DIVA: Deep Unfolded Network from Quantum Interactive Patches for Image Restoration

Sayantan Dutta, Adrian Basarab, Bertrand Georgeot, Denis Kouamé

DIVA: Deep Unfolded Network from Quantum Interactive Patches for Image Restoration

Sayantan Dutta, *Student Member, IEEE,* Adrian Basarab, *Senior Member, IEEE,* Bertrand Georgeot, and Denis Kouamé, *Senior Member, IEEE*

Abstract—This paper presents a deep neural network called DIVA unfolding a baseline adaptive denoising algorithm (De-QuIP), relying on the theory of quantum many-body physics. Furthermore, it is shown that with very slight modifications, this network can be enhanced to solve more challenging image restoration tasks such as image deblurring, super-resolution and inpainting. Despite a compact and interpretable (from a physical perspective) architecture, the proposed deep learning network outperforms several recent algorithms from the literature, designed specifically for each task. The key ingredients of the proposed method are on one hand, its ability to handle non-local image structures through the patch-interaction term and the quantum-based Hamiltonian operator, and, on the other hand, its flexibility to adapt the hyperparameters patch-wisely, due to the training process.

Index Terms—Quantum many-body interaction, Schrödinger equation, Unfolding, Deep learning, Image restoration, Quantum image processing.

1 Introduction

Restoring a high-quality image from a degraded observation is a classic but still major challenge in imaging applications, such as medical imaging, remote sensing, low-level vision, surveillance, to cite few. Such a degradation process can be formulated as $\boldsymbol{Y} = \boldsymbol{O}\boldsymbol{X} + \boldsymbol{e}$, where, $\boldsymbol{Y}$ and $\boldsymbol{X}$ denote the low quality observation and the image of interest, respectively, $\boldsymbol{O}$ denotes the degradation operator, and $\boldsymbol{e}$ is associated with an additive noise. The goal is to recover the underlying high-quality image $\boldsymbol{X}$ from the observation $\boldsymbol{Y}$. Depending on the degradation operator $\boldsymbol{O}$, different restoration problems occur. For example, if $\boldsymbol{O}$ is the identity operator, the resulting problem is an image denoising [1], [2], [3], [4] problem. If $\boldsymbol{O}$ is a blurring operator then restoration becomes a deblurring [5], [6], [7], [8], [9], or a super-resolution (SR) task [10], [11], [12] if $\boldsymbol{O}$ includes a subsampling operator. In practice, estimating $\boldsymbol{X}$ from $\boldsymbol{Y}$ by mitigating the effect of the degradation operator $\boldsymbol{O}$ is a challenging ill-posed inverse problem. Over the past few decades, image restoration techniques have been extensively studied, yet remain an active field of research.

Traditionally, the restoration process is framed as a model-based optimization problem from a Bayesian perspective, in which the desired solution is obtained by minimizing the sum of a regularization and a data fidelity term. Over the time, numerous model-based regularizers have been introduced in the literature including total variation [13], sparsity-based transformations [1], sparse models [2], [14], [15] and spatial filtering [16], [17], in particular non-local self-similarity (NLSS) filters [18], [19], anisotropic diffusion filters [20], [21], guided filters [22], etc. In particular, non-local regularization approaches [15], [23], [24], [25] blending the NLSS and low-rank regularity, such as BM3D [3], NLM [18], LSSC [19], NCSR [15], etc., have been extensively discussed due to their state-of-the-art restoration performances. Integration of non-local information in the process of retrieving a particular region of the image is the key to the success of the NLSS models. In general, the model-based approaches are fairly successful in tackling a variety of image retrieval tasks, including proper interpretation of their roles. However, these schemes require conducting a costlier computation process and manual tuning of several hyperparameters, which are the primary challenges of these strategies.

Based on deep convolutional neural networks (CNN), deep learning (DL)-based strategies brought an alternative to the well-established model-based methods to counter such imaging problems. DL algorithms [26], [27], [28], [29], [30], [31], [32], [33], [34], [35] achieved state-of-the-art performances in recent years by learning the mapping functions from observed degraded or low-resolution (LR) images to the original or high-resolution (HR) images. CSF [36], TNRD [37], DnCNN [38], Super-ONN [39], DWDN [40], SRCNN [41], RDN [32], DRLN [42], etc., are some of the well-known DL networks with proven efficiency in image restoration over the conventional model-based approaches, exploiting a training dataset in the learning process. However, training a CNN is not straightforward. The performance largely depends on the number of layers, the kernel size and the learning rate. Deeper network structures may provide better results but exponentially increase the training complexity [28]. Thus, network structures are in most cases determined empirically, which makes them suffer from a lack of inter-

- *S. Dutta is with the IRIT, Université de Toulouse, CNRS, Toulouse INP, UT3, Toulouse, France and the Laboratoire de Physique Théorique, Université de Toulouse, CNRS, UPS, France. E-mail: sayantan.dutta@irit.fr.*
- *A. Basarab is with the Université de Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1294, F-69621, Villeurbanne, France.*
- *B. Georgeot is with the Laboratoire de Physique Théorique, Université de Toulouse, CNRS, UPS, France.*
- *D. Kouamé is with the IRIT, Université de Toulouse, CNRS, Toulouse INP, UT3, Toulouse, France.*

(Corresponding author: Sayantan Dutta).

pretation of their true functionality.

Benefiting from CNN's powerful representation ability, a new concept, known as unfolding [43], gathering the advantages of both model and DL-based approaches, is currently gaining more attention due to its explanatory properties. The main idea of such frameworks is to construct a DL network starting from a classical algorithm. This approach has recently been successfully explored in the literature, leading to superior restoration performance over the classical peer, such as BM3D-NET [44], LKSVD [45], FBPConvNet [46], DRED-DUN [47], CORONA [48], USRNet [49], to cite few.

In this work, we advocate novel CNN architectures for image restoration problems, unfolding our recently introduced quantum mechanics-based adaptive denoising algorithm called De-QuIP [50], [51]. In the last decade, quantum mechanics-based frameworks have been explored in the field of image processing and analysis, such as the single-particle quantum theory, especially for image segmentation [52], [53], denoising [54], [55], deblurring [56], [57] or others [58]. Despite the promising performances, these single-particle-based frameworks cannot benefit from the structural features of the image like NLSS algorithms. In contrast, De-QuIP is based on the theory of many-body quantum systems, where each image patch behaves like a single particle system and interacts with its neighbors. This phenomenon of interaction preserves the image similarity/features from a local neighborhood. Indeed, absorption of this concept of interaction in De-QuIP brings an intrinsic non-local structure to the algorithm that notably enhances the denoising performance and has been extensively studied in [51]. Despite its interesting performances, De-QuIP struggles with costly computational processes (*e.g.*, hyperparameters tuning and eigenvalue decomposition) like many other model-based schemes, which may limit its practical use.

In this paper, we introduce a novel DL network unfolding the baseline De-QuIP algorithm, denoted as DIVA (Deep denoising by quantum InteractiVe pAtches) for image denoising problem. We further extend the network architecture to conduct a general image restoration task. The inclusion of the quantum interaction theory brings a non-local structure to the proposed CNN architecture. Indeed, in our depicted DL models, the fundamental aspects of quantum theory from the baseline De-QuIP algorithm are essentially preserved. Furthermore, the DL model efficiently resolves the hyperparameter tuning problem of the original De-QuIP scheme, harnessing the power of back-propagation. The integration of the key attributes of DL and quantum theory significantly enhances the functionality of our proposed networks due to their intrinsic versatility and enables our models to exhibit state-of-the-art performances for several restoration tasks.

An initial illustration of this work was presented as a conference report that portrays preliminary results on Gaussian denoising [59]. Herein, we extend our preliminary model to a robust generalized formalism by incorporating additional contents in significant ways: (i) we extend the initially proposed DL model, primarily designed for denoising, to more complex image restoration tasks such as deblurring, super-resolution and inpainting, with a resilient generalized network architecture; (ii) we conduct a detailed investigation regarding the network diagram and add considerable analysis of the incorporated quantum background, tunable parameter number, and run time; (iii) we report a comprehensive survey of image restoration performance against benchmark methods for various imaging problems.

The remainder of the paper is organized as follows. Sec. 2 reminds briefly the concepts of the baseline De-QuIP algorithm for self-consistency reasons. Sec. 3 first presents the proposed DIVA network for denoising and then extends it to an advanced model for other imaging tasks. The experimental settings and extensive evaluations are reported in Sec. 4. Sec. 5 outlines the overall remarks and possible future perspectives. Finally, Sec. 6 draws the conclusions.

2 Brief Review of Quantum-Interactive-Patches-Based Denoising

To facilitate the understanding of the proposed method, we briefly revisit the baseline De-QuIP algorithm for image denoising and its main properties.

2.1 The De-QuIP Scheme

Built on an underlying nonlocal architecture, De-QuIP [50], [51] offers an adaptive way of image denoising based on the theory of quantum many-body interaction. The theory of quantum many-body physics subscribes many-body quantum systems, where inevitably particle-to-particle interactions emerge. De-QuIP provides a framework for extending this concept of interaction to imaging problems. Effectively, De-QuIP divides an image into small patches, and each image patch acts as a single-particle system while interacting with its neighbors, i.e., with neighboring patches, inside the whole image, similarly to a many-body system. Indeed, these interactions between neighbors reflect their mutual similarities that enhance the denoising performance of De-QuIP significantly.

Similar to any denoising method, the goal is to estimate the underlying clean image $\boldsymbol{X} \in \mathbb{R}^{M\times N}$ from a noisy observation $\boldsymbol{Y} \in \mathbb{R}^{M\times N}$. The respective vectorized representations of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are denoted by $\boldsymbol{x} \in \mathbb{R}^{MN}$ and $\boldsymbol{y} \in \mathbb{R}^{MN}$ in lexicographical order. Based on the many-body quantum physics, the primary idea of De-QuIP algorithm is to construct an adaptive transformation using the wave solutions of the Schrödinger equation $H\boldsymbol{\psi}(z) = E\boldsymbol{\psi}(z)$, where the wave function $\boldsymbol{\psi}(z)$ describes a particle with energy E in a potential V, z being the spatial coordinate. In a many-body system, denoting by I the interaction, the Hamiltonian operator is $H = -(\hbar^2/2m)\nabla^2 + V + I$, where ∇ and $(\hbar^2/2m)$ are respectively the gradient operator and a function of the Planck's constant (this function acts as a hyperparameter in this formalism). For this patch-based imaging scheme, the potential V is represented by the original pixels' values of the image patch and the patch-similarity measures act as the interaction I. The set of eigenvectors of the Hamiltonian operator gives the adaptive transformations for the respective patch. Thus, for a system with multiple particles, the Hamiltonian operator $\boldsymbol{H}_a$ for the a-th patch is defined by:

$$\boldsymbol{H}_a = -(\hbar^2/2m)\nabla^2 + \boldsymbol{J}_a + \boldsymbol{I}_a, \tag{1}$$

where $\boldsymbol{J}_a$ and $\boldsymbol{I}_a$ are respectively the pixels' values and interaction term for the a-th patch. The corresponding set of

eigenvectors $\boldsymbol{B}_a$ of $\boldsymbol{H}_a$ acts as the quantum adaptive basis for the a-th patch. The key steps of De-QuIP algorithm are as follows.

Patch extraction: The patch extraction step primarily uncoils small patches from the observed image and assimilates their neighbors into their respective local groups. Let us denote by $\boldsymbol{J}_a$ a patch of size n^2 whose upper-left pixel position is a, and by Ω the set containing all such patches extracted from the image $\boldsymbol{y}$. For all $\boldsymbol{J}_a \in \Omega$, one creates a window of size $W \times W$ centered on $\boldsymbol{J}_a$ and accumulates all patches inside the window in a set denoted by $S_{\boldsymbol{J}_a}$ to create local groups.

Total interaction: The goal of the interaction step is to preserve local structures/similarities by exploiting the local groups through a notion akin to the interactions in quantum mechanics. This step computes the interactions $\boldsymbol{L}_{ab}$, for all $\boldsymbol{J}_b \in S_{\boldsymbol{J}_a}$ and all $\boldsymbol{J}_a \in \Omega$, using power laws of physics [51], *i.e.*, interaction is linearly proportional to the pixel-wise difference $\boldsymbol{K}_{ab}^k = |\boldsymbol{J}_a^k - \boldsymbol{J}_b^k|$ for $k = 1, \cdots, n^2$ and inversely proportional to the square of the Euclidean distance D_{ab} between the patches. Summing over b gives the total interaction for the a-th patch

$$\boldsymbol{I}_a = p \sum_b \boldsymbol{L}_{ab} = p \sum_b \boldsymbol{K}_{ab}/D_{ab}^2, \quad \forall \boldsymbol{J}_a \in \Omega. \tag{2}$$

In this construction the proportionality constant p acts as a hyperparameter.

Hamiltonian operator and adaptive basis: This step formulates the energy or Hamiltonian operators of the extracted patches by incorporating their total interaction with their neighbors in the local group using (1). The associated set of eigenvectors $\boldsymbol{B}_a$ of the Hamiltonian operator $\boldsymbol{H}_a$ operates as the adaptive basis for the current image patch $\boldsymbol{J}_a$.

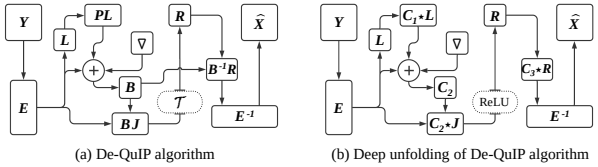
Thresholding: The thresholding is processed on the coefficients resulting from patch projections onto their respective adaptive basis. Hence, the noise is attenuated by projecting $\boldsymbol{J}_a$ onto $\boldsymbol{B}_a$ and performing hard/soft-thresholding $\mathcal{T}$ in energy. Finally, reverse projecting the truncated coefficients reinstates the denoised patch $\hat{\boldsymbol{J}}_a, \forall \boldsymbol{J}_a \in \Omega$.

