared to standard outputs and advanced pipelines like RSTCANet.

While INRID achieves state-of-the-art reconstruction fidelity, its computational demands highlight opportunities for further optimization. Future work will focus on improving efficiency through approaches such as multiresolution encoding and dictionary-based representations, and scaling to gigapixel images using advanced frameworks like MINER. Extending INRID to tasks like super-resolution and inpainting represents a promising direction, leveraging its capacity to adapt to diverse input characteristics while maintaining high fidelity.

In conclusion, INRID demonstrates the potential of implicit neural representations to not only improve demosaicking quality but also tackle joint problems such as denoising and deblurring, paving the way for their integration into advanced image reconstruction pipelines.

CRediT authorship contribution statement

Tomáš Kerepecký: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Filip Šroubek:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Jan Flusser:**

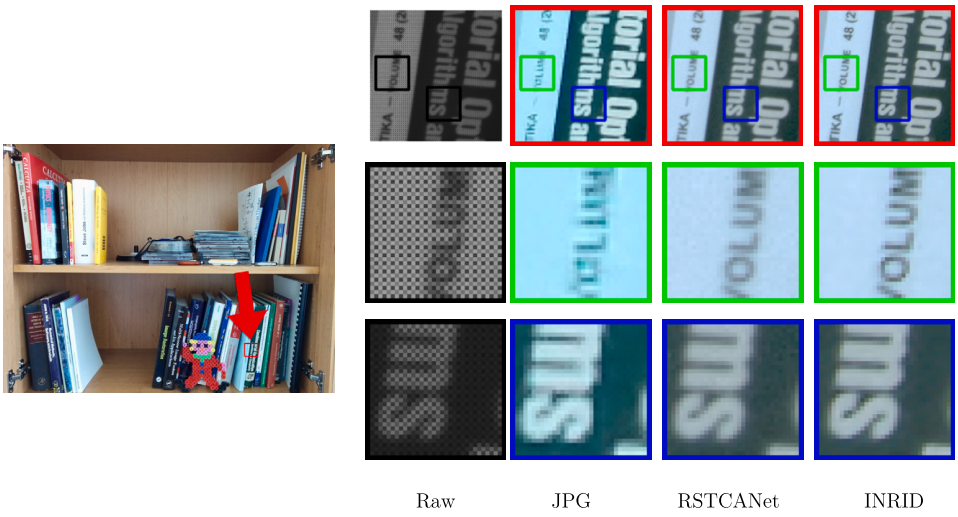


Fig. 12. Joint Demosaicking and Deblurring on real data from an LG Nexus 5 camera. The red arrow in the large image highlights the cropped area. The first column shows raw Bayer data, the second is the JPEG output, the third is the result from RSTCANet, and the fourth is INRID with initial RSTCANet estimate and deblurring using PSF priors. Green and blue borders in the first row show close-ups of specific regions.

Funding acquisition, Project administration, Supervision, Writing – review & editing.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used Grammarly and ChatGPT solely to improve the readability and language of the manuscript. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Czech Science Foundation grant 25-15933S.

Data availability

The code repository and details, along with the publicly available Kodak and McM datasets used, are provided in the Experimental section with the GitHub link included in the manuscript.

