Polynomial Phase Signal Denoising Connecting Semantic Information Based on Deep Neural Networks

Liang Zhang[1,*], Cheng-Yu Hung[2], Hongyu Yang[3] and Rui Huang[3]

[1]University of Arizona Tucson, AZ, USA
[2]University of Minnesota Minneapolis, MN, USA
[3]Civil Aviation University of China, Tianjin, China

E-mail: liangzh@email.arizona.edu

Abstract. This paper considers the problem of polynomial phase signal (PPS) denoising. To prove that proper use of semantic information can further improve the denoising performance based on deep neural networks, we propose an architecture combining the segmentation network and the denoising network. The vision semantic information is extracted from the segmentation network first. Then, that information connecting the time-frequency representation of noisy signal are fed into the denoising network for reconstructing signal. To effectively apply the semantic information, three connection strategies and the corresponding lower bound are presented and compared. The proposed method does not require the pre-identification of signal noise conditions and is suitable for a wide range of Signal-to-Noise-Ratio (SNR) scenarios. Simulation results demonstrate that the F1 scores of the spectrum segmentation results are over 0.98 and the proposed method connecting vision semantics for PPS denoising tasks outperforms the baseline and state-of-the-art architectures, when the SNR is larger than -8dB.

1. Introduction

Polynomial phase signals (PPSs) have been widely used in numerous signal processing applications such as those in radar, sonar, radio communications, medicine, speech and music signals, dispersive seismic signals, where the phase is continuously modulated [1-3]. Therefore, PPS denoising is a very worthwhile problem in many fields.

The effectiveness of various image denoising and semantic segmentation algorithms based on deep neural networks (DNN) have been extensively demonstrated. Combining convolutions, batch normalization, rectified linear unit and residual learning, Denoising CNN (DnCNN) [4] was proposed to deal with image denoising, super-resolution, and JPEG image deblocking. For improving denoising speed and process blind denoising, a fast and flexible denoising CNN (FFDNet) [5] was presented in which the noisy image patch is chosen as the input of this network. Encoder-Decoder networks have shown their superior performance in image denoising and restoration [6-8]. Using DNN for these challenging problems is surpassing other traditional approaches by a large margin in terms of accuracy [9-11]. Segnet [12] composed by the backbone of VGG16 [13] also demonstrated that the promising success for image semantic segmentation.

The successful technologies of DNN in the field of image denoising, especially CNN, auto-encoder, etc., can be able to extended to the applications of signal processing. In recent years, DNN have been increasingly introduced into denoising tasks of various special signals, such as ECG signals, seismic

signals, speech signals, etc. [14-21]. These algorithms usually considered the time, frequency, and phase information of noisy signals as the basis for denoising tasks.

Semantic information is applied in some image processing tasks, such as saliency detection, image denoising, image matching, etc. in some recent studies [22-26], which further improved the performance of these tasks. However, to the best of our knowledge, there is still a lack of research on how to apply semantic information to the complex signal denoising problem. Leveraging the similarity between the signal spectrum in the time-frequency domain and 2D images, we propose an architecture connecting the semantic information of signals and signal denoising based on DNN in this paper. And, the recovered region of noisy signal spectrum is selected as spatial semantics. Three strategies to connect semantic information that can further improve the quality of signal denoising and the corresponding low bound are presented. The specific training method for the two networks is also given. The effectiveness of the proposed method will be demonstrated in Simulations.

2. Problem

2.1. Signal model

The following model can describe general PPSs:

$$s(t) = A\exp\left(j\sum_{m=0}^{p} a_m t^m \right) \tag{1}$$

where A is the signal amplitude, p is the order of the polynomial phase, and the phase coefficients $\{a_m \mid m \in [0, p]\}$ are the real parameters. To make it easier to understand, we especially choose the second-order representation of the PPS modelling, that is, the chirp signal as an example to explain our method. However, the proposed method can be applied to higher-order situations, we will discuss it in the simulation section.

