

RESEARCH ARTICLE

Unlocking Signal Processing With Image Detection: A Frequency Hopping Detection Scheme for Complex EMI Environments Using STFT and CenterNet

ZIYI CHEN^{ID}, (Student Member, IEEE), YAOWU SHI, YINGWEI WANG, XINBO LI^{ID},
XIAOHUI YU^{ID}, AND YIRAN SHI^{ID}, (Member, IEEE)

College of Communication Engineering, Jilin University, Changchun 130022, China

Corresponding authors: Xiaohui Yu (yuxiaohui@jlu.edu.cn) and Yiran Shi (shiyiran@jlu.edu.cn)

ABSTRACT Accurate detection and parameter estimation of frequency hopping (FH) signals remain challenging in FH signal-based transmission systems. This study proposes a scheme combining time-frequency analysis (TFA) and deep learning (DL)-based image processing algorithms to alleviate the degradation of detection accuracy and estimation performance caused by complex electromagnetic interference (EMI). A short-time Fourier transform (STFT) was used to obtain the signal spectrogram, which reflects the signal energy in a concentration-dependent manner. Then, a CenterNet-based deep network was employed to identify each FH hop's shape and position, reducing the computational burden via a lightweight neural network while maintaining high recognition accuracy. Inverse mapping from the coordinates to the spectrogram was used to perform parameter estimation in the time-frequency (TF) domain. The estimation error was reduced by precisely locating the centroid of the signal energy using CenterNet. The simulation results demonstrate that the proposed scheme can accurately estimate the FH signal at a low signal-to-noise ratio (SNR) with complex EMI. Furthermore, appropriately determining the optimal parameters of CenterNet to ensure the estimator performance provides a novel approach for integrating DL into signal detection and estimation in complex EMI environments.

INDEX TERMS FH signal, TFA, CenterNet, complex EMI, image detection.

I. INTRODUCTION

Frequency hopping (FH) signals have been increasingly utilized in civilian and military communications. Since their invention, the robust anti-interference capability, low interception probability, and excellent anti-fading effect have made it possible to diversify the human use of electromagnetic waves [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12]. For instance, Code Division Multiple Access-based 3G cellular communications significantly enhance the channel capacity; however, the multipath effect may distort the frequency identification of FH carriers, leading to an elevated Bit Error Rate. While military FH communications

require less channel capacity and spectrum utilization, the adversary's intense Electromagnetic Interference (EMI) usually focuses on covering the hopping bands, resulting in false detection or even communication disruption at the receiver. Thus, the precise detection and estimation of the received FH signals are essential for effective information transmission. To address this issue, an adaptive kernel was utilized in [13] to maximize the mutual information between the input signal and the measurement, thus improving the detection performance after compressive sampling of the full FH spectrum. In [14], a spectrum estimation algorithm for FH signals underlying compressive sensing was introduced, which reconstructed FH signals with missing observations by exploiting the inherent structure of the signal. A compression-aware measurement was presented in [15] to decode FH

The associate editor coordinating the review of this manuscript and approving it for publication was Wen Chen^{ID}.

signals without prior knowledge and pseudo-random code, and an adaptive kernel was designed for compression and decoding. After the immediate exclusion of gross errors, [16] proposed a low-complexity multiple-target detection method for FH signals, based on a small set of random samples. In [17], FH signals were assumed to be blindly separated by exploiting the polarized frequency correlation, and the estimation accuracy was improved by descrambling reconstruction.

The solutions above consider only the case of FH signals under the assumption that no other EMI exists. However, in realistic environments, there are interferences from the nature, numerous electronic devices around the link, and deliberately released disturbances. Furthermore, the FH signal is time-varying, making it difficult to describe its changing patterns visually and to identify FH signals from interferences using traditional time-domain or frequency-domain analysis, especially in complex EMI environment. Fortunately, approaches based on time-frequency analysis (TFA) enable the visualization of signal characteristics in the time-frequency (TF) domain, providing an effective solution for this problem [18], [19], [20], [21], [22], [23].

Among the various TFA methods, the widely used ones include the short-time Fourier transform (STFT) [24], [25], [26], [27], [28], [29], the Wigner-Ville distribution (WVD) [30], [31], [32], [33], and its modified solution, known as the smooth pseudo Wigner-Ville distribution (SPWVD) [3], [34], [35], [36], [37]. However, although the WVD has the highest resolution, its sizeable cross-term interference requires numerous computationally intensive matrix operations. On the other hand, the SPWVD can handle cross-term interference but has limited noise tolerance. In contrast, the STFT is a computationally simple process that does not correlate the signal and prevents cross-term interference. Additionally, because STFT is a windowed transform, apparent edges in the spectrogram with high resolution can be achieved with a smaller window, thus making it easier to analyze the signal features using image processing algorithms. However, in complex EMI environments, interfering signals often contaminate the TF representation (TFR) of the STFT, causing a noticeable drop in detection accuracy. Consequently, accurately identifying FH signals from generated images has become a pressing concern.

With the advent of deep learning (DL) and rapid development of hardware capabilities, image-processing algorithms based on DL have emerged. Various networks focusing on target detection have been proposed in the literature [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], using convolutional kernels to extract image features and improve recognition accuracy by increasing the network depth and innovating structural design. However, these algorithms require abundant network parameters, resulting in a high computational burden. Moreover, these algorithms focus on the entire image and pay insufficient attention to the energy concentration of FH signals in spectrograms. In contrast, CenterNet [48], [49], [50], [51] is a lightweight network with a target centroid as

the primary measurement criterion for recognition, ensuring a high recognition accuracy and speed.

Based on the above analysis, this study proposes an FH signal estimation scenario using STFT and CenterNet, which can accurately estimate FH signals after detecting them in complex EMI. The scenario first transforms the signal into a spectrogram using STFT to show energy aggregation, and then uses CenterNet to detect the FH signals from other interferences by labeling the aggregation area of each hop in the spectrogram. Finally, mapping the pixel coordinates to the TF grid provides an accurate parameter estimation of the FH hops. The contributions of this study are as follows:

- 1) Making a more intuitive distinction, the scenario converts statistical features of both signals and interferences into image features;
- 2) By introducing DL-based image detection methods into signal processing, the fast iterative properties of DL algorithms can rapidly improve the accuracy and anti-interference capability of signal estimation;
- 3) This scenario enriches the toolbox of signal detection and parameter estimation by taking advantage of specific parameters from image detection, such as intersection over union (IoU) analysis;
- 4) CenterNet boosts the computing speed by leveraging parallel computing for machine learning.

The remainder of this paper is organized as follows: Section II establishes an FH signal model with complex EMI. Section III describes TFA for the FH signal, target detection, and parameter estimation. Section IV presents the simulation setup and results evaluation, and Section V provides concluding remarks.

II. SIGNALS AND ENVIRONMENT MODELING

This section analyses the EMI environment and creates a signal model in the TF domain.

A. COMPLEX EMI ENVIRONMENT

Interfering signals and noise always coexist with FH signals in an EMI environment. These interferences typically consist of three categories: fixed-frequency, linear-frequency modulation (LFM), and burst signals. Fixed-frequency signals such as voice broadcasts and TV signals typically have a permanent frequency and usually last for seconds. LFM signals, also known as chirp signals, are signals whose frequency varies linearly with time and whose appearance is typically periodic, such as radar signals. In contrast, burst signals are random and irregular with fixed frequencies and short durations. These interferences and background noise constitute the main obstacles to identifying FH signals [52].

