2018 15th International Conference on
Control, Automation, Robotics and Vision (ICARCV)
Singapore, November 18-21, 2018

Time-frequency Image Enhancement of Frequency Modulation Signals by using Fully Convolutional networks*

Xuan Xia, Fengqi Yu, Chuanqi Liu, Jiankang Zhao, and Tianzhun Wu

Abstract— **The uncertainty principle and cross-term can lead to blur, fake signal components and energy oscillation in time-frequency distribution, deteriorate the results of signal tracking, radar/sonar imaging and parameter estimation. Hence in this paper, we propose a time-frequency image enhancement method based on convolutional neural networks for clearer instantaneous frequency curve. The training data are generated by a frequency modulation signal generator, and then an end-to-end training is performed between Wigner-Ville distributions and time-frequency images. Our networks not only extract underlying features of Wigner-Ville distribution, but also understand the semantic of instantaneous frequency curve and use the priori knowledge of the modulation mode. Therefore, it can correctly recognize and eliminate the cross-terms, and transform the Wigner-Ville distribution to an image that can accurate represent the instantaneous frequency curve. The method is tested by three kinds of frequency modulation signals randomly with Gaussian noise. The results show that it can work properly in most cases and has the generalization ability of multi-component signals.**

I. Introduction

Time-frequency distribution (TFD) is an important tool for signal processing and automatic control. The physical meaning of TFD is the signal energy distribution around the instantaneous frequency (IF) on the time-frequency plane. However, there is no single TFD can achieve it for real-world signals due to the uncertainty principle and the cross-terms. For decades, researchers focus on the design of spectrograms, filter-banks and quadratic methods related to the Wigner-Ville distribution (WVD) for optimal representation of signals [1]. But classic methods such as smoothed pseudo Wigner-Ville distribution (SPWVD) [2], time-frequency reassignment method [3] and B-distribution [4] are hard to meet the requirements of TF resolution, anti-noise ability and cross-item suppression at the same time. Hence many image processing techniques have been introduced into this field [5], resulting in a number of time-frequency image enhancement approaches such as singular value decomposition (SVD) based method [6], wavelet based methods [7], viterbi algorithm based method [8], compressive sensing based method [9] and so on.

In recent years, deep learning has made the pattern recognition of non-stationary signal achieve fruitful results. Convolutional neural networks (CNNs) can achieve high detection and classification accuracies of speech signals [10], biological signals [11] and frequency modulation (FM) signals [12] now. However, how to enhance TFD by deep learning, i.e., overcome the ambiguity of TFD and achieve better IF estimation result by deep learning is still an unexplored problem as far as we know. Related works are mainly focus on the art image processing such as sketch simplification [13] [14] [15] and image synthesis [16] [17]. In common, both of TFD enhancement and sketch simplification can be seen as an image enhancement or an image debluring problem: Amplifying meaningful information and suppressing noise. The network architecture of [13] (SIGGRAPH 2016 model) is the most similar to our approach: it relies on an encoder-decoder architecture for semantic segmentation and is able to generate refined sketches by rough sketches, and it can process any size image since the networks are fully convolutional networks (FCNs) [18]. Yet the sketches don't have geometric structures that can be mathematically described, the training results depend on the quality of the objective data set. In contrast, our method can achieve better training results since the generation of our objective data is supported by theoretical derivation. And our networks are based on U-net for better low-level information delivery [19].

In summary, we propose a time-frequency image enhancement method for FM signals. Our fully convolutional neural network can transform the WVD to a TF image that can accurate represent the IF curve, achieve cross-term suppression and WVD deblurring.

All the WVD and TF images in this paper are inverted for better visual effect. Our networks and a test signal generator are available at github.com/flyfeatherok/TF-Enhancement.

