A TRAINABLE BOUNDED DENOISER USING DOUBLE TIGHT FRAME NETWORK FOR SNAPSHOT COMPRESSIVE IMAGING

BaoShun Shi [1,2*], *Yuxin Wang* [1,2], *Qiusheng Lian* [1,2]

[1] School of Information Science and Engineering, Yanshan University, China
[2] Hebei Key Laboratory of Information Transmission and Signal Processing, Hebei province, China

ABSTRACT

Recently, the PnP-GAP algorithm has achieved remarkable reconstruction quality for snapshot compressive imaging (SCI), and its convergence has been proven based on the condition of diminishing noise levels and the assumption of bounded denoisers. However, most of deep denoisers are difficult to be proven as bounded denoisers due to the lack of interpretability of deep network architectures. To address this issue, a trainable bounded denoiser using double tight frame network for SCI is proposed. Firstly, to achieve higher denoising ability, we extend the single-layer tight frame to the two-layer one dubbed as double tight frame. Then, we elaborate a deep shrinkage network (DSN) for improving the generalization ability of the plugged deep denoiser to noise levels. Finally, we employ the double tight frame equipped with DSN to construct a Gaussian denoiser that can be trained in a supervised learning manner. We prove this trainable denoiser is bounded theoretically, and demonstrate that the PnP-GAP framework plugged by this denoiser can achieve competitive reconstruction quality compared with benchmark SCI algorithms empirically.

Index Terms— double tight frame, computational imaging, image processing, deep learning

1. INTRODUCTION

Snapshot compressive imaging (SCI) system is an effective route to meet the challenge of acquiring high-dimensional (HD) signals in recent years [1]. The lower-dimensional measurement is captured by 2D detectors. Recovering the original HD signals from the measurement with limited information is a challenge. Since the inverse problem of SCI is ill-posed, image priors are often employed. So far, numerous algorithms using different priors, such as the non-local self-similarity in DeSCI [2] and gradient sparsity in GAP-TV [3], were developed for SCI. However, such traditional iterative algorithms suffer from high computational burden.

Inspired by deep learning techniques, researchers attempted to train an end-to-end network for SCI inversion [4, 5] and explored deep priors to promote the reconstruction quality [6, 7]. These deep learning-based algorithms greatly improve the reconstruction quality and reduce the computational burden, but lack of flexibility. Plug-and-play methods [8] can address this issue, which are promising trade-offs between traditional iterative algorithms and deep learning-based algorithms [9, 10]. Recently, the generalized alternating projection (GAP) [11] algorithms plugged by deep denoisers such as FFDNet [12] and FastDVDNet [13], not only can explore deep priors for recovering high-quality images, but also consume less running time compared with iterative algorithms using traditional priors.

On the theoretical side, the convergence of PnP-GAP [9] has been proven based on the condition of diminishing noise levels and the assumption of bounded denoisers [14]. For a bounded denoiser, the energy of the difference between the input image and its filtered version by the denoiser has an upper bound which is proportional to the noise variance. However, most of deep denoisers are difficult to be proven as bounded denoisers due to the complicated deep neural network architectures. Recently, to analyze the deep neural networks, theoretical foundations of deep learning via sparse representations were discussed in [15]. Under some conditions, convolutional neural networks (CNN) can be regarded as multilayered convolutional sparse coding (MLCSC) models [15]. Inspired by this spirit, we come up with a trainable bounded denoiser using the double tight frame network for SCI, the main contributions are summarized as follows.

- We extend the single-layer tight frame to the two-layer one termed double tight frame. We propose a supervised double tight frame learning framework, and show that the denoiser using the double tight frame has higher denoising ability than that using the single one empirically.
- We propose a deep shrinkage network (DSN) for filtering the frame coefficients adaptively. Differing from learning non-adaptive thresholds in [16], we elaborate a constant net (Cnet), the core unit of DSN, to extract the proportional constants of the thresholds from the instance adaptively. The double tight frame equipped with DSN is the so-called DoubleTFCnet whose architecture is depicted in Fig. 1 (a). We prove DoubleTFCnet is a bounded denoiser theoretically.
- We demonstrate the effectiveness of the proposed DoubleTFCnet on image denoising and SCI tasks. For image denoising, experiments demonstrate the DoubleTFCnet can

This work was supported by the National Natural Science Foundation of China (61901406), Natural Science Foundation of Hebei Province (F2020203025 and F2019203318), Young Talent Program of Universities and Colleges in Hebei Province (BJ2021044).