References

[1] H.S. Malvar, L.-w. He, R. Cutler, High-quality linear interpolation for demosaicing of Bayer-patterned color images, in: 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 3, IEEE, 2004, pp. iii–485.
[2] D. Menon, S. Andriani, G. Calvagno, Demosaicing with directional filtering and a posteriori decision, IEEE Trans. Image Process. 16 (1) (2006) 132–141.
[3] K. Hirakawa, T.W. Parks, Joint demosaicing and denoising, IEEE Trans. Image Process. 15 (8) (2006) 2146–2157.
[4] C.J. Schuler, M. Hirsch, S. Harmeling, B. Schölkopf, Non-stationary correction of optical aberrations, in: 2011 International Conference on Computer Vision (CVPR), IEEE, 2011, pp. 659–666.
[5] L. Condat, S. Mosaddegh, Joint demosaicking and denoising by total variation minimization, in: 2012 19th IEEE International Conference on Image Processing (ICIP), IEEE, 2012, pp. 2781–2784.
[6] H.Q. Luong, B. Goossens, J. Aelterman, A. Pižurica, W. Philips, A primal-dual algorithm for joint demosaicking and deconvolution, in: 2012 19th IEEE International Conference on Image Processing (ICIP), IEEE, 2012, pp. 2801–2804.
[7] F. Heide, M. Steinberger, Y.-T. Tsai, M. Rouf, D. Pajak, D. Reddy, O. Gallo, J. Liu, W. Heidrich, K. Egiazarian, et al., Flexisp: a flexible camera image processing framework, ACM Trans. Graph. 33 (6) (2014) 1–13.
[8] D.S. Yoo, M.K. Park, M.G. Kang, Joint deblurring and demosaicing using edge information from Bayer images, IEICE Trans. Inf. Syst. 97 (7) (2014) 1872–1884.
[9] T. Kerepecky, F. Šroubek, D3net: joint demosaicking, deblurring and deringing, in: 2020 25th International Conference on Pattern Recognition (ICPR), IEEE, 2021, pp. 1–8.
[10] F. Kokkinos, S. Lefkimmiatis, Deep image demosaicking using a cascade of convolutional residual denoising networks, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 303–319.
[11] W. Xing, K. Egiazarian, Residual swin transformer channel attention network for image demosaicing, in: 2022 10th European Workshop on Visual Information Processing (EUVIP), IEEE, 2022, pp. 1–6.
[12] M. Gharbi, G. Chaurasia, S. Paris, F. Durand, Deep joint demosaicking and denoising, ACM Trans. Graph. 35 (6) (2016) 1–12.
[13] D.S. Tan, W.-Y. Chen, K.-L. Hua, Deepdemosaicking: adaptive image demosaicking via multiple deep fully convolutional networks, IEEE Trans. Image Process. 27 (5) (2018) 2408–2419.
[14] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, R. Timofte, Plug-and-play image restoration with deep denoiser prior, IEEE Trans. Pattern Anal. Mach. Intell. 44 (10) (2021) 6360–6376.
[15] Y. Zhang, K. Li, B. Zhong, Y. Fu, Residual non-local attention networks for image restoration, in: International Conference on Learning Representations (ICLR), 2019.
[16] Y. Xie, T. Takikawa, S. Saito, O. Litany, S. Yan, N. Khan, F. Tombari, J. Tompkin, V. Sitzmann, S. Sridhar, Neural fields in visual computing and beyond, in: Computer Graphics Forum, vol. 41, Wiley Online Library, 2022, pp. 641–676.
[17] X. Li, B. Gunturk, L. Zhang, Image demosaicing: a systematic survey, in: Visual Communications and Image Processing 2008, vol. 6822, SPIE, 2008, pp. 489–503.
[18] S. Farsiu, M. Elad, P. Milanfar, Multiframe demosaicing and super-resolution of color images, IEEE Trans. Image Process. 15 (1) (2005) 141–159.
[19] J.J. Park, P. Florence, J. Straub, R. Newcombe, S. Lovegrove, Deepsdf: learning continuous signed distance functions for shape representation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 165–174.
[20] M. Tancik, P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. Barron, R. Ng, Fourier features let networks learn high frequency functions in low dimensional domains, Adv. Neural Inf. Process. Syst. 33 (2020) 7537–7547.
[21] B. Mildenhall, P.P. Srinivasan, M. Tancik, J.T. Barron, R. Ramamoorthi, R. Ng, Nerf: representing scenes as neural radiance fields for view synthesis, Commun. ACM 65 (1) (2021) 99–106.
[22] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, G. Wetzstein, Implicit neural representations with periodic activation functions, Adv. Neural Inf. Process. Syst. 33 (2020) 7462–7473.
[23] V. Saragadam, D. LeJeune, J. Tan, G. Balakrishnan, A. Veeraraghavan, R.G. Baraniuk, Wire: wavelet implicit neural representations, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 18507–18516.