II. Theoretical Background

Given a non-stationary analytical signal

$$s(t) = A(t)e^{j\varphi(t)} \tag{1}$$

where $A(t)$ is the modulated amplitude of the signal, $\varphi(t)$ is the phase of the signal. The density of its instantaneous frequency is given by

$$P(f_i) = \int_{-\infty}^{+\infty} \delta\left[2\pi f_i - \varphi'(t)\right] |s(t)|^2 \mathrm{d}t \tag{2}$$

where $f_i = f_i(t) = \varphi'(t)/2\pi$ is the instantaneous frequency of signal, and $\delta(\cdot)$ is the Dirac distribution. $P(f_i)$ is a marginal

*Research supported by China Postdoctoral Science Foundation (2018M633185), supported in part by the National Key Technology R&D Program (Grant 2016YFC0105002 and 61674162), Guangdong Innovative and Entrepreneurial Research Team Program (No. 2013S046) and Shenzhen Peacock Plan, Shenzhen Key Lab for RF Integrated Circuits, Shenzhen Shared Technology Service Center for Internet of Things, Guangdong government funds (Grant 2013S046 and 2015B 010104005), Shenzhen government funds (Grant JCYJ20160331192843950 and CXZZ20150601160410510), and Shenzhen Peacock Plan.

Xuan Xia, Fengqi Yu and Tianzhun Wu are with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences (University of Chinese Academy of Sciences), Shenzhen, CO 518055 China (e-mail: xuan.xia@siat.ac.cn; fq.yu@siat.ac.cn; tz.wu@siat.ac.cn).

Chuanqi Liu and Jiankang Zhao are with the Department of Instrument Science and Engineering, Shanghai JiaoTong University, Shanghai, CO 200240 China (e-mail: chuanqil@sjtu.edu.cn; zhaojiankang@sjtu.edu.cn).

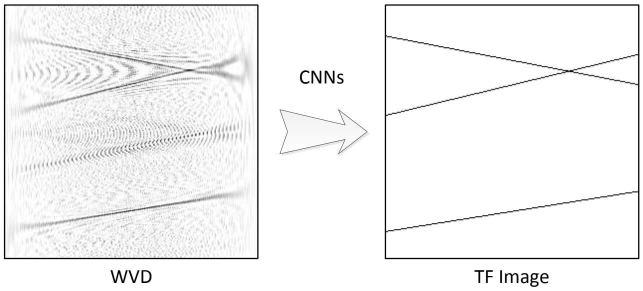

Figure. 1. The purpose of our method.

spectrum that for a given $f_i(t)$, choose only the values $|s(t)|^2$ for which $2\pi f_i = \varphi'(t)$.

The WVD of signal $s(t)$ is defined as

$$W_S(t,f) = \int_{-\infty}^{+\infty} s(t+\tau/2) s^*(t-\tau/2) e^{-j2\pi f\tau} \mathrm{d}\tau \tag{3}$$

WVD is an approach that can distribute signal energy in time-frequency plane. Ideally, we hope that the energy density of f_i in $W_S(t,f)$ can be equal to $P(f_i)$, i.e., the WVD of signal $s(t)$ should represent the signal energy distribution around the IF. However, two factors make them never equal. The first factor is the uncertainty principle defined by the Heisenberg-Gabor inequality [20]

$$\sigma_t \sigma_f \geq \frac{1}{4\pi} \tag{4}$$

where σ_t and σ_f are the time and frequency standard deviations of the signal energy distribution, respectively. The essence of the uncertainty principle is that the resolutions of time and frequency cannot be optimal at the same time. Hence there is always inevitable energy leakage that leads to lower energy density values and WVD blur [21]. The other factor is the cross-term comes from quadratic superposition principle,

$$W_{S_1+S_2}(t,f) = W_{S_1}(t,f) + W_{S_2}(t,f) + 2\mathrm{Re}[W_{S_1S_2}(t,f)] \tag{5}$$

where

$$W_{S_1S_2}(t,f) = \int_{-\infty}^{+\infty} s_1(t+\tau/2) s_2{}^*(t-\tau/2) e^{-j2\pi f\tau} \mathrm{d}\tau \tag{6}$$

is the cross-term of signal $s_1(t)$ and $s_2(t)$. Cross-term can lead to fake signal components or energy oscillation in WVD, resulting in the difficulty of IF estimation.

As shown in Figure. 1, there are three signal components in the WVD. The uncertainty principle leads to the blur of IF curve and the cross-terms lead to fake signal components. Humans can easily recognize real signal components by visual and priori knowledge, but it is hard to translate this cognitive mechanism into a mathematical model that computers can understand. This is because this cognition involves not only the underlying pixel structure, but also the understanding of image semantic. That is just what deep learning can handle. Hence, the purpose of our method is to transform the WVDs to TF images that can accurate represent IF curves of FM signals.

