

# Deep unfolding multi-scale regularizer network for image denoising

Jingzhao Xu<sup>1</sup>, Mengke Yuan<sup>2,3</sup>, Dong-Ming Yan<sup>2,3</sup>, and Tieru Wu<sup>1</sup> (✉)

© The Author(s) 2022, corrected publication 2023.

**Abstract** Existing deep unfolding methods unroll an optimization algorithm with a fixed number of steps, and utilize convolutional neural networks (CNNs) to learn data-driven priors. However, their performance is limited for two main reasons. Firstly, priors learned in deep feature space need to be converted to the image space at each iteration step, which limits the depth of CNNs and prevents CNNs from exploiting contextual information. Secondly, existing methods only learn deep priors at the single full-resolution scale, so ignore the benefits of multi-scale context in dealing with high level noise. To address these issues, we explicitly consider the image denoising process in the deep feature space and propose the *deep unfolding multi-scale regularizer network* (DUMRN) for image denoising. The core of DUMRN is the *feature-based denoising module* (FDM) that directly removes noise in the deep feature space. In each FDM, we construct a multi-scale regularizer block to learn deep prior information from multi-resolution features. We build the DUMRN by stacking a sequence of FDMs and train it in an end-to-end manner. Experimental results on synthetic and real-world benchmarks demonstrate that DUMRN performs favorably compared to state-of-the-art methods.

**Keywords** image denoising; deep unfolding network; multi-scale regularizer; deep learning

<sup>1</sup> School of Mathematics, Jilin University, Changchun 130012, China. E-mail: J. Xu, xujz19@mails.jlu.edu.cn; T. Wu, wutr@jlu.edu.cn (✉).

<sup>2</sup> National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China. E-mail: M. Yuan, mengke.yuan@nlpr.ia.ac.cn; D.-M. Yan, yandongming@gmail.com.

<sup>3</sup> School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China.

Manuscript received: 2022-01-12; accepted: 2022-02-22

## 1 Introduction

Image denoising is a fundamental problem in low-level vision since corruption by noise is inevitable during the image acquisition process. Image denoising aims to recover the latent clean image  $x$  from the corresponding noisy observation  $y$ . Mathematically, the degradation model for the denoising problem can be formulated as

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

where  $n$  is generally assumed to be *additive white Gaussian noise* (AWGN) with standard deviation  $\sigma$ . Due to the ill-posed nature of the image denoising problem, many conventional methods use various image prior terms based on the statistics of natural images, including sparse models [1–3], non-local self-similarity models [4–8], low-rank models [9, 10], and Markov random field models [11, 12]. Despite their significant progress, these model-based methods usually reconstruct latent clean images by solving complicated optimization problems, which limits their practical application. Other methods [9, 10, 13] sacrifice flexibility and efficiency to achieve high performance.

With the development of deep *convolutional neural networks* (CNNs), many learning-based methods [14–18] have been proposed for image denoising. With the powerful representation ability of deep CNNs, these methods outperform traditional model-based methods by a large margin. However, most learning-based methods directly learn the mapping between noisy and clean image pairs without considering a physical model of the noise process, which makes them more efficient but less interpretable than model-based methods.

Taking advantages of model-based and learning-based methods, many deep unfolding methods

incorporate standard optimization methods into deep CNNs. They unfold the image denoising problem through various optimization algorithms (e.g., gradient descent [19, 20], alternating direction method of multipliers [21], and primal-dual [22]), and implement the regularization term using deep CNNs, which can implicitly learn deep priors in the feature space. By integrating the image degradation constraint into CNNs, deep unfolding methods maintain the efficiency and improve the interpretability of deep learning. However, the performance of current deep unfolding methods is still limited for two main reasons. Firstly, in each iteration step of deep unfolding methods, noise is removed in the standard image space, but deep priors are learned in feature space. The transformation from deep feature space to image space limits the depth and receptive field of CNNs, which prevents CNNs from extracting non-local dependencies within images. Secondly, existing deep unfolding methods only learn deep priors at the original full-resolution, and thus cannot effectively capture spatial contextual information and restore clear edges for images that suffer from heavy noise [23].

In this paper, we propose the *deep unfolding multi-scale regularizer network* (DUMRN) to more closely integrate model-based and learning-based methods. To reduce the number of space transformations and improve information flow within the network, we explicitly consider the image denoising process in the deep feature space, and propose a *feature-based denoising module* (FDM) based on the image degradation model. By mimicking the gradient descent optimization process, a sequence of FDMs is stacked to build the DUMRN, so that we can obtain a deep CNN with a large receptive field and train it in an end-to-end manner. In each FDM, we construct a *multi-scale regularizer block* (MSRB) to learn deep prior information from multi-resolution features, which is able to capture local details at high resolution and large-scale contextual information at low resolution.

To summarize, the main contributions of this paper are three-fold:

- explicit consideration of the image denoising process in deep feature space, leading to a feature-based denoising module (FDM) based on the iterative optimization steps of the image

degradation model, and a deep unfolding multi-scale regularizer network (DUMRN) produced by stacking a series of FDMs, which can effectively leverage deep features and learn deep prior information,

- a multi-scale regularizer block (MSRB) to learn deep prior information from features of different resolutions, which can capture fine-scale detail information as well as long-range contextual information, and
- comprehensive experiments to demonstrate that DUMRN achieves competitive performance for synthetic and real-world image denoising; DUMRN is still effective and robust on the challenging blind Gaussian denoising task.

## 2 Related work

Image denoising methods can be roughly categorized as traditional model-based methods, deep learning-based methods, and deep unfolding methods. In this section, we briefly review these methods.

### 2.1 Traditional methods

Since image denoising is an ill-posed problem, various regularization or prior terms have been proposed to constrain the solution space. For example, Elad and Aharon [1] enforced sparsity on image patches by constructing highly over-complete dictionaries. Dabov et al. [6] proposed the well-known block-matching and 3D filter (BM3D) method to combine non-local self-similarity with sparsity for image denoising. Chen et al. [13] learned a Gaussian mixture model prior from external image patches and utilized it to find similar patches in input noisy images for denoising. However, hand-crafted priors are not strong enough to characterize complex image structures and usually involve non-convex and time-consuming optimization problems.

### 2.2 Deep learning methods

Motivated by the great success of CNNs in high-level vision tasks, various learning-based methods have been proposed for image denoising. Zhang et al. [14] proposed DnCNN which incorporates residual learning [24] and batch normalization [25] in a CNN to learn residuals between the noisy input and the corresponding clean image. To increase the flexibility of the network to deal with noise of

different levels, Zhang et al. [15] utilized a noise level map as input and performed denoising in the down-sampled sub-image space. Inspired by DenseNet [26], Zhang et al. [17] utilized dense connections to construct a residual dense network (RDN) for image restoration and achieved state-of-the-art results. For real-world image denoising, Guo et al. [27] proposed the convolutional blind denoising network (CBDNet) with a noise estimation subnetwork, while Kim et al. [28] constructed the adaptive instance normalization network (AINDNet) to transfer the Gaussian denoiser to various real noisy scenes. To decrease the computational cost on low-noise images, Yu et al. [29] proposed a multi-path CNN with a dynamic path selection module to adaptively select appropriate routes for different image regions. Recently, several methods [16, 30–32] have employed multi-scale strategies to enlarge the receptive field and improve the performance of deep networks. For example, Chang et al. [16] incorporated multi-size dilated convolutions into a U-Net [33] structure to capture multi-scale contextual information, which helps to restore rich details in complex scenes.

In order to increase the modeling capacity of deep priors, we utilize deep CNNs to learn deep prior information at different scales, which helps to enlarge the receptive field and capture long-range contextual information.

### 2.3 Deep unfolding methods

Deep unfolding methods integrate the advantages of model-based methods (e.g., good interpretability) and learning-based methods (e.g., efficiency and strong representation capability). They usually unfold iterative optimization algorithms as a cascade scheme with a fixed number of steps. Deep CNNs are utilized as regularizers in each step, which implicitly learn deep image priors. Many unfolding methods have been proposed for various image restoration tasks, including image denoising [34–36], image deblurring [19, 20, 37], and image super-resolution [38]. Schmidt and Roth [37] unrolled the half-quadratic optimization procedure into an end-to-end learning framework and proposed a random field-based architecture to learn an image restoration regularizer. Chen and Pock [19] proposed a trainable nonlinear reaction diffusion (TNRD) model by unfolding the gradient descent procedure

to a fixed number of iterations. To integrate sparsity regularization with deep CNNs, Simon and Elad [36] unfolded the convolutional sparse coding model by the learned iterative shrinkage threshold algorithm. By incorporating deep CNNs into a fully parameterized gradient descent scheme, Gong et al. [20] proposed to learn a universal gradient descent optimizer and construct a recurrent gradient descent network (RGDN) for image restoration. Recently, Ren et al. [39] incorporated the adaptive consistency prior into the maximum a posterior framework, and proposed DeamNet based on unfolding the optimization problem. Driven by a large training set, deep unfolding methods optimize the parameters in an end-to-end manner and surpass model-based methods.

