PROXIMAL GRADIENT-BASED LOOP UNROLLING WITH INTERSCALE THRESHOLDING

Ruiki KOBAYASHI*, Shogo MURAMATSU*, and Shunsuke ONO†
* Niigata University, Niigata, Japan
† Tokyo Institute of Technology University, Tokyo, Japan
E-mail: f20c050g@ieee.org, shogo@eng.niigata-u.ac.jp, ono@c.titech.ac.jp

***Abstract*—This work proposes an image restoration technique by introducing an interscale thresholding with a structured convolutional dictionary into a loop unrolled network based on the proximal gradient descent (PGD) method. A non-separable oversampled lapped transforms (NSOLT) in tree structure is adopted as the dictionary, where the interscale linear expansion of thresholds (LET) is applied as a Gaussian denoiser using the plug-and-play technique. The design parameters of the dictionary and thresholding function are made trainable, and image restoration systems are designed with the deep learning approach. Through the simulation of denoising and single image super-resolution (SISR), it is confirmed that the proposed method gives a high-performance feed-forward image restoration process with few design parameters.**

I. INTRODUCTION

With the development of measurement technologies, it is now possible to acquire a large amount of different physical data, such as tomographic imaging data and multi-dimensional time-series data. Simultaneously, the demand for high performance signal restoration is increasing. For high-efficiency signal restoration, it is advisable to use a generative model that can effectively represent the original signal. A generative process is a mathematical expression of a prior knowledge about the target signal model and it is employed in a lot of image restoration techniques to enable sparse modeling, wherein the generative model assumes that the essential information of the target signal can sparsely be reprensented [1], [2]. A generative model is provided by a dictionary consisting of atomic waveforms, which has a role to synthesize signals from sparse features. In image restoration using sparse modeling, iterative optimization processes are frequently used [3]–[5]. While the interpretability of this approach is clear, the processing speed and learning flexibility are disadvantages.

In the last decade, artificial intelligence (AI) has advanced remarkably. In particular, deep learning, such as design for convolutional neural networks (CNNs), has been applied in various fields, including signal restoration [6], [7]. CNN is a feed-forward-type neural network involving multiple layers of convolution and nonlinear functions that extract the local features of signals. Conventional CNNs can take the role of a generative model, but it requires a large amount of data for training. Unlike sparse modeling, most of the CNNs proposed for image restoration face the challenge of poor interpretability because they confuse the generative process with the observation process, making them black boxes. In addition, a large memory capacity is required for the large number of design parameters, and overfitting is likely to occur if there is insufficient training data.

Algorithm 1 Proximal gradient descent (PGD)

Input: $\mathbf{x}^{(0)}$
Output: $\mathbf{x}^{(n)}$
1: **while** A stopping criterion is not satisfied **do**
2: $\quad\mathbf{x}^{(n+1)} = \mathrm{prox}_{\gamma g}\left(\mathbf{x}^{(n)} - \gamma\nabla f(\mathbf{x}^{(n)})\right)$
3: $\quad n \leftarrow n+1$
4: **end while**

Gregor *et al.* developed the learned iterative thresholding algorithm (LISTA), which unrolls the iterative loop operations for solving a sparse-model-based problem. LISTA can train a network containing synthesis dictionaries, observation processes and regularized denoisers [8]. Furthermore, advanced methods have been proposed to improve the recovery performance and to reduce the number of parameters [9], [10]. LISTA is able to solve problems of the iteration and learning limitations in the sparse modeling approach. However, since generative process and gradient descent steps are trained together, their interpretability is still controversial.

In order to gain the quality and to reduce the design parameters in the loop unrolling approach, this study proposes to introduce an interscale thresholding with a structured convolutional dictionary to a proximal gradient-based loop unrolling network. The proposed method can effectively reduce the redundancy of parameters by using the knowledge of the filter bank and interscaling. Therefore, it solves the problem of the number of design parameters in conventional unrolling networks. The proposed method is a network that explicitly takes account of the generative model and the observation process. Since the optimization problem can be clearly understood, the proposed method can solve the interpretability issue of the conventional unrolling networks.

