

DDUNet: Dense Dense U-Net with Applications in Image Denoising

Fan Jia* Wing Hong Wong* Tiejong Zeng
Department of Mathematics, The Chinese University of Hong Kong
Shatin, Hong Kong

{jiafan, whwong, zeng}@math.cuhk.edu.hk

Abstract

The investigation of CNN for image denoising has arrived at a serious bottleneck and it is extremely difficult to design an efficient network for image denoising with better performance and fewer parameters. A nice starting point for this is the cascading U-Nets architecture which has been successfully applied in numerous image-to-image tasks such as image denoising and segmentation. However, the previous related models often focused on the local architecture in each U-Net rather than the connection between U-Nets, which strictly limits their performances. To further improve the connection between U-Nets, we propose a novel cascading U-Nets architecture with multi-scale dense processing, named Dense Dense U-Net (DDUNet). The multi-scale dense processing connects the feature maps in each level cross cascading U-Nets, which has several compelling advantages: they alleviate the vanishing gradient problem, strengthen feature propagation and encourage feature reuse. Furthermore, we develop a series of related important techniques to improve model performance with fewer parameters. Extensive experimental results on both synthetic and real noisy datasets demonstrate that the proposed model achieves outstanding results with fewer parameters. Meanwhile, experimental results show clearly that the proposed DDUNet is good at edge recovery and structure preservation in real noisy image denoising.

1. Introduction

Image denoising aims to recover the latent clean image from the degraded observation. As a fundamental problem in computer vision, image denoising is a prerequisite step for many high-level computer vision tasks, such as image segmentation [42], object detection [40], and video denoising [12, 56]. It also has a wide range of application such as medical imaging [23, 27, 43], satellite imaging [36, 44], and image compression [25, 45].

*indicates equal contributions.

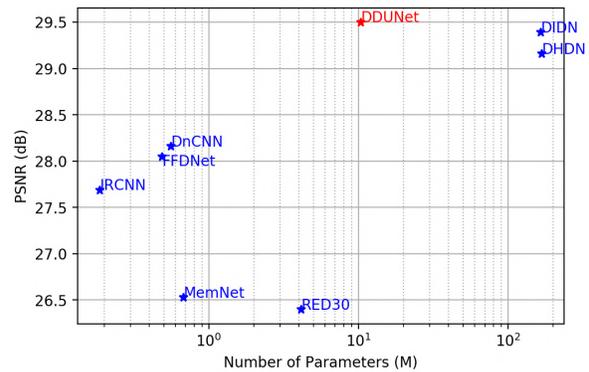


Figure 1: **Parameters vs. Performance.** The PSNRs are evaluated on color Urban100 with the $\sigma = 50$. All models do not share parameters except MemNet [46]. Our method has leading performance with reasonable parameters.

Along with the era of neural networks, many methods have been proposed based on prior knowledge and machine learning [6, 9], which often require solving complex optimization problems and tuning parameters manually. In the meantime, convolutional neural network (CNN) has also been carefully studied for image denoising, and the first state-of-the-art CNN model, DnCNN [58], was proposed by Zhang *et al.* in 2016. Since the success of DnCNN, neural networks have become the mainstream research directions towards image denoising. However, the research of CNN for image denoising has reached a serious bottleneck. Indeed, in the development of CNN, many denoising models such as MWCNN [31, 32] and DHDN [38] have adopted U-Net structure [42] for its consideration of multi-scale features, yet models with better performance need more parameters in general and their PSNR only advance marginally. Meanwhile, some cascading U-Net architecture like MCU-Net [4] employ different techniques such as residual learning and local dense connection for better performance. However, none of them fully consider the multi-scale feature cross different U-Nets except for upsampling, which is useful for mining more image structures and keep

more structures in the target image. Moreover, to seek good performance, some excellent models such as DHDN [38] also use hundreds of millions of parameters, which is impractical in some real-world applications due to the large storage and the risk of over-fitting. This raises a natural question. Can we have an efficient neural network model which 1) exhausts the multi-scale features with 2) better performance and 3) fewer parameters?

In this work, we give an affirmative answer to the above question. As illustrated, the cascading U-Nets architecture can be regarded as a nice starting point to design an effective network for image denoising. However, we need to pay attention to three important related issues. First, we need to smartly address the multi-scale information to keep more structures in the target image. Second, the trade-off between the efficiency and the number of parameters needs to be carefully balanced. Generally speaking, the networks with fewer parameters usually have better generalization abilities. Third, we use suitable techniques to further improve model performance effectively.

Based on the above observations, we now present a novel Dense Dense U-Net model (DDUNet), where the first “dense” refers to the multi-scale dense processing which connects every level in the dense U-Net blocks while the second “dense” refers to the local residual dense block used in the dense U-Net block. In multi-scale dense processing, dense connections are applied to connect every subnetwork of the U-Nets at every level. This is helpful to exploit more information from the feature maps in different scales rather than upsampling only and utilize the gradient flow. Furthermore, we develop a series of important techniques to improve model performance with fewer parameters, and these techniques should be inspiring for other related tasks. Indeed, to reduce the number of parameters, we keep the number of channels unchanged after downsampling and the experimental results show that the number of parameters is about 1/17 of doubling the number of channels while the performance drops slightly only. Moreover, we use discrete wavelet transform as our up- and downsampling operators for its ability to capture both frequency and location information of feature maps, which is helpful in preserving detailed texture [10, 11]. We use a modified version of the local residual dense block in the dense U-Net blocks to further enrich feature representation and reduce the computational burden. As an important U-Net variation, the number of parameters of our DDUNet is only 2/5 of that of MWCNN [32], see Figure 1, and our DDUNet outperforms the state-of-the-art algorithms in gray-scale image denoising, color image denoising and real noisy image denoising. Our research also benefits the model selection in neural architecture search [13]. To sum up, the main contributions of this work are as follows

- We propose a Dense Dense U-Net (DDUNet) for im-

age denoising with a multi-scale dense processing (MDP) mechanism for U-Nets, which fully utilizes the location and frequency information at every level. To the best of our knowledge, the multilevel dense connection did not appear in any previous work, and could be useful for other deep learning tasks and neural architecture search.