Patch accumulation: This step accumulates all the denoised patches to their original positions and normalizes them to reconstruct the estimated denoised image $\hat{\boldsymbol{x}}$. In the following, the patch extractor operator is denoted by $\boldsymbol{E}$, while the operation of accumulating the patches to form the denoised image is denoted by $\boldsymbol{E}^{-1}$.

In the De-QuIP framework, the preserved spatial information by the patch interaction phenomenon coherently passes through the Hamiltonian operator to the quantum adaptive basis and enables the algorithm to handle denoising tasks regardless of the noise intensity, statistics and image nature. Its application field is not limited to denoising tasks [50], [51], and its efficiency has been illustrated in various imaging problems such as despeckling [60] and super-resolution [61]. Fig. 1(a) depicts the De-QuIP architecture, where interaction, proportionality constant, adaptive basis and thresholded coefficients are denoted by $\boldsymbol{L}$, $\boldsymbol{P}$, $\boldsymbol{B}$ and $\boldsymbol{R}$ respectively.

2.2 Shortcomings of De-QuIP

The major challenge of De-QuIP is its high computational cost of tuning the hyperparameters p, $(\hbar^2/2m)$ and energy threshold. In [51], the influence of these hyperparameters and strategies to optimize their values were discussed. Despite some rules that guided the choice of the hyperparameters, they remain general for the whole image and are not optimized to be applied patch-wise. Although De-QuIP demonstrates favorable outcomes despite these drawbacks, the intrinsic non-local architecture of the algorithm raises an obvious question of assigning patch-dependent hyperparameter values, which can further enhance the adaptability of the model. However, manually tuning all the hyperparameters separately for each patch is practically impossible. The power of DL architecture removes this barrier by involving many parameters that can be learned during the training process.



(a) De-QuIP algorithm (b) Deep unfolding of De-QuIP algorithm

Fig. 1. Architectural comparison between De-QuIP and its DL counterpart.

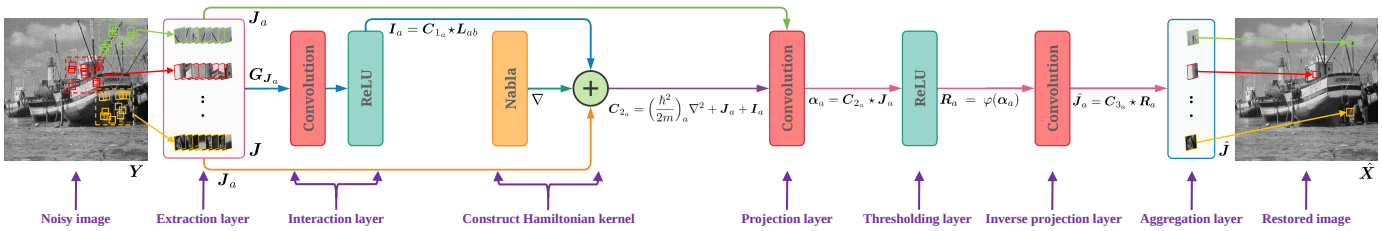
Another challenge of De-QuIP is the computationally-expensive task of adaptive basis vector computation from the Hamiltonian operator. Furthermore, this adaptive basis is exploited to calculate the projection coefficients, bringing additional computational burden. A deep learning model can bypass all these bottlenecks by directly estimating the projection coefficients with the help of convolutional kernels. The subsequent section focuses on this deep-learning prospect of the De-QuIP algorithm, the main contribution of this paper.

3 Proposed Deep Architectures for Image Restoration

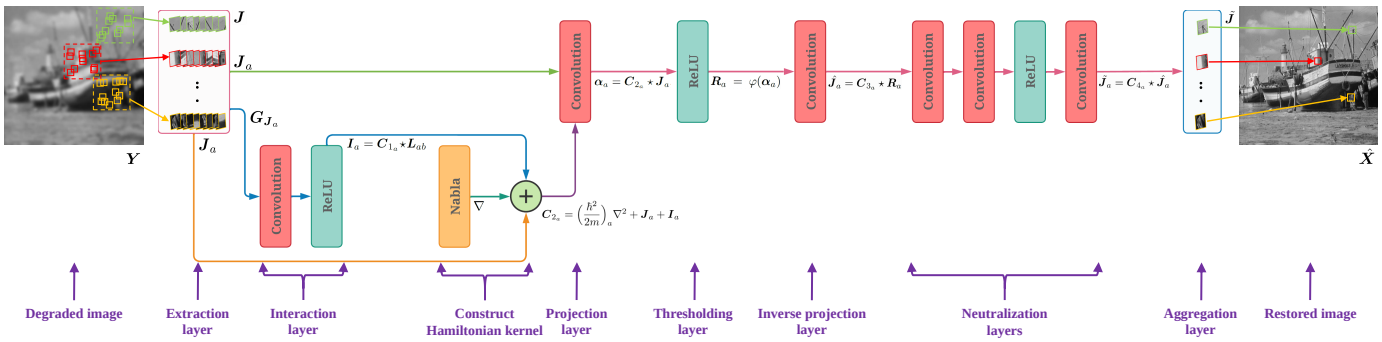
This section presents deep unfolding strategies for image restoration problems built on the baseline De-QuIP algorithm. Depending on the image degradation operator $\boldsymbol{O}$, various imaging problems arise. If $\boldsymbol{O}$ is an identity operator, the image restoration problem is equivalent to a denoising task, whereas, depending on $\boldsymbol{O}$, it may turn into deblurring, super-resolution or inpainting, addressed herein. In the following, two deep architectures are introduced: the first addresses denoising, and the second more complex image restoration tasks. The first proposed network, referred to as DIVA, is a direct translation of the baseline De-QuIP algorithm into a deep learning model to handle denoising. To handle non-identity degradation operators $\boldsymbol{O}$, DIVA architecture is slightly modified and denoted by DIVA advanced (DIVA-A). The subsequent subsections illustrate these two network architectures.

3.1 Proposed DIVA Architecture

The main idea behind the proposed unfolding strategy is to replace the matrix multiplication steps in De-QuIP by convolution layers. The analogy between the original algorithm and its unfolded version is illustrated in Fig. 1.



(a) Proposed DIVA network for image denoising.



(b) Proposed DIVA-A network for image restoration.

Fig. 2. The architectures of the proposed deep learning models. The corresponding operations for a patch $\boldsymbol{J}_a$ are indicated next to each block.

The proposed DIVA network primarily stands upon eight main pillars.

Extraction layer: Similar to the De-QuIP algorithm, the extraction layer in DIVA assembles all patches from a local window of size $W \times W$ centered at $\boldsymbol{J}_a$ in a local patch-group denoted as $S_{\boldsymbol{J}_a}, \forall \boldsymbol{J}_a \in \Omega$. Let the cardinality of $S_{\boldsymbol{J}_a}$ be κ, $\forall \boldsymbol{J}_a \in \Omega$ and ζ be the cardinality of Ω. The patch extraction operation from the local window can be defined as a matrix multiplication by $\boldsymbol{E}_{\boldsymbol{J}_a} \in \mathbb{R}^{n^2\kappa \times MN}$ for each $\boldsymbol{J}_a$. Therefore, mathematically, $\boldsymbol{G}_{\boldsymbol{J}_a} = \boldsymbol{E}_{\boldsymbol{J}_a}\boldsymbol{y}$, where $\boldsymbol{G}_{\boldsymbol{J}_a} \in \mathbb{R}^{n^2\kappa}$ is the concatenated vectorized local patch group for each $\boldsymbol{J}_a$. Thus, for the whole image, the patch extractor operator $\boldsymbol{E} \in \mathbb{R}^{\zeta n^2 \kappa \times MN}$ is constructed by concatenating $\boldsymbol{E}_{\boldsymbol{J}_a} \in \mathbb{R}^{n^2\kappa \times MN}$ $\forall \boldsymbol{J}_a \in \Omega$. Finally, $\boldsymbol{J}_a$ and $\boldsymbol{G}_{\boldsymbol{J}_a}$ $\forall \boldsymbol{J}_a \in \Omega$ are concatenated and reshaped to construct matrices $\boldsymbol{J} \in \mathbb{R}^{\zeta \times n^2}$ and $\boldsymbol{G} \in \mathbb{R}^{\zeta \times n^2 \kappa}$, further considered as inputs for the next layer.

Interaction layer: This layer computes the interactions between patches for each local group $\boldsymbol{G}_{\boldsymbol{J}_a}$ following the power laws discussed in Sec. 2.1. But rather than considering a fixed hyperparameter value p as in (2), for each local group $\boldsymbol{G}_{\boldsymbol{J}_a}$ a different set of p^k_{ab} values is assigned for each pixel k ($k = 1, \cdots, n^2$) and patch b ($b = 1, \cdots, \kappa; \neq a$) respectively. Therefore, the total interaction can be expressed as

$$\boldsymbol{I}^k_a = \sum_{b=1, b\neq a}^{\kappa} p^k_{ab} \frac{\boldsymbol{K}^k_{ab}}{D^2_{ab}} = \sum_{b=1, b\neq a}^{\kappa} p^k_{ab} \boldsymbol{L}^k_{ab}, \text{for each } \boldsymbol{G}_{\boldsymbol{J}_a}. \quad (3)$$

In matrix notation, $\boldsymbol{I}_a = \boldsymbol{P}_{ab}\boldsymbol{L}_{ab}$, for each $\boldsymbol{G}_{\boldsymbol{J}_a}$, where $\boldsymbol{I}_a \in \mathbb{R}^{n^2}$, $\boldsymbol{P}_{ab} \in \mathbb{R}^{n^2 \times n^2(\kappa-1)}$, and $\boldsymbol{L}_{ab} \in \mathbb{R}^{n^2(\kappa-1)}$ respectively denote the total interaction for patch $\boldsymbol{J}_a$, proportionality constant in local group $\boldsymbol{G}_{\boldsymbol{J}_a}$, and interaction between $\boldsymbol{J}_a$ and $\boldsymbol{J}_b$ patches. At this point, the main challenge is to tune the values of $\boldsymbol{P}_{ab}$ so that $\boldsymbol{I}_a$ can efficiently preserve the local information and incorporate them into the Hamiltonian. One may note that this process is equivalent to a convolution between $\boldsymbol{L}_{ab}$ and a learnable filter $\boldsymbol{C}_{1_a}$ of appropriate size. Hence, the local operation in the layer is,

$$\boldsymbol{I}_a = \boldsymbol{C}_{1_a} \star \boldsymbol{L}_{ab}, \quad \forall \boldsymbol{G}_{\boldsymbol{J}_a}, \quad (4)$$

where $\star$ indicates the convolution product. This convolution layer is followed by a Rectified Linear Unit (ReLU) to truncate the insignificant contributions of the interactions. Finally, by concatenating $\boldsymbol{I}_a, \forall \boldsymbol{G}_{\boldsymbol{J}_a}$, one obtains $\boldsymbol{I} \in \mathbb{R}^{\varsigma \times n^2}$.

Construct the Hamiltonian kernel: In the baseline architecture of De-QuIP, for each $\boldsymbol{J}_a$, the Hamiltonian/energy operator depends on the hyperparameter $(\hbar^2/2m)$ (*i.e.*, the Planck constant), the total interaction $\boldsymbol{I}_a$ and the original potential/pixels' values $\boldsymbol{J}_a$. This operator gives the adaptive basis $\boldsymbol{B}_a$ on which the noisy patch $\boldsymbol{J}_a$ is projected. The integration of the local interactions, bringing a non-local dimension to the formalism, is a core feature of De-QuIP.

This physical attribute of the Hamiltonian operator is preserved in this step by constructing a kernel

$$\boldsymbol{C}_{2_a} = (\hbar^2/2m)_a \nabla^2 + \boldsymbol{J}_a + \boldsymbol{I}_a, \quad \forall \boldsymbol{J}_a \in \Omega, \quad (5)$$

where different learnable values of $(\hbar^2/2m)_a$ are allotted instead of a constant one. This kernel $\boldsymbol{C}_{2_a}$ mimics the role of the adaptive basis $\boldsymbol{B}_a$ in the next layer in the shadow of a convolutional process. Note that throughout the learning process the kernel retains its original Hamiltonian structure which is a key ingredient of the original De-QuIP algorithm.

Projection layer: This layer deals with the adaptive transformation of the noisy patch $\boldsymbol{J}_a$ on the associative quantum adaptive basis $\boldsymbol{B}_a$ for each $\boldsymbol{J}_a \in \Omega$, *i.e.*, $\boldsymbol{\alpha}_a = \boldsymbol{B}_a\boldsymbol{J}_a$, where $\boldsymbol{\alpha}_a \in \mathbb{R}^{n^2}$ are the projection coefficients of $\boldsymbol{J}_a$. In our proposed deep architecture, this process is conducted by performing convolution operations on $\boldsymbol{J}_a$ using a learnable kernel $\boldsymbol{C}_{2_a}$ built in the previous step, as:

$$\boldsymbol{\alpha}_a = \boldsymbol{C}_{2_a} \star \boldsymbol{J}_a, \quad \forall \boldsymbol{J}_a \in \Omega. \quad (6)$$

Exploiting the power of a deep network, the convolution operation (6) removes the algebraically expensive processes, such as the computation of adaptive basis and projection coefficients, and uses the training dataset to directly estimate the projection coefficients. Finally, all $\boldsymbol{\alpha}_a$ are concatenated to form $\boldsymbol{\alpha} \in \mathbb{R}^{\zeta \times n^2}$, serving as input to the next layer.

Thresholding layer: The thresholding layer handles the process of trimming the projection coefficients $\boldsymbol{\alpha}$. A nonlinear ReLU activation function φ is used as a thresholding function, which makes the denoising process more robust by adding more flexibility than the baseline scheme, where thresholding was done in energy. Therefore, the shrunk coefficients $\boldsymbol{R}_a = \varphi(\boldsymbol{\alpha}_a)$ are obtained for each $\boldsymbol{J}_a \in \Omega$, further concatenated into $\boldsymbol{R} \in \mathbb{R}^{\zeta \times n^2}$, before stepping to the next layer.