*Corresponding author: Baoshun Shi(shibaoshun@ysu.edu.cn)

achieve denoising ability on-par with previous deep denoisers. We incorporate DoubleTFCnet into the PnP-GAP framework, termed as PnP-DoubleTFCnet. Fig. 1 (b) depicts its schematic diagram. Experimental results show that PnP-DoubleTFCnet can achieve results comparable to PnP-FFDNet.

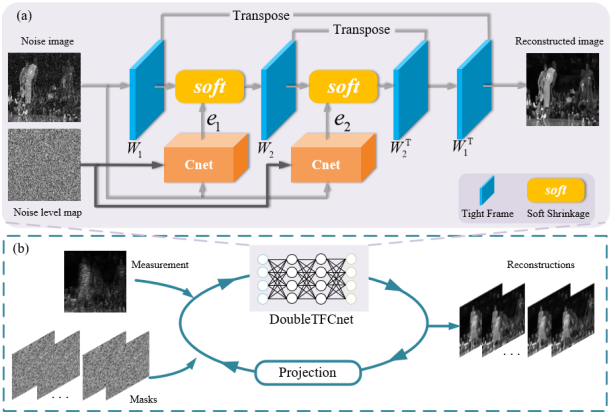


Fig. 1. (a) The architecture of DoubleTFCnet. (b) The GAP framework plugged by DoubleTFCnet. The raw measurement and the masks are sent to the PnP-GAP framework, and the estimated images denoised by DoubleTFCnet in each iteration.

2. THE PROPOSED DOUBLE TIGHT FRAME NETWORK

2.1. Tight frame learning: from single to double

Traditional tight frame learning methods use alternating iterative optimization to learn tight frames, which prevent the involvement of modern deep learning tricks. Differing from these spirits, we train tight frames via supervised learning. Let $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{n}$ represent the version of clean image $\boldsymbol{x} \in \mathbb{R}^N$ corrupted by Gaussian noise $\boldsymbol{n} \in \mathbb{R}^N$. The image filtered by a single tight frame $\boldsymbol{W} \in \mathbb{R}^{M\times N}(M \geq N)$ can be described as $\boldsymbol{W}^{\mathrm{T}} T_{\varepsilon}(\boldsymbol{W}\boldsymbol{z})$. Here $T_{\varepsilon}(\bullet)$ is the soft thresholding function defined as $soft_{\varepsilon}(u) = sign(u)max(|u|-\varepsilon, 0)$, and ε is the threshold. To learn this tight frame $\boldsymbol{W}$, the loss function can be defined as

$$L(\boldsymbol{W}, \varepsilon) = ||\boldsymbol{x} - \boldsymbol{W}^{\mathrm{T}} T_{\varepsilon}(\boldsymbol{W}\boldsymbol{z})||_2^2 + \lambda||\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W} - \boldsymbol{I}||_F^2, \quad (1)$$

where λ is a trade-off parameter. In Eqn. (1), the first term promotes the sparsity of the image over the tight frame, while the second term enforces the tight property onto the frame to be learned.

As a new interpretation of CNN, the MLCSC model consisting of a cascade of convolutional sparse layers has higher capacity of representing images, compared with the traditional single CSC model [15]. Inspired by this idea, we extend the single tight frame to the double one, and design a double tight frame denoiser

$$\mathcal{D}(\boldsymbol{z}) = \boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{W}_2^{\mathrm{T}} T_{\varepsilon_2}[\boldsymbol{W}_2 T_{\varepsilon_1}(\boldsymbol{W}_1\boldsymbol{z})], \quad (2)$$

where the tight frames $\boldsymbol{W}_1 \in \mathbb{R}^{M\times N}$ and $\boldsymbol{W}_2 \in \mathbb{R}^{M\times M}$ admit the tight property $\boldsymbol{W}_1{}^{\mathrm{T}}\boldsymbol{W}_1 = \boldsymbol{I}$ and $\boldsymbol{W}_2{}^{\mathrm{T}}\boldsymbol{W}_2 = \boldsymbol{I}$ respectively. In order to learn such denoiser, we can define a loss function as

$$\begin{aligned} L(\boldsymbol{W}_1, \boldsymbol{W}_2, \varepsilon_1, \varepsilon_2) &= ||\boldsymbol{x} - \mathcal{D}(\boldsymbol{z})||_2^2 \\ &+\lambda_1||\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{W}_1 - \boldsymbol{I}||_F^2 + \lambda_2||\boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{W}_2 - \boldsymbol{I}||_F^2. \end{aligned} \quad (3)$$

The first term enforces that the output of $\mathcal{D}(\bullet)$ is matched with the ground truth, and the last two terms constrain the tight property onto the two learned frames.

