Attention-based Pyramid Dilated Lattice Network for Blind Image Denoising

Mohammad Nikzad, **Yongsheng Gao** and **Jun Zhou**
Institute for Integrated and Intelligent Systems (IIIS), Griffith University, Australia
m.nikzaddehaji@griffithuni.edu.au, yongsheng.gao, jun.zhou@griffith.edu.au

Abstract

Though convolutional neural networks (CNNs) with residual and dense aggregations have obtained much attention in image denoising, they are incapable of exploiting different levels of contextual information at every convolutional unit in order to infer different levels of noise components with a single model. In this paper, to overcome this shortcoming we present a novel attention-based pyramid dilated lattice (APDL) architecture and investigate its capability for blind image denoising. The proposed framework can effectively harness the advantages of residual and dense aggregations to achieve a great trade-off between performance, parameter efficiency, and test time. It also employs a novel pyramid dilated convolution strategy to effectively capture contextual information corresponding to different noise levels through the training of a single model. Our extensive experimental investigation verifies the effectiveness and efficiency of the APDL architecture for image denoising as well as JPEG artifacts suppression tasks.

1 Introduction

Image denoising is an active topic in image processing and computer vision. In fact the performances of many computer vision systems are highly dependent on the quality of their input images [Chatterjee and Milanfar, 2009]. Thus, restoring the clean image from a given noisy observation using image denoising techniques is essential. The goal of image denoising is to attain a clean image x from a noisy image x_n which is formulated by $\mathrm{x}_n = \mathrm{x} + \mathrm{v}$. As the most common assumption in the literature [Dabov *et al.*, 2007; Zhang *et al.*, 2017; Zhang *et al.*, 2018], which we also follow in this paper, the noise signal, v, is additive white Gaussian noise (AWGN) with zero mean and standard deviation σ.

Previously, modeling image priors was a prominent approach for image denoising for which methods such as sparsity based models BM3D [Dabov *et al.*, 2007] and WNMM [Gu *et al.*, 2014] have been developed. Although these approaches are effective for image denoising, they suffer from two main shortcomings [Zhang *et al.*, 2017; Zhang *et al.*, 2018; Peng *et al.*, 2019]. First, they are nonconvex models and their parameters need to be manually chosen. Second, most of the prior-based models involve a complex optimization problem during testing, resulting in poor computational efficiency.

On the other hand, convolutional neural networks (CNNs) have shown outstanding performance on many computer vision and image processing tasks [Gu *et al.*, 2018]. CNNs effectively address the limitation of prior-model based denoising approaches. Such learning-based frameworks like denoising CNN (DnCNN) [Zhang *et al.*, 2017], and fast and flexible denoising CNN (FFDNet) [Zhang *et al.*, 2018] represent a significant leap in denoising performance over previous approaches.

Residual (ResNets) [He *et al.*, 2016] and dense (DenseNets) [Huang *et al.*, 2017] aggregations of layer outputs have presented further noticeable progress in image restoration domain. Residual links facilitate passing information and gradients efficiently through the network. Dense aggregations offer direct feature re-usage, as deeper layers have access to the outputs of shallower layers. Many recent image denoising methods such as RED [Mao *et al.*, 2016], pyramid attention networks (PANET) [Mei *et al.*, 2020], residual non-local attention networks [Zhang *et al.*, 2019b], and residual dense network (RDN) [Zhang *et al.*, 2020] tried to take the advantages of residual and/or dense aggregations techniques and obtained the state-of-the-art image restoration performance.

Despite the success of both residual and dense skip connections, both aggregation types have drawbacks. As shown in [Nikzad *et al.*, 2020], these techniques may cause information loss and/or parameter over allocating as the networks get deeper or wider. Networks like RDNs [Zhang *et al.*, 2020] combine the benefits of both aggregation types to extract hierarchical features from all the layers for image restoration tasks. However, RDNs follow a similar dense aggregation strategy to DenseNets and possess the same drawback inherent from DenseNets. Furthermore, most of the existing discriminative learning based methods [Zhang *et al.*, 2020; Zhang *et al.*, 2019b; Plötz and Roth, 2018] are limited and tailored for some specific given noise levels while in a practical situation, the noise level is mostly unknown (blind image denoising). In addition, they are not capable of effectively handling a wide range of noise levels with training of a sin-

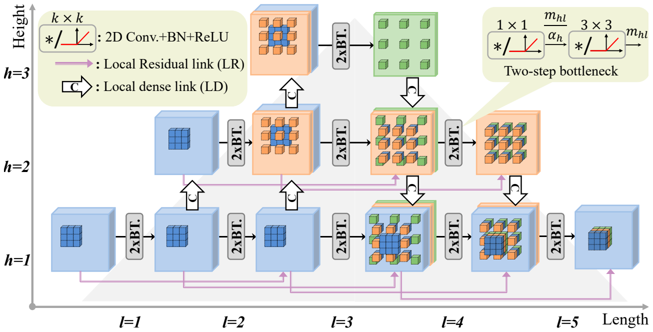


Figure 1: The proposed pyramid dilated lattice (PDL) topology. Voxels with different colors demonstrate different levels of contextual information extracted by multiple receptive fields.

gle denoising model. In few cases, a single CNN model (*e.g.* DnCNN-B) is trained for blind denoising while it's performance is restricted to the preset noise range in training phase (*e.g.* [0, 50]). One reason might be that convolutional units within these frameworks are not capable of jointly exploiting different levels of contextual information to simultaneously infer different levels of noise components with a single model.

