LEARNING STRUCTURED SPARSITY FOR TIME-FREQUENCY RECONSTRUCTION

Lei Jiang, Haijian Zhang, Lei Yu

Signal Processing Lab., School of Electronic Information, Wuhan University, China

ABSTRACT

Compressed sensing based algorithms are utilized to obtain high-resolution time-frequency distribution (TFD) with negligible cross-terms (CTs), however, performance deteriorates when the signal is composed of closely-located or overlapped components. Moreover, there is still an impressive resolution gap between obtained and ideal TFDs. Aiming at eliminating CTs meanwhile preserving resolution as high as possible in various hard cases, we propose a new U-Net aided iterative shrinkage-thresholding algorithm (U-ISTA), where unfolded ISTA with structure-aware thresholds is exploited to reconstruct near-ideal TFD. Specifically, we regard the U-Net as an adaptive threshold block, and structured sparsity of TFD is learned from numerous training data, thus underlying dependencies among neighboring time-frequency coefficients are incorporated into reconstruction. Experimental results over synthetic and real-life signals demonstrate that the proposed U-ISTA achieves superior performance compared with state-of-the-art algorithms.

***Index Terms*—** time-frequency distribution, iterative shrinkage-thresholding, deep learning, structured sparsity

1. INTRODUCTION

High-resolution time-frequency distribution (TFD) is of importance in signal representation. Linear TFDs usually have low resolutions, thus the reassigned spectrogram (RSP) algorithm [1] was proposed, whereas it is hard to obtain accurate representations for signals with closely-located components. On the other hand, despite having high-resolution, nonlinear TFDs generally suffer from oscillatory artifacts also called cross-terms (CTs), which heavily hamper the intuitive interpretation of signal's TFD [2]. In order to suppress these artifacts, kernel function algorithms design low-pass filters in the ambiguity domain to obtain clean TFDs [3–8]. Nevertheless, CTs reduction comes at the expense of low resolution.

Given great performance on having both CTs reduction and high-resolution TFD, an increasing number of compressed sensing (CS) based algorithms was proposed to attain near-ideal TFDs [9–13]. Taking the 2-dimensional (2D) Fourier transform pair of quadratic ambiguity function (AF) [14] and Wigner-Ville distribution (WVD) [15] into consideration, Flandrin *et al.* [9] proposed the ℓ_1-app algorithm to reconstruct sparse signal's TFDs, i.e., auto-terms (ATs). Some investigations based on ℓ_1-app algorithm design adaptive sampling geometry to improve performance [10, 11, 13] on the signals with closely-located components, whereas they still perform poorly to reconstruct signals composed of overlapped components. Additionally, the ℓ_1-app algorithm neglects the TF correlation structure of the analyzed signal and inner CTs resulting from the nonlinearity. The structured sparsity of signal's TFDs can be considered to exploit correlation structure of the analyzed signal [12, 16]. However, the structure of CTs is too complicated to be captured especially in the case of the signal having overlapped components, thus deep learning algorithms can be introduced to solve this problem. Notably, TF analysis combined with deep learning comes into sight [17, 18].

In this paper, we propose a new deep learning based structured sparse TF reconstruction algorithm, which is referred to as U-Net aided iterative shrinkage-thresholding algorithm (U-ISTA). The basic model is an unfolded ISTA, truncated to a fixed number of iterations. Adaptive structure-aware threshold blocks, i.e., the U-Net [19], are introduced into each iteration to exploit the structured sparsity of signal's TFDs. Benefitting from a large number of training data, robust TF reconstruction can be achieved even under high noise levels.

Compared with previous works [9–13, 17, 18], the contribution of our proposed work is threefold. First, we design an end-to-end network to achieve sparse reconstruction in TF domain other than using traditional optimization or iterative algorithms [9–13]. Second, existing works merely consider the sparse characteristic of the TFD [9–11], and the structural sparsity of the signal in TF domain needs to be further explored, thus our proposed adaptive threshold block aims at exploring inter-dependencies among TF coefficients benefitting reconstruction. Third, deep learning models are employed in the TF domain as convolutional kernel functions [17, 18], while our algorithm is a CS-based deep model employed in the ambiguity domain, i.e., less sampling points are needed.