Inverse projection layer: In the original algorithm the denoised patch $\hat{\boldsymbol{J}}_a$ is revamped from the reduced coefficients $\boldsymbol{R}_a$ by inverse projecting onto the quantum adaptive basis $\boldsymbol{B}_a$ for each $\boldsymbol{J}_a \in \Omega$, *i.e.*, $\hat{\boldsymbol{J}}_a = \boldsymbol{B}_a^{-1}\boldsymbol{R}_a$. This step resembles a convolution process of $\boldsymbol{R}_a$ with a learnable kernel $\boldsymbol{C}_{3_a}$. Hence, the mathematical operation of the layer is defined as

$$\hat{\boldsymbol{J}}_a = \boldsymbol{C}_{3_a} \star \boldsymbol{R}_a, \quad \forall \boldsymbol{J}_a \in \Omega. \tag{7}$$

Finally, before proceeding to the following layer, all outputs $\hat{\boldsymbol{J}}_a$ are concatenated to $\hat{\boldsymbol{J}} \in \mathbb{R}^{\zeta \times n^2}$.

Note that in the baseline algorithm, the operator used in the inversion step was the inverse of the adaptive basis used in the projection process. This mutual dependence is highlighted in Fig. 1(a) by an arrow. In the proposed deep unfolded network, the learnable kernels $\boldsymbol{C}_{2_a}$ and $\boldsymbol{C}_{3_a}$ replaced respectively the original and inverse adaptive basis. The convolutional operations are useful to learn these kernels independently and are illustrated by removing the arrow in Fig. 1(b).

Aggregation layer: Akin to the De-QuIP scheme, this layer conducts the $\boldsymbol{E}^{-1}$ operation to accumulate all the denoised patches and put them back to their initial positions in the image after normalization, and reconstructs the denoised image $\hat{\boldsymbol{x}}$. Note that overlapping patches are considered in the proposed formalism. Fig. 2(a) illustrates the proposed DIVA network architecture, highlighting all the layers described above.

3.2 Proposed DIVA Advanced Network

An advanced version of the DIVA network introduced in the previous section is proposed hereafter. This network slightly differs from DIVA, and is adapted to image restoration tasks involving, in addition to noise, other degradation effects on the observed image $\boldsymbol{y}$, such as blur, pixel resolution loss or missing pixels. In the case of additive Gaussian noise, the effect of the noise and the additional degradation can be considered independently. Therefore, DIVA network of Sec. 3.1 is extended by additional convolutional layers after the inversion process. In this way, the first part of the network eliminates the noise, and the second part neutralizes the effects of a nonidentity degradation operator.

The modified network referred to as DIVA-A primarily plugs a neutralization layer between the inverse projection and aggregation layers, as highlighted in Fig. 2(b).

Neutralization layer: This layer corresponds to the restoration of the patch $\tilde{\boldsymbol{J}}_a$ by eliminating the influence of a degradation operator $\boldsymbol{O}_a$ from the patch $\hat{\boldsymbol{J}}_a$ reconstructed in the inverse projection layer for each $\boldsymbol{J}_a \in \Omega$, *i.e.*, $\tilde{\boldsymbol{J}}_a = \boldsymbol{O}_a^{-1}\hat{\boldsymbol{J}}_a$, where $\boldsymbol{O}_a$ denotes a degradation operator acting on a patch $\boldsymbol{J}_a, \forall \boldsymbol{J}_a \in \Omega$. This operation is analogous to a convolutional process of $\hat{\boldsymbol{J}}_a$ with a learnable kernel $\boldsymbol{C}_{4_a}$, defined as

$$\tilde{\boldsymbol{J}}_a = \boldsymbol{C}_{4_a} \star \hat{\boldsymbol{J}}_a, \quad \forall \boldsymbol{J}_a \in \Omega. \tag{8}$$

The proposed network conducts this operation by adding three convolutions with multiple learnable filters, and one ReLU function to remove any unwanted contribution (see Fig. 2(b)). The power of a CNN architecture is used to learn these filters that mimic the role of a degradation operator in this layer.

Before proceeding to the aggregation layer, all $\tilde{\boldsymbol{J}}_a$ are concatenated to obtain $\tilde{\boldsymbol{J}} \in \mathbb{R}^{\zeta \times n^2}$. Similar to the DIVA network, the aggregation layer assembles all recovered patches and outputs the restored image $\hat{\boldsymbol{x}}$.

3.3 Loss Function

The proposed networks are trained end-to-end, where the mean squared error (MSE) between the predicted and original residuals is adopted as the loss function [44]:

$$\mathcal{L}_\Theta = \frac{1}{MN} \left\| \mathcal{R}(\hat{\boldsymbol{x}}; \Theta) - (\boldsymbol{y} - \boldsymbol{x}) \right\|_2^2, \tag{9}$$

where $\mathcal{R}(\hat{\boldsymbol{x}}; \Theta)$ denotes the predicted residual by the network with parameter set Θ. This loss function allows our models to learn the disorders present in a distorted image without bothering about the features of the true image. Note that it is possible to use different other loss functions. [1]

4 Experimental Results

In this section, we analyze the proposed networks and illustrate their performance in various image restoration tasks, such as image denoising, deblurring, SR, and inpainting.

Sec. 4.1 briefly summarizes the experimental settings used in the different contexts. Sec. 4.2 gives an overview of various benchmark methods considered for comparison purposes. An ablation study with/without considering the interaction layer and the Hamiltonian kernel within the proposed networks is conducted in Sec. 4.3, with an additional discussion on the parameter number, run time, and the depth of the network. Finally, Sec. 4.4 presents a quantitative and qualitative evaluation of our DL models on various image restoration problems.

4.1 Experimental Settings

4.1.1 Image Denoising

Training data. The proposed DIVA network was trained for the Gaussian denoising task following [29], [37], [38], over a set of 400 gray-scale images of size 180×180 extracted from BSD400 dataset. All images were contaminated with additive white Gaussian noise (AWGN) with standard deviation σ, following two configurations: known

1. The Python code of the proposed the trained networks are available at github.com/SayantanDutta95/

and unknown σ. For the case of known σ, the training was conducted individually over six known noise levels, for $\sigma = 10, 15, 25, 50, 75$ and 100. To tackle an unknown noise level, DIVA was also trained blindly for a range of noise levels corresponding to $\sigma \in [5, 40]$. The corresponding model is referred as DIVA-blind.

Testing data: The trained networks were tested on five standard benchmark datasets Set12, BSD68, Kodak, LIVE1 and Urban100, widely-used for denoising problems [29], [38].

4.1.2 Image Deblurring

Training data. DIVA-A was trained separately for two types of blur kernels, *i.e.*, motion and Gaussian blur, using the recently released high-quality dataset DIV2K [62] that consists of 800 images. Eight real motion blur (MB) kernels [47], [63] and three Gaussian blur (GB) kernels [64] were considered with AWGN.

Testing data: The models trained for motion blur were tested on four benchmark datasets Set10, Levin, Sun *et al.*, and Set12, used in [30], [47]. The BSD100 and Set16 datasets were considered for the Gaussian case, following [64].

4.1.3 Single Image Super-Resolution

Training data. Similar to the deblurring model, the high-quality DIV2K [62] dataset was used as training data for image SR application. Two degradation models were used to simulate LR images for network training: (i) bicubic downsampling (BD), and (ii) Gaussian downsampling (GD). The scaling factor was set to x2, x3, and x4. For BD case [42], a LR image was simulated from the HR image by adopting Matlab *imresize* function, whereas for GD scenario, the HR image was blurred by a Gaussian kernel of size 7×7 with standard deviation 1.6 before downsampling, similar to [42].

Testing data: For testing, four widely-used benchmark datasets for image SR problem [42], [65], [66] Set5, Set14, BSD100, and Urban100, were used.

4.1.4 Image Inpainting

Training data: The same 400 gray-scale images [38] exploited by the denoising model were used to conduct the training of the proposed DIVA-A model for image inpainting. Random pixel missing model was considered to generate LR images from HR ones. 20%, 50% and 80% rates of missing pixels were used.

Testing data: Datasets Set5 and Set12 were used to evaluate the trained inpaining networks.

4.1.5 Quantitative Metrics

For the purpose of quantitative evalution, the peak-signal-to-noise-ratio (PSNR) and the structural similarity (SSIM) computed between the true and the restored images were used.

4.1.6 Training Settings

All HR and simulated LR images were clipped between 0 and 1. The patch size was set to $n = 15$ with a local window of size $W = 35$ for the proposed image denoising model with known σ. For DIVA-blind and inpainting applications, these parameters were slightly modified to $n = 25$ and $W = 50$. For deblurring and SR, larger patch and window sizes were used, $n = 35$ and $W = 70$, to preserve more spatial information from the local neighborhood. Finally, all LR-HR patch pairs were augmented randomly by rotating 90 degree and flipping horizontally or vertically to generate training data pairs. The proposed models were trained in a supervised manner by exploiting these patch-pairs.

To conduct the training, the ADAM optimizer with a mini-batch size of 128 was employed. More precisely, the models were trained with an exponentially decaying learning rate ranging from 10^{-3} to 10^{-6} over 60 epochs. The proposed network architectures were implemented under the Keras framework, and trained using NVIDIA GTX 1080 Ti GPU. The training process took about 6 hours for DIVA and 12 hours for DIVA-A to reach convergence for each experiment.

4.2 Comparison Methods

This subsection regroups the state-of-the-art methods used to conduct a comprehensive comparison to illustrate the potential of the proposed models in various imaging problems.

4.2.1 Image Denoising

The residual learning-based DnCNN [38] model is the benchmark for AWGN denoising, and its superiority over model-based (e.g., BM3D [3], NLM [18], etc.), and learning-based (e.g., TNRD [37], MLP [27], CSF [36] etc.) algorithms is well-established. In addition to DnCNN [38], our denoising model DIVA was also compared to two recently introduced DL-based networks, FFDNet [29] and IRCNN [31]. Furthermore, comparisons were carried out with a newly proposed deep unfolded scheme, BM3D-NET [44], as well as with the baseline De-QuIP [51] algorithm.

4.2.2 Image Deblurring and SR

For image deblurring and SR problems, newly published leading methods from the literature were considered to illustrate the accuracy of DIVA-A architecture. In the following, the relevant methods used for comparison purposes in different settings are listed. (i) MB model: IDD-BM3D [5], FDN [67], VEMNet [30], DWDN [40], DRED-DUN [47]; (ii) GB model: IDD-BM3D [5], Son *et al.* [68], DEBCNN [64]; (iii) BD model: LapSRN [69], MemNet [66], CARN [65], DRLN [42]; (iv) GD model: IRCNN [31], DFAN [70], RDN [32], DRLN [42].

In image SR problems, the DRLN [42] is the new benchmark in the literature. It is already shown in the seminal paper that the DRLN [42] exhibits reference state-of-the-art performance for image SR. Thus, the DRLN [42] was considered in the comparisons, thus avoiding to include all the other approaches. Similarly, for image deblurring, DWDN [40], DRED-DUN [47], and DEBCNN [64] were the best performing models in their fields. Hence, these models are selected for comparisons over other methods in the literature.

4.2.3 Image Inpainting

DIVA-A trained for image inpainting was compared against the DL prior based model IRCNN [31].

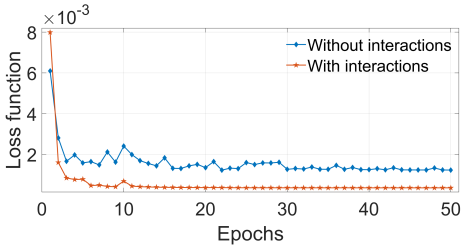


Fig. 3. Loss function (MSE) with respect to epochs. Two specific models are trained for image denoising with $\sigma = 15$, with and without integrating the interaction layer in the proposed DIVA architecture for the ablation study.

The pretrained models and the testing codes, made publicly available by the authors, were used for comparisons. Importantly, note that the proposed networks have been trained and tested exactly in the same conditions and on the same datasets as the comparison methods, thus ensuring a fair comparison.

4.3 Ablation Study and Model Analysis

This section regroups several ablation studies aiming at showing the importance of the layers inspired from quantum mechanics, and an in-depth analysis of the properties of the proposed networks.

4.3.1 *Influence of the Interaction Layer*

To show the effect of the interaction layer's integration in the Hamiltonian kernel, two versions of the DIVA network were trained for image denoising with $\sigma = 15$: the complete network as shown in Fig. 2(a), and the same network without the interaction layer. Fig. 3 plots the corresponding loss functions for these two network settings with respect to the number of epochs. One can see that using the interaction layer results into a faster and more stable convergence of the training process. Meanwhile, in the absence of this layer, a strong periodic fluctuation can be observed. This is caused by the absence of a non-local architecture in the network, which helps stabilizing the convergence process.

The same ablation study was conducted for different depths of the projection layer, using the Hamiltonian convolutional kernel constructed with and without the interaction layer. From Table 1, one can see a clear improvement in denoising performance in the presence of the interaction layer. In addition, the interaction layer significantly reduces the depth of the network by extracting the local similarities/structures from the neighboring patches. Indeed, more local information can be transferred through this non-local architecture, which helps network structures with lower depth to be more efficient. On the contrary, the network without the interaction layer improves while increasing the depth. This is expected since a deeper network consists of a larger set of tunable parameters. Although a bigger set of parameters leads to a better outcome, the learning process becomes more computationally expensive. Thus, the integration of the interaction layer enhances the network performance with a reduced computational cost, giving an edge to the proposed models.

