

Article

FFSCN: Frame Fusion Spectrum Center Net for Carrier Signal Detection

Hao Huang , Jiao Wang and Jianqing Li *

School of Electronic Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

* Correspondence: lijq@uestc.edu.cn

Abstract: Carrier signal detection is a complicated and essential task in many domains because it demands a quick response to the existence of several carriers in the wideband, while also precisely predicting each carrier signal's frequency centers and bandwidths, including single-carrier and multi-carrier modulation signals. Multi-carrier modulation signals, such as FSK and OFDM, could be incorrectly recognized as several single-carrier signals by using the spectrum center net (SCN) or FCN-based method. This paper designed a deep convolutional neural network (CNN) framework for multi-carrier signal detection by fusing the features of multiple consecutive frames of the broadband power spectra and estimating the information of each single-carrier or multi-carrier modulation signal in the broadband, called frame fusion spectrum center net (FFSCN), including FFSCN-R, FFSCN-MN, and FFSCN-FMN. FFSCN includes three base parts, the deep CNN-based backbone, the feature pyramid network (FPN) neck, and the regression network (RegNet) head. FFSCN-R and FFSCN-MN fusing the FPN out features, which use the Residual and MobileNetV3 backbone, respectively, and FFSCN-MN cost less inference time. To further reduce the complexity of FFSCN-MN, the designed FFSCN-FMN modifies the MobileNet blocks and fuses the features at each block of the backbone. The multiple consecutive frames of broadband power spectra not only preserve the high-resolution ratio of the broadband frequency, but also add the features of the signal changes in the time dimension. Extensive experimental results demonstrate that the proposed FFSCN can effectively detect multi-carrier and single-carrier modulation signals in the broadband power spectrum and outperform SCN in accuracy and efficiency.

Keywords: carrier signal detection; frame fusion; deep learning; convolutional neural network



Citation: Huang, H.; Wang, J.; Li, J. FFSCN: Frame Fusion Spectrum Center Net for Carrier Signal Detection. *Electronics* **2022**, *11*, 3349. <https://doi.org/10.3390/electronics11203349>

Academic Editors: Phivos Mylonas, Katia Lida Kermanidis and Manolis Maragoudakis

Received: 23 September 2022

Accepted: 14 October 2022

Published: 17 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Carrier signal detection in the wideband is usually the first and most vital step of blind communication signal processing. For further study, each sub-carrier signal demodulation, channel decoding, and other subsequent analyses, accurate carrier signal detection in the wideband is a prerequisite.

Similar to the primary signal detection in cognitive radio (CR) [1], carrier signal detection often requires the timely and precise detection of all sub-carrier signals in a non-cooperative communication environment in a wideband signal, which can be formulated as follows [2]:

$$Y(n) = \begin{cases} W(n), & H_0 \\ \sum_{i=1}^M S_i(n) + W(n), & H_1 \end{cases} \quad (1)$$

where $Y(n)$ denotes the received non-cooperative wideband signal, $S_i(n)$ is the i_{th} sub-carrier signal, M denotes the numbers of all sub-carrier signals in the received wideband signal, $W(n)$ denotes the received noise, which can be modeled as the zero-mean additive white Gaussian noise (AWGN), and H_0 and H_1 denote the hypothesis of the absence and the presence, respectively, of the sub-carrier signal in the received wideband signal.

There are many algorithms for carrier signal detection. Energy detection [3–5] is a non-coherent detection method that detects the carrier signal based on the sensed energy. Although it is simple and needs no prior knowledge, the detection performance is subject to the uncertainty of the received signal noise power. Cyclostationary feature detection [6–8] exploits the periodicity in the received narrowband signal to identify the presence of the carrier signal, and it is robust to noise uncertainties. In contrast, this method needs prior knowledge and has a high computational complexity and detection time. At the same time, energy detection and cyclostationary feature detection methods do not estimate the parameters of the carrier signal, only focusing on its presence.

Some improvements have been noted using the double-threshold method [9,10] to overcome these shortcomings. Moreover, by using signal properties such as the amplitude, slope, deflection width, or distance between neighboring deflections, Kim et al. [11] proposed using a slope-tracing-based algorithm to separate the intervals of the carrier signals. However, these thresholds methods have to face the critical issue of discovering the proper threshold values.

Recently, some studies [12–14] found that using deep learning for carrier signal detection achieves more robust and higher performance than threshold-based methods [9–11]. These deep-learning-based methods apply a broadband power spectrum as the input. Since the sub-carrier signals contain both single-carrier and multi-carrier modulation signals, the spectrum cannot drop the features between the two well-distinguished signals. There are some problems in mistakenly detecting the multi-carrier signal as several single-carrier signals, such as modulation by frequency-shift keying (FSK) and orthogonal frequency-division multiplexing (OFDM). Additionally, reducing the model complexity and inference time also remains to be solved.

Usually, people can easily distinguish multi-carrier and single-carrier modulation signals on a broadband short-time Fourier transform (STFT) spectrogram, because the time-frequency analysis of a non-cooperative signal provides us with a visual representation of the spectrum of frequencies of a signal as it varies with time [15]. However, directly using the STFT to detect the carrier signal converts the problem into a 2D picture object detection problem. While the broadband frequency is too large and the resolution of the input image is limited, it is necessary to crop the 2D image and cut it into multiple small pieces for detection. The long time of signal accumulation affects real-time detection, and the cropping of the picture makes it difficult to process the signal at the cut edges, all of which reduce the efficiency and accuracy of signal detection.

This paper describes a new deep CNN framework for carrier signal detection, which uses multiple consecutive frames of broadband power spectra as the input, called FFSCN, including FFSCN-FMN, FFSCN-MN, and FFSCN-R. FFSCN estimates the frequency and bandwidth information of each single-carrier and multi-carrier modulation signal by fusing the multiple frames of the input spectra. The multiple consecutive frames of broadband power spectra not only preserve the high-resolution ratio of the broadband frequency, but also add the features of the signal changes in the time dimension.

We aim to build a deep-learning-based model that optimizes the accuracy of the multi-carrier modulation signal detection in the wideband and improves the model's complexity. To accomplish this, in this work, we introduce (1) new data preprocessing steps for generating multiple consecutive frames of the broadband power spectra, (2) the three FFSCN model architectures for fusing the multiple frames of the inputs and reducing the model complexity, including the new backbone network design and modification of the FPN neck and RegNet head, and (3) new targets and loss functions.

Furthermore, we conducted extensive experiments to demonstrate the effectiveness and efficiency improvements of the proposed methods. Moreover, some ablation studies are presented to illustrate the choice of the model. Our experiments proved that the proposed FFSCN can effectively detect the multi-carrier and single-carrier modulation signals in the broadband and outperform SCN in accuracy and efficiency.

The remainder of this paper is organized as follows. We start with a discussion of related work in Section 2. Section 3 introduces the details of the proposed method. In Section 4, the experimental dataset, training setup, evaluation metrics, results, and some ablation studies are given. Finally, Section 5 concludes the paper.

2. Related Work

Artificial intelligence (AI) technologies, especially deep learning techniques, have now been applied in many areas, such as computer vision (CV), speech recognition, and natural language processing (NLP) [16]. Furthermore, in the wireless communication field, many researchers have performed considerable exploration of deep learning and its application [17–19], as well as communicational signal detection problems [20–22].

Inspired by fully connected networks (FCNs) [23,24] applied in two-dimensional (2D) object semantic segmentation, References [12,13] used an FCN-based model consisting of an encoder and a decoder for carrier signal detection in the broadband power spectrum. The FCN-based methods cannot correctly distinguish between the demarcation points when two or more neighboring subcarriers are very close. Moreover, the FCN-based methods need much post-processing, and their performance degrades severely as the signal-to-noise ratio (SNR) decreases. Reference [14] proposed SCN, an end-to-end deep-learning-based CNN model for carrier signal detection in the broadband power spectrum. SCN regards the carrier signal detection problem as a 1D object localization problem and uses an end-to-end CNN model to regress each sub-carrier's frequency center (FC) and bandwidth (BW) in the broadband power spectrum. It achieved better performance than the FCN-based methods, but cost much more inference time for its complex computation.