[24] A. Kazerouni, R. Azad, A. Hosseini, D. Merhof, U. Bagci, Incode: implicit neural conditioning with prior knowledge embeddings, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2024, pp. 1298–1307.
[25] S. Ramasinghe, S. Lucey, Beyond periodicity: towards a unifying framework for activations in coordinate-mlps, in: European Conference on Computer Vision (ECCV), Springer, 2022, pp. 142–158.
[26] T. Kerepecky, F. Sroubek, A. Novozamsky, J. Flusser, Nerd: neural field-based demosaicking, in: 2023 IEEE International Conference on Image Processing (ICIP), IEEE, 2023, pp. 1735–1739.
[27] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
[28] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, Springer, 2015, pp. 234–241.
[29] P. Blomgren, T.F. Chan, Color tv: total variation methods for restoration of vector-valued images, IEEE Trans. Image Process. 7 (3) (1998) 304–309.
[30] E. Kodak, Kodak lossless true color image suite (photocd pcd0992), http://r0k.us/graphics/kodak, 1993, 6.
[31] L. Zhang, X. Wu, A. Buades, X. Li, Color demosaicking by local directional interpolation and nonlocal adaptive thresholding, J. Electron. Imaging 20 (2) (2011) 023016.
[32] D. Geman, G. Reynolds, Constrained restoration and the recovery of discontinuities, IEEE Trans. Pattern Anal. Mach. Intell. 14 (03) (1992) 367–383.
[33] F. Šroubek, T. Kerepecky, J. Kamenicky, Iterative Wiener filtering for deconvolution with ringing artifact suppression, in: 2019 27th European Signal Processing Conference (EUSIPCO), IEEE, 2019, pp. 1–5.
[34] T. Müller, A. Evans, C. Schied, A. Keller, Instant neural graphics primitives with a multiresolution hash encoding, ACM Trans. Graph. 41 (4) (2022) 1–15.
[35] P. Wang, Z. Fan, T. Chen, Z. Wang, Neural implicit dictionary learning via mixture-of-expert training, in: International Conference on Machine Learning (ICML), in: PMLR, 2022, pp. 22613–22624.
[36] C. Reiser, S. Peng, Y. Liao, A. Geiger, Kilonerf: speeding up neural radiance fields with thousands of tiny mlps, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (CVPR), 2021, pp. 14335–14345.
[37] V. Saragadam, J. Tan, G. Balakrishnan, R.G. Baraniuk, A. Veeraraghavan, Miner: multiscale implicit neural representation, in: European Conference on Computer Vision (ECCV), Springer, 2022, pp. 318–333.
[38] J.N. Martel, D.B. Lindell, C.Z. Lin, E.R. Chan, M. Monteiro, G. Wetzstein, Acorn: adaptive coordinate networks for neural scene representation, ACM Trans. Graph. 40 (4) (2021) 1–13.
[39] E. Dupont, H. Kim, S.A. Eslami, D.J. Rezende, D. Rosenbaum, From data to functa: your data point is a function and you can treat it like one, in: International Conference on Machine Learning (ICML), in: PMLR, 2022, pp. 5694–5725.
[40] G. Yüce, G. Ortiz-Jiménez, B. Besbinar, P. Frossard, A structured dictionary perspective on implicit neural representations, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 19228–19238.
[41] N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. Hamprecht, Y. Bengio, A. Courville, On the spectral bias of neural networks, in: International Conference on Machine Learning (ICML), in: PMLR, 2019, pp. 5301–5310.

Tomáš Kerepecký earned his M.Sc. degree in Computational Physics from the Faculty of Nuclear Sciences and Physical Engineering at Czech Technical University in Prague in 2017. From 2017 to 2019, he worked as a Junior Researcher at ELI Beamlines in Dolní Břežany, Czechia, focusing on computational simulations in laser physics. He is currently pursuing a Ph.D. in Image Processing at the Institute of Information Theory and Automation, in collaboration with Czech Technical University, with research centered on inverse problems, particularly demosaicing and deconvolution in digital photography. He is also pursuing an M.A. in Leadership and Practical Theology at TCM International Institute, Austria. In 2021-2022, he was a Fulbright Visiting Scholar at Washington University in St. Louis, where he advanced his research in implicit neural representations.

Filip Šroubek received the M.Sc. degree in computer science from the Czech Technical University, Prague, Czech Republic in 1998 and the Ph.D. degree in computer science from Charles University, Prague, Czech Republic in 2003. From 2004 to 2006, he was on a postdoctoral position in the Instituto de Optica, CSIC, Madrid, Spain. In 2010/2011 he received a Fulbright Visiting Scholarship at the University of California, Santa Cruz. In 2014, he became a research professor in Physico-Mathematical Sciences (Informatics and Cybernetics) at the Czech Academy of Sciences. In 2016, he became an associate professor at the Faculty of Mathematics and Physics, Charles University. Currently he is the deputy head of the Department of Image Processing.

Jan Flusser received the M.Sc. degree in mathematical engineering from the Czech Technical University, Prague, Czech Republic, in 1985; the PhD degree in computer science from the Czechoslovak Academy of Sciences in 1990; and the DrSc. degree in technical cybernetics in 2001. Since 1985 he has been with the Institute of Information Theory and Automation, Czech Academy of Sciences, Prague. In 1995–2007, he was holding the position of a head of Department of Image Processing. In 2007–2017, he was a Director of the Institute. Since 2017 he has been at the position of Research Director. He is a full professor of computer science at the Czech Technical University, Faculty of Nuclear Science and Physical Engineering, and at the Charles University, Faculty of Mathematics and Physics, Prague, Czech Republic, where he gives undergraduate and graduate courses on Digital Image Processing, Pattern Recognition, and Moment Invariants and Wavelets. Jan Flusser's research interest covers moments and moment invariants, image registration, image fusion, multichannel blind deconvolution and super-resolution imaging. He has authored and coauthored more than 200 research publications in these areas, including the monographs Moments and Moment Invariants in Pattern Recognition (Wiley, 2009) and 2D and 3D Image Analysis by Moments (Wiley, 2016). In 2010, Jan Flusser was awarded by the SCOPUS 1000 Award. He received the Felber Medal of the Czech Technical University for excellent contribution to research and education in 2015 and the Praemium Academiae of the Czech Academy of Sciences for outstanding researchers in 2017.