Existing deep unfolding methods utilize CNNs to learn data-driven priors in deep feature space, but they remove noise in image space. Thus deep features are transformed into image space at each iteration step, which limits the depth of CNNs and makes it difficult from them to exploit contextual information within images. In order to make full use of deep CNNs, we explicitly consider the image denoising problem in deep feature space and construct a deep network for image denoising.

## 3 Proposed method

In this section, we first propose the feature-based denoising module (FDM) which performs image denoising in deep feature space. Then we introduce the multi-scale regularizer block (MSRB) which learns deep prior information from features at different resolutions. Finally, we stack several FDMs using an unfolding strategy to build the DUMRN for image denoising.

### 3.1 Feature-based denoising module

From a Bayesian perspective, the maximum a posterior (MAP) model for denoising can be formulated as

$$\hat{x} = \operatorname{argmin}_x \frac{1}{2} \|y - x\|_2^2 + \lambda \phi(x) \quad (2)$$

where the first fidelity term guarantees the solution  $\hat{x}$  is in accordance with the degradation process in Eq. (1),  $\phi(x)$  is the regularization term associated with prior information, and  $\lambda$  is the regularization parameter that controls the tradeoff between these terms. Eq. (2) can be solved by various optimization algorithms, such as gradient descent [19, 20],

alternating direction of multipliers [21], and the primal-dual method [22]. Here we utilize the momentum gradient descent method due to its simplicity and effectiveness [40, 41]. Thus,  $\hat{x}$  can be obtained through the iteration in Eqs. (3) and (4):

$$x^{t+1} = x^t - \alpha^t [(x^t - y) - \lambda \nabla \phi^t(x^t)] + \beta^t (x^t - x^{t-1}) \quad (3)$$

$$y = x^0 \quad (4)$$

where  $t$  denotes the step index,  $\alpha^t$  is the step size,  $\nabla \phi^t(\cdot)$  denotes the gradient of  $\phi(\cdot)$ ,  $\beta^t$  is the momentum parameter at step  $t$ , and  $(x^t - x^{t-1})$  is the momentum term. In order to learn deep image priors, deep unfolding methods employ CNNs to calculate the gradient of the regularizer  $\nabla \phi(\cdot)$ , which implicitly provides a deep image prior. Then the iterative procedure in Eq. (3) is unrolled with a fixed number of steps  $T$ . However, in each iteration step, noise in  $x^t$  is removed in the standard image space while the prior  $\nabla \phi(\cdot)$  is learned in deep feature space. The mapping from deep feature space to image space limits the depth and receptive field size of CNNs, making it difficult for unfolding methods to capture non-local dependencies inside images.

In order to make full use of deep CNNs and improve the information flow in the network, we explicitly consider removing noise in deep feature space. Specifically, we first utilize a feature extractor  $f(\cdot)$  to map image  $x$  to the feature space. Then we can use Eq. (3) to iteratively denoise  $f(x)$  without mapping it into image space. In order to scale the gradient and momentum terms adaptively, we replace the step size  $\alpha^t$  and momentum parameter  $\beta^t$  by  $\mathcal{A}^t(\cdot)$  and  $\mathcal{B}^t(\cdot)$  respectively. We let  $\mathcal{S}(\cdot)$  replace  $\nabla \phi(\cdot)$  to supplant the gradient of the regularizer;  $\mathcal{S}(\cdot)$  absorbs the trade-off weight  $\lambda$ .  $\mathcal{S}(\cdot)$  implicitly performs the role of the deep prior. Overall, the  $t$ -th feature-based denoising module (FDM) is formulated as

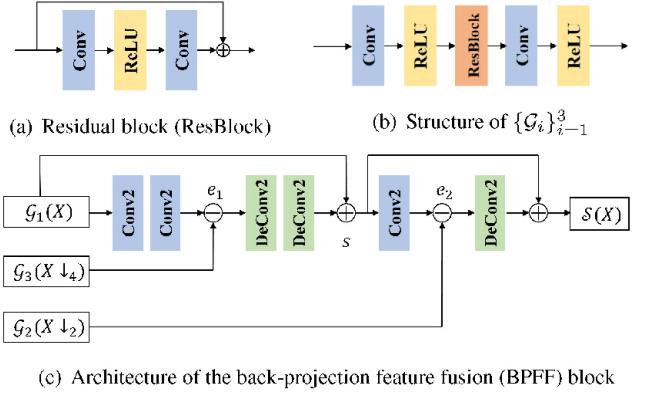
$$\mathbf{X}^{t+1} = \mathbf{X}^t - \mathcal{A}^t(\mathbf{X}^t - \mathbf{Y} + \mathcal{S}^t(\mathbf{X}^t)) + \mathcal{B}^t(\mathbf{X}^t - \mathbf{X}^{t-1}) \quad (5)$$

$$\mathbf{Y} = f(y) = f(x^0) = \mathbf{X}^0 \quad (6)$$

After  $T$  iterations, we utilize an image reconstructor  $g(\cdot)$  to reconstruct the final denoised image from  $\mathbf{X}^T$ :

$$\hat{x} = g(\mathbf{X}^T) \quad (7)$$

Considering the effectiveness of the residual block (ResBlock) in low-level vision tasks [42, 43], we use a single ResBlock to implement  $\mathcal{A}^t(\cdot)$  and  $\mathcal{B}^t(\cdot)$ . The structure of the ResBlock is shown in Fig. 1(a): each



**Fig. 1** (a) Structure of the residual block (ResBlock). (b)  $\{G_i(\cdot)\}_{i=1}^3$  in FDM share a common architecture of a CNN with different parameters to be learned. (c) Structure of the BPFF block. Conv2 and DeConv2 represent a convolutional layer and deconvolutional layer with stride of 2 respectively.

ResBlock contains two  $3 \times 3$  convolutional layers and a ReLU activation function [44]. For flexibility, we implement the feature extractor  $f$  and image reconstructor  $g$  using two convolutional layers with  $3 \times 3$  learnable kernels.

### 3.2 Multi-scale regularizer block

Many unfolding methods [19, 20, 38] implement  $\mathcal{S}(\cdot)$  using deep CNNs due to their strong learning capability. However, limited by the size of the convolution kernel, CNNs fail to capture diverse contextual information, and most existing deep unfolding methods (e.g., TNRD [19] and CSCNet [36]) perform poorly when the noise level is high. To overcome this problem, we adopt a multi-scale strategy by down-sampling features to different scales. On one hand, down-sampling can effectively enlarge the receptive field and enable models to exploit more spatial contextual information, which is helpful in denoising images that suffer from heavy noise [23]. On the other hand, as observed in Ref. [45], noise decreases while strong edges are less affected by down-sampling. To gain the advantages of a multi-scale strategy, we implement  $\mathcal{S}(\cdot)$  using a multi-scale regularizer block (MSRB) to extract useful feature information at multiple resolutions:

$$\mathcal{S}(\mathbf{X}) = f_{\text{MSRB}}(\mathbf{X}) = \mathcal{F}(g_1(\mathbf{X}), g_2(\mathbf{X} \downarrow_2), g_3(\mathbf{X} \downarrow_4)) \quad (8)$$

where  $f_{\text{MSRB}}(\cdot)$  represents the proposed MSRB function,  $\downarrow_k$  represents the down-sampling operator with a scaling factor of  $k$ ,  $\{g_i(\cdot)\}_{i=1}^3$  denotes a set of deep CNNs that learn useful prior information from

the features of different scales, and  $\mathcal{F}(\cdot)$  denotes the multi-scale feature fusion block utilized to aggregate features from multiple scales. In practice, we use three CNNs with the same structure but different channel dimensions to implement  $\{\mathcal{G}_i(\cdot)\}_{i=1}^3$ . The numbers of channels are set to 64, 128, 256 at resolutions 1, 1/2, 1/4 respectively. As shown in Fig. 1(b), each  $\mathcal{G}_i(\cdot)$  consists of two convolutional layers with ReLU activation functions and a ResBlock, and all kernel sizes are set to 3.