- We give a better trade-off between the efficiency and the number of parameters by our MDP. Experimental results show clearly that our network achieves state-of-the-art results in both gray-scale and color image denoising with fewer parameters.
- We develop a series of important techniques to further improve the model performance. As a result, our model achieves the best visualization results in real noisy image denoising. This shows the potential to use our model in real-world applications. Moreover, these techniques should be inspiring for other related tasks.

2. Related Works

2.1. Image Denoising

Image denoising aims to reconstruct a clear image x from a noisy image y . The degradation model is often formulated as

$$y = x + n, \quad (1)$$

where n is the additive white Gaussian noise (AWGN). Nowadays, CNN-based image denoising methods convert this problem into a learning problem.

Since 2009, many excellent CNN models have been invented for image denoising [21]. Early methods generally did not achieve state-of-the-art performance [21, 55]. To enhance the performance, many techniques, such as batch normalization and residual learning [3, 49, 58], dilated filters [59], skip connections in RED30 [34], memory-persistent gate unit in MemNet [46], fractional optimal control [22], residual dense blocks in RDN [62], attention mechanism [48], edge feature guidance mechanism [7, 16], graph convolution in GCDN [52, 53], have been invented for image denoising. Some methods [19, 32, 33, 38, 57] adopted hierarchical structure to enlarge the receptive field for better details preserving. To avoid training the network every time for every single noise level, FFDNet utilized the noise level map as an additional input to the network [60]. Apart from CNN, recurrent neural network (RNN) also plays a major role in image denoising. For instance, the NLRN [28] is an RNN model with non-local modules for image restoration. The non-local self-similarity information was also exploited in N³Net [39], which was inspired by the idea of nearest neighbors. For a contemporary overview of image denoising, one may refer to [15, 18, 47].

Beside image denoising, CNN denoisers have been widely incorporated into the model-based optimizations such as SISR and deblurring [59, 41]. Meanwhile, they have been also used as a pre-processing step in other deep-learning-based models such as image classification [29].

2.2. Variations of U-Net

In 2015, Ronneberger *et al.* [42] introduced the celebrated U-Net for 2D biomedical image segmentation. It has been extended to many applications soon, such as 3D image segmentation [37] and image restoration [32, 33, 38]. Researchers have been inventing many methods to improve the architecture of U-Net, such as cascading U-Nets [4, 30, 54], the skip connections [51] in UNet++ [64], discrete wavelet transform in MWCNN [32], and the inception-residual and dense connecting convolutional modules [63]. Since processing large images often brings large memory consumptions, invertible learnable up- and downsamplings are thus adopted in iUNets [14] for memory-efficient backpropagation. Unlike previous works, we connect several cascading U-Nets via a multilevel dense connection in our DDUNet, which exploit multi-scale features.

3. Method

In this section, we first give an overview of our DDUNet’s architecture. Next, we present the technical details of our local residual dense block and global residual dense structure. In the end, we give our loss function.

3.1. Network Structure

In Figure 2, we illustrate our proposed DDUNet, which consists of dense U-Net blocks with the global residual dense structure. The input is first sent to a convolution layer with a rectified linear unit (ReLU), and then sent to cascading dense U-Net blocks with multi-scale dense processing. After the dense U-Net blocks, a global feature fusion layer followed by a convolution is applied to predict the residual image. Unless other specified, all convolutions are 3×3 convolutions and the number of dense U-Net blocks is 5.

3.1.1 Dense U-Net Block

Our dense U-Net block, as shown in Figure 3, is a U-Net consisting of several local residual dense blocks. The input is first processed with a fusion layer, which consists of a concatenation layer followed by a convolution layer and a ReLU, to fuse the feature maps from all preceding U-Nets in the same level. Next, it is processed with a local residual dense block and downsampled by the discrete wavelet transform (DWT). After that, we process it with a fusion layer, a local residual dense block, and downsample it twice. After the last downsampling, a fusion layer followed by two local residual dense blocks is applied. Finally, the features

are repeatedly upsampled, fused with the features maps in the same U-Net and then processed with a local residual dense block. Unlike traditional U-Net [42], the number of channels of our dense U-Net blocks remains the same after downsampling to save parameters. In the up- and down-sampling part, we use Haar wavelet transform, which was proposed in [2, 32], for its excellent ability to capture the frequency and location information.

3.1.2 Local Residual Dense Block

To fuse information from all the convolution layers as much as possible, unlike [61], we apply two consecutive convolutions with ReLU for greater function space. Greater non-linearity allows richer choices of representation and thus beneficial to the performance. Moreover, we utilize several techniques including local feature fusion and local residual learning in our local residual dense block, as shown in Figure 4 to facilitate the information and gradient flow.

3.1.3 Global Residual Dense Structure

Apart from applying local residual dense structure in the local residual dense block, we further propose multi-scale dense processing, global feature fusion and global residual learning for better information flow.

Multi-scale dense processing. The idea behind our multi-scale dense processing (MDP) is to create a multi-level dense connection, which aims at fusing information from all levels of all preceding U-Nets to the subsequent U-Net. Direct fusion among feature maps from all previous layers is not practical due to memory and computational limit. Instead, we fuse all the feature maps from the last layer of all preceding U-Nets and the feature maps from the previous level in the same U-Net after DWT. Let $x_{0,0}$ be the input for the first U-Net, $x_{i,k}$ be the output of the i -th level in k -th U-Net for $k \geq 1$, $w_{i-1,n}$ be the output of the local residual dense block from the previous level. The input $y_{i,n}$ for LRDB in the i -th level in the n -th U-Net can be formulated as

$$\begin{cases} y_{0,n} = \text{ReLU}(W_{0,n}[x_{0,0}, x_{0,1}, \dots, x_{0,n-1}]), \\ y_{i,n} = \text{ReLU}(W_{i,n}[DWT(w_{i-1,n}), x_{i,1}, x_{i,2}, \dots, x_{i,n-1}]), \end{cases} \quad (2)$$

for $i = 1, 2, 3$, where $W_{i,n}$ is some convolutions, $[\cdot]$ is concatenation operation. Multi-scale dense processing allows the preceding U-Net directly connect all subsequent U-Nets. This is helpful for both feed-forward nature and local dense feature extraction.

