HNN: Hierarchical Noise-Deinterlace Net Towards Image Denoising

Amogh Joshi[1] Nikhil Akalwadi[1,3] Chinmayee Mandi[1] Chaitra Desai[1,3]
Ramesh Ashok Tabib[1,2] Ujwala Patil[2] Uma Mudenagudi[1,2]
[1]Center of Excellence in Visual Intelligence (CEVI) [2]School of Electronics and Communication Engineering
[3]School of Computer Science and Engineering
KLE Technological University, Hubballi, Karnataka, INDIA
{joshiamoghmukund, chinmayeemandi2001}@gmail.com
{nikhil.akalwadi, chaitra.desai, ramesh_t, ujwalapatil, uma}@kletech.ac.in

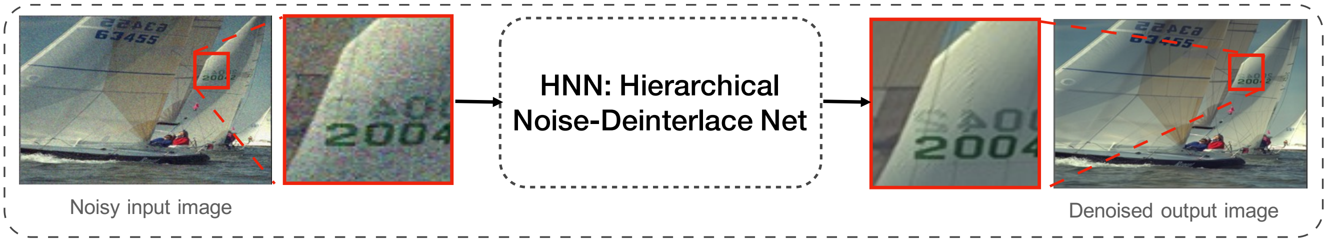


Figure 1. Framework of Hierarchical Noise-Deinterlace Net (HNN) for image denoising. The zoomed-in view (red color) shows specfic area of exemplar image.

Abstract

In this paper, we propose a hierarchical framework for image denoising and term it Hierarchical Noise-Deinterlace Net (HNN). Image denoising techniques aim to recover clean images from noisy observations by reducing unwanted noise and artifacts to enhance the clarity and introduce spatial coherence. Images captured during challenging scenarios suffer from granular noise, inducing fine-scale variations in the image, which occur due to the limitations of imaging technology or environmental conditions. This granular noise can significantly degrade the quality of the image, making it less useful for applications like object detection, image restoration/enhancement, face detection, and image super-resolution. From literature, we infer learning global-local features significantly contribute in reducing unwanted noise and artifacts within images. Typically, researchers rely on residual learning, Generative Adversarial Networks (GANs), and Attention Mechanisms to learn global-local features. However, these methods face challenges such as vanishing gradients, limited generalization of generators in GANs, lack of global context awareness and computation complexity in attention mechanisms leading to drop in performance. Towards this, we propose a hierarchical framework to process both global and local information across distinct levels of hierarchy. More specifically, we propose a hierarchical encoder-decoder network, with a distinct Global-Local Spatio-Contextual (GLSC) block for learning of fine-grained features and high-frequency details in an image. The proposed framework improves image denoising, as it allows the model to capture and utilize information from different scales, ensuring a comprehensive understanding of the image content. We demonstrate the efficacy of proposed HNN framework, on benchmark datasets in comparison with state-of-the-art methods with 5% (↑ in dB) increase in performance.

1. Introduction

In this paper, we propose a hierarchical framework for image denoising and term it Hierarchical Noise-Deinterlace Net (HNN). Images captured in challenging conditions introduce several degradations due to sensors constraints and low-lit scenes resulting in noisy observations. These degradations include noise, blur, camera mis-focus, and inappropriate exposure. For instance, hand-held cameras and smartphone cameras offer smaller apertures and sensors limiting the quality of the capture. Smaller apertures and sensors introduce noise of varying intensities resulting in noisy and unpleasant images. In view of this, image denoising methods aim to restore the clean image from the degraded observations.

Researchers in the domain, address image denoising in two ways: statistical and learning-based methods. Various statistical methods, such as sparse 3D transform-domain collaborative filtering (BM3D) [8], Expected Patch Log Likelihood (EPLL) [48], K-SVD [2], and Nonlocally Centralized Sparse Representations (NCSR) [11] are developed for image denoising. Typically, conventional methods for image denoising fail to retain high-frequency components leading to loss of spatial and contextual information. Towards this, learning-based techniques [30] [15] [47] [43] are proposed for exploiting the learning of spatial and contextual features. However, most learning-based methods incorporate training of CNNs on single/full-scale resolution. The image denoising methods employing single-scale features [40] [45] [10], suffer from limited receptive fields, and fail to capture both global-local structural and contextual information completely. For instance, CNN-based methods like [40] propose using ResNet [14] with batch normalization for image denoising. [41] employs a noise map as a clue towards denoising of image. A few methods also employ encoder-decoder, GAN-based [19] [3], U-Net based strategies for the task of image-denoising. Encoder-decoder and U-Net [26] [7][44] based methods progressively downsample the spatial resolution of the input image to a low resolution and revert to original resolution.

Statistical and single-scale methods suffice to capture global details, however are deficit to capture local spatial and contextual information. Loss of local spatial and contextual features, while reverting the low-resolution representation to the original resolution causes over-smoothening. To overcome this, several methods [6] [35] [46] propose to learn features in multiple scales encompassing global-local spatial and contextual information effectively. From literature, we infer multi-scale features encapsulate finer local contextual information in an image. Authors in, [38] and [6], employ multi-scale feature learning and show significant improvement in quality of denosied image. The key idea of these methods include receptive fields of varying sizes across different scales to extract substantially more information. However, these methods compute features along the same scale and not across the multiple scales. Authors in [46] and [25], propose learning through hierarchical approach for multispectral image analysis. More specifically [46] proves, combining information in the form of residues across different scales help capturing of finer local spatial and contextual information.

