Article

Lightweight Image Denoising Network for Multimedia Teaching System

Xuanyu Zhang [1] , Chunwei Tian [1,2,*] , Qi Zhang [3,4] , Hong-Seng Gan [5] and Tongtong Cheng [6] and Mohd Asrul Hery Ibrahim [4]

1 School of Software, Northwestern Polytechnical University, Xi'an 710129, China; xuanyuzhang@mail.nwpu.edu.cn
2 Research & Development Institute, Northwestern Polytechnical University, Shenzhen 518057, China
3 School of Economics and Management, Harbin Institute of Technology at Weihai, Weihai 264209, China; hit_zq910057@163.com
4 Faculty of Entrepreneurship and Business, Universiti Malaysia Kelantan, Kota Bharu 16100, Malaysia
5 School of AI and Advanced Computing, XJTLU Entrepreneurship College (Taicang), Xi'an Jiaotong-Liverpool University, Suzhou 215400, China
6 School of Power and Energy, Northwestern Polytechnical University, Xi'an 710129, China
* Correspondence: chunweitian@nwpu.edu.cn

Abstract: Due to COVID-19, online education has become an important tool for teachers to teach students. Also, teachers depend on a multimedia teaching system (platform) to finish online education. However, interacted images from a multimedia teaching system may suffer from noise. To address this issue, we propose a lightweight image denoising network (LIDNet) for multimedia teaching systems. A parallel network can be used to mine complementary information. To achieve an adaptive CNN, an omni-dimensional dynamic convolution fused into an upper network can automatically adjust parameters to achieve a robust CNN, according to different input noisy images. That also enlarges the difference in network architecture, which can improve the denoising effect. To refine obtained structural information, a serial network is set behind a parallel network. To extract more salient information, an adaptively parametric rectifier linear unit composed of an attention mechanism and a ReLU is used into LIDNet. Experiments show that our proposed method is effective in image denoising, which can also provide assistance for multimedia teaching systems.

Keywords: lightweight CNN; dynamic convolution; adaptive activation function; image denoising; multimedia teaching system

MSC: 68T45

1. Introduction

Traditional teaching requires students to learn knowledge through face-to-face methods. Although it has good effects, it has higher requirements for students in terms of time and space. To break these limitations, online education has been developed. It mainly depends on a multimedia system (platform) to complete teaching tasks. Also, obtained images in the multimedia system constitute important media for human-to-human interaction. However, these images often suffer from some challenges from noise caused by camera shake, hardware quality, weather [1], etc. After analyzing the process of collecting and disseminating relevant teaching resources, teaching images often suffer from challenges, i.e., noise from collection equipment. To address these mentioned drawbacks, image denoising techniques are also applied.

An image denoising technique is a classical low-level technique and has been applied in various fields, i.e., activities recognition [2] and remote sensing [3]. For instance, an expected patch log likelihood (EPLL) [4] used a mixed Gaussian model to learn prior

knowledge from many natural image blocks for image denoising. Also, block matching and three-dimensional filtering (BM3D) [5] utilized collaborative filtering on similar two-dimensional image blocks to remove noise. A weighted nuclear norm minimization (WNNM) algorithm can exploit an image's non-local self-similarity to extract more information for image denoising [6]. Although these methods can restore images, they face some challenges. That is, they excessively rely on manual adjustment of parameters and complex parameters. Due to strong expressive ability, convolutional neural networks (CNNs) have obtained abilities of feature extraction. Thus, CNNs have been applied in the field of image denoising. For instance, a denoising convolutional neural network (DnCNN) first utilized convolution and residual learning operations to complete denoising work [7]. To suppress the influence of the background on noise, an attention mechanism is fused in a CNN to separate background and foreground to suppress noise [8]. To address image denoising under complex scenes, a dynamic convolution is used in a CNN to achieve an adaptive denoiser, according to different noisy images [9]. To obtain a better denoising effect, a combination of an omni-dimensional dynamic convolution and attention mechanisms is integrated into a CNN to enhance the expressive ability of a denoising network, which can enhance interaction quality of the multimedia teaching system between students and teachers. Le et al. uses two phases, i.e., a feature augmentation stage and a feature refinement stage, to design a CNN to extract more accurate structural information for image denoising [10]. To reduce the complexity of a denoiser, Lin et al. simplified the residual spatial–spectral module and knowledge distillation to achieve a lightweight method to accelerate noise removal [11] Alternatively, a combination of a non-local algorithm and a residual CNN achieves a lightweight CNN to suppress noise [12]. That is, we present a lightweight image denoising network as well as LIDNet for multimedia teaching systems. LIDNet uses a parallel sub-network to mine complementary information for image denoising. To achieve an adaptive CNN, a dynamic convolution based on kernel information and input channel number and output channel number fused into an upper network can automatically adjust parameters to achieve a robust CNN, according to different input noisy images. That also enlarges differences in network architecture, which can improve the denoising effect. To refine the obtained structural information, a serial network is set behind a parallel network. To extract more salient information, an adaptively parametric rectifier linear unit composed of an attention mechanism and a ReLU is used in LIDNet. Experiments show that our proposed method is effective in image denoising, which may also provide assistance for multimedia teaching systems.