2.2. Deep shrinkage network

The thresholds ε_1 and ε_2 in Eqn. (2) are hyper-parameters that can be learned in a supervised learning manner. However, this method cannot cope with spatially-variant noise levels, resulting in poor generalization ability of the network. In [17], Mallat proposed the so-called "universal threshold theorem". Based on this theorem, Isogawa et al. [16] employed soft shrinkage as activation functions of DnCNN [18], and the thresholds of soft shrinkage are proportional to the input noise level σ, i.e. $\varepsilon_{l,i} = c_{l,i}\cdot\sigma$. Here $\varepsilon_{l,i}$ is the threshold for the i-th channel of feature map given the l-th layer, and $c_{l,i}$ is the corresponding proportional constant to be learned.

Since the image components or pixels are correlated in the spatial domain, the thresholds should be different for each coefficient. Based on this fact, we propose a DSN in which the thresholds are proportional to the noise levels and varied spatially. We propose the so-called Cnet that determines the proportional constants by using a deep network. Formally, the thresholds can be defined as

$$\boldsymbol{e}_l = \boldsymbol{c}_l \odot \boldsymbol{m}, l = 1, 2, \quad (4)$$

where $\odot$ denotes the element-wise product and l represents the l-th layer tight frame. The elements in $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ are utilized for the corresponding thresholds ε_1 and ε_2 in Eqn. (2) respectively. By doing so, the thresholds are spatially varied. To solve the problem of dimension mismatch, we stretch the noise level σ into a noise level map $\boldsymbol{m}$ whose size is the same as $\boldsymbol{c}_l$. The $\boldsymbol{c}_l$ is the proportional constant of threshold $\boldsymbol{e}_l$, which can be determined by Cnet. To avoid arbitrary, the elements of $\boldsymbol{c}_l$ are limited to $[c_{min}, c_{max}]$. Fig. 2 shows the architecture of Cnet. In order to pay attention to the channels with much information of image, we incorporate the residual channel attention block (RCAB) [19] into the Cnet. The main part of Cnet consists of a series of 3×3 convolutional layers and multiple RCAB cascades. The final output of the Cnet is the proportional constant $\boldsymbol{c}_l$.

2.3. The proposed DoubleTFCnet: a bounded denoiser

In [14], Chan et al. defined a class of bounded denoisers and proved the fixed-point convergence of the plug-and-play algorithms for image restoration. Subsequently, based on the assumption of bounded denoisers, Yuan et al. [9] proved the convergence of PnP-GAP under the condition of diminishing noise levels. However, how to design a trainable bounded

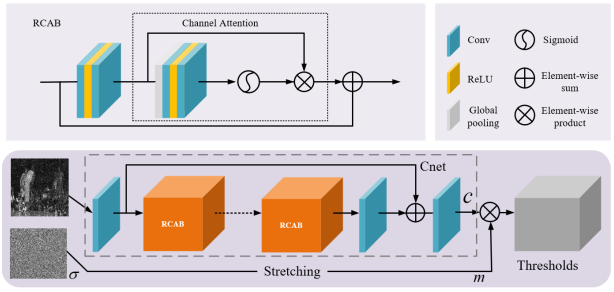


Fig. 2. The architecture of Cnet.

denoiser is a challenge. We address this issue by using the double tight frame network and DSN. We incorporate DSN into the double tight frame network to elaborate a denoiser. We term this denoiser equipped with Cnet as DoubleTFCnet, which can be described as

$$\mathcal{D}(\boldsymbol{z}) = \boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{W}_2^{\mathrm{T}}T[\boldsymbol{W}_2T(\boldsymbol{W}_1\boldsymbol{z}, \boldsymbol{e}_1), \boldsymbol{e}_2]. \tag{5}$$

Here, $T(\boldsymbol{W}_l\boldsymbol{z}, \boldsymbol{e}_l)$ indicates that each element of $\boldsymbol{e}_l$ is the threshold for shrinking each corresponding element of $\boldsymbol{W}_l\boldsymbol{z}$, $l = 1, 2$.

