Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

*Digital Object Identifier 10.1109/ACCESS.2017.DOI*

# A Non-local Enhanced Network for Image Restoration

**YUAN HUANG[1], XINGSONG HOU[1], YUJIE DUN[1], ZAN CHEN.[2], and XUEMING QIAN[1]**
[1]School of Information and Communication Engineering, Xi'an Jiaotong University, Xi'an 710049, China
[2]College of Information Engineering, Zhejiang University of Technology, Hangzhou 310014, China

Corresponding author: Xingsong Hou (e-mail: houxs@mail.xjtu.edu.cn).

**ABSTRACT** Non-local modules have been widely studied in image restoration (IR) tasks since they can learn long-range dependencies to enhance local features. However, most existing non-local modules still focus on extracting long-range dependencies within a single image or feature map. On the other hand, most IR methods simply employ a single type of non-local module in the network. A combination of various types of non-local modules to enhance local features can be more effective. In this paper, we propose a batch-wise non-local module to explore richer non-local dependencies within images. Furthermore, we combine various non-local extractors (different attention modules) with the proposed batch-wise non-local module as the Enhanced Batch-wise Non-local Attentive module (EBNA). Besides exploring richer non-local information, we build the Non-local and Local Information extracting Block (NLIB), in which we combine the EBNA with DEformable-Convolution Block (DECB) to utilize richer non-local and adaptive local information. Finally, We embed the NLIB within a U-net-like structure and build the Non-local Enhanced Network (NLENet). Extensive experiments on synthetic image denoising, real image denoising, JPEG artifacts removal, and real image super resolution tasks demonstrate that our proposed network achieves state-of-the-art performance on several IR benchmark datasets.

**INDEX TERMS** image restoration, non-local information, synthetic image denoising, real image denoising, JPEG artifacts removal, real image super resolution

## I. INTRODUCTION

Image restoration is a classic computer vision task that aims to restore high-quality image from its various degradation. It has been widely applied in many practical applications, such as medical image processing [1], [2], surveillance [3]–[5], synthetic aperture radar (SAR) image processing [6]–[8], image compression [9], and so on.

Traditional methods build handcrafted models to solve the image restoration problem based on specific degradation prior knowledge [10], [11]. However, such kinds of methods, including Block-Matching and 3D filtering (BM3D) [12], non-local means (NLM) [13], sparse coding [14], usually have limited robustness towards real-world data. To remedy this problem, the recent deep neural network (DNN) based methods tend to learn the parameters of the model using massive paired data in specific degradation, such as SRCNN [15], DnCNN [16] and RDN [17].

SRCNN [15] first introduced the convolution neural network to IR for the image super resolution task. Recent development in DNN showed that larger receptive fields could learn informative features from a larger neighborhood in the image. Therefore, more and more researchers try to build deeper and wider networks to improve the restoration performance, such as RDN [17], VDSR [18], and EDSR [19]. Another way of learning from larger receptive fields is employing wavelet decomposition to generate multi-scale inputs. Such as in divide-and-conquer framework [20], authors first decomposed images to multiple subspaces according to the visual importance and used different models to preserve texture details based on prior knowledge. In another work, authors tried to use CNN-based models for sparse coding (DCSC) [21]. In DCSC, the features were sparsely coded by employing CNN models instead of the handcrafted coding method. In MWCNN [22], authors employ wavelet decomposition and reconstruction to generate multi-scale features.

We can acquire non-local information in multiple ways, such as employing self non-local modules (based on non-local means [13]), global pooling modules (channel-wise and spatial-wise attention), multi-scale inputs, long-range connections and recurrent connections in the network. In

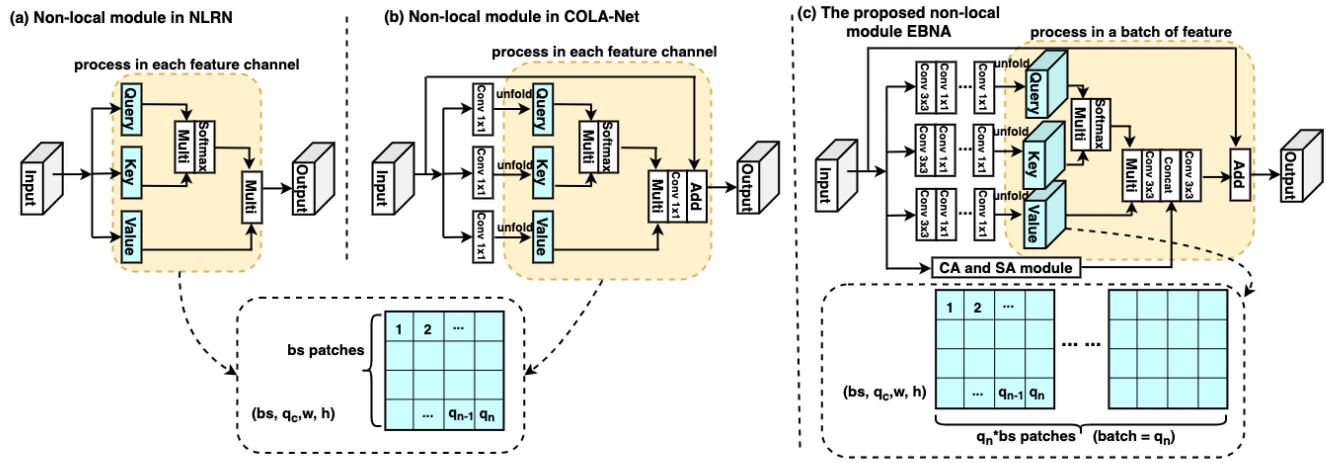


**FIGURE 1.** A comparison of the existing non-local modules with the proposed batch-wise non-local module in EBNA.

IR approaches, for example, many networks employ global pooling to extract long-range dependencies, such as RCAN [23], [24] and RIDNet [25]. In these attention modules, they squeeze the feature map (channel-wise or spatial-wise) and generate attentive weights [26]–[28] based on the whole channel of the feature map or the spatial location of the feature map (which is used as the non-local features). For networks like MemNet [29] and Rednet [30], they pass non-local information through recurrent connections to obtain richer features. COLA-Net [31] builds a non-local module that uses patches from a single image (or feature map) to extract long-range dependencies.

However, there still exist several problems in the above IR approaches. First of all, self non-local information has been explored in methods such as COLA-Net [31] and NLRN [32], in which they ignore the helpful patch-wise non-local information among multiple images. Secondly, most IR models employ a single type of non-local module in the network, limiting the feature extracting ability. Combining various types of non-local modules to enhance local features can be more effective for IR. Finally, besides the non-local features that capture the long-range dependency, more sophisticated local features can complement the restored local texture.

We propose the batch-wise non-local module to extract sophisticated non-local information and build long-range dependencies to tackle the above problems. Different from the previous self non-local modules [31], [33], our proposed batch-wise non-local module can fuse the prior and relevance from a batch of images (or feature maps), which can restore more contextual details.

To further explore diverse information from non-local regions, we propose a novel block named EBNA, which combines the proposed batch-wise non-local module and various existing non-local modules. In the proposed batch-wise non-local module, we intend to extract richer information from a batch of images instead of using only one image. Especially in situations when self-similarity is limited within one image, the chance of extracting more relevant information from a batch of images is increased. Unlike the non-local modules that employ global pooling to extract non-local relatedness, the batch-wise non-local module generates the non-local feature based on patch-wise relation within feature maps. In contrast, channel-wise attention (CA) and spatial-wise attention (SA) module extract non-local features by global pooling weights among channels and spatial location within the features. Combining the three could enhance each other and generate more diverse non-local features to improve the restoration of the textural and contextual details in the images.

FIGURE 1 shows the difference between our proposed EBNA module and several existing non-local modules. The highlighted parts in FIGURE 1 demonstrate the difference between the self non-local module and the proposed batch-wise non-local module. They match the patches within a single channel of features and a batch of features, respectively. We can observe that the batch-wise non-local modules in EBNA can explore long-range dependencies within images and extract more sophisticated non-local features. In EBNA, the combination of various non-local modules can generate more diversified features compared to existing networks which employ a single type of non-local module.

Furthermore, based on EBNA, we build a novel block named NLIB to collaborate the local and non-local features. In NLIB, we employ a DEformable-Convolution Block (DECB) to extract local features. Deformable convolution can learn local features from the adaptive receptive field, but the limited receptive field size still restricts the module from learning non-local information. Such cooperation between DECB and EBNA can extract more enhanced features from both local and non-local regions of the image.

Finally, we stack the NLIBs in a U-net-like multi-scale structure model and build the Non-local Enhanced Network (NLENet). Extensive experiments show that NLENet achieves state-of-the-art performance on several IR task benchmark datasets.

The main contributions of the paper can be summarized as follows:

- We propose a novel batch-wise non-local module to explore non-local dependencies among images and build a novel block called EBNA that combines various complementary non-local information.
- To cooperate non-local information with adaptive local information, we further employ EBNA together with DEformable-Convolution Block (DECB) as a new module NLIB. Based on the NLIB, we propose our final model, NLENet, which utilizes various non-local information and the adaptive local feature to improve IR performance.
- Extensive experiments on synthetic image denoising, real image denoising, JPEG artifacts removal and real image super resolution tasks show that the proposed model achieves state-of-the-art performance. Furthermore, the ablation study also demonstrates the superiority of the proposed network.

The rest of this paper is organized as follows. In section II, we introduce the related works. In section III, we present the structural details of our proposed model. Extensive experiments are conducted in section IV to evaluate the effectiveness of the proposed network on synthetic image denoising, real image denoising, JPEG artifacts removal and real image super resolution tasks. Furthermore, ablation study is presented in section V. The conclusion is given in section VI.

## II. RELATED WORKS

In this section, we give a brief review of the works related to our proposed network. We first list the typical traditional model-based and recent state-of-the-art DNN based IR methods. Then, we briefly introduce the typical non-local operations and local feature extraction approaches in IR.

### A. IMAGE RESTORATION

Image restoration, as a fundamental component in the image processing area, has been widely studied for decades. Traditional IR methods like BM3D [12], SA-DCT [34], and TNRD [35] have provided reasonable results on both accuracy and robustness. However, these algorithms usually have drawbacks, such as high complexity and limited generalization.

Recently DNN based IR methods have gained considerable attention and significant performance improvement. Researchers develop deeper, wider models to acquire larger receptive fields and extract pixel-wise relations from a larger region. SRCNN [15] first introduced CNN for IR tasks. Based on the CNN structure, researchers developed VDSR [18]. In VDSR, a structure that consists of several cascading filters was proposed to broaden the receptive field and increase the depth of the model. In building VDSR, the authors found that increasing the depth of the model could bring performance improvement. To solve the gradient descent problem in training deeper models, DRCN [36] proposed a deeper model together with a gradient clipping and recursive-supervision method, which increased the IR performance significantly. DnCNN [16] introduced the residual connection to ease the propagating of feature flow and solve the gradient vanishing problem in deep IR models. In another work RCAN [23], a deeper model with a residual in residual structure was proposed. RCAN also introduced the channel-wise attention module within the residual in residual structure to obtain a deep network and learn more adaptive features simultaneously. To increase the IR model's efficiency, RDN [17] employed dense connection, feature fusing, and residual connection to make full use of the features from different scales. In this paper, we propose a novel network that employs both non-local and local features to improve IR performance, as shown in section III.