Note that, in absence of the interaction layer in the proposed model, the network does not consider the influence of neighboring patches on the target patch and loses its non-local nature. Thus, each patch behaves as a single particle quantum system, and all patches are independent. Hence, in this circumstance, the network without an interaction layer becomes an unfolded DL scheme of the baseline QAB algorithm [55], originally proposed for image denoising based on single-particle quantum theory. Recently, in [51], it has been shown that the baseline De-QuIP outperforms the conventional QAB algorithm significantly. This observation by the traditional algorithm [51] is also consistent in our unfolded DL models, as reported in Fig. 3 and Table 1. Therefore, the consideration of the quantum interaction concept clearly enhances the model performance of both conventional and DL architectures.

4.3.2 *Depth of the Projection Layer*

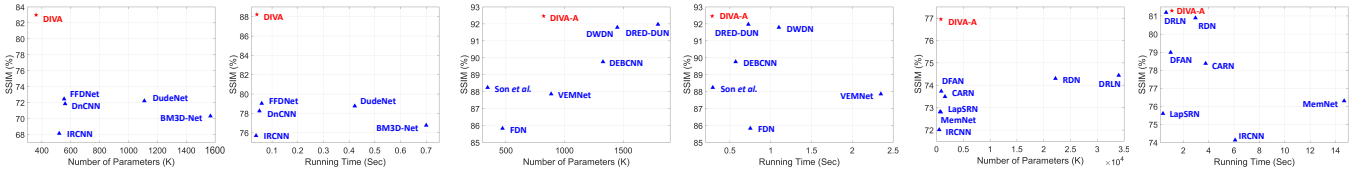
Table 1 reports denoising performance on Set12 for $\sigma = 15$ for different depths of the projection layer within the Hamiltonian kernel. As expected, the denoising performance increases with the depth of the network, but this increment is less significant beyond depth 3. Assessing the trade-off between the network efficiency and the computational complexity, a depth of 2 was considered in the proposed DL models.

4.3.3 *Ablation Study on the Hamiltonian Kernel*

In the proposed models, the objective is to construct a Hamiltonian kernel to conduct the projection operation, while preserving the original attributes of the proposed Hamiltonian operator in the baseline De-QuIP algorithm [51]. This Hamiltonian kernel is a sum of the nabla operator, original pixels' values of the patch and the interactions with its neighbors, following equation (5). To illustrate the importance of this Hamiltonian structure in the proposed networks, an ablation investigation of this Hamiltonian kernel was conducted, through three network settings: (i) without the Hamiltonian kernel and interaction layer, (ii) with the Hamiltonian kernel but without the interaction layer, and (iii) with the Hamiltonian kernel including the interaction layer. For all settings, the depth of the projection layer was set to 2. Table 2 regroups the denoising results on Set12 for AWGN with $\sigma = 15$ for all these three configurations. From these results, one may observe that the accuracy of the network is significantly improved in the case where the Hamiltonian shape is preserved and includes the interactions between neighboring patches. This improvement is even further illustrated by the SSIM, that is more sensitive to the image structure than the PSNR, and thus more suitable to reflect the contribution of the interaction-based Hamiltonian operator. Furthermore, one may notice that without none of these two ingredients, the denoising performance is largely decreased. This can be explained by the fact that in this case, the resulting netwrok, very similar to DnCNN [38], needs far more layers to achieve good denoising results. Indeed, a network depth of 17 is suggested in [38], while, as mentionned previously, the proposed network depth can be reduced to 2. Therefore, the exploitation of the local information through the patch interaction, originally proposed in the baseline De-QuIP, and the attributes of the Hamiltonian kernel, make the proposed DL networks easily adaptable but resilient even

TABLE 1
Ablation investigation of the projection layer's depth using Hamiltonian kernel with or without the interaction layer. The results (PSNR/SSIM) are obtained on Set12 contaminated with AWGN with $\sigma = 15$, in 50 epochs.

	Depth of the projection layer using Hamiltonian convolutional kernel									
	1	2	3	4	5	1	2	3	4	5
Interaction layer	✗	✗	✗	✗	✗	✓	✓	✓	✓	✓
PSNR(dB)/SSIM(%)	30.38/87.64	31.61/89.22	31.95/90.74	32.17/91.61	32.28/91.88	32.09/93.68	32.92/95.41	32.95/95.52	32.96/95.55	32.98/95.60



(a) Left: Denoising performance vs parameter number comparisons are presented on the BSD68 dataset with $\sigma = 50$. Right: Denoising performance vs run time comparisons are presented on the Set12 dataset with $\sigma = 50$.

(b) Left: Deblurring performance vs parameter number comparisons are presented on the Levin dataset with motion blur and $\sigma = 7.65$. Right: Deblurring performance vs run time comparisons presented on the Levin dataset with motion blur and $\sigma = 7.65$.

(c) Left: SR performance vs parameter number comparisons are presented on the BSD100 dataset for 4X SR. Right: SR performance vs run time comparisons are presented on the Urban100 dataset for 4X SR.

Fig. 4. Performance versus parameter number and run time versus performance are presented for different methods for different tasks. The proposed methods give high performances in terms of SSIM(%) with fewer number of parameters and low computation time.

TABLE 2
Ablation study with/without using the Hamiltonian kernel in the network. The results (PSNR/SSIM) are obtained in 50 epochs on Set12 images contaminated with AWGN ($\sigma = 15$).

	Contribution of different components		
Hamiltonian kernel	✗	✓	✓
Interaction layer	✗	✗	✓
PSNR(dB)/SSIM(%)	29.30/86.82	31.61/89.22	32.92/95.41

for lower depth. In conclusion, this experiment illustrates the significance of the inclusion of the Hamiltonian kernel with the interaction layer in the proposed models.

4.3.4 *Analysis of the Parameter Number and Runtime*

The number of hyperparamers of a DL network plays a crucial role in its efficiency. Generally, a larger pool of parameters drives the model more resilient and leads to better performance. However, it also imposes an important computational load, in particular within the training process. Furthermore, excess baggage of parameters may lead to an over-fitting problem. Hence, a balanced trade-off between the learnable parameter number, the performance, and the computational cost becomes a crucial factor for an efficient DL model.

As detailed in the previous ablation studies, the proposed models exploit the Hamiltonian kernel, which is enriched by an intrinsic non-local architecture through the interaction layer. As a result, the resulting DL networks are able to process more information through fewer parameters and significantly reduce the cost of training with high efficiency. Fig. 4 provides the performance in terms of SSIM(%) versus the number of parameters and the runtime of the proposed models against state-of-the-art methods, in the context of different image restoration problems. One can observe a significant gain in performance of DIVA model for image denoising (see Fig. 4(a)). DIVA increases SSIM by 10%, with almost half the number of parameters of its closest competitors FFDNet [29] and DnCNN [38]. For image deblurring problem (see Fig. 4(b)), DIVA-A requires only half of the parameters compared to its nearest rival DRED-DUN [47], but offers a 1% better SSIM value. Similarly, from Fig. 4(c), one can report a gain of 1-2% in SSIM for image SR by DIVA-A compared to the recently introduced DRLN network, whereas our model has 40 times less parameters than DRLN. Naturally, the proposed networks that need a reduced number of parameters to perform well, also offer a signaficantly reduced training cost. Fig. 4 presents the runtime comparisons against other standard models in various imaging tasks, showing that the proposed models are significantly faster. Note that similar results are achieved for image inpainting, but are not reported here since the comparison network is IRCNN, already included in the SR experiments. Hence, harnessing the benefits of the interaction layer and of the Hamiltonian kernel, the proposed DL models demonstrate better performance for image restoration with fewer parameters and more efficient computational costs.

4.4 Qualitative and Quantitative Image Restoration Results

4.4.1 *Image Denoising*

Table 3 summarizes the average PSNR and SSIM results of the different methods on six commonly used testing datasets with six different noise levels. One might notice that the proposed DIVA model uniformly outperforms all the state-of-the-art approaches, irrespective of the noise level and dataset. Compared to the deep unfolded BM3D network BM3D-NET, our model exhibits much better denoising performance with an average increment of 1.5dB PSNR and 4.5% SSIM for low noise levels and up to 2dB PSNR and 13% SSIM for higher σ. Note also that BM3D-NET was only available for four levels of noise. One can observe that the performance gain is much higher over the benchmark DnCNN and FFDNet networks for high

TABLE 3
Image denoising results in terms of average PSNR(dB) and SSIM(%) values for five benchmark datasets contaminated by six noise levels ($\sigma = 10, 15, 25, 50, 75, 100$). For each experiment, the best values are in bold and the second best values are underlined.

Dataset	σ	Input	Methods						
			DnCNN [38]	FFDNet [29]	IRCNN [31]	BM3D-NET [44]	De-QuIP [51]	DIVA	DIVA-blind
Set12	10	28.16/82.87	34.76/92.69	34.64/92.71	33.62/91.83	33.27/91.97	33.45/91.03	**34.80/96.77**	34.68/94.56
	15	24.64/69.97	32.84/90.23	32.75/90.27	32.77/88.08	31.65/88.96	31.15/87.30	**32.92/95.41**	32.79/93.61
	25	20.20/49.68	30.42/86.14	30.42/86.34	30.38/84.23	29.77/85.09	28.65/81.23	**30.47/93.00**	30.36/90.73
	50	14.18/24.87	27.16/78.25	27.32/79.03	27.14/75.70	25.78/76.77	25.28/70.43	**27.45/88.22**	-/-
	75	10.66/14.75	25.15/71.71	25.49/73.52	23.75/67.46	-/-	23.44/63.69	**25.63/84.31**	-/-
	100	8.16/9.64	23.87/64.28	24.20/69.26	21.95/59.70	-/-	22.21/58.02	**24.43/81.17**	-/-
BSD68	10	28.15/83.57	33.87/92.71	33.75/92.66	33.74/90.57	32.74/91.73	32.67/90.65	**33.94/96.21**	33.80/94.38
	15	24.63/70.99	31.73/89.06	31.63/89.02	31.63/87.98	31.42/88.77	30.24/85.38	**31.79/94.04**	31.64/92.74
	25	20.19/50.70	29.22/82.78	29.19/82.89	29.15/79.51	28.95/81.42	27.83/77.35	**29.34/90.07**	29.19/87.44
	50	14.17/25.08	26.22/71.85	26.29/72.45	26.16/68.13	25.73/70.31	24.88/64.25	**26.33/82.99**	-/-
	75	10.65/14.61	24.63/64.69	24.78/65.86	22.87/60.05	-/-	23.33/56.55	**24.87/77.81**	-/-
	100	8.15/9.41	23.16/55.46	23.77/60.96	19.46/49.47	-/-	22.27/51.23	**23.93/74.21**	-/-
Kodak	10	28.14/81.24	34.86/92.17	34.81/92.20	34.76/87.91	32.39/91.01	33.56/89.95	**34.91/96.35**	34.82/94.75
	15	24.62/67.32	32.84/88.82	32.72/88.90	32.63/83.40	30.82/87.68	31.27/85.13	**32.93/94.49**	32.78/93.02
	25	20.18/45.89	30.43/83.15	30.37/83.42	30.29/78.07	28.55/81.62	28.83/77.64	**30.55/91.16**	30.30/87.89
	50	14.16/21.13	27.47/73.53	27.61/74.34	27.44/69.24	25.91/72.15	25.71/65.76	**27.70/85.41**	-/-
	75	10.64/11.91	25.77/67.34	25.96/68.80	23.85/61.75	-/-	24.07/59.02	**26.16/81.36**	-/-
	100	8.14/7.54	23.99/55.99	24.88/64.74	20.38/51.29	-/-	22.92/53.74	**25.22/78.66**	-/-
LIVE1	10	28.14/83.19	34.24/92.95	34.13/92.96	33.02/88.09	32.77/91.83	32.39/90.98	**34.27/96.54**	32.19/94.81
	15	24.62/70.46	32.11/89.68	32.01/89.71	30.42/81.32	30.46/88.74	29.96/85.96	**32.19/94.65**	31.97/92.69
	25	20.18/50.19	29.55/83.91	29.53/84.08	27.22/75.04	27.61/82.14	27.44/78.00	**29.62/91.12**	29.46/88.32
	50	14.16/25.03	26.40/73.34	26.51/74.03	23.05/66.92	24.75/71.60	24.28/64.73	**26.63/84.54**	-/-
	75	10.64/14.74	24.70/66.14	24.92/67.59	21.21/57.58	-/-	22.62/56.66	**24.99/79.65**	-/-
	100	8.14/9.59	22.39/50.10	23.81/62.74	19.59/48.28	-/-	21.51/50.87	**23.99/76.23**	-/-
Urban100	10	28.15/87.17	34.43/95.74	34.45/94.89	32.93/91.35	32.53/94.52	31.25/93.26	**34.75/97.84**	34.52/95.37
	15	24.63/77.13	32.17/93.36	32.42/92.73	30.30/88.77	30.65/91.99	29.53/88.60	**32.51/96.52**	32.26/94.11
	25	20.19/60.04	29.27/88.42	29.92/88.87	27.01/83.09	27.68/86.63	25.75/82.53	**30.01/93.73**	29.75/91.89
	50	14.17/34.98	25.46/77.82	26.52/80.57	22.79/71.51	23.99/75.34	22.81/68.02	**26.67/87.80**	-/-
	75	10.65/22.46	23.23/68.69	24.52/73.65	20.81/61.21	-/-	20.59/58.92	**24.80/82.10**	-/-
	100	8.15/15.38	22.04/62.85	23.08/67.59	18.79/53.57	-/-	19.65/50.51	**23.39/77.37**	-/-

noise cases. Precisely, DIVA outperforms these competing methods by 0.05-1.2dB PSNR and 4-18% SSIM in average and achieves the best denoising yields. Moreover, our blind denoising model DIVA-blind that, in contrast to the other networks, is not trained for a given (known) noise level, but for a range of σ, still gives comparable PSNR values and improved SSIM compared to the state-of-the-art approaches. In all the cases, one can observe a considerable improvement in SSIM enabled by the proposed network, which proves that it is better equipped for image structure and pattern preservation than other models. Utilization of this local information from neighboring patches enables our network to be resilient and adapted to high and low-level noise, giving us an edge over other models.