As mentioned earlier, the denoiser with bounded property is crucial to prove the convergence of PnP-GAP. In this study, we prove the proposed DoubleTFCnet is bounded theoretically. With the definition of the threshold $\boldsymbol{e}_l = \boldsymbol{c}_l \odot \boldsymbol{m}, l = 1, 2$, we have the following theorem.

Theorem 1: *For any $\boldsymbol{x} \in \mathbb{R}^N$ whose elements admit $x_i \in [0, 1]$, and each element of proportional constant $\boldsymbol{c}_l$ has a limited range $c_i \in [c_{min}, c_{max}]$. For some universal constant L independent of M and noise level σ, DoubleTFCnet is a bounded denoiser such that*

$$\frac{1}{M}||\boldsymbol{x} - \boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{W}_2^{\mathrm{T}}T[\boldsymbol{W}_2T(\boldsymbol{W}_1\boldsymbol{x}, \boldsymbol{e}_1), \boldsymbol{e}_2]||_2^2 \leq \sigma^2 L. \tag{6}$$

Proof: See the supplementary material file at https://github.com/shibaoshun/proof.

3. DOUBLE TIGHT FRAME NETWORK FOR SCI

3.1. Mathematical model of SCI

We assume that a B-frame (grayscale) video $\boldsymbol{X} \in \mathbb{R}^{n_x \times n_y \times B}$ is modulated by B sensing matrices (masks) $\boldsymbol{P} \in \mathbb{R}^{n_x \times n_y \times B}$. The single snapshot compressive measurement $\boldsymbol{Y} \in \mathbb{R}^{n_x \times n_y}$ captured by the 2D detector can be modeled as

$$\boldsymbol{Y} = \sum_{b=1}^{B} \boldsymbol{X}_b \odot \boldsymbol{P}_b + \boldsymbol{N}, \tag{7}$$

where $\odot$ denotes the Hadamard (element-wise) product, $\boldsymbol{P}_b = \boldsymbol{P}(:, :, b) \in \mathbb{R}^{n_x \times n_y}$ represents the b-th mask and $\boldsymbol{X}_b = \boldsymbol{X}(:, :, b) \in \mathbb{R}^{n_x \times n_y}$ is the original video frame, $\boldsymbol{N} \in \mathbb{R}^{n_x \times n_y}$ denotes the noise. Using vectoring operator, we can define $\boldsymbol{y} = vec(\boldsymbol{Y}) \in \mathbb{R}^{n_x n_y}$, $\boldsymbol{n} = vec(\boldsymbol{N}) \in \mathbb{R}^{n_x n_y}$, $\boldsymbol{x}_b = vec(\boldsymbol{X}_b) \in \mathbb{R}^{n_x n_y}$ and $\boldsymbol{D}_b = diag(vec(\boldsymbol{P}_b)) \in \mathbb{R}^{n_x \times n_y}$. Let $\boldsymbol{x} = [\boldsymbol{x}_1^{\mathrm{T}}, \boldsymbol{x}_2^{\mathrm{T}}, \cdots, \boldsymbol{x}_B^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{n_x n_y B}$ and $\boldsymbol{H} = [\boldsymbol{D}_1, \boldsymbol{D}_2, \cdots, \boldsymbol{D}_B]$. Therefore, Eqn. (7) can be rewritten as

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}. \tag{8}$$

3.2. PnP-DoubleTFCnet for SCI

Based on the maximum posterior probability (MAP) rule, we formulate the SCI optimization problem using a regularization model as follows

$$\min_{\boldsymbol{x}} \frac{1}{2}||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}||_2^2 + \beta g(\boldsymbol{x}), \tag{9}$$

where $\frac{1}{2}||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}||_2^2$ is the data-fidelity term and $g(\boldsymbol{x})$ is a regularization term, β is the regularization parameter. Via using the framework of GAP and by introducing an auxiliary variable $\boldsymbol{v}$, the problem (9) can be rewritten as [9]

$$\min_{\boldsymbol{x}, \boldsymbol{v}} \frac{1}{2}||\boldsymbol{x} - \boldsymbol{v}||_2^2 + \beta g(\boldsymbol{v}), s.t. \boldsymbol{y} = \boldsymbol{H}\boldsymbol{x}. \tag{10}$$

Problem (10) can be attacked by the following two steps (for the $(t+1)$-th iteration).