In this paper, we introduce a new structure to fully exploit the potential of lattice topology for addressing the aforementioned shortcomings of CNN based image denosing methods. To this end, first we adopt a unique pyramid dilated convolution (PDC) strategy to effectively widen the receptive field. As shown in Fig. 1, multiple feature maps which carry contextual information from different receptive fields are reused by local dense aggregations through the height of the lattice. This property improves blind image denoising ability effectively as the network can exploit diverse context to predict image details. Further an attention mechanism is employed to fully exploit information carried by inter-dependencies among feature channels. We refer to our deep network as *attention-based pyramid dilated lattice* network (APDL-Net) which consists of several stacked APDL blocks.

Our extensive experiments show that the new APDL-Net is able to produce higher blind image denoising performance than other benchmark image denosing networks without sacrificing computational efficiency in terms of speed and model size. To verify the generalization capability of the proposed APDL-Net, we also demonstrate its effectiveness in a JPEG compression artifact reduction task.

2 Related Works

In the past few years, significant progress has been made for image denoising by developing discriminative learning-based methods using deep convolutional neural networks (CNNs). These methods aim at learning a nonlinear mapping from a noisy image to its corresponding clear image. Benefited from CNNs, Mao *et al.* [Mao *et al.*, 2016] proposed an encoder-decoder CNN framework to suppress the noise and to restore the high resolution image. A trainable nonlinear reaction diffusion (TNRD) method [Chen and Pock, 2016] used a flexible convolutional framework to learn the image prior. With the aid of batch normalization, rectified linear unit (ReLU) and residual mapping (RL) techniques, DnCNN [Zhang *et al.*, 2017] achieved the state-of-the-art denoising performance. FFDNet [Zhang *et al.*, 2018] introduced a CNN model which utilizes tunable noise level map as input and downsampled sub-images to improve speed and process of the blind denosing. A universal denoising network (UNLNet) [Lefkimmiatis, 2018] used a CNN network which is trained based on local and non-local variational models and a constrained optimization method.

As more recent and advanced denoising methods, dilated dense fusion networks (DDFN) [Chen *et al.*, 2019] adopted a deep topology which takes the advantages of densely connected links and dilation technique. RDN [Zhang *et al.*, 2020] employed both residual and dense aggregations to fully exploit the local and global hierarchical features from all convolutional layers. Neural nearest neighbour networks (N^3Nets) [Plötz and Roth, 2018] employed a non-local processing network to exploit self-similarity of the image's patches. Likewise, non-local recurrent network (NLRN) [Liu *et al.*, 2018] attempted to incorporate non-local techniques into a recurrent neural network (RNN) for image restoration. Methods like residual non-local attention network (RNAN) [Zhang *et al.*, 2019b], pyramid attention networks (PANet) [Mei *et al.*, 2020] focused on adopting attention mechanism to learn the inter-channel relationships of the convolutional feature maps.

Although current state-of-the-art CNN-based image restoration algorithms improve denoising performance, they cannot present a good trade-off between performance, model size and running time which is a necessity for a practical denoising method. For example, N^3Nets [Plötz and Roth, 2018] and NLRN [Liu *et al.*, 2018] can obtain very good performances with small models while experienced a very slow inferring phase as they rely on conversion from the image domain to the patch domain and vice versa. RDN [Zhang *et al.*, 2020] and RNAN [Zhang *et al.*, 2019b] suffer from over-allocating parameters and relatively slow running time. In this paper, we show that the proposed APDL-Net is capable of addressing the aforementioned limitations effectively.

3 Attention-based Pyramid Dilated Lattice

3.1 Network Structure

The proposed APDL-Net is shown in Fig. 2. The network consists of B APDL blocks, each being constructed by a PDL block, a global residual (GR) aggregation, and an attention (Att.) unit. To extract shallow features, a composite function consisting of a 2D convolution operation followed by batch normalization (BN) and ReLU activation is adopted at the first layer of the proposed APDL-Net. We also use a simple 2D convolution unit as the last layer of the network to reconstruct residual map. More details about APDL blocks, attention mechanism, and global residual are given in Subsection 3.2.