For all that, it must be pointed out that we still cannot overcome the uncertainty principle. The IF curve is also enhanced by the priori knowledge of the modulation mode in our method, thus time-frequency resolution is not improved,

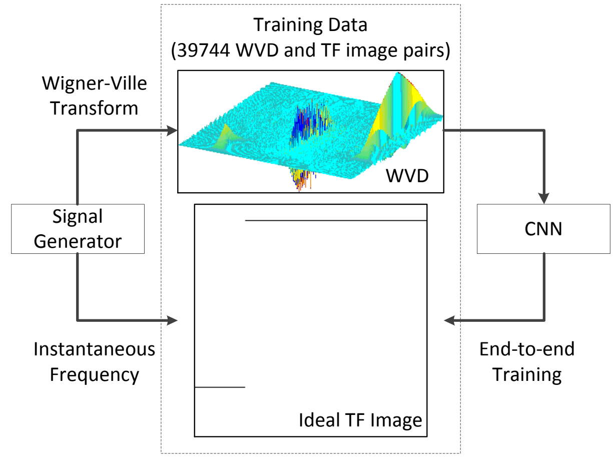

Figure. 2. The training data generation and the training method.

and our networks cannot handle a new FM mode it did not learn. Not only that, due to the black-box of neural networks model, we cannot find a mathematical tool to evaluate the accuracy and the unbiasedness of our method now. Hence we would rather call our method a kind of TF image enhancement method, but not an IF estimation method.

III. METHOD

A. Objective

Our method tries to learn a mapping M from x: $W_S(t,f)$ to a TF image y: $I(t,f)$ that

$$I(t,f) = \begin{cases} P[f_i(t)] = |s(t)|^2 = A^2(t), & \text{where } f = f_i(t) \\ 0, & \text{else.} \end{cases} \tag{7}$$

For a real world signal, x is the discrete WVD (DWVD) of a sampled signal $s(n)$

$$DW_s(n,m) = 2\sum_{k=-N}^{N} h(k) s(n+k) s^*(n-k) e^{-j4\pi mk/L} \tag{8}$$

Where N is the length of $s(n)$, $h(k)$ is a real symmetric window of length $L = 2N + 1$, and this DWVD is an $N\times N$ size matrix. Normally, $A(t)$ is regarded as a constant value in the window, and then $I(n,m)$ can be normalized to a binary image that only contains zeros and ones. Hence the loss of M can be measured by the $L2$ distance between $I(n,m)$ and y,

$$D_{L2} = \sum_{n=1}^{N}\sum_{m=1}^{N} [I(n,m) - y(n,m)]^2 \tag{7}$$

Then the objective of our method is

$$O = \arg\min(D_{L2}) \tag{8}$$

B. Data

The training data and test data are generated by MATLAB. Figure. 2 demonstrates the generation of training data and the training method. Firstly, according to a designed IF, signal generator outputs a corresponding signal. Secondly, the WVDs of this signal are calculated and ideal TF images according to the IF are produced. Thirdly, WVDs are normalized to [-1, 1], then they are put into the networks, resulting in the end-to-end training with TF images. Finally,

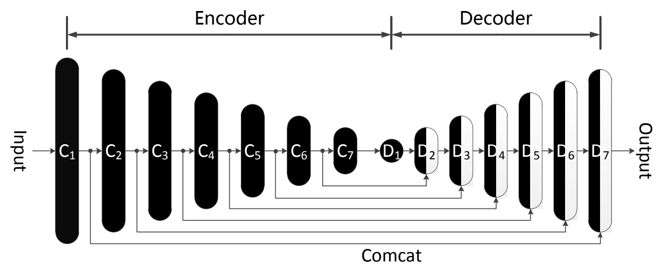

Figure. 3. The architecture of networks in this paper.

the CNN can learn a mapping from WVD to TF image after a number of training epochs.