Then, the noisy model can be given as

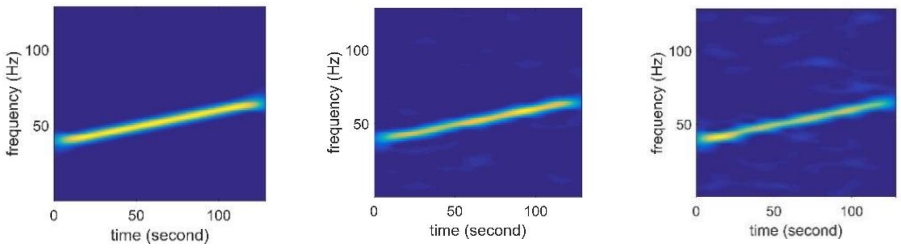
$$x(t) = s(t) + n(t) \tag{2}$$

where $s(t)$ is linear chirp signal and $n(t)$ is additive white Gaussian noise. Besides that, $s(t)$ can be further given by

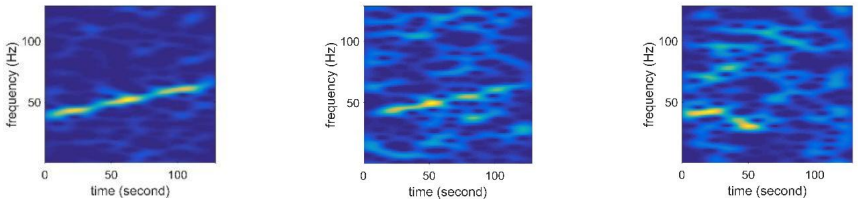
$$s(t) = a_0 e^{j\pi(2f_0 t + \mu_0 t^2) + j\varphi_0} \tag{3}$$

where a_0 is the amplitude, f_0 is the initial frequency, φ_0 is the initial phase, and μ_0 is the chirp rate determining the changing rate of frequency.

2.2. Differences between image and signal denoising



(a) (b) (c)



(d) (e) (f)

Figure 1. The spectrum of linear chirp signal in the time-frequency plane. (a) signal spectrum without noise (b) signal spectrum with SNR=8dB (c) signal spectrum with SNR=4dB (d) signal spectrum with SNR=0dB (e) signal spectrum with SNR=-6dB (f) signal spectrum with SNR=-10dB.

Although lots of great success has been achieved with deep learning in image denoising and semantic segmentation tasks, we face some differences while generalizing the image denoising architectures for signal denoising tasks:

1) Unlike the discretized integer representation of images constrained by the range of [0, 255], signals in the time-frequency domain are represented in the form of complex decimal numbers with both real and imaginary parts.

2) The spectrum shape of signal can be submerged or distorted by strong noises under low Signal-to-Noise-Ratio (SNR) conditions. For example, Figure 1 shows the linear chirp signal spectra under different SNRs. The spectrum of the signal is eroded more seriously and the noise spectrum creates more visual disturbance as SNR decreases. Furthermore, the shape and the value distribution of the signal spectrum under different SNR conditions are very different, which makes the realization of training a single architecture to denoise noise signals within an unknown and wide range of SNR more challenging.

The above differences require our architecture to accurately determine the semantic information of signals and reconstruct signals from spectrum shapes with varying degrees of erosion in the time-frequency domain. Thus, the robustness of the proposed algorithm is essential in this study.

3. Method

3.1. Spectrum segmentation

In the field of image processing, salient object detection and semantic segmentation are the two common high-level tasks which are highly correlated but essentially different in some ways. Salient object detection focuses on distinguishing salient objects from background, while semantic segmentation aims at separating objects of different semantic classes. Most related studies require a large number of noise-free images as the training set in which objects' shapes are complete and clear. Nevertheless, in signal denoising, more distorted inputs need to be considered. There may even be some false shapes visually, because of strong noises, such as (e) and (f) in Figure 1.

In light of the high accuracy of Encoder-Decoder networks with the VGG16 backbone when detecting target objects in images, we explore an architecture to segment the spectrum region in the time-frequency domain, which is served as the semantic prior knowledge for signal denoising tasks.

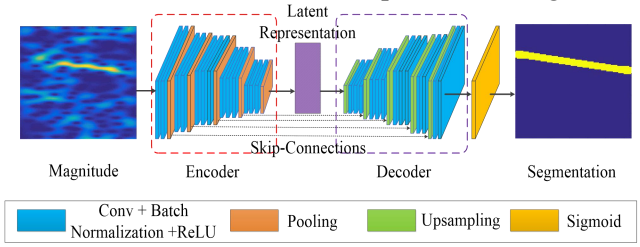


Figure 2. The segmentation network based on the VGG16 backbone.

The segmentation architecture for spectrum region segmentation is shown in Figure 2. To obtain a better grasp on the spectrum shape in the time-frequency plane, the magnitude is chosen as the input channel, and the spectrum region segmentation is the output. The encoder needs to locate the possible spectrum regions of signal from the magnitude of noisy signal spectrum, and predict and correct the region of signal spectrum eroded by noises. Then, this information containing the vision semantics is compressed and stored into the latent representation, which can be fed into the decoder to predict the probability of whether each time-frequency point is on the signal spectrum region by performing the Sigmoid classifier.

In the labelling set of spectrum region, we design a matrix as targets for training that has the same shape as noise-free GT coefficient matrices, in which 1 denotes the spectrum region and 0 occupies others, as shown in Figure 3.

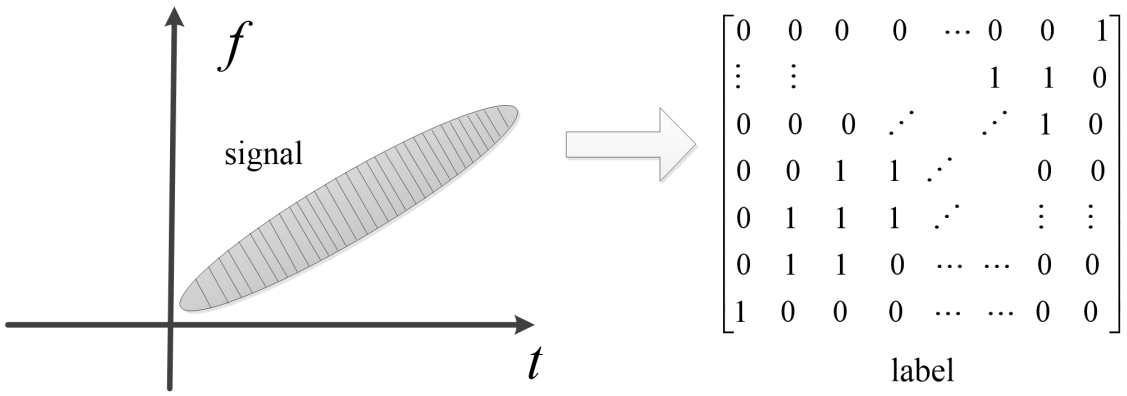


Figure 3. The signal spectrum in the time-frequency domain and the corresponding region label matrix.

In order to acquire the 2D time-frequency representation of signals, the time-frequency transform should be computed first. Gabor transform (GT) is selected as an example in this paper. And then, the output $o_x(t,f)$ of the spectrum region prediction based on the segmentation network h_{seg} can be interpreted by

$$o_x(t,f) = h_{seg}\left\{\left|G_x(t,f)\right|\right\} \tag{4}$$

where the GT $G_x(t,f)$ is defined as

$$G_x(t,f) = \int_{-\infty}^{+\infty} e^{-(\tau-t)^2/2} e^{-j2\pi f\tau} x(\tau)\, d\tau \tag{5}$$

and t and f are the variables in the time-frequency domain representing time and frequency points respectively.

Because the segmentation output $o_x(t,f)$ is not exactly 0 or 1 but the probability for each point, the threshold needs be set to determine the spectrum region. Then, we have the segmentation result according to

$$\begin{aligned} r_x(t,f) &= 1, \textit{when } o_x(t,f) \geq \textit{threshold} \\ r_x(t,f) &= 0, \textit{when } o_x(t,f) < \textit{threshold} \end{aligned} \tag{6}$$

The spectrum region matrix where $r_x(t,f) = 1$ in the output of Figure 2 corresponds to the yellow region, while the other areas in deep blue stand for the non-spectrum region. Note that, the threshold is an important parameter in our model, and we will discuss it further in the simulations.

3.2. Signal denoising

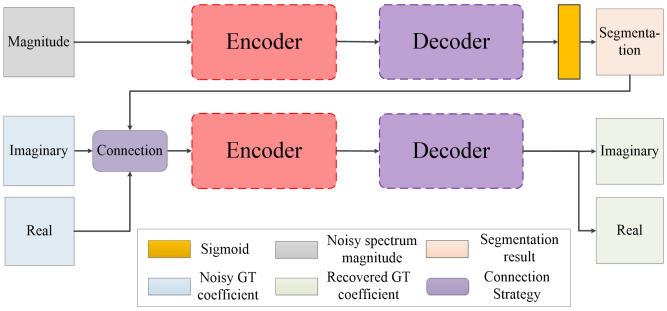


Figure 4. PPS denoising network connecting the spectrum region segmentation based on the VGG16 backbone.

The overview of the complete segmentation and denoising networks is shown in Figure 4. The real and imaginary parts of the noisy complex GT coefficient connecting the segmentation result are fed into the denoising network (the lower Encoder-Decoder architecture in Figure 4) which also uses VGG16 as the backbone. And, the output is the recovered GT coefficients.

In this architecture, in addition to the time, frequency, and phase used in the existing methods, we additionally introduce the semantic information of signals as the prior knowledge. To effectively use of semantic information, we propose the three connection strategies and the corresponding lower bound for the proposed method.

Strategy1: The connection function in Figure 4 concatenates the real and imaginary matrices of GT coefficients and the output matrix in (4) (the probability matrix without considering the threshold) which contains the complete segmentation information but may introduce more prediction error when SNR is low.

Strategy2: The connection function in Figure 4 concatenates the real and imaginary matrices of GT coefficients and the segmentation matrix in (6) (the 0-1 matrix) which preserves the estimated spectrum region.

Strategy3: The connection function in Figure 4 outputs the dot products of the real and imaginary matrices of GT coefficients and the segmentation matrix in (6), respectively, which only preserves the noisy GT coefficient information in the estimated spectrum region. Note that, this strategy returns GT coefficient matrices as the two input channels instead of the three channels in the two previous strategies.

Lower Bound of Our Method: The connection function in Figure 4 concatenates the real and imaginary matrices of GT coefficients and the noise-free spectrum region labelling matrix, which is the lower bound of our proposed method.

3.3. Training strategy, loss functions and time complexity

The training data set comes from a wide range of SNR, which is fed into the proposed networks for training after randomly shuffled. According to our test, the joint training for the two architectures is difficult to converge, thus the segmentation network and the denoising network are trained separately in our study. We train the segmentation network in an end-to-end manner and the weights of the well-trained segmentation net are fixed. Then, the spectrum magnitude of noisy signal is fed into the segmentation net to get the spectrum segmentation. After that, the noisy GT coefficients of this signal connecting the segmentation result are used as the input for the denoising network. Besides that, noise-free spectrum region labelling matrices in Figure 3 and noise-free GT coefficient matrices are the corresponding labels for the two networks, respectively. The weights of the two networks are updated according to the back-propagation algorithm during training.

The segmentation loss L_{seg} of the segmentation network is the Sigmoid Cross-Entropy loss between the segmentation network output $o_x(t,f)$ and the noise-free spectrum region label $l_x(t,f)$, which can be given by

$$L_{seg} = -l_x(t,f) \times \log\left(o_x(t,f)\right) - \left(1 - l_x(t,f)\right)\log\left(1 - o_x(t,f)\right) \tag{7}$$

While the denoising loss L_{den} of the denoising network is the mean squared error (MSE) between the denoising network output and the noise-free GT coefficients, which can be represented by

$$L_{den} = \left(h_{den}\left(g_x(t,f)\right) - g_{\bar{x}}(t,f)\right)^2 \tag{8}$$

where $x(t)$ is the noisy signal and $\bar{x}(t)$ is the noise-free signal. And h_{den} represents the denoising network.

The time complexity of the proposed method is mainly caused by all the convolutional layers which can be evaluated by [27]

$$O\left(\sum_{l=1}^{d} \alpha_{l-1} \cdot \beta_l^2 \cdot \alpha_l \cdot \gamma_l^2\right) \tag{9}$$

where l represents the index of a convolutional layer, and d is the depth (number of convolutional layers). αl is the number of filters (also known as "width") in the l-th layer. αl−1 can be also interpreted as the number of input channels of the l-th layer. βl is the spatial size (length) of the filter. γl is the spatial size of the output feature map. And this time complexity applies to both training and testing time, though with a different scale.

4. Simulations

The results of the second-order signal are shown and discussed in this section. However, the same conclusion can be easily generalized for higher-order representations. We also tested the PPSs denoising from third to fifth-order whose spectral shapes show varying degrees of bending by using the same simulation settings as chirp signals', and they all have very close results which demonstrates the effectiveness of the proposed method.

We generate a data set including 10 groups in the different SNRs from -10dB to 8dB with a 2dB interval. This data set contains 10,000 signals for training and 1,000 signals for testing. In the process of testing, there is no need to provide any other prior information except the sampled noise signal as the input for the two networks.

The instantaneous frequency $f = f_0 + \mu_0 t$ of each chirp signal is in the range of [10, 110] Hz during the sampling time by limiting the randomly generated ranges of f_0 and μ_0. Note that, due to the addition of additive white Gaussian noise in the generated signals, all the signals in the training set and the test set are different. Moreover, the number of samples is 128 and the window length of GT is chosen as 41.

The generated signals in the time domain are transformed into a complex 128×128 GT coefficient matrix first. Then, the magnitude of GT coefficient matrix is fed into the segmentation network which predicts the spectrum region. The segmentation result connecting the real and imaginary matrices of GT coefficients of the noisy signal through the proposed strategies are fed into the denoising network as the three input channels so that network learns from time, frequency, phase, and semantic information at the same time.

The learning rate for the proposed networks training is 0.001 using a RMSProp optimizer with the decay rate 0.995. The batch size is 32. Furthermore, the segmentation network is trained by 140 epochs and the denoising network is trained by 200 epochs. The denoising network without connecting the semantic information is chose as the baseline to perform the ablation study. In addition, the input and output for the DnCNN and FDDNet training are the noisy and recovered GT coefficient matrices without considering semantic information. The learning rates of the DnCNN and FDDNet trainings are 0.001 and the other hyperparameters are unchanged as they used in the original papers. Both the trainings for DnCNN and FDDNet are also proceeded for 200 epochs.

This section considers the evaluation of spectrum region prediction from the segmentation network. The Intersection over Union (IoU), precision, recall and F1 score are all the important metrics to

quantify the performance of deep neural networks for a vision segmentation task. The threshold in (6) is 0.8 in this test. The results for each metric under SNRs are shown in Table 1. The IoU achieves over 97% and the F1score is greater than 0.98 when SNR is higher than -8dB. In addition, according to our observation, most erroneous prediction appears in spectrum edge areas. Therefore, DNN-based image segmentation can be extended to signal spectrum segmentation tasks that are distorted by various degrees of noises. We expect that the high quality and accurate spectrum region prediction can be as the prior information to help improve signal denoising performance in denoising tasks.

Table 1. The result for each metric under SNRs

SNR\ metric	IoU	precision	recall	F1
-10 dB	0.9555	0.9794	0.9733	0.9762
-8 dB	0.9716	0.9897	0.9808	0.9852
-6 dB	0.9720	0.9897	0.9807	0.9851
-4 dB	0.9799	0.9941	0.9850	0.9895
-2 dB	0.9816	0.9952	0.9864	0.9908
0 dB	0.9850	0.9965	0.9887	0.9926
2 dB	0.9857	0.9962	0.9907	0.9934
4 dB	0.9871	0.9969	0.9912	0.9940
6 dB	0.9876	0.9969	0.9921	0.9945
8 dB	0.9890	0.9969	0.9933	0.9951

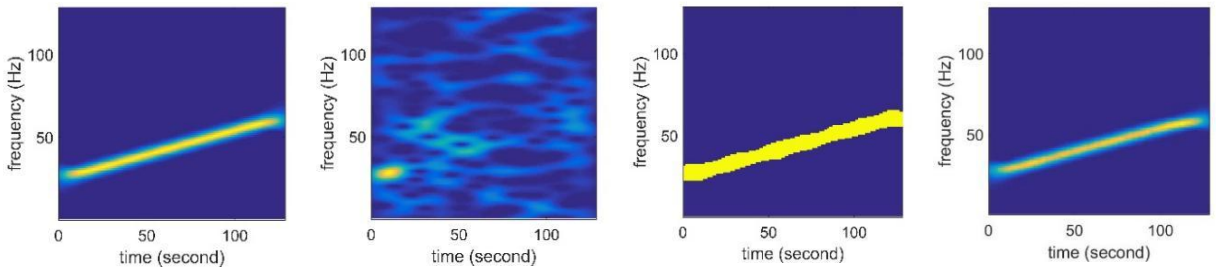


Figure 5. Example of using the strategy2 when SNR = -10dB. (a) noise-free signal (b) noisy signal (c) the segmentation result (d) the reconstructed signal.

Figure 5 illustrates an example of using the strategy2 for the PPS denoising. In the case where the spectral shape is severely eroded by noise (Figure 5 (b)), the segmentation net can still predict the spectral region corresponding to the noise-free signal relatively accurately (Figure 5 (c)). And the denoising net combining the spectral semantic information and the time-frequency coefficients of noisy signals to obtain a good denoising effect (Figure 5 (d)).

The Root-Mean-Square Error (RMSE) for GT coefficient recovery comparing with DnCNN, FDDNet, the proposed methods using the three strategies under the threshold 0.8, the baseline and the lower bound, are shown in the Figure 6. The lower bound has an overwhelming advantage, which proves that connecting semantic information can greatly improve the signal denoising performance. In addition, our proposed method and the baseline outperform state-of-the-art architectures of image denoising tasks for signal denoising. The proposed architecture shows better robustness when signals are eroded by noises in the time-frequency domain. Compared with the baseline, our proposed methods connecting semantic information have similar or better performance, when SNR is greater than -8dB. Especially, the strategy2 is advantageous than the other two strategies. This simulation demonstrates that the usage of the accurate spectrum region segmentation as the prior knowledge is beneficial to further improve the signal denoising performance. However, when the SNR is -10dB, the reduced accuracy of the spectrum region segmentation has the negative impact on the signal denoising performance.

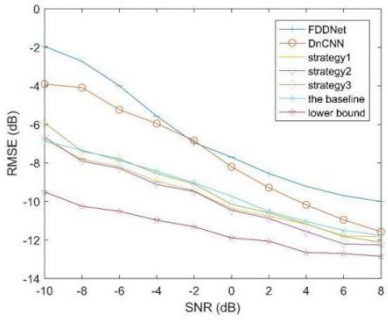


Figure 6. The RMSEs for signal denoising versus SNRs by the different architectures.

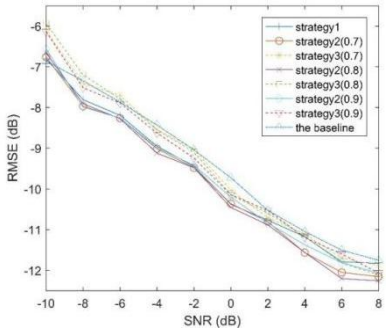


Figure 7. The RMSEs for signal denoising versus SNRs by the different thresholds.

We evaluate the impact of the different thresholds on the denoising performance of the proposed method. The result of this test is shown in Figure 7. The numbers in the parentheses of the legend represent the selected thresholds used for the connection strategy2 and strategy3. These tested methods using the three proposed strategies under the different thresholds are superior to the baseline in varying degrees when SNR is high. And, using the same threshold, the strategy2 is better than the strategy3.

5. Conclusion

This paper proposed a connected architecture design for PPS denoising in the time-frequency domain. The accurate spectrum region prediction from the segmentation network provided the semantic information that extracted from the magnitude of spectrum. Compared with the baseline and state-of-the-art architectures, the proposed method further improved the performance of PPS denoising tasks by combining semantic, time, frequency, and phase information from noisy signals while SNR is greater than -8dB. The simulations demonstrated that the proposed method has strong anti-noise characteristics in the two tasks of spectrum semantic segmentation and signal denoising without other knowledge. How to further improve the accuracy of spectrum region segmentation under strong noisy conditions is one of the goals that need to be solved in our future research. We will also study more complex situations in future work, such as multiple overlapping PPSs, and more complex signals, such as speech signals, etc.

References

[1] I. Djurović, and S. Marko. "Review of the quasi-maximum likelihood estimator for polynomial phase signals." *Digital Signal Processing*, vol. 72, pp. 59-74, 2018.

[2] I. Djurović, S. Marko, and W. Pu. "Cubic phase function: A simple solution to polynomial phase signal analysis." *Signal Processing*, vol. 135, pp. 48-66, 2017

[3] B. Bullecks, S. Resmi and R. Raghunathan. "Rapid impedance measurement using chirp signals for electrochemical system analysis." *Computers & Chemical Engineering*, vol. 106, pp. 421-436, 2017.

[4] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[5] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, Sep. 2018.

[6] X.-J. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," *in Proc. Adv. Neural Inf. Process. Syst.* (NIPS), 2016, p. 221.

[7] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-recurrent network for deep image deblurring," *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* (CVPR), Jun. 2018, pp.

8174–8182.
[8] H. Chen et al., “Low-dose CT with a residual encoder-decoder convolutional neural network,” *IEEE Trans. Image Process.*, vol. 36, no. 12, pp. 2524–2535, Dec. 2017.
[9] F. Ning, D. Delhomme, Y. LeCun, F. Piano, L. Bottou, and P. E. Barbano, “Toward automatic phenotyping of developing embryos from videos,” *IEEE Trans. on Image Processing*, vol. 14, no. 9, pp. 1360–1371, 2005.
[10] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, “Deep neural networks segment neuronal membranes in electron microscopy images,” *in Advances in neural information processing systems*, pp. 2843–2851, 2012.
[11] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, “Learning hierarchical features for scene labeling,” *IEEE trans. on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1915–1929, 2013.
[12] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation,” *in Proc. CoRR*, 2015.
[13] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
[14] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, and M. Adam, “Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals,” *Information Sciences*, vol. 415, pp. 190–198, 2017.
[15] P. Xiong, H. Wang, M. Liu, F. Lin, Z. Hou, and X. Liu, “A stacked contractive denoising auto-encoder for ECG signal denoising,” *Physiol. Meas*., vol. 37, no. 12, pp. 2214, Nov. 2016.
[16] H.-T. Chiang, Y.-Y. Hsieh, S.-W. Fu, K.-H. Hung, Y. Tsao, and S.-Y. Chien, “Noise reduction in ECG signals using fully convolutional denoising autoencoders,” *IEEE Access*, vol. 7, pp. 60806–60813, 2019.
[17] P. Xiong, H. Wang, M. Liu, S. Zhou, Z. Hou, and X. Liu, “ECG signal enhancement based on improved denoising auto-encoder,” *Eng. Appl. Artif. Intell.*, vol. 52, pp. 194–202, 2016.
[18] W. Zhu, S. M. Mousavi, and G. C. Beroza, “Seismic signal denoising and decomposition using deep neural networks,” *IEEE Trans. Geosci. Remote Sens*., vol. 57, no. 11, pp. 9476–9488, Nov. 2019.
[19] Y. Zhang, H. Lin, Y. Li, and H. Ma, “A patch based denoising method using deep convolutional neural network for seismic image,” *IEEE Access*, vol. 7, pp. 156 883–156 894, 2019.
[20] K. Minje, and P. Smaragdis. "Bitwise neural networks for efficient single-channel source separation." *in Proc. ICASSP. IEEE*, 2018.
[21] Z. Han, et al. "Convolutional-recurrent neural networks for speech enhancement." *in Proc. ICASSP. IEEE*, 2018.
[22] T. V. Nguyen,, K. Nguyen, and T Do, "Semantic prior analysis for salient object detection," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 3130-3141, June 2019.
[23] D. Liu, et al., "Connecting image denoising and high-level vision tasks via deep learning," *IEEE Trans. Image Process.*, vol. 29, pp. 3695-3706, 2020.
[24] L. Wang, et al., "Salient object detection with recurrent fully convolutional networks." *IEEE trans. pattern analysis and machine intelligence*, vol. 41, no. 7, pp. 1734-1746, 2018.
[25] T. Remez, et al. "Deep class-aware image denoising." *in Proc. SampTA IEEE*, 2017.
[26] W. Yu, et al., "Hierarchical semantic image matching using CNN feature pyramid." *Computer Vision and Image Understanding*, vol.169, pp. 40-51, 2018.
[27] He Kaiming, and Jian Sun. "Convolutional neural networks at constrained time cost." *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5353-5360. 2015.