### B. NON-LOCAL OPERATION

Non-local information has been explored in many areas, such as extracting relevance in video processing [37], [38], building long-range dependency among a sequence of words in natural language processing [39] and text summarization [40]. Besides extracting long-range dependency in the time domain, non-local operations also can build relevance in the space domain, such as in computer vision tasks.

In computer vision area, non-local operations that extracts long-range dependency among pixels, has been used for many tasks, for example object detection [33], [41], semantic segmentation [42], [43], video action recognition [44], image compressive sensing [45], and image restoration [13], [31], [46], [47]. To better understand the non-local operation's efficacy, we can observe it as an attention mechanism for pixel-to-pixel relation modeling. This relation is modeled as the dot-product between the features of two pixels. The larger the dot-product value indicates more relevance of the two pixels.

At first, traditional methods usually apply Additive White Gaussian Noise (AWGN), TV regularization [48], Fourier domain [49] or wavelet domain [50] coefficients transform in different IR tasks. However, it is the idea of non-local means (NLM) denoising [13] that brought the importance of long-range dependencies into IR tasks. Non-local means methods are built upon self-similarity and redundant information over realistic images. Later on, another non-local denoising approach BM3D [12] was developed.

As for DNN based methods, NLNet first introduced deep neural networks to perform non-local processing for color image denoising task which achieved remarkable performance. Non-local information is also widely explored in image super resolution area [51]–[53]. Such as in the NLSN [51], they proposed a Non-Local Sparse Attention (NLSA) with dynamic sparse attention pattern module to generate non-local attention with spherical locality sensitive hashing (LSH). Furthermore, in MHNAN [52], non-local information is extracted by a Mixed High-Order Attention (MHA) module. In another work, COLA-Net [31] tried to build a learnable non-local module to extract long-range dependencies within the degraded image. However, it only extracts relevance within one single image, which lacks non-local

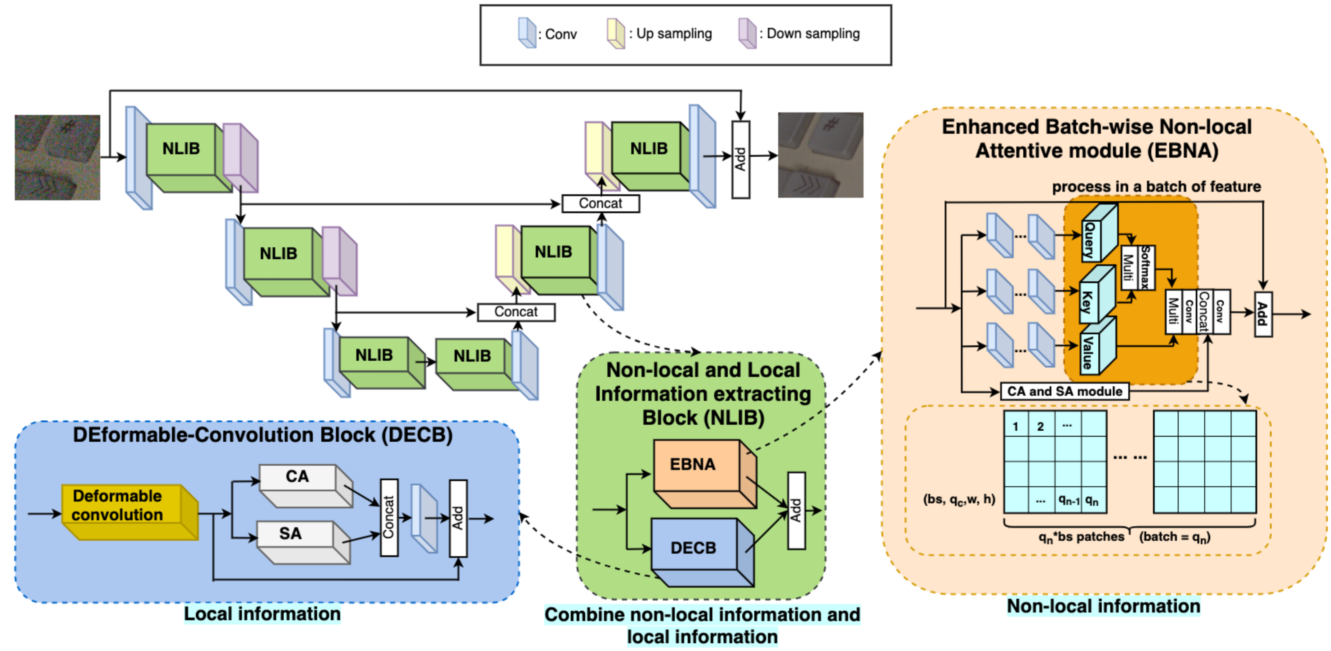


FIGURE 2. An overview of the proposed NLENet.The NLIB (Non-local and Local Information extracting Block) is build upon EBNA (Enhanced Batch-wise Non-local and Attentive modules) and DECB (DEformable-Convolution Block).

information among multiple images. In contrast, we proposed a novel batch-wise non-local module to extract the non-local information among multiple images, which has not been studied in the existing non-local methods. Based on the proposed batch-wise non-local module, we proposed the NLENet, which combines various non-local features to enhance the local feature and preserve more contextual details in IR.

## III. PROPOSED NETWORK

In this paper, we proposed a novel network, NLENet, for IR tasks. Here we present an overview of the proposed IR network, including the models for synthetic image denoising, real image denoising, JPEG artifacts removal and real image super resolution. FIGURE 2 illustrates the overall architecture of the proposed network, which is a multi-scale structure embedded with the proposed block NLIB. We can observe that NLIB consists of two proposed blocks EBNA and DECB. EBNA, based on the proposed batch-wise non-local module, collaborates with different types of non-local extractors to build an enhanced batch-wise non-local and attentive module. And DECB explores the local information by employing deformable convolution. With both modules, NLIB can combine the local feature and non-local feature. More concretely, (1) we propose a batch-wise non-local module to explore the relevance among images; (2) based on the proposed batch-wise non-local module, we propose EBNA, which provides enriched non-local features from various types of non-local modules; (3) to fully utilize the non-local and local information, we propose NLIB built upon EBNA and DECB. We stack NLIB in a U-net structure model to build NLENet, utilizing enriched non-local and local features to preserve better contextual details in the restored images.

### A. BATCH-WISE NON-LOCAL MODULE

Following the idea of non-local means operation [13], the generic non-local operation can be defined as:

$$y_j = \frac{1}{S} \sum_{i,j \in I} f(x_i, x_j) \cdot g(x_j), \tag{1}$$

in which the output patch $y_j$ at position $j$ has the same size as the input $x_j$, $S$ represents a normalization factor and $i, j$ represent different patches in image $I$. $f()$ is a scalar to compute the affinity between the patch $x_j$ and patch $x_i$, which represents the relationship between two patches. $g()$ is an embedding function that transforms the input $x_j$ to another representation domain. In this way, the non-local operation uses all the predictable information within a single image to restore the current patch. Further applying this idea in the DNN based models, the non-local module employs the same process within each channel of the feature maps to explore self-predictable information.

We extend this search region of predictable information from one single image to a batch of images in our work. Similar patches of pixels (or feature maps) are searched to generate more abundant predictable information. We reform the single image non-local operation to batch-wise non-local (BNL) operation as follows:

$$y_j = \frac{1}{S} \sum_{i,j \in I_{batch}} f(x_i, x_j) \cdot g(x_j), \tag{2}$$

where $i, j$ are from patches of a batch of images $I_{batch}$. Different from Equation 1, our proposed batch-wise non-local

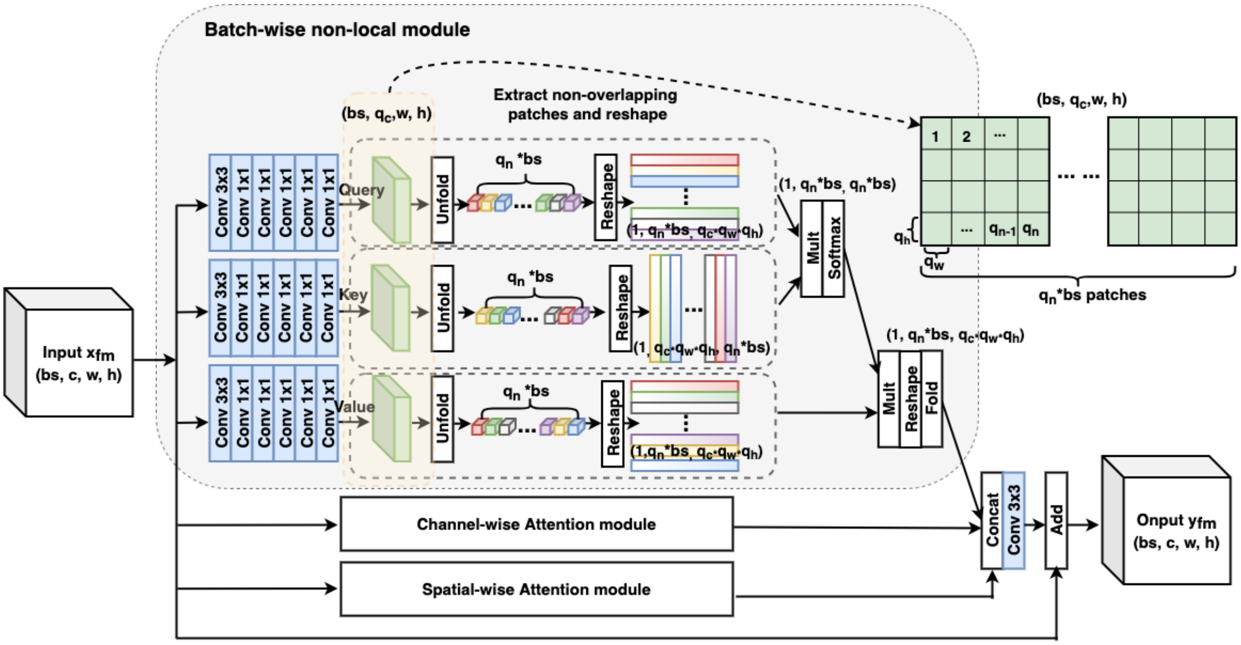


**FIGURE 3.** An overview of the proposed Enhanced Batch-wise Non-local and Attentive (EBNA) module. It combines the proposed batch-wise non-local module with various attention modules to extract richer non-local features in the IR model.

module, in Equation 2, expand the 'non-local region' from a single image to a batch of images. In the existing self non-local module, patches from a single image are cropped and perform the patch matching process. While in the proposed batch-wise non-local module, the patches are extracted from a batch of images where more relevant information can be found.

FIGURE 3 shows the batch-wise non-local module in detail. We take a feature map of size $(bs, c, w, h)$ as an input and $q_n \times bs$ represent as the number of patches unfolded from a batch of feature maps ($w.r.t.$as Query, Key and Value). $q_c$ represents the number of channels of the feature map, and $q_w$ and $q_h$ are the width and height of the patches (we set patch size as $4 \times 4$). Then the Query feature map is reshaped and multiplied with the Key feature map to generate a weight matrix. As an evaluation of the relevance among patches, the weight matrix is multiplied with the value feature map to generate the non-local feature map.