The contributions of the proposed method can be summarized as follows:

1. A dynamic convolution based on kernel information and input channel number and output channel number is used to adaptively mine more useful information, according to different input images.
2. A combination of attention mechanism and ReLU is set behind each convolutional layer in addition to the final convolutional layer to enhance the same distributions of training samples for pursuing better denoising performance.
3. Our denoising method is useful for enhancing the interaction quality of a multimedia teaching system between teacher and student.

The remainder of this paper is organized as follows. Section 2 lists related work about image denoising based on dual networks and dynamic networks. Section 3 provides detailed information of the proposed method. Section 4 presents analysis of our proposed method and results. Section 5 gives the conclusion of this paper.

2. Related Work

2.1. A Dual Network for Image Denoising

To extract complementary information, dual networks are developed in image denoising [13]. For instance, Tian et al. [13] presented a dual denoising network with sparse mechanism as well as DudeNet to extract complementary information to enhance denoising effects. Alternatively, Bai et al. [14] achieved a dual network via encoder–decoder and

channel attention architecture to extract local and non-local information for image denoising, where image spatial details and semantic information can be obtained by a criss-cross attention. To extract more information, Holla et al. [15] used edge information to design a CNN to capture high-frequency information in image denoising. Zhang et al. fused different masks into a CNN to facilitate complementary information to suppress noise [16]. To mine more high-frequency information, Qiao et al. combined two different networks and a sharpening loss function to improve the quality of visual denoising images [17]. Liu et al. used a wavelet decomposition technique to achieve a wide CNN to prevent vanishing and exploding gradient problems [18]. To extract salient noise information, Chen et al. fused a CNN and a transformer to implement a parallel network to extract structural information and key information based on pixel relations for improving denoising effects [19]. For medical noisy image denoising, Jiang et al. [20] used residual connections and dilated convolutions to achieve a heterogeneous dual network to mine more complementary information to suppress noise. According to the mentioned illustrations, we can see that dual networks are useful for image denoising. Inspired by that, we design a dual network architecture for image denoising in this paper.

2.2. Dynamic Networks for Image Denoising

To enhance the robustness of the image denoiser, a dynamic network is created [21]. For instance, Song et al. [21] combined dynamic convolutions and residual learning operations into a CNN to dynamically adjust parameters to obtain a robust denoising network, according to different input images. Du et al. [22] exploited a dynamic attention mechanism to better extract salient information for image denoising. Alternatively, Shen et al. [23] fused a spatial module and dynamic convolution to obtain more spatial context information to obtain better denoising performance. Tian et al. [9] used dynamic convolution and wavelet transform to extract more useful information to improve denoising effects. According to the mentioned descriptions, we can see that dynamic convolution is effective in image denoising. Motivated by that, we use a dynamic convolution in this paper, according to different kernel and channel information.

3. Proposed Method

3.1. Network Architecture

The proposed 17-layer LIDNet combines a parallel and series architecture. The parallel architecture is composed of a 6-layer block called the dynamic feature extraction block (DFEB) and a 6-layer block named complementary feature extraction block (CFEB). The series architecture contains an 11-layer block called the cascaded purification block (CPB), which is shown in Figure 1. DFEB uses a dynamic convolutional layer to adaptively extract structural information, including kernel information and channel information. To extract complementary information, CFEB use several stacked convolutional layers, BN and a combination of attention mechanism and activation function to extract complementary salient information. Also, a residual learning operation is used to connect the obtained information from a parallel network. To prevent over enhancement, a 11-layer CPB is designed behind the parallel network. To construct a clean image, a residual learning operation is used to act between an input image and output image of LIDNet. This process can be shown as Equation (1).

$$\begin{aligned} I_C &= LIDNet(I_N) \\ &= CPB(DFEB(I_N) + CFEB(I_N)) + (I_N), \end{aligned} \tag{1}$$

where I_C represents an output of LIDNet, which is regraded to a denoised image. I_N denotes the input noisy image, and $LIDNet()$ expresses a function of LIDNet. $DFEB$, $CFEB$, and CPB stand for functions of DFEB, CFEB, and CPB, respectively. $+$ is a residual learning operation, which is also shown as $\oplus$ in Figure 1. Furthermore, the MSE loss function of LIDNet is introduced in Section 3.2.
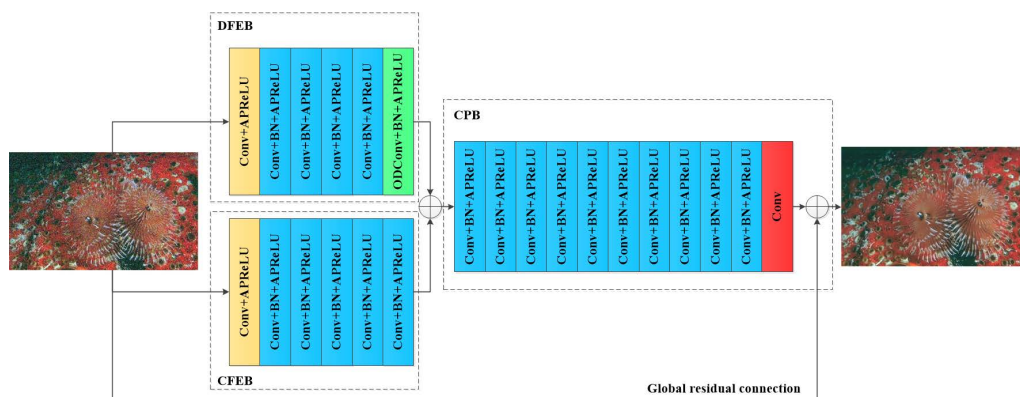
Figure 1. Network architecture of LIDNet.