II. IMAGE RESTORATION BY SPARSE MODELING

In this section, let us review sparsity-aware image restoration with the proximal gradient descent (PGD) method.

A. PGD algorhtm

The PGD algorithm is a popular primitive method in sparsity-aware image restoration [11], [12]. PGD can solve problems in the following form:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathbb{R}^L} f(\mathbf{x}) + g(\mathbf{x}), \tag{1}$$

where $f\colon \mathbb{R}^L \to \mathbb{R}\cup\{\infty\}$ and $g\colon \mathbb{R}^L \to \mathbb{R}\cup\{\infty\}$ are the proper lower semi-continuous convex functions; $\nabla f(\cdot)$ is the μ-Lipschitz continuous. Note that $g(\cdot)$ is preferable to be a function with a closed-form proximal operator.

Algorithm 1 shows the steps of PGD, where $\mathrm{prox}_h(\cdot)$ denotes the proximal map of a function $h(\cdot)$, and γ is a step size satisfying the condition $0 < \gamma \le 2/\mu$.

B. Restoration with synthesis dictionary

Let us consider applying the PGD algorithm to a sparsity-aware image restoration problem. Sparse modeling is a technique to solve an inverse problem with a prior knowledge that the essential information of the signal of interest is represented sparsely. By using an appropriate synthesis dictionary $\mathbf{D}\in\mathbb{R}^{N\times L}$, an image $\mathbf{u}\in\mathbb{R}^N$ can sparsely be represented. The synthesis of $\mathbf{u}$ is expressed as

$$\mathbf{u} = \mathbf{D}\mathbf{x}, \tag{2}$$

where $\mathbf{x}\in\mathbb{R}^L$ is a coefficient vector.

In this study, we assume an observation model expressed as

$$\mathbf{v} = \mathbf{P}\mathbf{u} + \mathbf{w}, \tag{3}$$

where $\mathbf{P}\in\mathbb{R}^{M\times N}$ is a linear measurement process, $\mathbf{w}\in\mathbb{R}^M$ is an additive white Gaussian noise (AWGN), and $\mathbf{v}\in\mathbb{R}^M$ is a contaminated observation.

A regularized least-squares problem is a typical application of (1). A sparsity-aware image restoration problem with a synthesis dictionary $\mathbf{D}$ is represented by

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathbb{R}^L} \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{v}\|_2^2 + \lambda\rho(\mathbf{x}), \tag{4}$$

where $\mathbf{A} = \mathbf{P}\mathbf{D}\in\mathbb{R}^{M\times L}$, $\|\cdot\|_2$ is the standard norm, $\rho(\cdot)$ is the regularizer, and λ is the regularization parameter. When the regularizer is the ℓ_1 norm, i.e., $\rho(\cdot) = \|\cdot\|_1$, (4) reduces to the least absolute shrinkage and selection operator (LASSO) [13]. The optimization problem in (4) can be solved by PGD in (1) by letting

$$f(\mathbf{x}) = \frac{1}{2}\left\|\mathbf{A}\mathbf{x}-\mathbf{v}\right\|_2^2, \tag{5a}$$

$$g(\mathbf{x}) = \lambda\rho(\mathbf{x}). \tag{5b}$$

Then, the second step in Algorithm 1 is expressed by

$$\mathbf{x}^{(n+1)} = \mathrm{prox}_{\gamma\lambda\rho}\left(\mathbf{x}^{(n)} - \gamma\mathbf{A}^{\mathsf{T}}\left(\mathbf{A}\mathbf{x}^{(n)}-\mathbf{v}\right)\right). \tag{6}$$

Since

$$\mathrm{prox}_{\gamma g}(\mathbf{x}) \triangleq \arg\min_{\mathbf{y}} \frac{1}{2\gamma}\|\mathbf{y}-\mathbf{x}\|_2^2 + g(\mathbf{y}) \tag{7}$$

can be interpreted as a maximum aposteriori (MAP) estimation of a signal $\mathbf{x}\sim p(\mathbf{x})\propto\exp(-g(\mathbf{x}))$ contaminated by AWGN of standard deviation $\sqrt{\gamma}$, it is convenient to introduce another expression of the proximal operator as a Gaussian denoiser

$$\mathfrak{G}_g(\mathbf{x},\sqrt{\gamma}) = \mathrm{prox}_{\gamma g}(\mathbf{x}). \tag{8}$$