With the motivation to capture finer local and spatial contextual information, we propose HNN, a hierarchical network to learn multi-scale features, for simultaneous exchange of information. The hierarchy aids in minimizing the loss of local/fine spatial and contextual information and maintain high-resolution spatial details. HNN extracts shallow features with the hierarchical encoder as shown in Figure 2. The shallow features are passed through the proposed Global-Local Spatio-Contextual (GLSC) block, to generate a set of deep features. Unlike [38], we propose incorporating a dedicated Global-Local Spatio-Contextual (GLSC) block for learning the deep features hierarchically across branches. The deep features are decoded with the hierarchical decoder and are passed through a series of convolution layers to obtain the denoised image of the original resolution.

The main contributions of this work include:

- We propose a hierarchy-based framework (HNN) for image-denoising (Section 2.1).
  - We propose a novel hierarchical feature encoder to obtain features from three distinct scales. (Section 2.2)
  - We propose a Global-Local Spatio-Contextual (GLSC) block for learning of fine-grained features and high-frequency details. (Section 2.3)
- We introduce $\mathcal{L}_{HNN}$ as a weighted combination of $L1$ loss, VGG-19 perceptual loss, and MS-SSIM to exploit local spatial and contextual information across scales while keeping the original resolution of the image intact (Section 2.5).
- We demonstrate the results of image denoising on benchmark, real and synthetic datasets, and compare the performance with SOTA methods using quantitative metrics (Section 3.3).

2. Hierarchical Noise-Deinterlace Net (HNN)

In this section, we propose a novel hierarchical framework (HNN) for image denoising. In Section 2.2, we discuss the mechanism of encoding hierarchical features. In Section 2.3, we focus on the methodology for learning deep features. In Section 2.4, we discuss the methodology for decoding the deep features with hierarchical decoder.

2.1. HNN Framework

The proposed HNN framework includes three main modules *i.e.,* hierarchical feature encoder, Global-Local Spatio-Contextual (GLSC) block, and hierarchical decoder as shown in Figure 2. Typically, image-denoising networks employ feature scaling for varying the sizes of the receptive fields. The varying receptive fields facilitate learning of local-to-global variances in the features. With this motivation, for HNN, we espouse learning contextual information from multi-scale features while preserving high-resolution spatial details. We achieve this via a hierarchical style encoder-decoder network with residual blocks as the backbone for learning.

Given a noisy input image x, the proposed multi-scale hierarchical encoder extracts shallow features in three distinct scales and is given as,

$$F_{si} = M_{E_s}(x) \tag{1}$$

where, F_{si} are the shallow features extracted at i^{th} scale

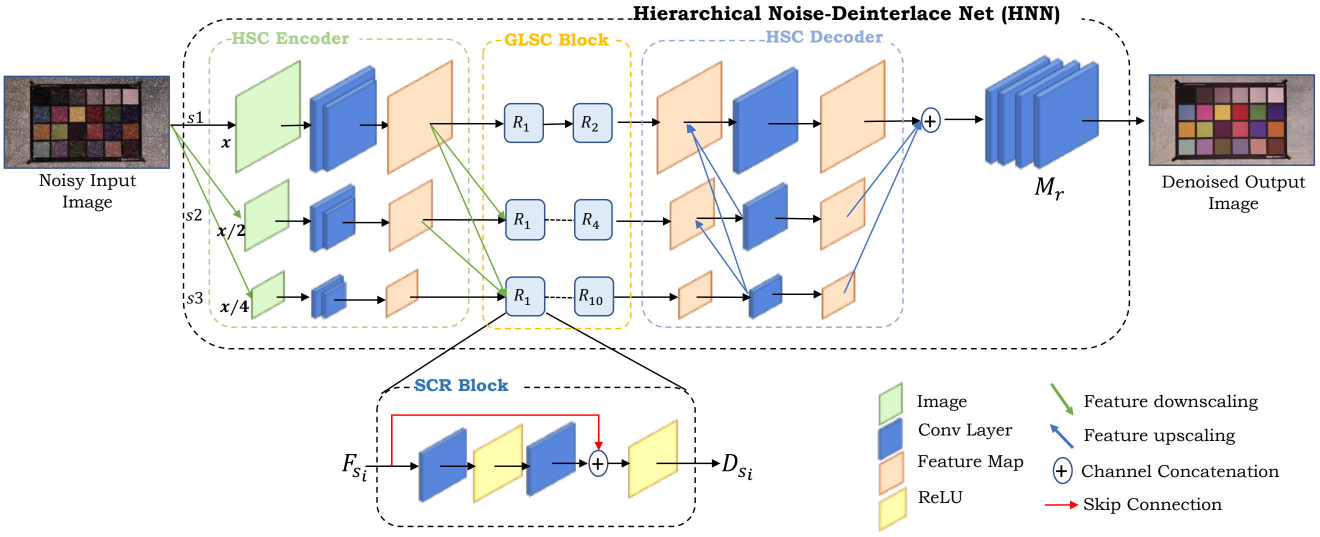


Figure 2. Overview of proposed Hierarchical Noise-Deinterlace Network (HNN). At encoder, we extract the features in three distinct scales, and pass the corresponding information across the hierarchies (as depicted in green color dashed box). We learn fine-grained global-local saptial and contextual information through the proposed GLSC Block (as depicted in orange color dashed box). At decoder, we exchange the information in reverse hierachies (as depicted in blue color dashed box).

from sampled space of input image x and, M_{E_s} represents the hierarchical encoder.

To learn the global-to-local representations from these shallow-level features, we propose Global-Local Spatio-Contextual $(GLSC)$ block, with residual blocks as the backbone. The learnt deep features are respresented as,

$$D_{si} = GLSC_{si}(F_{si}) \tag{2}$$

where D_{si} is the deep feature at i^{th} scale and F_{si} are the shallow features extracted at i^{th} scale and, $GLSC_{si}$ represent residual blocks at respective scales.