3.2. Loss Function

To fairly compare with the famous denoising benchmark of DnCNN, a mean squared error (MSE) [24] is chosen as the loss function to train LIDNet. Specifically, MSE uses pairs of $\{I_N^i, I_C^i\}$ $(1 \leq i \leq n)$ to train our LIDNet in a supervised way, where I_N^i and I_C^i are defined as the *i*-th noisy and clean image, respectively. *n* represents the number of image pairs in the training dataset. LIDNet also uses the popular Adam [25] to obtain reasonable parameters. The mathematical expression of the loss function is as follows:

$$L(\theta) = \frac{1}{2n} \sum_{i=1}^{n} \left\| LIDNet(I_N^i) - I_C^i \right\|^2, \tag{2}$$

where *L* is a loss function of MSE and θ stands for learned parameters.

3.3. Dynamic Feature Extraction Block

The first layer in DFEB consists of a convolutional and a rectified linear unit (ReLU) [26] operation. The following 4 layers are composed of a convolutional, a batch normalization (BN) operation and an adaptively parametric rectified linear unit (APReLU) [27]. And the final layer has an omni-dimensional dynamic convolution (ODConv) [28], BN and APReLU. In terms of parameter setting, the input channel number of the first convolutional operation is the same with the channel of the input images. If the input image is color, input channel number of LIDNet is 3. otherwise, input channel number of LIDNet is 1. Other numbers of input and output channels of all the layers are 64. Every size of convolutional kernels in LIDNet is set to 3×3. And the output of the DFEB is fused with the output of the CFEB via a concatenation connection. Mathematical expression of the DFEB is shown as follows:

$$\begin{aligned} O_{DFEB} &= DFEB(I_N) \\ &= Conv_6(4Conv_2(Conv_1(I_N))) \\ &= APReLU(BN(ODConv(4APReLU(4BN(4Conv(ReLU(Conv(I_N)))))))), \end{aligned} \tag{3}$$

where O_{DFEB} is the output of DFEB. $DFEB()$ expresses a function of DFEB. $Conv_1$ means the first layer of the DFEB, $4Conv_2$ means 4 stacked layers in DFEB, which form the second layer to the fifth layer, and $Conv_6$ means the last layer of the DFEB. $Conv$ stands for a function of a convolutional operation, $ReLU$ stands for an activation function of ReLU, BN stands for the batch normalization operation, $APReLU$ stands for another activation function of APReLU, and $ODConv$ stands for ODConv operation. $4APReLU(4BN(4Conv()))$ is the equation for the $4Conv_2$.

3.4. Complementary Feature Extraction Block

The lower branch in a parallel architecture is CFEB, which is responsible for extracting complementary features by a different network architecture. The first layer in CFEB consists of a convolutional and a ReLU operation. And the following 5 layers are composed of a

stacked combination of convolutional, BN, and APReLU operations. As shown in Figure 1, the difference between DFEB and CFEB is mainly reflected on the last convolutional layer. Specifically, the CFEB uses a common convolution operation to replace the ODConv as the final layer in the DFEB. Input and output channel numbers of the final convolutional layer are both 64. The mathematical expression of CFEB is as follows:

$$
\begin{aligned}
O_{CFEB} &= CFEB(I_N) \\
&= 5Conv_2(Conv_1(I_N)) \\
&= 5APReLU(5BN(5Conv(ReLU(Conv(I_N))))),
\end{aligned}
\tag{4}
$$

where O_{CFEB} is the output of CFEB. $CFEB()$ expresses a function of CFEB. $Conv_1$ means the first layer of the CFEB and $5Conv_2$ means 5 stacked layers in the DFEB, which is as the second layer to the sixth layer in the CFEB. $5APReLU(5BN(5Conv()))$ is equal to $5Conv_2$.