2. THE PROPOSED U-ISTA

It is well known that true TF components are mainly focused around the origin while CTs are mainly located at the high frequency area in the domain of AF. One can suppress the CTs and reconstruct the TFD from the AF samples in the low

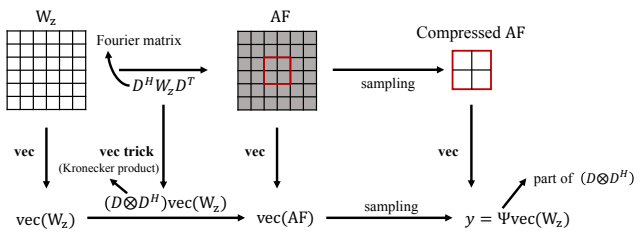
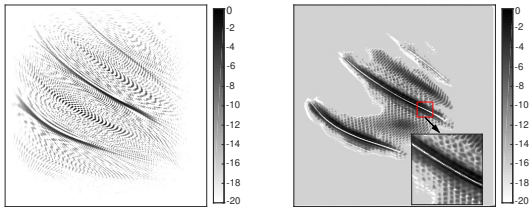


Fig. 1. Compressed sensing in the ambiguity domain.



(a) (b)

Fig. 2. (a) The WVD of a real-life bat echolocation signal. (b) Adaptive thresholds of the first layer.

frequency area. Therefore, in the following, we first build up a typical CS model in ambiguity domain, and then our proposed U-ISTA is described in detail.

2.1. Compressed Sensing in the Ambiguity Domain

Let us consider a non-stationary signal with l components $s(t) = \sum_{i=1}^{l} s_i(t)$, $l > 1$, and its WVD is defined as [20]:

$$\mathrm{W}_z(t,f) = \int_{-\infty}^{\infty} z\left(t+\frac{\tau}{2}\right) z^*\left(t-\frac{\tau}{2}\right) e^{-j2\pi f\tau} d\tau, \quad (1)$$

where $z(t)$ is the analytic associate of $s(t)$, and WVD is the sum of ATs and CTs [11]. Intuitively, only ATs are needed in practical applications, while oscillatory artifacts damage the interpretation of WVD. Through 2D Fourier transform of WVD, the ATs and CTs can be separated [2]:

$$\mathrm{A}_z(\nu,\tau) = \mathcal{F}_{\mathrm{2D}}\Big\{\mathrm{W}_z(t,f)\Big\}, \quad (2)$$

where $\mathcal{F}_{\mathrm{2D}}$ is 2D Fourier transform. Due to the sparsity of signal's TFD, CS-based algorithms can be employed to greatly reduce CTs. As shown in Fig. 1, compressed AF is centered at the origin of AF, and is regarded as the observation vector. H denotes the conjugate transpose operator, and the transpose operator is denoted as T. Thanks to a trick which is called Kronecker product, the vectorization of compressed AF can be expressed as:

$$\boldsymbol{y} = \boldsymbol{\Psi}(\boldsymbol{\omega} + \mathrm{error}) = \boldsymbol{\Psi}\boldsymbol{\omega} + \boldsymbol{\epsilon}, \quad (3)$$

where the measurement matrix $\boldsymbol{\Psi}$ consists of the corresponding Fourier coefficients, the real WVD vector $\mathrm{vec}(\mathrm{W}_z)$ can be replaced by a near-ideal WVD vector $\boldsymbol{\omega}$ with an error, and $\boldsymbol{\epsilon}$ denotes the noise vector.

The solution of (3) can be found by optimizing the following unconstrained problem with ℓ_1 norm regularization [21]:

$$\widehat{\boldsymbol{\omega}} = \arg\min_{\boldsymbol{\omega}} \frac{1}{2}||\boldsymbol{y} - \boldsymbol{\Psi}\boldsymbol{\omega}||_2^2 + \lambda||\boldsymbol{\omega}||_1, \quad (4)$$

where $\lambda > 0$ denotes a regularization parameter.