We decode the deep features obtained at various scales, with the proposed heirarchical decoder and is given by,

$$d_{si} = M_{d_{si}}(D_{si}) \tag{3}$$

where D_{si} is the deep feature at i^{th} scale and d_{si} is decoded feature at i^{th} scale and, $M_{d_{si}}$ represents the hierarchical decoder.

The decoded features and upscaled features at each scale, are passed to the reconstruction layers M_r to obtain the denoised image $\hat{y}$. The upscaled features from each scale are stacked, and represented as,

$$\mathbb{P} = d_{s1} + d_{s2} + d_{s3} \tag{4}$$

where, d_{s1}, d_{s2} and, d_{s3} are decoded features at three distinct scales, $\mathbb{P}$ represents the final set of features pased to reconstruction layers to obtain the denoised image $\hat{y}$.

$$\hat{y} = M_r(\mathbb{P}) \tag{5}$$

where, $\hat{y}$ is the denoised image obtained from reconstruction layers M_r.

We optimize the learning with the proposed $\mathcal{L}_{HNN}$. $\mathcal{L}_{HNN}$ is a weighted combination of three distinct losses. $L1$ loss to minimize error at pixel level, perceptual loss to efficiently restore contextual information between the ground truth image and the denoised image, and structural dissimilarity loss to restore structural details. The aim is to minimize the weighted combinational loss $\mathcal{L}_{HNN}$ as,

$$L(\theta) = \frac{1}{N} \sum_{i=1}^{N} \|HNN(x_i - y_i)\| \, L_{HNN} \tag{6}$$

where, θ denotes the learnable parameters of the proposed framework, N is the total number of training pairs, x and y are the noisy input and denoised image respectively, and $HNN(\cdot)$ is our proposed framework. $\mathcal{L}_{HNN}$ is the proposed loss function.

2.2. Hierarchical Spatio-Context Encoder (HSCE)

Inspired from visual science, primate visual cortex primarily contributes for processing of visual information. However, in primate visual cortex, the visual information is processed in distinct visual areas at different scales with varying sizes of receptive fields [16] [17] [25] [28]. We replicate the mechanism of processing information in various scales in the proposed hierarchical encoder as shown in Figure 2. The mechanism of hierarchy learns significantly more spatial information [46]. Unlike [46], we incorporate downscaling/upscaling the spatial resolution of the image towards capturing information in distinct scales.

The extracted features at distinct scales is given by:
at scale 1, the features are,

$$f_{s1} = M_{E_{s1}}(x) \tag{7}$$

at scale 2, the features are,

$$f_{s2} = M_{E_{s2}}(\frac{x}{2}) \tag{8}$$

at scale 3, the features are,

$$f_{s3} = M_{E_{s3}}(\frac{x}{4}) \tag{9}$$

where, f_{si} is the shallow feature at i^{th} scale, $M_{E_{si}}(\cdot)$ performs convolution at i^{th} scale on the noisy input image x downscaled by a factor of $\frac{x}{2}$, and $\frac{x}{4}$ respectively.

The multi-scale information is hierarchically exchanged across the scales as shown below,
at hierarchy 1, the features are,

$$F_{s1} = f_{s1} \tag{10}$$

at hierarchy 2, the features are,

$$F_{s2} = f_{s2} + f_{s1} \tag{11}$$

at hierarchy 3, the features are,

$$F_{s3} = f_{s3} + f_{s2} + f_{s1} \tag{12}$$

where, F_{s1}, F_{s2} and F_{s3} are input features to the GLSC Block for learning of deep features as shown in Equations 10, 11, and 12. In Equations 11 and 12, f_{s1} and f_{s2} are the hierarchically exchanged information across scales 2 and 3 respectively.

2.3. Global-Local Spatio-Contextual (GLSC) Block

The features extracted from the hierarchical encoder, are provided to N number of Spatio-Contextual Residual Blocks (SCRB) as shown in Figure 2. The SCRBs yield a set of deep features of varying scales as shown in 13, 14, and 15. The number of SCR Blocks is set to two for $s1$, four for $s2$, and ten for $s3$ (An ablation study is provided in the Section 3.3). We start with two SCRBs for $s1$ as CNNs are known to capture finer spatial details. As we progressively downscale the input image x, we increase the number of SCRBs to maximize learning of global-local spatial and contextual information. The deep features learnt at each scale are represented as,
at scale 1, the deep features are,

$$D_{s1} = GLSC_{s1}(F_{s1}) \tag{13}$$

at scale 2, the deep features are,

$$D_{s2} = GLSC_{s2}(F_{s2}) \tag{14}$$

at scale 3, the deep features are,

$$D_{s3} = GLSC_{s3}(F_{s3}) \tag{15}$$

where, D_{si} is the deep feature at i^{th} scale, $GLSC_{si}$ is the function (GLSC Block) to learn the deep features F_{si}, at respective scales.

2.4. Hierarchical Spatio-Context Decoder (HSCD)

The deep features F_{si}, learnt in the GLSC block are reconciled from multiple scales to match the original resolution. We upscale and concatenate the learnt spatio-contextual features of varying scales, in the reverse order of encoder. The decoded features for each scale are represented as,
at scale 1, the decoded features are,

$$d_{s1} = M_{d_{s1}}(D_{s1} + D_{s2} + D_{s3}) \tag{16}$$

at scale 2, the decoded features are,

$$d_{s2} = M_{d_{s1}}(D_{s2} + D_{s3}) \tag{17}$$

at scale 3, the decoded features are,

$$d_{s3} = M_{d_{s1}}(D_{s3}) \tag{18}$$

where, D_{si} is the decoded feature at i^{th} scale, $M_{d_{si}}$ is a function to decode deep features at i^{th} scale. Here D_{s2} and D_{s3} are the reverse hierarchical information from scales 2 and 3 respectively.