In the past few years, many researchers have engaged in designing a small deep neural network architecture for an optimal trade-off between accuracy and efficiency, such as Xception network [25], SqueezeNet [26], ShuffleNet [27], CondenseNet [28], ShiftNet [29], and MobileNets serious [30–32]. Among these methods, MobileNetV1 [30] employs depthwise separable convolution to improve computation efficiency substantially. MobileNetV2 [31] expands on this by introducing a resource-efficient block with inverted Residuals and linear bottlenecks. Moreover, MobileNetV3 [32] uses a combination of hardware-aware network architecture search (NAS) complemented by the NetAdapt algorithm and subsequently improved through novel architecture advances.

In this study, we propose the FFSCN models. As an upgrade to SCN, we replaced the ResNet backbone with a MobileNetV3 base backbone to reduce network computation; moreover, we incorporated a frequency center (FC) shift regression in RegNet to correct the FC prediction. In particular, we created a fusion block based on the MobileNetV3 block, which is utilized in the MobileNet backbone. Extensive experimental results demonstrated that the proposed FFSCN can effectively detect multi-carrier and single-carrier modulation signals in the broadband and outperform current deep-learning-based approaches in accuracy and efficiency.

3. Methodology

3.1. Data Preprocessing

In this study, we employed multiple consecutive frames of the broadband power spectra as the input of the proposed FFSCN for carrier signal detection. The Welch method [33,34] was used to obtain the broadband power spectrum.

First, the N -point frame of the received signal sequence $Y(n)$ is subdivided into K overlapping segments, each with a length of M . Thus, the l_{th} data segments can be represented as

$$Y_l(n) = Y(n + iD) \quad \begin{matrix} n = 0, 1, \dots, M - 1 \\ l = 0, 1, \dots, K - 1 \end{matrix} \quad (2)$$

where iD is the starting point for the l_{th} data segments and $M - D$ is the overlap between each of two neighbor segments.

Then, for each segment, a window function $w(n)$ of length M is used to window the data prior to computing the periodogram. The result is

$$\tilde{P}_{xx}^{(l)}(\omega) = \frac{1}{MU} \left| \sum_{n=0}^{M-1} Y_l(n)w(n)e^{-j\omega n} \right|^2 \quad l = 0, 1, \dots, K-1 \quad (3)$$

where ω is the frequency of the received signal and U is a normalization factor for the power in the window function and is selected as

$$U = \frac{1}{M} \sum_{n=0}^{M-1} w^2(n) \quad (4)$$

The Welch power spectrum estimation is the average of the K -modified periodograms, namely:

$$P_{xx}^W(\omega) = \frac{1}{K} \sum_{l=0}^{K-1} \tilde{P}_{xx}^{(l)}(\omega) \quad (5)$$

Next, the network input is a matrix that can be formulated by

$$\mathbf{P} = 10 \cdot \lg \left(\left[P_{xx}^W(\omega_0), P_{xx}^W(\omega_1), \dots, P_{xx}^W(\omega_r) \right]^T \right) \quad (6)$$

where r denotes the number of consecutive frames of the broadband power spectra; furthermore, the logarithmic transformation is used to convert power to decibels, which scales the numerical range of the spectra.

Finally, we adopted zero mean normalization to normalize the network input matrix, which follows as

$$\tilde{\mathbf{P}} = \frac{\mathbf{P} - \bar{\mathbf{P}}}{\sigma(\mathbf{P})} \quad (7)$$

where $\bar{\mathbf{P}}$ and $\sigma(\mathbf{P})$ are the mean and standard deviation of all the elements in the matrix \mathbf{P} , respectively.

3.2. FFSCN Architecture

FFSCN was built based on the SCN model, which consists of three main parts: the deep-CNN-based backbone, the FPN neck, and the RegNet head, as Figure 1 shows. In this work, to accurately detect the single-carrier and multi-carrier signals, we used multiple consecutive frames of the broadband power spectra in the wideband to replace the single-frame input in SCN. Therefore, the backbone and FPN neck will extract more valuable features of the sub-carriers, and our aim was to fuse these features, which is the difference between the FFSCN and SCN methods.

In FFSCN-R, we added an adaptive average pooling layer between the FPN neck and the RegNet head to fuse the output features of the FPN neck output features. However, compared to SCN, multi-frame input requires significantly more processing in the Residual backbone and the FPN neck, which reduces the network's inference speed and prevents it from responding quickly to the burst signals in the wideband. Therefore, FFSCN-MN uses the MobileNetV3 backbone, not the Residual backbone, to reduce the amount of the network. FFSCN-R and FFSCN-MN both fuse the features before RegNet. Although this is an effective solution for multi-carrier signal detection, it still sacrifices too much time in the backbone and the FPN neck parts. To further improve the network's performance, we modified the MobileNetV3 backbone and propose FFSCN-FMN, which fuses the multi-frame input features in all the blocks of the Fusion-MN backbone. Moreover, FFSCN-FMN optimizes the network complexity and improves the detection performance.

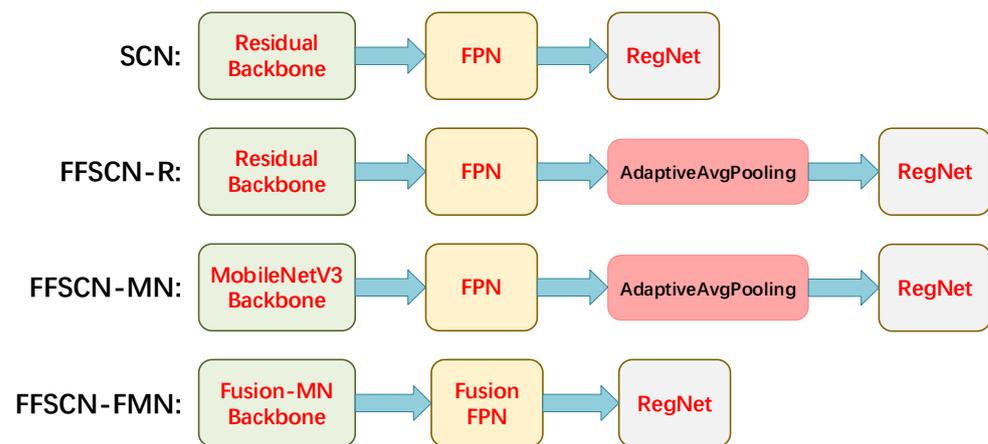


Figure 1. The architecture comparison of the FFSCN and SCN models.

3.2.1. Network Backbones

The backbone network is the fundament of a deep learning object detection model. Figure 2 shows the basic block specification architecture of three deep-CNN-based backbone networks in our work.