2.2. The Proposed U-ISTA Network

2.2.1. *From ℓ_1 to Re-Weighted ℓ_1 Norm Regularization*

Obviously, the limitation of the above optimization problem lies in that it gives unfair penalty to each nonzero coefficient, i.e., a large value is imposed with severe penalty while a small value is imposed with slight one. In order to have a more "democratic" penalty, re-weighted ℓ_1 norm regularization is denoted as [22]:

$$\widehat{\boldsymbol{\omega}} = \arg\min_{\boldsymbol{\omega}} \frac{1}{2}||\boldsymbol{y} - \boldsymbol{\Psi}\boldsymbol{\omega}||_2^2 + \lambda \sum_{i=1}^{N^2} g\big(|\omega_i|\big), \quad (5)$$

where $g(\cdot)$ denotes a concave, non-decreasing function called merit function. The ISTA is one of the commonly used algorithms to iteratively solve (5) and is written as [23]:

$$\boldsymbol{\omega}^{(k+1)} = h_\theta\left(\boldsymbol{\omega}^{(k)} - \frac{1}{L}\boldsymbol{\Psi}^T\left(\boldsymbol{\Psi}\boldsymbol{\omega}^{(k)} - \boldsymbol{y}\right)\right), \quad (6)$$

where $k = 0, 1, 2, ..., h_\theta(\boldsymbol{\omega}) = \mathrm{sign}(\boldsymbol{\omega})(\max(|\boldsymbol{\omega}| - \theta, 0))$ denotes the soft shrinkage function, and adaptive thresholds are denoted as $\theta = \lambda g'\big(\boldsymbol{\omega}\big)/L$, in which L is the largest eigenvalue of $\boldsymbol{\Psi}^T\boldsymbol{\Psi}$.

Contrary to traditional iterative algorithms fixing the measurement matrix, many model-based iterative algorithms choose to learn the weights in the matrix as well as thresholds in Equation (6), e.g., learned ISTA (LISTA) [24]. Besides, extended investigations based on LISTA examine that theoretically fast convergence rate and low error rate can be gained due to learning algorithms [24–27].

2.2.2. *Structure-Aware Threshold Block*

The ATs in most common TFDs often occupy sparse but continuous regions, e.g., the WVD of a real-life bat echolocation signal in Fig. 2(a). It can be seen that the ATs are smooth components with continuous structure, while the CTs have oscillatory characteristic. Though successful, the re-weighted ISTA [22] is unsuitable for TF representation. The reason is that it only defines element-wise adaptive thresholds, ignoring the dependencies among TF coefficients. Therefore, we propose the U-ISTA network to investigate the structure sparsity of the TFD.

A large effective receptive field is prone to be employed since long-range dependencies among TF coefficients can be attained. Then the U-Net, as an effective deep learning technology with a large effective receptive field, is used to learn

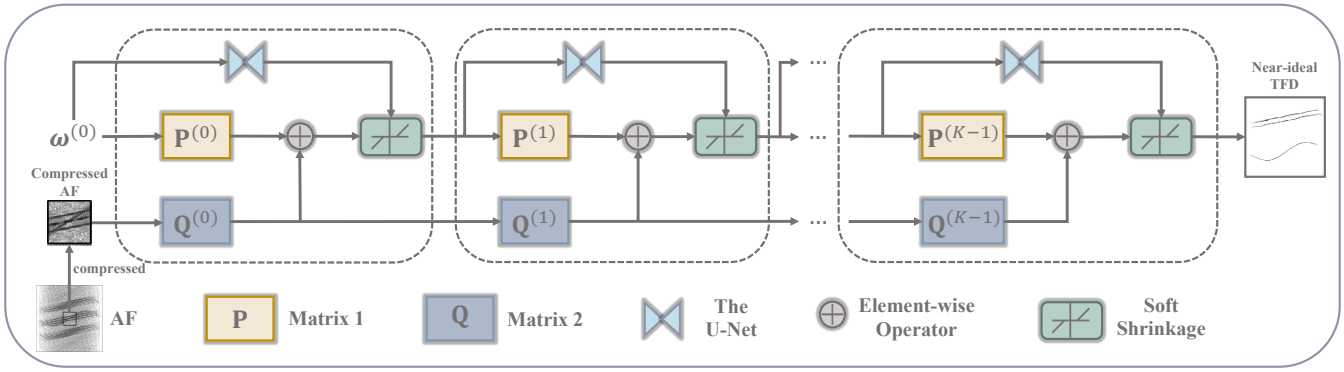


Fig. 3. The network architecture of the proposed U-ISTA.