Furthermore, a notable gain of an average of 1.5-3dB PSNR and 5-26% SSIM is observed compared to the baseline De-QuIP method. This is a consequence of finely tuned hyperparameters values for each patch by harnessing the power of the backpropagation architecture.

Fig. 5 illustrates denoising results for three images, *Girl*, *Castle* and *Parrot*, from three datasets, for $\sigma = 25$, 50 and 75 respectively. The qualitative analysis of the denoised images confirms the superiority of the proposed model. Indeed, all competing methods fail to recover the original textures around the eye and lips in *Girl* image, the sharp edges and peaks around the windows and roof in *Castle* image, and the patterns in *Parrot* image. IRCNN restores blurred edges, and BM3D-NET and De-QuIP generate some small artifacts. DnCNN and FFDNet give comparable PSNR, but low SSIM, caused by over-smoothed results, which were not able to retrieve small details. In contrast, DIVA is faithful to the ground truths and restores the images with the right consistency by capturing the subtle details more reliably.

Visual and quantitative inspections indicate that DIVA model conclusively outperforms its baseline method De-QuIP, as well as other advanced DL methods by a significant margin in terms of PSNR and SSIM. The DnCNN and FFDNet are the closest to DIVA, but struggle to preserve image textures accurately, mainly because of a smoothing effect. DIVA preserves most of the image fragments and textures in a better way without creating any visible artifacts and thus provides a denoised image closer to the ground truth.

TABLE 4
Deblurring results in terms of average PSNR(dB) and SSIM(%) values for two datasets degraded with three GB kernels and AWGN.

Dataset	kernel_σ	noise_σ	Methods			
			IDD-BM3D [5]	Son *et al.* [68]	DEBCNN [64]	DIVA-A
BSD100	1.6	2	27.17/86.14	23.18/73.47	28.47/87.90	**29.97/89.65**
	3	10	24.16/76.66	22.88/68.14	25.34/78.11	**26.57/80.16**
	5	10	22.75/71.74	22.17/65.92	22.79/71.94	**23.73/74.09**
Set16	1.6	2	30.85/93.41	29.87/93.29	31.34/94.39	**32.38/95.37**
	3	10	26.37/85.78	25.20/82.34	26.93/86.91	**27.38/89.31**
	5	10	24.23/82.24	23.63/80.55	27.28/82.76	**28.11/87.22**

4.4.2 *Image Deblurring*

Image deblurring results for GB are illustrated on two benchmark datasets degraded with three different GB kernel settings of size 25×25: (i) GB kernel with standard deviation of 1.6 and AWGN with $\sigma = 2$, (ii) GB kernel with standard deviation of 3 and AWGN with $\sigma = 10$, (iii) GB kernel with standard deviation of 5 and AWGN with $\sigma = 10$. Table 4 regroups the average PSNRs and SSIMs obtained

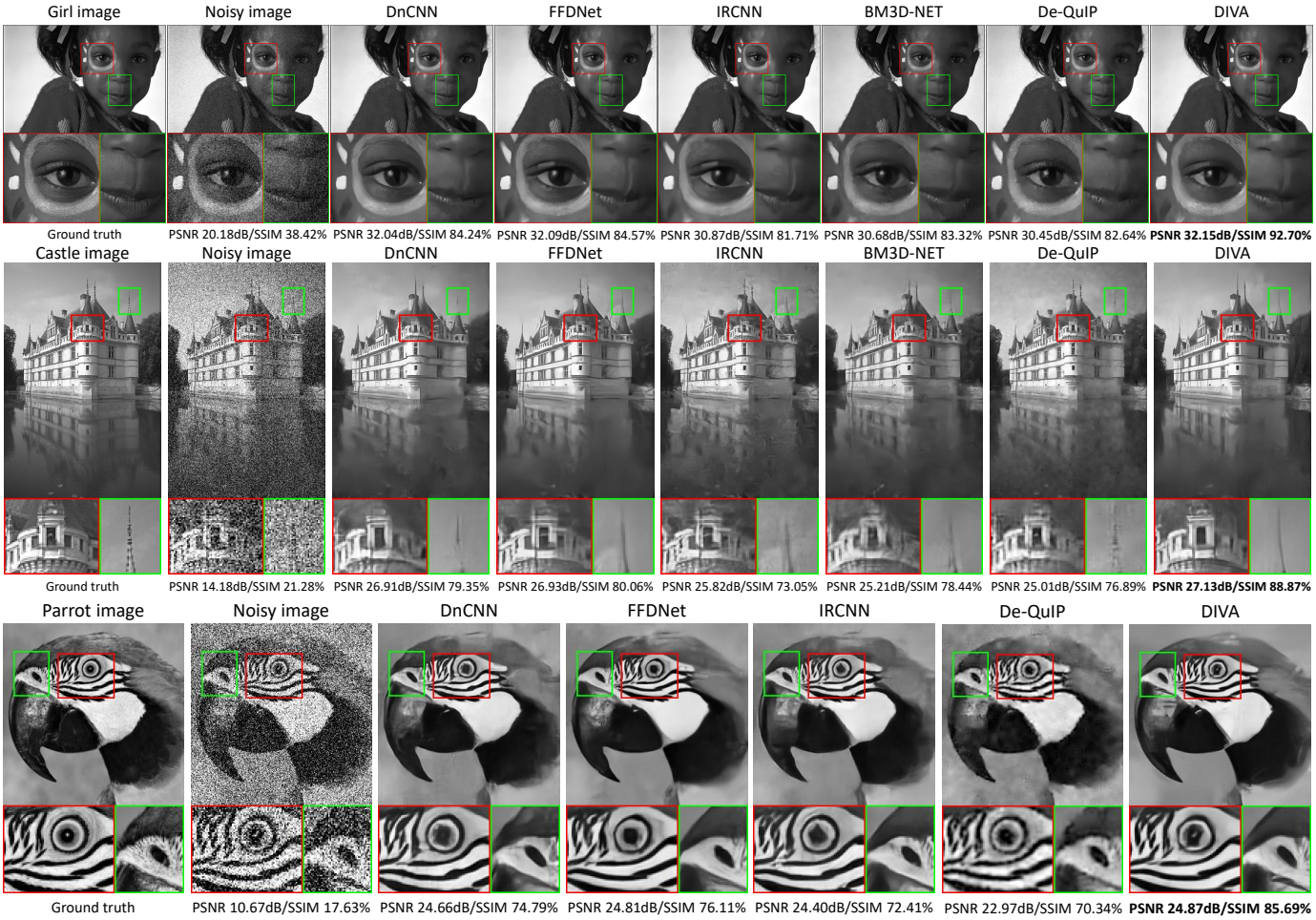


Fig. 5. Denoising image results using different methods. The *Girl* image (Top), *Castle* image (middle), and *Parrot* image (bottom) are respectively contaminated with AWGN with $\sigma = 25$, $\sigma = 50$, and $\sigma = 75$.

TABLE 5
Deblurring results in terms of average PSNR(dB) and SSIM(%) values for four datasets degraded with standard MB kernels and AWGN.

Dataset	noise$_\sigma$	Methods					
		IDD-BM3D [5]	FDN [67]	VEMNet [30]	DWDN [40]	DRED-DUN [47]	DIVA-A
Set10	0	36.24/89.24	-/-	-/-	**43.95**/96.49	43.67/96.38	43.54/**96.67**
	2.55	30.75/86.63	-/-	31.71/89.95	**33.28**/93.12	33.16/92.97	33.03/**93.54**
	7.65	27.25/77.76	-/-	28.27/82.51	29.61/88.07	**29.80**/88.48	29.38/**90.03**
	12.75	25.71/71.38	-/-	26.62/77.68	26.92/83.16	**27.49**/84.05	27.42/**85.79**
Levin	0	37.48/94.68	-/-	-/-	46.13/97.63	45.56/97.27	**46.19/97.76**
	2.55	33.75/92.19	34.05/93.35	34.31/94.31	**36.90/96.14**	36.02/95.79	36.19/95.86
	7.65	29.26/85.78	29.77/85.83	30.50/87.86	32.77/91.79	32.87/91.97	**33.12/92.46**
	12.75	27.33/78.92	27.94/81.39	28.52/82.73	30.77/88.57	**30.89**/88.79	30.80/**89.87**
Sun *et al.*	0	37.14/90.42	-/-	-/-	**43.10**/97.19	42.49/97.08	42.65/**97.36**
	2.55	32.24/87.79	32.63/88.87	32.73/90.13	34.05/92.25	34.43/92.97	**34.44/93.49**
	7.65	28.74/77.86	28.97/78.42	29.41/81.08	29.11/86.31	29.88/87.28	**30.30/89.14**
	12.75	27.30/73.24	27.62/74.52	28.04/77.89	27.81/80.85	**28.20**/81.59	27.95/**83.36**
Set12	0	-/-	-/-	-/-	-/-	-/-	**43.48/96.39**
	2.55	31.43/88.14	31.43/89.17	31.93/90.19	-/-	-/-	**33.77/92.58**
	7.65	27.56/80.09	27.89/80.86	28.47/82.78	-/-	-/-	**28.97/87.89**
	12.75	25.95.74.88	26.28/76.24	26.77/78.13	-/-	-/-	**27.28/84.45**

$*$*The symbol -/- denotes that the results were not provided in the original paper for a particular experiment.*

by all competing methods. One can observe that the benchmark DEBCNN [64] method performs much better than the model-based IDD-BM3D [5] and learning-based Son *et al.* [68] schemes. DIVA-A outperforms DEBCNN by 1.1 dB in PSNR and 2% in SSIM and 0.8 dB in PSNR and 2.6% in SSIM on average for BSD100 and Set16 datasets, respectively.

In Fig. 6, a qualitative evaluation shows that the proposed method not only generates better image contrast but

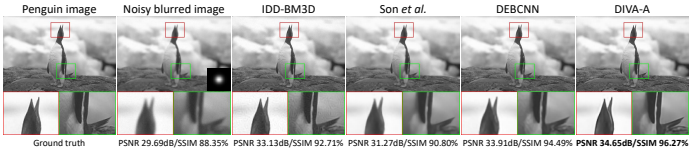


Fig. 6. Image deblurring results for *Penguin* image degraded by a 25×25 GB kernel of standard deviation 1.6 with random AWGN of standard deviation 2.

also retrieves sharp edges with more details than the other approaches, like IDD-BM3D and Son *et al.* [68], where random artifacts and blurred edges are visible in the deblurred outputs. Our DL model restores the *Penguin* image with much sharper and more precise edges than the DEBCNN, for which edges look hazy. Thus, though DEBCNN and DIVA-A are the two best models in this setting, our model uniformly outperforms the sophisticated DEBCNN method for GB problems.

Table 5 gives the average deblurring performance of our method in terms of PSNRs and SSIMs in contrast to other standard models from the literature under eight commonly used MB kernels [47], [63] and four different noise levels. One should note that the code or trained models provided by the authors are used to generate these results. As the

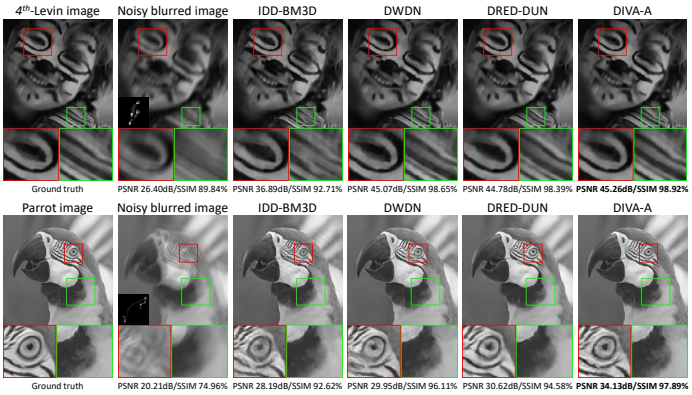


Fig. 7. Deblurring results for MB kernels. The first row shows restored *4th*-image from the Levin dataset with 17×17 MB kernel, and second row shows restored *Parrot* images with 25×25 MB kernel.

TABLE 6
SR results in terms of average PSNR(dB) and SSIM(%) values for 4 benchmark datasets degraded with bicubic downsampling with downsampling factors of 2, 3 and 4.

Dataset	Scale	Methods				
		LapSRN [69]	MemNet [66]	CARN [65]	DRLN [42]	DIVA-A
Set5	2x	37.52/95.91	37.78/95.97	37.76/95.90	**38.27**/96.16	37.42/**97.43**
	3x	33.82/92.27	34.09/92.48	34.29/92.55	**34.78**/93.03	33.14/**93.36**
	4x	31.54/88.50	31.74/88.93	32.13/89.37	**32.63/90.02**	30.87/**90.02**
Set14	2x	33.08/91.30	33.28/91.42	33.52/91.66	**34.28**/92.31	33.67/**93.69**
	3x	29.87/83.20	30.00/83.50	30.29/84.07	**30.73**/84.88	29.18/**85.34**
	4x	28.19/77.20	28.26/77.23	28.60/78.06	**28.94**/79.00	27.74/**80.66**
BSD100	2x	31.80/89.50	32.08/89.78	32.09/89.78	**32.44**/90.28	32.00/**90.49**
	3x	28.82/79.80	28.96/80.01	29.06/80.34	**29.36**/81.17	28.91/**82.15**
	4x	27.32/72.80	27.40/72.81	27.58/73.49	**27.83**/74.44	27.66/**76.95**
Urban100	2x	30.41/91.00	31.31/91.95	31.51/93.12	**33.37/93.90**	31.48/93.06
	3x	27.07/82.80	27.56/83.76	27.38/84.04	**29.21/87.22**	27.54/85.31
	4x	25.21/75.60	25.50/76.30	26.07/78.37	**26.98**/81.19	25.39/**81.29**

first observation, one can see that DWDN and DRED-DUN outperform the conventional IDD-BM3D, FDN and VEMNet for the Set10, Levin and Sun *et al.* datasets, which is consistent with the findings in [47]. Secondly, DWDN performs better in the case of low/no noise in terms of PSNRs compared to DRED-DUN and our proposed model. DRED-DUN is more accurate for high levels of noise. On the contrary, our proposed model exhibits the best SSIMs with a gain up to 0.15-1.8% against the DWDN and DRED-DUN for low as well as high noise levels and this efficiency increases with noise intensity. In terms of PSNR values, our model often stays in the top two and only fails to do so for Set10, where the average PSNR gaps between the best two methods and our model is very small. Noticeably, although DIVA-A sometimes offers slightly worse PSNRs than DWDN and DRED-DUN, it requires only half of the tunable parameters (shown in Fig. 4(b)). Finally for Set12, our model unilaterally dominates the comparison and exceeds its nearest rival VEMNet by up to an average of 1dB PSNR and 3.5% SSIM.