2.5. Loss Function

Unlike [38] and [29], we propose a weighted combinational loss function, comprising of $L1$ loss, MS-SSIM [31], and VGG-16 perceptual loss [20] towards image denoising. The inspiration for the loss function is taken from [9, 12]. However, unlike [9], we set the weights to the loss components experimentally. In tasks like image restoration and enhancement, using $L1$ loss alone produces a sharper and unpleasant image. To overcome the drawback of $L1$ loss, we propose a weighted combinational loss function employing VGG-19 perceptual loss, MS-SSIM and $L1$ loss. We use VGG-19 perceptual loss to efficiently restore features like color and overall structure in the image. MS-SSIM Loss helps in restoring of local and global structural information and $L1$ loss to minimize the generalized error between the pixels.

VGG-19 Perceptual Loss. Spatial and contextual information refers to features like colors/contrast, brightness, and local-global structural details. Spatial and contextual information in an image contributes majorly for image denoising. Towards this, we consider VGG-19 perceptual loss for restoration of spatial and contextual information. VGG-19 perceptual loss is a content-based loss built on the ReLU activation layers of the pre-trained 19-layer VGG network and is given as,

$$\mathcal{L}_{VGG} = \frac{1}{WH}\sum_{i=1}^{W}\sum_{j=1}^{H}(\phi(\hat{y})_{i,j} - \phi(x)_{i,j}) \tag{19}$$

Table 1. Performance comparisons of HNN framework with SOTA methods on CBSD68 [21], Kodak24K [27], and McMaster [42] datasets with varying levels of σ. Cells highlighted in . represents highest values and, cells highlighted in . represents second highest values respectively.

Datasets	**CBSD68** [21]				**Kodak24K** [27]				**McMaster** [42]			
Noise Levels	$\sigma = 5$	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$	$\sigma = 5$	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$	$\sigma = 5$	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$
DnCNN [39] (2017)	23.89	21.86	19.94	18.11	22.31	21.84	19.56	17.40	24.09	22.84	20.11	18.60
CBDNet [13] (2019)	29.21	28.54	26.16	23.19	28.46	26.19	25.16	22.34	29.11	27.49	26.16	23.47
FFDNet [41] (2019)	33.81	32.59	29.66	27.42	32.69	31.09	28.63	25.94	33.94	32.17	30.45	28.44
SADNet [6] (2020)	35.31	34.04	34.10	31.74	36.10	35.44	32.71	31.44	36.40	34.31	32.75	29.97
CycleISP [34] (2020)	37.12	36.48	32.81	29.49	38.45	37.21	35.87	33.01	38.47	36.46	35.02	33.89
DAGL [22] (2021)	34.10	31.69	29.45	27.14	34.49	32.70	31.11	28.94	33.05	31.64	28.74	25.40
HNN (Ours)	**42.91**	**41.63**	**40.22**	**39.79**	**41.55**	**40.48**	**39.58**	**38.94**	**42.16**	**41.03**	**40.29**	**39.82**

where, $\phi(.)$ is the activation of j^{th} layer of network ϕ when processing on image x. W and H are width and height of image.

Multi-scale Structural Similarity Index Measure (MS-SSIM). VGG-19 perceptual loss, emphasises more on restoration of contextual information demanding for improvement in structural information. Towards this, we consider using MS-SSIM as loss, to learn the global and local structural information. MS-SSIM computes SSIM in multiple as shown in the Equation 20.

$$\mathbb{MSSSIM} = l_m(x,y)^{\alpha m} \cdot \textstyle\prod_{j=1}^{m}[c_j(x,y)]^{\beta j}[s_{j(x,y)}]^{\gamma j} \tag{20}$$

where, l_m is luminance at m^{th} scale, c_j is contrast parameter, s_j is structure parameter. β_j and γ_j define relative importance between luminance, contrast and structure parameters.

We formulate $\mathcal{L}_{MSSSIM}$ as,

$$\mathcal{L}_{MSSSIM} = 1 - \mathbb{MSSSIM} \tag{21}$$

We define $\mathcal{L}_{HNN}$ as,

$$\mathcal{L}_{HNN} = (\alpha * L_1) + (\beta * \mathcal{L}_{VGG}) + (\gamma * \mathcal{L}_{MSSSIM}) \tag{22}$$

where, α, β, and γ are the weights. We experimentally set the weights to $\alpha = 0.5$, $\beta = 0.7$, and $\gamma = 0.5$.

3. Results and Discussions

In this section, we evaluate the performance of the proposed HNN both qualitatively and quantitatively (Section 3.3). We compare the results of the proposed HNN with SOTA methods on benchmark datasets.

3.1. Dataset Description

For gaussian (synthetic) image denoising, we use CBSD68 [21], Kodak24K [27], and McMaster [42] as test datasets.

For real image denoising, we utilize the Smartphone Image Denoising Dataset (SIDD) [1], which comprises 320 image pairs for training and 1280 image pairs for validation. For testing, we employ the Darmstadt Noise Dataset (DND) [23], consisting of 50 images. Note: ground-truth images for the DNDataset [23] are not publicly available; reference-based quantitative scores on the DNDataset are obtained using the online submission system [23].

3.2. Implementation Details

We train the proposed HNN using Python (v3.8) and PyTorch framework on 2x NVIDIA RTX 3090 GPUs with AMD Ryzen ThreadRipper 3960X CPU. Our model comprises 16 residual blocks split into three layers, generating 256 feature maps per scale. Training is conducted on SIDD [1] with 600x400 patches using Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$) and a learning rate of $lr = 0.0002$, for 200 epochs, optimizing with proposed $\mathcal{L}_{HNN}$.

3.3. Comparison with SOTA methods

We demonstrate the results of proposed HNN using reference-based quantitative metrics i.e., PSNR and SSIM on SIDD [1] and DND [23] as shown in Figure 3. We quantitatively compare the performance of proposed HNN framework with SOTA methods like DnCNN, MLP, BM3D, CBDNet, DAGL [22], RIDNet, AINDNet, VDN [32], DeamNet, SADNet [6], DANNet, CycleISP [34], MPRNet [36], Restormer [37], MIRNet-v2 [38] as shown in Table 2.