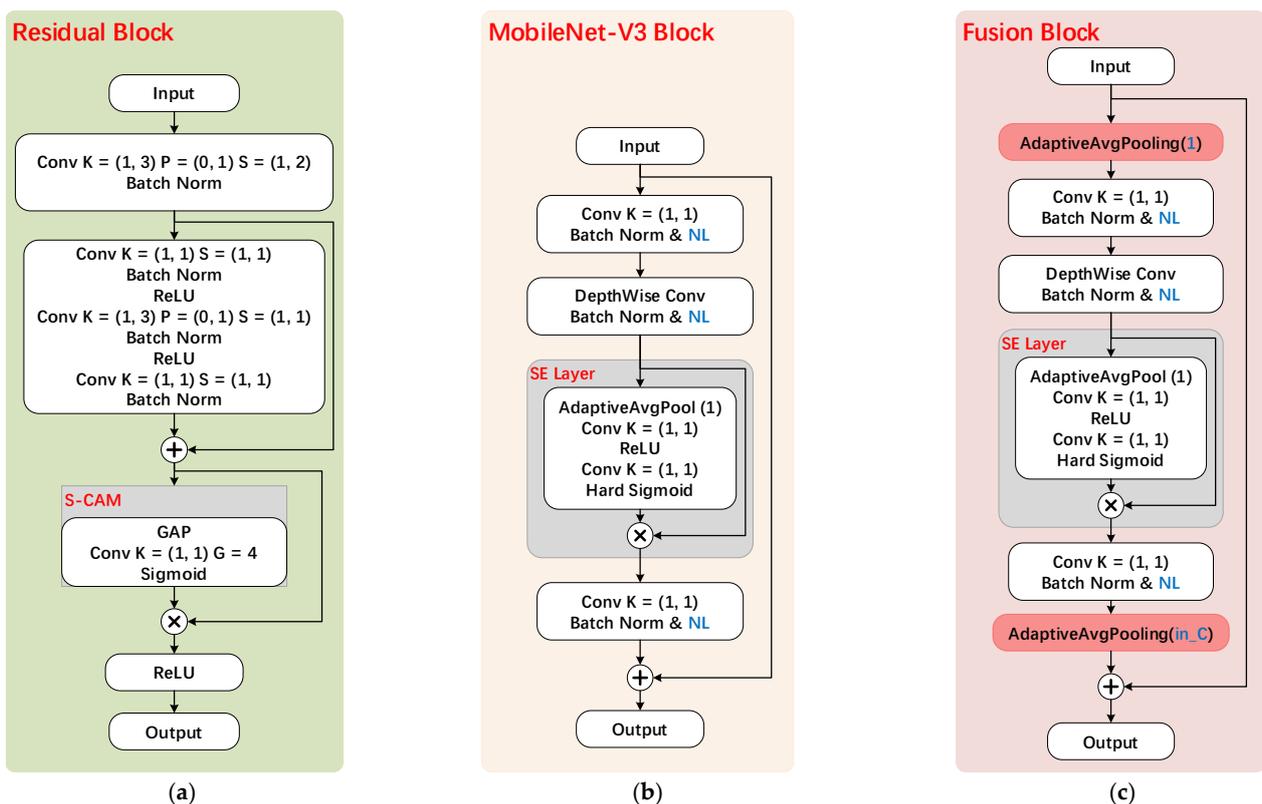


Figure 2. The basic block structure of the deep-CNN-based backbone network. (a) The Residual block. (b) The MobileNetV3 block. (c) The proposed Fusion block, consistent with the MobileNetV3 block, uses a different nonlinear (NL) activation depending on the layer, and in_C denotes the input channels.

Firstly, both FFSCN-R and SCN use the same Residual backbone network, which is modified by the deep residual network (ResNet) [35], and we added a simplified channel attention module (S-CAM) [36] prior to the last nonlinear activation of the Residual block. In SCN [14], we elaborated on the specification structure of the Residual backbone and block.

Then, by replacing the Residual backbone network with the MobileNetV3 backbone, we propose the FFSCN-MN model, and Figure 2b shows the MobileNetV3 block structure. MobileNets are based on a streamlined architecture that uses depthwise separable convolutions to build lightweight deep neural networks. MobileNetV3 adds Squeeze-and-Excite [37] in the inverted Residual with the linear bottleneck of the MobileNetV2 block. Furthermore, MobileNetV3 uses the hard-swish nonlinear activation to enhance the network inference speed [32,38]. The hard-swish function is as follows:

$$f(x) = \begin{cases} 0, & \text{if } x \leq -3 \\ x, & \text{if } x \geq +3 \\ \frac{x(x+3)}{6} & \text{otherwise} \end{cases} \quad (8)$$

Next, in FFSCN-FMN, we added two adaptive average pooling layers at the beginning and the end of the MobileNetV3 block, referred to as the Fusion block, as illustrated in Figure 2c. After adding the two adaptive average pooling layers, the features of the consecutive frames of inputs are no longer kept as separate, but are fused to be a whole. Meanwhile, because the first adaptive average pooling layer fuses the multiple frames of inputs into one frame, compared with the MobileNetV3 block, the Fusion block also reduces the amount of computation.

However, the original MobileNet only downsampled the input scale five times, which is not enough to extract the useful features for the carrier detection task, according to the experimentation in SCN [14]. Therefore, we added some Fusion blocks to increase the downsample times. All the added Fusion blocks' strides were set to 2, and the nonlinear activations were hard-swish functions. The specification for the FFSCN-FMN backbone network is shown in Table 1, and FFSCN-MN uses the same structure, but with MobileNetV3 blocks in the operator.

Table 1. Specification for the FFSCN-FMN backbone network.

Input Shape	Operator	Exp Size	Out Shape	Out Scale	SE	NL	Stride
1 × 10 × 16,384	Conv2d	-	16 × 10 × 8192	P1	-	HS	2
16 × 10 × 8192	Fusion block, 1 × 3	16	16 × 10 × 8192	-	-	RE	1
16 × 10 × 8129	Fusion block, 1 × 3	64	24 × 10 × 4096	P2	-	RE	2
24 × 10 × 4096	Fusion block, 1 × 3	72	24 × 10 × 4096	-	-	RE	1
24 × 10 × 4096	Fusion block, 1 × 5	72	40 × 10 × 2048	P3	✓	RE	2
40 × 10 × 2048	Fusion block, 1 × 5	120	40 × 10 × 2048	-	✓	RE	1
40 × 10 × 2048	Fusion block, 1 × 5	120	40 × 10 × 2048	-	✓	RE	1
40 × 10 × 2048	Fusion block, 1 × 5	240	80 × 10 × 1024	P4	-	HS	2
80 × 10 × 1024	Fusion block, 1 × 5	200	80 × 10 × 1024	-	-	HS	1
80 × 10 × 1024	Fusion block, 1 × 5	184	80 × 10 × 1024	-	-	HS	1
80 × 10 × 1024	Fusion block, 1 × 5	184	80 × 10 × 1024	-	-	HS	1
80 × 10 × 1024	Fusion block, 1 × 5	480	112 × 10 × 1024	-	✓	HS	1
112 × 10 × 1024	Fusion block, 1 × 5	672	112 × 10 × 1024	-	✓	HS	1
112 × 10 × 1024	Fusion block, 1 × 5	672	160 × 10 × 512	P5	✓	HS	2
160 × 10 × 512	Fusion block, 1 × 5	480	160 × 10 × 512	-	✓	HS	1
160 × 10 × 512	Fusion block, 1 × 5	480	160 × 10 × 512	-	✓	HS	1
160 × 10 × 512	Fusion block, 1 × 5	480	80 × 10 × 256	P6	✓	HS	2
80 × 10 × 256	Fusion block, 1 × 5	480	80 × 10 × 128	P7	✓	HS	2
80 × 10 × 128	Fusion block, 1 × 5	480	80 × 10 × 64	P8	✓	HS	2
80 × 10 × 64	Fusion block, 1 × 5	480	80 × 10 × 32	P9	✓	HS	2
80 × 10 × 32	Fusion block, 1 × 5	480	80 × 10 × 16	P10	✓	HS	2
80 × 10 × 16	Fusion block, 1 × 5	480	80 × 10 × 8	P11	✓	HS	2

Here, Exp size denotes the number of channels in the expansion layer. SE denotes whether there is a Squeeze-and-Excite in that block. NL denotes the type of nonlinear activation used. HS denotes hard-swish. RE denotes ReLU.

In SCN, we found that the performance of SCN-11× is almost comparable to that of SCN-13×, but the inference time is shorter [14]. Therefore, in this paper, the downsample times of our proposed FFSCN models were set to 11.

3.2.2. The FPN Neck

In this work, as Figure 3a shows, an adaptive average pooling layer was added at the end of the original FPN neck in [14], which is used to fuse the multiple consecutive broadband power spectra features and is still called the FPN neck. The FPN neck fuses all the top-to-bottom scale features of the backbone in FFSCN-R and FFSCN-MN. Compared with the original SCN, the amount of computation of the FPN neck is times the number of consecutive frames of broadband power spectra inputs. To further optimize the efficiency of the FFSCN-FMN model, we propose the Fusion FPN neck by applying an adaptive average pooling layer after each backbone feature input of the original FPN neck, shown in Figure 3b. The Fusion FPN neck fuses the consecutive frames of input features first and keeps the same amount of computation as the original SCN. Additionally, in FFSCN-R, there are 256 Conv layer channels in the FPN neck, while in FFSCN-MN and FFSCN-FMN, there are 64 Conv layer channels.