For visual assessment, restored images from images degraded with three different MB kernels are shown in Figs. 7, 8. Fig. 7 shows that for the *4th*-Levin and *Parrot* images respectively under a MB kernel of size 17×17 and 25×25, the quality of the restored images by our model is considerably improved compared to the other methods. In particular, the finer texture of the images is severely smoothed out by IDD-BM3D, DWDN and DRED-DUN, as shown in the zoomed boxes. Furthermore, the overall visual impression of the restored images is improved, as visible on the facial decorations and minute patterns that are better preserved with DIVA-A. Finally, Fig. 8 offers a similar conclusion for the restored *3th*-Levin image under a 23×23 MB kernel with AWGN with $\sigma = 2.55$. Thus, under MB kernels DIVA-A demonstrates a better efficiency in recovering edges and patterns of the original images. DWDM and DRED-DUN produce comparative results compared to our model, but with lower contrast. Quantitatively, our method is always among the best two approaches in this context.

4.4.3 Single Image Super-Resolution (SR)

This subsection presents SR results for two standard downsampling operators, bicubic downsampling (BD) and Gaussian downsampling (GD). Tables 6, 7 regroup average PSNR and SSIM values of different methods on four datasets for BD and GD respectively. One may observe that the recently introduced benchmark method DRLN [42] provides the best performance in both contexts. DRLN has a complex network architecture with dense residual Laplacian modules powered by 34 million parameters. In contrast, the proposed model has a much simpler architecture, and requires only 850K parameters approximately. Nevertheless, our model obtains the best SSIM for three datasets (*e.g.*, Set5, Set14 and BSD100) and among the top two SSIM for Urban100 images for BD. One can see an average gain of 1.5% SSIM by our method over DRLN in the BD scenario. For GD problems, our method struggles to produce competitive results against benchmark DRLN, RDN and DFAN approaches. Note that for SR problem our method upsamples the observed LR image by bicubic interpolation to obtain an initial HR image before enhancing it using the trained DL network.

The visual inspections of Figs. 9, 10 and 11 illustrate the potential of our method for SR. Figs. 9 and 10 correspondingly display the restored HR images from the LR BD *Zebra* and *Baby-face* images with scale factors of 3 and 4. The visual effects of HR images recovered by our method are better than others and higher in accuracy. For example, in our retrieved HR images the stripes on the zebra's body, in *Baby-face* image the textures and shapes of eye, lips and nose have better specifications than the other methods. Fig. 11 shows the reconstructed HR images from the LR *Book-cover* image obtained by GD with scale factor of 2. Observation reveals that our method efficiently recovers the edges, patterns and texts of the original image from LR data. Moreover, our method strongly competes with the benchmark DRLN and beats it in some respects, especially in terms of overall visual quality and preservation of the image structure.

4.4.4 Image Inpainting

Table 8 illustrates our model performance in terms of average PSNR and SSIM on Set5 and Set12 datasets compared to the standard IRCNN network for image inpainting problems. Our model outperforms IRCNN in almost all situations when 20%, 50%, and 80% of random pixels are missing in the degraded observed images. DIVA-A provides a improvement of 0.2-1 dB in PSNR and 0.6-5.5% in SSIM over IRCNN and this gain increases with data lacking.

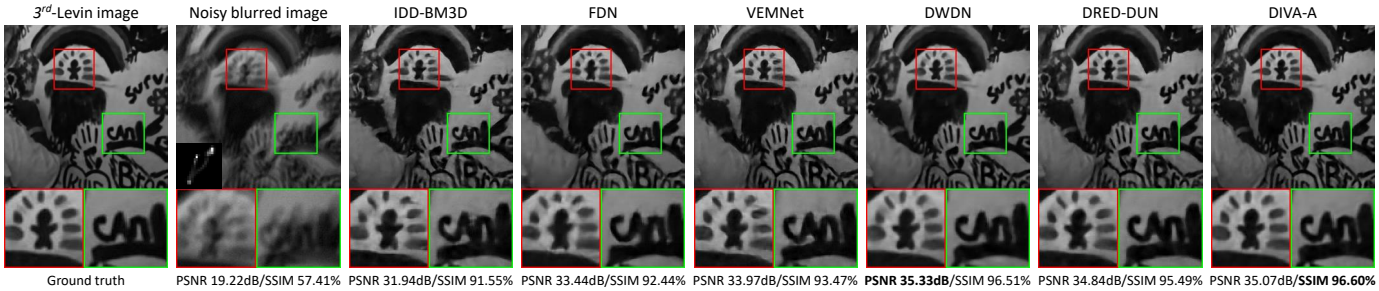


Fig. 8. Deblurring results for the *3rd*-image from the Levin dataset with motion blur kernel of size 23×23 and random AWGN $\sigma = 2.55$.

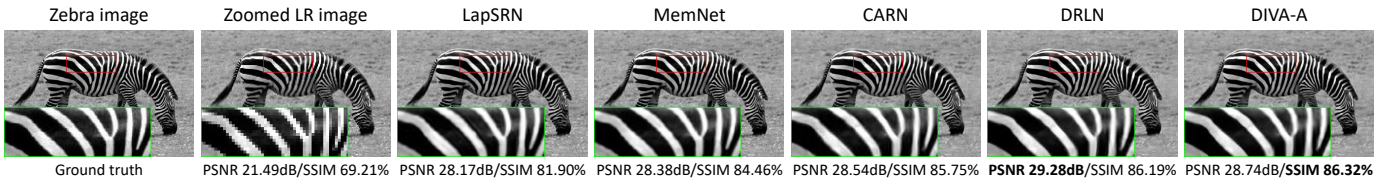


Fig. 9. SR results for *Zebra* image for a bicubic downsampling with scaling factor 3.

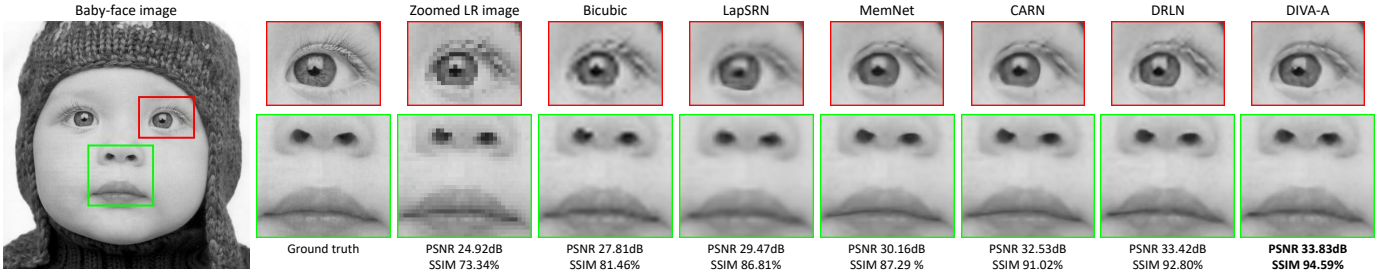


Fig. 10. Two zoomed regions of the restored HR *Baby-face* images, extracted from SR results for a bicubic downsampling with scaling factor 4.

TABLE 7
SR results in terms of average PSNR(dB) and SSIM(%) values for 4 benchmark datasets degraded with GD by using a 7×7 GB kernel of standard deviation 1.6 with scaling factors of 2, 3 and 4.

Dataset	Scale	Methods				
		IRCNN [31]	DFAN [70]	RDN [32]	DRLN [42]	DIVA-A
Set5	2x	**35.34**/93.04	-/-	-/-	-/-	33.62/**93.79**
	3x	33.38/91.82	34.50/92.74	34.58/92.80	**34.81/92.97**	32.70/91.45
	4x	**30.76**/85.47	-/-	-/-	-/-	29.02/**85.76**
Set14	2x	**31.98**/88.49	-/-	-/-	-/-	30.88/**90.65**
	3x	29.63/82.81	30.43/84.19	30.53/84.47	**30.81/84.87**	28.97/83.47
	4x	**27.73**/74.12	-/-	-/-	-/-	26.86/**76.01**
BSD100	3x	28.65/79.22	29.17/80.58	29.23/80.79	**29.40/81.21**	28.26/80.65
Urban100	3x	26.77/81.54	28.27/85.26	28.46/85.82	**29.11/86.97**	27.72/84.92

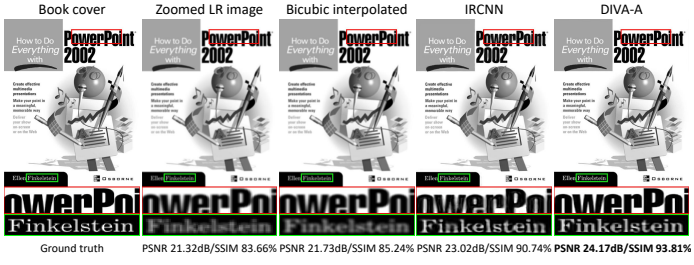


Fig. 11. Restored HR *Book-cover* images from LR images generated by GD under a 7×7 GB kernel of standard deviation 1.6 with scaling factor 2.

The visual analysis of Fig. 12 confirms the quantitative results. From the restored *F-16 Jet* image, it appears that our model efficiently reproduces the F-16 logo, borders and sharp edges despite 50% of data missing, whereas IRCNN fails to do so and loses/distorts many details in the restored output. Hence, our model can gather local information from the image neighborhood quite promisingly and delivers a high-quality restored image even with limited pixels available. [2]

5 Discussions

In this section, we briefly recap the benefits and limitations of our proposed networks and future prospects in this regard.

Advantages: With the quantum principles of the baseline De-QuIP algorithm, our proposed DIVA/DIVA-A network provides an efficient DL method for image restoration following the deep unfolding philosophy. Indeed, the use of quantum concepts like patch interaction layer and Hamiltonian kernel makes our models better equipped than others. The local structure/similarities in an image neighborhood are preserved through the interaction layer exploiting the local patch groups that convey an intrinsic non-local network architecture. Processing of the local information by this

2. More visual results can be found in the supplementary material.

TABLE 8
Image inpainting results in terms of average PSNR(dB) and SSIM(%) values for two benchmark datasets for respectively 20%, 50% and 80% pixels missing.

Dataset	Missing pixels'	Input	Methods	
			IRCNN [31]	DIVA-A
Set5	20%	13.33/38.61	41.62/98.67	**41.85/99.24**
	50%	9.34/23.44	35.57/95.87	**36.08/97.84**
	80%	7.29/12.40	29.41/88.54	**30.38/94.01**
Set12	20%	12.46/27.93	**39.06**/98.29	38.57/**99.15**
	50%	8.48/14.45	32.82/94.53	**33.02/97.21**
	80%	6.44/6.71	26.75/84.53	**27.73/91.92**

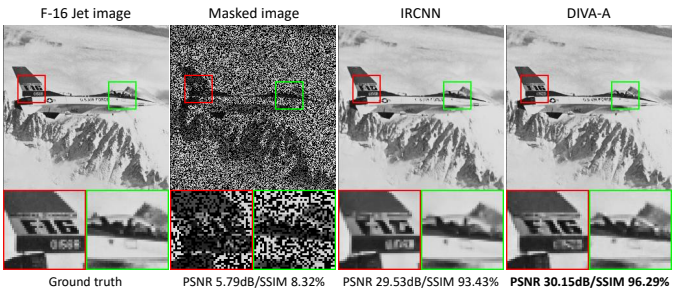


Fig. 12. Restored *F-16 Jet* images, when 50% pixels' are missing.

interaction layer significantly enhances the performances of the network. It even yields a smaller network depth, leading to a good trade-off between the performance and computational cost, as portrayed in Sec. 4.3. Harnessing the power of back-propagation, our networks uniquely tune all hyperparameters, such as proportionality constant, Planck constant and thresholding energy, for each patch. This enables network adaptability with several image restoration tasks, and leads to promising performances.

Limitations: In the case of a challenging image degradation task, our method may sometimes struggle to produce a better recovered image than other benchmarks. To restore a Gaussian downsampled LR image, we notice that our DL model fails to compete in quantitative data against benchmark methods, like DRLN, RDN, and DFAN, as noted in Table 7. However, the overall visual efficiency of our method is quite good, as depicted in Fig. 11. Perhaps in presence of a strong decay, such as 'blur+downsampling', our method does not match the true pixels' intensity, which seems to be the main reason for the lower quantitative measures. Instead, our method utilizes the interaction layer to provide better visual quality by preserving the image structure, patterns, and textures with more details. Furthermore, our proposed models are trained in an end-to-end supervised manner, *i.e.*, we need the clean-degraded image pairs for training. However it is worth-noting that the proposed method is much simpler and not specialized in a specific task as is the case for the other methods.

Future perspectives: The quantum mechanics-based imaging methods open up a broad spectrum of future prospects. Following the limitations, the obvious direction would be an unsupervised DL network design, that essentially solves the training data problem and extends our reach to real-life applications more reliably [71], [72]. Another possibility is to design a versatile network by stacking the proposed DIVA to build a deep and more complex architecture like UNet [47] and offer some attention mechanisms [42] to make the patch interaction robust while preserving the core philosophy. This complex network system should enhance the capacity of the proposed network in challenging degradation operators and even for blind imaging problems. Furthermore, the idea of quantum interaction can also be treated as a transformer in a deep architecture [73]. Another interesting prospect would be to explore imaging problems beyond the Gaussian model since baseline De-QuIP is well-adapted for such tasks without modifying the global architecture. Combining graph signal processing model with the proposed quantum-based interaction framework is also an interesting perspective [74].