3.5. Concatenated Purification Block

To refine fused structural information from DFEB and CFEB, CPB is set as the last part of LIDNet. Specifically, its first 10 layers in CPB are composed of convolutional, BN, and APReLU operations. And its last layer is simply a common convolutional operation, which is used to construct clean images. To construct a clean image, a residual learning operation is used to act between an input image and output image of LIDNet. The numbers of input and output channels are 64 except the output channel number of the final convolutional layer, which is the same as the channel of the input image. The mathematical expression of CPB is as follows:

$$
\begin{aligned}
O_{CPB} &= CPB(O_{DFEB} + O_{CFEB}) \\
&= Conv_{11}(10Conv_1(O_{DFEB} + O_{CFEB})) \\
&= Conv(10APReLU(10BN(10Conv(O_{DFEB} + O_{CFEB})))),
\end{aligned}
\tag{5}
$$

$$
I_C = I_N - O_{CPB},
\tag{6}
$$

where O_{CPB} is the output of CPB. $CPB()$ expresses a function of CPB. $10Conv_1$ means the 10 stacked layers in CPB, which form the first layer to the tenth layer, and $Conv_{11}$ means the last layer of the CPB. $10APReLU(10BN(10Conv()))$ is equal to $10Conv_1$.

4. Experiments

4.1. Datasets

The video quality of many courses inevitably declines due to the impact of the environment and equipment during shooting. To achieve better performance in multimedia, we propose LIDNet to denoise these teaching images. The architecture of our LIDNet is shown in Figure 1.

For image denoising with Gaussian noise, 400 images with sizes of 180×180 from Ref. [29] are used to train a denoising model. Three different denoising models with noise levels of 15 and 25 can be trained, respectively. To train a blind denoising model with noise levels from 0 to 55, a blind model is trained. Specifically, patch sizes are set to 40×40.

To fairly test denoising performance, public BSD68 [30], Set 12 [31], Kodak24 [32], and collected educational images from the Internet are used as test datasets. Guassian noise with noise levels of 15 and 25 is added on BSD68, Set12, Kodak24, and collected educational images from the Internet to test the denoising performance of the proposed method.

4.2. Parameter Setting

This paper has the following experimental settings. The number of training epochs is 180. The original learning rate is 1×10^{-3} and it will decline to 0.2 times when the epoch is 30, 60, and 90, respectively. Batch size is set to 128. Adam is used to optimize parameters [25], where β_1 is 0.9 and β_2 is 0.999. More parameters can be found in Ref. [13].

The LIDNet can be trained on a PC with Intel Xeon Gold 6330 Processor and one Nvidia GeForce RTX 3090. Furthermore, all the codes run on Ubuntu 20.04 with Python 3.8, PyTorch 1.11.0, and CUDA 11.7.

4.3. Network Analysis

This paper uses a parallel network architecture to extract complementary information for image denoising, where a parallel network consists of an upper network (also regarded as dynamic feature extraction block, DFEB) and lower network (also regarded as complementary feature extraction block, CFEB). It connects a serial architecture (concatenated purification block, CPB) to extract more hierarchical structural information. Also, each branch in the parallel network is composed of six layers of stacked architecture. The upper network is composed of a Conv + APReLU, four Conv + BN + APReLU, and a ODConv + BN + APReLU, where APReLU [27] is composed of an attention mechanism and a ReLU is used to extract salient information and nonlinear information. Also, OD-Conv [28] utilizes convolutional kernel information and channel information to dynamically learn parameters to adaptively train a denoising model for different given noisy images. 'LIDNet without global residual connection and ODConv + BN + APReLU' has an improvement of 0.013dB compared to 'LIDNet with only Conv + APReLU in the upper network' in Table 1, which describes the effectiveness of four Conv + BN + APReLU in the upper network for image denoising. Also, the denoising effect of 'Conv + APReLU' in the upper network is verified by 'The combination of lower network and CPB' and 'LIDNet with only Conv + APReLU in upper network and without global residual connection' in Table 1. To test the denoising performance of DFEB, we use 'The combination of lower network and CPB' and 'LIDNet without global residual connection' to conduct comparative experiments. As shown in Table 1, we can see that 'LIDNet without global residual connection' exceeds 'The combination of lower network and CPB' in terms of PSNR. That shows that DFEB in the parallel network is effective for image denoising. Additionally, to test complementarity of two sub-networks, 'The combination of lower network and CPB' is superior to 'CPB' in terms of image denoising in Table 1, which shows the superiority of a parallel network for image denoising. To prevent the interference of upper and lower networks, a serial network is set behind a parallel network to refine the obtained structural information for image denoising. Finally, a global residual connection is employed between outputs of the first layer in a lower network and the last layer in the CPB to construct clean images.

Table 1. Denoising results (average PSNR (dB)) of several networks on BSD68 for noise level of 25.