In the batch-wise non-local module, richer information can be extracted from a batch of images instead of only one single image. Especially in situations when self-similarity is limited within one image, the chance of extracting more relevant information from a batch of images is increased.

### B. ENHANCED BATCH-WISE NON-LOCAL AND ATTENTIVE MODULE

In this section, we describe the proposed EBNA module in detail. The main idea of developing the EBNA module is to employ diverse non-local information to enhance the local features. Besides the proposed batch-wise non-local module (based on patch-wise relevance among feature maps), we use different modules that employ global pooling operations to extract diverse non-local information.

Non-local relevance can be extracted through various operations, such as global pooling operation (channel-wise attention [54] and spatial-wise attention [55] et al.). A set of weights as long-range dependencies are built on a specific channel or spatial location feature in CA and SA.

To take advantage of the above non-local modules, we cooperate CA and SA with the proposed batch-wise non-local module to build a novel block EBNA. FIGURE 3 shows the overview of the proposed EBNA module.

Taking feature map $x_{fm} \in \mathbb{R}^{bs \times c \times w \times h}$ as input and $y_{fm} \in \mathbb{R}^{bs \times c \times w \times h}$ as output, the EBNA module can be defined as follows:

$$\begin{aligned} y_{fm} =& Conv(Concat(F_{BNL}(x_{fm}), F_{CA}(x_{fm}), F_{SA}(x_{fm}))) \\ &+ x_{fm}, \end{aligned} \tag{3}$$

where $Conv()$ and $Concat()$ represents the convolution operation and feature maps concatenation operation. $F_{BNL}()$ represents the proposed batch-wise non-local module, $F_{CA}()$ and $F_{SA}()$ represent channel-wise and spatial-wise attention modules, respectively.

In EBNA, the proposed batch-wise non-local module can extract patch-wise dependencies from a batch of images and build relevance weights among image patches. In contrast, CA and SA modules can extract more general dependencies across channels and spatial locations. Combining the proposed batch-wise non-local module and attention modules can generate enhanced non-local features and extract more diverse non-local information among features.

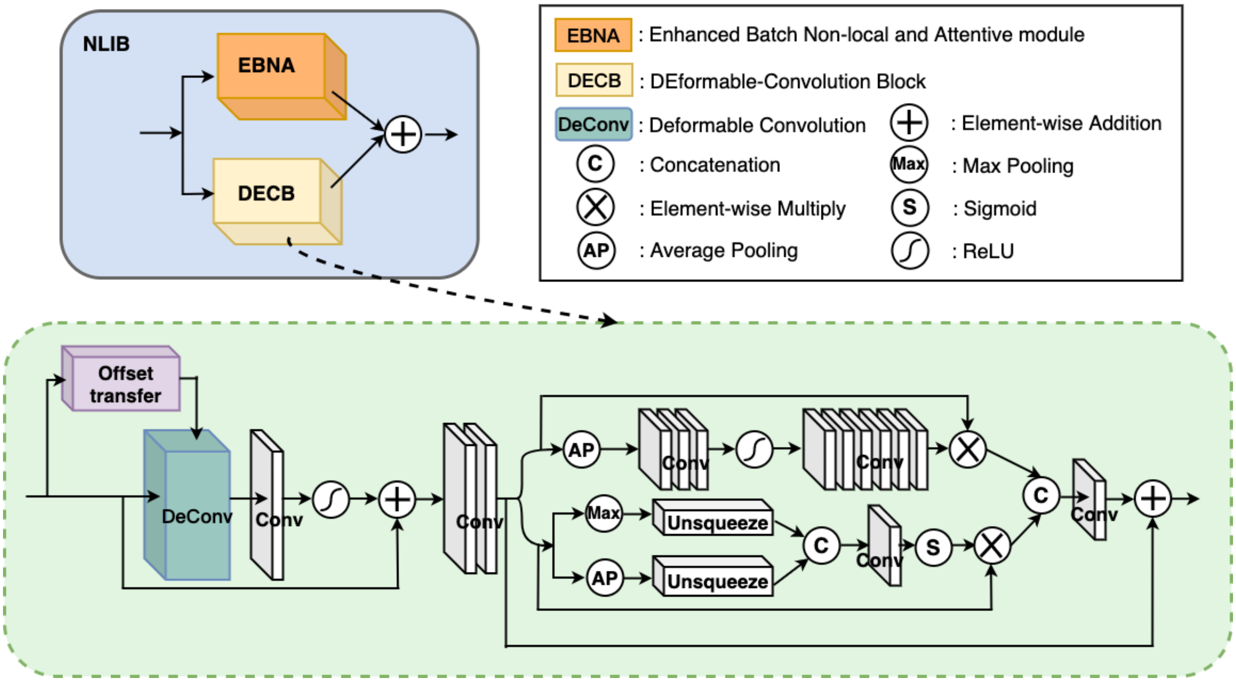


FIGURE 4. An overview of the proposed Non-local and Local Information extracting Block (NLIB) along with the structure details of DECB. It combines the proposed non-local module EBNA with a adaptive local module DECB to employ adaptive local features and long range dependencies in the feature domain.

### C. NLIB

Besides using abundant non-local information, local features are also essential in IR models. Based on the above concept we build a block named as NLIB which combine various non-local and adaptive local information from the degraded image.

As shown in FIGURE 4, the proposed NLIB consists of two parts, in which EBNA can extract various non-local dependencies and DECB can extract local features. We build the local feature extractor based on the deformable convolution. The DECB includes the deformable convolution and an attentive operation. The deformable convolution can learn from adaptive receptive fields, while the attentive operation with a residual connection can help extract adaptive features with focus. Thus, sophisticated local features can be acquired.

NLIB can be defined as follows:

$$y_{fm} = Conv(F_{EBNA}(x_{fm}) + F_{DECB}(x_{fm})), \quad (4)$$

where $Conv()$ represents the convolution operation. $F_{EBNA}()$ represents the proposed EBNA module, $F_{DECB}()$ represents DECB module. The output feature maps of EBNA and DECB are added together through a convolution layer. In NLIB, various non-local features can enhance the local features without the limits of the receptive field and learn from long-range dependencies. Both non-local and local information extracted by NLIB can help to restore more structure and texture details. Furthermore, as the basic component of NLENet, we apply NLIB at every down-sampling and up-sampling stage to extract richer features at each scales.

During training, given the corrupted images $\{\hat{I}_i\}_{i=1}^N, \hat{I}_i \in \mathbb{R}^{H\times W\times C}$ ($H$ as the height, $W$ as the width and $C$ as the channel of the image) as inputs, NLENet learns a mapping function $f_\theta$ with a set of parameters $\theta$ in generating the corresponding restored images $\{I_i\}_{i=1}^N, I_i \in \mathbb{R}^{H\times W\times C}$, by employing $\ell_2$ loss function formulated as follows:

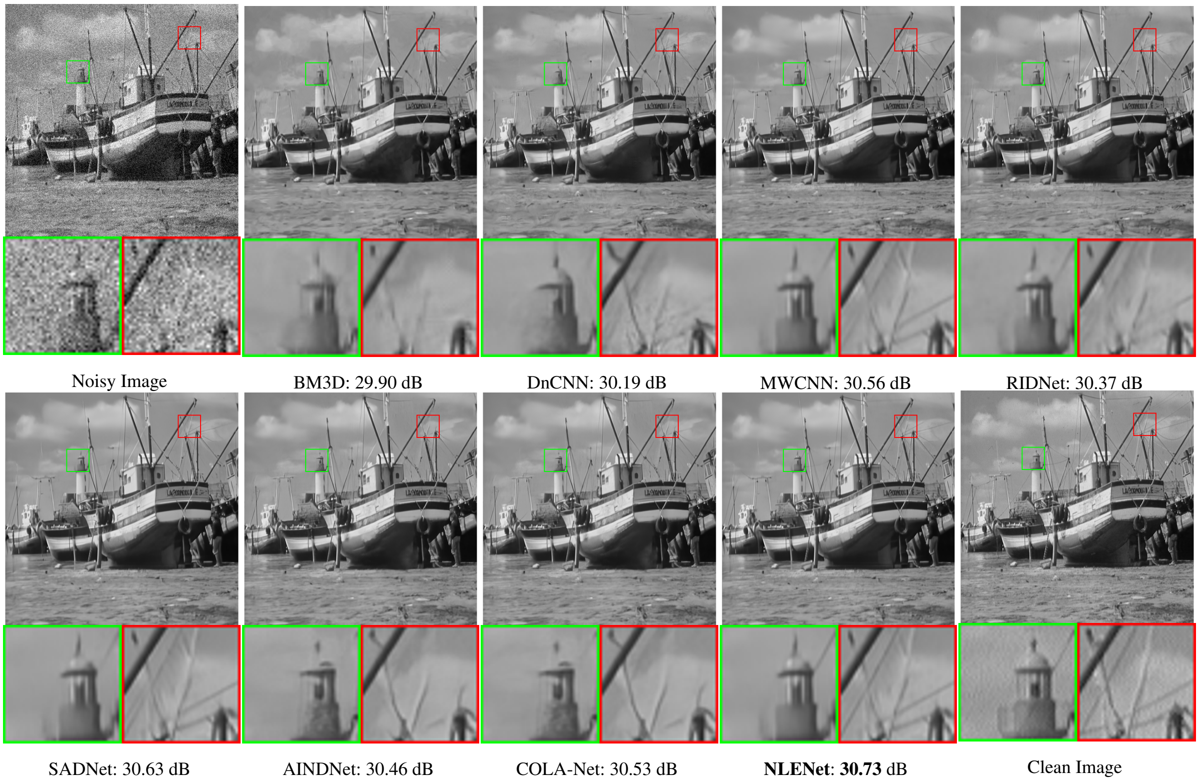
$$\mathcal{L} = \arg\min_\theta \sum_{i=1}^{N} \|f_\theta(\hat{I}_i) - I_i\|^2. \quad (5)$$

## IV. EXPERIMENTS

We perform extensive experiments to demonstrate the proposed NELNet's effectiveness on four IR tasks in this section. The evaluated IR tasks including (a) synthetic image denoising, (b) real image denoising, (c) JPEG artifacts removal, and (d) real image super resolution. We test on several benchmark datasets in each task to give a thorough performance evaluation of the proposed network.

The model is trained with the Adam optimizer under setting $\beta_1 = 0.9$, $\beta_2 = 0.999$. We train the models with an initial learning rate $1\times10^{-4}$ and gradually decrease to $1\times10^{-6}$. During training, we apply data augmentation (including random horizontal and vertical flipping) for better performance. The training batch size set as 6 with the patch size 256 × 256. Similar settings are used in all four IR tasks. The experiments are conducted on NVIDIA Tesla V100 with the PyTorch library [56].

We list the results comparisons in terms of PSNR, SSIM in TABLE 1,2,3 (for methods which code and model are not available, we can only compare their published results from


Noisy Image

BM3D: 29.90 dB

DnCNN: 30.19 dB

MWCNN: 30.56 dB

RIDNet: 30.37 dB

SADNet: 30.63 dB

AINDNet: 30.46 dB

COLA-Net: 30.53 dB

**NLENet**: **30.73** dB

Clean Image

FIGURE 5. Synthetic image denoising results (“boats.png” from *set12*) using different denoising networks. The AWGN noise level is ($\sigma = 25$). The best result is highlighted in bold.

the original paper and "NA" is placed because of missing results in some cases.). We will release the pre-trained models along with the source code upon the acceptance of the paper.