CBSD68 Dataset [21]: Consists of 68 noisy images with $\sigma = (5, 10, 15, 25, 50)$ in each set respectively. We compare the performance of HNN with SOTA methods for varying σ values. In Figure 4, we show the denoising performance of HNN on image *0000.png* with $\sigma = 5$ from CBSD68 Dataset [21]. Results of authors in [6] and [22] show oversmoothening of fine structural details in the exemplar (tail of the airplane). Results from authors in [13] and [34] retain substantially more structral information in comparison with other methods. We observe HNN outperfoms SOTA methods as shown in Figure 4 and consistantly retains fine-grained structural information.

Kodak24K Dataset [27]: Consists of 24 high resolution images. We synthetically add noise to these images with $\sigma = (5, 15, 25, 50)$, and use the same for testing. We

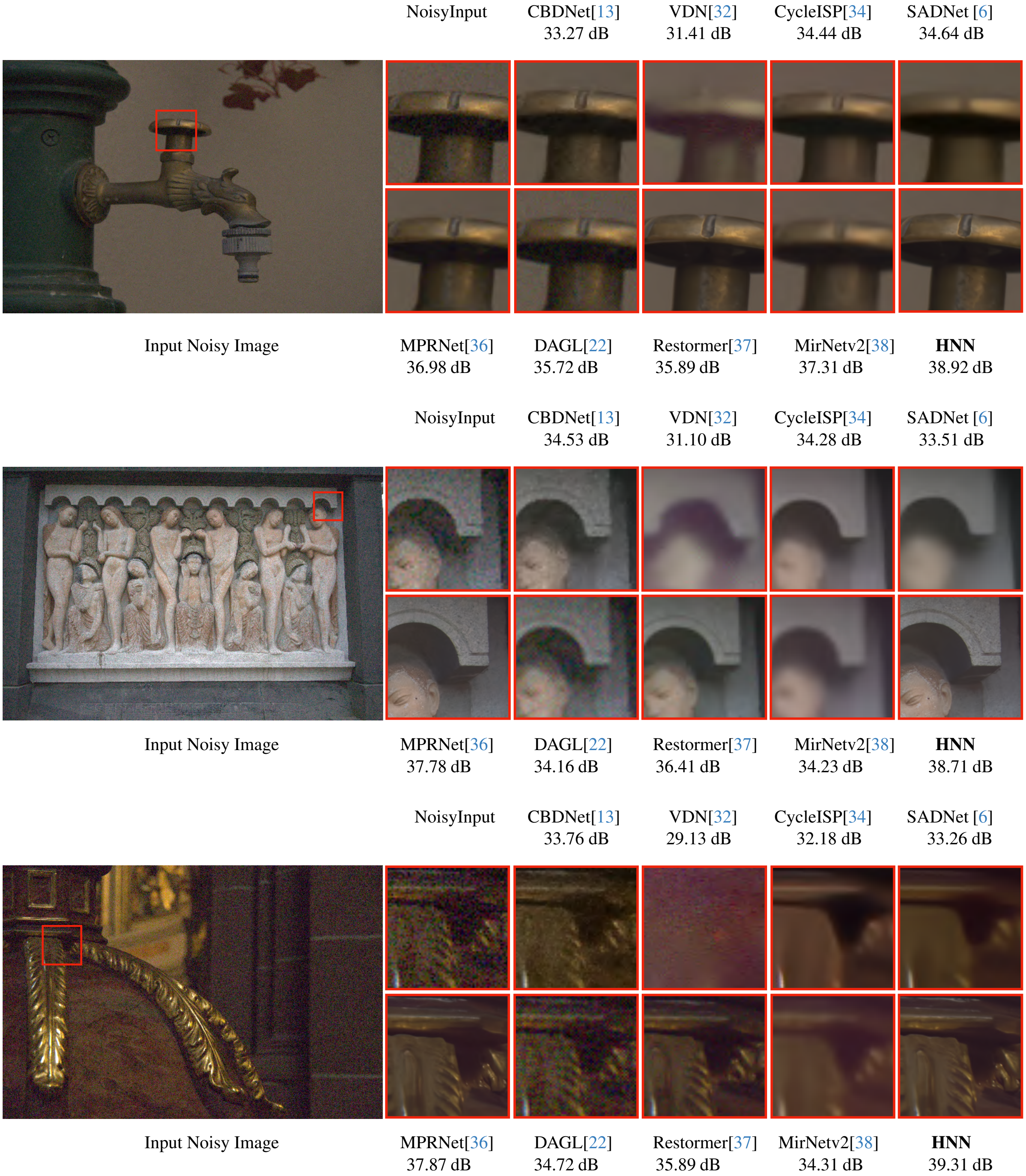


Figure 3. Qualitative comparisons of HNN framework with SOTA methods on DND [23] dataset. The left most image is an input noisy image. The zoomed-in view of the highlighted area is shown in 1^{st} row (Noisy input). We observe, the proposed HNN consistently preserves fine spatial and contextual information in the denoised images.

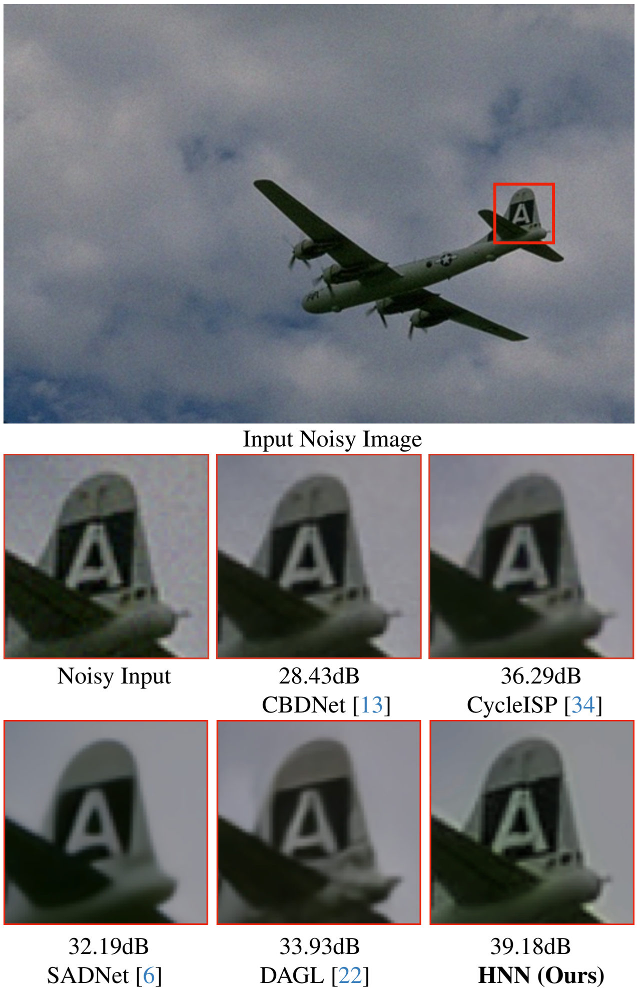
Table 2. Performance comparisons of HNN with SOTA methods on SIDD [1] and DND [23] datasets. † indicates the methods using additional data during training. **Note: the proposed HNN is trained on SIDD [1] dataset and tested on DND [23]** dataset. Cells highlighted in 🟧 . represents highest values and, cells highlighted in 🟦 . represents second highest values respectively.