Different Networks	PSNR(dB)
CPB	28.937
The combination of lower network and CPB	28.944
LIDNet with only Conv + APReLU in upper network and without global residual connection	29.219
LIDNet without global residual connection and ODConv + BN + APReLU	29.232
LIDNet without global residual connection	29.237
LIDNet (Ours)	29.247

4.4. Comparisons with State of the Art

To test the effectiveness of proposed method, we choose several popular denoising methods, i.e., EPLL, BM3D, WNNM, DnCNN, image restoration CNN (IRCNN) [33], fast and flexible denoising network (FFDNet) [34], and a cascade of shrinkage fields (CSF) [35] as comparative methods on the BSD68 and Set12 to conduct experiments. As shown in Table 2, we can see that our LIDNet has obtained the best denoising result on the BSD68 for $\sigma = 15$ and $\sigma = 25$. For instance, our LIDNet has an improvement of 0.11 dB compared to

IRCNN for $\sigma = 15$. That shows that our method is effective for gray noisy image denoising. To verify good denoising performance for a single gray noisy image, different methods on Set12 are used to conduct denoising effects. As illustrated in Table 3, we can see that our LIDNet has obtained the best denoising effect for single noisy image denoising. For instance, our LIDNet has obtained an improvement of 0.09 dB compared to a popular denoising method, i.e., WNNM for a noise level of 15. That shows that our method is a good denoising tool for low-frequency noisy image denoising. Our LIDNet has obtained an improvement of 0.06 dB compared to a popular denoising method, i.e., IRCNN for noise level of 25. That shows that our method is a good denoising tool for high-frequency noisy image denoising. According to that, we can see that our method is effective for single noisy image denoising. Furthermore, to further demonstrate the denoising performance of our LIDNet on color images, Table 4 records the denoising results from different models with different noise levels. Compared with popular methods, i.e., IRCNN, FFDNet, D-BSN, FL(NLM), and FL(BM3D), our LIDNet has also achieved improvements in denoising performance for color noisy images. This also proved the effectiveness of LIDNet in processing color noisy images.

Table 2. Average PSNR (dB) results of several networks on BSD68 for noise levels of 15 and 25.

Methods	EPLL [4]	BM3D [5]	WNNM [6]	DnCNN [7]	IRCNN [23]	FFDNet [34]	CSF [35]	LIDNet (Ours)
$\sigma = 15$	31.21	31.07	31.37	31.73	31.63	31.62	31.24	31.74
$\sigma = 25$	28.68	28.57	28.83	29.23	29.15	29.19	28.74	29.25

**Table 3.** PSNR (dB) results of different methods on Set12 with noise levels of 15 and 25.

Images	C.man	House	Peppers	Starfish	Monarch	Airplane	Parrot	Lena	Barbara	Boat	Man	Couple	Average
Noise level							15						
EPLL [4]	31.8	34.17	32.64	31.13	32.10	31.19	31.42	33.92	31.38	31.93	32.00	31.93	32.14
BM3D [5]	31.91	34.93	32.69	31.14	31.85	31.07	31.37	34.26	33.10	32.13	31.92	32.10	32.37
WNNM [6]	32.17	35.13	32.99	31.82	32.71	31.39	31.62	34.27	33.60	32.27	32.11	32.17	32.70
IRCNN [33]	32.55	34.89	33.31	32.02	32.82	31.70	31.84	34.53	32.43	32.34	32.40	32.40	32.77
FFDNet [34]	32.43	35.07	33.25	31.99	32.66	31.57	31.81	34.62	32.54	32.38	32.41	32.46	32.77
CSF [35]	31.95	34.39	32.85	31.55	32.33	31.33	31.37	34.06	31.92	32.01	32.08	31.98	32.32
LIDNet (Ours)	31.93	35.03	33.24	32.21	33.09	31.73	31.93	34.54	32.57	32.40	32.29	32.46	32.79
Noise level							25						
EPLL [4]	29.26	32.17	30.17	28.51	29.39	28.61	28.95	31.73	28.61	29.74	29.66	29.53	29.69
BM3D [5]	29.45	32.85	30.16	28.56	29.25	28.42	28.93	32.07	30.71	29.90	29.61	29.71	29.97
WNNM [6]	29.64	33.22	30.42	29.03	29.84	28.69	29.15	32.24	31.24	30.03	29.76	29.82	30.26
IRCNN [33]	30.08	33.06	30.88	29.27	30.09	29.12	29.47	32.43	29.92	30.17	30.04	30.08	30.38
FFDNet [34]	30.10	33.28	30.93	29.32	30.08	29.04	29.44	32.57	30.01	30.25	30.11	30.20	30.44
CSF [35]	29.48	32.39	30.32	28.80	29.62	28.72	28.90	31.79	29.03	29.76	29.71	29.53	29.84
LIDNet (Ours)	30.20	33.10	30.83	29.39	30.40	29.09	29.52	32.39	30.04	30.18	30.04	30.08	30.44

To comprehensively test the denoising effect of our proposed method, we use qualitative analysis to measure the effects of visual images. Specifically, we choose one area of denoising images from BM3D, FFDNet, IRCNN, and LIDNet as observation areas. If the observation area is clearer, its corresponding method shows better denoising performance. As shown in Figures 2–4, we can see that our LIDNet is clearer than the results of other methods. In Figure 3, other methods can obtain more incorrect texture information. Because real noisy images are difficult to obtain in the world, we choose Guassian noise added on educational images to test the performance of the proposed method for educational image denoising. In Figure 4, we can see that our method can obtain clearer detailed information for noisy educational image denoising. Thus, that not only shows that our method is

superior to other methods in terms of qualitative analysis, but also that it is robust for different scenes in terms of image denoising.