## A. SYNTHETIC IMAGE DENOISING

This section shows the comparison results of the proposed NLENet for AWGN denoising on grayscale images. We train the synthetic image denoising models with the training set of DIV2K [66] in grayscale. Then we evaluate the trained models on Set12, BSD68 [58], and Urban100 [59] datasets, which are commonly used in synthetic images denoising task. To fully validate the proposed network's denoising ability, we train models with AWGN under different levels of noise, *i.e.*, $\sigma$=15, 25, 50, and 75 (standard deviation $\sigma$) and compared with the SOTA methods, which are listed in TABLE 1.

TABLE 1 presents quantitative comparisons of PSNR and SSIM [57], where we can observe that the proposed NLENet outperforms the traditional and latest SOTA CNN based denoising methods at most noise levels. Specifically, compared to the latest non-local networks COLA-Net, our algorithm demonstrates a performance improvement of 0.02 to 0.1 dB in PSNR at different noise levels on all test sets.

We also give a visual compassion of the denoised results from the proposed method and latest methods in FIGURE 5. From the FIGURE 5, we can easily find that methods like DnCNN lose fine details. In the visual results of RIDNet and AINDNet, the restored images obtain blurred edges. Compared to the latest non-local network COLA-Net, the denoised result lose part of the lines in the zoom-in area. While our NLENet preserves clear lines. Therefore, NLENet is able to reconstruct the structural information and fine texture of the noisy image.

## B. REAL IMAGE DENOISING

To further demonstrate the merits of our proposed method, we compare the proposed network with several SOTA real image denoising approaches. Unlike synthetic image denoising, real images are corrupted by realistic noise during capturing, and we have no prior knowledge of the noise distribution. We train the real image denoising model with the training set of SIDD medium [67]. And SIDD [67] and DND [68] test sets are used as evaluation. The training set of SIDD medium [67] contains 320 very high-resolution image pairs captured by smartphones under different environments. Moreover, the test set of SIDD contains 1280 images of size 512×512. And DND [68] which has 1000 images of size 512×512.

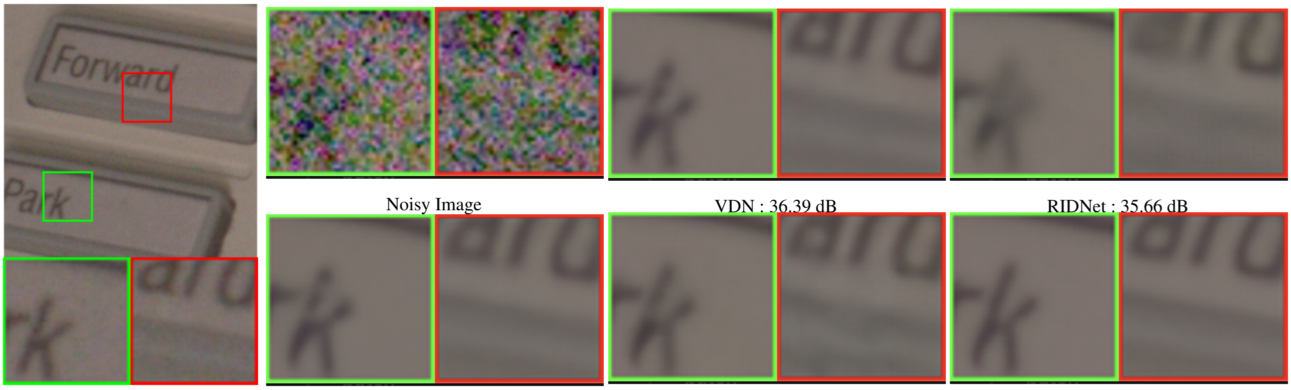
Results comparisons are summarized in Table 2. We can observe that our NLENet outperforms the latest methods on SIDD and achieves competitive results on DND. For instance, compared with another non-local network COLA-Net and latest method GNSCNet, the PSNR results of NLENet are about 0.6∼0.7 dB (in PSNR) higher on SIDD. As for SSIM, GNSCNet achieves the highest result in the comparing methods, NLENet achieves the second-highest result. How-

**TABLE 1.** Average PSNR (dB) and SSIM [57] results of different denoising methods for grayscale synthetic image denoising (AWGN), evaluate on *Set12*, *BSD68* [58], *Urban100* [59] with noise levels $\sigma = \{15, 25, 50, 75\}$. The best results are highlighted in **bold**. "NA" means "Not Available" due to unavailable code or model.

| Dataset | $\sigma$ | BM3D | DnCNN | MWCNN | NLRN | SADNet | RIDNet | AINDNet | GNSCNet | DudeNet | GCDN | DAGL | COLA-Net | Proposed |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | [12] | [16] | [22] | [32] | [60] | [25] | [61] | [62] | [63] | [64] | [65] | [31] | |
| *Set12* | 15 | 32.37 | 32.86 | 33.15 | 33.16 | 33.18 | 32.91 | 32.92 | 33.19 | 32.94 | 33.14 | 33.28 | 33.27 | **33.29** |
| | | 0.8952 | 0.9027 | 0.9088 | 0.9070 | 0.9127 | 0.9085 | NA | 0.9019 | NA | 0.9072 | 0.9100 | 0.9097 | **0.9136** |
| | 25 | 29.97 | 30.43 | 30.79 | 30.80 | 30.88 | 30.60 | 30.62 | 30.80 | 30.52 | 30.78 | 30.93 | 30.90 | **30.99** |
| | | 0.8505 | 0.8618 | 0.8711 | 0.8689 | 0.8752 | 0.8694 | NA | 0.8715 | NA | 0.8687 | 0.8720 | 0.8716 | **0.8773** |
| | 50 | 26.72 | 27.18 | 27.74 | 27.64 | 27.80 | 27.43 | 27.51 | 27.72 | 27.30 | 27.60 | 27.81 | 27.77 | **27.91** |
| | | 0.7676 | 0.7827 | 0.8056 | 0.7980 | 0.8069 | 0.7944 | NA | 0.8049 | NA | 0.7957 | 0.8042 | 0.8032 | **0.8091** |
| | 75 | 24.91 | 25.20 | 25.88 | NA | 26.02 | 25.72 | NA | NA | NA | NA | NA | 25.92 | **26.09** |
| | | 0.6950 | 0.7095 | 0.7487 | NA | 0.7542 | 0.7406 | NA | NA | NA | NA | NA | 0.7398 | **0.7561** |
| *BSD68* | 15 | 31.08 | 31.72 | 31.86 | 31.88 | 31.89 | 31.81 | 31.69 | 31.90 | 31.78 | 31.83 | 31.93 | 31.92 | **31.94** |
| | | 0.8722 | 0.8906 | 0.8947 | 0.8932 | 0.9008 | 0.8982 | NA | 0.8956 | NA | 0.8933 | 0.8953 | 0.8968 | **0.9012** |
| | 25 | 28.57 | 29.23 | 29.41 | 29.41 | 29.46 | 29.34 | 29.26 | 29.43 | 29.29 | 29.46 | 29.35 | 29.46 | **29.51** |
| | | 0.8017 | 0.8278 | 0.8360 | 0.8331 | 0.8431 | 0.8381 | NA | 0.8367 | NA | 0.8332 | 0.8366 | 0.8368 | **0.8452** |
| | 50 | 25.62 | 26.23 | 26.48 | 26.47 | 26.50 | 26.40 | 26.32 | 26.52 | 26.31 | 26.38 | 26.51 | 26.52 | **26.60** |
| | | 0.6869 | 0.7189 | 0.7366 | 0.7298 | 0.7382 | 0.7314 | NA | 0.7364 | NA | 0.7389 | 0.7334 | 0.7340 | **0.7423** |
| | 75 | 24.21 | 24.64 | 24.98 | NA | 25.05 | 24.89 | NA | NA | NA | NA | NA | 24.98 | **25.10** |
| | | 0.6139 | 0.6401 | 0.6707 | NA | 0.6742 | 0.6639 | NA | NA | NA | NA | NA | 0.6637 | **0.6766** |
| *Urban100* | 15 | 32.34 | 32.67 | 33.17 | 33.45 | 33.21 | 33.09 | NA | 33.34 | NA | 33.47 | **33.79** | 33.73 | 33.65 |
| | | 0.9220 | 0.9250 | 0.9088 | 0.9354 | 0.9104 | 0.9364 | NA | 0.9370 | NA | 0.9358 | 0.9393 | 0.9387 | **0.9419** |
| | 25 | 29.70 | 29.97 | 30.66 | 30.94 | 30.71 | 30.53 | NA | 30.81 | NA | 30.95 | **31.39** | 31.33 | 31.37 |
| | | 0.8777 | 0.8792 | 0.9026 | 0.9018 | 0.9033 | 0.9009 | NA | 0.9042 | NA | 0.9020 | 0.9093 | 0.9086 | **0.9143** |
| | 50 | 25.94 | 26.28 | 27.42 | 27.49 | 27.75 | 27.05 | NA | 27.50 | NA | 27.41 | 27.97 | 27.84 | **28.12** |
| | | 0.7791 | 0.7869 | 0.8371 | 0.8279 | 0.8380 | 0.8242 | NA | 0.8392 | NA | 0.8160 | 0.8423 | 0.8372 | **0.8526** |
| | 75 | 23.91 | 23.94 | 25.52 | NA | 25.95 | 25.22 | NA | NA | NA | NA | NA | 25.81 | **26.25** |
| | | 0.6950 | 0.6989 | 0.7810 | NA | 0.7958 | 0.7639 | NA | NA | NA | NA | NA | 0.7561 | **0.8020** |

**TABLE 2.** Average PSNR (dB) and SSIM [57] of different methods for real image denoising evaluate on the *SIDD* [67] and the *DND* [68]. The best results are highlighted in **bold**. "NA" means "Not Available" due to unavailable code or model.

| Dataset | CBM3D | TNRD | MLP | DnCNN | CBDNet | SADNet | RIDNet | VDN | AINDNet | GNSCNet | COLA-Net | Proposed |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | [69] | [35] | [70] | [16] | [71] | [60] | [25] | [72] | [61] | [62] | [31] | |
| *SIDD* | 25.65 | 24.73 | 24.71 | 23.66 | 30.78 | 39.36 | 38.71 | 39.28 | 39.15 | 39.31 | 39.10 | **39.86** |
| | 0.685 | 0.643 | 0.641 | 0.583 | 0.754 | NA | 0.914 | 0.909 | 0.905 | **0.955** | 0.905 | 0.916 |
| *DND* | 34.51 | 33.65 | 34.23 | 37.90 | 38.06 | 39.37 | 39.26 | 39.38 | **39.77** | 39.43 | 39.64 | 39.48 |
| | 0.851 | 0.831 | 0.833 | 0.943 | 0.942 | 0.954 | 0.952 | 0.952 | **0.955** | 0.953 | 0.954 | 0.952 |



Noisy Image

VDN : 36.39 dB

RIDNet : 35.66 dB

Clean Image

SADNet : 36.72 dB

COLA-Net : 36.14 dB

**NLENet**: **37.30** dB

**FIGURE 6.** Real image denoising results ("*11_4.png*" from *SIDD* [67]) using different denoising networks. The best result (PSNR in dB) is highlighted in bold.