6 Conclusions

This paper introduces a novel neural network approach to solve image denoising problems, further extended to general image restoration tasks relying on the philosophy of quantum many-body theory. Our model recasts the baseline De-QuIP algorithm into a DL framework and optimizes the relevant parameters by exploiting the power of back-propagation approach. The proposed unfolded CNN architecture inherently employs various quantum mechanical components, such as interaction and Hamiltonian operator, from its baseline method to boost up the network performance while significantly reducing the training cost. Integration of these key features from the quantum theory enables our proposed model to be well-adapted for handling several imaging problems efficiently. We conduct thorough ablation investigations and present extensive assessments regarding the network design. Finally, we perform comprehensive evaluations of our proposed DL methods for various imaging problems, such as denoising, deblurring, single image super-resolution, and inpainting. In all cases, notable improvements were shown in the image restoration performance, especially overall visual quality, compared to standard well-established techniques from the literature.

Acknowledgment

Sayantan Dutta would like to thank Mr. Nishchal Prasad, Ph.D. student at Institut de Recherche en Informatique de Toulouse (IRIT), UMR CNRS 5505, Université de Toulouse, France, for his valuable inputs in this work. Authors also thank CNRS for funding through the 80 prime program.

References

[1] D.L. Donoho and J.M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
[2] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, 2006.
[3] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, 2007.
[4] W. Dong, G. Shi, and X. Li, "Nonlocal image restoration with bilateral variance estimation: A low-rank approach," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 700–711, 2013.
[5] A. Danielyan, V. Katkovnik, and K. Egiazarian, "Bm3d frames and variational image deblurring," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1715–1728, 2012.

[6] M.V. Afonso, J.M. Bioucas-Dias, and M.A.T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345–2356, 2010.

[7] R.H. Chan, M. Tao, and X. Yuan, "Constrained total variation deblurring models and fast algorithms based on alternating direction method of multipliers," *SIAM J. Imag. Sci.*, vol. 6, no. 1, pp. 680–697, 2013.

[8] M.S.C. Almeida and M. Figueiredo, "Deconvolving images with unknown boundaries using the alternating direction method of multipliers," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3074–3086, 2013.

[9] Z. Chen, A. Basarab, and D. Kouamé, "Compressive deconvolution in medical ultrasound imaging," *IEEE Trans. Med. Imag.*, vol. 35, no. 3, pp. 728–737, 2016.

[10] X. Gao, K. Zhang, D. Tao, and X. Li, "Image super-resolution with sparse neighbor embedding," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3194–3205, 2012.

[11] S.P. Mudunuri and S. Biswas, "Low resolution face recognition across variations in pose and illumination," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 38, no. 5, pp. 1034–1040, 2016.

[12] N. Zhao, Q. Wei, A. Basarab, N. Dobigeon, D. Kouamé, and J.-Y. Tourneret, "Fast single image super-resolution using a new analytical solution for $\ell_2 - \ell_2$ problems," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3683–3697, 2016.

[13] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, "An iterative regularization method for total variation-based image restoration," *Multiscale Model. Simulation*, vol. 4, no. 2, pp. 460–489, 2005.

[14] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006.

[15] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1620–1630, 2013.

[16] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. IEEE Int. Conf. Computer Vis.*, 1998, pp. 839–846.

[17] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 257–266, 2002.

[18] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *SIAM Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, 2005.

[19] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2009, pp. 2272–2279.

[20] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 7, pp. 629–639, 1990.

[21] Yongjian Y. and S.T. Acton, "Speckle reducing anisotropic diffusion," *IEEE Trans. Image Process.*, vol. 11, no. 11, pp. 1260–1270, 2002.

[22] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 35, no. 6, pp. 1397–1409, 2013.

[23] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1838–1857, 2011.

[24] J. Sun, J. Sun, Z. Xu, and H.-Y. Shum, "Gradient profile prior and its applications in image super-resolution and enhancement," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1529–1542, 2011.

[25] A.M. Teodoro, J.M. Bioucas-Dias, and M.A.T. Figueiredo, "Image restoration and reconstruction using variable splitting and class-adapted image priors," in *Proc. IEEE Int. Conf. Image Process.*, 2016, pp. 3518–3522.

[26] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, and L. Bottou, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, pp. 3371–3408, 2010.

[27] H.C. Burger, C.J. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with BM3D?," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 2392–2399.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, June 2016, pp. 770–778.

[29] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for cnn-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, 2018.

[30] Y. Nan, Y. Quan, and H. Ji, "Variational-em-based deep learning for noise-blind image deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 3623–3632.

[31] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep cnn denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 2808–2817.

[32] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image restoration," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 43, no. 7, pp. 2480–2495, 2021.

[33] W. Dong, P. Wang, W. Yin, G. Shi, F. Wu, and X. Lu, "Denoising prior driven deep neural network for image restoration," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 41, no. 10, pp. 2305–2318, 2019.

[34] Z. Zha, B. Wen, X. Yuan, J.T. Zhou, J. Zhou, and C. Zhu, "Triply complementary priors for image restoration," *IEEE Trans. Image Process.*, vol. 30, pp. 5819–5834, 2021.

[35] Z. Zha, B. Wen, X. Yuan, J. Zhou, C. Zhu, and A.C. Kot, "Low-rankness guided group sparse representation for image restoration," *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1–15, 2022.

[36] U. Schmidt and S. Roth, "Shrinkage fields for effective image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, June 2014, pp. 2774–2781.

[37] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 39, no. 6, pp. 1256–1272, 2017.

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

[39] J. Malik, S. Kiranyaz, and M. Gabbouj, "Image denoising by super neurons: Why go deep?," *arXiv preprint arXiv:2111.14948*, 2021.

[40] J. Dong, S. Roth, and B. Schiele, "Deep wiener deconvolution: Wiener meets deep learning for image deblurring," in *Proc. NeurIPS*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, Eds. 2020, vol. 33, pp. 1048–1059, Curran Associates, Inc.

[41] C. Dong, C.C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 38, no. 2, pp. 295–307, 2016.

[42] S. Anwar and N. Barnes, "Densely residual laplacian super-resolution," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 44, no. 3, pp. 1192–1204, 2022.

[43] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proc. Int. Conf. Mach. Learn.*, 2010, p. 399–406.

[44] D. Yang and J. Sun, "BM3D-Net: A convolutional neural network for transform-domain collaborative filtering," *IEEE Signal Process. Lett.*, vol. 25, no. 1, pp. 55–59, 2018.

[45] M. Scetbon, M. Elad, and P. Milanfar, "Deep k-svd denoising," *IEEE Trans. Image Process.*, vol. 30, pp. 5944–5955, 2021.

[46] K.H. Jin, M.T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4509–4522, 2017.

[47] S. Kong, W. Wang, X. Feng, and X. Jia, "Deep red unfolding network for image restoration," *IEEE Trans. Image Process.*, vol. 31, pp. 852–867, 2022.

[48] O. Solomon, R. Cohen, Y. Zhang, Y. Yang, Q. He, J. Luo, R.J.G. van Sloun, and Y.C. Eldar, "Deep unfolded robust PCA with application to clutter suppression in ultrasound," *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 1051–1063, 2020.

[49] K. Zhang, L.V. Gool, and R. Timofte, "Deep unfolding network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 3217–3226.

[50] S. Dutta, A. Basarab, B. Georgeot, and D. Kouamé, "Image denoising inspired by quantum many-body physics," in *Proc. 28th IEEE Int. Conf. Image Process. (ICIP)*, 2021, pp. 1619–1623.

[51] S. Dutta, A. Basarab, B. Georgeot, and D. Kouamé, "A novel image denoising algorithm using concepts of quantum many-body theory," *Signal Processing*, vol. 201, pp. 108690, 2022.

[52] Ç. Aytekin, S. Kiranyaz, and M. Gabbouj, "Quantum mechanics in computer vision: Automatic object extraction," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2013, pp. 2489–2493.

[53] A. Youssry, A. El-Rafei, and S. Elramly, "A quantum mechanics-based framework for image processing and its application to

image segmentation," *Quantum Inf. Process.*, vol. 14, no. 10, pp. 3613–3638, 2015.
[54] Z. Kaisserli, T.-M. Laleg-Kirati, and A. Lahmar-Benbernou, "A novel algorithm for image representation using discrete spectrum of the schrödinger operator," *Digit. Signal Process.*, vol. 40, pp. 80–87, 2015.
[55] S. Dutta, A. Basarab, B. Georgeot, and D. Kouamé, "Quantum mechanics-based signal and image representation: Application to denoising," *IEEE Open J. of Signal Process.*, vol. 2, pp. 190–206, 2021.
[56] S. Dutta, A. Basarab, B. Georgeot, and D. Kouamé, "Plug-and-play quantum adaptive denoiser for deconvolving poisson noisy images," *IEEE Access*, vol. 9, pp. 139771–139791, 2021.
[57] S. Dutta, A. Basarab, B. Georgeot, and D. Kouamé, "Poisson image deconvolution by a plug-and-play quantum denoising scheme," in *Proc. 29th Eur. Signal Process. Conf. (EUSIPCO)*, 2021, pp. 646–650.
[58] Y. Altmann, S. McLaughlin, M.J. Padgett, V.K. Goyal, A.O. Hero, and D. Faccio, "Quantum-inspired computational imaging," *Science*, vol. 361, no. 6403, pp. eaat2298, 2018.
[59] S. Dutta, A. Basarab, B. Georgeot, and D. Kouamé, "Deep unfolding of image denoising by quantum interactive patches," in *Proc. 29th IEEE Int. Conf. Image Process. (ICIP)*, 2022.
[60] S. Dutta, A. Basarab, B. Georgeot, and D. Kouamé, "Despeckling ultrasound images using quantum many-body physics," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, 2021, pp. 1–4.
[61] S. Dutta, K.T. Nwigbo, J. Michetti, B. Georgeot, D.-H. Pham, A. Basarab, and D. Kouamé, "Quantum denoising-based super-resolution algorithm applied to dental tomography images," in *Proc. Int. Symp. Biomed. Imaging (ISBI)*, 2022.
[62] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. workshops (CVPRW)*, 2017, pp. 126–135.
[63] A. Levin, Y. Weiss, F. Durand, and W.T. Freeman, "Efficient marginal likelihood optimization in blind deconvolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 2657–2664.
[64] R. Wang and D. Tao, "Training very deep cnns for general non-blind deconvolution," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2897–2910, 2018.
[65] N. Ahn, B. Kang, and K. Sohn, "Fast, accurate, and lightweight super-resolution with cascading residual network," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 252–268.
[66] Y. Tai, J. Yang, X. Liu, and C. Xu, "Memnet: A persistent memory network for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 4549–4557.
[67] J. Kruse, C. Rother, and U. Schmidt, "Learning to push the limits of efficient fft-based image deconvolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 4596–4604.
[68] H. Son and S. Lee, "Fast non-blind deconvolution via regularized residual networks with long/short skip-connections," in *Proc. IEEE Int. Conf. Comput. Photography*, 2017, pp. 1–10.
[69] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 5835–5843.
[70] S. Li, G. Zhang, Z. Luo, and J. Liu, "Dfan: Dual feature aggregation network for lightweight image super-resolution," *Wireless Commun. Mobile Comput.*, vol. 44, no. 3, pp. 1530–8669, 2022.
[71] F. de Morsier, M. Borgeaud, V. Gass, J.-P. Thiran, and D. Tuia, "Kernel low-rank and sparse graph for unsupervised and semi-supervised classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3410–3420, 2016.
[72] M. Pereyra and S. McLaughlin, "Fast unsupervised bayesian image segmentation with adaptive spatial regularisation," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2577–2587, 2017.
[73] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li, "Uformer: A general u-shaped transformer for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, June 2022, pp. 17683–17693.
[74] F. Hua, C. Richard, J. Chen, H. Wang, P. Borgnat, and P. Gonçalves, "Learning combination of graph filters for graph signal modeling," *IEEE Signal Process. Lett.*, vol. 26, no. 12, pp. 1912–1916, 2019.

SUPPLEMENTARY MATERIAL: FOR DIVA: DEEP UNFOLDED NETWORK FROM QUANTUM INTERACTIVE PATCHES FOR IMAGE RESTORATION

Sayantan Dutta[1,2,*], Adrian Basarab[3], Bertrand Georgeot[2], and Denis Kouamé[1]

[1]Institut de Recherche en Informatique de Toulouse, UMR CNRS 5505, Université de Toulouse, France

[2]Laboratoire de Physique Théorique, Université de Toulouse, CNRS, UPS, France

[3]Université de Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1206, Villeurbanne, France.

[*]Corresponding author: Sayantan Dutta (Email: sayantan.dutta@irit.fr; sayantan.dutta110@gmail.com).

7 INTRODUCTION

In this work, we introduce a novel deep-learning (DL) network unfolding the baseline Denoising by Quantum Interactive Patches (De-QuIP) [50], [51] algorithm, denoted as DIVA (Deep denoising by quantum InteractiVe pAtches) for image denoising problem. We further extend the network architecture to conduct a general image restoration task and the respective network denoted as DIVA advanced (DIVA-A). The integration of the key attributes of DL and quantum theory significantly enhances the functionality of our proposed networks due to its intrinsic versatility and enables our models to exhibit state-of-the-art performances for several restoration tasks such as denoising, deblurring, super-resolution, inpainting, etc.

In the original manuscript, we extensively study the network architecture and present comprehensive comparisons with benchmark approaches. The detailed quantitative and qualitative analyses are reported in the original manuscript. In this supplementary material, we depict more restored images for the image deblurring, super-resolution, and inpainting problems to give better insights into the visual qualities of the images recovered by our proposed networks.