B. MODELING

Assume that there is a mixed-signal $r(t)$ consisting of M FH signals $c_m(t)$, H chirp signals $f_h(t)$, L fixed-frequency signals $g_l(t)$, Q burst signals $j_q(t)$, and $v(t)$ at the receiver during the

observation time T_0 , denoted as

$$r(t) = \sum_{m=1}^M c_m(t) + \sum_{h=1}^H f_h(t) + \sum_{l=1}^L g_l(t) + \sum_{q=1}^Q j_q(t) + v(t) \quad (1)$$

where $v(t)$ represents the additional Gaussian white noise (AWGN) with zero mean and variance of σ^2 .

Given that the m th FH signal has a hopping period of T_m , there are B hops in the observation time T_0 . The carrier frequency corresponding to the b th hop is denoted as f_{m_b} . Because the start time of the first hop cannot be determined and the hop may not be complete, the duration is generally considered to be αT_m with $0 < \alpha < 1$. Then, $c_m(t)$ can be expressed as

$$c_m(t) = a_m \sum_{b=1}^B e^{j(2\pi f_{m_b} t' + \phi_{m_b})} \text{rect}\left(\frac{t'}{T_m}\right) \quad (2)$$

where $t' = t - (b - 1)T_m - \alpha T_m$. a_m is the amplitude of $c_m(t)$ and ϕ_{m_b} is the initial phase of the b th hop with $b \in \{1, 2, \dots, B\}$. $\text{rect}(\bullet)$ is a unit of rectangular pulse function.

From (1), it can be seen that there are no correlations between the FH signals and others; thus, signals other than FH can be unified as additive interference $n(t)$ to be represented as

$$n(t) = \sum_{h=1}^H f_h(t) + \sum_{l=1}^L g_l(t) + \sum_{q=1}^Q j_q(t) + v(t). \quad (3)$$

From (1) and (3), we have

$$r(t) = \sum_{m=1}^M c_m(t) + n(t). \quad (4)$$

III. METHOD

This section provides an overview of FH signal detection, covering spectrogram generation, deep network image detection, and parameter estimation using the detection outputs.

A. SCHEME OVERVIEW

The structure of the proposed scheme is illustrated in Fig. 1. The time-domain input signal is converted into a spectrogram via the STFT. The spectrogram is then fed into the CenterNet as an image for detection, yielding the center point position and bounding box (Bbox) of the FH signal. Finally, the frequencies and hopping period of the FH signal are estimated through inverse mapping based on the coordinates from the image.

B. STFT-BASED SPECTROGRAM GENERATION

STFT is a Fourier transform with added windowing, where a time-domain signal is processed separately by a sliding window and then switched into the TF domain as stacked

spectra. The STFT of the signal $r(t)$ is defined as

$$\begin{aligned} R(t, f) &= \text{STFT}(r(t)) \\ &= \int_{-\infty}^{\infty} r(t)h(\tau - t)e^{-j2\pi f\tau} d\tau \end{aligned} \quad (5)$$

where $h(t)$ denotes a window function. Performing the STFT on both ends of (4) simultaneously yields

$$\begin{aligned} R(t, f) &= \int_{-\infty}^{\infty} \left(\sum_{m=1}^M c_m(t) + n(t) \right) h(\tau - t) e^{-j2\pi f\tau} d\tau \\ &= \int_{-\infty}^{\infty} \sum_{m=1}^M c_m(t) h(\tau - t) e^{-j2\pi f\tau} d\tau \\ &\quad + \int_{-\infty}^{\infty} n(t) h(\tau - t) e^{-j2\pi f\tau} d\tau \\ &= \text{STFT}\left(\sum_{m=1}^M c_m(t)\right) + \text{STFT}(n(t)) \\ &= C_m(t, f) + N(t, f). \end{aligned} \quad (6)$$

To facilitate 2-D image processing, we denote the TF values using the following notation:

$$|R(t, f)| = |C_m(t, f)| + |N(t, f)|. \quad (7)$$

This representation makes it easier to detect the FH signal in a spectrogram, which is the process of detecting $|C_m(t, f)|$ from $|R(t, f)|$.

After converting the time-domain signal into a spectrogram using the STFT, different signals can be easily distinguished from various interferences. As shown in Fig. 2, the FH signal appears as a short horizontal line of consistent length in the spectrogram, whereas the fixed-frequency signal appears as a long horizontal line. The chirp signal appears as diagonal segments with the same slope, frequency band, and period, whereas the burst signal appears as a short horizontal line of random length. The random noise signal covers the entire spectrogram as irregular dots, and the lower the signal-to-noise ratio (SNR), the higher the energy and brightness of the dots in the spectrogram. According to the above analysis, detecting FH signals in a spectrogram is transformed into an image detection problem. The objective is to detect regular short horizontal lines from other shapes in the image.

Current transmitters usually select a relatively clean band to avoid overlapping interference. Therefore, we assume that the FH signal does not overlap with the interfering signals in the spectrogram, which allows for the accurate detection and parameter estimation of the FH signal using the proposed scheme.

C. NETWORK ARCHITECTURE

The architecture of the CenterNet used in this study is illustrated in Fig. 3, which comprises two main components: a convolution-based backbone network and an upsampling network. The backbone network extracts features from the image, and the upsampling network merges these features

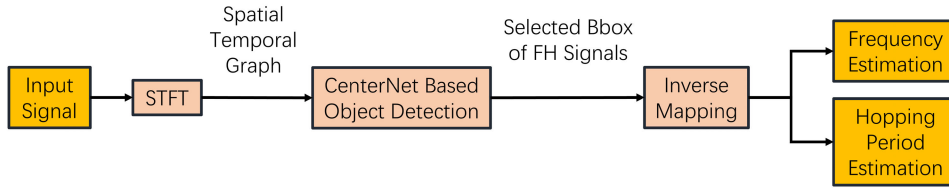


FIGURE 1. Structure of the proposed scheme.

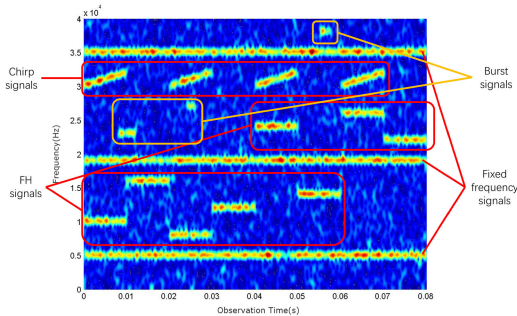


FIGURE 2. Spectrogram of the FH signal with interferences (SNR = 1 dB).

into feature maps. The feature maps are then fed into three different convolutional heads to produce the detection results of the FH signal. These three heads are responsible for the following:

- 1) Extracting the heat map of the features for classification;
- 2) Determining the height and width of the target Bbox;
- 3) Recording the offset between the center point and the target's actual position.

One advantage of CenterNet over traditional CNN-based networks (such as R-CNN and Faster R-CNN) is its insensitivity to the initial size and stretching of the image. Instead of traversing the image using numerous alternate anchors, CenterNet focuses solely on the center point position and anchor size, making the network design straightforward. In this study, the input image was preprocessed from 900×1200 to 512×512 pixels in the RGB format, which contains three channels.

The backbone network is typically selected from ResNet, DLA, and Hourglass [48]. In this study, we selected ResNet-50 as the backbone network, which improves processing speed while ensuring learning accuracy and avoids gradient disappearance and explosion caused by network deepening [53]. After feature extraction using the backbone network, the dimensions of the feature maps were 16×16 with 2048 channels. Subsequently, deconvolution of the feature maps was used as an upsampling tool to obtain heat maps, acquiring feature maps at a dimension of 128×128 pixels with 64 channels [42], [54].

The obtained feature maps are duplicated into three parallel heads to output parameters. First, the convolution depth is equal to the number of desired classifications. As the FH signal is the unique class in this study, the convolution depth of the head was 1. Second, because the Bbox is determined

by the width and height, the convolution depth was 2. Finally, to compensate for the offset during feature extraction, a fine-tuning of the centroid position is performed with a convolution depth of 2. The specific architecture of the detection network is presented in Table 1.