The representation in (8) allows for the extension of the algorithm through a plug-and-play (PnP) approach [14]. For $\rho(\cdot) = \|\cdot\|_1$, the denoiser reduces to the soft-thresholding, i.e.,

$$\mathfrak{G}_{\lambda\|\cdot\|_1}(\mathbf{x},\sqrt{\gamma}) = \mathrm{sign}(\mathbf{x})\odot\max(\mathrm{abs}(\mathbf{x})-\gamma\lambda\mathbf{I},\mathbf{0}), \tag{9}$$

where $\odot$ denotes the element-wise multiplication, i.e., the Hadamard product. The PGD algorithm repeats this soft-thresholding operation and gives us the iterative shrinkage thresholding algorithm (ISTA) [3]. The PGD algorithm is based on an explicit model, which is highly interpretable, and has the advantage that it does not require an inverse matrix and can approach the exact solution by repeating simple operations. However, there are some issues such as lack of flexibility in repeating the same operation and slow latency due to iteration.

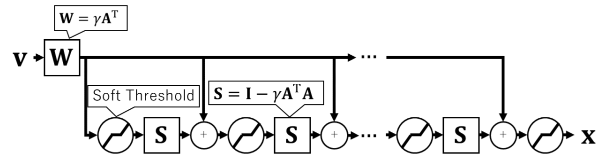

Fig. 1. Architecture of LISTA, where ISTA's loop is unrolled. The matrices $\mathbf{W}$ and $\mathbf{S}$, are learned, so as to minimize the approximation error of the optimal sparse codes on a given dataset [8].

C. Learned ISTA (LISTA)

CNNs are capable of extracting features of input images by reducing the dimensionality through operations such as convolution, downsampling, and nonlinear activation. Zhang *et al.* applied a multilayer network consisting of transpose convolution, upsampling, nonlinear activation to the extracted features to achieve image denoising [6]. Various proposals have been made for image restoration using CNNs. Most CNN-based image restoration methods show high performance. However, they lack interpretability and have a problem of overfitting easily without a large amount of training data due to the large number of design parameters.

Gregor *et al.* developed the learned ISTA (LISTA), a pioneering method of loop unrolling [8]. The loop unrolling approach is an image restoration method with characteristics intermediate between model-based iterative algorithms and learning-based deep CNNs.

Fig. 1 depicts an example architecture of LISTA [8]. LISTA unloops the iteration steps of ISTA and builds a feed-forward network to learn the parameters of each layer independently and build an image restoration unit with fewer iteration steps (number of layers). In LISTA, (6) is rewritten as

$$\begin{aligned}\mathbf{x}^{(n+1)} &= \mathrm{prox}_{\gamma\lambda\|\cdot\|_1}\left(\mathbf{S}\mathbf{x}^{(n)}+\mathbf{W}\mathbf{v}\right)\\ &= \mathfrak{G}_{\lambda\|\cdot\|_1}\left(\mathbf{S}\mathbf{x}^{(n)}+\mathbf{W}\mathbf{v},\sqrt{\gamma}\right),\end{aligned} \tag{10}$$

Algorithm 2 Inference by LISTA [8]

Input: $\mathbf{v}$
Output: $\mathbf{x}^{(T)}$
$\mathbf{b} = \mathbf{W}\mathbf{v}$
$\mathbf{x}^{(0)} = \mathrm{prox}_{\gamma g}(\mathbf{b})$
for $t = 0$ to $T-1$ **do**
  $\mathbf{x}^{(t+1)} = \mathfrak{G}_{\lambda\|\cdot\|_1}\left(\mathbf{S}\mathbf{x}^{(t)} + \mathbf{b}, \sqrt{\gamma}\right)$
end for