Since the residual mapping technique is used $\mathcal{R}(\mathrm{x}_n) \approx -\mathrm{v}$, the clean image is found as: $\mathrm{x} = \mathrm{x}_n + \mathcal{R}(\mathrm{x}_\mathrm{n})$. The resid-

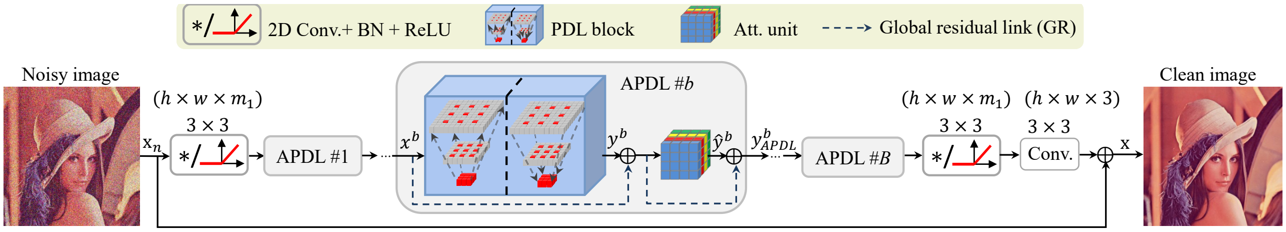


Figure 2: Overview of the proposed APDL-Net for image denoising. $\bigoplus$ indicates element-wise sum operator.

ual mapping, $\mathcal{R}(\mathrm{x}_n)$, can be obtained by:

$$\mathcal{R}(\mathrm{x}_n) = H_{APDL-Net}(\mathrm{x}_n) \tag{1}$$

where $H_{APDL-Net}$ denotes the function of the proposed APDL-Net. We also adopt the following loss function to optimize the model:

$$\mathcal{L}(\Theta) = \frac{1}{2M}\sum_{i=1}^{M} \|H_{APDL-Net}(\mathrm{x^i}_n, \Theta) - (\mathrm{x^i} - \mathrm{x^i}_n)\|_2^2 \tag{2}$$

where $\{(\mathrm{x^i}_n, \mathrm{x^i})\}_{i=1}^{M}$ and Θ indicate M noisy-clean patch pairs and trainable parameters in APDL-Net respectively.

3.2 Attention Pyramid Dilated Lattice Block

Our proposed APDL block consists of three main parts: pyramid dilated lattice (PDL), global residual (GR) aggregations, and attention mechanism. They are explained in detail as follows.

Pyramid dilated lattice. PDL is a triangular lattice of convolutional units. The location of each convolutional unit, C_{hl}, within the lattice, is specified by height and length coordinates (h, l):

$$C_{hl} = \begin{cases} C_{hl}^{\triangleleft}, & \text{if } h \le l,\ l \le H, \\ C_{hl}^{\triangleright}, & \text{if } h \le 2H - l,\ H < l \le L, \\ \emptyset, & \text{otherwise}, \end{cases} \tag{3}$$

where $h = 1, 2, ..., H$, and, $l = 1, 2, ..., L$. $C_{hl}^{\triangleleft}$ and $C_{hl}^{\triangleright}$ are convolutional units that exist in the left and right triangles of the lattice, respectively. $\emptyset$ indicates that the convolutional unit does not exist. The number of convolutional units in each block is denoted by N, where N is a square number and $N \ge 4$. The height of the lattice is $H = \sqrt{N}$, and the length is $L = 2\sqrt{N} - 1$. Local dense aggregations of convolutional unit outputs are formed strictly over the *height* of the lattice. While PDL block does not allow for total feature reusage, densely aggregating only a subset of previous outputs has been shown to be beneficial. Local residual links are also adopted, to improve intra block gradient flow

Pyramid Dilated Convolution

Capturing more contextual information by widening the receptive fields is desired for enhancing the performance of the CNN. Using a larger kernel size and stacking more convolutional layers are two common ways to expand the receptive field. However, these techniques increase the number of parameters and can cause problems such as overfitting. Here, we propose a new lattice topology by leveraging a simple and effective pyramid dilated convolution schema. The idea behind the new lattice framework is illustrated in Fig. 1. Specifically, the receptive field of a convolutional unit in the lattice block is controlled via the dilatation rate, $d = h$. Unlike other commonly used dilated convolution techniques like DSNets [Peng *et al.*, 2019] which focus only on expanding receptive fields, here, we enlarge receptive fields alongside aggregating different levels of contextual information. This dilation schema elevates the capacity of the network to explore and to process wide range of contextual information exploited by multiple receptive fields. As can be seen from Fig. 1, the input to a convolutional unit in the left triangle of the lattice, $x_{hl}^{\triangleleft}$, is the dense aggregation of the outputs at length $l - 1$, and heights $h, h - 1, ..., 1$:

$$x_{hl}^{\triangleleft} = \begin{cases} x_{11}, & \text{if } l = h = 1 \\ y_{h(l-1)}, & \text{if } l > 1, h = 1 \text{ for } C_{hl}^{\triangleleft} \\ \left[y_{h(l-1)}, x_{(h-1)l}\right], & \text{if } l > h, h > 1 \text{ for } C_{hl}^{\triangleleft} \\ x_{(h-1)l}, & \text{if } l = h, h > 1 \text{ for } C_{hl}^{\triangleleft}, \end{cases} \tag{4}$$

where $[.]$ denotes the concatenation operation, and y_{hl} is the output of the convolution unit at (h, l). We also define function $CI^{+}(.)$ which returns a set that contains levels of contextual information aggregated by $x_{hl}^{\triangleleft}$:

$$CI^{+}(x_{hl}) = \begin{cases} \{1\}, & \text{if } l = h = 1 \\ \{d\}, & \text{if } l > 1, h = 1 \text{ for } C_{hl}^{\triangleleft} \\ CI^{+}(x_{(h-1)l}) \cup \{d\} & \text{if } l > h, h > 1 \text{ for } C_{hl}^{\triangleleft} \\ CI^{+}(x_{(h-1)l}), & \text{if } l = h, h > 1 \text{ for } C_{hl}^{\triangleleft}, \end{cases} \tag{5}$$