D. TRAINING AND PREDICTION OF THE SPECTROGRAMS

1) POSITION PREDICTION OF THE FH SIGNAL

The keypoint prediction of the FH signal was referenced from [55]. Given a particular FH signal c in the input image $I \in \mathfrak{R}^{W \times H \times 3}$ with width W and height H , the keypoint of the ground truth is $Y \in [0, 1]^{\frac{W}{S} \times \frac{H}{S} \times C}$ with the position of $p \in \mathfrak{R}^2$, which is scaled by the same ratio as the position of $\tilde{p} = \lfloor \frac{p}{S} \rfloor$ and is dispersed by the Gaussian kernel in (8) on the heat map. The output scaling stride S was set to 4 according to the network structure, and the number of classes C was set to 1 in this study. The variance σ_p^2 of the heat map is calculated from the radius r , which is determined using IoU settings.

$$Y_{xyc} = e^{-\frac{(x-\tilde{p}_x)^2 + (y-\tilde{p}_y)^2}{2\sigma_p^2}} \quad (8)$$

Let the corresponding keypoint $\hat{Y} \in [0, 1]^{\frac{W}{S} \times \frac{H}{S} \times C}$ be given by the network prediction, whose position is $\hat{p} \in \mathfrak{R}^2$. We use $\hat{Y}_{xyc} = 1$ to denote that the keypoint is acquired, whereas $\hat{Y}_{xyc} = 0$ denotes the background. Moreover, we use the focal loss for logistic regression, and the loss function can be expressed as

$$L_{pt} = -\frac{1}{M} \sum_{xyc} \begin{cases} (1 - \hat{Y}_{xyc})^\alpha \log(\hat{Y}_{xyc}), & Y_{xyc} = 1 \\ (1 - Y_{xyc})^\beta (\hat{Y}_{xyc})^\alpha \\ \log(1 - \hat{Y}_{xyc}), & otherwise \end{cases} \quad (9)$$

where α and β are the hyperparameters of focal loss. In this study, we take $\alpha = 2$ and $\beta = 4$, respectively. M is the number of keypoints in an image when normalizing the focal loss.

In addition, for the head used for offset prediction, let $\hat{O} \in \mathfrak{R}^{\frac{W}{S} \times \frac{H}{S} \times 2}$ be the offset prediction shared by all classes with offset coordinates $(\delta\hat{x}, \delta\hat{y})$. When using L1 loss to participate in the training, we have

$$L_{off} = \frac{1}{M} \sum_p \left| \hat{O}_{\tilde{p}} - \left(\frac{p}{S} - \tilde{p} \right) \right|. \quad (10)$$

2) SIZE PREDICTION OF THE FH SIGNAL

For the m th FH hop c_m with Bbox coordinates $(x_1^{(m)}, y_1^{(m)}, x_2^{(m)}, y_2^{(m)})$, the center point position p_m and the

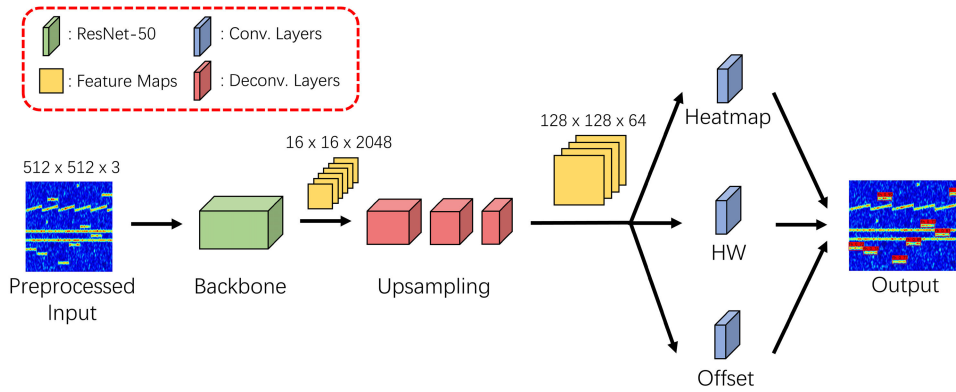


FIGURE 3. Architecture of the CenterNet-based detection network.

TABLE 1. Architecture specifications of the image detection network.

Layer name	Layer details	Output size ($H \times W \times C$)
ResNet-50	Block 1 Convolution: 7×7 , 64 channels, stride 2, BN, ReLU	$256 \times 256 \times 64$
	Block 2 $\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix}^a$ $\times 3$, Max Pool: 3×3 , stride 2	$128 \times 128 \times 256$
	Block 3 $\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix}$ $\times 4$, stride 2 on first Convolution layer	$64 \times 64 \times 512$
	Block 4 $\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix}$ $\times 6$, stride 2 on first Convolution layer	$32 \times 32 \times 1024$
	Block 5 $\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix}$ $\times 3$, stride 2 on first Convolution layer	$16 \times 16 \times 2048$
Upsampling	Deconvolution: 4×4 , 256 channels, stride 2	$32 \times 32 \times 256$
	Deconvolution: 4×4 , 128 channels, stride 2	$64 \times 64 \times 128$
	Deconvolution: 4×4 , 64 channels, stride 2	$128 \times 128 \times 64$
Heat map	Convolution: 3×3 , 64 channels, BN, ReLU	$128 \times 128 \times 1$
	Convolution: 1×1 , 1 channel	
HW	Convolution: 3×3 , 64 channels, BN, ReLU	$128 \times 128 \times 2$
	Convolution: 1×1 , 2 channels	
Offset	Convolution: 3×3 , 64 channels, BN, ReLU	$128 \times 128 \times 2$
	Convolution: 1×1 , 2 channels	

^a[] denotes the Bottleneck block in ResNet-50 [41].

size of the Bbox $s_m(w^{(m)}, h^{(m)})$ are

$$p_m = \left(\frac{x_1^{(m)} + x_2^{(m)}}{2}, \frac{y_1^{(m)} + y_2^{(m)}}{2} \right) \quad (11)$$

and

$$s_m = \left(x_2^{(m)} - x_1^{(m)}, y_2^{(m)} - y_1^{(m)} \right), \quad (12)$$

respectively. Let the size prediction be $\hat{S} \in \mathfrak{R}^{\frac{W}{R} \times \frac{H}{R} \times 2}$, and use L1 loss again to participate in the training. Then we have

$$L_{\text{size}} = \frac{1}{M} \sum_{m=1}^M \left| \hat{S}_{p_m} - s_m \right|. \quad (13)$$

Overall, the loss for the entire training can be expressed as

$$L_{\text{train}} = L_{\text{pt}} + \lambda_{\text{size}} L_{\text{size}} + L_{\text{off}} \quad (14)$$

where the coordinates use the pixel positions in the original image. The weights assigned to each part of the loss were

defined by $\lambda_{\text{size}} = 0.1$. Finally, the Bbox predicted by the network can be expressed as

$$\left(\hat{x} + \delta\hat{x} - \frac{\hat{w}}{2}, \hat{y} + \delta\hat{y} - \frac{\hat{h}}{2}, \hat{x} + \delta\hat{x} + \frac{\hat{w}}{2}, \hat{y} + \delta\hat{y} + \frac{\hat{h}}{2} \right). \quad (15)$$

where \hat{w} and \hat{h} denote the width and height of the Bbox, respectively.

E. PARAMETER ESTIMATION OF THE FH SIGNAL

The FH signal c_m is obtained from network prediction, and the frequency limit of its image representation is displayed between the upper limit F_{max} and the lower limit F_{min} , respectively. The inverse mapping from the image to the signal is calculated as

$$\frac{T_m}{\hat{w}} = \frac{T_0}{W} \quad (16)$$

and

$$\frac{F_{\text{max}} - F_{\text{min}}}{H} = \frac{f_{m_b} - F_{\text{min}}}{\hat{y} + \delta\hat{y}}. \quad (17)$$

The hopping period T_m and carrier frequency f_{m_b} of c_m can be converted to

$$T_m = \frac{\hat{w}}{W} T_0 \quad (18)$$

and

$$f_{m_b} = \frac{\hat{y} + \delta\hat{y}}{H} (F_{\max} - F_{\min}) + F_{\min}. \quad (19)$$

When there is only one FH signal in the environment, the hop duration for the M hop FH signal in a single observation window can be estimated as

$$\bar{T} = \frac{1}{M} \sum_{m=1}^M T_m. \quad (20)$$

IV. NUMERAL EXPERIMENTS AND ANALYSIS

In this section, numerical results are used to demonstrate the effectiveness of the proposed scheme.

A. EXPERIMENT SETTINGS

In this study, the carrier signal operates within the very-low frequency (VLF) band of the International Radio Consultative Committee (CCIR) and the long-wave (LW) band of the International Special Committee on Radio Interference CISPR-25. This band is well suited for air-sea integrated communications and navigation because of the electromagnetic waves transmitted over longer distances, allowing for communication over several kilometers. Furthermore, the band is highly stable, providing robustness against day and night, weather, climate, cosmic rays, and other interferences. Additionally, the band is more penetrative underwater, with less attenuation up to 100 m.

However, the VLF band still faces complex EMI issues, particularly in FH communication. Chirp signals are commonly used in maritime radar systems that may interfere with VLF band communication, as well as the fixed-frequency radiation emitted by electrical devices such as motors and engine blades on aircraft, ships, and offshore platforms. Furthermore, the burst emissions of maritime lightning and ionospheric fluctuations can also affect VLF band communication. Therefore, identifying FH signals in such complex EMI is of great significance.

This study focused on baseband signal processing to simplify the analysis. The SNR of AWGN ranged from -10 dB to 10 dB, with an observation time of 0.08 s and a 128-point Hamming window for the STFT. The parameters of the FH, fixed-frequency, chirp, and burst signals presented in the complex EMI are listed in Table 2. In addition, to guarantee the statistical convenience and accuracy of the experimental results, Monte Carlo samples considering the algorithm's robustness are usually selected by discrete SNR with a step of 1 dB. However, to better understand the prediction results between selected SNRs, it is crucial to conduct sufficient Monte Carlo samples at each SNR. Therefore, this study generated 500 Monte Carlo samples per dB per step for a hopping period of the FH signal between 0.01 s and 0.02 s

with a step of 0.002 s, which contained a total dataset of 63,000 spectrograms.

For the CenterNet parameters, the training, validation, and test sets were segmented at 50%, 25%, and 25%, respectively. The model was pre-trained on MS COCO [56] for weight initialization, and the Adam algorithm with 0.94 momenta and 0.0001 weight decay was used for training. We used a learning rate segmentation strategy to avoid misconvergence. The learning rate was initially set to 1×10^{-4} with a batch size of 64, and then decayed to 1×10^{-5} after 50 epochs with a batch size of 16, and the training terminus was set to 100 epochs.

B. RESULTS EVALUATION

1) STFT PERFORMANCE VERIFICATION

Fig. 4 compares the spectrograms and contour plots of STFT with WVD and SPWVD for the same mixed signal at an SNR of 0 dB. Comparing Fig. 4(a) and 4(b) with Fig. 4(c) and 4(d) reveals that the WVD extensively aggregated the signal representation, as evidenced by the slim and bright outlines, and the noise was effectively suppressed. However, owing to the presence of cross terms, the WVD of signal and interference were represented as non-independent, causing a severe TF overlap. Furthermore, because WVD is the Fourier transform of the autocorrelation function, the values lose linearity over time, resulting in a more confusing TFR, making it challenging to achieve signal detection using image processing algorithms. Comparing Fig. 4(c) and 4(d) with Fig. 4(e) and 4(f), we can see that the SPWVD reduced the TF aggregation and overlap. Nevertheless, as the cross terms remain, the TF overlap generated substantial interference and few correct outlines, which differs from the STFT. The results confirm the incomparable advantage of the STFT in this scheme.

2) RESULTS OBSERVATION OF THE PROPOSED SCHEME

To illustrate the procedure of the proposed scheme more intuitively, we showcase the detection outputs of the FH signal in a mixed signal with AWGN at an SNR of 5 dB, as shown in Fig. 5. The mixed signal in the time domain is shown in Fig. 5(a). The 3-D mesh in the TF domain after applying STFT highlights the concentration of the signal energy, as shown in Fig. 5(b). After 2-D processing to obtain Fig. 5(c), the FH signal can be visually identified from other interferences with the energy distribution. Finally, the Bbox coordinates of the FH hops are revealed in the image through CenterNet, as shown in Fig. 5(d). After the screening, an inverse mapping is performed to derive the frequency and hopping period estimations of the FH signal.

3) NETWORK TRAINING PERFORMANCE

In Fig. 6, we present the trend of the total loss and validation loss with respect to the epoch of our DL architecture, observing that both losses decreased smoothly as the number of training epochs increased. The result indicates

TABLE 2. Parameters of the signal and complex EMI.

Signal type	Lasting time(s)	Jump speed(s)	Frequency (kHz)	Bandwidth (kHz)	Quantity
FH signal	No less than 0.08	[0.01:0.002:0.02]	[2:2:40]	N/A	N/A
Fixed-frequency signals	No less than 0.08	N/A	[2:2:10]	N/A	[1:1:5]
Chirp signals	No less than 0.08	[0.01:0.01:0.04]	N/A	[2:10] ^b	1
Burst signals	[0.001:0.001:0.004] ^a	N/A	[2:2:40]	N/A	[0:1:3]

^a] denotes alternatives for random selection, which forms the value as Minimum: Step: Maximum.

^b] denotes bandwidth alternatives for random selection, which forms the value as Minimum: Maximum.

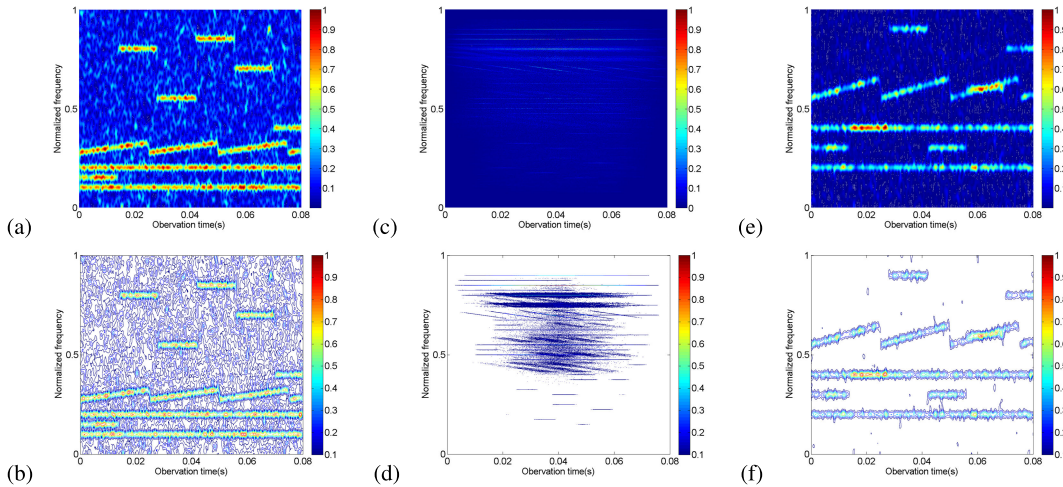


FIGURE 4. Comparison of the TFA algorithms (SNR = 0 dB). (a) STFT spectrogram. (b) STFT contour. (c) WVD spectrogram. (d) WVD contour. (e) SPWVD spectrogram. (f) SPWVD contour.

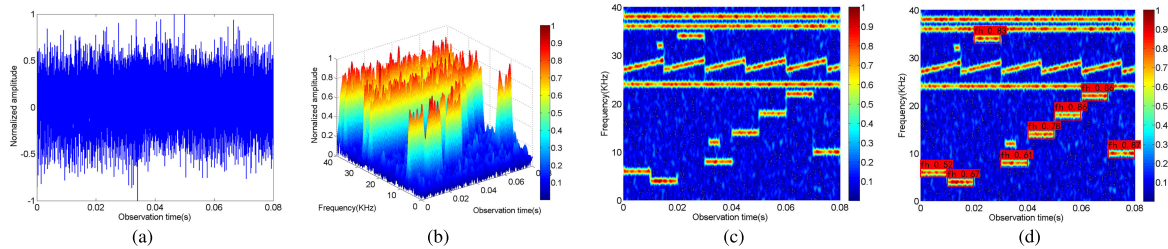


FIGURE 5. A proceeding example of proposed scheme (SNR = 5 dB). (a) Original time-domain signal. (b) 3-D mesh after the STFT. (c) Spectrogram after 2-D Processing. (d) Image with Bbox after CenterNet Detection.

that the architecture parameters were well selected without overfitting. Moreover, the losses reduced steadily when the epoch was between 30 and 50 and decreased sharply again when the learning rate was adjusted to 1×10^{-5} at the 50th epoch, demonstrating an appropriate setting. Finally, the losses became smooth again after 90 epochs and were reduced to the magnitude of 10^{-1} , indicating quick convergence and appropriate fitting of the architecture.

4) PERFORMANCE COMPARISON OF THE DL ARCHITECTURES

In this part, we conducted two experiments to compare the performance of CenterNet with YOLOv7 [47] and YOLOX [46]. The results of the mean average precision (mAP), AP_{50}^{test} , and AP_{75}^{test} obtained from training and testing these three DL architectures on the dataset generated in Section IV-A are summarized in Table 3. As can be seen, the average precision (AP) performance of all architectures is significantly higher than that of MS COCO [56], with APs exceeding 90%. This result originates from the

continued training for the dedicated application in this study based on the pre-trained model, which improves the generalization capability of the network parameters for FH signal detection. In addition, CenterNet obtained 99.25% mAP, 99.91% AP_{50}^{test} , and 99.75% AP_{75}^{test} , outperforming YOLOv7 and YOLOX by 6.31%, 0.01%, 0.03% and 8.39%, 0.03%, 1.21%, respectively. This is because YOLOv7 and YOLOX are both anchor-based architectures that require different aspect ratios of anchors to traverse the image. In contrast, CenterNet is an anchor-free architecture that uses a heat map of the features to represent the probability of the center point location, thus obtaining a better representation of the features of the object. On the other hand, YOLOv7 and YOLOX primarily concentrate on optimization for small-target detection, and the improvement in performance for regular-target detection is comparatively limited. For the application in this study, the FH signal will not appear as a small target in the spectrogram as long as the TF scale is set appropriately; thus, the performance is generally satisfactory.

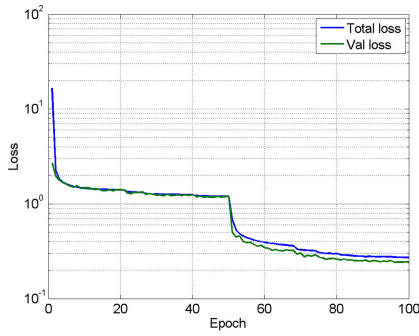


FIGURE 6. Training losses of the CenterNet.

TABLE 3. AP comparison of the DL architectures.

Architecture	mAP(%)	AP ₅₀ ^{test} (%)	AP ₇₅ ^{test} (%)
CenterNet	99.25	99.91	99.75
YOLOv7	92.94	99.90	99.72
YOLOX	90.86	99.88	98.54

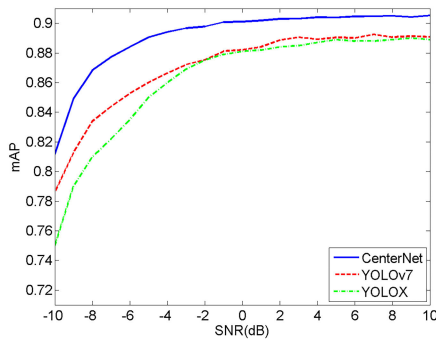


FIGURE 7. mAP comparison of the DL architectures by SNR.

Furthermore, we compared the mAPs of the three DL architectures under different SNRs, as shown in Fig. 7. As can be seen, the mAP of all architectures gradually increased as the SNR increased. The mAP of CenterNet exceeded 90% when the SNR reached -1 dB, whereas the mAPs of both YOLOv7 and YOLOX exceeded 88% when the SNR reached 3 dB. Subsequently, the mAPs of all the architectures converged. It was also evident that the mAPs of YOLOv7 and YOLOX were closer under all SNRs, whereas the mAP of CenterNet was consistently higher by at least 2%. Thus, we can conclude that CenterNet outperforms YOLOv7 and YOLOX in terms of the prediction performance under all SNRs.

5) PERFORMANCE IMPACT OF THE IOU

The key to detecting a signal using image detection is correctly identifying the target representation from the image. Therefore, the next step is to verify the detection accuracy of the proposed method. Here, we chose IoU as the criterion and used miss probability as an index of recognition accuracy. The miss probability P_{miss} is the quantity ratio of detected FH hops N_{detect} to total FH hops N_{total} .

$$P_{miss} = 1 - \frac{N_{detect}}{N_{total}} \quad (21)$$

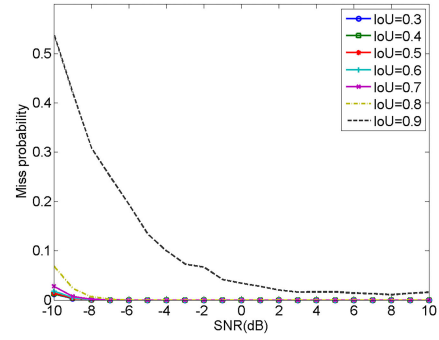
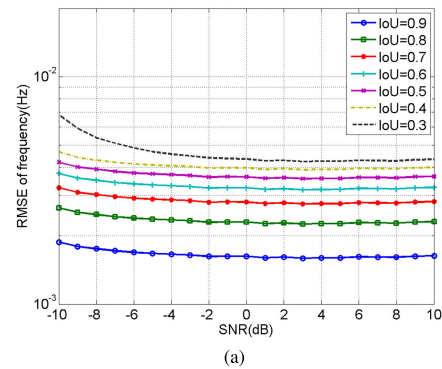
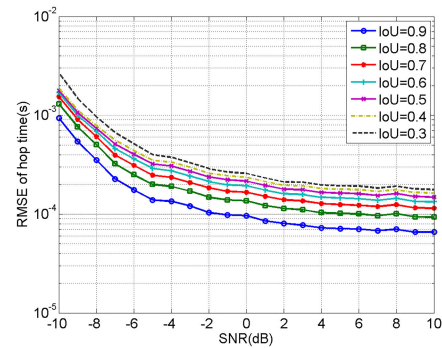


FIGURE 8. SNR-miss probability graph with different IoU thresholds. The detection accuracy is evaluated by miss probability.



(a)



(b)

FIGURE 9. Evaluation error analysis. (a) Frequency error of the FH signal. (b) Hopping period error of the FH signal.

The SNR-miss probability graph for different IoU thresholds is shown in Fig. 8. We observed that if the IoU was too large, the miss probability also became significant. When the IoU decreased gradually, the miss probability also decreased and stabilized until it almost stopped changing when the IoU was less than 0.6. These results can be explained by the fact that, although the Bbox of the detected FH hop is close to the ground truth, there may still be some errors in the exact position. Thus, setting the IoU between 0.3 and 0.6 will ensure the reliability of the detection results to avoid discarding of approximate predictions, which would decrease detection accuracy by having a high threshold value.

In addition, considering $IoU = 0.5$, as an example, the miss probability decreased as the SNR increased. When the SNR exceeded -7 dB, the miss probability dropped to 0, demonstrating that 100% of the FH hops were detected. Furthermore, the miss probability reduced to within 2% when the

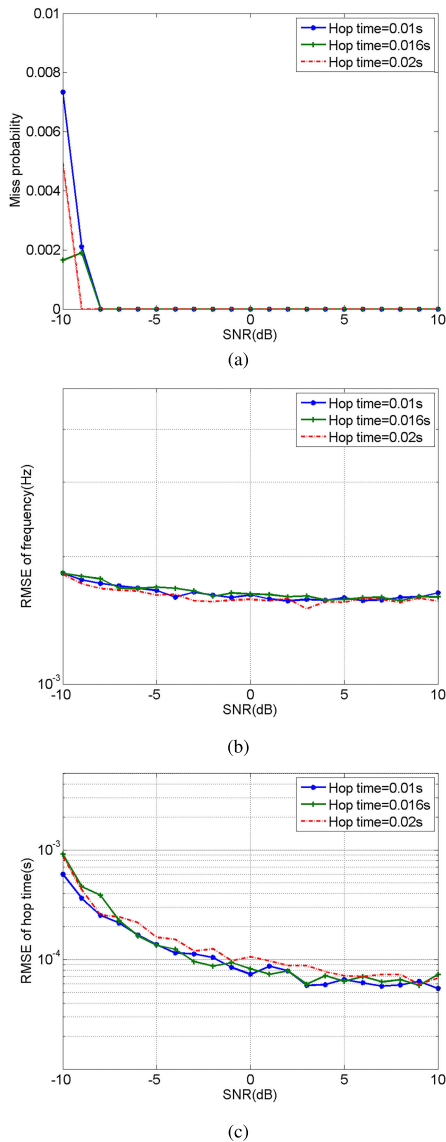


FIGURE 10. Performance impact of different hopping periods (IoU = 0.5). (a) SNR-Miss probability. (b) SNR-RMSE of frequency. (c) SNR-RMSE of hopping period.

SNR reached -10 dB, indicating that the proposed scheme can accurately detect FH signals in complex EMI environments with a low SNR.

Furthermore, we evaluate the estimator using the root mean squared error (RMSE) metric and present the results in Fig. 9. Fig. 9(a) and 9(b) illustrate that the estimation of the frequency and hopping period reached the magnitudes of 10^{-3} and 10^{-4} , respectively. The RMSE of the frequency tended to stabilize after a slight decrease as the SNR increased, converging after the SNR reached -1 dB. On the other hand, the RMSE of the hopping period decreased as the SNR increased and converged after the SNR reached -9 dB, indicating that the proposed scheme achieves desirable estimation accuracy. Furthermore, increasing the IoU decreases the RMSE because a higher IoU implies a stricter screening of the prediction results, which selects a Bbox closer to the

ground truth and lowers the RMSE. However, combined with the analysis of Fig. 8, the reduction in estimation error comes at the cost of recognition accuracy, verifying that an IoU between 0.3 and 0.6 is the optimal choice for performance compromise.

6) PERFORMANCE IMPACT OF THE HOPPING PERIOD

In this part, we compare the miss probability and the RMSE of the proposed scheme at different hopping periods. The results with hopping periods of 0.01 s, 0.016 s, and 0.02 s at IoU of 0.5 are shown in Fig. 10. As can be seen from Fig. 10, the three curves overlapped with each other with an irregular trend, indicating that the hopping period change does not affect the experimental results. It is worth noting that the statistics of other hopping periods still obey the same conclusion. Therefore, Fig. 10 does not present the statistical results for all hopping periods simultaneously to maintain conciseness and to facilitate comparison.

V. CONCLUSION

This paper presents a novel scheme for estimating FH signals in complex EMI environments by combining STFT analysis and CenterNet image detection. The optimal parameters of CenterNet were obtained to ensure the detection accuracy and estimation precision. Simulations showed promising results in detecting and estimating the FH signal in complex EMI at a low SNR, providing a reference for designing and optimizing other DL-based image detection algorithms for detecting and estimating FH signals. Future work will focus on exploring the effects of STFT window size, discovering other TFA algorithms, fingerprinting multiple FH transmitters, detecting interfering signals, analyzing the minimum size of the dataset for model training, and designing adaptive spectrum switching or FH code schemes to mitigate the effects of active or passive interference. The performance of the proposed scheme under more complex background noise requires further exploration. Moreover, the current performance evaluation focuses mainly on the accuracy of the offline scheme, and we will develop an online version in the future to comprehensively evaluate its real-time performance.

REFERENCES

- [1] A. Ephremides, J. E. Wieselthier, and D. J. Baker, "A design concept for reliable mobile radio networks with frequency hopping signaling," *Proc. IEEE*, vol. 75, no. 1, pp. 56–73, Jan. 1987, doi: 10.1109/PROC.1987.13705.
- [2] Q. Chen, E. S. Sousa, and S. Pasupathy, "Multicarrier CDMA with adaptive frequency hopping for mobile radio systems," *IEEE J. Sel. Areas Commun.*, vol. 14, no. 9, pp. 1852–1858, Dec. 1996, doi: 10.1109/49.545707.
- [3] S. Barbarossa and A. Scaglione, "Parameter estimation of spread spectrum frequency-hopping signals using time-frequency distributions," in *Proc. 1st IEEE Signal Process. Workshop Signal Process. Adv. Wireless Commun.*, Apr. 1997, pp. 213–216, doi: 10.1109/SPAWC.1997.630288.
- [4] D. J. Torrieri, "Mobile frequency-hopping CDMA systems," *IEEE Trans. Commun.*, vol. 48, no. 8, pp. 1318–1327, Aug. 2000, doi: 10.1109/26.864169.
- [5] S. Srinivasa and S. Jafar, "Cognitive radios for dynamic spectrum access—The throughput potential of cognitive radio: A theoretical perspective," *IEEE Commun. Mag.*, vol. 45, no. 5, pp. 73–79, May 2007, doi: 10.1109/MCOM.2007.358852.