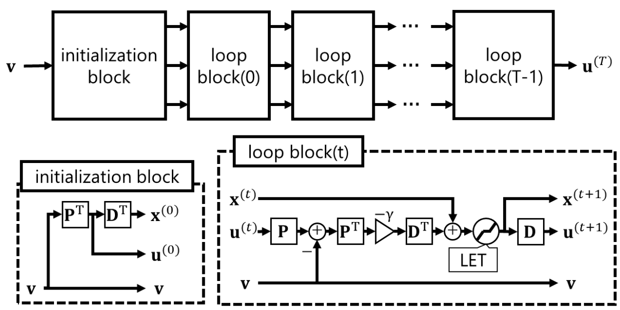

Fig. 2. Architecture of proposed loop unrolling network, where $\mathbf{D}$ is a synthesis dictionary and $\mathbf{P}$ is a measurement process. $\mathbf{D}^\mathsf{T}$ is the adjoint of $\mathbf{D}$ and has the role of analyzer. Note that $\mathbf{D}$ and $\mathbf{D}^\mathsf{T}$ are independently trained so that the backpropagation is available.

where

$$\mathbf{S} = \mathbf{I} - \gamma\mathbf{A}^\mathsf{T}\mathbf{A}, \tag{11a}$$

$$\mathbf{W} = \gamma\mathbf{A}^\mathsf{T}, \tag{11b}$$

where $\mathbf{I}$ is the identity. The corresponding inference prodedure of T-stage LISTA is represented as in Algorithm 2.

III. Loop Unrolling with Inter-scale Thresholding

Since the proximal operator in (7) can be interpreted as a regularized Gaussian denoiser as in (8), we can modify the operation by assuming an implicit regularizer $g(\cdot)$. In this study, we propose to introduce the interscale LET [16] as a more effective Gaussian denoiser $\mathfrak{G}_g$ and a hierarchical convolutional dictionary based on NSOLT as a trainable synthesis dictionary $\mathbf{D}$ into the unrolled PGD method.

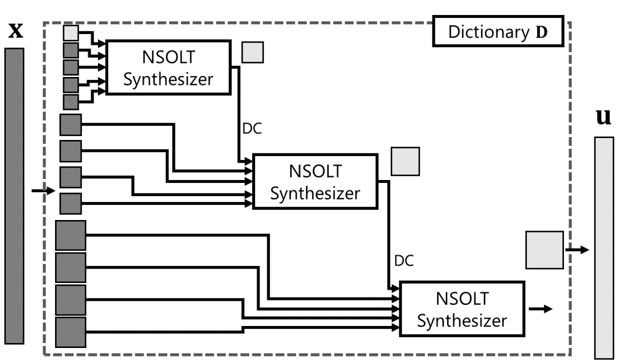

Fig. 3. A 3-level tree structure of multiscale synthesis NSOLT [15]

The gradient $\nabla f(\mathbf{x})$ of $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{v}\|_2^2$ is expressed by

$$\nabla f(\mathbf{x}) = \mathbf{A}^\mathsf{T}(\mathbf{A}\mathbf{x} - \mathbf{v}) = \mathbf{D}^\mathsf{T}\mathbf{P}^\mathsf{T}(\mathbf{P}\mathbf{D}\mathbf{x} - \mathbf{v}). \tag{12}$$

Then, the corresponding inference procedure of the proposed PGD method can be shown as in Algorithm 3. Note that the measurement process $\mathbf{P}$, synthesis dictionary $\mathbf{D}$, adjoint dictionary $\mathbf{D}^\mathsf{T}$ and Gaussian denoiser $\mathfrak{G}_g$ are all pluggable. Fig. 2 shows the unrolled network architecture, where we propose to replace the soft-thresholding in LISTA with the interscale LET.