Summation and concatenation are the most commonly-used strategies for feature fusion, but directly applying them to multi-scale feature fusion is not effective [46]. Motivated by the success of back-projection technique in image super-resolution [47–49], we implement  $\mathcal{F}(\cdot)$  using a back-projection feature fusion (BPFF) block [46] to aggregate multi-scale features. As shown in Fig. 1(c), we first down-sample  $\mathcal{G}_1(\mathbf{X})$  to the same resolution as  $\mathcal{G}_3(\mathbf{X} \downarrow_4)$ , and compute their difference:

$$e_1 = \mathcal{G}_1(\mathbf{X}) \downarrow_4 - \mathcal{G}_3(\mathbf{X} \downarrow_4) \quad (9)$$

Then we enhance the prior information  $\mathcal{G}_1(\mathbf{X})$  using the back-projected difference:

$$s = \mathcal{G}_1(\mathbf{X}) + (e_1) \uparrow_4 \quad (10)$$

We obtain the final  $\mathcal{S}(\mathbf{X})$  by applying a similar update procedure to integrate  $s$  and  $\mathcal{G}_3(\mathbf{X} \downarrow_4)$ :

$$e_2 = (s) \downarrow_2 - \mathcal{G}_2(\mathbf{X} \downarrow_2) \quad (11)$$

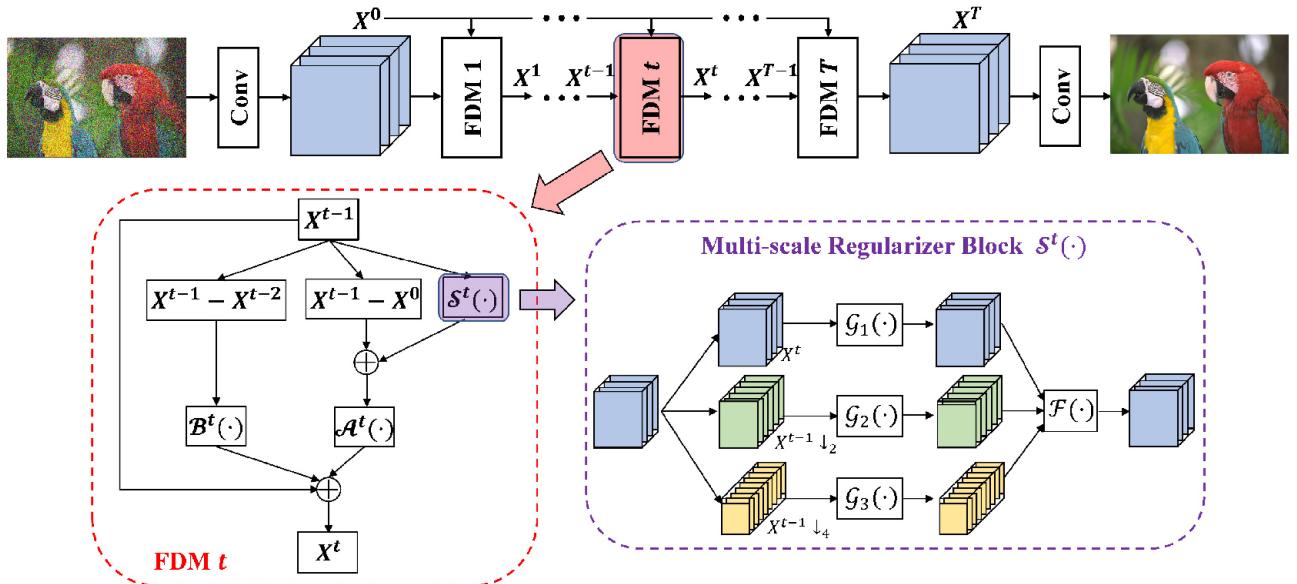
$$\mathcal{S}(\mathbf{X}) = s + (e_2) \uparrow_2 \quad (12)$$

In MSRB, we use convolutional layers and deconvolutional layers to implement down-sampling and up-sampling operators respectively; the strides of both convolutional layers and deconvolutional layers are set to 2. By utilizing the BPFF block in MSRB, we can effectively aggregate prior information at different resolutions into a stronger prior.

### 3.3 Overall architecture

The overall architecture of our proposed deep unfolding multi-scale regularizer network (DUMRN) is shown in Fig. 2. Let  $y$  and  $x$  denote the input noisy image and the corresponding ground-truth image respectively. The feature extractor  $f$  first extracts features  $\mathbf{Y}$  from  $y$ . The extracted features  $\mathbf{Y}$  are set as the initial value  $\mathbf{X}^0$  for the feature denoising process.  $T$  FDMs are stacked to remove noise in feature space: we update  $\mathbf{X}^t$  via  $T$  steps using Eq. (5). Finally, we utilize the image reconstructor  $g$  to reconstruct the final image from the denoised feature  $\mathbf{X}^T$ .

The DUMRN is optimized by minimizing the difference between the input noisy images and the corresponding ground-truth counterparts. To assess the effectiveness of the proposed network, we adopt the same  $L_2$  loss function as previous works [14–



**Fig. 2** Architecture of DUMRN. The first convolutional layer extracts feature  $\mathbf{X}^0$  from the noisy input, and then  $T$  feature-based denoising modules (FDMs) are stacked to remove noise in deep feature space; the structure of FDM is based on the momentum gradient descent algorithm. Benefiting from FDM, there is no need to transform deep features into image space at each step. The last convolutional layer converts the denoised feature  $\mathbf{X}^T$  to a clean image. A multi-scale regularizer block  $\mathcal{S}(\cdot)$  learns deep prior information from features at different resolutions;  $\downarrow_k$  represents down-sampling with a scaling factor of  $k$ .

16, 18]. Given a training dataset  $\{y_i, x_i\}_{i=1}^N$ , where  $N$  is the number of the training patch pairs, we obtain the optimal parameters by minimizing the objective function in Eq. (13):

$$\mathcal{L}(\Theta) = \frac{1}{N} \sum_{i=1}^N \|f_{\text{DUMRN}}(y_i) - x_i\|^2 \quad (13)$$

where  $f_{\text{DUMRN}}(\cdot)$  represents the DUMRN function, and  $\Theta$  represents all learnable parameters in DUMRN.

## 4 Experimental results

In this section, we first provide the training and implementation details of the proposed network. Then, we compare our proposed DUMRN with several state-of-the-art methods on synthetic and real-world image denoising tasks. Our source code is available at <https://github.com/Xujz19/DUMRN>.

### 4.1 Datasets and implementation

#### 4.1.1 Datasets

Following RDN [17] and SADNet [16], we adopt 800 high-resolution training images from the DIV2K dataset [50] to train our models for Gaussian denoising at four different noise levels ( $\sigma = 10, 30, 50, 70$ ). In addition, we use noisy images with varying noise levels  $\sigma \in [0, 75]$  to train a single model for blind Gaussian denoising; it is referred as DUMRN-B. All synthetic Gaussian noisy images were obtained by adding Gaussian noise of different levels to clean images. For real-world image denoising, following DeamNet [39], we trained our model on the SIDD medium dataset [51] and RENOIR dataset [52].

To evaluate DUMRN for gray-scale image denoising, we used three benchmark datasets: Set12 [14], BSD68 [53], and Urban100 [54]. For color image denoising, we chose Kodak24 [55], CBSD68, and Urban100 as test datasets. For real image denoising, we evaluated our model on the SIDD Benchmark dataset [51] and Darmstadt Noise Dataset (DnD) [56]. We used peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [57] as the metrics for quantitative evaluation.

#### 4.1.2 Implementation details

We implemented our proposed model in the PyTorch environment and adopted the ADAM optimizer [62] with default parameters to optimize the network parameters. We trained our network on an NVIDIA

TITAN Xp GPU for a total of 15,000 epochs. For each iteration, we set the batch size to 16 and randomly cropped the noisy and sharp images to size  $96 \times 96$ . Like other denoising methods [14–18, 59], we performed data augmentation on the training images, using random flipping and rotation. We set the initial learning rate to  $10^{-4}$  and halved it every 2500 epochs. All trainable parameters were initialized using the Xavier method [63]. For the trade-off between efficiency and performance, we set the number of FDMs  $T$  to 6. More details of this choice are provided in Section 5.1.

### 4.2 Synthetic noise removal

To evaluate denoising performance on synthetic noisy images, we compared our proposed DUMRN with several state-of-the-art denoising methods including classical model-based methods (BM3D [6] and CBM3D [60]), deep unfolding methods (TNRD[19], CSCNet [36], and DeamNet [39]), and deep-learning based methods (DnCNN [14], FFDNet [15], NLRN[58], SADNet [16], DudeNet [59], COLANet [18], Neb2Neb [61], and RDN [17]).