structure-aware thresholds. Additionally, shallow features are combined with semantic ones so that a more robust result can be gained. That is, multi-scale implicit TF structure is extracted, and mutual complementation as well as improvement is achieved by concatenation. We employ only three layers in encoder and decoder path to balance between complexity and high-resolution, meanwhile channels are considerably reduced compared with the original U-Net. Benefiting from exploring TF correlation structure of the signal, adaptive thresholding values are prone to be small in the useful area entries while large values are learned in the area of artifacts. We visualize adaptive thresholds of the first layer. It is shown in Fig. 2(b) that our algorithm can well exploit TF structural characteristic of a real-life bat echolocation signal, i.e., entries with small values emerge as a continuous curve.

The network architecture of U-ISTA is shown in Fig. 3. Firstly, we unfold ISTA as a resemble recurrent neural network (RNN) inspired by LISTA [24], as it is worthwhile introducing structure-aware threshold blocks. Other than learning matrix $\boldsymbol{\Psi}$ from training data, we choose to fix it in every iteration since Fourier transform matrix has an explicit physical significance. Furthermore, with a proper selection of threshold, ATs will be retained while CTs will be reduced, for which artifacts are cut off by large thresholds. In the framework of U-ISTA, the solution to the problem in (6) is written as:

$$\begin{cases} \mathbf{P}^{(k)} = \mathbf{I} - t^{(k)} \boldsymbol{\Psi}^T \boldsymbol{\Psi} \\ \mathbf{Q}^{(k)} = t^{(k)} \boldsymbol{\Psi} \\ \boldsymbol{\omega}^{(k+1)} = h_{\theta^{(k)}} \left(\mathbf{P}^{(k)} \boldsymbol{\omega}^{(k)} + \mathbf{Q}^{(k)} \boldsymbol{y} \right) \end{cases} \tag{7}$$

where $k = 0, 1, 2, ..., K-1$, resulting in a K-layer network and $t^{(k)}$ is a scalar which is a key to the convergence rate of network [26], and $\mathbf{I}$ denotes identity matrix. Hence, the structure-aware threshold is defined as:

$$\theta^{(k)} = U\left\{\boldsymbol{\omega}^{(k)}\right\}, \quad k = 0, 1, 2, ..., K-1 \tag{8}$$

where $U\{\cdot\}$ represents the mapping from the k^{th} iteration signal TF reconstruction to the structure-aware thresholds at the k^{th} U-Net block. Generally, in order to achieve a near-ideal TF representation, the proposed U-ISTA controls thresholds adaptively according to the analyzed signal structure by learning local dependencies among signal amplitudes, which makes a compensation for simple sparsity.

3. NUMERICAL EXPERIMENTS

We train the whole U-ISTA with an end-to-end manner. For parameter settings, the number of iterations K is set to 10, and the proposed U-ISTA is based on the knowledge of the 51×51 samples around the origin of ambiguity domain to train the network. The training dataset includes 256-sample synthetic three-component signals, which are composed of overlapped frequency modulated signals with signal-to-noise ratio ranging from 5 to 25 dB. Commonly-used smooth ℓ_1 loss function is utilized to train our model. Additionally, the number of samples can be changed depending on the size of the signal's TFD in test stage, e.g., for a 400-sample real bat echolocation signal [2], we choose 81×81 samples to recover the signal. The U-ISTA is compared to state-of-the-art algorithms including ADTFD [7], ℓ_1-app [9], RSP [1], MDD [6], AOK [5], EMBD [4], and RGK [8]. In the case of synthetic data, ℓ_1 distance to ideal model [9] is utilized as an assessment criteria. For real-life data, Rényi entropy [9] of each TFD is computed as smaller value of Rényi entropy indicates more concentrated TFD. For all reconstruction results, amplitudes are coded logarithmically with a dynamic range of 20 dB.