The signal generator can generate three kinds of typical FM signals: frequency hopping (FH) signal, linear frequency modulation (LFM) signal and sinusoidal frequency modulation (SFM) signal. It can generate a mono-component or a multi-component signal randomly with Gaussian noise in any signal-to-noise-ratio (SNR) and in any duration. This ensures that the training set can approach the real distribution of FM signals as close as possible. All the networks in this paper are trained by random signals in 2dB SNR with 4 seconds (39744 WVD and TF image pairs). For each FM signal, we trained two models: mono-component signal model and two-component signal model. This means that our model should run with right priori knowledge, e.g., LFM signal model should process LFM signals but not SFM signals.

C. Network architectures

As shown in Figure.3, we use an U-net as the basic architecture in our method. The encoder can extract semantic information from input, and the decoder can translate the semantic information to a desired output. Meanwhile, the architecture of U-net can make the low-level information shared between the input and output by adding skip connections. Our method is able to process WVD of any size since our networks are FCNs. The pooling layer is replaced by letting stride equal to 2. All layers are batch normalized [22] and activated by Leaky ReLu functions [23] except the last layer D_7. The output of our networks I_M are limited in [-1, 1] due to the Tanh function, hence the TF image we need should be

$$I = 127.5 \times (I_M + 1) \tag{9}$$

TABLE I. PARAMETERS OF NETWORKS

Layer	Input size	Kernel size	Kernel number	Activation function
C1	N*N*1	8*8*1	8	Leaky ReLu
C2	(N/2)*(N/2)*8	6*6*8	16	Leaky ReLu
C3	(N/4)*(N/4)*16	4*4*16	32	Leaky ReLu
C4	(N/8)*(N/8)*32	4*4*32	64	Leaky ReLu
C5	(N/16)*(N/16)*64	4*4*64	64	Leaky ReLu
C6	(N/32)*(N/32)*64	4*4*64	64	Leaky ReLu
C7	(N/64)*(N/64)*64	4*4*64	64	Leaky ReLu
D1	(N/128)*(N/128)*64	4*4*64	64	Leaky ReLu
D2	(N/64)*(N/64)*(64+64)	4*4*128	64	Leaky ReLu
D3	(N/32)*(N/32)*(64+64)	4*4*128	64	Leaky ReLu
D4	(N/16)*(N/16)*(64+64)	4*4*128	32	Leaky ReLu
D5	(N/8)*(N/8)*(32+32)	4*4*64	16	Leaky ReLu
D6	(N/4)*(N/4)*(16+16)	4*4*32	8	Leaky ReLu
D7	(N/2)*(N/2)*(8+8)	4*4*16	1	Tanh

The detailed parameters of our net are shown in Table. 1. Our parameters of networks are much lesser than the SIGGRAPH 2016 model from [13], to meet the real-time requirements better.

IV. RESULTS AND ANALYSIS

Some TF image enhancement results of FM test signals are showed in Figure.4. The first row demonstrates the TF images enhanced by mono-component signal model, the second row demonstrates the TF images enhanced by two-component signal model, and the third row also demonstrates the TF images enhanced by two-component signal model.

The test data are also generated by our random FM signal generator. In these tests, sampling time is 0.1ms, WVD size is 256, and SNR is 2dB. The random FH signal we used can be written as

$$s(t) = \sum_{n=1}^{M} \exp\{ j2\pi [f_n + m_n(t)] t \},\ t \in [0, T] \tag{10}$$

where f_n is the frequency of the n^{th} signal components, $m_n(t)$ is the n^{th} pseud random sequence, T is the signal duration.

The random LFM signal we used can be written as

$$s(t) = \sum_{n=1}^{M} \exp\left\{ j2\pi \int_0^T [f_n + m_n(t)] \mathrm{d}t \right\} \tag{11}$$

And the random SFM signal we used can be written as

$$s(t) = \sum_{n=1}^{M} \exp\left\{ j2\pi \left[f_n t + m_{an}(t) \int_0^T \sin\left[2\pi \int_0^T m_{fn}(t) \mathrm{d}t \right] \mathrm{d}t \right] \right\} \tag{12}$$

where $m_{an}(t)$ is the n^{th} amplitude modulation of SFM, $m_{fn}(t)$ is the n^{th} sinusoidal frequency of SFM.

Our method transforms WVDs to TF images successfully. CNN can correctly recognize and eliminate the cross-terms in all cases. The blur on the edge of WVD caused by the uncertainty principle is also resolved and TF image shows IF curves clearly. Although the models in the second row are only trained by two-component signals, it can still work with three-component signals in the third row (but not include SFM signal). This proves our net indeed learned “something” and has generalization ability.