Tables 1 and 2 summarize quantitative results for different methods on gray-scale and color image denoising, respectively. Our proposed DUMRN achieves the best PSNR under all experiment settings, demonstrating the superiority of our model. Specifically, DUMRN outperforms the most representative traditional method BM3D by 0.74–1.84 dB in different settings. For gray-scale image denoising on the Urban100 dataset with noise level  $\sigma = 50$ , DUMRN achieves 2.11dB/0.0678 and 0.31dB/0.0061 improvements over the previous deep unfolding methods TNRD and DeamNet, respectively. Benefiting from the incorporation of the physical model and deep CNNs, our model also outperforms state-of-the-art deep learning-based methods DnCNN, SADNet, COLANet, and RDN. Taking color image denoising with noise level  $\sigma = 50$  as an example, our DUMRN obtains 0.65dB/0.0207, 0.45dB/0.0161, and 1.37dB/0.0320 improvements over DnCNN on Kodak24, CBSD68, and Urban100 respectively. Due to the proposed FDM and multi-scale strategy that enlarge the receptive field to exploit more spatial contextual information, DUMRN is especially effective when the noise level is high ( $\sigma = 50, 70$ ). Specifically, on the challenging Urban100 dataset, DUMRN outperforms RDN by 0.42dB/0.0175 on

**Table 1** Average PSNR (dB) and SSIM results for different methods on the gray-scale datasets for noise levels  $\sigma = 10, 30, 50$ , and  $70$ . Best results are highlighted in bold. “—” means that a result is unavailable

Method	Set12				BSD68				Urban100			
	10	30	50	70	10	30	50	70	10	30	50	70
BM3D [6]	34.38	29.13	26.72	25.22	33.23	27.74	25.61	24.42	34.47	28.75	25.94	24.27
	0.9233	0.8315	0.7676	0.7174	0.9148	0.7721	0.6855	0.6318	0.9459	0.8567	0.7791	0.7163
TNRD [19]	—	—	26.88	—	—	—	26.01	—	—	—	25.71	—
	—	—	0.7717	—	—	—	0.7057	—	—	—	0.7757	—
CSCNet [36]	—	—	27.15	—	—	—	26.23	—	—	—	26.21	—
	—	—	0.7823	—	—	—	0.7199	—	—	—	0.7957	—
DnCNN [14]	34.78	29.52	27.18	25.52	33.88	28.36	26.23	24.90	34.73	28.88	26.28	24.36
	0.9270	0.8420	0.7827	0.7273	0.9270	0.7999	0.7189	0.6567	0.9486	0.8566	0.7869	0.7178
FFDNet [15]	34.64	29.61	27.32	25.81	33.76	28.39	26.29	25.04	34.45	29.03	26.52	24.86
	0.9271	0.8465	0.7903	0.7451	0.9266	0.8032	0.7245	0.6700	0.9489	0.8707	0.8057	0.7495
NLRN [58]	—	—	27.64	—	—	—	26.47	—	—	—	27.49	—
	—	—	0.7980	—	—	—	0.7298	—	—	—	0.8279	—
SADNet [16]	—	29.94	27.67	26.19	—	28.58	26.49	25.23	—	29.83	27.42	25.82
	—	0.8551	0.8032	0.7622	—	0.8106	0.7340	0.6819	—	0.8876	0.8347	0.7891
DudeNet [59]	34.65	29.54	27.21	—	33.76	28.37	26.25	—	34.41	29.00	26.46	—
	0.9256	0.8430	0.7839	—	0.9251	0.7991	0.7167	—	0.9444	0.8629	0.7945	—
COLANet [18]	34.97	29.81	27.50	25.98	33.93	28.47	26.35	25.10	35.37	30.01	27.47	25.79
	0.9310	0.8486	0.7929	0.7490	0.9284	0.8054	0.7269	0.6704	0.9545	0.8864	0.8289	0.7782
RDN [17]	35.06	29.94	27.60	26.05	33.91	28.54	26.40	25.10	35.42	30.04	27.44	25.69
	0.9315	0.8537	0.7998	0.7558	0.9281	0.8071	0.7275	0.6723	0.9550	0.8886	0.8314	0.7801
DeamNet [39]	—	—	27.74	—	—	—	26.53	—	—	—	27.51	—
	—	—	0.8052	—	—	—	0.7364	—	—	—	0.8374	—
DUMRN	<b>35.12</b>	<b>30.06</b>	<b>27.83</b>	<b>26.33</b>	<b>34.04</b>	<b>28.65</b>	<b>26.56</b>	<b>25.30</b>	<b>35.44</b>	<b>30.19</b>	<b>27.82</b>	<b>26.11</b>
	<b>0.9328</b>	<b>0.8566</b>	<b>0.8068</b>	<b>0.7663</b>	<b>0.9298</b>	<b>0.8120</b>	<b>0.7372</b>	<b>0.6835</b>	<b>0.9555</b>	<b>0.8931</b>	<b>0.8435</b>	<b>0.7976</b>
DUMRN-B	34.94	29.98	27.77	26.32	33.93	28.60	26.52	25.27	35.01	30.00	27.66	26.10
	0.9300	0.8549	0.8047	0.7648	0.9284	0.8107	0.7351	0.6816	0.9519	0.8899	0.8397	0.7957

gray-scale image denoising and  $0.37\text{dB}/0.0112$  on color image denoising when the noise level is  $\sigma = 70$ .

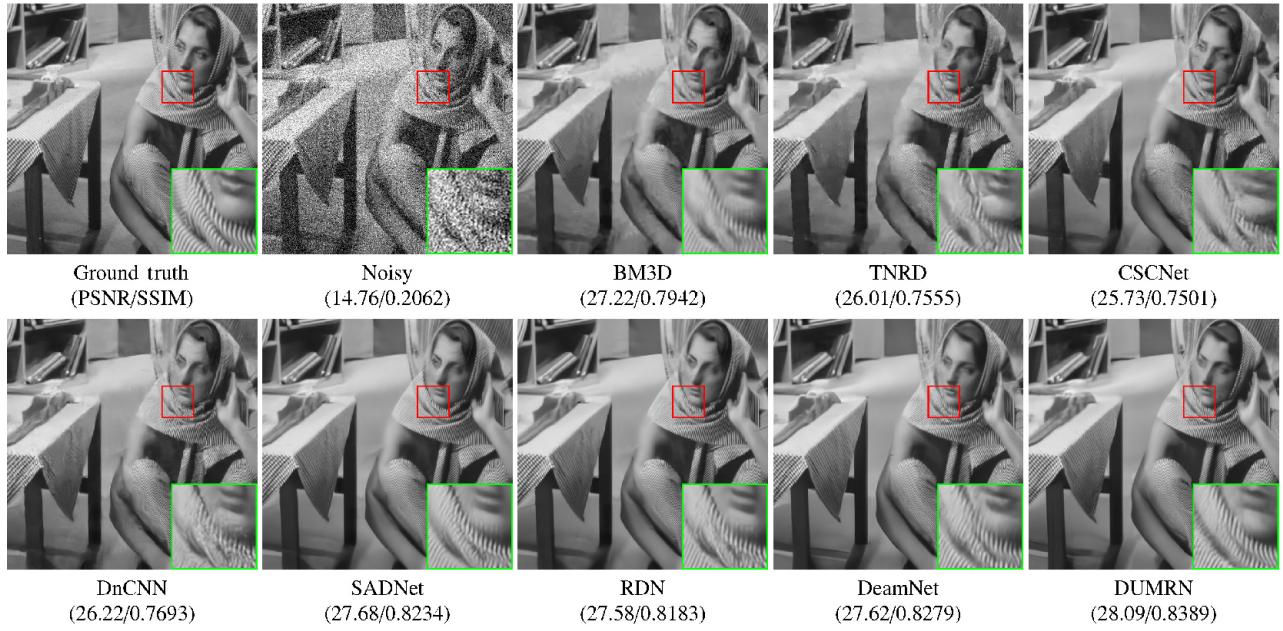
Figure 3 shows a visual comparison of results from different methods on gray-scale image denoising. We can observe that TNRD and CSCNet generate results with severe distortion and noticeable artifacts, and no competing methods recover a clear edge of the chin in the photo, while DUMRN produces more faithful results. Visual comparisons for color images are shown in Figs. 4–6. It can be seen that our DUMRN produces sharper edges and recovers more details while other methods suffer from over-smoothing, demonstrating the powerful representation ability of our DUMRN model.

Quantitative results for our blind denoising model

are also given in Tables 1 and 2. The performance of DUMRN-B slightly decreases due to the unknown noise level, but it is still higher than most deep learning-based methods which are trained for a known specific noise level, indicating that our model is robust in blind image denoising. Taking a noise level  $\sigma = 50$  as an example, DUMRN-B outperforms the non-blind RDN by  $0.14\text{dB}/0.0026$ ,  $0.14\text{dB}/0.0020$ , and  $0.20\text{dB}/0.0056$  on the Kodak24, CBSD68, and Urban100 datasets, respectively. As shown in Figs. 4–6, DUMRN-B also generates more faithful results than the other competing methods, further demonstrating that our model can effectively tackle the challenging blind Gaussian denoising task.