As shown in Fig. 4(a), for a closely located synthetic signal, ℓ_1-app and RSP both have difficulty in separating the close components, and undesirable distortion mainly stems from the interaction of parallel linear FM (LFM), while ADTFD achieves great performance in this case as directions of ATs are exploited. However, There are still a lot of CTs remaining in the result of ADTFD. High-resolution signal's TFD without CTs is gained by the U-ISTA. In particular, the interaction of parallel LFM components is reconstructed with negligible artifacts. Additionally, the lowest ℓ_1 distance is obtained by the U-ISTA, i.e., near-ideal TFD is gained.

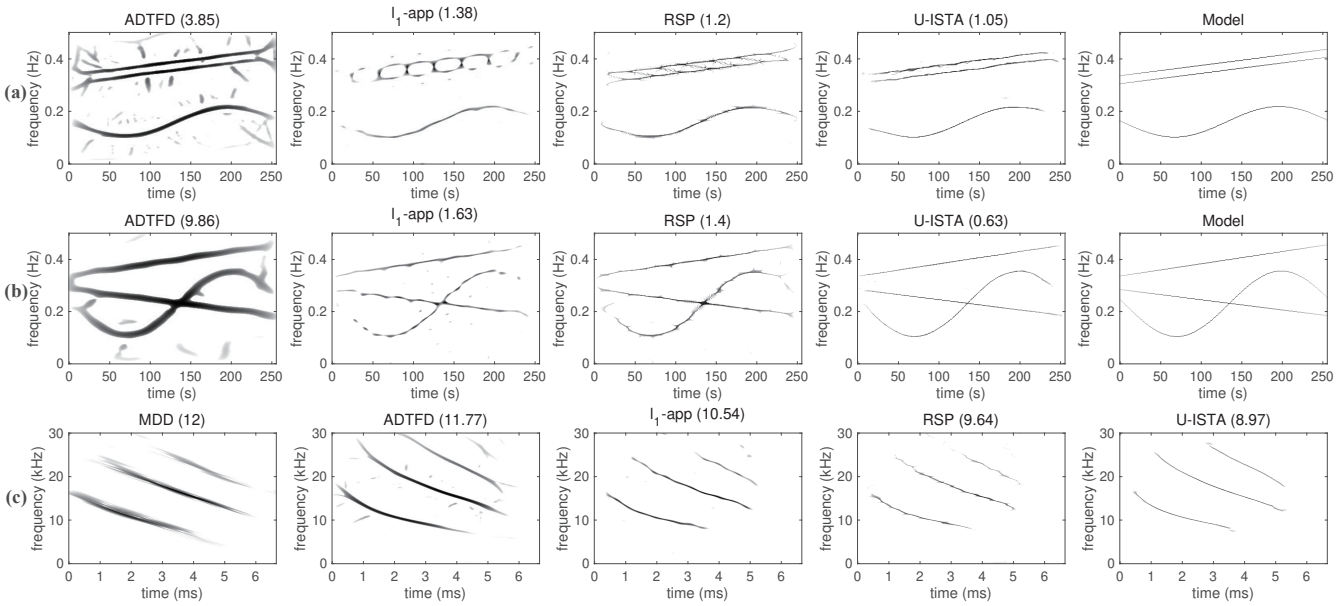


Fig. 4. TF results of synthetic test signals and a real-life bat echolocation signal with **SNR = 10 dB**. ℓ_1 distance to model and Rényi entropy are annotated above each image for synthetic and real-life signals respectively. (a) a synthetic test signal with closely located components. (b) a synthetic test signal with overlapped components. (c) a real-life bat echolocation signal.

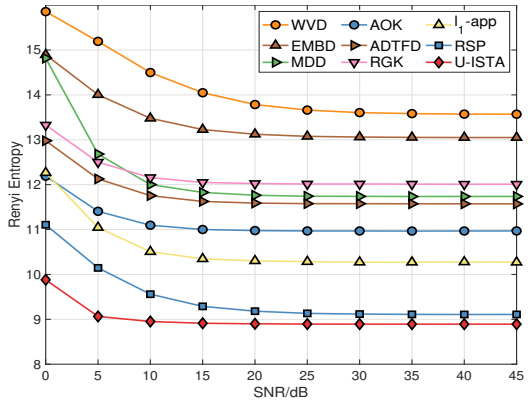


Fig. 5. Rényi entropy performance with various SNR levels using different TF reconstruction algorithms over the real-life bat echolocation signal.