Method	**SIDD** [1]		**DND** [23]	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑
BM3D [8] (2007)	25.65	0.685	34.51	0.851
MLP [5] (2012)	24.71	0.641	34.23	0.833
DnCNN [39] (2017)	23.66	0.583	32.43	0.790
CBDNet† [13] (2019)	31.25	0.801	38.06	0.942
RIDNet† [4] (2019)	38.74	0.951	39.26	0.953
VDN [32] (2019)	39.28	0.956	39.38	0.952
AINDNet† [18] (2020)	38.74	0.952	39.37	0.951
SADNet† [6] (2020)	38.97	0.957	39.59	0.952
DANet+† [33] (2020)	39.15	0.957	39.58	0.955
CycleISP† [34] (2020)	39.52	0.957	39.56	0.956
DAGL [22] (2021)	38.94	0.953	39.77	0.956
DeamNet† [24] (2021)	38.79	0.957	39.63	0.953
MPRNet [36] (2021)	39.71	0.958	39.80	0.954
MIRNet-v2 [38] (2022)	39.84	0.959	39.86	0.955
HNN (ours)	**43.82**	**0.968**	**41.06**	**0.956**

evaluate the performance of proposed HNN in comparison with SOTA methods using appropriate quantitative metrics as shown in Table 1. Results in [13] show inconsistancy in denoising the image. Results in [34] and [6] lose fine structural details leading to over-smoothening of image. Results from authors in [22] retain minimal structural details. In contrast, the results of proposed HNN consistantly retains fine-grain local structural information (wrinkles within the fabric of the sail) as shown in Figure 5. **McMaster Dataset** [42]: Contains 12 high resolution images. We synthetically add noise to these images with $\sigma = (5, 15, 25, 50)$ and use the same for testing. We evaluate the performance of proposed HNN in comparison with SOTA methods using appropriate quantitative metric as shown in Table 1. Results from authors in [13] and [22] denoise the image to a certain extent. Results from authors in [34] and [6] show over-smoothening and retain limited local structural information. On the contrary, results of proposed HNN retains fine-grained structural information with respect to the embroidery area in the cloth, and outperfoms as shown in Figure 6.

3.4. Ablation study

Influence of hierarchy in the network: Hierarchical connections facilitate information exchange across scales. The proposed HNN with hierarchy exhibits superior performance compared to its absence, enhancing both intra and inter-scale feature correspondence as shown in Table 3.


Input Noisy Image

Noisy Input

28.43dB
CBDNet [13]

36.29dB
CycleISP [34]

32.19dB
SADNet [6]

33.93dB
DAGL [22]

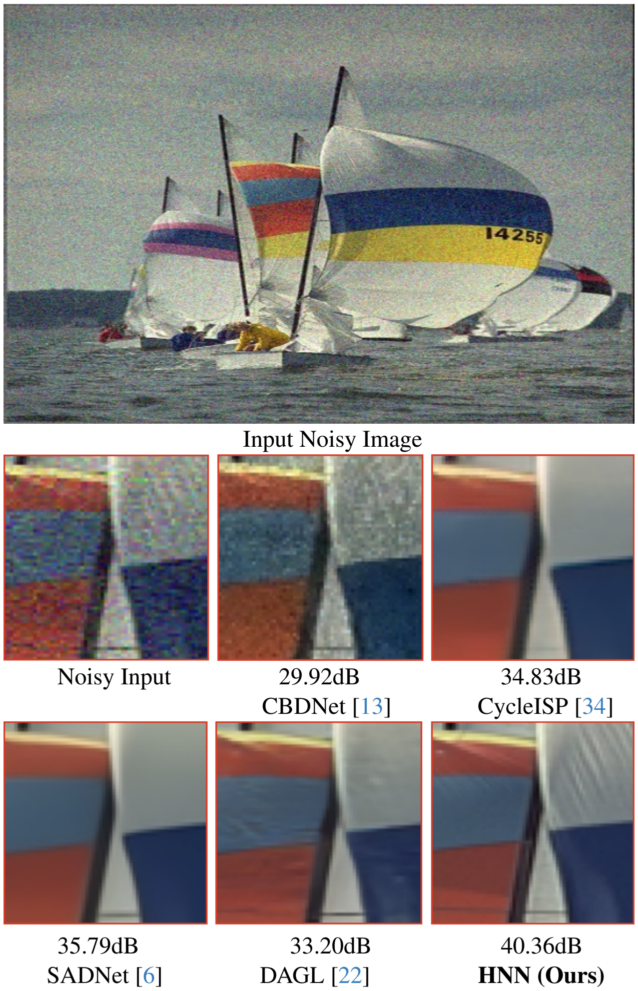
39.18dB
HNN (Ours)

Figure 4. Qualitative comparisons of HNN with SOTA methods on CBSD68 [21] dataset. The above image is an input noisy image. The following images represent comparison of SOTA methods with HNN. We observe, the proposed HNN preserves fine spatial-contextual information in the denoised images.