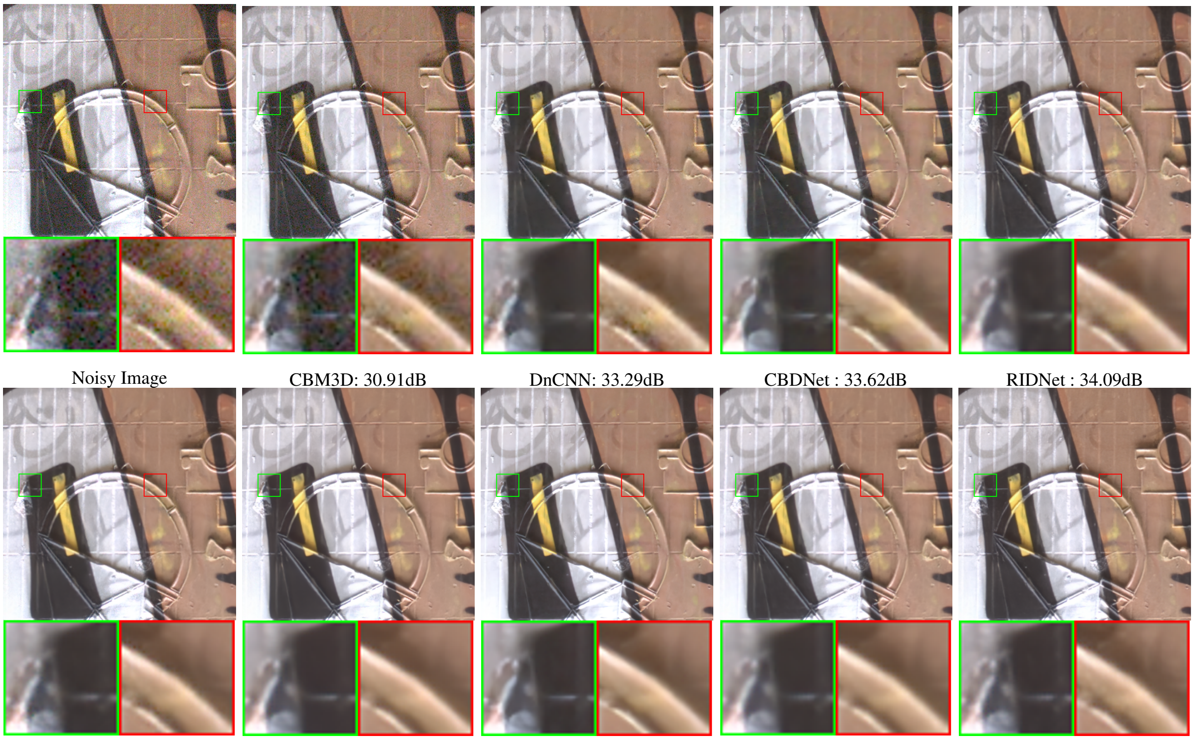


**FIGURE 7.** Real image denoising results ("*0002_19.png*" from *DND* [68]) using different denoising networks. The best result is highlighted in bold.

**TABLE 3.** Average PSNR (dB) and SSIM [57] results of different methods for JPEG artifacts removal, evaluate on *Classic5*, *LIVE1* [73] with quality factors $Q = \{10, 20, 30, 40\}$. The best results are highlighted in bold. "NA" means "Not Available" due to unavailable code or model.

| Dataset | $Q$ | JPEG | SA-DCT<br>[34] | ARCNN<br>[74] | DnCNN<br>[16] | MemNet<br>[29] | RNAN<br>[75] | MWCNN<br>[22] | QGAC<br>[76] | $D^2$CIR<br>[77] | DUN<br>[78] | DAGL<br>[65] | COLA-Net<br>[31] | Proposed |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Classic5* | 10 | 27.82 | 28.88 | 29.03 | 29.04 | 29.69 | 29.96 | 30.01 | 29.84 | 30.05 | 29.95 | 30.08 | 30.03 | **30.26** |
| | | 0.7595 | 0.8071 | 0.7929 | 0.8026 | 0.8107 | 0.8178 | 0.8195 | **0.8370** | 0.8202 | 0.8171 | 0.8196 | 0.8184 | 0.8329 |
| | 20 | 30.12 | 30.92 | 31.15 | 31.63 | 31.90 | 32.11 | 32.16 | 31.98 | 32.20 | 32.11 | 32.35 | 32.28 | **32.44** |
| | | 0.8344 | 0.8663 | 0.8517 | 0.8610 | 0.8658 | 0.8693 | 0.8701 | **0.8850** | 0.8707 | 0.8689 | 0.8719 | 0.8705 | 0.8823 |
| | 30 | 31.48 | 32.14 | 32.51 | 32.91 | 33.20 | 33.38 | 33.43 | 33.22 | 33.44 | 33.33 | 33.59 | 33.54 | **33.64** |
| | | 0.8744 | 0.8914 | 0.8806 | 0.8861 | 0.8980 | 0.8924 | 0.8930 | **0.9070** | 0.8935 | 0.8916 | 0.8942 | 0.8935 | 0.9030 |
| | 40 | 32.43 | 33.00 | 33.34 | 33.96 | 34.06 | 34.27 | 34.27 | NA | 34.29 | 34.10 | 34.41 | 34.38 | **34.47** |
| | | 0.8911 | 0.9055 | 0.8953 | 0.9047 | 0.9052 | 0.9061 | 0.9061 | NA | 0.9066 | 0.9045 | 0.9069 | 0.9066 | **0.9151** |
| *LIVE1* | 10 | 27.77 | 28.65 | 28.96 | 29.19 | 29.45 | 29.63 | 29.69 | 29.53 | 29.68 | 29.61 | 29.70 | 29.66 | **29.87** |
| | | 0.7595 | 0.8093 | 0.8076 | 0.8123 | 0.8193 | 0.8239 | 0.8254 | **0.8400** | 0.8253 | 0.8232 | 0.8245 | 0.8234 | 0.8358 |
| | 20 | 30.07 | 30.81 | 31.29 | 31.59 | 31.83 | 32.03 | 32.04 | 31.86 | 32.04 | 31.98 | 32.12 | 32.06 | **32.23** |
| | | 0.8512 | 0.8781 | 0.8733 | 0.8802 | 0.8846 | 0.8877 | 0.8885 | **0.9010** | 0.8882 | 0.8869 | 0.8887 | 0.8880 | 0.8977 |
| | 30 | 31.41 | 32.08 | 32.67 | 32.98 | 33.24 | 33.45 | 33.45 | 33.23 | 33.44 | 33.38 | 33.54 | 33.48 | **33.62** |
| | | 0.9000 | 0.9078 | 0.9043 | 0.9090 | 0.9187 | 0.9149 | 0.9153 | **0.9250** | 0.9154 | 0.9142 | 0.9156 | 0.9152 | 0.9231 |
| | 40 | 32.35 | 32.99 | 33.63 | 33.96 | 34.27 | 34.47 | 34.45 | NA | 34.44 | 34.32 | 34.53 | 34.49 | **34.62** |
| | | 0.9173 | 0.9240 | 0.9198 | 0.9247 | 0.9187 | 0.9061 | 0.9301 | NA | 0.9301 | 0.9289 | 0.9305 | 0.9300 | **0.9369** |

ever, in the synthetic image denoising task, NLENet gains over 0.1 dB in terms of PSNR and 0.02∼0.03 in terms of SSIM higher on all the test sets and noise levels compared to GNSCNet.

As an IR structure with generalization ability, NLENet shows superior or comparable performance on different IR tasks and test sets. In the DND dataset, we still achieve a competitive denoising result, while AINDNet and COLA-Net achieve the higher performance because of employing extra training data. AINDNet is specially designed for real image denoising and trained with extra data (more data besides the SIDD training set) for better performance. COLA-Net also employs extra training data during training. While we still employ only the SIDD training set as most methods did. However, when COLA-Net and AINDNet are trained with the same dataset as NLENet, NLENet still achieves the highest PSNR and SSIM in all the noise level of synthetic image denoising datasets. The visual comparison of the results are shown in FIGURE 6 and FIGURE 7 in which we can see that the NLENet recovers cleaner outlines and preserves more textural details than other competitors' approaches.

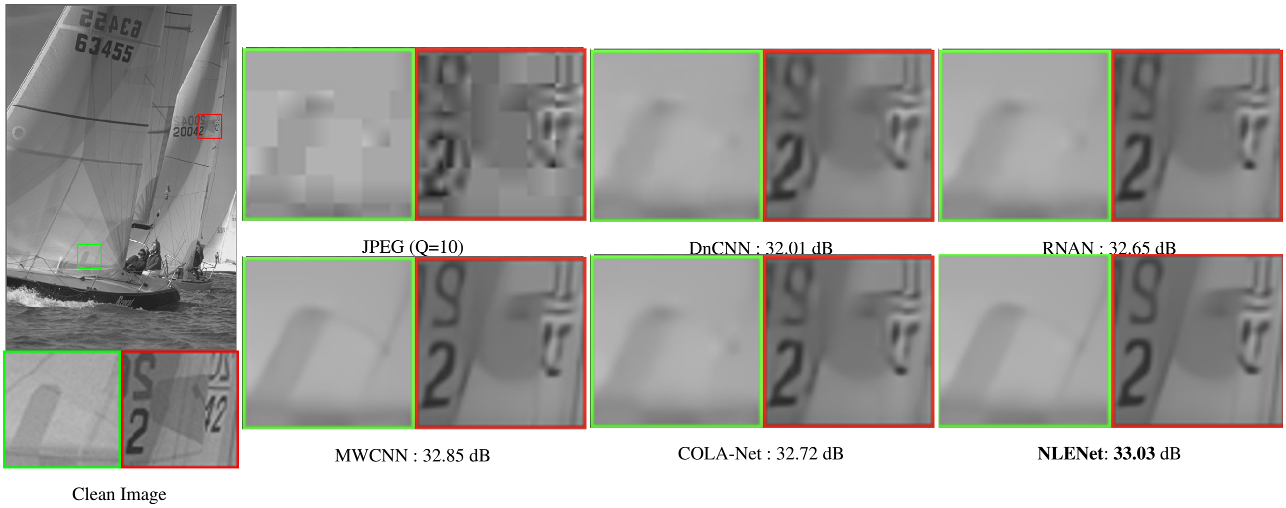

JPEG (Q=10)
DnCNN : 32.01 dB
RNAN : 32.65 dB
MWCNN : 32.85 dB
COLA-Net : 32.72 dB
**NLENet**: **33.03** dB
Clean Image

FIGURE 8. JPEG artifacts removal (Q=10) results ("*sailing3*" from *LIVE1* test set) using different JPEG artifacts removal networks.The best result (PSNR in dB) is highlighted in bold.