8 EXPERIMENTAL RESULTS

In this section, we analyze the qualitative performance of our proposed networks in various image restoration tasks, such as image deblurring, super-resolution, and inpainting.

8.1 Quantitative Metrics

For the purpose of quantitative evalution, the peak-signal-to-noise-ratio (PSNR) and the structural similarity (SSIM) computed between the true and the restored images were used.

8.2 Qualitative Image Restoration Results

8.2.1 Image Deblurring

In Fig. 13, through qualitative evaluation, one can notice that our method not only generates better image contrast but also retrieves sharp edges with more details than other models. For example, in the *Horse* image, IDD-BM3D [5] produces better contrast, but random patterns are visible in the deblurred outputs, whereas Son *et al.* [68] fails to preserve sharp edges

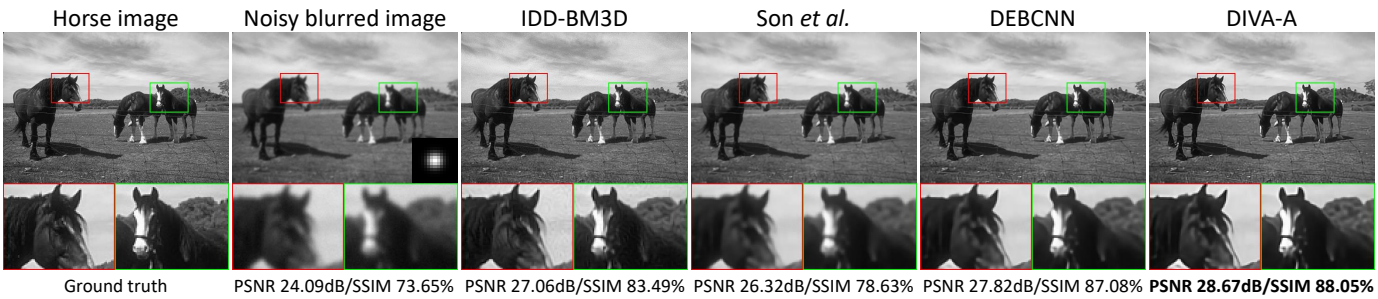


Fig. 13. Image deblurring results for *Horse* image degraded by a 25×25 Gaussian blur kernel of standard deviation 1.6 with random Gaussian noise of standard deviation 2.

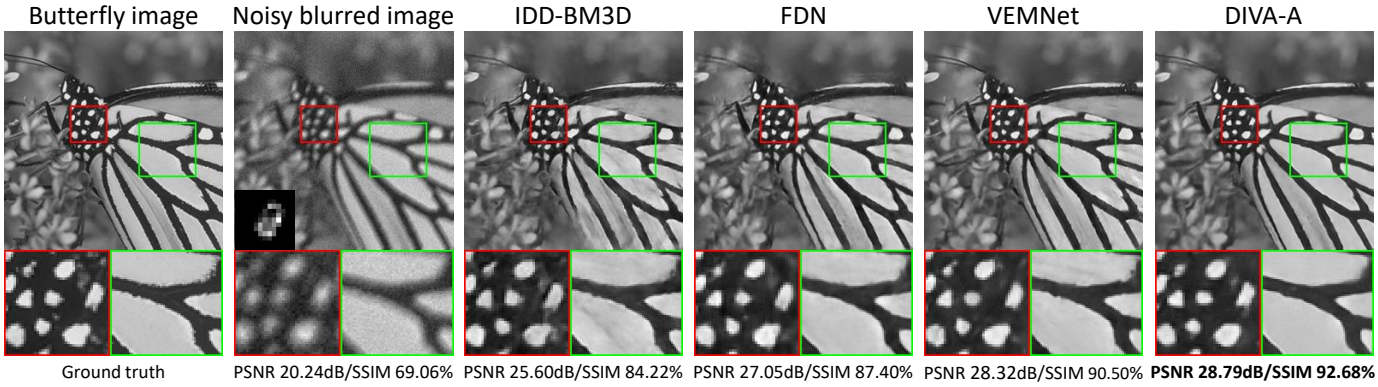


Fig. 14. Deblurring results for motion blur kernel. The restored *Butterfly* images with 13×13 motion blur kernel and Gaussian noise of standard deviation 7.65.

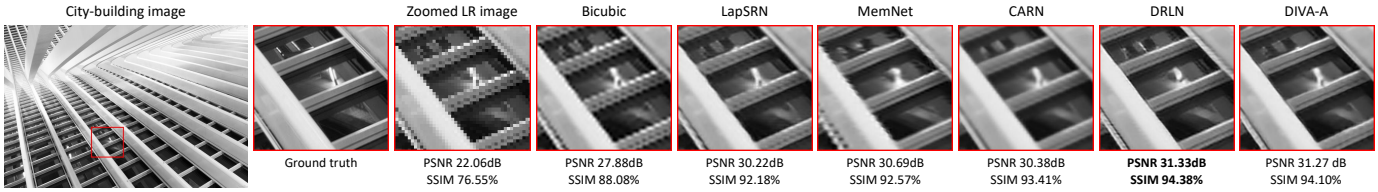


Fig. 15. A zoomed regions of the restored HR *City-building* images, extracted from SR results for a bicubic downsampling with scaling factor 3.

and the overall images appear blurred compared to others. Our DL model restores the head with much sharp and precise edges than the DEBCNN [64], where the edges look hazy. Thus, though DEBCNN [64] and our DIVA-A are the two best models in this setting, our model uniformly dominates the sophisticated DEBCNN [64] method in the Gaussian deblurring problems.

In Fig. 14, the *Butterfly* image is degraded by a moderate size 13×13 motion blur kernel with random Gaussian noise of standard deviation 7.65. Our proposed model for retrieving original image quality is significantly better than other competitors. For example, IDD-BM3D [5], FDN [67], and VEMNet [30] fail to properly restore the pattern on the butterfly's wings and body, and show many distortions in the restored images, as visible in the zoomed boxes. Compared to its rivals, our model has the capability to restore these subtle attributes like patterns on the wings and body, and preserves the sharp edges with finer precision, as shown in the zoomed boxes.

Thus, under both Gaussian and motion blur kernels the overall visual quality of the recovered deblurred images by our proposed model is the best among all the tested methods.

8.2.2 Single Image Super-Resolution (SR)

The visual inspections of Figs. 15, 16 and 17 illustrate the potential of our method for SR. Figs. 15 and 16 display the restored high-resolution (HR) images from the low-resolution (LR) bicubic down-sampled *City-building* and *Fish* images with scale factors of 3 and 4, respectively. The visual effects of HR images recovered by our method are better than others and higher in accuracy. For example, in our retrieved HR *City-building* image the sharp edges of the windows, in *Fish* image the patterns on the fish and the shapes of the seagrass have better specifications than the other methods, such as LapSRN [69], MemNet [66], CARN [65]. Although, the benchmark DRLN [42] provides better quantitative data in some aspects, the proposed DIVA-A not only gives comparable results but also outperforms DRLN [42] in terms of quantitative and visual assessments in some cases.

Fig. 17 shows the reconstructed HR images from the LR *Flowers* image obtained by Gaussian downsampling with scale factor of 3. In the degraded image, the small-scale details are nearly unrecognizable. Observation reveals that our method efficiently recovers the edges and patterns of the original image from LR data compared to the state-of-the-art IRCNN [31], DFAN [70] and RDN [32] methods. Moreover, our method strongly competes with the benchmark DRLN [42] and beats it in some respects, especially in terms of overall visual quality and preservation of the image structure.

Hence qualitatively, our method is always among the best two approaches in this context. These results can be explained by the ability of DIVA-A to exploit the local structures/attributes via the interaction layer. This layer enables our DL network to efficiently conduct super-resolution tasks while correclty restoring patterns, sharp edges and other small-scale details.

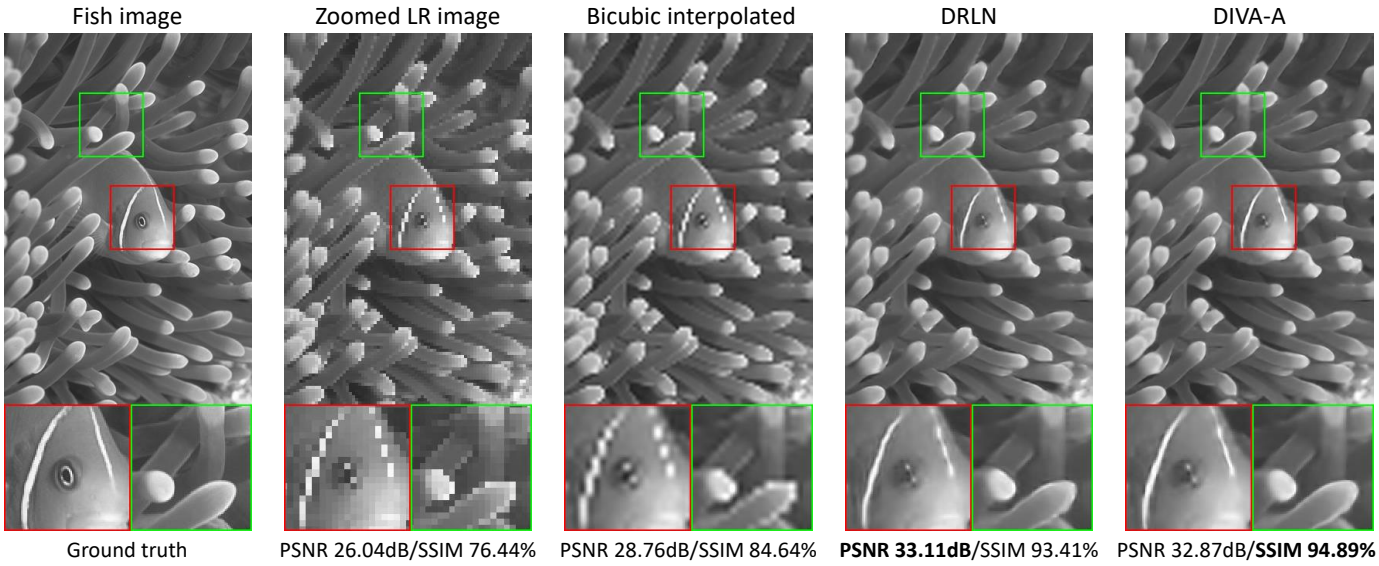


Fig. 16. The restored HR *Fish* images from LR images generated by bicubic downsampling with scaling factor 4.

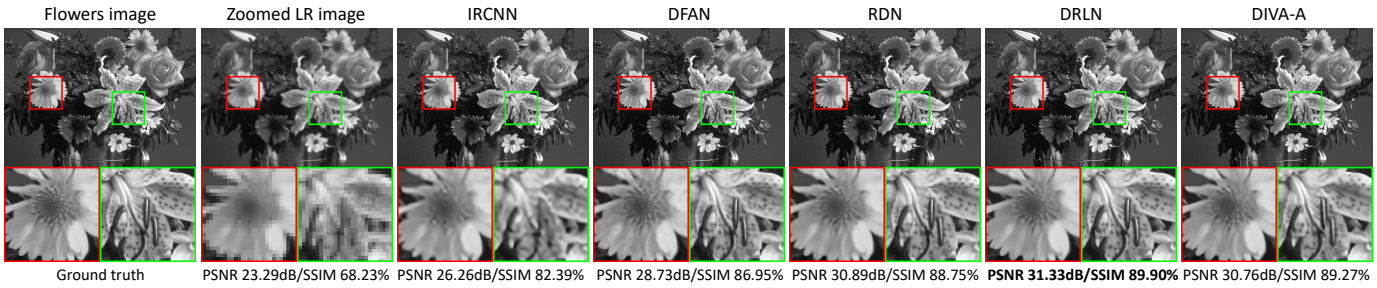


Fig. 17. Restored HR *Flowers* images from LR images generated by Gaussian downsampling under a 7×7 Gaussian blur kernel of standard deviation 1.6 with scaling factor 3.

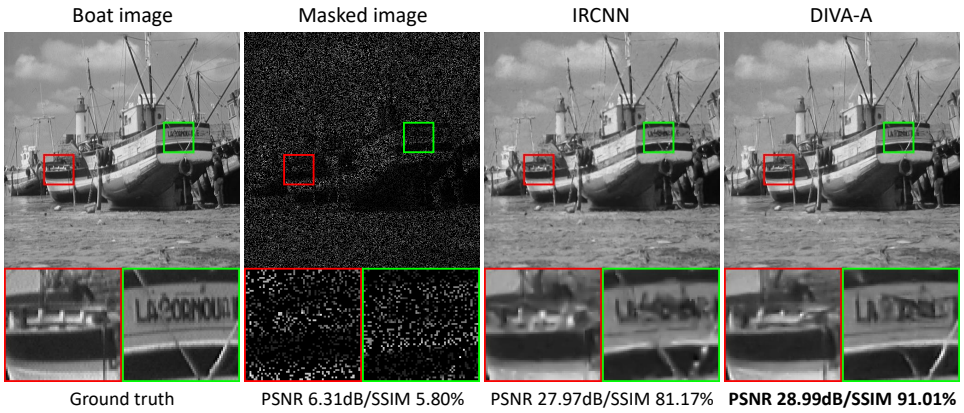


Fig. 18. Restored *Boat* images, when 80% pixels' are missing.

8.2.3 Image Inpainting

The visual analysis of Fig. 18 confirms the excellence of our DL model in inpainting tasks. In the *Boat* image, despite 80% of data missing our model recovers minute details like the ropes and structures on the deck. On the contrary, the image restored by IRCNN [31] is more blurry and loses/distorts many details, such as the borders, sharp edges, and ropes in the restored output. Hence, our model can gather local information from the image neighborhood quite promisingly through the quantum interaction layer and delivers a high-quality restored image even with limited pixels available.