Table 3. Comparison of proposed HNN framework with and without hierarchical connections using PSNR (in dB) as quantitative metrics on benchmark datasets. We observe proposed HNN shows significant improvement in PSNR (dB) with hierarchy.

Datasets	**With Hierarchy**	**Without Hierarchy**
SIDD [1] (2018)	43.82	37.19
DND [23] (2017)	41.06	36.28
CBSD68 [21] (2001)	42.91	37.10
Kodak24K [27] (2015)	41.55	36.09
McMaster [42] (2011)	42.16	37.32

Influence of $\mathcal{L}_{HNN}$: While $L1$ loss optimizes pixel-to-pixel error with $\alpha = 0.5$, varying β and γ evaluates denoising effects. β sensitivity is observed in local contextual information, whereas γ sensitivity affects fine-grained structural information as shown in Table 4.



Input Noisy Image

Noisy Input

29.92dB
CBDNet [13]

34.83dB
CycleISP [34]

35.79dB
SADNet [6]

33.20dB
DAGL [22]

40.36dB
HNN (Ours)

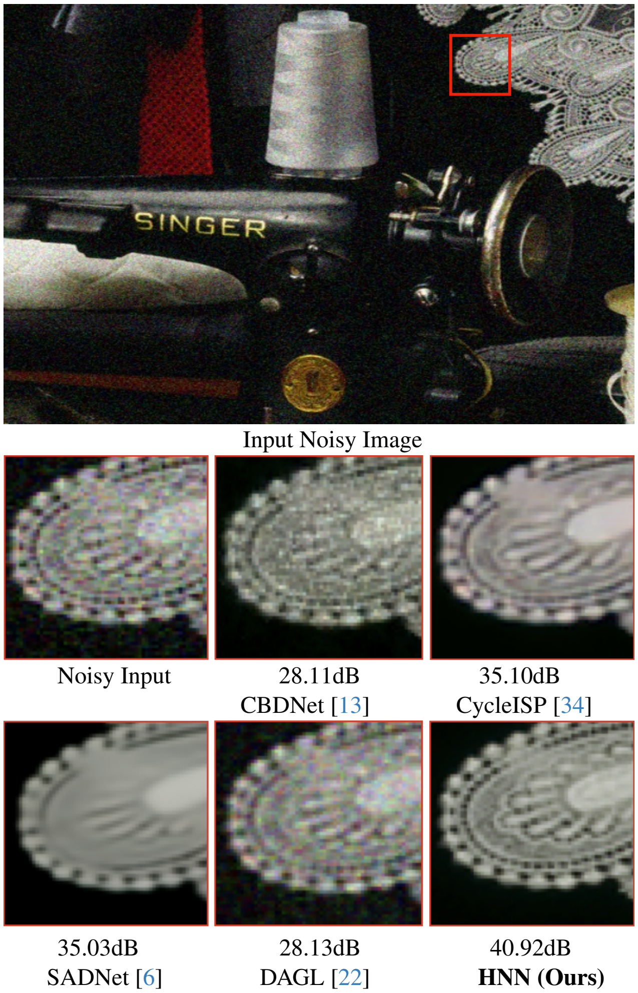
Figure 5. Qualitative comparisons of HNN with SOTA methods on Kodak24K [27] dataset. The above image is an input noisy image. The following images represent comparison of SOTA methods with HNN. We observe, the proposed HNN preserves fine spatial-contextual information in the denoised images.

Table 4. Influence of weights in proposed $\mathcal{L}_{HNN}$ loss. We show the performance comparison of proposed HNN influenced by varying α, β, and γ parameters on benchmark datasets.

	Loss weights			**SIDD**[1]		**DND**[23]	
Experiment	α	β	γ	**PSNR**	**SSIM**	**PSNR**	**SSIM**
1	0.5	0.5	-	39.71	0.863	36.88	0.810
2	0.5	-	0.5	38.25	0.894	37.48	0.849
3	0.5	0.7	0.5	44.12	0.902	39.76	0.948
4 (Ours)	**0.5**	**0.5**	**0.7**	**43.82**	**0.968**	**41.06**	**0.956**

4. Conclusions

In this work, we have proposed HNN, a hierarchy based network towards image denoising. Unlike existing methods, we have proposed exchanging of information arcoss scales through hierarchy improving the learning of fine-grained information. We have proposed $\mathcal{L}_{HNN}$, a weighted combinational loss including $L1$ loss, VGG-19 perceptual loss, and MS-SSIM for learning fine spatio-contextual information. We have demonstrated the performance of the proposed HNN on benchmark datasets and have compared with SOTA methods using appropriate quantitative metrics.



Input Noisy Image

Noisy Input

28.11dB
CBDNet [13]

35.10dB
CycleISP [34]

35.03dB
SADNet [6]

28.13dB
DAGL [22]

40.92dB
HNN (Ours)

Figure 6. Qualitative comparisons of HNN with SOTA methods on McMaster [42] dataset. The above image is an input noisy image. The following images represent comparison of SOTA methods with HNN. We observe, the proposed HNN preserves fine spatial-contextual information in the denoised images.

5. Acknowledgement

This project is partly funded by the Department of Science and Technology (DST) under the ICPS programme, specifically the ”Digital Poompuhar” project (DST/ICPS/Digital Poompuhar/2017).