3.2. Loss Function

Denote Θ the total network parameters of DDUNet, and let $F(\mathbf{y}; \Theta)$ be the network output. Let $\{(\mathbf{y}_i, \mathbf{x}_i)\}_{i=1}^N$ be a training set, where \mathbf{y}_i is the i -th input image, \mathbf{x}_i is the

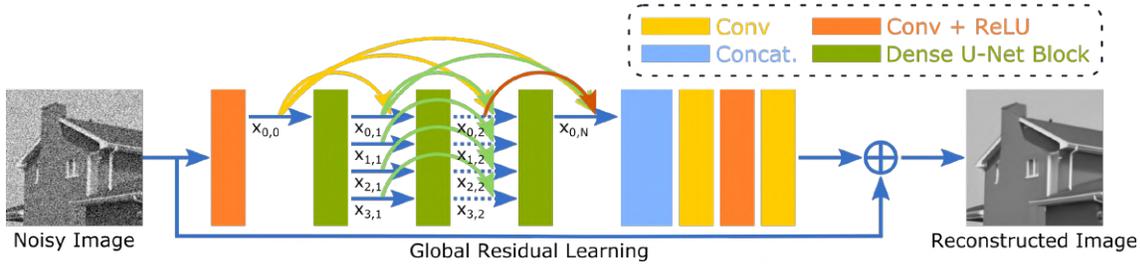


Figure 2: Architecture of Dense Dense U-Net. Figure 3 shows the detailed structure of a dense U-Net block.

Dataset	No. of Param.	Set12			BSD68			Urban100		
		15	25	50	15	25	50	15	25	50
BM3D [9]	-	32.37/.8952	29.97/.8505	26.72/.7676	31.08/.8722	28.57/.8017	25.62/.6869	32.34/.9220	29.70/.8777	25.94/.7791
TNRD [6]	-	32.50/.8962	30.05/.8515	26.82/.7677	31.42/.8822	28.92/.8148	25.97/.7021	31.98/.9187	29.29/.8731	25.71/.7756
DnCNN [58]	556K	32.86/.9027	30.44/.8618	27.18/.7827	31.73/.8906	29.23/.8278	26.23/.7189	32.67/.9250	29.97/.8792	26.28/.7869
IRCNN [59]	186K	32.77/.9008	30.38/.8601	27.14/.7804	31.63/.8881	29.15/.8249	26.19/.7171	32.49/.9244	29.82/.8839	26.14/.7927
RED30 [34]	4.13M	32.83/.9030	30.49/.8637	27.34/.7897	31.72/.8910	29.26/.8300	26.35/.7245	32.75/.9282	30.21/.8901	26.48/.7991
MemNet [46]	677K	32.78/.9015	30.50/.8635	27.38/.7931	31.66/.8883	29.26/.8290	26.35/.7249	32.61/.9257	30.23/.8902	26.64/.8024
FFDNet [60]	485K	32.75/.9027	30.43/.8634	27.32/.7903	31.63/.8902	29.19/.8289	26.29/.7245	32.43/.9273	29.92/.8886	26.52/.8057
MWCNN [32]	24.92M	33.15/.9088	30.79/.8711	27.74/.8056	31.86/.8947	29.41/.8360	26.53/.7366	33.17/.9357	30.66/.9026	27.42/.8371
DIDN [57]	165M	33.14/.9076	30.83/.8706	27.77/.8043	31.85/.8933	29.40/.8332	26.47/.7312	33.32/.9353	30.94/.9045	27.66/.8387
DDUNet (ours)	10.36M	33.23/.9101	30.89/.8728	27.83/.8071	31.94/.8958	29.50/.8379	26.58/.7382	33.48/.9383	31.11/.9087	27.83/.8461

Table 1: Number of parameters and average PSNR(dB)/SSIM results of the competing methods for gray-scale image denoising with noise levels $\sigma = 15, 25$ and 50 on datasets Set12, BSD68 and Urban100. The red and purple color indicate the best and the second-best performance respectively. Our DDUNet has leading performance with reasonable parameters.

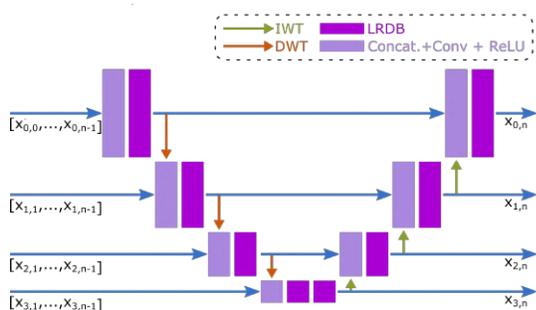


Figure 3: Architecture of a dense U-Net block, where LRDB denotes a local residual dense block shown in Figure 4. DWT and IWT are the discrete wavelet transform and inverse wavelet transform respectively. The number of channels is 64 except possibly the concatenation layers.

ground-truth image. We adopt the common l^2 -loss function defined by

$$\mathcal{L}_\Theta(\mathbf{y}, \mathbf{x}) = \frac{1}{N} \sum_{i=1}^N \|F(\mathbf{y}_i; \Theta) - \mathbf{x}_i\|_2^2 \quad (3)$$

during training. We adopt the ADAM algorithm [24] to minimize the above objective function.