LET provides a shrinkage process that can optimize the shape simulating a soft threshold. Luisier *et al.* proposed the point-wise thresholding:

$$\mathcal{T}(x; \{a_k\}_k) = \sum_{k=1}^{K} a_k x e^{-(k-1)\frac{x^2}{2\tau^2}}, \tag{13}$$

where x is a wavelet coefficient. In this function, the authors suggested to use $K = 2$ and $\tau = \sqrt{6}\sigma$, where σ is the standard deviation of noise $\mathbf{w}$ [16]. The denoising process is completely characterized by a set of parameters. The authors of [16] also proposed the following form of the shrinkage function:

$$\mathcal{T}(x, x_p; \{a_k\}_k, \{b_k\}_k) = e^{-\frac{x_p^2}{12\sigma^2}} \sum_{k=1}^{K} a_k x e^{-(k-1)\frac{x^2}{12\sigma^2}} + \left(1 - e^{-\frac{x_p^2}{12\sigma^2}}\right) \sum_{k=1}^{K} b_k x e^{-(k-1)\frac{x^2}{12\sigma^2}}, \tag{14}$$

where x_p is an interscale prediction of x obtained from the wavelet parent-child relationship. The authors use the parent x_p as a discriminator between high and low SNR wavelet coefficients. The parameters a_k and b_k are linearly solved for minimizing some cost function. In order to use interscale LET, it is necessary to compensate for the group delay characteristics between scales. In this paper, we use NSOLT [17], which guarantees the symmetry of filter kernels, as a convolutional dictionary $\mathbf{D}$ to avoid group delay compensation. NSOLT can construct a wavelet-like hierarchical structure that guarantees the no-DC-leakage property [15]. Fig. 3 illustrates a synthesis NSOLT architecture in tree level 3. All design parameters of the synthesizer, analyzer, and their component layers are kept independent so that the loop unrolling network can be trained by the backpropagation method [18]. The parameters a_k, b_k, and σ of the interscale LET are also trained independently.

IV. Performance Evaluation

In this section, we conduct image restoration to confirm the effectiveness of the proposed method. The configurations of NSOLT and the proposed network used in this simulation are shown in Tab. I, where SFTH denotes the soft thresholding. From Tab. I, the number of parameters in the proposed method is less than 17k, while that in MobileNet [19], a small-scale

Algorithm 3 Inference by PGD w/ synthesis dictionary

Input: $\mathbf{v}$
Output: $\mathbf{u}^{(T)}$

$\mathbf{u}^{(0)} = \mathbf{P}^\intercal \mathbf{v}$
$\mathbf{x}^{(0)} = \mathbf{D}^\intercal \mathbf{u}^{(0)}$
for $t = 0$ to $T - 1$ **do**
  $\mathbf{r} = -\gamma \mathbf{P}^\intercal (\mathbf{P}\mathbf{u}^{(t)} - \mathbf{v})$
  $\mathbf{c} = \mathbf{D}^\intercal \mathbf{r}$ ▷ Analysis process
  $\mathbf{x}^{(t+1)} = \mathfrak{G}_g \left(\mathbf{x}^{(t)} + \mathbf{c}, \sqrt{\gamma} \right)$ ▷ Gaussian denoising
  $\mathbf{u}^{(t+1)} = \mathbf{D}\mathbf{x}^{(t+1)}$ ▷ Synthesis process
end for

TABLE I
CONFIGURATION OF THE UNROLLED PGD METHOD AND NSOLT

Unrolled PGD	
# of blocks (Iterations)	3
Thresholding	SFTH, LET
NSOLT	
# of channels	$8+8$
Decimation factor	2×2
Polyphase order	$2+2$
No DC-leakage	True
# of Tree levels	4

network, is around 2M. In addition, parameters in proximal operators such as λ and σ are to be learned. For training, we use a patch decomposition of a set of case study images in Fig. 4. Training specifications are shown in Tab. III.

A. Denoising

In denoising, the measurement process $\mathbf{P}$ is the identity, and the observation image is contaminated only by AWGN $\mathbf{w}$. In this evaluation, we assume AWGN $\mathbf{w}$ with standard deviation of $\sigma_\mathrm{w} = 50/255$.