Updating $\boldsymbol{x}$: $\boldsymbol{x}^{(t+1)}$ is updated via an Euclidean projection of $\boldsymbol{v}^{(t)}$ on the linear manifold $\mathcal{M} : \boldsymbol{y} = \boldsymbol{H}\boldsymbol{x}$. That is

$$\boldsymbol{x}^{(t+1)} = \boldsymbol{v}^{(t)} + \boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}})^{-1}(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{v}^{(t)}), \tag{11}$$

where $\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}$ is a diagonal matrix, thus, its inverse matrix can be calculated bluntly [2, 3, 9].

Updating $\boldsymbol{v}$: Given $\boldsymbol{x}^{(t+1)}$, updating $\boldsymbol{v}$ can be regarded as a denoising problem, which can be solved by the proposed DoubleTFCnet in Eqn. (5), i.e.

$$\boldsymbol{v}^{(t+1)} = \boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{W}_2^{\mathrm{T}}T[\boldsymbol{W}_2T(\boldsymbol{W}_1\boldsymbol{x}^{(t+1)}, \boldsymbol{e}_1), \boldsymbol{e}_2]. \tag{12}$$

In summary, to obtain the satisfied solution of problem (10), we iterate Eqn. (11) and Eqn. (12) until the algorithm reaches the maximum iteration t_{max}. In Theorem 1, we have proved that DoubleTFCnet is a bounded denoiser, thus, the convergence of PnP-DoubleTFCnet can be proved under the condition of diminishing noise levels [9].

4. EXPERIMENTS

4.1. Effectiveness of DoubleTFCnet

We train the DoubleTFCnet model in a supervised learning manner. The training settings are as follows.

Dataset: We follow [12] to prepare 128×8000 input-output patches extracted from the Waterloo Exploration Database, where the batch size is 128 and patch size is 50×50. The noisy inputs are generated by adding AWGN of noise level $\sigma \in [0, 75]$ to clean image patches.

Table 1. The average PSNR values achieved by different denoisers on BSD68 with noise levels 5, 10, 20, 30, 40 and 50. The best two results are highlighted in red and blue colors, respectively.The listed denoisers are implemented by Python.

Denoiser	$\sigma = 5$	$\sigma = 10$	$\sigma = 20$	$\sigma = 30$	$\sigma = 40$	$\sigma = 50$
BM3D	37.54	33.33	29.66	27.78	26.57	25.69
DnCNN-B	37.74	33.73	30.18	28.32	27.09	26.19
FFDNet	37.73	33.74	30.23	28.40	27.20	26.32
DoubleTFCnet	37.84	33.79	30.25	28.39	27.17	26.28

Table 2. The average results of PSNR values (left) and SSIM (right) by different algorithms on six benchmark datasets. The best two results are highlighted in red and blue colors, respectively. The listed algorithms are implemented by Python.

Algorithm	Kobe		Traffic		Runner		Drop		Crash		Aerial		Average	
GAP-TV	26.92	0.8378	20.66	0.6905	29.81	0.8949	34.95	0.9664	24.48	0.7988	24.81	0.8105	26.94	0.8332
E2E-CNN	29.02	0.8612	23.45	0.8375	34.43	0.9580	36.77	0.9740	26.40	0.8858	27.52	0.8822	29.26	0.8999
PnP-FFDNet	30.33	0.9252	24.01	0.8353	32.44	0.9313	39.68	0.9864	24.67	0.8330	24.29	0.8198	29.21	0.8876
PnP-DoubleTFCnet	30.36	0.9257	23.73	0.8328	33.38	0.9432	41.33	0.9897	25.24	0.8519	25.12	0.8326	29.86	0.8960

Training: The size of filters for constructing double tight frames is 9×9, and the number is 81. The initialization of tight frame filters is trained by the traditional method [20]. The range of the proportional constants of the thresholds is $(0, 10]$. We use Eqn. (3) as the loss function, and set the trade-off parameters $\lambda_1 = \lambda_2 = 0.001$. The ADAM algorithm is employed as the optimizer whose initial learning rate is set to 2×10^{-4} and reduces to 2×10^{-5} after the 20-th epoch. We set the total number of iterations to 50 epochs. All experiments were performed in a computer with the NVIDIA GTX 3080ti GPU based on PyTorch.