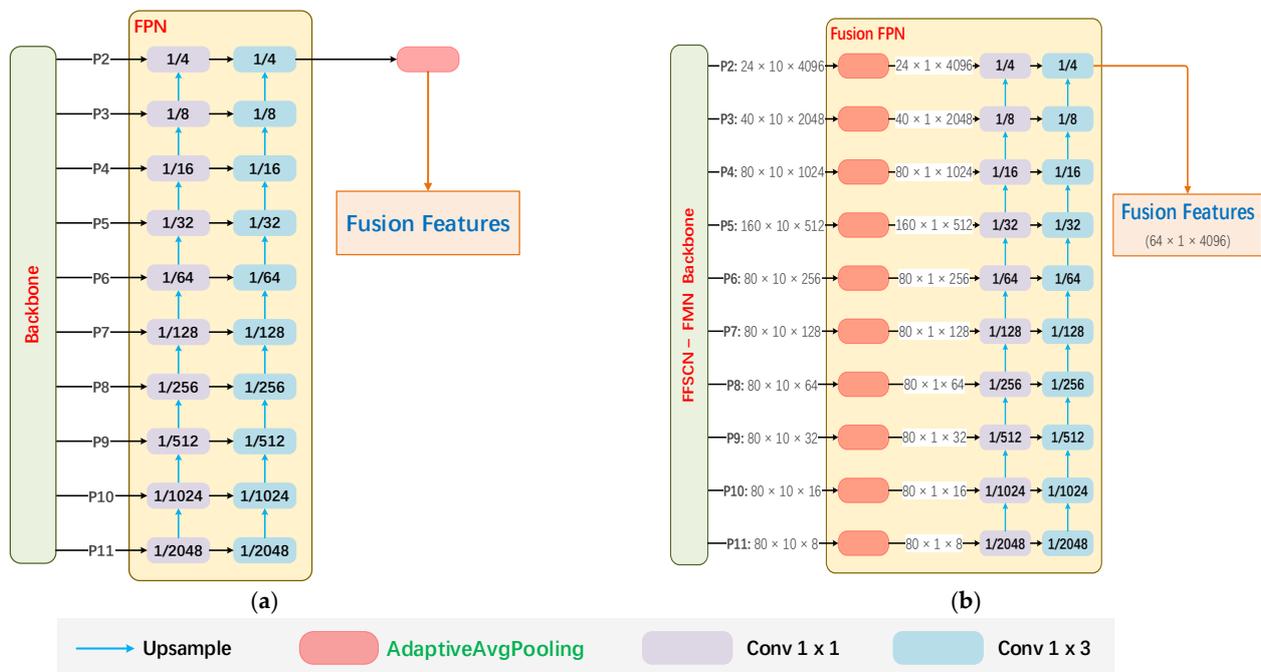


Figure 3. (a) The FPN architecture is used in FFSCN-R and FFSCN-MN. (b) The Fusion FPN architecture is used in FFSCN-FMN.

3.2.3. The Regression Network Head

Figure 4 gives the RegNet head of this work, and we added a frequency center (FC) shift regression branch compared with that in the original SCN. Considering that we regressed the FC prediction in the 1/4 scale of the input length, a shifting bias exists when we set the target FC point as an integer. Therefore, the FC shift regression is to fix the bias, and it consists of the same structure as FC regression and BW regression, a depthwise separable convolutional layer [25] with 256 channels, rectified linear unit (ReLU) [39], and a 1 × 1 Conv with one channel in common.

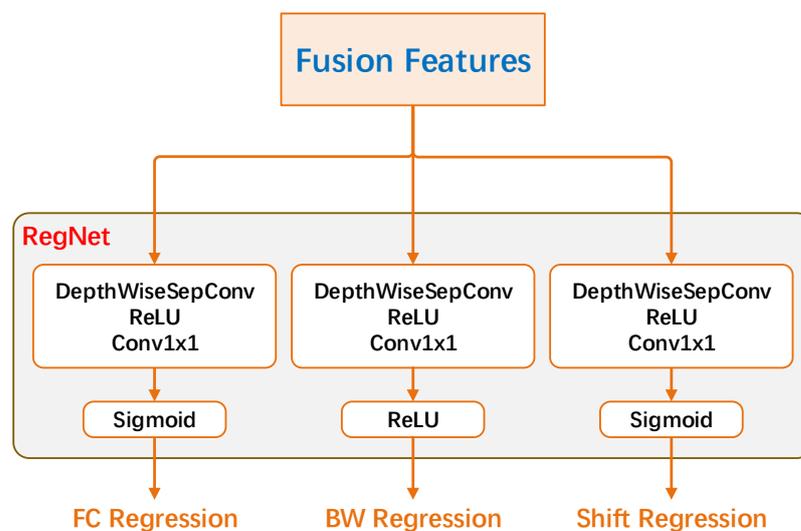


Figure 4. The RegNet head in the FFSCN model.

3.3. FFSCN Targets and Loss Function

In this work, the proposed FFSCN regresses three sets of prediction key points, the power spectrum distribution (PSD) prediction for all subcarrier FC positions and the corresponding BW and FC Shift bias predictions. Here, the PSD and BW predictions are the same as those in the original SCN [14], and the loss functions are formulated as follows:

$$L_{psd} = -\frac{1}{N} \sum_{i=1}^L \begin{cases} (1 - P_i)^\alpha \log(P_i) & \text{if } Y_i = 1 \\ (1 - Y_i)^\beta (P_i)^\alpha \log(1 - P_i) & \text{otherwise} \end{cases} \tag{9}$$

$$L_{bw} = -\frac{1}{N} \sum_{k=1}^N |\hat{W}^k - W^k| \tag{10}$$

$$\hat{W}^k = \frac{L \times BW^k}{BSW} \tag{11}$$

where L_{psd} and L_{bw} denote the PSD loss and BW loss, N denotes the number of all subcarriers in the power spectra input, L denotes the input spectrum length, α and β are hyper-parameters and set to 2 and 4, respectively, P_i denotes the score at the i_{th} point in the predicted PSD, and Y_i denotes the ground-truth PSD. BSW denotes the broadband power spectrum bandwidth; BW^k denotes the k_{th} subcarrier bandwidth; \hat{W}^k . and W^k are the BW ground-truth and prediction, respectively.

Let Pos^k be the k_{th} subcarrier FC in the input broadband spectrum, and we can formulate the corresponding FC shift ground-truth and the whole FC shift loss as follows:

$$\hat{S}^k = \frac{Pos^k}{4} - \left\lfloor \frac{Pos^k}{4} \right\rfloor \tag{12}$$

$$L_{shift} = -\frac{1}{N} \sum_{k=1}^N |\hat{S}^k - S^k| \tag{13}$$

where \hat{S}^k and S^k denote the FC shift ground-truth and prediction. $\lfloor \cdot \rfloor$ represents rounding down. L_{shift} denotes the FC shift loss. Like the BW loss, we applied the L1 loss and only focused on the subcarriers' center point.

To balance the three losses, we used two constants λ_{bw} and λ_{shift} to scale the BW and FC shift losses, respectively. The overall training loss is as follows:

$$L_{det} = L_{psd} + \lambda_{bw} L_{bw} + \lambda_{shift} L_{shift} \tag{14}$$

where we set $\lambda_{bw} = 0.01$ and $\lambda_{shift} = 0.1$ in all our experiments.

4. Experiments

We describe the dataset and evaluation metrics in detail. We report the experimental results and compared the performance with other methods to demonstrate the effectiveness of FFSCN models. Moreover, some ablation studies are shown to shed light on the effects of various design decisions.

4.1. Dataset Description

Table 2 shows detailed information on the dataset used in this work. We used Matlab to generate all 1000 time domain signals, which are all complex. Each signal sample rate was 3.2 MHz, and the time duration as 200 ms. Because the time domain signal is complex, the broad signal bandwidth equals the sample rate. To demonstrate the effectiveness of FFSCN for multi-carrier modulation signals, we used Matlab to generate multi-carrier modulation signals and single-carrier modulation signals, where multi-carrier modulation consisted of 2FSK and OFDM, and the single-carrier modulation consisted of binary phase-shift keying (BPSK), 16 quadrature amplitude modulation (16-QAM), and Gaussian minimum-shift keying (GMSK). Moreover, for each sub-carrier, the narrow signal bandwidth range was 4~117 kHz, the SNR range was $-4\sim 14$ dB, and the time duration range was 20~200 ms. When the sub-carrier signal time duration is 200 ms, the signal is called a constant signal; otherwise, it is called a burst signal.

Table 2. Specification for the dataset in this work.

Sample Nums	1000
Sample Rate	3.2 MHz
Sample Time Duration	200 ms
Broad Signal Bandwidth	3.2 MHz
Sub-Carrier Signal Modulation	2FSK, OFDM, BPSK, 16QAM, GMSK
Sub-Carrier Signal bandwidth	4~117 kHz
Sub-Carrier Signal Time Duration	20~200MS
Sub-Carrier Signal SNR	$-4\sim 14$ dB
FFT Length	16,384
Window Function	Hanning Window
Consecutive Frame Nums	10
Single Frame Time Domain Signal Length	3200

We used a length of 3200 time domain signals to calculate the single frame broadband power spectrum; the FFT length was to 16,384, and the window function was selected as the Hanning window. We set the nums of consecutive frames of broadband power spectra inputs in the training phase to 10.

4.2. Training Setup

The training setup was mostly the same as SCN [14]. We implemented our models in the PyTorch [40] library on a machine with 2 NVIDIA GeForce RTX 3080Ti graphic process units (GPUs) and an Intel(R) Bronze 3204 CPU, with the Ubuntu 20.04 operation system. We used a cosine annealing warm restarts [41] learning rate strategy, with an initial value of 1×10^4 , $T_0 = 10$, $T_{mult} = 2$, and a batch size of 16. We used the Adam optimization method [42] to optimize the overall training loss and adopted Dropout [43] prior to RegNet to reduce overfitting. All the FFSCN models were trained for 150 epochs, and were applying the same data preprocessing steps described in Section 3.

4.3. Evaluation Metrics

In accordance with [14], we also used the intersection-over-unit (IoU) on carriers to decide the correctness of each sub-carrier on the broadband power spectrum, as shown and defined in Figure 5.

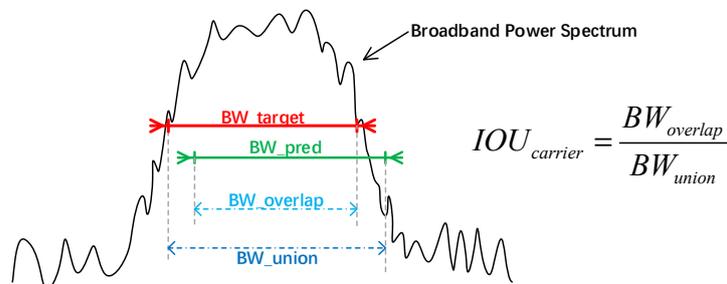


Figure 5. The carrier IoU.

During the evaluation, when the detected sub-carrier IoU is greater than the IoU threshold, it is referred to as true positive (TP), otherwise as true negative (TN). Furthermore, false negative (FN) represents the sub-carrier that is not detected but a ground-truth. We calculate the harmonic means of the average precision rate (AP) and average recall rate (AR) to quantify and compare the performance of different trained models, called the *F-Score* [44], using the following formula:

$$AP = \frac{\sum_{i=0}^N TP}{\sum_{i=0}^N TP + \sum_{i=0}^N TN} \tag{15}$$

$$AR = \frac{\sum_{i=0}^N TP}{\sum_{i=0}^N TP + \sum_{i=0}^N FN} \tag{16}$$

$$F-Score = \frac{2 \times AP \times AR}{AP + AR} \tag{17}$$

4.4. Results

Firstly, to demonstrate the effectiveness of the proposed FFSCN models, we compared the performance with other deep-learning-based methods, including SCN [14], FCN [12], and SigdetNet [13]. As can be seen in Table 3, our proposed FFSCN-FMN models outperformed the other models. The performances of SigdetNet and FCN degraded more than other models. With the IoU threshold increasing, all the model’s detection performances degraded, but the proposed FFSCN models performed more robustly than our previous SCN model overall. Moreover, from the table, we also concluded that the Residual backbone performed better than the MobileNetV3 backbone. However, the fusion MobileNetV3 backbone achieved the best performance, which indicates that the multiple time feature fusion is superior to the one-time fusion strategy.

Then, to further demonstrate the multi-carrier modulation signal detection performance of our proposed FFSCN models, Figure 6 shows the performances of the multi-carrier and single-carrier modulation signal detection on the validation dataset. The proposed models outperformed other methods on both the multi-carrier and single-carrier modulation signal validation datasets. Moreover, the performance of all models degraded as the IoU threshold increased. However, SigdetNet and FCN degraded more severely, especially when the SNR was lower in multi-carrier modulation signal detection performance. Therefore, the proposed FFSCN model predictions were more exact than the others. Moreover, from Table 3 and Figure 6, FFSCN-FMN achieved better performance than FFSCN-R and FFSCN-MN.

Table 3. Performance comparison of the proposed FFSCN and other deep-learning-based methods on the whole validation dataset.

MODEL	AP60	AR60	F-S60	AP70	AR70	F-S70	AP80	AR80	F-S80	AP90	AR90	F-S90
FFSCN-FMN	99.33	94.82	97.02	98.19	93.73	95.91	96.44	92.06	94.20	93.43	89.19	91.26
FFSCN-MN	99.43	90.42	94.71	98.71	89.76	94.02	96.52	87.77	91.94	91.27	82.99	86.93
FFSCN-R	99.04	92.57	95.69	97.70	91.31	94.40	94.95	88.75	91.75	89.72	83.85	86.69
SCN	99.08	90.75	94.73	98.12	89.87	93.81	95.84	87.78	91.63	90.95	83.30	86.96
SigdetNet	83.76	95.11	89.07	81.59	92.64	86.77	77.10	87.55	81.99	65.10	73.92	69.23
FCN	34.32	73.06	46.70	31.49	67.03	42.85	28.05	59.69	38.17	20.98	44.64	28.55

Where F-S denotes the F-Score and the numbers after AP, AR, and F-S denote the IoU threshold. The downsample times of the SCN and FCN were also set to 11.

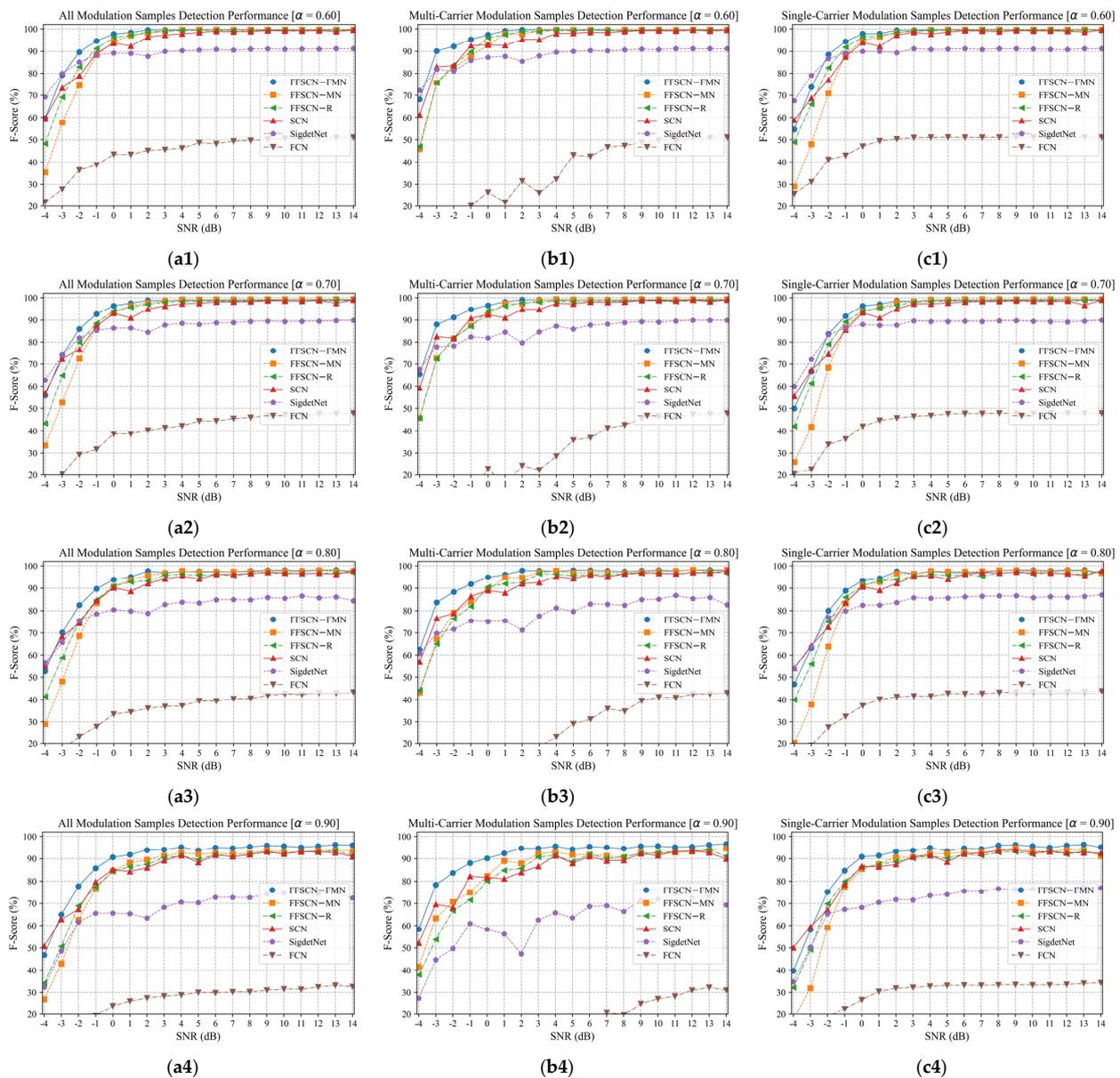


Figure 6. The multi-carrier and single-carrier modulation signal detection performances of our proposed FFSCN models and other deep-learning-based methods: (a1–a4) all modulation samples’ detection performance; (b1–b4) multi-carrier modulation samples’ detection performance; (c1–c4) single-carrier modulation samples’ detection performance. α denotes the IoU threshold and increases from 0.6 to 0.9 from (1) to (4).

Next, Table 4 shows the complexity comparison between the FFSCN models and other deep-learning-based methods. Compared to FFSCN-R and SCN, by adopting the MobileNet backbone, the floating-point operations (FLOPs) and inference time cost of FFSCN-MN decreased obviously. Furthermore, by applying the Fusion FPN neck, the FLOPs and inference time of FFSCN-FMN decreased further than FFSCN-MN. Furthermore, even though the FLOPs of FFSCN-FMN were much larger than FCN and SigdetNet, they consumed a comparable inference time. Therefore, FFSCN-FMN improved the efficiency of SCN.

Table 4. The complexity comparison between the proposed FFSCN method and the other deep-learning-based methods.

MODEL	FLOPS (M)	Parameters (K)	Time Cost (ms)
FFSCN-FMN	2427.08	2803.83	7.82
FFSCN-MN	7814.48	2589.96	10.71
FFSCN-R	15,680.69	2410.66	21.42
SCN	2043.88	2342.56	17.35
SigdetNet	454.98	297.52	8.80
FCN	9.43	110.11	7.16

Finally, from the performance and complexity comparison, our proposed FFSCN models can effectively detect multi-carrier and single-carrier signals in the broadband power spectrum. Compared to other deep-learning-based methods, the FFSCN models achieved better detection performance. Moreover, FFSCN-FMN not only achieved the best detection performance, but also cost a comparable inference time to that of the FCN-based methods, dramatically improving the model complexity of SCN. Meanwhile, SCN, FCN, and SigdetNet use broadband power spectra as the model input and can only detect the frequency locations. As Figures 7 and 8 show, since the multiple consecutive frames of the broadband power spectra were the model input, FFSCN can distinguish the burst signals from the constant signals and locate the frequency and time position. Note that the time location accuracy correlated with the number of consecutive and overlap frames. In Figures 7 and 8, we used 10 consecutive frames inputs without overlap. Compared to the ground-truth, our frequency location predictions were rather good, but some errors existed in the time location prediction, especially when the SNR was low.

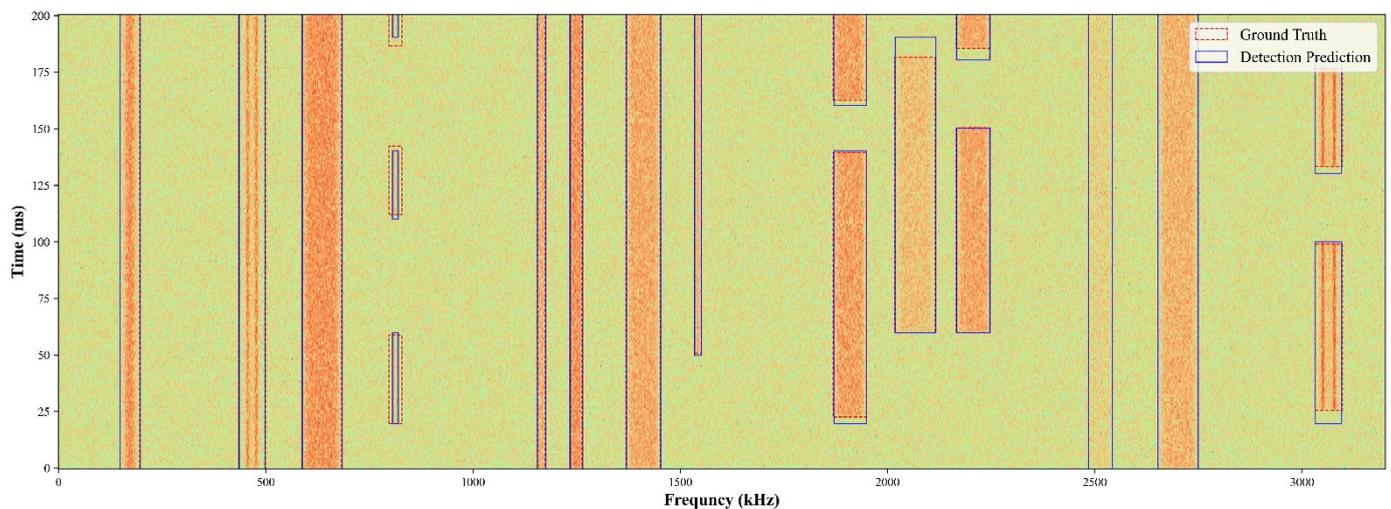


Figure 7. The first example of the FFSCN-FMN detection results. Here, the signal time duration is 200 ms, and we split it into 200 consecutive frames of the broadband power spectra and grouped them into 20 FFSCN-FMN inputs without frame overlap.