The simulation results are shown in Fig. 5 and Tab. IV. For comparison, we adopted Approximate Convolutional Sparse Coding (ACSC) [2], a noise restoration method based on LISTA, Denoising convolutional neural network (DnCNN) [6], a CNN-based restoration method, and DRUNet [20], a method integrating residual blocks into U-Net for effective denoiser prior modeling. ACSC, DnCNN and DRUNet are trained under the specifications shown in Tab. III. The number of parameters for each network is shown in Tab. II. From Fig.5, we see that high-quality recovery is possible even with fewer than 17k parameters. The performance of the proposed method compared to ACSC and DRUnet indicates that there is room for improvement in the configuration of the dictionary $\mathbf{D}$. The comparison between soft thresholding and inter-scale LET shows the effectiveness of the latter.

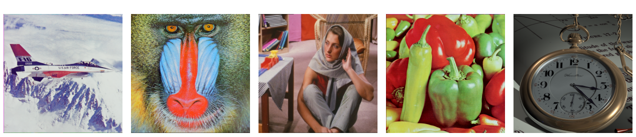

Fig. 4. Case study image set. The images are degraded, and patches of size 128×128 are extracted for training.

TABLE II
COMPARISON OF THE NUMBER OF PARAMETERS BETWEEN THE CONVENTIONAL AND PROPOSED METHOD

Network name	# of learnable parameters	
	Denoising	SISR
DnCNN	555k	-
ACSC	45k	-
CSCN	-	42k
VDSR	-	665k
DRUnet	33M	
Soft Th.	15k	
LET	17k	

TABLE III
TRAINING SPECIFICATIONS

Optimizer	Adam
# of epochs	50
Patch size	128×128
Batch size	16
# of patchs per image	128
# of images	5
# of iterations	2000

B. Single image super resolution (SISR)

In single image super resolution (SISR), the measurment process $\mathbf{P}$ is modeled by convolution and downsampling. For comparison, we adopted the cascade of sparse coding based network (CSCN) [21], a noise restoration method based on LISTA, Very Deep Super Resolution (VDSR) [22], a CNN-based restoration method and DPIR [20], a detailed algorithm of HQS-based plug-and-play IR with deep denoiser prior. ACSC and CSCN are trained under the specifications shown in Tab. III. DPIR used DRUnet as a denoiser. The number of parameters for each network is shown in Tab. II. In this evaluation, $\mathbf{P}$ is set to Gaussian filter of size 9×9 with a standard deviation of $\sigma_\mathrm{p} = 2$ downsampling with factor 4×4, and $\mathbf{w}$ is set to AWGN with standard deviation of $\sigma_\mathrm{w} = 20/255$, respectivery.

The simulation results are shown in Fig. 6 and Tab. V. From Fig.6, CNN and conventional LISTA-based recovery methods showed little difference from bicubic. In these methods, the observation process $\mathbf{P}$ cannot be explicitly set. This result can be due to the fact that complex observation processes require a large amount of training images. The proposed method solves this problem by explicitly setting the observation process $\mathbf{P}$, and is able to extract image features even with a small number of training images. The interscale LET further shows a better restoration result than conventional soft thresholding.

TABLE IV
QUANTITATIVE RESULTS IN PSNR FOR DENOISING, WHERE UPGD DENOTES THE UNROLLED PGD.

Image	Fruits	Girl	Monar.	Parrot	Tulips	Ave.
DnCNN	25.60	25.40	25.51	26.31	25.22	25.62
ACSC	**28.31**	27.67	26.92	28.70	26.84	27.69
DRUnet	28.21	**27.75**	**27.72**	**28.78**	**26.92**	**27.88**
UPGD-SFTH	26.31	25.78	26.39	27.06	25.96	26.31
UPGD-LET	27.17	26.56	27.20	27.87	26.79	27.12

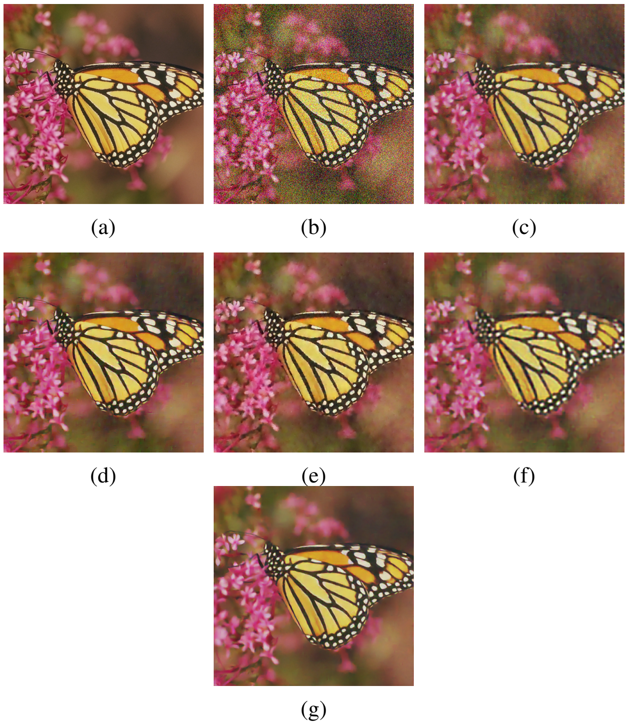

Fig. 5. Denoising results. (a) ground truth, (b) noisy observation with PSNR: 14.88dB, (c) DnCNN with PSNR 25.51dB, (d) ACSC with PSNR 26.92dB, (e) DRUnet with PSNR 27.72dB (f) soft-thresholding network result with PSNR: 26.39dB and (g) LET network result with PSNR: 27.20dB.

TABLE V
QUANTITATIVE RESULTS IN PSNR FOR SISR, WHERE UPGD DENOTES THE UNROLLED PGD.

Image	Fruits	Girl	Monar.	Parrot	Tulips	Ave.
Bicubic	20.77	21.38	19.31	21.33	19.74	20.51
CSCN	20.89	21.39	19.34	21.34	19.78	20.54
VDSR	20.82	21.88	19.32	21.31	19.77	20.51
DPIR	22.72	23.08	21.82	22.93	22.56	22.62
UPGD-SFTH	23.28	24.73	21.66	24.28	22.54	23.30
UPGD-LET	**24.30**	**25.43**	**22.50**	**24.97**	**23.37**	**24.11**

V. CONCLUSIONS

In this study, we proposed an image restoration technique by introducing an interscale thresholding technique with a structured convolutional dictionary into a loop unrolling network based on the PGD method. From some simulation results, we verified that comparable or better performance was obtained with fewer learning parameters. Especially, the experiments showed the effectiveness of the proposed method when the observation process $\mathbf{P}$ has degradation and is identified. However, the denoising exeriments showed that there is room for consideration in the generative process. It also should be noted that the gradient calculation of the parameters of NSOLT takes a long time, and the learning time has not been reduced yet. We plan to further improve the performance of the proposed method in terms of quality and speed by improving the dictionary structure and implementation.

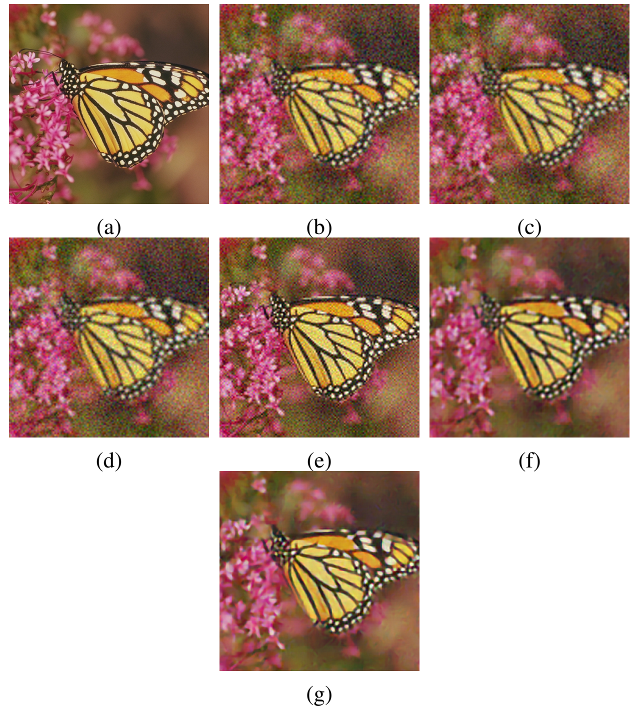

Fig. 6. SISR results. (a) ground truth, (b) bicubic interpolation with PSNR: 20.77dB, (c) CSCN result with PSNR 20.89dB, (d) VDSR result with PSNR 20.82dB, (e) DPIR with PSNR 22.72dB (e) soft thresholding network result with PSNR 21.66dB and (f) LET network result with PSNR: 24.30dB.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number JP19H04135.

REFERENCES

[1] Vishal Monga, Yuelong Li, and Yonina C. Eldar, “Algorithm Unrolling: Interpretable, Efficient Deep Learning for Signal and Image Processing,” *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 18–44, 3 2021.

[2] Hillel Sreter and Raja Giryes, “Learned Convolutional Sparse Coding,” in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. 9 2018, vol. 2018-April, pp. 2191–2195, Institute of Electrical and Electronics Engineers Inc.

[3] Ingrid Daubechies, Michel Defrise, and Christine De Mol, “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 7 2003.

[4] Amir Beck and Marc Teboulle, “A fast iterative shrinkage-thresholding algorithm with application towavelet-based image deblurring,” *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 693–696, 2009.

[5] BoydStephen, ParikhNeal, ChuEric, PeleatoBorja, and EcksteinJonathan, “Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers,” *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 1 2011.

[6] 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, 8 2016.

[7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 6 2017.

[8] Karol Gregor and Yann LeCun, "Learning Fast Approximations of Sparse Coding," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, Madison, WI, USA, 2010, ICML'10, pp. 399–406, Omnipress.

[9] Xiaohan Chen, Jialin Liu, Zhangyang Wang, and Wotao Yin, "Theoretical Linear Convergence of Unfolded ISTA and its Practical Weights and Thresholds," *Advances in Neural Information Processing Systems*, vol. 2018-December, pp. 9061–9071, 8 2018.

[10] Jialin Liu, Xiaohan Chen, Zhangyang Wang, and Wotao Yin, "ALISTA: analytic weights are as good as learned weights in lista," 6 2019.

[11] Atsushi Nitanda, "Stochastic Proximal Gradient Descent with Acceleration Techniques," *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[12] Morteza Mardani, Qingyun Sun, Shreyas Vasawanala, Vardan Papyan, Hatef Monajemi, John Pauly, and David Donoho, "Neural Proximal Gradient Descent for Compressive Imaging," *Advances in Neural Information Processing Systems*, vol. 2018-December, pp. 9573–9583, 6 2018.

[13] Robert Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[14] Shunsuke Ono, "Primal-Dual Plug-and-Play Image Restoration," *IEEE Signal Processing Letters*, vol. 24, no. 8, pp. 1108–1112, 8 2017.

[15] Shogo Muramatsu, "Structured dictionary learning with 2-D non-separable oversampled lapped transform," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 2624–2628, 2014.

[16] Florian Luisier, Thierry Blu, and Michael Unser, "A new SURE approach to image denoising: Interscale orthonormal wavelet thresholding," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 593–606, 3 2007.

[17] Shogo Muramatsu, Kosuke Furuya, and Naotaka Yuki, "Multidimensional Nonseparable Oversampled Lapped Transforms: Theory and Design," *IEEE Transactions on Signal Processing*, vol. 65, no. 5, pp. 1251–1264, 3 2017.

[18] Shogo Muramatsu, Masaki Ishii, and Zhiyu Chen, "Efficient parameter optimization for example-based design of nonseparable oversampled lapped transform," *Proceedings - International Conference on Image Processing, ICIP*, vol. 2016-August, pp. 3618–3622, 8 2016.

[19] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, Quoc V. Le, and Hartwig Adam, "Searching for MobileNetV3," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2019-October, pp. 1314–1324, 5 2019.

[20] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte, "Plug-and-Play Image Restoration with Deep Denoiser Prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, p. 1, 2021.

[21] Zhaowen Wang, Ding Liu, Jianchao Yang, Wei Han, and Thomas Huang, "Deep Networks for Image Super-Resolution with Sparse Prior," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 370–378.

[22] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1646–1654.