TABLE 4. Average PSNR (dB) and SSIM [57] results of different methods for real-world super resolution, tested on *RealSR* [79] dataset with scale factors $scale = \{2, 3, 4\}$. The best results are highlighted in red, respectively. "NA" means "Not Available" due to unavailable code or model.

| Training Dataset | *Scale* | Bicubic | VDSR [18] | SRResNet [80] | RCAN [23] | LP-KPN [79] | DDet [81] | CDC [82] | NLENet |
|---|---|---|---|---|---|---|---|---|---|
| *RealSR* | ×2 | 32.61 | 33.64 | 33.69 | 33.87 | 33.90 | 33.22 | 33.96 | **34.15** |
| | | 0.907 | 0.917 | 0.919 | 0.922 | 0.927 | NA | 0.925 | **0.930** |
| | ×3 | 29.34 | 30.14 | 30.18 | 30.40 | 30.42 | 30.62 | 30.99 | **31.05** |
| | | 0.841 | 0.856 | 0.859 | 0.862 | 0.868 | NA | 0.869 | **0.870** |
| | ×4 | 27.99 | 28.63 | 28.67 | 28.88 | 28.92 | 28.94 | 29.24 | **29.43** |
| | | 0.806 | 0.821 | 0.824 | 0.826 | 0.834 | NA | 0.827 | **0.839** |

### C. JPEG ARTIFACTS REMOVAL

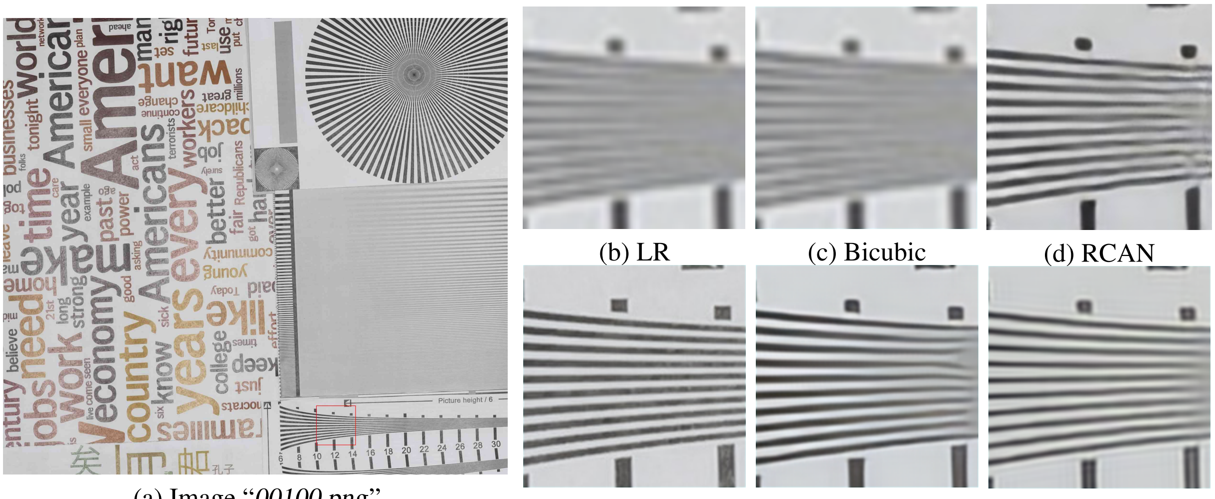
In this section, we evaluate our NLENet in JPEG artifacts removal task. We train the models with DIV2K [66] training set and test on classic JPEG artifacts removal test sets CLASSIC5 and LIVE1 [73].

We compare the NLENet with SOTA JPEG artifacts removal approaches in terms of PSNR and SSIM [57]. The results are shown in TABLE 3, in which we can see that our proposed method demonstrates the best results on all test sets and quality factors over previous approaches in PSNR. For instance, compared with the latest non-local network COLA-Net, NLENet achieves superior performance on both test sets. We can also observe that although NLENet achieves the highest PSNR, QGAC shows a slightly higher SSIM. Because NLENet is trained with L2 loss with one stage. While QGAC requires two-stage training, in which it requires L2 loss for initial training and then employs GAN loss to finetune the model. Training with GAN loss, which contains perception terms, will improve the SSIM results and better visual quality but decrease in terms of PSNR.

The visual quality comparison is shown in FIGURE 8, in which we can observe that the results from DnCNN and RNAN show an over smoothing in preserving texture details. In the results of MWCNN and the latest non-local networks, COLA-Net, a blurred outline is preserved in the restored image. In contrast, NLENet can preserve more subtle texture details and clear edges in the restored image, which further demonstrates its superiority.

### D. REAL IMAGE SUPER RESOLUTION

We apply the proposed NLENet to the real image super resolution task and compare with the SOTA SR algorithms (VDSR [18], SRResNet [80], RCAN [23], LP-KPN [79]) and CDC [82] on the RealSR test dataset with upscaling factors of ×2, ×3 and ×4. Note that all the comparing algorithms are trained on the training set of RealSR [79] (the comparing results are also provided from RealSR [79] and CDC [82]). RealSR [79] is a real image super resolution dataset, which contains LR-HR real-world image pairs captured by adjusting the cameras' focal length. RealSR has 183 (×2), 234 (×3), 178 (×4) very high resolution image pairs for training and 30 images pairs for testing in each scales. In the experiment, we compute the PSNR and SSIM [57] using the Y channel (in YCbCr color space), which is a common practice in SR [18], [23], [80]. The results are summarized in Table 4, and we can observe that NLENet shows a superior performance among the competitive methods. The PSNR improvement is around 0.06 to 0.19 dB, compared with the latest SOTA methods and 0.4∼0.5 dB compared with the classic methods like RCAN and LP-KPN. The visual comparison in FIGURE 9 proves more advantages of NLENet on restoring clear structural details, while the comparing



(a) Image "*00100.png*"

(b) LR (c) Bicubic (d) RCAN

(e) HR (f) LP-KPN (g) NLENet

**FIGURE 9.** Super resolution results of a typical image ("*00100.png*" from *RealSR* [79] test set) using different super resolution methods.

**TABLE 5.** Ablation study results on the proposed modules in NLENet for synthetic image denoising task.

| Index | self non-local | batch non-local | EBNA | NLIB | PSNR(dB)/SSIM |
|---|---|---|---|---|---|
| case1 | × | × | × | × | 30.15/0.8556/ |
| case2 | ✓ | × | × | × | 30.67/0.8679 |
| case3 | × | ✓ | × | × | 30.72/0.8717 |
| case4 | × | ✓ | ✓ | × | 30.82/0.8729 |
| case5 | × | ✓ | ✓ | ✓ | 30.99/0.8773 |
| case6 | ✓ | × | ✓ | ✓ | 30.85/0.8725 |
| case7 | × | ✓ | × | ✓ | 30.92/0.8739 |

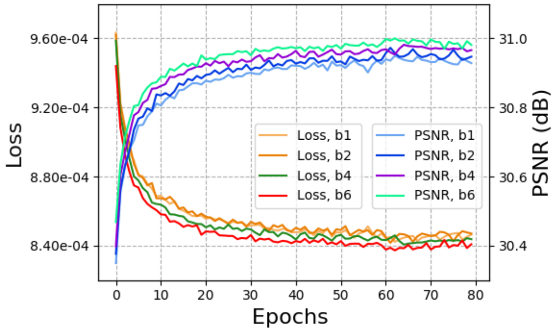


**FIGURE 10.** Ablation on batch-wise non-local module.

methods restore blurry edges.

## V. ABLATION STUDY

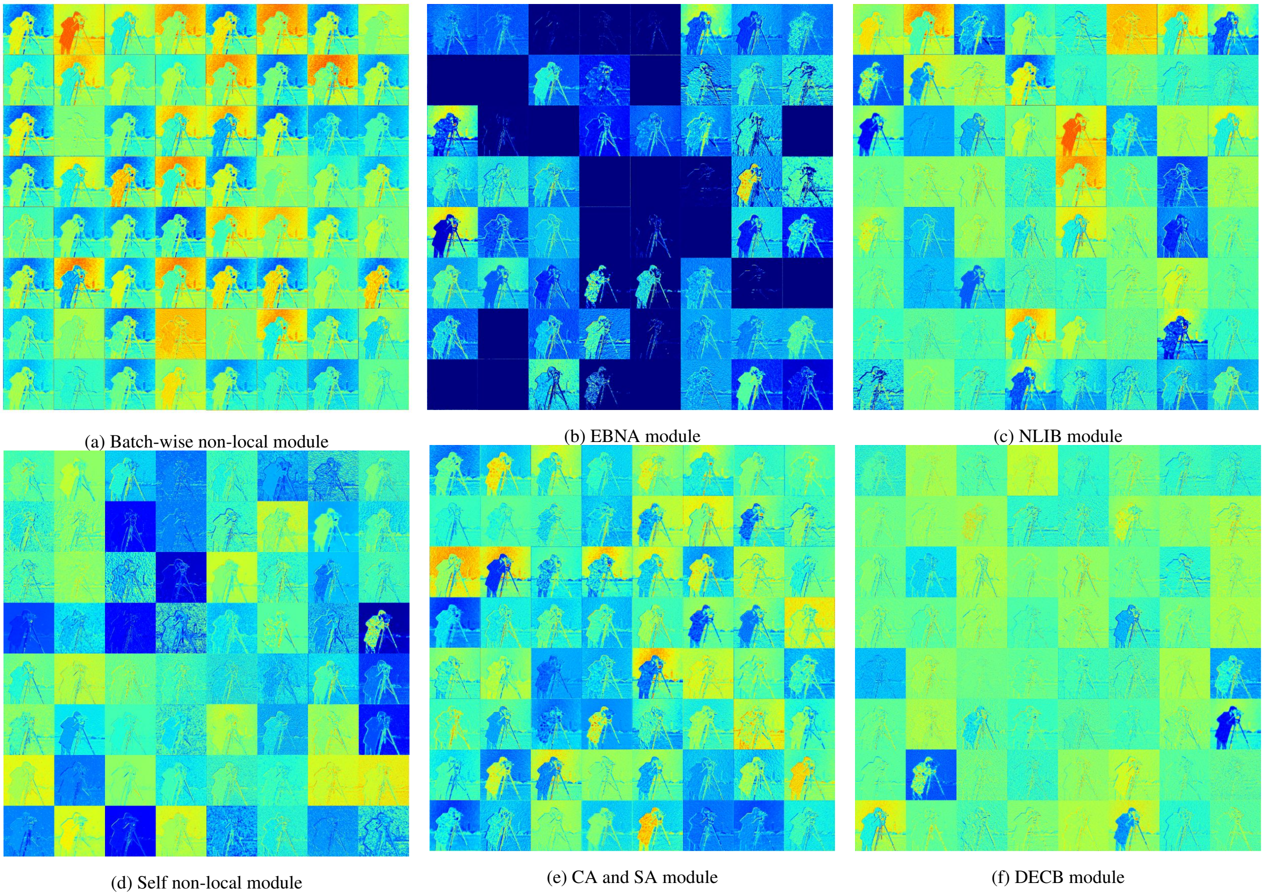
In this section, we further explore and investigate the effectiveness of the proposed NLENet. Here we study the impact and effectiveness of each proposed component on the final model performance. The ablation experiments are performed for the grayscale image synthetic denoising task with noise level $\sigma = 25$.