Co-author Statements

CO-AUTHOR STATEMENT 1

I hereby declare that I am aware that the work in the paper/manuscript entitled:

Filip Šroubek, Tomáš Kerepecký, and Jan Kamenický, "Iterative Wiener filtering for deconvolution with ringing artifact suppression," in *2019 27th European Signal Processing Conference (EUSIPCO)*. September 2019, pp. 1–5, IEEE

of which I am the main author, will form a part of the PhD dissertation by PhD student:

TOMÁŠ KEREPECKÝ

Description of the involvement:

The main contribution of Tomáš Kerepecký in this work was performing experiments and providing final feedback on the manuscript. These efforts helped refine the overall study and supported the validation of its methodologies.

Name:	Filip Šroubek
Institution:	Institute of Information Theory and Automation Czech Academy of Sciences
Date:	March 7th, 2025
Signature:	

CO-AUTHOR STATEMENT 2

I hereby declare that I am aware that the work in the paper/manuscript entitled:

Tomáš Kerepecký and Filip Šroubek, "D3net: Joint demosaicking, deblurring and deringing," in *2020 25th International Conference on Pattern Recognition (ICPR)*. January 2021, pp. 1–8, IEEE

of which I am the co-author, will form a part of the PhD dissertation by PhD student:

TOMÁŠ KEREPECKÝ

Description of the involvement:

The main contribution of Tomáš Kerepecký in this paper involved extensive research on deep unrolling, formulating the core idea, and the development of the methods described. Tomáš Kerepecký was responsible for preparing data from the relevant databases, thoroughly testing all mentioned algorithms, and drafting the initial version of the manuscript. The final version was then refined collaboratively by both authors.

Name:	Filip Šroubek
Institution:	Institute of Information Theory and Automation Czech Academy of Sciences
Date:	March 7th, 2025
Signature:	

CO-AUTHOR STATEMENT 3

I hereby declare that I am aware that the work in the paper/manuscript entitled:

Kerepecký, Tomáš, Jiaming Liu, Xuan Wei Ng, David W. Piston, and Ulugbek S. Kamilov, "Dual-cycle: Self-supervised dual-view fluorescence microscopy image reconstruction using cyclegan," in *2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. June 2023, pp. 1–5, IEEE

of which I am the co-author, will form a part of the PhD dissertation by PhD student:

TOMÁŠ KEREPECKÝ

Description of the involvement:

The main contribution of Tomáš Kerepecký in this work consisted of outlining the core conceptual framework, conducting and analyzing the principal experiments, and preparing the final draft of the manuscript. This involved designing the study methodology, gathering and evaluating the data, and ensuring the coherence of the paper's overall structure and presentation.

Name:	Jiaming Liu
Institution:	Department of Radiology Stanford University
Date:	March 7th, 2025
Signature:	Jiaming Liu

CO-AUTHOR STATEMENT 4

I hereby declare that I am aware that the work in the paper/manuscript entitled:

Tomáš Kerepecký, Filip Šroubek, Adam Novozámský, and Jan Flusser, "Nerd: Neural field-based demosaicking," in *2023 IEEE International Conference on Image Processing (ICIP)*. October 2023, pp. 1735–1739, IEEE

of which I am the co-author, will form a part of the PhD dissertation by PhD student:

TOMÁŠ KEREPECKÝ

Description of the involvement:

The main contribution of Tomáš Kerepecký in this paper involved extensive research on implicit neural representations, formulating the core idea, the development of the methods described, and thorough testing of the proposed algorithms. Tomáš Kerepecký was responsible for preparing relevant data, conducting experiments, and drafting the initial version of the manuscript. The final paper was then refined collaboratively with other co-authors, incorporating additional insights to achieve the published version.

Name:	Filip Šroubek
Institution:	Institute of Information Theory and Automation Czech Academy of Sciences
Date:	March 7th, 2025
Signature:	

CO-AUTHOR STATEMENT 5

I hereby declare that I am aware that the work in the paper/manuscript entitled:

Tomáš Kerepecký, Filip Šroubek, and Jan Flusser, "Implicit neural representation for image demosaicking," *Digital Signal Processing*, p. 105022, 2025, Elsevier

of which I am the co-author, will form a part of the PhD dissertation by PhD student:

TOMÁŠ KEREPECKÝ

Description of the involvement:

The main contribution of Tomáš Kerepecký in this paper involved formulating the core idea, conducting extensive research on implicit neural representations, and developing the methods described. Tomáš Kerepecký was also responsible for preparing relevant data, carrying out experiments, and drafting the initial manuscript. The final version was then refined collaboratively with other co-authors, incorporating additional insights to produce the published work.

Name:	Filip Šroubek
Institution:	Institute of Information Theory and Automation Czech Academy of Sciences
Date:	March 7th, 2025
Signature: