






# DSPNet: A Lightweight Dilated Convolution Neural Networks for Spectral Deconvolution With Self-Paced Learning

Hu Zhu , Member, IEEE, Yiming Qiao , Guoxia Xu , Member, IEEE, Lizhen Deng , Member, IEEE, and Yu-Feng Yu , Member, IEEE

**Abstract**—In the fields of industry research, infrared spectrometers are widely used in diverse applications. However, the spectrum often suffers from band overlap and random noise due to the distortion caused by the point spread function, especially for aging instruments. The problem of reconstructing the clear spectrum from the degraded spectrum is called spectrum deconvolution. Traditional partial differential equation (PDE) methods rely on distribution assumptions in the reconstructed process. This restriction makes PDE methods sensitive to tackle complex instrumental broadening effect in the dispersive IR spectrometers. Also, we need to spend much time setting the parameters of PDE models manually. These problems intuitively degrade the performances of PDE methods. In this article, we propose an end-to-end neural network framework for spectral deconvolution problem. The novelty of this article lies in its strong robustness from dilated deconvolution and self-paced learning procedure to challenge the complicated degraded spectra. Actually, the deconvolution problem is tailored to a dense prediction problem in this article. Inspired by the extensive use and excellent effects of dilated convolutions in dense prediction, a lightweight dilated convolution module is given to detect the overlaps of degraded spectra. Experimental results demonstrate that the proposed solution has an outstanding performance against many other approaches.

Such improvements have the potential to facilitate industrial applications and further exploration of an unknown chemical mixture. Our framework has a good performance on feature extracting and spectrum reconstruction, even in the case of low signal-to-noise ratio.

**Index Terms**—Convolution neural network (CNN), infrared (IR) spectrometer, spectral deconvolution.

## I. INTRODUCTION

INFRARED (IR) spectroscopic is one of the most vital tools in analyzing the chemical structure. It relies on its great accuracy and high efficiency in the judgment of molecular structure and bonding. In the past few decades, it has achieved a multitude of excellent achievements. For example, detecting gas leaks [1], educational robot IR vision sensing [2], and investigation of chemical structures [3].

Although IR spectra are useful and powerful, they often suffer from band overlaps and random noise due to the distortion from the broadening effect of the point spread function (PSF), especially for aging instruments. To tackle it, deconvolution problem is proposed [4]. It is well known that deconvolution is a severely ill-posed problem due to the existence of random noise [5]. The goal of deconvolution is to remove the broadening effect of PSF, e.g., Gaussian blur kernel and Lorentzian blur kernel. Generally, almost all spectral deconvolution problems can be divided into three types: blind deconvolution (BD), non-BD (NBD), and semi-BD (SBD) [6]. BD methods never have any information about the blur kernel, SBD indicates that the kernel is given without specific parameters, and NBD-based methods can operate the given blur kernel for deconvolution.

There are two main directions to resolve the deconvolution problem, which both rely on prior knowledge to some extent. Filtering methods are proposed first and easy to implement, but the performance is easy to be affected by random noise [5]. They are a branch of SBD, and this kind of methods could not be applied if the blur kernel is unknown completely. Meanwhile, partial differential equation (PDE) methods are good at designing various regularization terms according to spectrum priors, the design of regularization terms of optimization problem is highly dependent on the analysis of the characteristics of the blur kernel [6]. Due to the deficiency of prior knowledge, the existing deconvolution methods often should be given certain

Manuscript received August 17, 2019; revised September 18, 2019 and December 2, 2019; accepted December 9, 2019. Date of publication December 19, 2019; date of current version September 18, 2020. This work was supported by the National Natural Science Foundation of China under Grant 61701259. Paper no. TII-19-3790. (Hu Zhu and Yiming Qiao contributed equally to this work.) (Corresponding author: Lizhen Deng.)

H. Zhu is with the Jiangsu Province Key Laboratory of Image Processing and Image Communication, Nanjing University of Posts and Telecommunications, Nanjing 210003, China (e-mail: peter.hu.zhu@gmail.com).

Y. Qiao is with Bell Honors School, Nanjing University of Posts and Telecommunications, Nanjing 210003, China (e-mail: yimingqiao3163@gmail.com).

G. Xu is with the Department of Imaging and Interventional Radiology, The Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: gxu.re@gmail.com).

L. Deng is with the National Engineering Research Center of Communication and Network Technology, Nanjing University of Posts and Telecommunications, Nanjing 210003, China (e-mail: alicedenglzh@gmail.com).

Y.-F. Yu is with the Department of Statistics, Guangzhou University, Guangzhou 510006, China (e-mail: yuyufeng220@163.com).

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TII.2019.2960837

preconditions against complex spectral characteristics in the real world [5]. Besides, the high computational complexity of PDE methods is also a problem, which is caused by the restriction of fine-tuning the parameters in the model manually. These limitations weaken the generality and adaptability of PDE methods in real industrial applications.

More recently, the convolution neural network (CNN) is a new trend in signal processing [7], [8]. It is noticeable that the neural networks are dependent on the data-driven manner for training according to the sample data automatically and adaptively. Among the diverse neural networks, dilated mechanism has drawn much attention in the community. Dilated convolution layer is a derivation of convolution layer for enhancing the perception of learning patterns, which could intuitively extract the information from diverse scales, and expand the receptive field of feature map [7]. This technology is usually applied in natural language processing and dense prediction problems [9]. In spectral deconvolution, we need to distinguish the adjacent peaks and split the overlap, which agree with the intention of dilated CNN. Another novel component of our motivation is self-paced learning (SPL). It is a new type of curriculum learning methodology by stimulating the learning process of human being [10]. The SPL-based realization scheme has been designed for computer vision and pattern recognition. This method has a cutting edge to find a good local minimum for achieving a better generalization ability [11].

In this article, we focus on the CNN learning based method to break the reliance on prior knowledge and empirical handcrafted design of the model. Our proposed method could extract features and put it into subsequent use automatically instead of hand-engineered. In general, this article commences a new direction of spectrum BD. Rather than depending on mathematical and physical characteristics of spectral theories, the new direction of spectral deconvolution is a data-driven system that only relies on spectrum samples and has potential to achieve better performance. Experiments show its robustness and high efficiency. In summary, the main contributions can be summarized as follows.

- 1) A data-driven CNN, namely DSPNet, is proposed for spectral deconvolution problem, as illustrated in Fig. 1. DSPNet is the first method to tackle spectrum deconvolution problem from the perspective of deep learning. Compared with past works, DSPNet can extract features in an end-to-end manner without hand-interfering. In addition, the proposed end-to-end lightweight network architecture can achieve good performance even under complicated conditions without too much distribution assumptions.
- 2) Specifically, instead of using normal layer, a dilated convolution layer is employed for spectrum processing, which is helpful to achieve exponentially expanding receptive fields without losing precision and coverage [12]. Meanwhile, the spatial positional relationship of the spectrum could be considered, which makes it easier to detect the overlap from the gradient information.
- 3) An objective function is constructed to be embedded into the SPL framework, in which the heuristic methodology of SPL is used for training the neural networks. The enhancement accelerates the speed of function

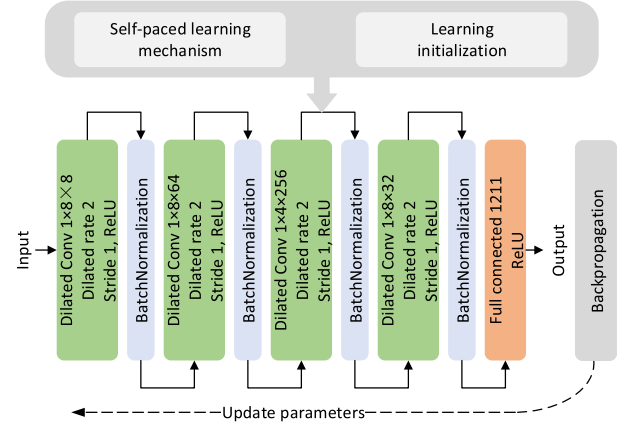


Fig. 1. Framework of DSPNet for spectrum deconvolution with dilated convolution layer and SPL. After the parameters initialization, the instances are selected by SPL mechanism for each iteration. Thus, DSPNet can learn simple and universal knowledge structures firstly. Common convolution layer are replaced by the dilated convolution layer for larger receptive fields.

convergence. Empirical results show that the better generalization ability is achieved, DSPNet can have a more satisfied performance on the validation set.

## II. RELATED WORKS

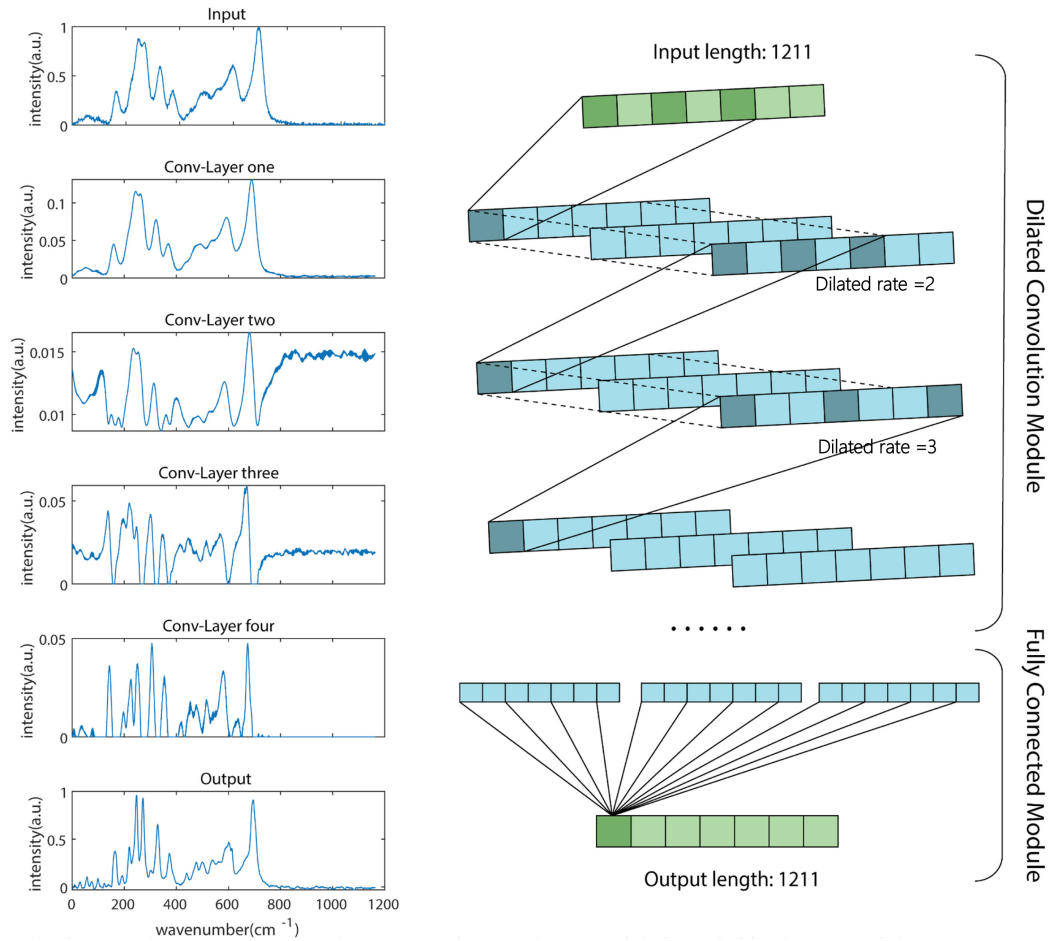
Given the clear spectrum  $y(u)$ , nonnegative blur kernel  $h(u)$ , and random noise  $\epsilon(u)$ ,  $u$  denotes the wavenumber,  $\otimes$  denotes the convolution operation, and the degraded spectrum  $x(u)$  is defined by

$$x(u) = y(u) \otimes h(u) + \epsilon(u). \quad (1)$$

Many researches focus on optimizing the function  $E = \frac{1}{2} \|y(u) \otimes h_{\theta}(u) - x(u)\|_2$ , where  $h_{\theta}$  indicates the pseudoinverse kernel. Spectrum deconvolution is an ill-posed problem due to the existence of  $\epsilon(u)$ .

To address this problem, many methods, which are proposed, can be roughly divided into following three categories.

- 1) *Filtering methods*: For spectral deconvolution problem, the easiest way is to do inverse filtering directly. One of the simplest of filtering methods is Fourier self-deconvolution [13], but noise limits its effect and performance in many cases. As to the Wiener deconvolution method, the spectrum degradation is assumed as a wide-sense stationary process strictly. Richardson–Lucy method imposed a Poisson distribution on signal noise [12]. Hence, the initial setting of a blur kernel should be given for implementation, which limits its applicability greatly.
- 2) *PDE methods*: In the recent years, deconvolution methods based on PDE are popular, such as fast blind instrument function estimation method proposed by Liu *et al.* [5]. In the deconvolution with modified Tikhonov regularization method [14], an adaptive regularization term is introduced to distinguish the certain specific region from others in the spectrum. The aforementioned two methods assume that



**Fig. 2.** Overall schematic diagram of the network structure. The convolution module has four hidden layers. Each layer connects to a batch normalization layer. The number of convolution kernels in the convolution module increases layer by layer, which are 8, 64, and 256. A flatten layer is connected between two modules, and the fully connected layer finally outputs the restored spectrum.

the noise obeys Poisson distribution. Hence, the performance may not be as good as expected if noise exists. To preserve spectral details and enhance the stability of kernel estimation, trimmed  $\varphi_{HS}$  regularization and weighted  $\varphi_{HS}$  regularization are proposed in SBD, but the same problem still exists [6]. Although SBD methods do not need to know the value of PSF in advance, they still rely on a PSF assumption (i.e., it is a Gaussian blur kernel). It should be noted that all of these nonblind and semiblind methods have a limited range of application due to diverse assumptions. These problems motivate us to find a solution, which is robust, no matter the kinds of blur kernel and noise.

- 3) *CNN methods*: Deep learning has attracted much attention of signal processing researchers for the amazing performance and applicability in recent years. The end-to-end network architecture contains some convolution layers, which extracts the knowledge from priors in the offline training process. For these advantages, it has greatly accelerated the development and progress of all aspects of image processing and achieved great success [9]. Xu *et al.* [15] proposed a CNN to deblur the image, this network

uses a small convolution kernel and deeper convolution layer to get the same effect as the big kernel has. Different from the work in [15], DSPNet makes use of SPL and has relatively shallower architecture, which helps DSPNet to perform spectrum deconvolution operator.

Dilated convolution layer is an improvement on the common convolution layer, which aims to increase the receptive field. Yu and Koltun found the fact that dilated convolution can increase the resolution of output feature maps without reducing the receptive field of individual neurons [7]. According to the original definition by Yu and Koltun, let  $F : \mathbb{Z} \rightarrow \mathbb{R}$  be a discrete function. Let  $\Omega_r = [-r, r] \cap \mathbb{Z}$ ,  $r \in \mathbb{Z}_+$  and  $k : \Omega_r \rightarrow \mathbb{R}$  be a discrete filter of size  $(2r + 1)$ .  $l$  denotes the dilated rate, then the dilated convolution  $\otimes_l$  is defined by

$$(F \otimes_l k)(p) = \sum_{s+lt=p} F(s)k(t) \quad (2)$$

where  $s, t$ , and  $p$  are independent variables of corresponding function. In the right part of Fig. 2, the process of dilated convolution on a spectrum is illustrated by dark blue squares.

Inspired by the aforementioned analysis, we make use of the dilated convolution layer to detect the high-order difference

sequence information. We find that the difference sequence of spectrum data is an important indicator for peaks. Each spectrum could be considered as a discrete sequence, and the difference sequence near the overlaps has a similar pattern, which is different from that of the flat region. Besides, we design an objective function according to the methodology of SPL, which accelerates the speed of function convergence and is able to gain better generalize results. Compared with the filtering and PDE methods, DSPNet has following advantages.

- 1) DSPNet does not need any kernel distribution assumption.
- 2) DSPNet has robust performance under complex environment.
- 3) It is an end-to-end network and easy to apply without any hand-engineered work.

### III. PROPOSED DSPNET FOR SPECTRUM DECONVOLUTION

#### A. Overview

For past PDE methods, it is so sensitive of spectrum that could be disturbed and the band-side artifacts always challenge the performance. It is important to adjust the parameters due to the formidable energy function with nonlinear degradation [8]. Luckily, DSPNet has the potential property to deal with these problems from a deep learning manner.

First, the detailed layers and parameters of the proposed network are shown in Fig. 1. Second, the objective function in SPL framework  $\mathbb{F}$  is introduced, which is important in our network. Third, we introduce the training and analysis of DSPNet.

#### B. Network Architecture

The motivation of architecture of the DSPNet is the CNN for image deblur proposed by Xu *et al.* [15], which consists of two modules: deconvolution module and artifact removal module. Moreover, batch normalization layer is added to increase the convergence speed [16], and we use multiple small kernels to simulate large kernel [17].

The overall schematic diagram of the network structure is shown in Fig. 2. The input of this network is a degraded spectrum. The convolution module is responsible for extracting information for the subsequent module from multiple scales. The recovery effect is associated with the completeness of information extraction. The fully connected layer is responsible for reconstructing the peaks and valleys in the original spectrum from the information extracted from the degraded spectrum. In each iteration, each layer selects the less degraded spectra according to the SPL mechanism. Moreover, different scale features are interacted via a dilated convolution operation. We assume that  $x_j^\ell$  is the  $j$ th output map in the  $\ell$ th layer, and  $k_{ij}^\ell$  denotes the convolution kernel in the  $\ell$ th layer. Each output map is given an additive bias  $b_j^\ell$ ,  $\sigma$  indicates the activation function ReLU,  $f_{\text{BN}}(\cdot)$  denotes the batch normalization operation, and  $M_j$  represents a selection of input maps. Mathematically, the dilated convolution module of out spectrum deconvolution network can be expressed

as

$$x_j^\ell = f_{\text{BN}}\left(\sigma\left(\sum_{i \in M_j} x_i^{\ell-1} \otimes_l k_{ij}^\ell + b_j^\ell\right)\right). \quad (3)$$

Note that the input maps are convolved with distinct kernels for a particular output map. The fully connected module is as simple as many other works, which is responsible for handling the flatten output of the previous module.

#### C. Objective Function With SPL

Given a DSPNet  $f(\cdot)$ ,  $\omega$  represents all the parameters in the DSPNet,  $N$  denotes the number of instances in dataset,  $y_i$  is the  $i$ th clear spectrum, as described earlier, and  $x_i$  indicates the  $i$ th degraded spectrum, which works as the input of DSPNet, then the basic *Loss* is defined by

$$\text{Loss} = \frac{1}{2N} \sum_i (f(w, x_i) - y_i)^2 + \frac{\alpha}{2} \|w\|_2^2. \quad (4)$$

Apparently,  $\frac{1}{2N} \sum_i (f(w, x_i) - y_i)^2$  is the mean-square error (MSE) term of spectrum pairs,  $\frac{\alpha}{2} \|\omega\|_2^2$  is the  $l_2$  regularization term, which has the ability of limiting the hypothesis space and reducing the risk of over-fitting.

In experimenting, results show that the flat area (actually, the region is numerically close to zero) is disturbed by artificial noise. A penalty term  $\frac{\beta}{2} \sum_i \sum_j |f(w, x_i)_j - a|^2$  is proposed to deal with this problem.  $\beta$  is a hyperparameter, which controls the weight of the penalty term.  $f(w, x_i)_j$  indicates the  $j$ th point value of the  $i$ th spectrum in a training batch.  $a$  represents the average value of the flat area, it is the same for two different flat areas in a certain type of spectrum. In this article, the point values in all flat area are close to zero after normalization, thus  $a = 0$  is considered in this article. The *Loss* is transformed by

$$\begin{aligned} \text{Loss}' &= \frac{1}{2N} \sum_i (f(w, x_i) - y_i)^2 + \frac{\alpha}{2} \|w\|_2^2 \\ &\quad + \frac{\beta}{2} \sum_i \sum_j |f(w, x_i)_j|^2 \\ \text{s.t. } N, i, j &\in \mathbb{N}_+, |f(w, x_i)_j| \leq t \end{aligned} \quad (5)$$

where  $t$  is a threshold, which indicates the intensity range of flat area. For  $f(w, x_i)_j > t$ , it means that the corresponding point in  $y_i$  does not belong to flatten area. The value of point in flat area can be reduced by learning and smoothing the spectrum in this way. Therefore, the goal is to find the  $\omega_{\text{goal}}$  by optimizing

$$\begin{aligned} \omega_{\text{goal}} &= \arg \min_{\omega} \text{Loss} + \frac{\beta}{2} \sum_i \sum_j |f(w, x_i)_j|^2 \\ \text{s.t. } N, i, j &\in \mathbb{N}_+, |f(w, x_i)_j| \leq t. \end{aligned} \quad (6)$$

However, only a sound *Loss'* cannot obtain the best result, we need to embed it into SPL framework.

In this case, our work uses the methodology of SPL in the training of neural networks. SPL is an advanced method designed from the scale of simulating human/animal's learning



principles. The core idea of curriculum learning and SPL is to learn simple and universal knowledge structures by simulating human cognitive mechanisms, and then the learning system could gradually increase the difficulty and transition to learn more complex and specialized knowledge [10].

This improvement could help our proposed model automatically to adjust its course materials and establish a sound SPL system. Thus, our model becomes more comprehensively, which can be presented as a relatively powerful generalization to spectrum deconvolution problem. Here, we reformulate the objective function of CNN with a soft-thresholding variable  $v$  ( $1/0$ ), which is introduced to indicate the fact whether the sample is going to be learned. Thus, the objective function in SPL framework  $\mathbb{F}(v, w; \eta)$  can be expressed as

$$\mathbb{F}(v, w; \eta) = \sum_{k=1}^m |v_k| \text{Loss}'(x_k, y_k, w) - \eta \sum_{k=1}^m (|v_k|^2 + |v_k|). \quad (7)$$

In this objective function,  $\eta$  is a threshold that controls the learning pace (difficulty of learning), which corresponds to the “age” of network and gets larger along with the increasing iteration.  $\sum_{k=1}^m (|v_k|^2 + |v_k|)$  is a regularize term, which consists of  $l_1$ -norm and  $l_2$ -norm of  $v_k$ . The use of norm commonly increases the generalization ability for overparameterized neural network [18].  $v_k$  is defined by

$$v_k = \begin{cases} 1, & \eta > \text{Loss}'(x_k, y_k, w) \\ 0, & \text{otherwise} \end{cases}. \quad (8)$$

The difficulty of samples is quantified by calculating the MSE between the degraded spectrum and the original spectrum. We assume that the MSE is positively related to the severity of spectral degradation. Given  $\eta > 0$ , train samples with bigger loss (the parameter  $\eta$  controls the criteria of selection) are labeled as “not selected,” and others are marked as “selected.” Then, the network updates parameters  $w$  based on the samples selected. In the early stage,  $\eta$  is relatively small (“young” network), the size of “course” (samples labeled “selected”) for the network is a relatively small and lower loss. Along with the increasing of  $\eta$  (grow “mature”), the network could get in touch with more samples with higher loss. Although the “hard” samples are not learned in the early stage, the loss for the “hard” samples is still getting smaller along with the growth of the network due to the generalization of the network.

#### D. Training and Analysis of DSPNet

We train this network on an artificially degraded spectral dataset. Dataset is generated based on the blur degradation model in (1). Original clear spectra are gained from high-precision IR spectrometer. Different kernels are applied to blur these spectra, such as Gaussian blur and Lorentz blur [5]. The noise we add is random white noise. There is a batch normalization layer between two dilated convolution layers. The feature vectors are normalized here, which could accelerate convolution network training by reducing internal covariate shift [16]. After several layers of feature extraction, the fully connected module gets the processed information and reconstructs the spectrum  $\mathbf{S}$ .

Then, we compare the output spectrum  $\mathbf{S}$  with the original clear spectrum  $\mathbf{O}$  by computing objective function and update the parameters by back-propagation. The process is illustrated by Figs. 1 and 2, the detailed computational process is shown as (3).

**1) Dilated Convolution Layer:** Our network consists of four dilated convolution layers and a fully connected layer. Dilated convolution layer is an improvement on the common convolution layer, which aims to increase the receptive field, as described by (2). The architecture relies on the fact that dilated convolution supports for increasing the resolution of output feature maps without reducing the receptive field of individual neurons [7]. In this article, the special convolution layer is used to enhance the multiscale spectral information and integrate into the subsequent network, which extracts hidden details of the original spectrum. This network has a fully connected network after the dilated convolution module to complete an end-to-end design architecture. Actually, the deconvolution problem is tailored to a dense prediction problem in this article. The use of dilated convolution layer facilitates the processing of spectrum recovery. The proposed method shows a better performance at steep derivatives, such as peaks. In deconvolution problem, the detailed information such as peak and valley from the degraded spectrum need to be restored. The peak is often hidden in the details of several adjacent points, which requires higher detail processing ability. The application of the dilated CNN can detect the overlap from the gradient information and achieve precise information.

**2) Small Convolution Kernel:** Meanwhile, it is important to adopt any possible enhancement to accelerate the training of network as the number of parameters in network is usually far more than number of instances. This work uses multiple small kernel convolution layers to achieve the same effect of the relatively large kernel. As Fig. 1 shows, the size of kernels are  $1 \times 8$ ,  $1 \times 8$ ,  $1 \times 4$ , and  $1 \times 8$  for each layer ( $1 \times 8$  is small relative to the length of spectra), whereas the number of kernels in each layer is 8, 64, 256, and 32, respectively. This enhancement could speed up the training of this network as introduced in [17]. Compared with traditional convolution network, this network has faster convergence speed and fewer parameters.

## IV. EXPERIMENTAL SETTING AND RESULTS

### A. Experiments Settings

**1) Baseline Methods:** In order to objectively and comprehensively evaluate the proposed approach, this work will compare our proposed approach with other advanced deconvolution methods in recent years, including the following.

- 1) *DCTR-BD* [5]: Fast blind instrument function deconvolution with discrete cosine transform regularization. This method is motivated by the observation that the discrete cosine transform coefficient distribution of the ground-truth spectrum is sparser than that of the observed spectrum.
- 2) *FBPSR* [13]: Fast blind Poissonian spectral reconstruction. This method considers spectrum deconvolution as a maximum a posterior in a loss minimization prototype.

- 3) *SBD-HS* [19]: SBD with  $\phi_{HS}$ . A regularization term in form of a convex function  $\phi_{HS}$  is included in the spectral deconvolution model to enhance spectral resolution.
- 4) *MaxEntD* [20]: Maximum entropy deconvolution approach. This method is outstanding for using generalized negative Burg-entropy (ItakuraSato discrepancy) generalized for difference spectra.

2) *Evaluation Indicators*: This article mainly uses the following indexes to quantitatively measure the quality of the recovered spectra: root MSE (RMSE), Pearsons correlation coefficient (CC), and weighted CC (WCC). The specific formulas and calculation of these evaluation indicators are given by Liu *et al.* [5].

Pearsons CC is used to describe the degree of linear correlation. In this article, it represents the average similarity between the peaks and valleys of the original and reconstructed spectrum. The value of CC is ranging from 0 to 1 and  $CC = 1$  means that it matches seamlessly [14]. The high WCC value can be achieved from the good quality of recovered spectrum. It reflects the recovery effect of peaks and valleys intuitively.

3) *Dataset and Blur Kernel*: Based on (1), we simulate the degraded process on the clear IR spectral (as the output of network) to gain degraded spectra (as the input of network). These clear spectra are gained from the high-precision IR spectrometer. The degradation has two parts: First, convolute clear spectrum with blur kernel. Second, add noise to spectrum after convolution. The number of clear spectra is 880, and the number of samples in the dataset is 10 560 after degrading with different kernels and noise. The different values of  $\rho$  are tried,  $\rho \in \{5, 6, 7\}$  for each kernel, the operation order of convolution is not important for that convolution operation is commutative.

Different from some research works, where the blur kernel is set as Gaussian  $L_\rho(u)$  or Lorentzian  $G_\rho(u)$  [4], not only do we exploit the single kernel for testing, but also further set the mixed kernel to evaluate the robustness of all compared approaches and our proposed method. The definitions of Gaussian and Lorentzian blur kernels are shown in (9) and (10), respectively

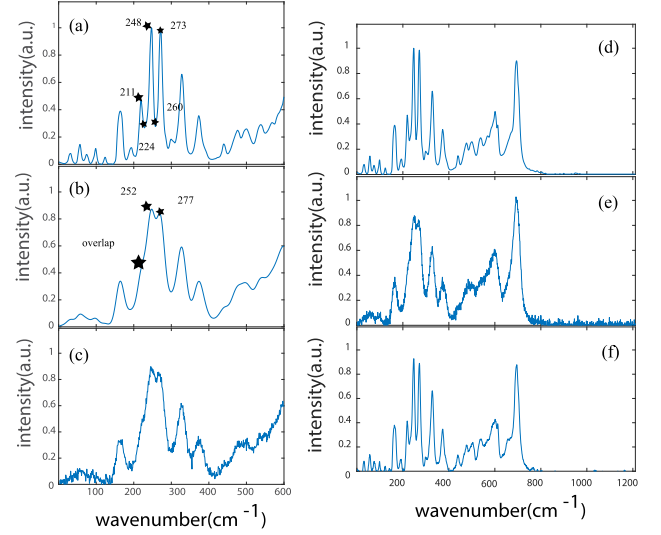
$$L_\rho(u) = \frac{1}{\pi} \frac{\rho}{(u^2 + \rho^2)} \quad (9)$$

$$G_\rho(u) = \frac{1}{\sqrt{2\pi}\rho} \exp\left(-\frac{u^2}{2\rho^2}\right) \quad (10)$$

where  $\rho$  is the half width at half maximum and  $u$  denotes the spectral data point. The discrete support of kernel is set to  $6\rho + 1$ . In these experiments, the support is far smaller than spectrum size. Besides, the signal-to-noise ratio (SNR) of spectrum is 27.

The motivation of using mixed kernel is that many classic approaches work well under strict kernel distribution assumptions, whereas our proposed approach does not need them. We want to find whether these previous works can still have a good performance under complex environment.

4) *Hyperparameters of Networks*: TensorFlow 1.4.1 is used to implement for our experiments, and we utilize GPU computation with NVIDIA GTX 1050Ti to accelerate computation. The optimizer of the network is adaptive moment estimation (Adam) with a batch size of 32, the initial learning rate is 0.001 with exponential decay, and the decay rate is 0.985, which



**Fig. 3.** IR spectra data degradation and recovery process (single blur kernel). (a) Part of clear IR spectrum. (b) Part of the convoluted IR spectrum. (c) Part of the convoluted spectrum white noise (degraded spectrum). (d) Clear IR spectrum. (e) Degraded IR spectrum. (f) Spectrum recovered by DSPNet ( $l = 2$ ).

are empirically determined. Other hyperparameters are default. The wavenumber of input and output spectrum are 1211. The proportion of training and testing set is 7:3 and partition is performed by random completely.

### B. Single Blur Kernel Spectral Deconvolution Experiments

The simulation degraded process is illustrated in Fig. 3(a)–(c) and they show the spectrum from 1 to 600  $\text{cm}^{-1}$  at the resolution of 1  $\text{cm}^{-1}$ . Fig. 3(b) is the convolution result by a Gaussian kernel with  $\rho = 9$ . Fig. 3(c) is the noisy degraded spectrum with SNR=27 white noise. The comparison between Fig. 3(a) and (b) shows that it is extremely hard to distinguish the peaks and valleys at 211, 248, and 273  $\text{cm}^{-1}$ . The valley at 260  $\text{cm}^{-1}$  almost disappears and the valley at 224  $\text{cm}^{-1}$  cannot be distinguished completely. The degraded spectrum is obviously much smoother than the original spectrum.

We choose a small segment of the degraded spectrum that is to be deconvoluted by four different methods. The effect map is shown in Fig. 4. We can see that in Fig. 4(a) and (b), the recovery spectrum does not split the overlap well compared with the original spectrum, and there are large errors from the naked eye between the peak and the trough. In Fig. 4(c), the restored spectrum and the original clear spectrum are almost coincident, whereas there is residual noise in the flat area of reconstructed spectrum. In Fig. 3(a), we choose three peaks from the original spectrum as the ground-truth peaks, the positions of them are 79, 93, and 123  $\text{cm}^{-1}$ , respectively. In Fig. 4, we demonstrate the comparison between the recovery spectrum and the original spectrum of several methods by visualization, especially the change of the peaks (position and intensity) in Table I. These pictures could describe the fact that DSPNet could get the best and delicate details among these four methods. According to

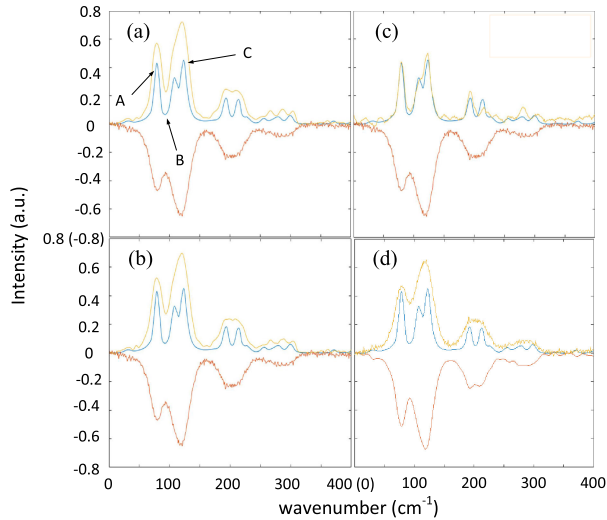


Fig. 4. Spectrum experiments under strong noise condition. The blue line indicates clear spectra, red line denotes degraded spectra, and yellow line indicates recovery spectra. (a) Performance of DCTR-BD. (b) Performance of FBPSR. (c) Performance of DSPNet (rate = 2). (d) Performance of SBD-HS.

TABLE I

PEAK DISTORTIONS COMPARISON IN SPECTRA DECONVOLUTION IN FIG. 4

Peak positions	A	B	C	RMSE
Position <sup>1</sup>				
DCTR-BD	-1	+2	-3	2.1602
FBPSR	-1	+1	-2	1.4142
SBD-HS	-1	+2	-3	2.1602
DSPNet	<b>0</b>	<b>+2</b>	<b>0</b>	<b>1.1547</b>
Intensity <sup>2</sup>				
DCTR-BD	+0.092	+0.204	+0.244	0.1912
FBPSR	+0.138	+0.442	+0.271	0.3089
SBD-HS	+0.072	+0.197	+0.321	0.2214
DSPNet	<b>-0.021</b>	<b>+0.014</b>	<b>+0.048</b>	<b>0.0313</b>

<sup>1</sup> Obtained from the band maximum ( $\text{cm}^{-1}$ ).

<sup>2</sup> "+" or "-" indicates larger or smaller than the original, respectively.

The boldface to show the performance of our proposed method compared with other methods.

these pictures and tables, our data-oriented approach has better performance than the traditional methods based on PDE and others.

### C. Mixed Blur Kernel Spectral Deconvolution Experiments

We design the mixed blur kernel-degraded spectral deconvolution experiment with low SNR to verify the robustness of DSPNet, as shown in Table II. It is a comparison table of the performances for different methods under various conditions. The spectrum is degraded with the Gaussian and Lorentzian blur kernel together. Additionally, it is clear that our method exhibits excellent performance in the case of mixed convolution kernels in Table III, no matter noise is added or not.

We present the comprehensive performance of the DSPNet with various mixed blur kernels in the case of strong noise or not. The spectral dataset has 28 spectra, each length is  $1211 \text{ cm}^{-1}$ , and the average results are taken. It can be found that the proposed method is robust. We also give the comparison pictures

TABLE II  
PERFORMANCE OF DSPNET IN VARIOUS SITUATION

$\rho_{Gau}$	$\rho_{Lor}$	Noise level	RMSE	CC	WCC
5	7	Noise-free	0.0191	0.9959	0.9981
		SNR=27	0.0197	0.9948	0.9977
5	6	Noise free	0.0196	0.9949	0.9976
		SNR=27	0.0198	0.9935	0.9971
6	5	Noise free	0.0203	0.9944	0.9972
		SNR=27	0.0214	0.9930	0.9967
6	6	Noise free	0.0194	0.9958	0.9980
		SNR=27	0.0198	0.9945	0.9975
7	5	Noise free	0.0197	0.9956	0.9979
		SNR=27	0.0197	0.9945	0.9974
7	6	Noise free	0.0199	0.9958	0.9977
		SNR=27	0.0200	0.9948	0.9973
7	7	Noise free	0.0210	0.9948	0.9966
		SNR=27	0.0215	0.9933	0.9958

TABLE III

AVERAGE PERFORMANCES COMPARISON UNDER MIXED KERNELS ( $\rho_{gau} = \rho_{lor} = 6$ )

Noise level	Method	RMSE	CC	WCC
Noise- free	DCTR-BD	0.1168	0.9243	0.9387
	FBPSR	0.1208	0.9428	0.9479
	SBD-HS	0.0702	0.9694	0.967
	SBD-MHS	0.0697	0.9811	0.9803
	MaxEntD	0.0719	0.9953	0.9952
	DSPNet ( $l=5$ )	0.0262	0.9735	0.9911
SNR=27	DSPNet ( $l=2$ )	<b>0.0194</b>	<b>0.9958</b>	<b>0.998</b>
	DCTR-BD	0.1172	0.9224	0.9378
	FBPSR	0.1215	0.9398	0.9464
	SBD-HS	0.0706	0.9682	0.9659
	SBD-MHS	0.07	0.9783	0.9774
	MaxEntD	0.0732	0.9941	0.9934
	DSPNet ( $l=5$ )	0.0281	0.9729	0.9904
	DSPNet ( $l=2$ )	<b>0.0198</b>	<b>0.9945</b>	<b>0.9975</b>

The boldface to show the performance of our proposed method compared with other methods.

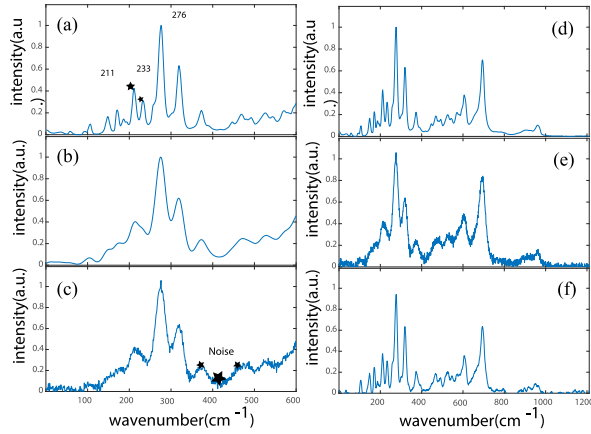
to illustrate the performance of DSPNet in the right part of Figs. 5 and 6. There are some overlaps in the degraded spectrum. It is difficult for human to distinguish the overlaps and split them but DSPNet performs well. The effect of DSPNet is satisfied by comparing the original and reconstructed spectrum.

### D. Real IR Spectrum Experiments

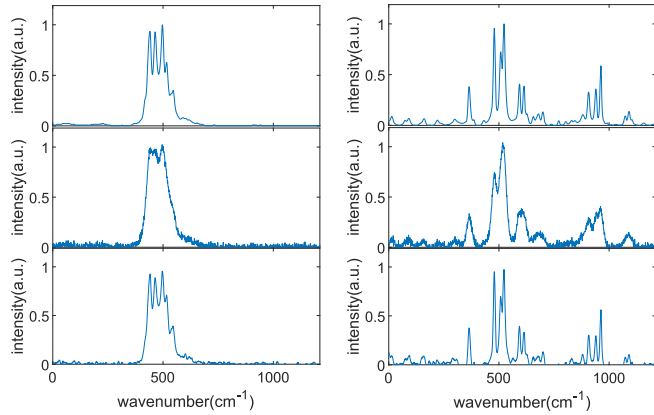
In the final stage of this work, we conduct a real spectral experiment test on the effectiveness of our proposed method. These real spectra are obtained from aged Fourier-transform IR spectrometer.

1) *Real Degraded Spectrum*: These two spectra are good materials for testing the performances of DSPNet. In Fig. 7(a), the marked part of the spectrum is too smooth and the peaks seem to have been hidden seriously. It is extremely possible to find peaks here. Additionally, in Fig. 7(c), there are two obvious overlaps in the spectrum that is marked. We are looking forward to seeing the good performance of DSPNet in these parts.

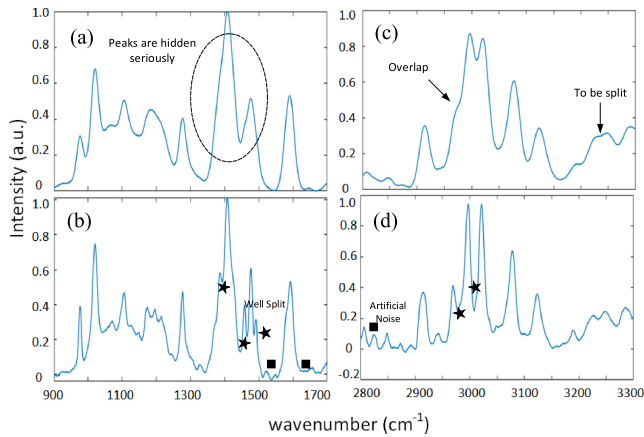
2) *Real Spectrum Results*: As we can see in Fig. 7(b), there are three hidden peaks that are found by DSPNet, and the split peaks are marked with a star. In Fig. 7(d), the original shallow peak is deeper now, and the overlap is split well. The split



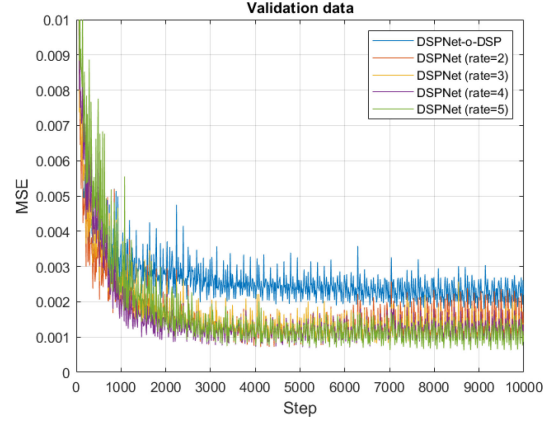
**Fig. 5.** Spectrum experiments under strong noise condition. (Mixed blur kernel). (a) Part of clear IR spectrum. (b) Part of the convoluted IR spectrum. (c) Part of convoluted spectrum with strong white noise (degraded spectrum). (d) Whole clear spectrum. (e) Whole degraded spectrum. (f) Spectrum recovered by DSPNet ( $l = 2$ ).



**Fig. 6.** Two spectra comparisons. The top picture illustrates the original clear spectrum, the middle picture describes the degraded spectrum, and the reconstructed spectrum is shown in the bottom picture.  $\rho_{gau} = \rho_{lor} = 6$ , SNR is equal to 27.



**Fig. 7.** Real Raman spectrum experiments. (a) Part of spectrum L(+)-Arabinofuranose from 2800 to 3300  $\text{cm}^{-1}$ . (b) Result of DSPNet. (c) Part of spectrum D(-)-Ribose from 900 to 1700  $\text{cm}^{-1}$ . (d) Result of DSPNet.



**Fig. 8.** Performances on validation data of different networks.

peaks are beneficial to indicate the property of the spectrum. However, there are still some problems, such as artificial noise and remaining overlap. DSPNet does not get a good grade in the part to be split in Fig. 7(b). There is artificial noise in the reconstructed spectra, which is marked with a square.

In a word, DSPNet could detect most overlap in real spectra though some disadvantages exist. DSPNet has a competitive performance compared with traditional approaches.

### E. Parameter Sensitivity Analysis

Finally, to verify the reliability of the model, we performed a sensitivity analysis on the dilated rate, an important hyperparameter of the dilated convolution layer. According to Fig. 8, it shows that the dilation series networks have a better performance than the normal CNN (also called DSPNet-o-DSP) on the validation dataset obviously.

On the validation data, from Fig. 8, we could find that the DSPNet ( $l = 5$ ) has better results. The networks of other dilated rate are over-fitting to some extent. In general, dilation series are stable and effective, and even the worst situation still achieved a better result than the normal CNN.

### F. Limitations

Although the DSPNet works well for the peak-to-valley restoration in the spectrum robustly. There is still one problem: the residual noise in the flat region cannot be removed well, even worse. It is widely used that the residual noise is white Gaussian noise denoted by  $n(t)$  for analyzing conveniently, where  $t$  denotes time. We have

$$r_{nn} = \mathbb{E}\{n(t)n(t - \tau)\} = \delta(t) \quad (11)$$

where  $r_{nn}$  denotes the autocorrelation function and  $\delta(t)$  is the Dirac delta function. According to the Fourier analysis, the power spectrum of  $n(t)$  is equal to a constant, which means the noise has much intensity in high frequencies domain. Xu *et al.* proposed that CNNs often fit target functions from low to high frequencies called frequency principle [21]. In other words, CNNs are not good at learning high frequencies information. Our experiment results conform to the analysis of the frequency principle.



## V. CONCLUSION

In this article, we proposed a new method based on SPL with dilated CNNs. The primary contribution of our work demonstrated the possibility of tackling spectral deconvolution problems from neural network perspective. Our system novelly replaced normal convolution layer with dilated convolution layer to increase receptive field and utilized the SPL framework, achieving robust performance. Compared with the PDE methods, the proposed end-to-end neural network does not rely on prior knowledge and could design features based on spectral information automatically instead of hand-engineered, thanks to delicate network structure. Extensive experiments demonstrated that the proposed solution has an outstanding performance compared with past methods.

In much research, blur kernel was considered as a fixed function, but it may change along with the time in practical. It is easier for deep learning based approaches to handling nonfixed blur kernel compared with classic approaches as data-driven approaches need not limit the blur kernel into certain distribution. We will focus on nonfixed blur kernel deconvolution problem in the future.

## REFERENCES

- [1] F. Chraim, Y. B. Erol, and K. Pister, "Wireless gas leak detection and localization," *IEEE Trans. Ind. Informat.*, vol. 12, no. 2, pp. 768–779, Apr. 2016.
- [2] T. Liu, H. Liu, Y. Li, Z. Zhang, and S. Liu, "Efficient blind signal reconstruction with wavelet transforms regularization for educational robot infrared vision sensing," *IEEE/ASME Trans. Mechatronics*, vol. 24, no. 1, pp. 384–394, Feb. 2019.
- [3] S.-B. Roh, S.-K. Oh, and W. Pedrycz, "Identification of black plastics based on fuzzy RBF neural networks: Focused on data preprocessing techniques through Fourier transform infrared radiation," *IEEE Trans. Ind. Informat.*, vol. 14, no. 5, pp. 1802–1813, May 2018.
- [4] P. A. Jansson, *Deconvolution of Images and Spectra*. North Chelmsford, MA, USA: Courier Corporation, 2014.
- [5] T. Liu, H. Liu, Z. Chen, and A. M. Lesgold, "Fast blind instrument function estimation method for industrial infrared spectrometers," *IEEE Trans. Ind. Informat.*, vol. 14, no. 12, pp. 5268–5277, Dec. 2018.
- [6] H. Zhu, L. Deng, G. Xu, Y. Chen, and Y. Li, "Spectral semi-blind deconvolution methods based on modified  $\varphi_{HS}$  regularizations," *Opt. Laser Technol.*, vol. 110, pp. 24–29, 2019.
- [7] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. 4th Int. Conf. Learn. Representations*, 2016. [Online]. Available: <https://openreview.net/forum?id=BygfgHAcYX>
- [8] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4509–4522, Sep. 2017.
- [9] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
- [10] L. Jiang, D. Meng, Q. Zhao, S. Shan, and A. G. Hauptmann, "Self-paced curriculum learning," in *Proc. 29th Conf. Artif. Intell.*, 2015, pp. 2694–2700.
- [11] C. Xu, D. Tao, and C. Xu, "Multi-view self-paced learning for clustering," in *Proc. 24th Int. Joint Conf. Artif. Intell.*, 2015, pp. 3974–3980.
- [12] J. Gu et al., "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, 2018.
- [13] H. Liu, Y. Li, Z. Zhang, S. Liu, and T. Liu, "Blind Poissonian reconstruction algorithm via curvelet regularization for an FTIR spectrometer," *Opt. Express*, vol. 26, no. 18, pp. 22837–22856, 2018.
- [14] H. Liu, L. Yan, T. Huang, S. Liu, and Z. Zhang, "Blind spectral signal deconvolution with sparsity regularization: An iteratively reweighted least-squares solution," *Circuits, Syst., Signal Process.*, vol. 36, no. 1, pp. 435–446, 2017.
- [15] L. Xu, J. S. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, 2014, vol. 1, pp. 1790–1798.
- [16] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, vol. 37, pp. 448–456.
- [17] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. 29th IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 2818–2826.
- [18] S. S. Du, X. Zhai, B. Póczos, and A. Singh, "Gradient descent provably optimizes over-parameterized neural networks," in *Proc. 7th Int. Conf. Learn. Representations*, 2019. [Online]. Available: <https://openreview.net/forum?id=BygfgHAcYX>
- [19] H. Zhu, L. Deng, H. Li, and Y. Li, "Deconvolution methods based on convex regularization for spectral resolution enhancement," *Comput. Elect. Eng.*, vol. 70, pp. 959–967, 2018.
- [20] V. A. Lorenzonfria and E. Padros, "Maximum entropy deconvolution of infrared spectra: Use of a novel entropy expression without sign restriction," *Appl. Spectrosc.*, vol. 59, no. 4, pp. 474–486, 2005.
- [21] Z.-Q. J. Xu, Y. Zhang, T. Luo, Y. Xiao, and Z. Ma, "Frequency principle: Fourier analysis sheds light on deep neural networks," Jan. 2, 2020, *arXiv:1901.06523*.



**Hu Zhu** (M'17) received the B.S. degree in mathematics and applied mathematics from Huaibei Coal Industry Teachers College, Huaibei, China, in 2007, and the M.S. and Ph.D. degrees in computational mathematics and pattern recognition and intelligent systems from the Huazhong University of Science and Technology, Wuhan, China, in 2009 and 2013, respectively.

He is currently an Associate Professor in Nanjing University of Posts and Telecommunications, Nanjing, China. His research interests include pattern recognition, image processing, and computer vision.



**Yiming Qiao** is currently working toward the bachelor's degree in information security with Bell Honors School, Nanjing University of Posts and Telecommunications, Nanjing, China.

He is currently an Intern with the Oracle Corporation, Redwood Shores, CA, USA. His research interests include machine learning and data mining.



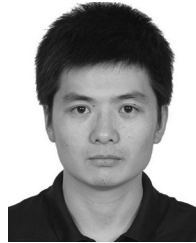
**Guoxia Xu** (M'19) received the B.S. degree in computer and information science from the Department of Mathematics, Yancheng Teachers University, Yancheng, China, in 2015, and the M.S. degree in computer science and technology from the Department of Computer and Information, Hohai University, Nanjing, China, in 2018.

He was a Research Assistant in Electronic Engineering with the City University of Hong Kong, Hong Kong. He is currently a Research Assistant with the Department of Imaging and Interventional Radiology, The Chinese University of Hong Kong, Hong Kong. His research interests include machine learning and image and video processing.



**Lizhen Deng** (M'17) received the B.S. degree in electronic information science and technology from Huaibei Coal Industry Teachers College, Huaibei, China, in 2007, the M.S. degree in communication and information systems from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2010, and the Ph.D. degree in electrical engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2014.

She is currently an Associate Professor in Nanjing University of Posts and Telecommunications, Nanjing, China. Her current research interests include image processing, computer vision, pattern recognition, and spectral data processing.



**Yu-Feng Yu** (M'19) received the Ph.D. degree in statistics from Sun Yat-Sen University, Guangzhou, China, in 2017.

He is currently an Assistant Professor with the Department of Statistics, Guangzhou University, Guangzhou. From 2016 to 2017, he was a Visiting Scholar with the Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV, USA. From 2017 to 2018, he was a Senior Research Associate with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong. His research interests include image processing, statistical optimization, pattern recognition, and machine learning.