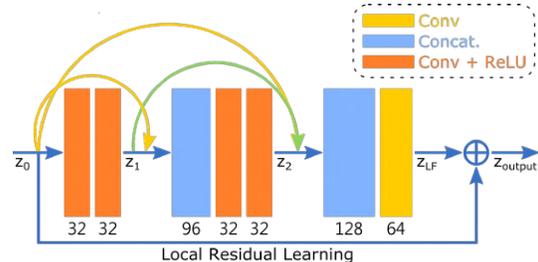


Figure 4: Architecture of a local residual dense block. The numbers of channels of input and output are both 64.

4. Experiments

In this work, we choose the additive white Gaussian noise as our research object due to its extensiveness and practicality. Experiments are conducted for performance evaluation and ablation study.

4.1. Experimental Setting

4.1.1 Datasets

We use a large training set constructed from DIV2K [1]. DIV2K contains 800 images with 2K resolution for training, 100 images for validation, and 100 images for testing. In the

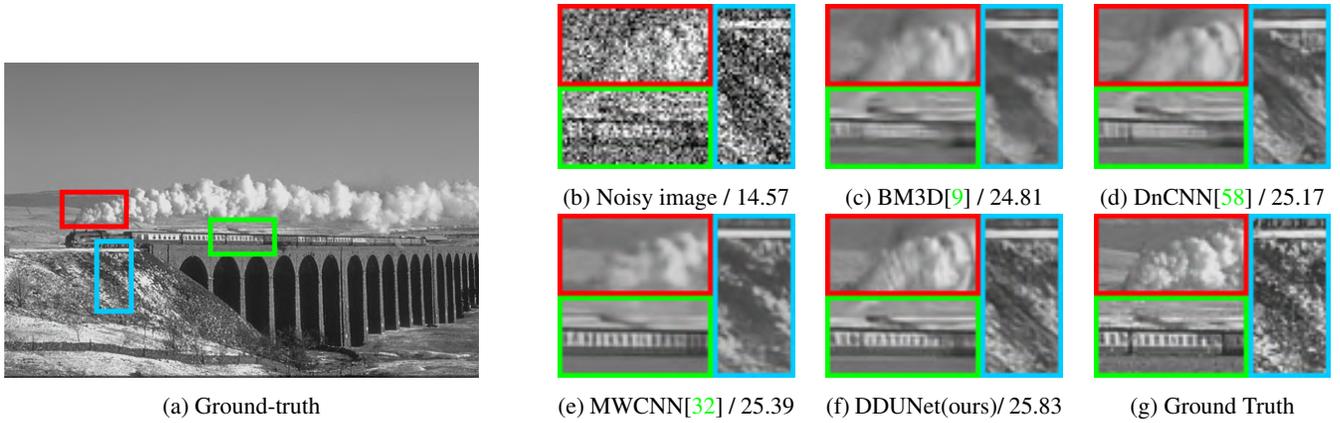


Figure 5: *Gray-scale image denoising* results of “Test033” (BSD68) with noise level of 50 in PSNR (dB). The proposed DDUNet method has higher PSNR value and better details.

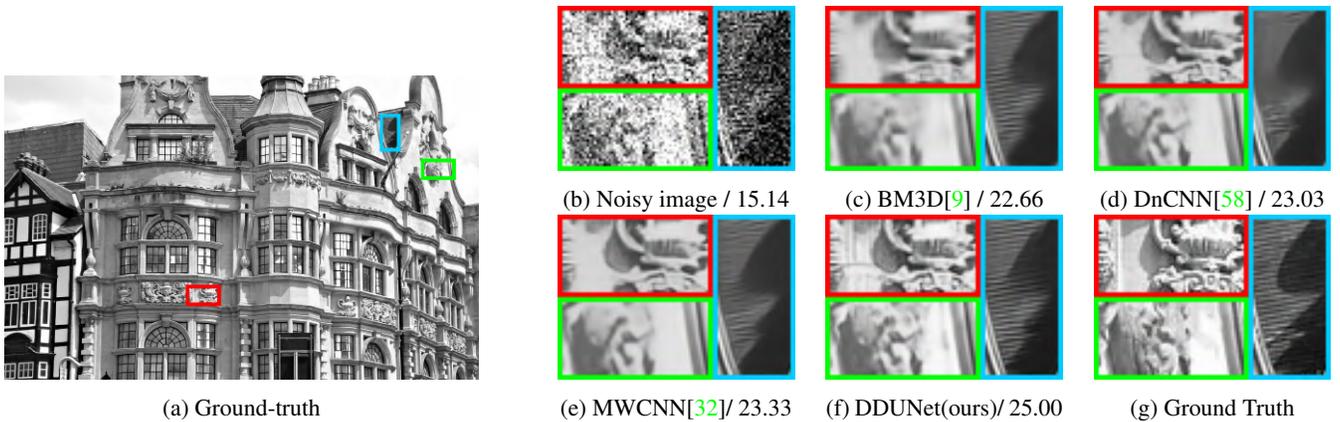


Figure 6: *Gray-scale image denoising* results of “img_053” (Urban100) with noise level of 50 in PSNR (dB). The proposed DDUNet method has higher PSNR value and better details.

training stage $N = 66,884$ patches with a size of 256×256 are cropped from the 800 training images with stride 128.

For gray-scale image denoising, Gaussian noise with a specific noise level is added to a clean patch, and DDUNet is trained to learn a mapping from noisy image to denoising result. We consider three noise levels $\sigma = 15, 25$ and 50 , and evaluate our denoising method on three datasets, including Set12 [58], BSD68 [35], and Urban100 [20].

Similarly, for color image denoising, we still add Gaussian noise with a specific noise level to a clean patch and train our DDUNet to learn the residual map. We consider three noise level $\sigma = 30, 50$ and 70 , and evaluate our denoising method on three datasets, including Kodak24 [17], CBSD68 [35] and Urban100 [20].