- [6] Y. Yuan, Z. Huang, and X. Wang, "Detection of frequency-hopping radio frequency-switch transients," *Electron. Lett.*, vol. 50, no. 13, pp. 956–957, Jun. 2014.
- [7] Y.-A. Jung, D. Shin, and Y.-H. You, "Efficient estimation algorithm of carrier frequency offset for LTE machine-type communication using frequency hopping," *IEEE Access*, vol. 7, pp. 177274–177283, 2019.
- [8] J. Ma, Y. Yang, H. Li, and J. Li, "FH-BOC: Generalized low-ambiguity anti-interference spread spectrum modulation based on frequency-hopping binary offset carrier," *GPS Solutions*, vol. 24, no. 3, p. 70, Jul. 2020, doi: [10.1007/s10291-020-00982-3](https://doi.org/10.1007/s10291-020-00982-3).
- [9] C. Li, P. Qi, D. Wang, and Z. Li, "On the anti-interference tolerance of cognitive frequency hopping communication systems," *IEEE Trans. Rel.*, vol. 69, no. 4, pp. 1453–1464, Dec. 2020, doi: [10.1109/TR.2020.3002105](https://doi.org/10.1109/TR.2020.3002105).
- [10] J. Ye, J. Zou, J. Gao, G. Zhang, M. Kong, Z. Pei, and K. Cui, "A new frequency hopping signal detection of civil UAV based on improved K-means clustering algorithm," *IEEE Access*, vol. 9, pp. 53190–53204, 2021, doi: [10.1109/ACCESS.2021.3070491](https://doi.org/10.1109/ACCESS.2021.3070491).
- [11] K. Wu, J. Andrew Zhang, X. Huang, Y. Jay Guo, and R. W. Heath Jr., "Waveform design and accurate channel estimation for frequency-hopping MIMO radar-based communications," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 1244–1258, Oct. 2020, doi: [10.1109/TCOMM.2020.3034357](https://doi.org/10.1109/TCOMM.2020.3034357).
- [12] H. Chougrani, S. Kisseleff, and S. Chatzinotas, "Efficient preamble detection and time-of-arrival estimation for single-tone frequency hopping random access in NB-IoT," *IEEE Internet Things J.*, vol. 8, no. 9, pp. 7437–7449, May 2021, doi: [10.1109/JIOT.2020.3039004](https://doi.org/10.1109/JIOT.2020.3039004).
- [13] F. Liu, M. W. Marcellin, N. A. Goodman, and A. Bilgin, "Compressive sampling for detection of frequency-hopping spread spectrum signals," *IEEE Commun. Lett.*, vol. 64, no. 21, pp. 5513–5524, Nov. 2016, doi: [10.1109/TSP.2016.2597122](https://doi.org/10.1109/TSP.2016.2597122).
- [14] S. Liu, Y. D. Zhang, T. Shan, and R. Tao, "Structure-aware Bayesian compressive sensing for frequency-hopping spectrum estimation with missing observations," *IEEE Trans. Signal Process.*, vol. 66, no. 8, pp. 2153–2166, Apr. 2018, doi: [10.1109/TSP.2018.2806351](https://doi.org/10.1109/TSP.2018.2806351).
- [15] S. Ghanem, "Decoding and measurement of frequency-hopping spread spectrum signals using an adaptive algorithm-based compressive sensing," *Int. J. Commun. Syst.*, vol. 34, no. 3, p. e4675, Feb. 2021, doi: [10.1002/dac.4675](https://doi.org/10.1002/dac.4675).
- [16] B. F. Lo, S. Torborg, and C. K. A. Yeung, "HopSAC: Frequency hopping parameter estimation based on random sample consensus for counter-unmanned aircraft systems," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Nov. 2019, pp. 355–360, doi: [10.1109/MILCOM47813.2019.9020841](https://doi.org/10.1109/MILCOM47813.2019.9020841).
- [17] J. L. Loof and T. G. Pratt, "Frequency-hopped signal source identification in frequency-selective channels," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 6, pp. 3316–3329, Dec. 2019.
- [18] A. Kanaa and A. Z. Sha'ameri, "A robust parameter estimation of FHSS signals using time-frequency analysis in a non-cooperative environment," *Phys. Commun.*, vol. 26, pp. 9–20, Feb. 2018, doi: [10.1016/j.phycom.2017.10.013](https://doi.org/10.1016/j.phycom.2017.10.013).
- [19] L. Zhi, Z. Jianhua, C. Hao, G. Xu, and L. Jian, "Parameter estimation of frequency hopping signals based on analogue information converter," *IET Commun.*, vol. 13, no. 13, pp. 1886–1892, Aug. 2019, doi: [10.1049/iet-com.2019.0057](https://doi.org/10.1049/iet-com.2019.0057).
- [20] J. Wan, D. Zhang, W. Xu, and Q. Guo, "Parameter estimation of multi frequency hopping signals based on space-time-frequency distribution," *Symmetry*, vol. 11, no. 5, p. 648, May 2019. [Online]. Available: <https://www.mdpi.com/2073-8994/11/5/648>
- [21] J. Wu and F. Guo, "Time-frequency parameter estimation method of frequency hopping signal based on morphology method under low SNR," in *Proc. IEEE 6th Int. Conf. Signal Image Process. (ICSIP)*, Oct. 2021, pp. 734–738, doi: [10.1109/ICSIP52628.2021.9688633](https://doi.org/10.1109/ICSIP52628.2021.9688633).
- [22] H. Wang, B. Zhang, H. Wang, B. Wu, and D. Guo, "Hopping time estimation of frequency-hopping signals based on HMM-enhanced Bayesian compressive sensing with missing observations," *IEEE Commun. Lett.*, vol. 26, no. 9, pp. 2180–2184, Sep. 2022, doi: [10.1109/LCOMM.2022.3184173](https://doi.org/10.1109/LCOMM.2022.3184173).
- [23] Z. Jiang, S. Wang, Y. Chen, P. Wang, and L. Gao, "Frequency hopping signal parameter estimation algorithm based on time-frequency point correlation," in *Proc. IEEE 10th Joint Int. Inf. Technol. Artif. Intell. Conf. (ITAIC)*, Jun. 2022, pp. 740–744, doi: [10.1109/ITAIC54216.2022.9836485](https://doi.org/10.1109/ITAIC54216.2022.9836485).
- [24] B. Hinman, J. Bernstein, and D. Staelin, "Short-space Fourier transform image processing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 1984, pp. 166–169, doi: [10.1109/ICASSP.1984.1172374](https://doi.org/10.1109/ICASSP.1984.1172374).
- [25] L. Tong, T. Yinhuai, and L. Jun, "Parameter estimation of FH signals based on STFT and music algorithm," in *Proc. Int. Conf. Comput. Appl. Syst. Model. (ICCASM)*, Oct. 2010, pp. 22–24, doi: [10.1109/ICCASM.2010.5619186](https://doi.org/10.1109/ICCASM.2010.5619186).
- [26] T. Feng and C. W. Yuan, "Blind parameter estimation of frequency-hopping signals based on the time-frequency distribution maxima," *Acta Electronica Sinica*, vol. 39, no. 12, pp. 2921–2925, 2011.
- [27] W. Fu, X. Li, N. Liu, Y. Hei, and J. Wei, "Parameter blind estimation of frequency-hopping signal based on time-frequency diagram modification," *Wireless Pers. Commun.*, vol. 97, no. 3, pp. 3979–3992, Dec. 2017, doi: [10.1007/s11277-017-4710-5](https://doi.org/10.1007/s11277-017-4710-5).
- [28] B. Kaplan, I. Kahraman, A. Gorcin, H. A. Cirpan, and A. R. Ekti, "Measurement based FHSS-type drone controller detection at 2.4 GHz: An STFT approach," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 1–6, doi: [10.1109/VTC2020-Spring48590.2020.9129525](https://doi.org/10.1109/VTC2020-Spring48590.2020.9129525).
- [29] S. Kumawat, M. Verma, Y. Nakashima, and S. Raman, "Depthwise spatio-temporal STFT convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 4839–4851, Apr. 2022, doi: [10.1109/TPAMI.2021.3076522](https://doi.org/10.1109/TPAMI.2021.3076522).
- [30] T. Feng and C. W. Yuan, "Combination of Wigner-Ville distribution and its application to blind parameter estimation of frequency-hopping signals," *J. Xidian Univ.*, vol. 37, no. 6, pp. 1137–1142, 2010.
- [31] T.-C. Chen, "Joint signal parameter estimation of frequency-hopping communications," *IET Commun.*, vol. 6, no. 4, pp. 381–389, 2012.
- [32] Z. Leyuan, L. Chuanhui, L. Bihui, L. Faping, and K. Jiafang, "Research on time-frequency energy distribution characteristics of PSWFs signals based on WVD," in *Proc. 4th Int. Conf. Inf. Commun. Signal Process. (ICICSP)*, Sep. 2021, pp. 134–138, doi: [10.1109/ICICSP54369.2021.9611851](https://doi.org/10.1109/ICICSP54369.2021.9611851).
- [33] Z. Li, X. Zhang, Q. Yang, Y. Xiao, H. An, H. Yang, J. Wu, and J. Yang, "Hybrid SAR-ISAR image formation via joint FrFT-WVD processing for BFSAR ship target high-resolution imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5215713, doi: [10.1109/TGRS.2021.3117280](https://doi.org/10.1109/TGRS.2021.3117280).
- [34] P. Flandrin, "Some features of time-frequency representations of multicomponent signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 1984, pp. 266–269, doi: [10.1109/ICASSP.1984.1172741](https://doi.org/10.1109/ICASSP.1984.1172741).
- [35] Z. Wang, Y. Li, and W. Xu, "A blind parameter estimation method of frequency hopping signal with low SNR," *Int. J. Circuits, Syst. Signal Process.*, vol. 15, pp. 248–253, Apr. 2021.
- [36] S. K. Khare, V. Bajaj, and U. R. Acharya, "SPWVD-CNN for automated detection of schizophrenia patients using EEG signals," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021.
- [37] A. Kamble, P. H. Ghare, and V. Kumar, "Deep-learning-based BCI for automatic imagined speech recognition using SPWVD," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–10, 2023, doi: [10.1109/TIM.2022.3216673](https://doi.org/10.1109/TIM.2022.3216673).
- [38] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587, doi: [10.1109/CVPR.2014.81](https://doi.org/10.1109/CVPR.2014.81).
- [39] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
- [40] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788, doi: [10.1109/CVPR.2016.91](https://doi.org/10.1109/CVPR.2016.91).
- [41] 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, doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031).
- [42] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988, doi: [10.1109/ICCV.2017.322](https://doi.org/10.1109/ICCV.2017.322).
- [43] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
- [44] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully convolutional one-stage object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9626–9635, doi: [10.1109/ICCV.2019.00972](https://doi.org/10.1109/ICCV.2019.00972).
- [45] I. Oksuz, J. R. Clough, B. Ruijsink, E. P. Anton, A. Bustin, G. Cruz, C. Prieto, A. P. King, and J. A. Schnabel, "Deep learning-based detection and correction of cardiac MR motion artefacts during reconstruction for high-quality segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 12, pp. 4001–4010, Dec. 2020, doi: [10.1109/TMI.2020.3008930](https://doi.org/10.1109/TMI.2020.3008930).
- [46] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021, *arXiv:2107.08430*.