For an overlapped synthetic signal, as illustrated in Fig. 4(b), there is a trade-off between high-resolution and CTs reduction when applying ADTFD. On the other hand, it is hard for ℓ_1-app and RSP to accurately represent the intersection of the TFD though high concentration is achieved. Furthermore, the visualized as well as quantitative analysis indicate that the U-ISTA has the best performance on resolution improvement and CT reduction tasks, i.e., ℓ_1 distance drops as low as 0.63 which is much lower compared to other algorithms.

Though we only train our network using synthetic data, the proposed U-ISTA also gives a desirable representation of a real-life bat echolocation signal [2]. As seen in Fig. 4(c), most algorithms can achieve CT-free TFDs, whereas smooth TFD with high concentration is only gained by the U-ISTA, i.e., ℓ_1-app gives a smoothing representation while resolution is relatively low compared with the U-ISTA, and RSP attains a high-resolution but rough TFD. Quantitative evaluation in terms of Rényi entropy is carried out, and the U-ISTA attains the highest TF resolution when SNR = 10 dB. In this example, Rényi entropy performance using various algorithms with SNR ranging from 0 to 45 dB is shown in Fig. 5, and 100 Monte Carlo runs are implemented at each SNR level. Intuitively, RSP and U-ISTA perform much better than others. Furthermore, it is obvious that the U-ISTA is more robust to noise than RSP as it gives a better representation with low SNR, and Fig. 4(c) illustrates this case as well.

4. CONCLUSION

We propose a new U-ISTA, which unfolds the typical ISTA as a resemble RNN. Structured sparsity is introduced by regarding the U-Net as an adaptive threshold block, thus the TF correlation structure of the analyzed signal can be exploited so that more prior knowledge can be used to reconstruct near-ideal TFD. Experimental results on synthetic and real-life signals validate the effectiveness of our algorithm. Even though the signal is composed of closely located or overlapped components, a CT-free as well as high-resolution TFD is gained. Moreover, despite our model is only trained with synthetic data, the desirable result is also gained on a real-life signal, which examines that our algorithm has generalization performance to some extent.

References

[1] F. Auger and P. Flandrin, "Improving the Readability of Time-Frequency and Time-Scale Representations by the Reassignment Method," *IEEE Trans. Signal Process.*, vol. 43, no. 5, pp. 1068–1089, May 1995.

[2] B. Boashash and S. Ouelha, "Designing High-Resolution Time–Frequency and Time–Scale Distributions for the Analysis and Classification of Non-Stationary Signals: a Tutorial Review with a Comparison of Features Performance," *Digit. Signal Process.*, vol. 77, pp. 120–152, 2018.

[3] B. Barkat and B. Boashash, "A High-Resolution Quadratic Time-Frequency Distribution for Multicomponent Signals Analysis," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2232–2239, Oct. 2001.

[4] B. Boashash, G. Azemi, and J. M. O'Toole, "Time-Frequency Processing of Nonstationary Signals: Advanced TFD Design to Aid Diagnosis with Highlights From Medical Applications," *IEEE Signal Process. Mag.*, vol. 30, no. 6, pp. 108–119, Nov. 2013.

[5] D.L. Jones and R.G. Baraniuk, "An Adaptive Optimal-Kernel Time-Frequency Representation," *IEEE Trans. Signal Process.*, vol. 43, no. 10, pp. 2361–2371, 1995.

[6] B. Boashash and S. Ouelha, "An Improved Design of High-Resolution Quadratic Time-Frequency Distributions for the Analysis of Nonstationary Multicomponent Signals Using Directional Compact Kernels," *IEEE Trans. Signal Process.*, vol. 65, no. 10, pp. 2701–2713, May 2017.

[7] N. A. Khan and B. Boashash, "Multi-Component Instantaneous Frequency Estimation Using Locally Adaptive Directional Time Frequency Distributions," *Int. J. Adapt. Control Signal Process.*, vol. 30, no. 3, pp. 429–442, Mar. 2016.