**Table 2** Average PSNR (dB) and SSIM results for different methods on the color datasets for noise levels  $\sigma = 10, 30, 50$ , and  $70$ . Best results are highlighted in bold. “—” means the result is unavailable

Method	Kodak24				CBSD68				Urban100			
	10	30	50	70	10	30	50	70	10	30	50	70
CBM3D [60]	36.57 0.9425	30.89 0.8450	28.63 0.7764	27.27 0.7290	35.91 0.9507	29.73 0.8422	27.38 0.7627	26.00 0.7073	36.00 0.9586	30.36 0.8930	27.94 0.8404	26.31 0.7934
TNRD [19]	— —	— —	27.24 0.7210	— —	— —	— —	26.01 0.7013	— —	— —	— —	25.64 0.7696	— —
DnCNN [14]	36.99 0.9472	31.39 0.8611	29.16 0.7985	27.64 0.7504	36.31 0.9552	30.40 0.8642	28.01 0.7925	26.56 0.7404	36.21 0.9607	30.28 0.8922	28.16 0.8490	26.17 0.7924
FFDNet [15]	36.81 0.9461	31.39 0.8595	29.10 0.7947	27.68 0.7473	36.14 0.9540	30.31 0.8603	27.96 0.7881	26.53 0.7332	35.77 0.9585	30.53 0.8983	28.05 0.8476	26.39 0.8028
SADNet [16]	— —	31.80 0.8707	29.58 0.8140	28.22 0.7718	— —	30.62 0.8697	28.31 0.8041	26.91 0.7540	— —	31.26 0.9108	28.97 0.8712	27.45 0.8368
DudeNet [59]	36.72 0.9452	31.41 0.8605	29.10 0.7952	— —	36.06 0.9531	30.33 0.8616	27.96 0.7893	— —	35.65 0.9545	30.63 0.8962	28.16 0.8454	— —
Neb2Neb [61]	36.17 0.9403	31.24 0.8594	28.93 0.7896	— —	35.49 0.9492	30.22 0.8611	27.86 0.7839	— —	34.85 0.9541	30.22 0.8961	27.82 0.8423	— —
RDN [17]	37.31 0.9498	31.94 0.8714	29.65 0.8132	28.19 0.7659	36.47 0.9565	30.67 0.8695	28.31 0.8024	26.85 0.7481	36.69 0.9635	31.69 0.9147	29.29 0.8748	27.64 0.8377
DUMRN	<b>37.33</b> <b>0.9504</b>	<b>32.02</b> <b>0.8743</b>	<b>29.81</b> <b>0.8192</b>	<b>28.44</b> <b>0.7779</b>	<b>36.50</b> <b>0.9570</b>	<b>30.76</b> <b>0.8725</b>	<b>28.46</b> <b>0.8086</b>	<b>27.06</b> <b>0.7593</b>	<b>36.71</b> <b>0.9640</b>	<b>31.81</b> <b>0.9174</b>	<b>29.53</b> <b>0.9174</b>	<b>28.01</b> <b>0.8810</b>
DUMRN-B	37.24 0.9500	31.99 0.8740	29.79 0.8191	28.44 0.7777	36.42 0.9565	30.73 0.8722	28.45 0.8084	27.06 0.7593	36.44 0.9626	31.71 0.9164	29.49 0.8804	28.01 0.8487



**Fig. 3** Gray-scale image denoising results on image 09 in Set12 with noise level  $\sigma = 50$ .

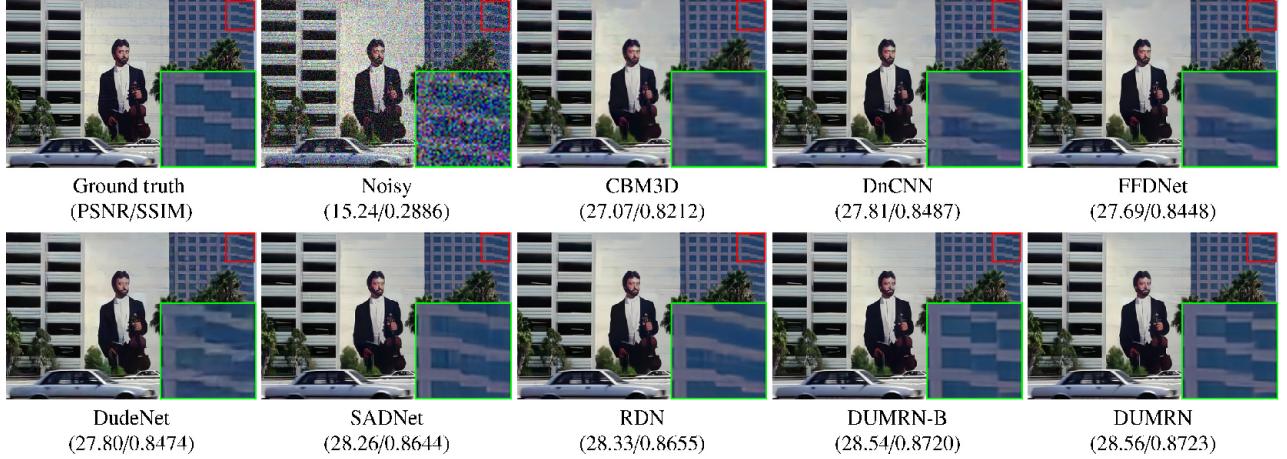
#### 4.3 Real-world noise removal

Actually, it is hard to model real-world noise precisely. Since investigating the degradation processes of real

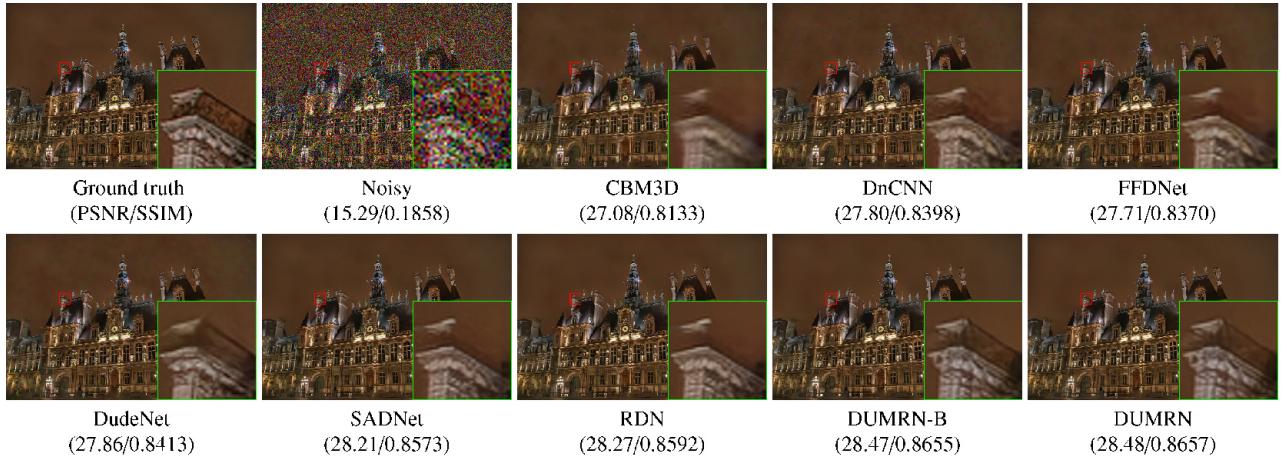
noise is not our main goal, we simply assume it is a kind of additive noise that can be removed by Eq. (2). We compare the DUMRN to several state-of-



**Fig. 4** Color image denoising results on image kodim04 from Kodak24 with noise level  $\sigma = 50$ .



**Fig. 5** Color image denoising results on image 119082 from CBSD68 with noise level  $\sigma = 50$ .



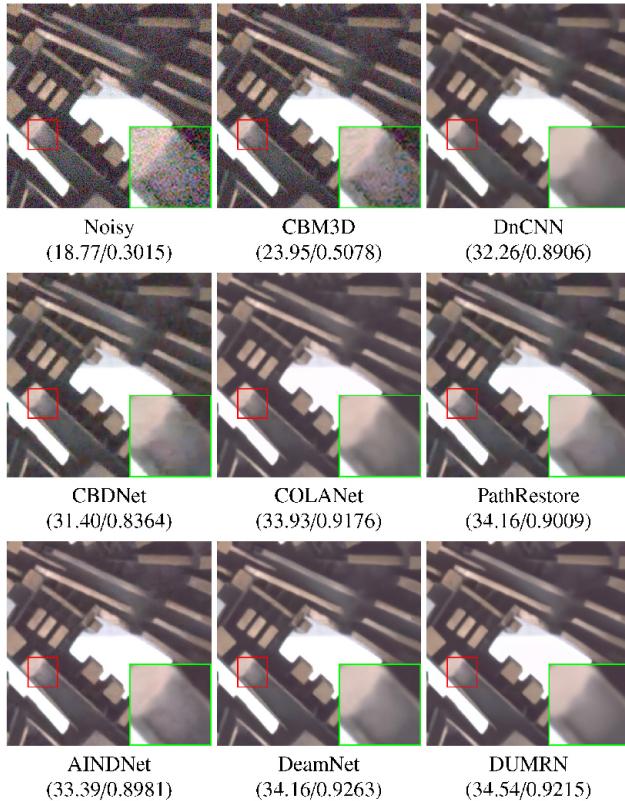
**Fig. 6** Color image denoising results on image img\_022 from Urban100 with noise level  $\sigma = 50$ .

the-art methods, including CBM3D [60], DnCNN[14], PathRestore [29], COLANet [18], R<sup>2</sup>Net [64], and FFDNet [15], CBDNet [27], AINDNet(TF) [28], DeamNet [39].