Table 4. Average PSNR (dB) results of different methods on CBSD68 and Kodak24 datasets with noise levels of 15 and 25.

| Datasets	Models	$\sigma = 15$	$\sigma = 25$
CBSD68	CBM3D [5]	33.52	30.71
	DnCNN [7]	33.98	31.31
	IRCNN [33]	33.86	31.16
	FFDNet [34]	33.80	31.18
	D-BSN [30]	33.56	30.61
	FL (NLM) [12]	-	31.01
	FL (BM3D) [12]	-	31.13
	LIDNet (Ours)	33.99	31.37
Kodak24	CBM3D [5]	34.28	31.68
	DnCNN [7]	34.73	32.23
	IRCNN [33]	34.56	32.03
	FFDNet [34]	34.55	32.11
	D-BSN [30]	33.74	31.64
	FL (NLM) [12]	-	32.11
	FL (BM3D) [12]	-	32.26
	LIDNet (Ours)	34.57	32.12
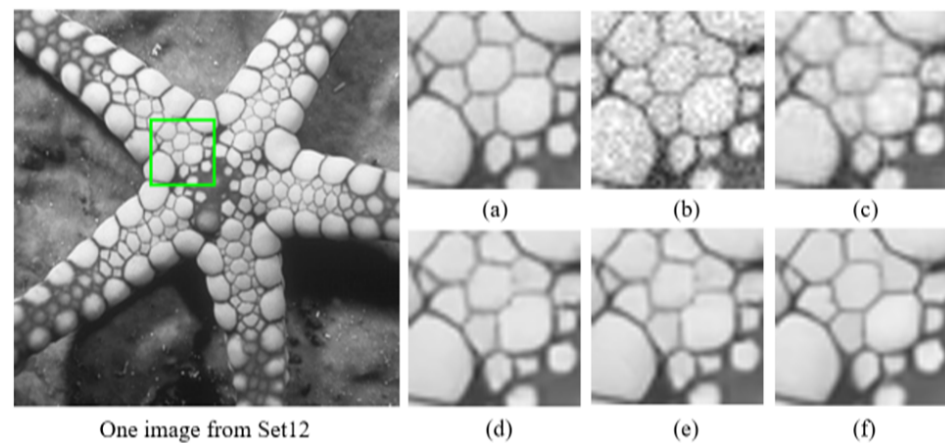
Figure 2. Visual effects of several denoising methods on an image from Set12 for noise level of 15. (**a**) Original image, (**b**) noisy image (24.60 dB), (**c**) BM3D [4] (30.98 dB), (**d**) FFDNet [34] (31.99 dB), (**e**) IRCNN [33] (32.02 dB), and (**f**) LIDNet (Ours) (32.21 dB).
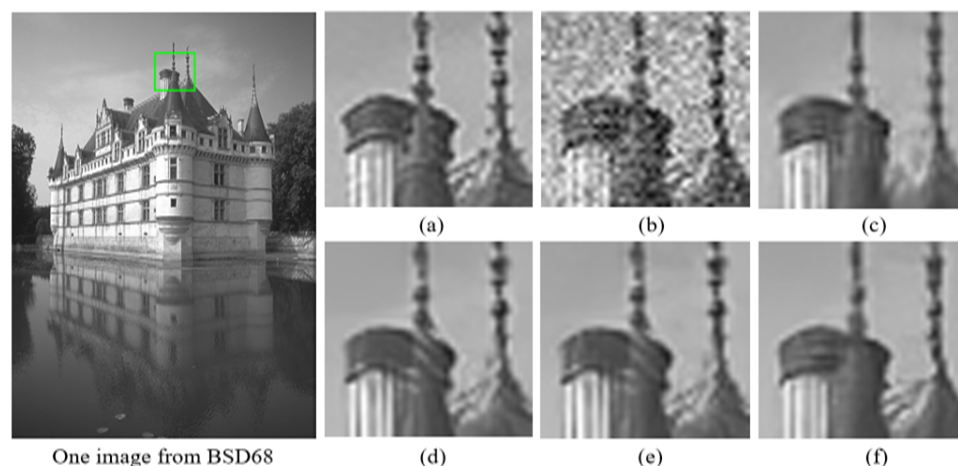
Figure 3. Visual effects of several denoising methods on an image from BSD68 for noise level of 25. (**a**) Original image, (**b**) noisy image (20.19 dB), (**c**) BM3D [4] (29.49 dB), (**d**) FFDNet [34] (30.04 dB), (**e**) IRCNN [33] (30.07 dB), and (**f**) LIDNet (Ours) (30.14 dB).
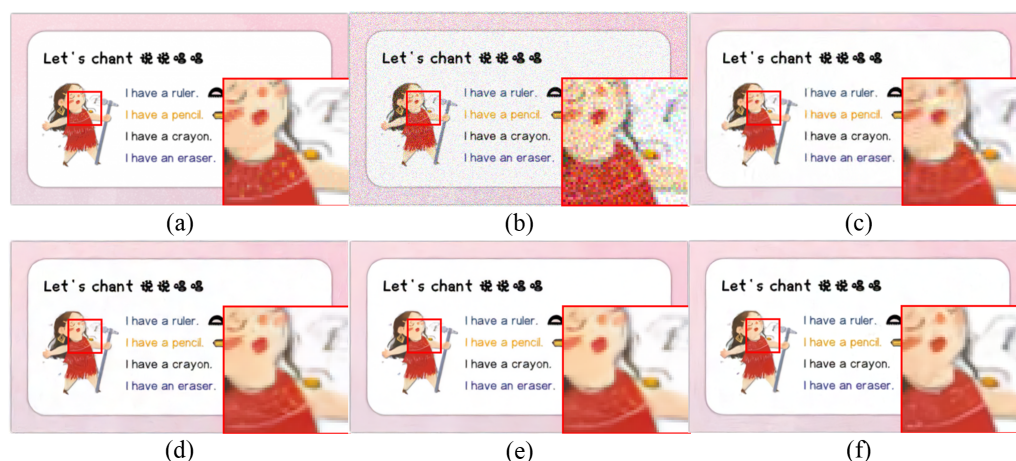
Figure 4. Visual results of several denoising methods on a real teaching image for noise level of 25. (**a**) Original image, (**b**) noisy image (20.18 dB), (**c**) BM3D [4] (36.22 dB), (**d**) IRCNN [33] (36.50 dB), (**e**) FFDNet [34] (36.73 dB) and (**f**) LIDNet (Ours) (36.82 dB).