The local dense aggregations in the left triangle of the lattice allow for multiple concise outputs with different level of contextual information to be progressively formed. Note that level of contextual information is proportional to the dilation rate, d, which manages the receptive field of each convolution unit.

The input to a convolutional unit in the right triangle of the lattice, $x_{hl}^{\triangleright}$, is the dense aggregation of the outputs at length $l - 1$, and heights $h, h + 1, ..., H$:

$$x_{hl}^{\triangleright} = \begin{cases} \left[y_{h(l-1)}, y_{(h+1)(l-1)}\right], & \text{if } h = 2H - l \text{ for } C_{hl}^{\triangleright} \\ \left[y_{h(l-1)}, x_{(h+1)l}\right], & \text{if } h < 2H - l \text{ for } C_{hl}^{\triangleright}. \end{cases} \tag{6}$$

In the right triangle of the lattice, the outputs are progressively amalgamated into a single output. Similar to Eq. 5, we can define function $CI^{-}(.)$ which returns the set of different levels of contextual information aggregated by $x_{hl}^{\triangleright}$:

$$CI^{-}(x_{hl}) = \begin{cases} \{d+1, d\}, & \text{if } h = 2H - l \text{ for } C_{hl}^{\triangleright} \\ CI^{-}(x_{(h+1)l}) \cup \{d\}, & \text{if } h < 2H - l \text{ for } C_{hl}^{\triangleright}. \end{cases} \quad (7)$$

By densely aggregating outputs over the height of the lattice, the input size to deeper convolutional units within the block is also limited and avoids over-parameter allocating. We refer to the proposed dilated convolution schema as *pyramid dilated convolution* (PDC) as the feature maps represented by $x_{hl}^{\triangleleft}$ and $x_{hl}^{\triangleright}$ are stacked and ordered correspond to $CI^{+}(.) = \{1, 2, ..., h\}$ and $CI^{-}(.) = \{H, H-1, ..., h\}$, respectively. In particular, for the left half of the PDL the sizes of receptive fields increase from small to large range while for the second half of the lattice the trend is opposite as the height decreases.

To bring more computational efficiency, we use two-step bottleneck layers. In particular, each layer is a composite function consisting of three operations, 2D dilated convolutions followed by batch normalisation and ReLU activation. The first layer has a kernel and output size of $k = 1 \times 1$, and $\frac{m_h}{\alpha_h}$ (In this paper we set α_h as $\alpha_1 = 2$ and $\alpha_h = 1 (h \geq 2)$.), respectively. The second layer uses a kernel size of $k = 3 \times 3$ and output size of m_h (m_h indicates the convolutional unit output size at each height), as depicted in Fig. 1.

Local residual aggregations. To improve the flow of gradients over the length of the lattice, local residual links are adopted:

$$y_{hl} = \begin{cases} y_{hl} + x_{h(l-1)}, & \text{if } C_{hl}^{\triangleleft} \text{ or } C_{hl}^{\triangleright},\ l > h \\ y_{hl}, & \text{if } C_{hl}^{\triangleleft} \text{ or } C_{hl}^{\triangleright},\ l \leq h. \end{cases} \quad (8)$$

When the number of channels of y_{hl} and $x_{h(l-1)}$ are non-identical, the residual link is weighted so that $x_{h(l-1)}$ is of the same number of channels as y_{hl}.

Attention mechanism. Typically, all channel-wise features are treated equally by CNN based image denoising networks. This policy hinders the network's ability to deal with different types of information. Here, we adopt a channel-wise attention mechanism by applying squeeze and excitation operations [Hu *et al.*, 2018] at the end of every APDL block. Specifically, the squeeze operation encodes the global contextual information and yields m_1 feature descriptors of size 1×1 as:

$$z_c = \frac{1}{h \times w} \sum_{i=1}^{h} \sum_{i=1}^{w} y_c(i, j), \quad (9)$$

where m_1 denotes the input channel dimension. $z_c(c = 1, ...m_1)$ and $y_c(i, j)$ denote the c-th channel descriptor and the feature value at position (i, j) respectively. The excitation operator acts as a gating mechanism to learn nonlinear synergies between channels alongside non-mutually-exclusive relationship. Here, we adopt the gating mechanism by forming a bottleneck with two fully-connected (FC) layers. The first FC layer followed by ReLU non-linearity (δ) reduces the channel dimension with ratio r (here, we set $r = 8$) and then an up-sampling FC layer followed by sigmoid activation returns the input channel dimension (m_1). So that the output of the attention mechanism can be formalized as:

$$p_c = \alpha(U(\delta(D(z_c)))), \quad (10)$$

where D and U represent the FC layers for channel reduction and up-sampling operators respectively. Finally, rescaled input feature maps, $\hat{y_c}$, are adaptively formed by p_c as:

$$\hat{y_c} = p_c \times y_c, \quad (11)$$

Global residual aggregations. Global residual links are adopted, to further enhance the propagation of information between PDL and attention blocks. Thus the output of the APDL block, y_{APDL}^{b}, is formulated as:

$$y_{APDL}^{b} = (x^b + y^b) + \hat{y}^b, \quad (12)$$

where the superscript is added to the notation to indicate the block index, $b = 1, 2, ..., B$. x^b, y^b, and $\hat{y}^b$ denote the input, output of the PDL block b, and output of the attention unit respectively.

3.3 Implementation Details

In this work, we set the convolutional unit output size at each height to $m_h = \frac{m_1}{2^{h-1}}$, where m_1 is the output size at $h = 1$. This setting ensures that a reduced number of parameters are used for feature re-usage. The total number of convolutional units for each pyramid dilated lattice (PDL) block is set to $N = 16$ (hence, $H = 4$ and $L = 7$). We experiment APDL-Net with configurations $\{B = 3, m_1 = 64\}$ and $\{B = 6, m_1 = 64\}$ which lead to two networks as APDL-Net$_{(3,64)}$ and APDL-Net$_{(6,64)}$ with sizes of 0.785 and 1.53 million parameters, respectively.

We adopt Adam optimizer with default hyper-parameters and 10^{-5} as the weight decay for the training of the proposed models. All models are trained for 100 epochs with mini-batch size of 32 and initial learning rate of 0.001 which is decayed to 10^{-5} by adopting cosine annealing technique . All models are trained using a single NVIDIA Geforce Titan RTX GPU card.

4 Experimental Results and Discussion

4.1 Datasets

We use 400 images of size 180 × 180 from Berkeley segmentation dataset (BSD500) [Arbelaez *et al.*, 2010] as the training data for the image restoration tasks. Moreover, sub-image patches with the size of 60 × 60 are randomly cropped from the training images. For test images, three widely-used datasets for evaluation of Gaussian denoising methods, "Set12", "BSD68", and "Kodak24" [Franzen, 1999] are adopted. We train APDL-Nets for blind color image denoising (APDL-Net-B) and JPEG image artifact reduction tasks. Following [Zhang *et al.*, 2017], we adopt the color version of "BSD68" (*i.e.* "CBSD68") for evaluating the color image denoising task. Further, we use the "Classic5" [Zeyde *et al.*, 2010] and "LIVE1" [Sheikh *et al.*, 2005] datasets for testing the JPEG deblocking task as in [Zhang *et al.*, 2017].

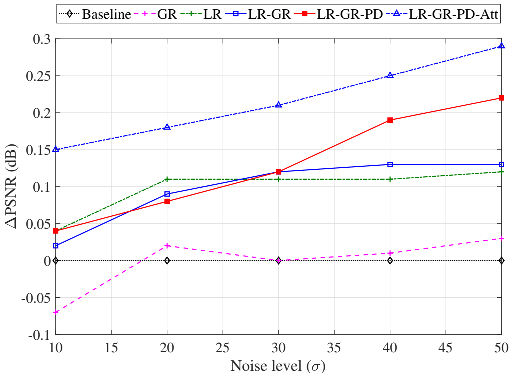


Figure 3: PSNR difference (ΔPSNR (dB)) attained by six APDL-Net$_{(3,64)}$ configuration types: Baseline, GR, LR, LR-GR, LR-GR-PD, and LR-GR-PD-Att on "Set12" dataset.

4.2 Ablation Study of APDL-Net

In this section we conduct an ablation analysis on the effects of four different concepts of the proposed APDL-Net over a wide range of noise levels: local residual links (LR), global residual links (GR), pyramid dilation strategy (PD) and attention units (Att). To this end, six APDL-Net$_{(3,64)}$ configurations are examined. The PSNR difference (ΔPSNR (dB)) of each configuration with different noise levels, σ, on dataset "Set12" is reported in Fig. 3.