Finally, to assess our denoising method on real noisy image denoising, we choose RNI6 and RNI15 [26] as the test datasets, which have no ground truth as reference.

4.1.2 Implementation Details

We adopt the ADAM algorithm [24] with $\beta_1 = 0.9, \beta_2 = 0.999$ and $\epsilon = 10^{-8}$ for optimization and use a mini-batch size of 16. In each batch, we randomly extract 16 patches with a size of 128×128 as input for gray-scale denoising, and $128 \times 128 \times 3$ as input for color image denoising. The learning rate is 2.048×10^{-4} in the first 25 epochs reduce to half every 25 epochs. The total number of epochs is 150. Rotation, flip-based or/and zooming data augmentation [50] is used during mini-batch learning. All experiments are conducted with Nvidia GTX1080Ti GPU.

4.2. Gray-scale Image Denoising

We compare our DDUNet with two classic denoising methods, i.e., BM3D [9] and TNRD [6], and 7 CNN-based methods, i.e., DnCNN [58], IRCNN [59], RED30 [34], MemNet [46], FFDNet [60], MWCNN [32] and DIDN [57]. Table 1 lists the average PSNR/SSIM results of the com-

Dataset σ	No. of Param.	Kodak24			CBSD68			Urban100		
		30	50	70	30	50	70	30	50	70
CBM3D [8]	-	30.89	28.63	27.27	29.73	27.38	26.00	30.36	27.94	26.31
TNRD [6]	-	28.83	27.17	24.94	27.64	25.96	23.83	27.40	25.52	22.63
DnCNN [58]	556K	31.39	29.16	27.64	30.40	28.01	26.56	30.28	28.16	26.17
IRCNN [59]	186K	31.24	28.93	20.65	30.22	27.86	20.61	30.28	27.69	20.69
RED30 [34]	4.13M	29.71	27.62	26.36	28.46	26.35	25.09	29.02	26.40	24.74
MemNet [46]	677K	29.67	27.65	26.40	28.39	26.33	25.08	28.93	26.53	24.93
FFDNet [60]	485K	31.39	29.10	27.68	30.31	27.96	26.53	30.53	28.05	26.39
DIDN [57]	165M	31.97	29.72	28.26	30.71	28.35	26.89	31.70	29.39	27.77
DHDN [38]	168M	31.95	29.67	-	30.41	28.02	-	31.58	29.16	-
DDUNet (ours)	10.36M	32.01	29.80	28.41	30.76	28.45	27.04	31.72	29.50	27.94

Table 2: Number of parameters and average PSNR(dB) results of the competing methods for color image denoising with noise levels $\sigma = 30, 50$ and 70 on datasets Kodak24, CBSD68 and Urban100. The red and purple color indicate the best and the second-best performance respectively. The proposed DDUNet method has higher PSNR values with fewer parameters.

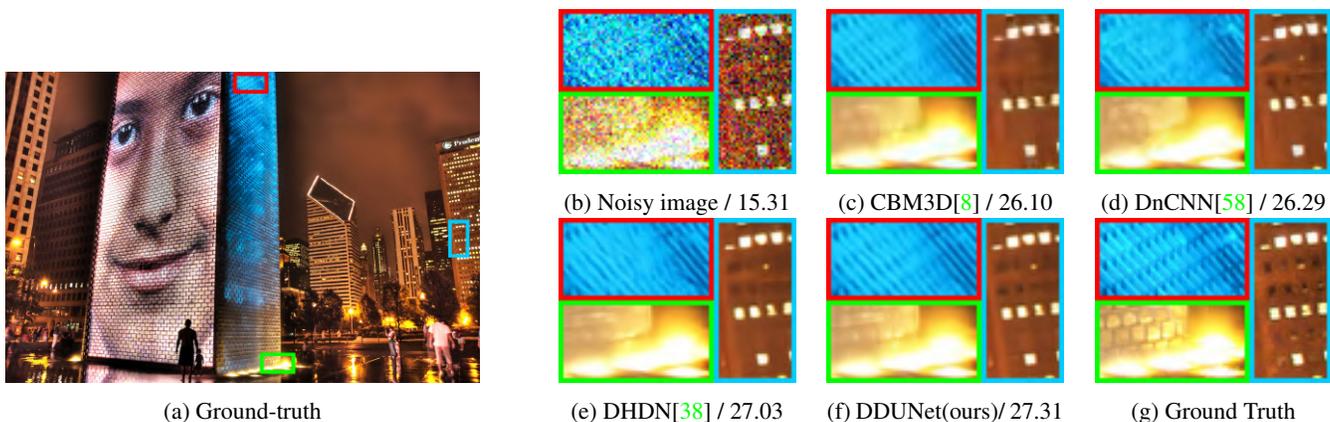


Figure 7: *Color image denoising* results of “img_076” (Urban100) with noise level of 50 in PSNR (dB). The proposed DDUNet method has higher PSNR values and better visual effect.

peting methods on three datasets. For unavailable data, we use the symbol ‘-’. Experimental results show that our DDUNet outperforms state-of-the-art algorithms in all datasets and noise levels in both PSNR and SSIM. In particular, our DDUNet achieves excellent performance on Urban100 when the noise level is high, which is about 0.2 dB higher than DIDN. Since Urban100 has a higher resolution than other datasets, this demonstrates the potential of our DDUNet in the restoration of high-resolution images. Figure 5 and Figure 6 give visual comparisons with BM3D, DnCNN and MWCNN when the noise level is 50. Clearly, our method achieves the best results.