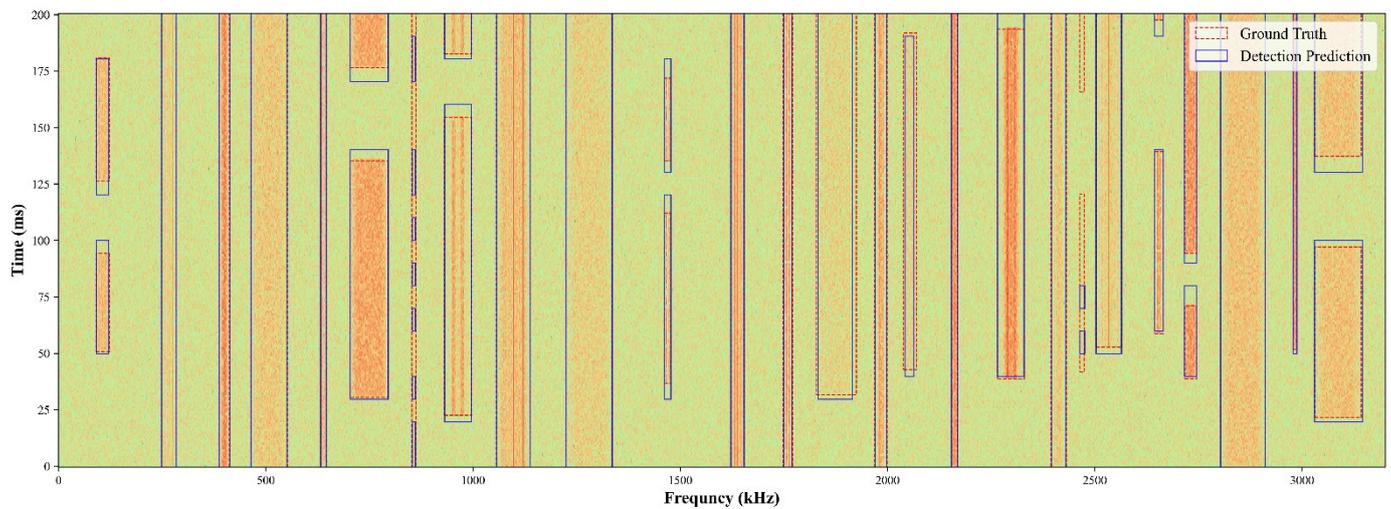


Figure 8. The second example of the FFSCN-FMN detection results. Here, the signal time duration is 200 ms, and we split it into 200 consecutive frames of the broadband power spectra and grouped them into 20 FFSCN-FMN inputs without frame overlap.

4.5. Ablation Study

4.5.1. Impact of the Two Adaptive Pooling Layer Types

The effects of the two different adaptive pooling layer types used in the FFSCN models are depicted in Figure 9. When the sub-carrier SNR was larger than 2 dB, the two types of adaptive pooling layers performed comparably well in the three FFSCN models. However, when the sub-carrier SNR was lower than 2 dB, the models using the adaptive average pooling layer outperformed the models using the adaptive maximum pooling layer.

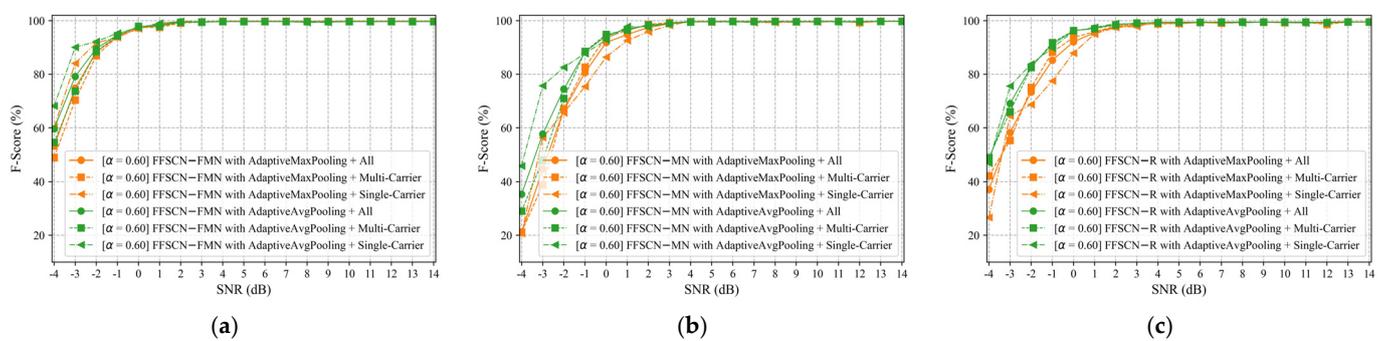


Figure 9. Effects of the two different adaptive pooling layer types used in the FFSCN models: (a) FFSCN-FMN; (b) FFSCN-MN; (c) FFSCN-R. Here, α denotes the IoU threshold, and we fixed it to 0.6.

4.5.2. Impact of Downsample Times

In Table 5, we show the performance comparison of different downsample times used in the FFSCN-FMN model. We can see that FFSCN-FMN₁₁ \times achieved the best AR and F-Score and that FFSCN-FMN₁₃ \times achieved the best AP, but the increase was feeble compared to FFSCN-FMN₁₁ \times . Furthermore, the multi-carrier and single-carrier modulation signal detection performances of different downsample times in the FFSCN-FMN model are shown in Figure 10. We found that the main reason that FFSCN-FMN₁₃ \times had the best AP was that it performed better in the single-carrier modulation sample detection results. Considering the bigger gap in the AR and F-Score between FFSCN-FMN₁₃ \times and FFSCN-FMN₁₁ \times and the more downsample times used, complexity, and inference time cost, we think that using 11 downsample times is the better choice.

Table 5. Performance comparison of different downsample times used in the FFSCN-FMN model.

MODEL	AP60	AR60	F-S60	AP70	AR70	F-S70	AP80	AR80	F-S80	AP90	AR90	F-S90
FFSCN-FMN_9×	99.46	91.31	95.21	98.54	90.47	94.33	96.57	88.66	92.45	93.08	85.46	89.11
FFSCN-FMN_11×	99.33	94.82	97.02	98.19	93.73	95.91	96.44	92.06	94.20	93.43	89.19	91.26
FFSCN-FMN_13×	99.48	94.65	97.01	98.25	93.54	95.84	96.49	91.74	94.05	93.44	88.93	91.13

F-S denotes the F-Score, and the numbers after AP, AR, and F-S denote the IoU threshold. We compared the performance of the FFSCN-FMN model using 9, 11, and 13 downsample times.

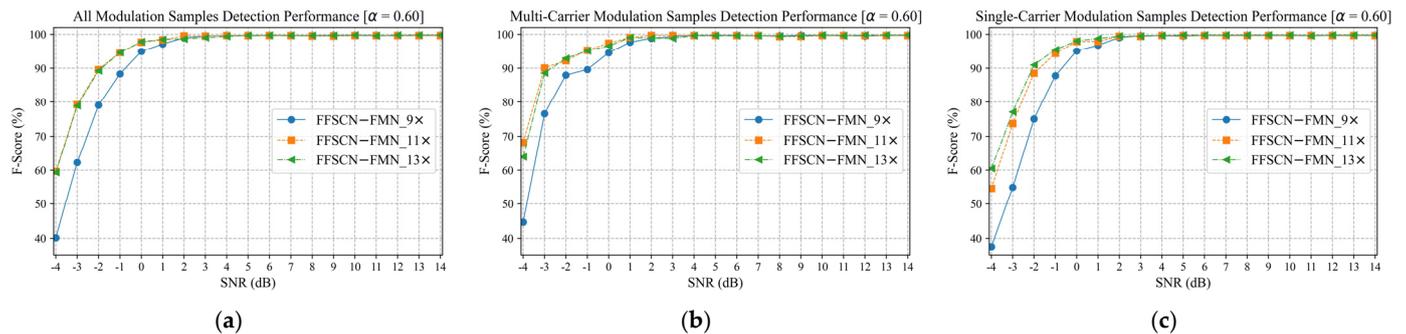


Figure 10. The multi-carrier and single-carrier modulation signal detection performances of different downsample times used in the FFSCN-FMN model: (a) all modulation samples' detection performance; (b) multi-carrier modulation samples' detection performance; (c) single-carrier modulation samples' detection performance. Here, α denotes the IoU threshold, and we fixed it to 0.6. We compared the performance of the FFSCN-FMN model using 9, 11, and 13 downsample times.

5. Conclusions

This paper introduced the FFSCN-FMN, FFSCN-MN, and FFSCN-R models for carrier signal detection. As an upgrade to SCN, by using multiple frames of the broadband power spectra as the model input rather than one, the model can extract the features of the broadband power spectra of frequencies as they vary with time, so that it can effectively detect the multi-carrier and single-carrier modulation signals. FFSCN-R adds an adaptive average pooling layer between the FPN neck and RegNet head in SCN. FFSCN-MN replaces the FFSCN-R backbone network with the MobileNetV3 backbone to reduce the complexity of the model. FFSCN-FMN further modifies the MobileNetV3 backbone and FPN neck to the Fusion backbone and Fusion FPN neck to design a more lightweight model. Extensive experimental results suggested that the proposed FFSCN models outperformed the other deep-learning-based methods SCN in accuracy and efficiency, and the FFSCN-FMN model performed the best. As it remains a problem to detect the burst signals in a timely manner, we will engage in solving this in future work.