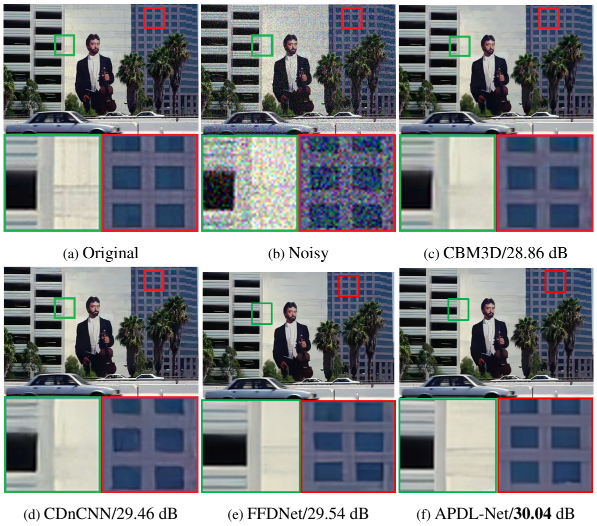
Local and global aggregations analysis. As described in Subsection 3.2, two aggregation types are used in the APDL-Net topology, including local residual links (LR) and global residual links (GR). By adding either LR or GR to the baseline (no LR and GR), it can be seen that a higher overall PSNR values can be attained compared with the baseline. However, the GR configuration obtained the lowest PSNR for noise levels (σ) below 20dB. This shows that inter block information propagation incorporate well in removing strong noise and vice versa. On the other hand, LR aggregation improves the baseline's PSNR values about +0.1dB for most levels of noise. Utilising both LR and GR could enhance overall blind denoising performance. This phenomenon demonstrates that enhanced intra block gradient flow and inter block information propagation are beneficial to the training of an APDL-Net for blind image denoising.

Pyramid dilation strategy (PD) analysis. We further verify the role of pyramid dilation strategy. As can be seen in Fig. 3, PD can elevate overall denoising performance of LR-GR. In particular, expanding receptive fields through PD is more effective in enhancing the restoration performance of the APDL-Net for stronger noise levels $\sigma \geq 30$ (Fig. 3).

Attention mechanism (Att) analysis. We also investigate the contribution of the proposed attention mechanism (Att). As demonstrated in Fig. 3, the positive impact of considering inter-dependencies among feature maps boosts the restoration performance considerably.

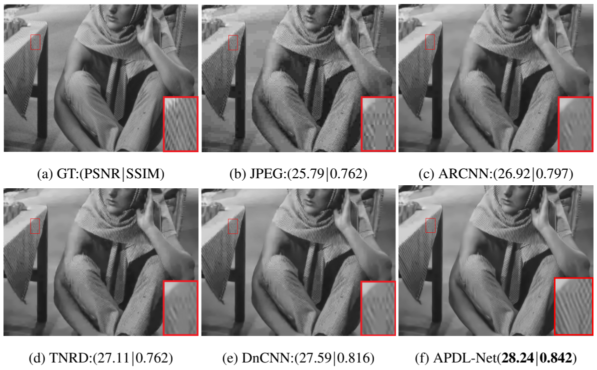
4.3 Blind Image Denoising Performance

We evaluate the proposed APDL-Net on restoring noisy images corrupted by AWGN. Table 1 reports comparisons with five state-of-the-art color blind denoising models on the "CBSD68" for a wide range of noise levels. Experimental results show that the proposed APDL-Net outperforms all of studied benchmarks by an obvious margin. This phenomenon demonstrates that APDL-Nets utilise larger receptive fields than others and have more capability of removing different levels of noise. The denoised color image produced by APDL-Net-B$_{(6,64)}$ is illustrated in Fig. 4 (f). It can be seen that the APDL-Net shows superior performance with lower level of detail distortion. Moreover we evaluate the robustness of the proposed APDL-Net-B$_{(6,64)}$ when the noise level goes beyond the level used during training. As demonstrated in Table 1, APDL-Net-B$_{(6,64)}$ is robust against higher noise levels, $\sigma \geq 55$, that is not trained for.


(a) Original (b) Noisy (c) CBM3D/28.86 dB

(d) CDnCNN/29.46 dB (e) FFDNet/29.54 dB (f) APDL-Net/**30.04** dB

Figure 4: Denoising results of an image from "BSD68" dataset at the noise level of $\sigma = 35$.


(a) GT:(PSNR|SSIM) (b) JPEG:(25.79|0.762) (c) ARCNN:(26.92|0.797)

(d) TNRD:(27.11|0.762) (e) DnCNN:(27.59|0.816) (f) APDL-Net(**28.24**|**0.842**)

Figure 5: Visual comparison on a JPEG compressed image from "Classic5" dataset for quality factor (QF) of 10.

4.4 Blind JPEG Image Deblocking Performance

We further use APDL-Net-B$_{(6,64)}$ to remove artifacts generated by JPEG image compression process. We compare our model with four state-of-the-art methods as listed in Table 2. PSNR and structural similarity index metric (SSIM) are adopted for quantitative assessment. The low quality com-