4.3. Color Image Denoising

We compare with two classic denoising methods, i.e., CBM3D [8] and TNRD [6], and 7 CNN-based methods, i.e., DnCNN [58], IRCNN [59], RED30 [34], MemNet [46], FFDNet [60], DIDN [57] and DHDN [38], Table 2 lists the average PSNR results of the competing methods on three datasets. We do not compare the SSIM results since most

of the compared methods did not list their SSIMs in the original literature. For unavailable data, we use the symbol ‘-’. Experimental results show that our DDUNet outperforms state-of-the-art algorithms in all datasets and noise levels. In particular, our DDUNet achieves excellent performance when the noise level is 50 and 70, which is 0.1-0.2 dB higher than the state-of-the-art method. Figure 7 gives a visual comparison with CBM3D, DnCNN, and FFDNet at the noise level 50. Again, our proposed DDUNet method has higher PSNR values and better visual effect.

4.4. Real Noisy Image Denoising

Many reasons cause image noise in reality, such as camera imaging pipelines (shot noise, read noise, and quantization noise), demosaicking, white balance, scanning, and lossy compression [5]. All these different types of noise are usually non-uniform and non-Gaussian. These make the task of real image denoising difficult. To assess the practicability of our DDUNet, we evaluate it on real noisy images. In particular, we choose RNI6 and RNI15 [26] as



Figure 8: Real noisy image denoising. Left to Right: noisy images, denoised images by DnCNN, denoised images by FFDNet, and denoised images by our DDUNet. We set $\sigma = 15$ and $\sigma = 50$ for gray-scale and color images respectively. One can observe that our DDUNet is good at edge recovery and the preservation of the detailed structure.

MDP		✓				✓	✓	✓
GRL			✓		✓		✓	✓
GFF				✓	✓	✓		✓
PSNR	33.26	33.33	33.34	33.34	33.36	33.35	33.35	33.37
SSIM	.9363	.9370	.9371	.9369	.9370	.9370	.9370	.9371
No. of Param.	5.18M	5.62M	5.18M	5.36M	5.36M	5.80M	5.62M	5.80M

Table 3: Ablation investigation of multi-scale dense processing (MDP), global residual learning (GRL), and global feature fusion (GFF). We observe the best performance (PSNR) on gray-scale Urban100 with $\sigma = 15$ in 100 epochs. The red color indicates the best performance.

the test datasets, which have no ground truth as reference. Therefore, the performance can be assessed only by visual comparison. We choose DnCNN and FFDNet for comparison for their wide acceptance as the benchmark for image denoising. In Figure 8, we provide a visual comparison of

these models. Since our DDUNet is a non-blind denoising model, we set $\sigma = 15$ and $\sigma = 50$ for gray-scale and color images respectively by our rough estimation. One can observe that our DDUNet is good at edge recovery and preservation of the detailed structure.

No. of Channels after Downsampling	$\times 1$	$\times 2$
PSNR	33.37	33.52
SSIM	.9371	.9386
No. of Param.	5.80M	102M

Table 4: Ablation investigation of the number of channels after downsampling. We observe the best performance (PSNR) on gray-scale Urban100 with $\sigma = 15$ in 100 epochs, where ' $\times 1$ ' and ' $\times 2$ ' means the number of channels are $(64 \rightarrow 64 \rightarrow 64 \rightarrow 64 \rightarrow 64)$ and $(64 \rightarrow 128 \rightarrow 256 \rightarrow 512 \rightarrow 1024)$ respectively.

No. of U-Nets	2	3	4	5
PSNR	33.29	33.38	33.44	33.48
SSIM	.9365	.9373	.9378	.9383
No. of Param.	3.75M	5.80M	8.01M	10.36M

Table 5: Investigation of the number of U-Net blocks. We observe the best performance (PSNR) on gray-scale Urban100 with $\sigma = 15$ in 100 epochs.

4.5. Ablation Study

4.5.1 Study of the Number of Channels

Table 4 shows the relation between the number of channels and the performance of DDUNet. Both networks have the same number of U-Nets ($N = 3$). Experimental results indicate that doubling the number of channels after downsampling only improves the PSNR for about 0.15 dB but the number of parameters increases 17 times. This justifies our choice of a deeper network instead of a wider one.

4.5.2 Study of Number of U-Nets

In this subsection, we investigate the basic network parameter: the number of U-Nets (denote N for short). Table 5 shows that the performance of our DDUNet increases with the number of U-Nets. Note that the number of parameters increases roughly linearly with the number of U-Nets. Since the improvement decelerates when the number of U-Net blocks attaining $N = 5$ and a single GPU does not have enough memory to train a DDUNet for $N = 6$, we adopt $N = 5$ throughout the whole paper unless other specified.

4.5.3 Study of the Global Residual Dense Structure

Table 3 shows the ablation investigation on the effects of multi-scale dense processing (MDP), global residual learning (GRL) and global feature fusion (GFF). The eight networks have the same number of U-Nets ($N = 3$). Experimental results show that the global residual dense structure is essential for our network. The number of parameters increases only 1/8 after adopting all these dense techniques.

4.6. Number of Parameters

The number of parameters is crucial in determining the potential of a neural network. More parameters usually lead to better performance but it requires larger storage space and often risks over-fitting. In Table 1-5, we compare the number of parameters and the performance in PSNR (dB) under different setting. Experimental results demonstrate that our DDUNet outperforms DIDN [57], DHDN [38] and MWCNN [32] with fewer parameters. Figure 1 gives a visualization of different models' number of parameters against their performance. Note that MemNet [46] is the only parameter-sharing models.

5. Conclusion

The investigation of CNN for image denoising has arrived at serious a bottleneck as models with better performance usually need more parameters in general and their PSNR can only be improved marginally in recent years. It is an extremely challenging task to design an efficient network for image denoising with better performance and fewer parameters. In order to solve this problem, in this work, we propose Dense Dense U-Net (DDUNet) to utilize the hierarchical feature in each scale and a better trade-off between the number of parameters and efficiency. To this end, we keep the number of channels of the U-Net the same after downsampling and connect several dense U-Net by the multilevel dense connection, which directly connects every level cross the U-Nets. To further utilize the global feature, we apply global feature fusion (GFF) and global residual learning (GRL). Inside the dense U-Net, the invertibility of the Haar wavelet transform allows better preservation of the location and frequency information during up- and downsampling. With the usage of local residual dense blocks, the flow of gradient and information has further improvement. By combining local and global features together, our DDUNet surpasses state-of-the-art models in gray-scale, color and real noisy image denoising with fewer parameters. In future work, we aim to extend our DDUNet to other image-to-image tasks like single image super-resolution, image restoration and image segmentation. We will also investigate to incorporate our multi-scale dense processing in neural architecture search.