Author Contributions: Conceptualization, H.H. and J.L.; methodology, H.H.; software, H.H. and J.W.; validation, J.W.; writing—original draft preparation, H.H.; writing—review and editing, H.H. and J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The dataset can be accessed at onedrive (<https://1drv.ms/u/s!AvmAsnk8HBhuiwMRK8kQWKzktiIw?e=YRZFay>) or gitee (https://gitee.com/yulongpo/ffscn_dataset).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Ian, F.; Brandon, F.; Ravikumar, B. Cooperative spectrum sensing in cognitive radio networks: A survey. *Phys. Commun.* **2011**, *4*, 40–62.
- Ian, F.; Won-Yeol, L.; Kaushik, R. CRAHNS: Cognitive radio ad hoc networks. *Ad Hoc Netw.* **2009**, *7*, 810–836.

3. Digham, F.F.; Alouini, M.-S.; Simon, M.K. On the energy detection of unknown signals over fading channels. *IEEE Trans. Commun.* **2007**, *55*, 21–24.
4. Urkowitz, H. Energy detection of unknown deterministic signals. *Proc. IEEE* **1967**, *55*, 523–531. [[CrossRef](#)]
5. Zhi, Q.; Cui, S.; Sayed, A.H.; Poor, H.V. Optimal multiband joint detection for spectrum sensing in cognitive radio networks. *IEEE Trans. Sig. Process.* **2009**, *57*, 1128–1140.
6. Gardner, W.A. Signal interception: A unifying theoretical framework for feature detection. *IEEE Trans. Commun.* **1988**, *36*, 897–906. [[CrossRef](#)]
7. Fehske, A.; Gaeddert, J.; Reed, J.H. A new approach to signal classification using spectral correlation and neural networks. In Proceedings of the First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks DySPAN 2005, Baltimore, MD, USA, 8–11 November 2005; pp. 144–150.
8. Lunden, J.; Koivunen, V.; Huttunen, A.; Poor, H.V. Collaborative cyclostationary spectrum sensing for cognitive radio systems. *IEEE Trans. Commun.* **2009**, *57*, 4182–4195. [[CrossRef](#)]
9. Vartiainen, J.; Lehtomaki, J.J.; Saarnisaari, H. Double-threshold based narrowband signal extraction. In Proceedings of the 2005 IEEE 61st Vehicular Technology Conference VTC, Stockholm, Sweden, 30 May–1 June 2005; Volume 2, pp. 1288–1292.
10. Vartiainen, J. Localization of multiple narrowband signals based on the FCME algorithm. In Proceedings of the Nordic Radio Symposium NRS, Oulu, Finland, 16–18 August 2004; Volume 1, p. 5.
11. Kim, J.; Kim, M.; Won, I.; Yang, S.; Lee, K.; Huh, W. A biomedical signal segmentation algorithm for event detection based on slope tracing. In Proceedings of the 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, MN, USA, 3–6 September 2009; pp. 1889–1892.
12. Huang, H.; Li, J.Q.; Wang, J.; Wang, H. FCN-Based Carrier Signal Detection in Broadband Power Spectrum. *IEEE Access* **2020**, *8*, 113042–113051. [[CrossRef](#)]
13. Lin, M.; Zhang, X.; Tian, Y.; Huang, Y. Multi-Signal Detection Framework: A Deep Learning Based Carrier Frequency and Bandwidth Estimation. *Sensors* **2022**, *22*, 3909. [[CrossRef](#)]
14. Huang, H.; Wang, P.; Wang, J.; Li, J. Deep Learning-Based End-to-End Carrier Signal Detection in Broadband Power Spectrum. *Electronics* **2022**, *11*, 1896. [[CrossRef](#)]
15. Sejdčić, E.; Djurović, I.; Jiang, J. Time–frequency feature representation using energy concentration: An overview of recent advances. *Digit. Signal Process.* **2009**, *19*, 153–183. [[CrossRef](#)]
16. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
17. Chen, M.; Challita, U.; Saad, W.; Yin, C.; Debbah, M. Artificial Neural Networks-Based Machine Learning for Wireless Networks: A Tutorial. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 3039–3071. [[CrossRef](#)]
18. Luong, N.C.; Hoang, D.T.; Gong, S.; Niyato, D.; Wang, P.; Liang, Y.-C.; Kim, D.I. Applications of Deep Reinforcement Learning in Communications and Networking: A Survey. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 3133–3174. [[CrossRef](#)]
19. O’Shea, T.; Hoydis, J. An Introduction to Deep Learning for the Physical Layer. *IEEE Trans. Cogn. Commun. Netw.* **2017**, *3*, 563–575. [[CrossRef](#)]
20. Morozov, O.A.; Ovchinnikov, P.E. Neural Network Detection of MSK Signals. In Proceedings of the 2009 IEEE 13th Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop, Marco Island, FL, USA, 4–7 January 2009; pp. 594–596.
21. Li, Y.; Wang, B.; Shao, G.; Shao, S.; Pei, X. Blind Detection of Underwater Acoustic Communication Signals Based on Deep Learning. *IEEE Access* **2020**, *8*, 204114–204131. [[CrossRef](#)]
22. Yuan, Y.; Sun, Z.; Wei, Z.; Jia, K. DeepMorse: A Deep Convolutional Learning Method for Blind Morse Signal Detection in Wideband Wireless Spectrum. *IEEE Access* **2019**, *7*, 80577–80587. [[CrossRef](#)]
23. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention*; Springer: Munich, Germany, 2015; Volume 9351, pp. 234–241.
24. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [[CrossRef](#)]
25. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
26. Forrest, N.I.; Song, H.; Matthew, W.M.; Khalid, A.; William, J.D.; Kurt, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. *arXiv* **2014**, arXiv:1602.07360.
27. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6848–6856.
28. Huang, G.; Liu, S.; Laurens van der, M.; Kilian, Q.W. CondenseNet: An Efficient DenseNet Using Learned Group Convolutions. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, 18–22 June 2018; pp. 2752–2761.
29. Wu, B.; Wan, A.; Yue, X.; Jin, P.; Zhao, S.; Noah, G.; Amir, G.; Joseph, G.; Kurt, K. Shift: A Zero FLOP, Zero Parameter Alternative to Spatial Convolutions. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, 18–22 June 2018; pp. 9127–9135.

30. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.
31. Mark, S.; Andrew, H.; Zhu, M.; Andrew, Z.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520.
32. Andrew, H.; Mark, S.; Chu, G.; Chen, L.C.; Chen, B.; Tan, M.X.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for MobileNetV3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea, 27 October–2 November 2019; pp. 1314–1324.
33. Proakis, J.G.; Manolakis, D.G. *Digital Signal Processing: Principles, Algorithms and Applications*, 3rd ed.; Prentice-Hall: Hoboken, NJ, USA, 1996; pp. 910–913.
34. Welch, P. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* **1967**, *15*, 70–73. [[CrossRef](#)]
35. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 26 June–1 July 2016; pp. 770–778.
36. Woo, S.H.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the 15th European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
37. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
38. Avenash, R.; Viswanath, P. Semantic Segmentation of Satellite Images using a Modified CNN with Hard-Swish Activation Function. In Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019), Prague, Czech Republic, 25–27 February 2019; pp. 413–420.
39. Maas, A.L.; Hannun, A.Y.; Ng, A.Y. Rectifier nonlinearities improve neural network acoustic models. In Proceedings of the International Conference Machine Learning (ICML), Atlanta, GA, USA, 16–21 June 2013; p. 3.
40. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In Proceedings of the 33rd Conference Neural Information Processing Systems (NeurIPS), Vancouver, BC, Canada, 8–14 December 2019.
41. Loshchilov, I.; Hutter, F. SGDR: Stochastic Gradient Descent with Warm Restarts. *arXiv* **2016**, arXiv:1608.03983.
42. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
43. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
44. Goutte, C.; Gaussier, E. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In *Advances in Information Retrieval*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 345–359.