However, the quality of output is limited by the quality of WVD. IF curve interrupts occasionally at the end of IF curve or when the signal components are too close. And the tests of three-component SFM signals are ignored because the networks can’t work properly due to the strong energy oscillation.

Results comparison among SPWVD, TF reassigned spectrogram, SIGGRAPH 2016 model from [13] and our TF image enhancement in different SNRs are showed in Fig.5.

The test signal is

$$\begin{aligned} s(t) &= \exp\left[j2\pi \left(100t + 22888t^2 \right) \right] \\ &+ \exp\left[j2\pi (250t) + j19.2 \sin\left(\frac{\pi t}{0.128} \right) \right] \end{aligned} \tag{13}$$

and the signal duration is 256ms, sampling time is 1ms. The time smoothing window and the frequency smoothing window of SPWVD are Hamming windows with 9 and 63

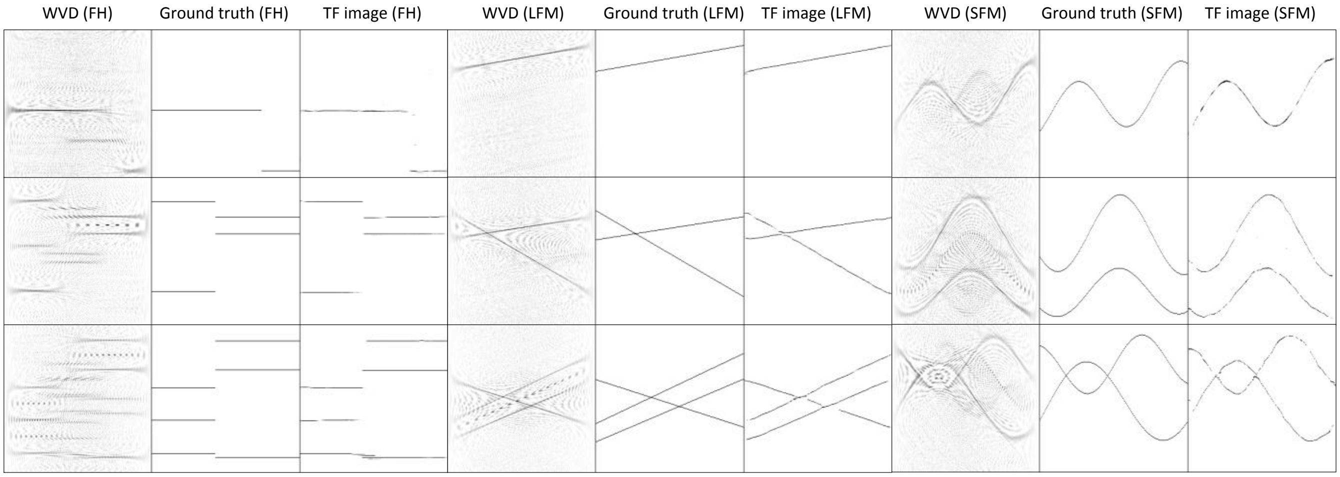

Figure. 4. Some TF image enhancement results of FM test signals.

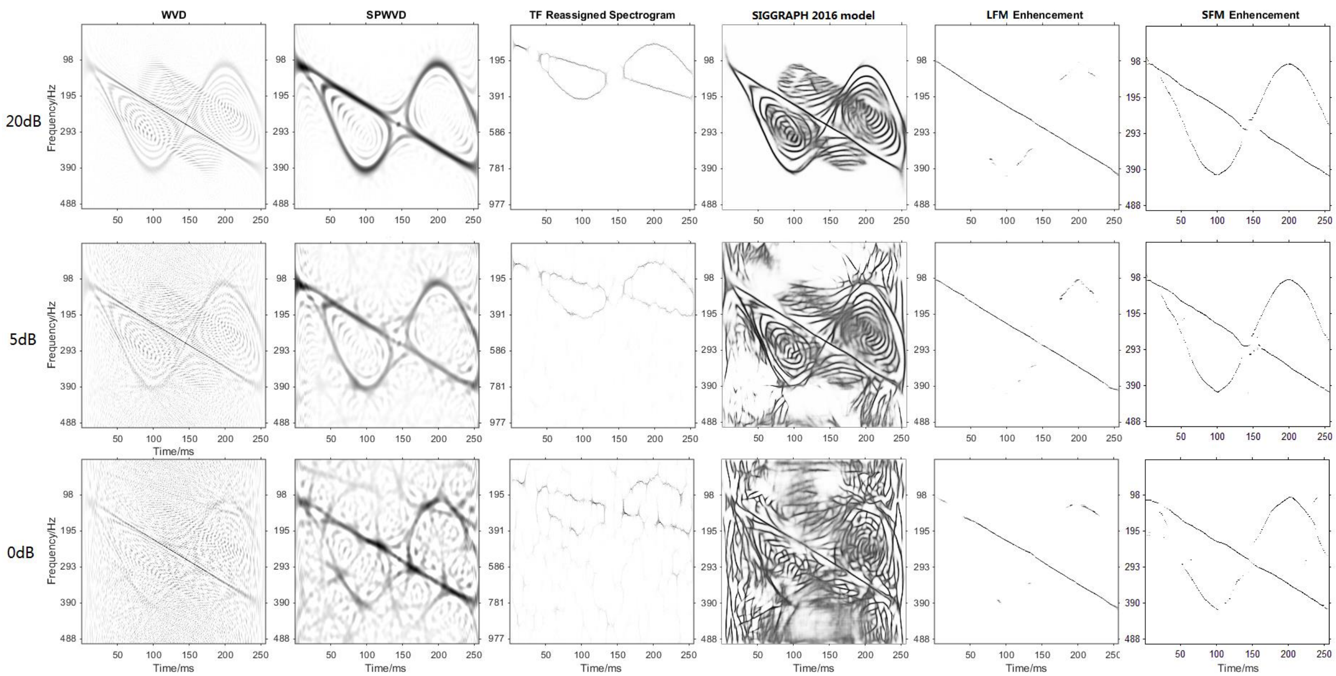

Figure. 5. Results comparison among different methods in different SNRs.

points length, respectively. The frequency smoothing window of TF reassigned spectrogram is a Hamming window with 17 points length. TF Images in "LFM Enhancement" are enhanced by the mono-component LFM signal model; TF Images in "SFM Enhancement" are enhanced by the two-component SFM signal model.

As we can see, SPWVD can bold IF curves and somehow suppress cross-items, meanwhile suffer lower TF resolution and higher uncertainty in low SNR. Reassigned spectrogram has wonderful resolution in high SNR but too sensitive to noise, led to unusefulness in practice. SIGGRAPH 2016 model produces bad results due to the processed images mismatch: the model wasn't trained by WVD. Hence it not only enhanced IF curve, but also cross-items and irrelevant noise. However, it still shows its ability of curve enhancement, the IF curves can be seen in the results. Our method shows the most clearly IF curves in all cases, cross-terms no longer appear in enhancement results and the uncertainty of IF curves were kept at a good level that was better than other methods. However, our networks can only enhance the signals it trained by. Hence only LFM signal appears in the "LFM Enhancement" and the LFM signal is fluctuant in "SFM Enhancement" because our networks think it is a SFM signal. Consider the depth and the parameters number of SIGGRAPH 2016 model, we believe that SIGGRAPH 2016 model can get better results than we do if it trained by WVD. However, our method gives a lightweight neural networks choice for real-time processing. This is also important in the signal processing field.

V. Conclusion

Different from TFD optimal representation or traditional image processing technology, our method has provided a new approach for TF image enhancement via FCNs. FCNs can transform the WVD to a TF image that can represent the IF curve more clearly. Test results have proved that our method has the ability to eliminate cross-terms, deblur the WVD blur caused by the uncertainty principle and has the generalization ability of multi-component signals. Hence its utilization potentiality of signal detection/tracking, radar/sonar imaging and parameter estimation are worth paying attention to.

REFERENCES

[1] Boashash, Boualem, and Samir Ouelha. "An improved design of high-resolution quadratic time-frequency distributions for the analysis of nonstationary multicomponent signals using directional compact kernels." *IEEE Transactions on Signal Processing* 65.10 (2017): 2701-2713.