6. Acknowledgments

This work was supported in part by the National Key R&D Program of China under Grant 2021YFE0203700, in part by NSFC/RGC Grant N_CUHK 415/19, in part by RGC Grant 14300219, 14302920, 14301121, and in part by CUHK Direct Grant for Research under Grant 4053405, 4053460.

References

- [1] Eirikur Agustsson and Radu Timofte. NTIRE 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 126–135, 2017. 4
- [2] Ali N Akansu, Paul A Haddad, Richard A Haddad, and Paul R Haddad. *Multiresolution signal decomposition: transforms, subbands, and wavelets*. Academic Press, 2001. 3
- [3] Saeed Anwar, Cong Phuoc Huynh, and Fatih Porikli. Identity enhanced residual image denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. 2
- [4] Long Bao, Zengli Yang, Shuangquan Wang, Dongwoon Bai, and Jungwon Lee. Real image denoising based on multi-scale residual dense block and cascaded U-Net with block-connection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 448–449, 2020. 1, 3
- [5] Tim Brooks, Ben Mildenhall, Tianfan Xue, Jiawen Chen, Dillon Sharlet, and Jonathan T Barron. Unprocessing images for learned raw denoising. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 11036–11045, 2019. 6
- [6] 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. 1, 4, 5, 6
- [7] Supakorn Chupraphawan and Chotirat Ann Ratanamahatana. Deep convolutional neural network with edge feature for image denoising. In *International Conference on Computing and Information Technology*, pages 169–179. Springer, 2019. 2
- [8] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Color image denoising via sparse 3D collaborative filtering with grouping constraint in luminance-chrominance space. In *2007 IEEE International Conference on Image Processing*, volume 1, pages 1–313. IEEE, 2007. 6
- [9] 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. 1, 4, 5
- [10] I. Daubechies. The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory*, 36(5):961–1005, 1990. 2
- [11] Ingrid Daubechies. *Ten lectures on wavelets*. SIAM, 1992. 2
- [12] Axel Davy, Thibaud Ehret, Jean-Michel Morel, Pablo Arias, and Gabriele Facciolo. A non-local CNN for video denoising. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 2409–2413. IEEE, 2019. 1
- [13] Thomas Elsken, Jan Hendrik Metzen, Frank Hutter, et al. Neural architecture search: A survey. *Journal of Machine Learning Research*, 20(55):1–21, 2019. 2
- [14] Christian Etmann, Rihuan Ke, and Carola-Bibiane Schönlieb. iUNets: Fully invertible U-Nets with learnable up-and downsampling. *arXiv preprint arXiv:2005.05220*, 2020. 3
- [15] Linwei Fan, Fan Zhang, Hui Fan, and Caiming Zhang. Brief review of image denoising techniques. *Visual Computing for Industry, Biomedicine, and Art*, 2(1):7, 2019. 2
- [16] Faming Fang, Juncheng Li, Yiting Yuan, Tiejong Zeng, and Guixu Zhang. Multilevel edge features guided network for image denoising. *IEEE Transactions on Neural Networks and Learning Systems*, 2020. 2
- [17] Rich Franzen. Kodak lossless true color image suite. *source: http://r0k.us/graphics/kodak*, 4(2), 1999. 5
- [18] Bhawna Goyal, Ayush Dogra, Sunil Agrawal, BS Sohi, and Apoorav Sharma. Image denoising review: From classical to state-of-the-art approaches. *Information Fusion*, 55:220–244, 2020. 2
- [19] Shuhang Gu, Yawei Li, Luc Van Gool, and Radu Timofte. Self-guided network for fast image denoising. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2511–2520, 2019. 2
- [20] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5197–5206, 2015. 5
- [21] Viren Jain and Sebastian Seung. Natural image denoising with convolutional networks. In *Advances in Neural Information Processing Systems*, pages 769–776, 2009. 2
- [22] Xixi Jia, Sanyang Liu, Xiangchu Feng, and Lei Zhang. Focnet: A fractional optimal control network for image denoising. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6054–6063, 2019. 2
- [23] Worku Jifara, Feng Jiang, Seungmin Rho, Maowei Cheng, and Shaohui Liu. Medical image denoising using convolutional neural network: a residual learning approach. *The Journal of Supercomputing*, 75(2):704–718, 2019. 1
- [24] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. 4, 5
- [25] Younghee Kwon, Kwang In Kim, James Tompkin, Jin Hyung Kim, and Christian Theobalt. Efficient learning of image super-resolution and compression artifact removal with semi-local Gaussian processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9):1792–1805, 2015. 1
- [26] Marc Lebrun, Miguel Colom, and Jean-Michel Morel. The noise clinic: a blind image denoising algorithm. *Image Processing On Line*, 5:1–54, 2015. 5, 6
- [27] Shutao Li, Haitao Yin, and Leyuan Fang. Group-sparse representation with dictionary learning for medical image denoising and fusion. *IEEE Transactions on Biomedical Engineering*, 59(12):3450–3459, 2012. 1
- [28] Ding Liu, Bihan Wen, Yuchen Fan, Chen Change Loy, and Thomas S Huang. Non-local recurrent network for image restoration. In *Advances in Neural Information Processing Systems*, pages 1673–1682, 2018. 2
- [29] Ding Liu, Bihan Wen, Jianbo Jiao, Xianming Liu, Zhangyang Wang, and Thomas S Huang. Connecting im-