Quantitative results for different methods on the SIDD benchmark and DnD dataset are provided in Table 3. All PSNR and SSIM results were obtained from public benchmark websites. It can be seen that our DUMRN achieves comparable performance to these competing methods. Visual comparisons of results of different methods on an image from the DnD dataset are shown in Fig. 7. The conventional

**Table 3** Quantitative results for different methods on the SIDD and DnD datasets. Best results are highlighted in bold

Dataset	SIDD		DnD		
	Metric	PSNR	SSIM	PSNR	SSIM
Noisy	23.70	0.480	29.84	0.7018	
CBM3D [6]	25.65	0.685	34.51	0.8507	
DnCNN [14]	23.66	0.583	37.90	0.9430	
FFDNet [15]	29.30	0.694	37.61	0.9415	
CBDNet [27]	33.28	0.868	38.06	0.9421	
AINDNet(TF) [28]	38.95	0.952	39.37	0.9505	
PathRestore[29]	38.21	0.946	39.00	<b>0.9542</b>	
COLANet [18]	38.99	0.951	39.45	0.9626	
R <sup>2</sup> Net [64]	37.87	0.943	39.25	0.9528	
DeamNet [39]	39.35	0.955	39.63	0.9531	
DUMRN	<b>39.44</b>	<b>0.956</b>	<b>39.66</b>	0.9529	



**Fig. 7** Real image denoising results of different methods on the image from DnD dataset.

CBM3D cannot effectively remove real-world noise, and most competing methods (e.g., DnCNN and CBDNet) over-smooth structures while removing noise. In contrast, DUMRN effectively removes noise in constant regions and recovers clear edges with few artifacts.

## 5 Model analysis

In this section, we conduct ablation studies to analyze the effects of different components, including the FDM, MSRB, and consider the unrolling length  $T$ . In addition, we also compare the computational requirements of our DUMRN network and competing methods.

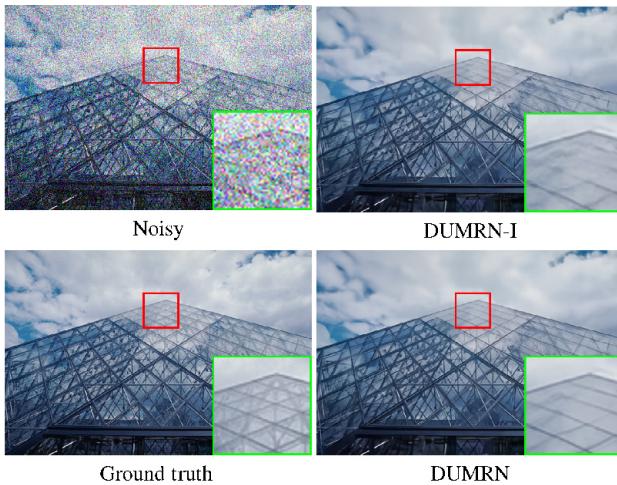
### 5.1 Analysis on FDM

The structure of the proposed DUMRN is similar to TNRD [19] and RGDN [20]; all are based on the gradient descent method. The main difference is that our DUMRN learns a data-driven prior from deep features, but TNRD and RGDN learn priors from the original images. The performance improvement shown in Table 4 of DUMRN over TNRD and RGDN illustrates the superiority of the proposed feature-based denoising framework. To further demonstrate the effect of the deep feature space, we set the feature extractor  $f(\cdot)$  and image reconstructor  $g(\cdot)$  as identity functions (DUMRN-I for short) to learn deep prior information from the raw image space. Table 4 shows that the PSNR of our whole model is 0.19 dB higher than DUMRN-I, which illustrates the importance of the feature-based denoising module. A visual comparison of results in Fig. 8 further demonstrates the effectiveness of the FDM. It can be seen that the proposed DUMRN with FDM generates fewer artifacts and sharper edges than DUMRN-I: FDM can learn more useful features for high-quality image denoising.

To better demonstrate the effect of the proposed FDM intuitively, we visualize some intermediate features  $\mathbf{X}^t$ . Using the visualization method in Ref. [65], we compute the top three principal

**Table 4** Analysis of the proposed MSRB. PSNR (dB) results of color image denoising are evaluated on the CBSD68 dataset with noise level  $\sigma = 50$

Method	TNRD [19]	RGDN [20]	DUMRN-I	DUMRN
PSNR (dB)	25.96	27.66	28.27	28.46



**Fig. 8** Effectiveness of the FDM on image denoising. Visual comparison on image 223061 from CBSD68 with noise level  $\sigma = 50$ .

components of the intermediate features and map them to the principal components of RGB space. The visualized features are shown in Fig. 9. It can be seen that noise is removed progressively from  $\mathbf{X}^2$  to  $\mathbf{X}^6$ . Feature  $\mathbf{X}^6$  contains the least noise and has the sharpest texture: the proposed FDM effectively removes noise and reconstructs details.

The number of FDMs determines the depth of the DUMRN. In Fig. 10, we show the variation in PSNR performance and inference time with increasing numbers of steps  $T$ . PSNR values and inference time both increase as  $T$  rises. As  $T$  increases from 5 to 6, an obvious PSNR improvement is still obtained. However, when  $T$  becomes bigger than 6, the curve in Fig. 10 becomes flatter, and PSNR gains become minor. Thus, we conclude that DUMRN has almost converged when  $T = 6$ . Although DUMRN achieves the best PSNR performance when  $T = 8$ , it takes

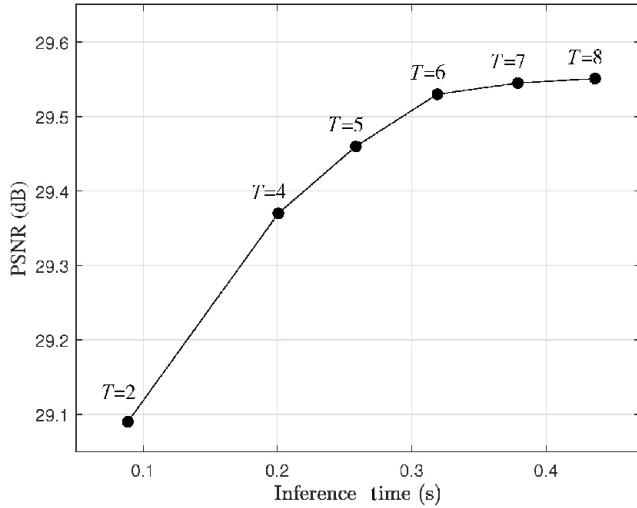
the most inference time. Considering the trade-off between the performance and inference time, we adopt  $T = 6$  in our experiments.

## 5.2 Analysis on MSRB

To investigate the effect of the proposed multi-scale strategy and the adopted BPFF block, we designed several baseline models. Specifically, we trained the following alternatives of our model: (i) removing  $\mathcal{G}_2(\cdot)$  and  $\mathcal{G}_3(\cdot)$  in MSRB (DUMRN-1 for short), (ii) removing  $\mathcal{G}_3(\cdot)$  in MSRB (DUMRN-2 for short), (iii) changing the number of filters in  $\mathcal{G}_2(\cdot)$  and  $\mathcal{G}_3(\cdot)$  to 64 and letting them learn deep prior information at the original full-resolution (DUMRN-3 for short), (iv) replacing BPFF block with concatenation (DUMRN-C for short), (v) replacing the BPFF block with summation (DUMRN-S for short). Quantitative results are shown in Table 5. Compared to DUMRN-1, DUMRN-2 has one more branch that learns deep prior information at the coarse resolution, and DUMRN-3 has two more branches that learn deep prior information at the original full-resolution. It can be seen that DUMRN-2 achieves better denoising performance than DUMRN-3, indicating that using features at different scales can provide more effective prior information. In addition, DUMRN outperforms DUMRN-1, DUMRN-2, and DUMRN-3, which further demonstrates that multi-scale information facilitates image denoising. Visual comparisons are given in Fig. 11; we can observe that a learning deep prior at the original full-resolution (DUMRN-1) is insufficient to recover fine texture details. Taking advantage of the multi-scale prior information, the denoised result of DUMRN contains fewer artifacts and much more detail than the results



**Fig. 9** Visualizations of intermediate features in DUMRN.



**Fig. 10** Varying the number  $T$  of FDMs. PSNR (dB) results for color image denoising are evaluated on the Urban100 dataset with noise level  $\sigma = 50$ .

of DUMRN-1 and DUMRN-2.

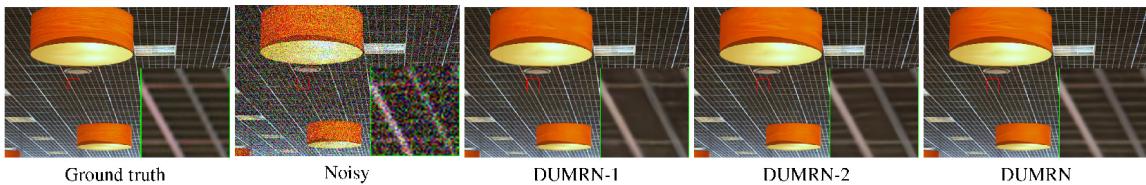
From Table 5, it can also be observed that DUMRN generates better results than DUMRN-C and DUMRN-S: simply using concatenation or summation can not effectively integrate multi-scale information.