- [47] C.-Y. Wang, A. Bochkovskiy, and H.-Y. Mark Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2022, *arXiv:2207.02696*.
- [48] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," *arXiv:1904.07850*, 2019.
- [49] Y. Sun, Z. Li, L. Wang, J. Zuo, L. Xu, and M. Li, "Automatic detection of vehicle targets based on CenterNet model," in *Proc. IEEE Int. Conf. Consum. Electron. Comput. Eng. (ICCECE)*, Jan. 2021, pp. 375–378.
- [50] Z. Dai, J. Yi, L. Jiang, S. Yang, and X. Huang, "Cascade CenterNet: Robust object detection for power line surveillance," *IEEE Access*, vol. 9, pp. 60244–60257, 2021, doi: [10.1109/ACCESS.2021.3072901](https://doi.org/10.1109/ACCESS.2021.3072901).
- [51] H. Zheng, Y. Cui, W. Yang, J. Li, L. Ji, Y. Ping, S. Hu, and X. Chen, "An infrared image detection method of substation equipment combining irectgroup structure and CenterNet," *IEEE Trans. Power Del.*, vol. 37, no. 6, pp. 4757–4765, Dec. 2022, doi: [10.1109/TPWRD.2022.3158818](https://doi.org/10.1109/TPWRD.2022.3158818).
- [52] Y.-J. Zhang, R.-Y. Liu, and H.-J. Song, "A method of the detection of frequency-hopping signal based on channelized receiver in the complicated electromagnetic environment," in *Proc. Int. Conf. Intell. Inf. Hiding Multimedia Signal Process. (IIH-MSP)*, Sep. 2015, pp. 294–297, doi: [10.1109/IIH-MSP.2015.65](https://doi.org/10.1109/IIH-MSP.2015.65).
- [53] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [54] B. Xiao, H. Wu, and Y. Wei, "Simple baselines for human pose estimation and tracking," 2018, *arXiv:1804.06208*.
- [55] H. Law and J. Deng, "CornerNet: Detecting objects as paired keypoints," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 734–750.
- [56] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Proc. ECCV*, 2014, pp. 740–755.



YINGWEI WANG received the B.Eng. degree in mechatronic engineering from the Shenyang Institute of Engineering, Shenyang, China, in 2016, and the M.Sc. degree in control engineering from Jilin University, Changchun, China, in 2019, where he is currently pursuing the Ph.D. degree in control science and engineering with the College of Communication Engineering.

His research interests include deep learning and control technology of acoustic levitation.



XINBO LI received the B.Eng. degree in automation and the M.Sc. and Ph.D. degrees in control theory and control engineering from Jilin University, Changchun, China, in 2002, 2005, and 2009, respectively.

From October 2007 to October 2008, he was a Visiting Ph.D. Student with the School of Electrical and Electronic Engineering, Nanyang Technological University (NTU), Singapore. He is currently a Professor with Jilin University. His

research interests include piezoelectric drive control technology, array signal processing, and time frequency analysis.



ZIYI CHEN (Student Member, IEEE) received the B.Eng. degree in automation from Jilin University, Changchun, China, in 2012, where he is currently the Ph.D. degree in control theory and control engineering with the College of Communication Engineering.

From September 2017 to September 2019, he was a Visiting Ph.D. Student with the School of Electrical and Computer Engineering, Georgia Institute of Technology (GATech), Atlanta, GA,

USA. His currently research interests include frequency hopping signal processing, time frequency analysis, and image processing using machine learning.

Mr. Chen has been a Student Member of the China Institute of Communication (CIC), China, since 2013.



XIAOHUI YU received the B.Eng. degree in automation and the M.Sc. and Ph.D. degrees in control theory and control engineering from Jilin University, Changchun, China, in 2001, 2004, and 2007, respectively.

She is currently a Lecturer with Jilin University. Her research interests include control theory, parameter estimation, and array signal processing.



YAOWU SHI received the B.Eng., M.Sc., and Ph.D. degrees in control theory and control engineering from the Jilin University of Technology, Changchun, China, in 1982, 1985, and 1994, respectively.

From 1986 to 1994, he was a Lecturer with the Jilin University of Technology, where he was promoted to an Associate Professor, in 1994. In 1998, he was a Professor with Jilin University. He is the author or coauthor of one book and several book

chapters, and more than 200 articles. His current research interests include modern cross-spectrum analysis, system identification, DOA estimation, harmonic parameter estimation, and chaotic signal processing.



YIRAN SHI (Member, IEEE) received the B.Eng. degree in control engineering from the Changchun University of Science and Technology, Changchun, China, in 2007, the M.Sc. degree in control engineering from the Lanzhou University of Technology, Lanzhou, China, in 2010, and the Ph.D. degree in control engineering from Jilin University, Changchun, in 2014.

He is currently a Professor with Jilin University. He is the author or coauthor of more than

30 articles. His current research interests include signal processing and model predictive control.

...