Methods	Noise Level (σ)											
	5	10	15	20	25	30	35	40	45	50	55*	65*
CBM3D [Dabov *et al.*, 2007]	40.24	35.88	33.49	31.88	30.68	29.71	28.86	28.06	27.82	27.36	26.97	26.29
CDnCNN [Zhang *et al.*, 2017]	40.10	36.11	33.88	32.36	31.22	30.31	29.57	28.94	28.39	27.91	27.46	26.40
CUNLNet$_5$ [Lefkimmiatis, 2018]	40.39	36.20	33.90	32.34	31.17	30.24	29.53	28.91	28.37	27.89	-	-
ADNet-B [Tian *et al.*, 2020]	-	-	33.79	-	31.12	-	29.48	-	-	27.83	-	-
DURR [Zhang *et al.*, 2019a]	-	-	-	-	31.25	-	29.63	-	28.48	-	27.57	26.83
APDL-Net-B$_{(6,64)}$	**40.45**	**36.41**	**34.18**	**32.66**	**31.54**	**30.65**	**29.92**	**29.30**	**28.78**	**28.33**	**27.89**	**27.03**

Table 1: Comparison of the blind color denoising for different noise levels on the "CBSD68" dataset. We measure the average PSNR (dB). Noise levels with * are not present in the training data. The best results are indicated in boldface.

Dataset	QF	ARCNN [Dong *et al.*, 2015]	TNRD [Chen and Pock, 2016]	DnCNN [Zhang *et al.*, 2017]	DDFN [Chen *et al.*, 2019]	APDL-Net-B$_{(6,64)}$
Classic5	10	29.03|0.793	29.28|0.799	29.40|0.803	29.55|0.808	**29.70**|**0.820**
	20	31.15|0.852	31.47|0.858	31.63|0.861	31.70|0.863	**31.93**|**0.874**
	30	32.51|0.881	32.78|0.884	32.91|0.886	33.03|0.888	**33.17**|**0.897**
LIVE1	10	28.96|0.808	29.15|0.811	29.19|0.812	29.39|0.818	**29.43**|**0.824**
	20	31.29|0.873	31.46|0.877	31.59|0.880	31.76|0.883	**31.84**|**0.891**
	30	32.67|0.904	32.84|0.906	32.98|0.909	33.19|0.911	**33.24**|**0.918**

Table 2: Comparison of the PSNR (dB)|SSIM of different JPEG image deblocking methods on the "Classic5" and "LIVE1" datasets. We evaluate the quantitative measurements for three quality factors (QF). The best results are indicated in boldface.

pressed images are produced by the PIL module of python using three JPEG quality factors (QF= 10, 20, 30). As it can be seen from Table 2, the proposed APDL-Net outperforms other compared methods in terms of PSNR and SSIM values on the "Classic5" and "LIVE1" datasets with all JPEG quality factors. However, our APDL-Net is trained through a blind fashion, the difference between the achieved SSIM values by APDL-Net and other methods are more noticeable than the PSNR values. This indicates that APDL-Net prioritizes structural information restoration and perceptual quality enhancement rather than absolute error minimization. Visual results of a JPEG compressed image with quality factors of 10 is shown in Fig. 5.

4.5 Model Size and Running Time Analysis

Fig. 6 visualizes the comparison of the model size, performance, and running time between the proposed APDL-Nets and other advanced image denoising deep networks. Although involving more computation is the inevitable cost of stacking multiple APDL blocks, APDL-Nets are still quite efficient in terms of running time. APDL-Nets improved the performance significantly compared with CDnCNN [Zhang *et al.*, 2017] and MemNet [Tai *et al.*, 2017] in a cost of negligible extra parameters. They also have a clear advantage in speed against the most recent state-of-the-arts RNAN [Zhang *et al.*, 2019b] and RDN [Zhang *et al.*, 2020]. Though techniques like attention mechanism, residual and dense aggregations have been also adopted in RNAN [Zhang *et al.*, 2019b] and RDN [Zhang *et al.*, 2020], the APDL-Net that requires 1.53 million parameters achieved a PSNR value very close to that of the RNAN [Zhang *et al.*, 2019b] and RDN [Zhang *et al.*, 2020] that require 4.8× and 14.3× as many parameters respectively. This shows that APDL-Nets employ these techniques effectively and efficiently, so that a great trade-off between model size, performance and running time is obtained.

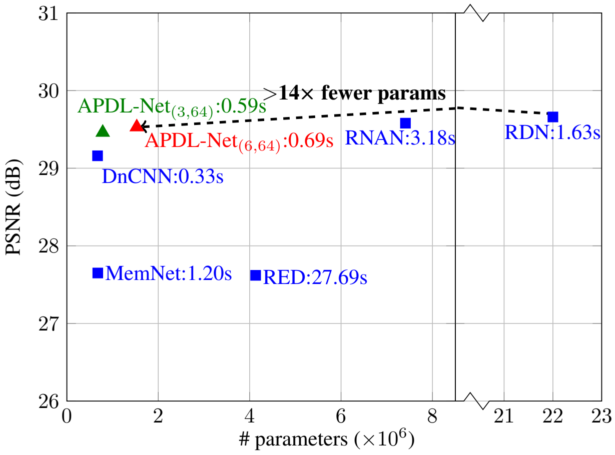


Figure 6: Comparison of model size, PSNR, and running time (in seconds) for noise level of $\sigma = 50$ on the "Kodak24" dataset.