5. Conclusions

Multimedia teaching systems have become a popular tool for online education. However, interacted images from a multimedia teaching system may suffer from noise. In this paper, we present a lightweight image denoising network as well as LIDNet for multimedia teaching systems. LIDNet uses a parallel network to mine complementary information. To improve robustness of the obtained denoiser, an omni-dimensional dynamic convolution is designed in one sub-network from the parallel network to automatically adjust parameters to achieve an adaptive CNN. That also enlarges the differences in network architecture, which can improve the denoising effect. To refine the obtained structural information, a serial network is set behind a parallel network. To extract more salient information, an adaptively parametric rectifier linear unit composed of an attention mechanism and a ReLU is used in LIDNet. Experiments show that our LIDNet is effective in image denoising, which can also provide assistance for multimedia teaching systems.

Author Contributions: Methodology, X.Z. and C.T.; software, X.Z., Q.Z. and H.-S.G.; validation, X.Z., C.T., Q.Z. and H.-S.G.; formal analysis, X.Z., C.T., Q.Z. and T.C.; writing—original draft, X.Z.; writing—review & editing, X.Z., C.T., Q.Z., T.C. and M.A.H.I.; supervision, Q.Z.; project administration, C.T.; funding acquisition, C.T. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Khorov, E.; Kureev, A.; Levitsky, I.; Akyildiz, I.F. A phase noise resistant constellation rotation method and its experimental validation for NOMA Wi-Fi. *IEEE J. Sel. Areas Commun.* **2022**, *40*, 1346–1354. [CrossRef]
2. Gu, F.; Khoshelham, K.; Valaee, S.; Shang, J.; Zhang, R. Locomotion activity recognition using stacked denoising autoencoders. *IEEE Internet Things J.* **2018**, *5*, 2085–2093. [CrossRef]
3. Liu, P.; Huang, F.; Li, G.; Liu, Z. Remote-sensing image denoising using partial differential equations and auxiliary images as priors. *IEEE Geosci. Remote Sens. Lett.* **2011**, *9*, 358–362. [CrossRef]
4. Zoran, D.; Weiss, Y. From learning models of natural image patches to whole image restoration. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 479–486.
5. Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Process.* **2007**, *16*, 2080–2095. [CrossRef]
6. Gu, S.; Zhang, L.; Zuo, W.; Feng, X. Weighted nuclear norm minimization with application to image denoising. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 2862–2869.
7. Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Trans. Image Process.* **2017**, *26*, 3142–3155. [CrossRef]
8. Tian, C.; Xu, Y.; Li, Z.; Zuo, W.; Fei, L.; Liu, H. Attention-guided CNN for image denoising. *Neural Netw.* **2020**, *124*, 117–129. [CrossRef] [PubMed]
9. Tian, C.; Zheng, M.; Zuo, W.; Zhang, B.; Zhang, Y.; Zhang, D. Multi-stage image denoising with the wavelet transform. *Pattern Recognit.* **2023**, *134*, 109050. [CrossRef]
10. Le T.H.; Lin, P.H.; Huang, S.C. LD-Net: An efficient lightweight denoising model based on convolutional neural network. *IEEE Open J. Comput. Soc.* **2020**, *1*, 173–181. [CrossRef]
11. Lin, Y.; Cai, Z.; Li, J.; Zhang, J. Lightweight Remote Sensing Image Denoising via Knowledge Distillation. In Proceedings of the 2022 IEEE 24th International Workshop on Multimedia Signal Processing (MMSP), Shanghai, China, 26–28 September 2022; pp. 1–7.
12. Guo, Y.; Davy, A.; Facciolo, G.; Morel, J.-M.; Jin, Q. Fast, nonlocal and neural: A lightweight high quality solution to image denoising. *IEEE Signal Process. Lett.* **2021**, *28*, 1515–1519. [CrossRef]