- age denoising and high-level vision tasks via deep learning. *IEEE Transactions on Image Processing*, 29:3695–3706, 2020. [3](#)
- [30] Hongying Liu, Xiongjie Shen, Fanhua Shang, Feihang Ge, and Fei Wang. CU-Net: Cascaded U-Net with loss weighted sampling for brain tumor segmentation. In *Multimodal Brain Image Analysis and Mathematical Foundations of Computational Anatomy*, pages 102–111. Springer, 2019. [3](#)
- [31] Pengju Liu, Hongzhi Zhang, Wei Lian, and Wangmeng Zuo. Multi-level wavelet convolutional neural networks. *IEEE Access*, 7:74973–74985, 2019. [1](#)
- [32] Pengju Liu, Hongzhi Zhang, Kai Zhang, Liang Lin, and Wangmeng Zuo. Multi-level wavelet-CNN for image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 773–782, 2018. [1](#), [2](#), [3](#), [4](#), [5](#), [8](#)
- [33] Wei Liu, Qiong Yan, and Yuzhi Zhao. Densely self-guided wavelet network for image denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 432–433, 2020. [2](#), [3](#)
- [34] Xiaojiao Mao, Chunhua Shen, and Yu-Bin Yang. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In *Advances in Neural Information Processing Systems*, pages 2802–2810, 2016. [2](#), [4](#), [5](#), [6](#)
- [35] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 416–423. IEEE, 2001. [5](#)
- [36] Mario Mastriani and Alberto E Giraldez. Microarrays denoising via smoothing of coefficients in wavelet domain. *International Journal of Electronics and Communication Engineering*, 1(2):386–393, 2007. [1](#)
- [37] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D vision (3DV)*, pages 565–571. IEEE, 2016. [3](#)
- [38] Bumjun Park, Songhyun Yu, and Jechang Jeong. Densely connected hierarchical network for image denoising. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019. [1](#), [2](#), [3](#), [6](#), [8](#)
- [39] Tobias Plötz and Stefan Roth. Neural nearest neighbors networks. In *Advances in Neural Information Processing Systems*, pages 1087–1098, 2018. [2](#)
- [40] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 779–788, 2016. [1](#)
- [41] Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (RED). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844, 2017. [3](#)
- [42] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015. [1](#), [3](#)
- [43] Sameera V Mohd Sagheer and Sudhish N George. A review on medical image denoising algorithms. *Biomedical Signal Processing and Control*, 61:102036, 2020. [1](#)
- [44] Vivek Soni, Ashish Kumar Bhandari, Anil Kumar, and Girish Kumar Singh. Improved sub-band adaptive thresholding function for denoising of satellite image based on evolutionary algorithms. *IET Signal Processing*, 7(8):720–730, 2013. [1](#)
- [45] Pavel Svoboda, Michal Hradis, David Barina, and Pavel Zemečik. Compression artifacts removal using convolutional neural networks. *International Conference in Central Europe on Computer Graphics and Visualization*, 24(2):63–72, 2016. [1](#)
- [46] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu. MemNet: A persistent memory network for image restoration. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4539–4547, 2017. [1](#), [2](#), [4](#), [5](#), [6](#), [8](#)
- [47] Chunwei Tian, Lunke Fei, Wenxian Zheng, Yong Xu, Wangmeng Zuo, and Chia-Wen Lin. Deep learning on image denoising: An overview. *Neural Networks*, 2020. [2](#)
- [48] 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. [2](#)
- [49] Chunwei Tian, Yong Xu, and Wangmeng Zuo. Image denoising using deep CNN with batch renormalization. *Neural Networks*, 121:461–473, 2020. [2](#)
- [50] Radu Timofte, Rasmus Rothe, and Luc Van Gool. Seven ways to improve example-based single image super resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1865–1873, 2016. [5](#)
- [51] Guofeng Tong, Yong Li, Huairong Chen, Qingchun Zhang, and Huiying Jiang. Improved U-Net network for pulmonary nodules segmentation. *Optik*, 174:460–469, 2018. [3](#)
- [52] Diego Valsesia, Giulia Fracastoro, and Enrico Magli. Image denoising with graph-convolutional neural networks. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 2399–2403. IEEE, 2019. [2](#)
- [53] Diego Valsesia, Giulia Fracastoro, and Enrico Magli. Deep graph-convolutional image denoising. *IEEE Transactions on Image Processing*, 29:8226–8237, 2020. [2](#)
- [54] Xide Xia and Brian Kulis. W-net: A deep model for fully unsupervised image segmentation. *Computing Research Repository*, abs/1711.08506, 2017. [3](#)
- [55] Junyuan Xie, Linli Xu, and Enhong Chen. Image denoising and inpainting with deep neural networks. In *Advances in Neural Information Processing Systems*, pages 341–349, 2012. [2](#)
- [56] Xiangyu Xu, Muchen Li, and Wenxiu Sun. Learning deformable kernels for image and video denoising. *Computing Research Repository*, abs/1904.06903, 2019. [1](#)
- [57] Songhyun Yu, Bumjun Park, and Jechang Jeong. Deep iterative down-up CNN for image denoising. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019. [2](#), [4](#), [5](#), [6](#), [8](#)

- [58] 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. [1](#), [2](#), [4](#), [5](#), [6](#), [7](#)
- [59] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep CNN denoiser prior for image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3929–3938, 2017. [2](#), [3](#), [4](#), [5](#), [6](#)
- [60] 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. [2](#), [4](#), [5](#), [6](#), [7](#)
- [61] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2472–2481, 2018. [3](#)
- [62] 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*, 2020. [2](#)
- [63] Ziang Zhang, Chengdong Wu, Sonya Coleman, and Dermot Kerr. Dense-inception U-net for medical image segmentation. *Computer Methods and Programs in Biomedicine*, 192:105395, 2020. [3](#)
- [64] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, 39(6):1856–1867, 2019. [3](#)