### 5.3 Speed

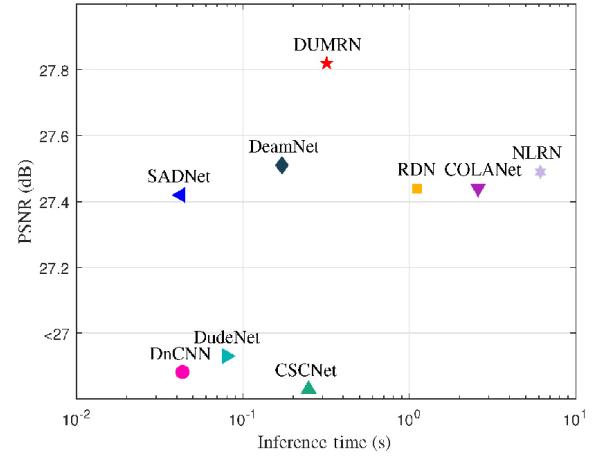
We further evaluated the inference time for different methods for processing a  $512 \times 512$  gray-scale image. All inference time was tested on an NVIDIA TITAN Xp GPU. Figure 12 shows that our DUMRN achieves better PSNR performance with lower inference time than the state-of-the-art methods RDN [17],

**Table 5** Analysis of the proposed MSRB. PSNR (dB) and SSIM results of color image denoising are evaluated on the Urban100 dataset with noise level  $\sigma = 50$

Method	PSNR	SSIM
DUMRN-1	28.81	0.8646
DUMRN-2	29.22	0.8747
DUMRN-3	29.07	0.8710
DUMRN-C	29.41	0.8779
DUMRN-S	29.38	0.8778
DUMRN	29.53	0.8810



**Fig. 11** Effectiveness of the MSRB on image denoising. Visual comparison on image img\_044 from Urban100 with noise level  $\sigma = 50$ .

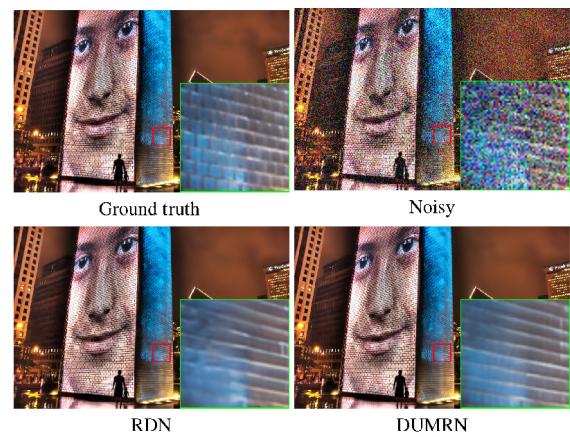


**Fig. 12** Inference time versus PSNR for different methods. PSNR (dB) results for gray-scale image denoising were evaluated on the Urban100 dataset with noise level  $\sigma = 50$ .

COLANet [18], and NLRN [58]. Although DUMRN is a little slower than DnCNN [14], DudeNet [59], and SADNet [16], it achieves much better denoising performance. Overall, DUMRN performs well in terms of both effectiveness and efficiency.

## 6 Limitations

Like other denoising methods, DUMRN may fail to reconstruct proper details in some challenging cases. As Fig. 13 shows, strong noise may make



**Fig. 13** Unsuccessful results. Visual comparison on image img\_076 from Urban100 with noise level  $\sigma = 50$ .

vertical textures imperceptible in the noisy input, and as a result DUMRN and the state-of-the-art RDN cannot correctly recover the vertical textures shown in the close-up. Severe corruption makes it difficult to restore the textures, and DUMRN generates the most likely texture patterns learned from the training dataset.

## 7 Conclusions

In this paper, we proposed the deep unfolding multi-scale regularizer network (DUMRN) to better integrate the traditional image denoising model with deep neural networks. We explicitly consider the image denoising process in deep feature space, and propose a feature-based denoising module (FDM) following the iterative optimization steps of the image degradation model. Benefiting from the FDM, we can construct a deep network with a large receptive field to effectively learn deep prior information. In addition, we proposed the multi-scale regularizer block (MSRB) to extract more spatial contextual information from features of different scales, which is beneficial for images that suffer from heavy noise. We also analyzed the effect of each component in the proposed DUMRN. Extensive experiments demonstrate that our proposed DUMRN can effectively and robustly denoise images, assessed both by quantitative metrics and visual quality. In future, we will investigate real noise models and modify DUMRN to achieve better results for real image denoising. We will also extend our model to other image restoration tasks, such as image deblurring and rain removal.

## Acknowledgements

This work was partially supported by the National Key R&D Program of China (No. 2020YFA0714101) and the National Nature Science Foundation of China (Nos. 61872162, 62102414, 62172415, and 52175493).

## Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

## References

- [1] Elad, M.; Aharon, M. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing* Vol. 15, No. 12, 3736–3745, 2006.
- [2] Mairal, J.; Bach, F.; Ponce, J.; Sapiro, G.; Zisserman, A. Non-local sparse models for image restoration. In: Proceedings of the IEEE 12th International Conference on Computer Vision, 2272–2279, 2019.
- [3] Dong, W. S.; Zhang, L.; Shi, G. M.; Li, X. Nonlocally centralized sparse representation for image restoration. *IEEE Transactions on Image Processing* Vol. 22, No. 4, 1620–1630, 2013.
- [4] Buades, A.; Coll, B.; Morel, J. M. A non-local algorithm for image denoising. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 60–65, 2005.
- [5] Buades, A.; Coll, B.; Morel, J. M. Nonlocal image and movie denoising. *International Journal of Computer Vision* Vol. 76, No. 2, 123–139, 2008.
- [6] Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing* Vol. 16, No. 8, 2080–2095, 2007.
- [7] Wu, X. C.; Zhou, B. Y.; Ren, Q. Y.; Guo, W. Multispectral image denoising using sparse and graph Laplacian Tucker decomposition. *Computational Visual Media* Vol. 6, No. 3, 319–331, 2020.
- [8] Xie Q.; Zhao Q.; Xu Z. B.; Meng D. Y. Color and direction-invariant nonlocal self-similarity prior and its application to color image denoising. *Science China Information Sciences* Vol. 63, Article No. 222101, 2020.
- [9] Dong, W. S.; Shi, G. M.; Li, X. Nonlocal image restoration with bilateral variance estimation: A low-rank approach. *IEEE Transactions on Image Processing* Vol. 22, No. 2, 700–711, 2013.
- [10] Gu, S. H.; Xie, Q.; Meng, D. Y.; Zuo, W. M.; Feng, X. C.; Zhang, L. Weighted nuclear norm minimization and its applications to low level vision. *International Journal of Computer Vision* Vol. 121, No. 2, 183–208, 2017.
- [11] Lan, X. Y.; Roth, S.; Huttenlocher, D.; Black, M. J. Efficient belief propagation with learned higher-order Markov random fields. In: *Computer Vision – ECCV 2006. Lecture Notes in Computer Science, Vol. 3952*. Leonardis, A.; Bischof, H.; Pinz, A. Eds. Springer Berlin Heidelberg, 269–282, 2006.
- [12] Roth, S.; Black, M. J. Fields of experts. *International Journal of Computer Vision* Vol. 82, No. 2, 205–229, 2009.
- [13] Chen, F.; Zhang, L.; Yu, H. M. External patch prior guided internal clustering for image denoising. In: Proceedings of the IEEE International Conference on Computer Vision, 603–611, 2015.