13. Tian, C.; Xu, Y.; Zuo, W.; Du, B.; Lin, C.-W.; Zhang, D. Designing and training of a dual CNN for image denoising. *Knowl.-Based Syst.* **2021**, *226*, 106949. [CrossRef]
14. Bai, Y.; Liu, M.; Yao, C.; Lin, C.; Zhao, Y. MSPNet: Multi-stage progressive network for image denoising. *Neurocomputing* **2023**, *517*, 71–80. [CrossRef]
15. Holla, S.; Park, N.; Lee, B. EFID: Edge-Focused Image Denoising Using a Convolutional Neural Network. *IEEE Access* **2023**, *11*, 9613–9626.
16. Zhang, D.; Zhou, F.; Jiang, Y.; Fu, Z. Mm-bsn: Self-supervised image denoising for real-world with multi-mask based on blind-spot network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 24–31 January 2023; pp. 4188–4197.
17. Qiao, S.; Yang, J.; Zhang, T.; Zhao, C. Layered input GradiNet for image denoising. *Knowl.-Based Syst.* **2022**, *254*, 109587. [CrossRef]
18. Liu, G.; Dang, M.; Liu, J.; Xiang, R.; Tian, Y.; Luo, N. True wide convolutional neural network for image denoising. *Inf. Sci.* **2022**, *610*, 171–184. [CrossRef]
19. Chen, Y.; Yin, M.; Li, Y.; Cai, Q. CSU-Net: A CNN-Transformer parallel network for multimodal brain tumour segmentation. *Electronics* **2022**, *11*, 2226. [CrossRef]
20. Jiang, X.; Jin, Y.; Yao, Y. Low-dose CT lung images denoising based on multiscale parallel convolution neural network. *Vis. Comput.* **2021**, *37*, 2419–2431. [CrossRef]
21. Song, Y.; Zhu, Y.; Du, X. Dynamic residual dense network for image denoising. *Sensors* **2019**, *19*, 3809. [CrossRef]

22.	Du, Y.; Han, G.; Tan, Y.; Xiao, C.; He, S. Blind image denoising via dynamic dual learning. *IEEE Trans. Multimed.* **2020**, *23*, 2139–2152. [CrossRef]

23.	Shen, H.; Zhao, Z.-Q.; Zhang, W. Adaptive Dynamic Filtering Network for Image Denoising. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; pp. 2227–2235.

24.	Douillard, C.; Jézéquel, M.; Berrou, C.; Electronique, D.; Picart, A.; Didier, P.; Glavieux, A. Iterative correction of intersymbol interference: Turbo-equalization. *Eur. Trans. Telecommun.* **1995**, *6*, 507–511. [CrossRef]

25.	Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.

26.	Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems 2012: 26th Annual Conference on Neural Information Processing Systems 2012, Lake Tahoe, NV, USA, 3–6 December 2012; Volume 25.

27.	Zhao, M.; Zhong, S.; Fu, X.; Tang, B.; Dong, S.; Pecht, M. Deep residual networks with adaptively parametric rectifier linear units for fault diagnosis. *IEEE Trans. Ind. Electron.* **2020**, *68*, 2587–2597. [CrossRef]

28.	Li, C.; Zhou, A.; Yao, A. Omni-dimensional dynamic convolution. *arXiv* **2022**, arXiv:2209.07947.

29.	Chen, Y.; Pock, T. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1256–1272. [CrossRef] [PubMed]

30.	Wu, X.; Liu, M.; Cao, Y.; Ren, D.; Zuo, W. Unpaired learning of deep image denoising. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 352–368.

31.	Mairal, J.; Bach, F.; Ponce, J.; Sapiro, G.; Zisserman, A. Non-local sparse models for image restoration. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009; pp. 2272–2279.

32.	Franzen, R. Kodak Lossless True Color Image Suite. 1999. Available online: https://r0k.us/graphics/kodak/ (accessed on 15 November 1999).

33.	Zhang, K.; Zuo, W.; Gu, S.; Zhang, L. Learning deep CNN denoiser prior for image restoration. In Proceedings of IEEE the Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3929–3938.

34.	Zhang, K.; Zuo, W.; Zhang, L. FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Trans. Image Process.* **2018**, *27*, 4608–4622. [CrossRef] [PubMed]

35.	Schmidt, U.; Roth, S. Shrinkage fields for effective image restoration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 2774–2781.