[2] Gosme, J., C. Richard, and P. Goncalves. "Adaptive diffusion as a versatile tool for time-frequency and time-scale representations processing: a review." *IEEE Transactions on Signal Processing* 53.11(2005):4136-4146.

[3] Auger, François, and Patrick Flandrin. "Improving the readability of time-frequency and time-scale representations by the reassignment method." *IEEE Transactions on signal processing* 43.5 (1995): 1068-1089.

[4] Barkat, Braham, and Boualem Boashash. "A high-resolution quadratic time-frequency distribution for multicomponent signals analysis." *IEEE Transactions on Signal Processing* 49.10 (2001): 2232-2239.

[5] Boashash, Boualem, Nabeel Ali Khan, and Taoufik Ben-Jabeur. “Time–frequency features for pattern recognition using high-resolution TFDs: A tutorial review,” *Digital Signal Processing*, pp. 1-30, 40, 2015.

[6] Iqbal, Naveed, et al. "Automated SVD filtering of time-frequency distribution for enhancing the SNR of microseismic/microquake events." *Journal of Geophysics and Engineering* 13.6 (2016): 964.

[7] Lee, Sin Ho, Jin Bae Park, and Yoon Ho Choi. "Blind spot reduction in wavelet transform-based time–frequency domain reflectometry using Gaussian chirp as mother wavelet." *IET Signal Processing* 8.7 (2014): 703-709.

[8] Li, Po, and Qing-Hai Zhang. "An improved Viterbi algorithm for IF extraction of multicomponent signals." *Signal, Image and Video Processing* 12.1 (2018): 171-179.

[9] Stanković, Ljubiša, et al. "Compressive sensing based separation of nonstationary and stationary signals overlapping in time-frequency." *IEEE Transactions on Signal Processing* 61.18 (2013): 4562-4572.

[10] Hui, Like, et al. "Convolutional maxout neural networks for speech separation." *Signal Processing and Information Technology (ISSPIT), 2015 IEEE International Symposium on. IEEE*, 2015: 24-27.

[11] Schirrmeister, Robin Tibor, et al. "Deep learning with convolutional neural networks for EEG decoding and visualization." *Human brain mapping* 38.11 (2017): 5391-5420.

[12] Dai, Ao, Haijian Zhang, and Hong Sun. "Automatic modulation classification using stacked sparse auto-encoders." *Signal Processing (ICSP), 2016 IEEE 13th International Conference on*. IEEE, 2016: 248-252.

[13] Simo-Serra, Edgar, et al. "Learning to simplify: fully convolutional networks for rough sketch cleanup." *ACM Transactions on Graphics (TOG)* 35.4 (2016): 121.

[14] Donati, Luca, Simone Cesano, and Andrea Prati. "An Accurate System for Fashion Hand-Drawn Sketches Vectorization." *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (2017): 2280-2286.

[15] Simo-Serra, Edgar, Satoshi Iizuka, and Hiroshi Ishikawa. "Mastering sketching: adversarial augmentation for structured prediction." *ACM Transactions on Graphics (TOG)* 37.1 (2018): 11.

[16] Sangkloy, Patsorn, et al. "Scribbler: Controlling deep image synthesis with sketch and color." *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Vol. 2. (2017): 5400-5409.

[17] Lu, Yongyi, et al. "Sketch-to-Image Generation Using Deep Contextual Completion." *arXiv preprint arXiv:1711.08972*(2017).

[18] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015: 3431-3440.

[19] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham, 2015: 234-241.

[20] Oliveira, Paulo M., and Victor Barroso. "Uncertainty in the time-frequency plane." *Statistical Signal and Array Processing, 2000. Proceedings of the Tenth IEEE Workshop on*. IEEE, 2000: 607-611.

[21] Xia, Xuan, et al. "Time–frequency image-based carrier tracking method for Global Navigation Satellite System signal with ultra-high dynamics." *IET Radar, Sonar & Navigation* 10.9 (2016): 1691-1701.

[22] Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." *arXiv preprint arXiv:1502.03167* (2015).

[23] Xu, Bing, Wang N, Chen T, & Li M. "Empirical evaluation of rectified activations in convolutional network." *arXiv preprint arXiv:1505.00853* (2015).