- [14] Zhang, K.; Zuo, W. M.; Chen, Y. J.; Meng, D. Y.; Zhang, L. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing* Vol. 26, No. 7, 3142–3155, 2017.
- [15] Zhang K.; Zuo W. M.; Zhang L. FFDNet: Toward a fast and flexible solution for CNN based image denoising. *IEEE Transactions on Image Processing* Vol. 27, No. 9, 4608–4622, 2018.
- [16] Chang, M.; Li, Q.; Feng, H.; Xu, Z. Spatial-adaptive network for single image denoising. In: *Computer Vision – ECCV 2020. Lecture Notes in Computer Science*, Vol. 12375. Vedaldi, A.; Bischof, H.; Brox, T.; Frahm, J. M. Eds. Springer Cham, 171–187, 2020.
- [17] Zhang, Y. L.; Tian, Y. P.; Kong, Y.; Zhong, B. N.; Fu, Y. Residual dense network for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 43, No. 7, 2480–2495, 2021.
- [18] Mou, C.; Zhang, J.; Fan, X. P.; Liu, H. F.; Wang, R. G. COLA-net: Collaborative attention network for image restoration. *IEEE Transactions on Multimedia* Vol. 24, 1366–1377, 2022.
- [19] Chen, Y. J.; Pock, T. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 39, No. 6, 1256–1272, 2017.
- [20] Gong, D.; Zhang, Z.; Shi, Q. F.; van den Hengel, A.; Shen, C. H.; Zhang, Y. N. Learning deep gradient descent optimization for image deconvolution. *IEEE Transactions on Neural Networks and Learning Systems* Vol. 31, No. 12, 5468–5482, 2020.
- [21] Yang, Y.; Sun, J.; Li, H. B.; Xu, Z. B. Deep ADMM-Net for compressive sensing MRI. In: Proceedings of the 30th International Conference on Neural Information Processing Systems, 10–18, 2016.
- [22] Adler, J.; Öktem, O. Learned primal-dual reconstruction. *IEEE Transactions on Medical Imaging* Vol. 37, No. 6, 1322–1332, 2018.
- [23] Feng, W. S.; Qiao, P.; Xi, X. Y.; Chen, Y. J. Image denoising via multiscale nonlinear diffusion models. *SIAM Journal on Imaging Sciences* Vol. 10, No. 3, 1234–1257, 2017.
- [24] He, K. M.; Zhang, X. Y.; Ren, S. Q.; Sun, J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770–778, 2016.
- [25] Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Proceedings of the 32nd International Conference on International Conference on Machine Learning, Vol. 37, 448–456, 2015.
- [26] Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K. Q. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2261–2269, 2017.
- [27] Guo, S.; Yan, Z. F.; Zhang, K.; Zuo, W. M.; Zhang, L. Toward convolutional blind denoising of real photographs. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 1712–1722, 2019.
- [28] Kim, Y.; Soh, J. W.; Park, G. Y.; Cho, N. I. 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, 3479–3489, 2020.
- [29] Yu, K.; Wang, X. T.; Dong, C.; Tang, X. O.; Loy, C. C. Path-restore: Learning network path selection for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 44, No. 10, 7078–7092, 2022.
- [30] Gu, S. H.; Li, Y. W.; van Gool, L.; Timofte, R. Self-guided network for fast image denoising. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2511–2520, 2019.
- [31] Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; Yang, M. H.; Shao, L. Learning enriched features for real image restoration and enhancement. In: *Computer Vision – ECCV 2020. Lecture Notes in Computer Science*, Vol. 12370. Vedaldi, A.; Bischof, H.; Brox, T.; Frahm, J. M. Eds. Springer Cham, 492–511, 2020.
- [32] Yu, X. J.; Fu, Z. X.; Ge, C. K. A multi-scale generative adversarial network for real-world image denoising. *Signal Image Video Processing* Vol. 16, No. 1, 257–264, 2022.
- [33] Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science*, Vol. 9351. Navab, N.; Hornegger, J.; Wells, W.; Frangi, A. Eds. Springer Cham, 234–241, 2015.
- [34] Barbu, A. Training an active random field for real-time image denoising. *IEEE Transactions on Image Processing* Vol. 18, No. 11, 2451–2462, 2009.
- [35] Sun, J.; Tappen, M. F. Learning non-local range Markov Random field for image restoration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2745–2752, 2011.
- [36] Simon, D.; Elad, M. Rethinking the CSC model for natural images. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems, 2274–2284, 2019.



- [37] Schmidt, U.; Roth, S. Shrinkage fields for effective image restoration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2774–2781, 2014.
- [38] Zhang, K.; van Gool, L.; Timofte, R. Deep unfolding network for image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 3214–3223, 2020.
- [39] Ren, C.; He, X. H.; Wang, C. C.; Zhao, Z. B. Adaptive consistency prior based deep network for image denoising. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 8592–8602, 2021.
- [40] Qian, N. On the momentum term in gradient descent learning algorithms. *Neural Networks* Vol. 12, No. 1, 145–151, 1999.
- [41] Sutskever, I.; Martens, J.; Dahl, G.; Hinton, G. On the importance of initialization and momentum in deep learning. In: Proceedings of the 30th International Conference on International Conference on Machine Learning, Vol. 28, III-1139–III-1147, 2013.
- [42] Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K. M. Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 1132–1140, 2017.
- [43] Nah, S.; Kim, T. H.; Lee, K. M. Deep multi-scale convolutional neural network for dynamic scene deblurring. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 257–265, 2017.
- [44] Nair, V.; Hinton, G. E. Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on International Conference on Machine Learning, 807–814, 2010.
- [45] Zontak, M.; Mosseri, I.; Irani, M. Separating signal from noise using patch recurrence across scales. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1195–1202, 2013.
- [46] Dong, H.; Pan, J. S.; Xiang, L.; Hu, Z.; Zhang, X. Y.; Wang, F.; Yang, M. H. Multi-scale boosted dehazing network with dense feature fusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2154–2164, 2020.
- [47] Irani, M.; Peleg, S. Improving resolution by image registration. *CVGIP: Graphical Models and Image Processing* Vol. 53, No. 3, 231–239, 1991.
- [48] Haris, M.; Shakhnarovich, G.; Ukita, N. Deep back-projecti networks for single image super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 43, No. 12, 4323–4337, 2021.
- [49] Dai, S. Y.; Han, M.; Wu, Y.; Gong, Y. H. Bilateral back-projection for single image super resolution. In: Proceedings of the IEEE International Conference on Multimedia and Expo, 1039–1042, 2007.
- [50] Agustsson, E.; Timofte, R. NTIRE 2017 challenge on single image super-resolution: Dataset and study. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 1122–1131, 2017.
- [51] Abdelhamed, A.; Lin, S.; Brown, M. S. A high-quality denoising dataset for smartphone cameras. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 1692–1700, 2018.
- [52] Anaya, J.; Barbu, A. RENOIR - A dataset for real low-light image noise reduction. *Journal of Visual Communication and Image Representation* Vol. 51, 144–154, 2018.
- [53] Martin, D.; Fowlkes, C.; Tal, D.; Malik, J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings of the 8th IEEE International Conference on Computer Vision, 416–423, 2001.
- [54] Huang, J. B.; Singh, A.; Ahuja, N. Single image super-resolution from transformed self-exemplars. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 5197–5206, 2015.
- [55] Franzen, R. Kodak lossless true color image suite. 1999. Available at <http://r0k.us/graphics/kodak>.
- [56] Plötz, T.; Roth, S. Benchmarking denoising algorithms with real photographs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2750–2759, 2017.
- [57] Wang, Z.; Bovik, A. C.; Sheikh, H. R.; Simoncelli, E. P. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing* Vol. 13, No. 4, 600–612, 2004.
- [58] Liu, D.; Wen, B. H.; Fan, Y. C.; Loy, C. C.; Huang, T. S. Non-local recurrent network for image restoration. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems, 1680–1689, 2018.
- [59] Tian, C. W.; Xu, Y.; Zuo, W. M.; Du, B.; Lin, C. W.; Zhang, D. Designing and training of a dual CNN for image denoising. *Knowledge-Based Systems* Vol. 226, 106949, 2021.
- [60] Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Color image denoising via sparse 3D collaborative filtering with grouping constraint in luminance-chrominance space. In: Proceedings of the IEEE International Conference on Image Processing, 313–316, 2007.

- [61] Huang, T.; Li, S. J.; Jia, X.; Lu, H. C.; Liu, J. Z. Neighbor2Neighbor: Self-supervised denoising from single noisy images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 14776–14785, 2021.
- [62] Kingma, D. P.; Ba, J. Adam: A method for stochastic optimization. In: Proceedings of the 3rd International Conference for Learning Representations, 2015.
- [63] Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the 13th International Conference on Artificial Intelligence and Statistics, 249–256, 2010.
- [64] Anwar, S.; Barnes, N.; Petersson, L. Attention-based real image restoration. *IEEE Transactions on Neural Networks and Learning Systems* doi: 10.1109/TNNLS.2021.3131739, 2021.
- [65] Liu, C.; Yuen, J.; Torralba, A. SIFT flow: Dense correspondence across scenes and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 33, No. 5, 978–994, 2011.



**Jingzhao Xu** is currently pursuing a Ph.D. degree at the School of Mathematics, Jilin University, where he received his B.S. degree of information and computing science in 2017. His research interests include image processing, pattern recognition, and deep learning.



**Mengke Yuan** is currently an assistant professor at the National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CASIA). He received his B.S. degree in applied mathematics and his M.S. degree in computational mathematics from Zhengzhou University in 2012 and 2015 respectively, and his Ph.D. degree in computer science from NLPR, CASIA in 2019. His research

interests lie in image processing, computer vision, and computer graphics.



**Dong-Ming Yan** is a professor with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. He received his Ph.D. degree in computer science from Hong Kong University in 2010, and his master and bachelor degrees in computer science and technology from Tsinghua University in 2005 and 2002, respectively. His research interests include computer graphics, computer vision, and pattern recognition.



**Tieru Wu** is a professor with the School of Mathematics, Jilin University. His research interests include computer graphics, CAGD, and deep learning.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made.

The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Other papers from this open access journal are available free of charge from <http://www.springer.com/journal/41095>. To submit a manuscript, please go to <https://www.editorialmanager.com/cvmj>.



清华大学出版社  
Tsinghua University Press



Springer