### A. ABLATION ON THE PROPOSED MODULES

We apply different combinations of the proposed modules to test their effectiveness in the proposed NLENet. TABLE 5 shows the comparison results test on Set12. In the experiment, the performance of a baseline multi-scale architecture (mostly based on stacked convolution layers) without using any proposed components are shown in case1.

In case2 and case3, we apply only the non-local module proposed by COLA-Net [31] (self non-local) and our proposed batch-wise non-local module respectively in the multi-scale structure to compare their influence on the IR performance. Case2 shows good performance gain, demonstrating the effectiveness of building long-range dependencies. The comparison of case2 and case3 demonstrates the superior performance of our proposed batch-wise non-local module compared to the self non-local module.

Subsequently, from case3 to case5, we add the proposed components gradually to explore the performance changes. In case4, we apply the proposed EBNA in the multi-scale model. We can observe that combining various non-local modules achieves better performance than using only one type of non-local module (compare to case3). Case5 is our proposed NLENet, which achieves the highest performance. To further compare the self non-local and the proposed batch-wise non-local module, we replace the batch-wise non-local



(a) Batch-wise non-local module

(b) EBNA module

(c) NLIB module

(d) Self non-local module

(e) CA and SA module

(f) DECB module

**FIGURE 11.** Feature visualization of different proposed modules in NLENet.

module in NLENet with the self non-local module [31] in case6. In case7, we remove the CA, SA module in EBNA and remain the other part in NLENet to explore the impact of combining CA, SA with the batch-wise non-local module. We can observe that without the CA and SA modules, NLENet still achieves superior performance.

Based on the results from the TABLE 5, we can summarize the following observations:

(1) Our proposed batch-wise non-local module can achieve better performance than the existing self non-local module (used in COLA-Net), as shown in case2 and case3.

(2) Adding the proposed EBNA and NLIB modules in the model can improve the performance, as shown in case4 and case5. Also, case5, as our proposed NLENet, achieves the best performance among all the cases.

(3) We replace the batch-wise non-local in NLENet with self non-local in case6, which further illustrates the superiority of our proposed batch-wise non-local module against COLA-Net [31] 's self non-local module.

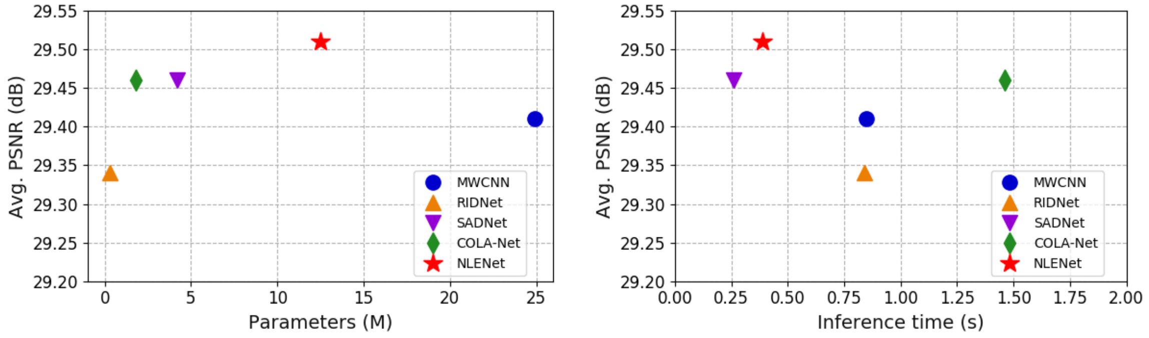
(4) Removing the CA and SA modules in EBNA, our NLENet still achieves competitive performance in case7.

### B. ABLATION ON BATCH-WISE NON-LOCAL MODULE

To further explore the impact of the proposed batch-wise non-local module, we train several models with a different number of images $b = 1, 2, 4, 6$ which are employed in the batch-wise non-local module. The results of loss and PSNR test on Set12 are shown in FIGURE 10. We can observe that larger $b$ accelerates the convergence of the model and achieves better performance. We can also find that when the $b$ increases the performance gain is getting moderate. However, larger $b$ also leads to more memory consumption in the patch matching process during training. In this case we set the $b = 6$ because the memory consumption reaches the limit capacity of the GPU we use. So we choose the largest $b$ we can to improve the performance.

### C. FEATURE VISUALIZATION OF THE PROPOSED MODULES

To further explore the proposed modules' influence, we visualize their feature map output in Figure 11. We can observe that the feature map extracted by our batch-wise non-local module (in (a)) is more clear and abundant compared to self non-local module (in (d)). We can also find that compared to the batch-wise non-local module, the CA and SA modules extract non-local features with a clear focus area



(a)

(b)

**FIGURE 12.** Parameters and inference time study on image of size $480 \times 320$ (from BSD68 test set).

in the feature map. Thus, we collaborate these modules to build EBNA (in (e)), which generates more sparsed non-local features. To further preserve more delicate local details in the restored images, we add a deformable convolution block (DECB) to cooperate with EBNA. We can see from (f) that the DECB extract more sophisticated local features. With the EBNA and DECB, NLIB can combine non-local and local features to generate more enriched features, as shown in (c). As mentioned in reference [83] that feature map with lower channel weight contains more noise-like information. Therefore NLIB can preserve more texture and structural information, since it generates higher channel weight feature maps (in (c)) after combining the non-local and local features.

### D. PARAMETERS AND INFERENCE TIME STUDY

FIGURE 12 shows the model parameter size and the GPU run time of the competing methods on synthetic denoising task. The Nvidia cuDNN-v7.0 deep learning library is adopted under Ubuntu 16.04 system.

From (a) in FIGURE 12, we can observe that despite the superior performance of NLENet, its parameter size is larger compared to RIDNet, SADNet, and COLA-Net. Because in NLENet, the proposed non-local module includes several extra convolution layers and attention modules.

While from (b) in FIGURE 12, the run-time evaluation demonstrates that our proposed model still achieves a competitive speed with an outstanding performance. Especially compared with COLA-Net, our NLENet can achieve significant speed improvement. We achieve such effectiveness because NLENet employs a multi-scale structure, saving time at a low-resolution scale. Furthermore, in COLA-Net, overlapping non-local patches are extracted and will cost more time in the non-local process. NLENet has a longer inference time than SADNet because NLENet has a more sophisticated structure. Besides the deformable convolution modules that explore adaptive local information, the EBNA module employs various non-local modules to generate enriched non-local features, which takes more time. Therefore, our multi non-local enhanced module can achieve better performance while keeping a competitive inference speed at the same time.

## VI. CONCLUSION

In this paper, we propose a batch-wise non-local module to explore long-range dependencies. We further build an EBNA module based on our proposed batch-wise non-local module, in which various non-local modules are combined to extract more enriched non-local features. Besides EBNA, we build a novel block named NLIB, which collaborates various non-local features with adaptive local features to acquire the capability of preserving fine contextual details. Finally, we embed the NLIB in a U-net-like structure named as NLENet. Extensive experiments show that NLENet consistently achieves state-of-the-art performance for several image restoration tasks, such as synthetic image denoising, real image denoising, JPEG artifacts removal and real image super resolution.

## REFERENCES

[1] L. Gondara, “Medical image denoising using convolutional denoising autoencoders,” in *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2016, pp. 241–246.

[2] L. Wu, M. Xu, L. Sang, T. Yao, and T. Mei, “Noise augmented double-stream graph convolutional networks for image captioning,” *IEEE Transactions on Circuits and Systems for Video Technology*, 2020.

[3] H. Jiang, G. Zhai, H. Cai, and J. Yang, “Scalable motion analysis based surveillance video de-noising,” in *2018 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2018, pp. 1–6.

[4] L. Zhang, S. Vaddadi, H. Jin, and S. K. Nayar, “Multiple view image denoising,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 1542–1549.

[5] S. Hao, X. Han, Y. Guo, X. Xu, and M. Wang, “Low-light image enhancement with semi-decoupled decomposition,” *IEEE Transactions on Multimedia*, vol. 22, no. 12, pp. 3025–3038, 2020.

[6] Z. Huang, Y. Zhang, Q. Li, T. Zhang, N. Sang, and H. Hong, “Progressive dual-domain filter for enhancing and denoising optical remote-sensing images,” *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 759–763, 2018.

[7] J. Zhang, Y. Liu, S. Zhang, R. Poppe, and M. Wang, “Light field saliency detection with deep convolutional networks,” *IEEE Transactions on Image Processing*, vol. 29, pp. 4421–4434, 2020.

[8] G. Li, Y. Yang, X. Qu, D. Cao, and K. Li, “A deep learning based image enhancement approach for autonomous driving at night,” *Knowledge-Based Systems*, p. 106617, 2020.

[9] Z. Chen, X. Hou, L. Shao, C. Gong, X. Qian, Y. Huang, and S. Wang, “Compressive sensing multi-layer residual coefficients for image coding,”

*IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, pp. 1109–1120, 2020.

[10] P. Zhuang, Y. Huang, D. Zeng, and X. Ding, "Mixed noise removal based on a novel non-parametric bayesian sparse outlier model," *Neurocomputing*, vol. 174, pp. 858–865, 2016.

[11] Y. Wei, Z. Zhang, Y. Wang, M. Xu, Y. Yang, S. Yan, and M. Wang, "Deraincyclegan: Rain attentive cyclegan for single image deraining and rainmaking," *IEEE Transactions on Image Processing*, vol. 30, pp. 4788–4801, 2021.

[12] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[13] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2005, pp. 60–65.

[14] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[15] C. Dong, C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *European conference on computer vision*. Springer, 2014, pp. 184–199.

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

[17] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, no. 99, pp. 1–1, 2020.

[18] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1646–1654.

[19] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 136–144.

[20] P. Zhuang and X. Ding, "Divide-and-conquer framework for image restoration and enhancement," *Engineering Applications of Artificial Intelligence*, vol. 85, no. Oct., pp. 830–844, 2019.

[21] X. Fu, Z. J. Zha, F. Wu, X. Ding, and J. Paisley, "Jpeg artifacts reduction via deep convolutional sparse coding," in *IEEE International Conference on Computer Vision*, 2019, pp. 1230–1239.

[22] P. Liu, H. Zhang, K. Zhang, L. Lin, and W. Zuo, "Multi-level wavelet-cnn for image restoration," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 773–782.

[23] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 286–301.

[24] Z. Chen, W. Guo, Y. Feng, Y. Li, C. Zhao, Y. Ren, and L. Shao, "Deep-learned regularization and proximal operator for image compressive sensing," *IEEE Transactions on Image Processing*, vol. 30, pp. 7112 – 7126, 2021.

[25] S. Anwar and N. Barnes, "Real image denoising with feature attention," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 3155–3164.

[26] X. Wang, L. Zhu, Y. Wu, and Y. Yang, "Symbiotic attention for egocentric action recognition with object-centric alignment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[27] H. Fan, L. Zhu, Y. Yang, and F. Wu, "Recurrent attention network with reinforced generator for visual dialog," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 16, no. 3, pp. 1–16, 2020.

[28] D. Yu, J. Fu, X. Tian, and T. Mei, "Multi-source multi-level attention networks for visual question answering," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 15, no. 2s, pp. 1–20, 2019.

[29] Y. Tai, J. Yang, X. Liu, and C. Xu, "Memnet: A persistent memory network for image restoration," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 4539–4547.