[8] R. G. Baraniuk and D. L. Jones, "Signal-dependent Time-Frequency Analysis Using a Radially Gaussian Kernel," *Signal Process.*, vol. 32, no. 3, pp. 263–284, Jun. 1993.

[9] P. Flandrin and P. Borgnat, "Time-Frequency Energy Distributions Meet Compressed Sensing," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 2974–2982, Jun. 2010.

[10] S. S. Moghadasian and S. Gazor, "Sparsely Localized Time-Frequency Energy Distributions for Multi-Component LFM Signals," *IEEE Signal Process. Lett.*, vol. 27, pp. 6–10, 2020.

[11] I. Volaric, V. Sucic, and S. Stankovic, "A Data Driven Compressive Sensing Approach for Time-Frequency Signal Enhancement," *Signal Process.*, vol. 141, pp. 229–239, Dec. 2017.

[12] N. Whitelonis and H. Ling, "Radar Signature Analysis Using a Joint Time-Frequency Distribution Based on Compressed Sensing," *IEEE Trans. Antennas. Propag.*, vol. 62, no. 2, pp. 755–763, 2014.

[13] I. Volaric, V. Sucic, and G. Bokelmann, "Sparse Time-Frequency Distribution Calculation with an Adaptive Thresholding Algorithm," in *International Symposium on Image and Signal Processing and Analysis (ISPA)*, Dubrovnik, Croatia, Sep. 2019, pp. 341–346.

[14] H. Zhang, G. Bi, S. G. Razul, and C. M. S. See, "Supervised modulation classification based on ambiguity function image and invariant moments," in *2014 9th IEEE Conference on Industrial Electronics and Applications*, 2014, pp. 1461–1465.

[15] J. R. Fonoliosa and C. T. Nikias, "Wigner Higher Order Moment Spectra: Definition, Properties, Computation and Application to Transient Signal Analysis," *IEEE Trans. Signal Process.*, vol. 41, no. 1, pp. 245, Jan. 1993.

[16] Q. Wu, Y. D. Zhang, and M. G. Amin, "Continuous structure based Bayesian compressive sensing for sparse reconstruction of time-frequency distributions," in *Int. Conf. Digital Signal Process.*, Hong Kong, China, Aug. 2014, pp. 831–836.

[17] L. Jiang, H. Zhang, L. Yu, and G. Hua, "Kernel Learning for High-Resolution Time-Frequency Distribution," *arXiv preprint arXiv: 2007.00322*, 2020.

[18] S. Zhang and Y. D. Zhang, "Crossterm-Free Time-Frequency Representation Exploiting Deep Convolutional Neural Network," *arXiv preprint arXiv: 2007.03570*, 2020.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Med. Image Comput. Comput.-Assisted Intervention*. Springer, 2015, pp. 234–241.

[20] B. Boashash, *Time-Frequency Signal Analysis and Processing: a Comprehensive Reference (Second Edition)*, Academic Press, 2016.

[21] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *J. R. Stat. Soc. Series. B Stat. Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.

[22] S. M. Fosson, "A Biconvex Analysis for Lasso ℓ_1 Reweighting," *IEEE Signal Process. Lett.*, vol. 25, no. 12, pp. 1795–1799, 2018.

[23] I. Daubechies, M. Defrise, and C. De Mol, "An Iterative Thresholding Algorithm for Linear Inverse Problems with a Sparsity Constraint," *Comm. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.

[24] K. Gregor and Yann LeCun, "Learning Fast Approximations of Sparse Coding," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 399–406.

[25] X. Chen, J. Liu, Z. Wang, and W. Yin, "Theoretical Linear Convergence of Unfolded ISTA and Its Practical Weights and Thresholds," in *Adv. Neural Inf. Process. Syst.*, 2018, pp. 9061–9071.

[26] J. Liu, X. Chen, Z. Wang, and W. Yin, "ALISTA: Analytic Weights Are as Good as Learned Weights in LISTA," in *International Conference on Learning Representations (ICLR)*, New Orleans, LA, USA, May 2019.

[27] Y. Jiang, L. Yu, H. Zhang, and Z. Liu, "Learning Cluster Structured Sparsity by Reweighting," *arXiv preprint arXiv: 2007.03570*, 2020.