We compare the noise reduction performance of BM3D [21], DnCNN-B [18], FFDNet [12] and the proposed DoubleTFCnet in the case of Gaussian noise. The peak signal-to-noise ratio (PSNR) values achieved by the benchmark denoisers and DoubleTFCnet on BSD68 are presented in Table 1. One can see from the table that the proposed DoubleTFCnet can obtain PSNR values on-par with the state-of-the-art deep denoisers.

We trained a model that using single tight frame with Cnet, denoted as SingleTFCnet. Meanwhile, we trained SingleTFC and DoubleTFC whose proportional constants of the thresholds are learned in a supervised learning manner. These three models share training settings of the DoubleTFCnet except for network structures. Fig. 3 shows the PSNR values achieved by SingleTFC, SingleTFCnet, DoubleTFC and DoubleTFCnet on Set12 with noise levels ranging from 20 to 40. From Fig. 3, we can draw the following conclusions: (1) the denoisers using the double tight frame have higher denoising ability than those using the single one; (2) the denoisers equipped with Cnet can achieve higher PSNR values compared with those achieved by the denoisers without Cnet.

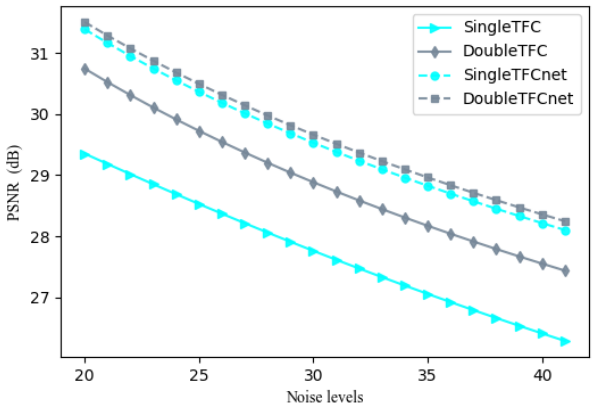
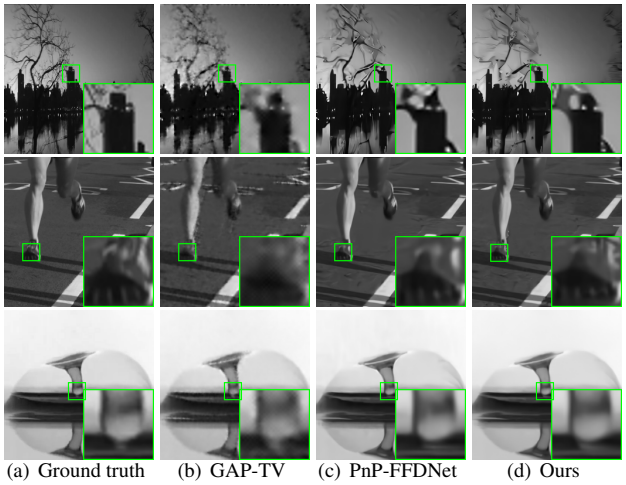


Fig. 3. The PSNR values achieved by four models at various noise levels.



(a) Ground truth (b) GAP-TV (c) PnP-FFDNet (d) Ours

Fig. 4. Reconstructed frames of PnP-GAP algorithms on three simulated benchmark video SCI datasets.

4.2. Effectiveness of PnP-DoubleTFCnet

We compare the performance of PnP-DoubleTFCnet with that of GAP-TV [3], E2E-CNN [5], and PnP-FFDNet [9] on simulation datasets. We follow [9] to evaluate PnP-DoubleTFCnet on six simulation datasets. Table 2 summarizes the PSNR values and structural similarity (SSIM) results using various SCI algorithms under the noise-free case. One can see from the table that the proposed PnP-DoubleTFCnet can achieve the highest average PSNR value. Moreover, the average PSNR value of PnP-DoubleTFCnet is 0.65 dB higher than that of PnP-FFDNet. This lies in the fact that DoubleTFCnet can extract thresholds from instances adaptively.

Fig. 4 presents selected frames of the three datasets using different algorithms. It can be observed that the proposed PnP-DoubleTFCnet can recover better edge contour information compared with GAP-TV. From visual perspective, the proposed PnP-DoubleTFCnet achieved reconstruction qualities similar to PnP-FFDNet.