[30] X. Mao, C. Shen, and Y. Yang, "Image restoration using convolutional auto-encoders with symmetric skip connections," in *Advances in Neural Information Processing Systems*, 2016, p. 2810–2818.

[31] C. Mou, J. Zhang, X. Fan, H. Liu, and R. Wang, "Cola-net: Collaborative attention network for image restoration," *IEEE Transactions on Multimedia*, pp. 60–76, 2021.

[32] D. Liu, B. Wen, Y. Fan, C. C. Loy, and T. S. Huang, "Non-local recurrent network for image restoration," in *Advances in Neural Information Processing Systems*, 2018, pp. 1673–1682.

[33] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7794–7803.

[34] Y. Li, F. Guo, R. Tan, and M. Brown, "A contrast enhancement framework with jpeg artifacts suppression," in *European conference on computer vision*. Springer, 2014, pp. 174–188.

[35] Y. Chen, W. Yu, and T. Pock, "On learning optimized reaction diffusion processes for effective image restoration," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 5261–5269.

[36] J. Kim, J. Kwon Lee, and K. Mu Lee, "Deeply-recursive convolutional network for image super-resolution," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1637–1645, 2016.

[37] A. Davy, T. Ehret, J.-M. Morel, P. Arias, and G. Facciolo, "A non-local cnn for video denoising," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 2409–2413.

[38] T. Chen, H. Liu, Z. Ma, Q. Shen, X. Cao, and Y. Wang, "End-to-end learnt image compression via non-local attention optimization and improved context modeling," *IEEE Transactions on Image Processing*, vol. 30, pp. 3179–3191, 2021.

[39] P. Liu, S. Chang, X. Huang, J. Tang, and J. C. K. Cheung, "Contextualized non-local neural networks for sequence learning," in *Proceedings of the Conference on Artificial Intelligence*, vol. 33, 2019, pp. 6762–6769.

[40] R. Jia, Y. Cao, H. Tang, F. Fang, C. Cao, and S. Wang, "Neural extractive summarization with hierarchical attentive heterogeneous graph network," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020, pp. 3622–3631.

[41] Z. Tu, Y. Ma, C. Li, J. Tang, and B. Luo, "Edge-guided non-local fully convolutional network for salient object detection," *IEEE Transactions on Circuits and Systems for Video Technology*, 2020.

[42] Z. Wang, N. Zou, D. Shen, and S. Ji, "Non-local u-nets for biomedical image segmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, 2020, pp. 6315–6322.

[43] Z. Zhu, M. Xu, S. Bai, T. Huang, and X. Bai, "Asymmetric non-local neural networks for semantic segmentation," in *Proceedings of the International Conference on Computer Vision*, 2019, pp. 593–602.

[44] G. Hu, B. Cui, and S. Yu, "Skeleton-based action recognition with synchronous local and non-local spatio-temporal learning and frequency attention," in *2019 IEEE International Conference on Multimedia and Expo*. IEEE, 2019, pp. 1216–1221.

[45] Z. Chen, X. Hou, X. Qian, and C. Gong, "Efficient and robust image coding and transmission based on scrambled block compressive sensing," *IEEE Transactions on Multimedia*, vol. 20, no. 7, pp. 1610–1621, 2018.

[46] J. Yu, J. Liu, L. Bo, and T. Mei, "Memory-augmented non-local attention for video super-resolution," *arXiv preprint arXiv:2108.11048*, 2021.

[47] X. Gao, Y. Wang, J. Cheng, M. Xu, and M. Wang, "Meta-learning based relation and representation learning networks for single-image deraining," *Pattern Recognition*, vol. 120, p. 108124, 2021.

[48] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[49] Y. Shen, Q. Liu, S. Lou, and Y.-L. Hou, "Wavelet-based total variation and nonlocal similarity model for image denoising," *IEEE Signal Processing Letters*, vol. 24, no. 6, pp. 877–881, 2017.

[50] W. Liu, Q. Yan, and Y. Zhao, "Densely self-guided wavelet network for image denoising," in *CVPR Workshops*, 2020, pp. 432–433.

[51] Y. Mei, Y. Fan, and Y. Zhou, "Image super-resolution with non-local sparse attention," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 3517–3526.

[52] X. Du, S. Jiang, Y. Si, L. Xu, and C. Liu, "Mixed high-order non-local attention network for single image super-resolution," *IEEE Access*, vol. 9, pp. 49 514–49 521, 2021.

[53] X. Du, J. Niu, and C. Liu, "Expectation-maximization attention cross residual network for single image super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 888–896.

[54] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[55] S. Woo, J. Park, J. Lee, and S. K., "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[56] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Curran Associates, Inc., 2019, pp. 8024–8035.

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

[58] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2, July 2001, pp. 416–423.

[59] J. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5197–5206.

[60] M. Chang, Q. Li, H. Feng, and Z. Xu, "Spatial-adaptive network for single image denoising," in *European Conference on Computer Vision*. Springer, 2020, pp. 171–187.

[61] Y. Kim, J. W. Soh, G. Y. Park, and N. I. Cho, "Transfer learning from synthetic to real-noise denoising with adaptive instance normalization," in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3482–3492.

[62] C. Wang, C. Ren, X. He, and L. Qing, "Deep recursive network for image denoising with global non-linear smoothness constraint prior," *Neurocomputing*, vol. 426, pp. 147–161, 2021.

[63] C. Tian, Y. Xu, W. Zuo, B. Du, C.-W. Lin, and D. Zhang, "Designing and training of a dual cnn for image denoising," *Knowledge-Based Systems*, vol. 226, p. 106949, 2021.

[64] D. Valsesia, G. Fracastoro, and E. Magli, "Deep graph-convolutional image denoising," *IEEE Transactions on Image Processing*, vol. 29, pp. 8226–8237, 2020.

[65] C. Mou, J. Zhang, and Z. Wu, "Dynamic attentive graph learning for image restoration," in *Proceedings of the International Conference on Computer Vision*, 2021.

[66] E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," in *The IEEE Conference on Computer Vision and Pattern Recognition Workshops*, July 2017, pp. 1230–1239.

[67] A. Abdelhamed, S. Lin, and M. Brown, "A high-quality denoising dataset for smartphone cameras," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2018, pp. 1230–1239.

[68] T. Plotz and S. Roth, "Benchmarking denoising algorithms with real photographs," in *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2750–2759.

[69] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Color image denoising via sparse 3D collaborative filtering with grouping constraint in luminance-chrominance space," in *IEEE International Conference on Image Processing*. IEEE, 2007, pp. 313–316.

[70] H. C. Burger, C. J. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with BM3D?" in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2392–2399.

[71] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, "Toward convolutional blind denoising of real photographs," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1712–1722.

[72] Z. Yue, H. Yong, Q. Zhao, D. Meng, and L. Zhang, "Variational denoising network: Toward blind noise modeling and removal," in *Advances in Neural Information Processing Systems*, 2019, pp. 1688–1699.

[73] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on image processing*, vol. 15, no. 11, pp. 3440–3451, 2006.

[74] K. Yu, C. Dong, C. Loy, and X. Tang, "Deep convolution networks for compression artifacts reduction," *arXiv preprint arXiv:1608.02778*, 2016.

[75] Y. Zhang, K. Li, K. Li, B. Zhong, and Y. Fu, "Residual non-local attention networks for image restoration," in *International Conference on Learning Representations*, 2019, pp. 1230–1239.

[76] M. Ehrlich, L. Davis, S.-N. Lim, and A. Shrivastava, "Quantization guided jpeg artifact correction," in *Proceedings of the European Conference on Computer Vision*. Springer, 2020, pp. 293–309.

[77] C. Ren, Q. Teng, X. He, L. Qing, and T. Q. Nguyen, "Compressed image restoration via deep deblocker driven unified framework," *Knowledge-Based Systems*, vol. 228, p. 107268, 2021.

[78] X. Fu, M. Wang, X. Cao, X. Ding, and Z.-J. Zha, "A model-driven deep unfolding method for jpeg artifacts removal," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2021.

[79] J. Cai, H. Zeng, H. Yong, Z. Cao, and L. Zhang, "Toward real-world single image super-resolution: A new benchmark and a new model," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 3086–3095.

[80] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, and Z. Wang, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4681–4690.

[81] Y. Shi, H. Zhong, Z. Yang, X. Yang, and L. Lin, "Ddet: Dual-path dynamic enhancement network for real-world image super-resolution," *IEEE Signal Processing Letters*, vol. PP, pp. 1–1, 2020.

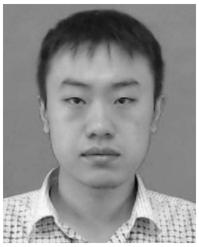
[82] P. Wei, Z. Xie, H. Lu, Z. Zhan, Q. Ye, W. Zuo, and L. Lin, "Component divide-and-conquer for real-world image super-resolution," in *Proceedings of the European conference on computer vision*, 2020, pp. 101–117.

[83] Y. Zhang, K. Li, K. Li, G. Sun, Y. Kong, and Y. Fu, "Accurate and fast image denoising via attention guided scaling," *IEEE Transactions on Image Processing*, vol. 30, pp. 6255–6265, 2021.

YUAN HUANG is currently pursuing the Ph.D. degree with Xi'an Jiaotong University under the supervision of Prof. X. Hou. Her research interests include image denoising, image super-resolution, and image retrieval.

XINGSONG HOU received the Ph.D. degree from Xi'an Jiaotong University, Xi'an, China, in 2005. From 2010 to 2011, he was a Visiting Scholar with Columbia University, New York, USA. He is currently a Professor with the School of Electronics and Information Engineering, Xi'an Jiaotong University. His research interests include video/image coding, wavelet analysis, sparse representation, compressive sensing, and radar signal processing.

YUJIE DUN received the Ph.D. degree in information and communication engineering from Xi'an Jiaotong University (XJTU), Xi'an, China, in 2016. After the Ph.D. degree, she visited as a visiting scholar at Washington University in St. Louis, U.S., in 2017/2018, and a Post-Doctoral Researcher at Washington University in St. Louis, U.S., in 2018/2019. Currently, she works as an Associate Professor in School of Information and Communication, Xi'an Jiaotong University. Her research interests include audio/speech signal processing and coding, statistical signal processing and modeling, biomedical signal processing, and machine learning.



ZAN CHEN received the B.S. degree and P.h.D from Xi'an Jiaotong University in 2012 and 2019, respectively. He was a visiting scholar at the University of East Anglia (UEA), Norwich, UK, in 2018. Now, he is an associate professor with the College of Information Engineering, Zhejiang University of Technology. His research interests include compressive sensing, computer vision, and medical image processing.

XUEMING QIAN received the B.S. and M.S. degrees from the Xi'an University of Technology, Xi'an, China, in 1999 and 2004, respectively, and the Ph.D. degree in electronics and information engineering from Xi'an Jiaotong University, Xi'an, in 2008. He was a Visiting Scholar with Microsoft Research Asia, Beijing, China, from 2010 to 2011. He was an Assistant Professor with Xi'an Jiaotong University, where he was an Associate Professor from 2011 to 2014, and is currently a Full Professor. He is also the Director of the Smiles Laboratory, Xi'an Jiaotong University. His research interests include social media big data mining and search.

• • •