5 Conclusion

In this paper, we present a novel attention-based pyramid dilated lattice topology (APDL) for image denoising. Unlike other CNNs that use both residual and dense aggregations, APDL-Nets take advantage of both aggregation types more efficiently and effectively. APDL-Nets utilize an effective attention-based pyramid dilation schema to efficiently exploit contextual information corresponding to different noise levels. These enable APDL-Nets to obtain a good trade-off between model size and running time alongside a superior performance than many benchmarks in blind image denoising and image deblocking methods. Further improvements on the APDL-Net topology and investigating its capabilities in other vision applications will be considered as future works.

References

[Arbelaez *et al.*, 2010] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):898–916, 2010.

[Chatterjee and Milanfar, 2009] Priyam Chatterjee and Peyman Milanfar. Is denoising dead? *IEEE Transactions on Image Processing*, 19(4):895–911, 2009.

[Chen and Pock, 2016] Yunjin Chen and Thomas Pock. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1256–1272, 2016.

[Chen *et al.*, 2019] C. Chen, Z. Xiong, X. Tian, Z. Zha, and F. Wu. Real-world image denoising with deep boosting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2019.

[Dabov *et al.*, 2007] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007.

[Dong *et al.*, 2015] Chao Dong, Yubin Deng, Chen Change Loy, and Xiaoou Tang. Compression artifacts reduction by a deep convolutional network. In *ICCV*, pages 576–584, 2015.

[Franzen, 1999] Rich Franzen. Kodak lossless true color image suite. *source: http://r0k. us/graphics/kodak*, 4, 1999.

[Gu *et al.*, 2014] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. In *CVPR*, pages 2862–2869, 2014.

[Gu *et al.*, 2018] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, and Tsuhan Chen. Recent advances in convolutional neural networks. *Pattern Recognition*, 77:354 – 377, 2018.

[He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016.

[Hu *et al.*, 2018] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *CVPR*, pages 7132–7141, 2018.

[Huang *et al.*, 2017] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *CVPR*, pages 4700–4708, 2017.

[Lefkimmiatis, 2018] Stamatios Lefkimmiatis. Universal denoising networks: a novel CNN architecture for image denoising. In *CVPR*, pages 3204–3213, 2018.

[Liu *et al.*, 2018] Ding Liu, Bihan Wen, Yuchen Fan, Chen Change Loy, and Thomas S Huang. Non-local recurrent network for image restoration. In *NeurIPS*, pages 1673–1682, 2018.

[Mao *et al.*, 2016] Xiaojiao Mao, Chunhua Shen, and Yu-Bin Yang. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In *NeurIPS*, pages 2802–2810, 2016.

[Mei *et al.*, 2020] Yiqun Mei, Yuchen Fan, Yulun Zhang, Jiahui Yu, Yuqian Zhou, Ding Liu, Yun Fu, Thomas S Huang, and Honghui Shi. Pyramid attention networks for image restoration. *arXiv preprint arXiv:2004.13824*, 2020.

[Nikzad *et al.*, 2020] Mohammad Nikzad, Aaron Nicolson, Yongsheng Gao, Jun Zhou, Kuldip K Paliwal, and Fanhua Shang. Deep residual-dense lattice network for speech enhancement. In *AAAI*, pages 8552–8559, 2020.

[Peng *et al.*, 2019] Yali Peng, Lu Zhang, Shigang Liu, Xiaojun Wu, Yu Zhang, and Xili Wang. Dilated residual networks with symmetric skip connection for image denoising. *Neurocomputing*, 345:67–76, 2019.

[Plötz and Roth, 2018] Tobias Plötz and Stefan Roth. Neural nearest neighbors networks. In *NeurIPS*, pages 1087–1098, 2018.

[Sheikh *et al.*, 2005] Hamid R Sheikh, Zhou Wang, Lawrence Cormack, and Alan C Bovik. Live image quality assessment database release 2 (2005). *URL http://live. ece. utexas. edu/research/quality*, 2005.

[Tai *et al.*, 2017] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu. Memnet: A persistent memory network for image restoration. In *CVPR*, pages 4539–4547, 2017.

[Tian *et al.*, 2020] Chunwei Tian, Yong Xu, Zuoyong Li, Wangmeng Zuo, Lunke Fei, and Hong Liu. Attention-guided cnn for image denoising. *Neural Networks*, 124:117–129, 2020.

[Zeyde *et al.*, 2010] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *ICCS*, pages 711–730. Springer, 2010.

[Zhang *et al.*, 2017] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017.

[Zhang *et al.*, 2018] Kai Zhang, Wangmeng Zuo, and Lei Zhang. FFDNet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018.

[Zhang *et al.*, 2019a] Xiaoshuai Zhang, Yiping Lu, Jiaying Liu, and Bin Dong. Dynamically unfolding recurrent restorer: A moving endpoint control method for image restoration. In *ICLR*, 2019.

[Zhang *et al.*, 2019b] Yulun Zhang, Kunpeng Li, Kai Li, Bineng Zhong, and Yun Fu. Residual non-local attention networks for image restoration. In *ICLR*, 2019.

[Zhang *et al.*, 2020] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2020.