5. CONCLUSION

In this paper, we proposed a novel double tight frame network termed DoubleTFCnet for PnP-GAP framework. The DoubleTFCnet serving as a denoiser is consisting of double tight frame and an elaborated deep shrinkage network. On the theoretical side, we have proven this denoiser is bounded. On the practical side, we have demonstrated its effectiveness on image denoising and SCI tasks. In the future, we will apply DoubleTFCnet to tackling different imaging tasks.

6. REFERENCES

[1] Xin Yuan, David J. Brady, and Aggelos K. Katsaggelos, "Snapshot compressive imaging: Theory, algorithms, and applications," *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 65–88, Mar. 2021.

[2] Yang Liu, Xin Yuan, Jinli Suo, David J. Brady, and Qionghai Dai, "Rank minimization for snapshot compressive imaging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 12, pp. 2990–3006, Dec. 2019.

[3] Xin Yuan, "Generalized alternating projection based total variation minimization for compressive sensing," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, Sept. 2016, pp. 2539–2543.

[4] Ziheng Cheng, Ruiying Lu, Zhengjue Wang, Hao Zhang, Bo Chen, and Xin Yuan, "BIRNAT: Bidirectional recurrent neural networks with adversarial training for video snapshot compressive imaging," in *Proceedings of the European Conference on Computer Vision (ECCV)*, Aug. 2020.

[5] Mu Qiao, Ziyi Meng, Jiawei Ma, and Xin Yuan, "Deep learning for video compressive sensing," *APL Photonics*, vol. 5, no. 3, pp. 030801, 2020.

[6] Haiquan Qiu, Yao Wang, and Deyu Meng, "Effective snapshot compressive-spectral imaging via deep denoising and total variation priors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2021, pp. 9127–9136.

[7] Shipeng Zhang, Lizhi Wang, Lei Zhang, and Hua Huang, "Learning tensor low-rank prior for hyperspectral image reconstruction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2021, pp. 12006–12015.

[8] Singanallur V. Venkatakrishnan, Charles A. Bouman, and Brendt Wohlberg, "Plug-and-play priors for model based reconstruction," in *Proceedings of the IEEE Global Conference on Signal and Information Processing*, Dec. 2013, pp. 945–948.

[9] Xin Yuan, Yang Liu, Jinli Suo, and Qionghai Dai, "Plug-and-play algorithms for large-scale snapshot compressive imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2020, pp. 1444–1454.

[10] Xin Yuan, Yang Liu, Jinli Suo, Fredo Durand, and Qionghai Dai, "Plug-and-play algorithms for video snapshot compressive imaging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Jul. 2021.

[11] Xuejun Liao, Hui Li, and Lawrence Carin, "Generalized alternating projection for weighted-ℓ_2, 1 minimization with applications to model-based compressive sensing," *SIAM Journal on Imaging Sciences*, vol. 7, no. 2, pp. 797–823, Apr. 2014.

[12] Kai Zhang, Wangmeng Zuo, and Lei Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4608–4622, Sept. 2018.

[13] Matias Tassano, Julie Delon, and Thomas Veit, "FastDVDnet: Towards real-time deep video denoising without flow estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2020, pp. 1351–1360.

[14] Stanley H. Chan, Xiran Wang, and Omar A. Elgendy, "Plug-and-play ADMM for image restoration: Fixed-point convergence and applications," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 84–98, Mar. 2017.

[15] Vardan Papyan, Yaniv Romano, Jeremias Sulam, and Michael Elad, "Theoretical foundations of deep learning via sparse representations: A multilayer sparse model and its connection to convolutional neural networks," *IEEE Signal Processing Magazine*, vol. 35, no. 4, pp. 72–89, Jul. 2018.

[16] Kenzo Isogawa, Takashi Ida, Taichiro Shiodera, and Tomoyuki Takeguchi, "Deep shrinkage convolutional neural network for adaptive noise reduction," *IEEE Signal Processing Letters*, vol. 25, no. 2, pp. 224–228, Feb. 2018.

[17] Stephane Mallat, "A wavelet tour of signal processing: The sparse way," *Elsevier Science*, 2008.

[18] 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*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[19] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu, "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, Sept. 2018, pp. 286–301.

[20] Jianfeng Cai, Ji Hui, Zuowei Shen, and Guibo Ye, "Data-driven tight frame construction and image denoising," *Applied and Computational Harmonic Analysis*, vol. 37, no. 1, pp. 89–105, Jul. 2014.

[21] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.