References

[1] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1692–1700, 2018. 5, 7, 8

[2] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, 54(11):4311–4322, 2006. 2

[3] Abeer Alsaiari, Ridhi Rustagi, Manu Mathew Thomas, Angus G Forbes, et al. Image denoising using a generative adversarial network. In *2019 IEEE 2nd international conference on information and computer technologies (ICICT)*, pages 126–132. IEEE, 2019. 2

[4] Saeed Anwar and Nick Barnes. Real image denoising with feature attention. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3155–3164, 2019. 7

[5] Harold C Burger, Christian J Schuler, and Stefan Harmeling. Image denoising: Can plain neural networks compete with bm3d? In *2012 IEEE conference on computer vision and pattern recognition*, pages 2392–2399. IEEE, 2012. 7

[6] Meng Chang, Qi Li, Huajun Feng, and Zhihai Xu. Spatial-adaptive network for single image denoising. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXX 16*, pages 171–187. Springer, 2020. 2, 5, 6, 7, 8

[7] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3291–3300, 2018. 2

[8] 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. 2, 7

[9] Chaitra Desai, Nikhil Akalwadi, Amogh Joshi, Sampada Malagi, Chinmayee Mandi, Ramesh Ashok Tabib, Ujwala Patil, and Uma Mudenagudi. Lightnet: Generative model for enhancement of low-light images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2231–2240, 2023. 4

[10] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015. 2

[11] Weisheng Dong, Lei Zhang, Guangming Shi, and Xin Li. Nonlocally centralized sparse representation for image restoration. *IEEE transactions on Image Processing*, 22(4): 1620–1630, 2012. 2

[12] Egor Ershov, Alex Savchik, Denis Shepelev, Nikola Banić, Michael S Brown, Radu Timofte, Karlo Koščević, Michael Freeman, Vasily Tesalin, Dmitry Bocharov, et al. Ntire 2022 challenge on night photography rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1287–1300, 2022. 4

[13] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1712–1722, 2019. 5, 6, 7, 8

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2

[15] Jun-Jie Huang and Pier Luigi Dragotti. Winnet: Wavelet-inspired invertible network for image denoising. *IEEE Transactions on Image Processing*, 31:4377–4392, 2022. 2

[16] David H Hubel and Torsten N Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1):106, 1962. 3

[17] Chou P Hung, Gabriel Kreiman, Tomaso Poggio, and James J DiCarlo. Fast readout of object identity from macaque inferior temporal cortex. *Science*, 310(5749):863–866, 2005. 3

[18] Yoonsik Kim, Jae Woong Soh, Gu Yong Park, and Nam Ik Cho. Transfer learning from synthetic to real-noise denoising with adaptive instance normalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3482–3492, 2020. 7

[19] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *The IEEE International Conference on Computer Vision (ICCV)*, 2019. 2

[20] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 4

[21] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, pages 416–423 vol.2, 2001. 5, 7

[22] Chong Mou, Jian Zhang, and Zhuoyuan Wu. Dynamic attentive graph learning for image restoration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4328–4337, 2021. 5, 6, 7, 8

[23] Tobias Plotz and Stefan Roth. Benchmarking denoising algorithms with real photographs. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1586–1595, 2017. 5, 6, 7, 8

[24] Chao Ren, Xiaohai He, Chuncheng Wang, and Zhibo Zhao. Adaptive consistency prior based deep network for image denoising. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8596–8606, 2021. 7

[25] Maximilian Riesenhuber and Tomaso Poggio. Hierarchical models of object recognition in cortex. *Nature neuroscience*, 2(11):1019–1025, 1999. 2, 3

[26] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. pages 234–241, 2015. 2

[27] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115:211–252, 2015. 5, 7, 8

[28] Thomas Serre, Lior Wolf, Stanley Bileschi, Maximilian Riesenhuber, and Tomaso Poggio. Robust object recognition with cortex-like mechanisms. *IEEE transactions on pattern analysis and machine intelligence*, 29(3):411–426, 2007. 3

[29] Hao Shen, Zhong-Qiu Zhao, and Wandi Zhang. Adaptive dynamic filtering network for image denoising. *arXiv preprint arXiv:2211.12051*, 2022. 4

[30] Chunwei Tian, Menghua Zheng, Wangmeng Zuo, Bob Zhang, Yanning Zhang, and David Zhang. Multi-stage image denoising with the wavelet transform. *Pattern Recognition*, 134:109050, 2023. 2

[31] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, pages 1398–1402. Ieee, 2003. 4

[32] Zongsheng Yue, Hongwei Yong, Qian Zhao, Deyu Meng, and Lei Zhang. Variational denoising network: Toward blind noise modeling and removal. *Advances in neural information processing systems*, 32, 2019. 5, 6, 7

[33] Zongsheng Yue, Qian Zhao, Lei Zhang, and Deyu Meng. Dual adversarial network: Toward real-world noise removal and noise generation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part X 16*, pages 41–58. Springer, 2020. 7

[34] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Cycleisp: Real image restoration via improved data synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2696–2705, 2020. 5, 6, 7, 8

[35] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*, pages 492–511. Springer, 2020. 2

[36] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *CVPR*, 2021. 5, 6, 7

[37] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5728–5739, 2022. 5, 6

[38] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for fast image restoration and enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2):1934–1948, 2022. 2, 4, 5, 6, 7

[39] 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. 5, 7

[40] 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. 2

[41] 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, 5

[42] Lei Zhang, Xiaolin Wu, Antoni Buades, and Xin Li. Color demosaicking by local directional interpolation and nonlocal adaptive thresholding. *Journal of Electronic imaging*, 20(2): 023016–023016, 2011. 5, 7, 8

[43] Qi Zhang, Jingyu Xiao, Chunwei Tian, Jerry Chun-Wei Lin, and Shichao Zhang. A robust deformed convolutional neural network (cnn) for image denoising. *CAAI Transactions on Intelligence Technology*, 2022. 2

[44] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM international conference on multimedia*, pages 1632–1640, 2019. 2

[45] 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*, 43(7):2480–2495, 2020. 2

[46] Yuzhi Zhao, Lai-Man Po, Qiong Yan, Wei Liu, and Tingyu Lin. Hierarchical regression network for spectral reconstruction from rgb images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 422–423, 2020. 2, 3

[47] Menghua Zheng, Keyan Zhi, Jiawen Zeng, Chunwei Tian, and Lei You. A hybrid cnn for image denoising. *Journal of Artificial Intelligence and Technology*, 2(3):93–99, 2022. 2

[48] Daniel Zoran and Yair Weiss. From learning models of natural image patches to whole image restoration. In *2011 international conference on computer vision*, pages 479–486. IEEE, 2011. 2