A New Deep Learning Framework for HF Signal Detection in Wideband Spectrogram

Weihao Li, Keren Wang, Ling You, and Zhitao Huang

***Abstract*—Detection of high frequency (HF) signal in the wideband is challenging since the HF environment is chaotic. Recent works adopt deep learning-based object detectors to capture signals in wideband spectrogram, but the task of signal detection exhibits different characteristics from that of generic object detection, which causes the classical deep learning-based detectors to have defects such as limited receptive field and prior anchor mismatch. Based on the task analysis, this letter proposes a deep learning framework which extracts features along the time axis at each frequency bin, and predicts multiple characteristics of the signal, including the center frequency and the shape attributes. The unique advantages of the framework are the utilization of all time features and no prior anchor which adapt to the slender shape of signal with a strong generalization ability. The numerical studies on simulation signal + real-world background prove the superiority of the proposed framework both in accuracy and speed.**

***Index Terms*—HF signal detection, wideband spectrogram, deep learning, object detection, features along time.**

I. Introduction

HIGH frequency (HF) communication plays an important role in military, aviation, navigation, etc. HF bandwidth is approximate 27 MHz, which is quite wide compared to HF signal bandwidth such as 3 kHz or less. Nowadays HF band contains a large number of signals, coupled with the multipath delay and fading, the HF environment is getting chaotic. The purpose of this letter is to detect multiple interested HF signals in the wideband, with the output of time-frequency location and signal classes.

Traditional signal detection methods are performed in the narrowband and detect the presence of signals, such as optimal maximum-likelihood [1], suboptimal linear zero-forcing (ZF) [2], minimum mean square error (MMSE) [3], and deep learning-based algorithms [4], [5]. With the fast development of hardware computational ability, signal detection covering the entire band becomes possible. Channelization + narrowband detection which splits the entire band into narrow channels with a fixed frequency step and then implements narrow band detection is a solution to entire-band detection [6]. However, channelization for tens of thousands of channels is quite computationally expensive, and channelization parameters, such as the bandwidth and the frequency step, are quite difficult to choose for signals with various bandwidth and frequency gaps. Therefore, this letter decides to split the entire band into wideband and detect multiple interested signals in the wideband. Wideband detection can detect signals with various bandwidths and save calculated amount compared to narrowband detection. A third-party finer identification can be further performed on the results of wideband detection.

To the best of our knowledge, most wideband detection methods are based on the spectrogram calculated by the short-time Fourier transform (STFT), which can clearly present the time-frequency distribution of signals. Due to the spectrogram being a special type of image, the task of signal detection in the spectrogram is quite similar to the object detection in the image. Most existing works indeed go along the idea of object detection: a combination of energy detection and convolutional neural network (CNN) classifier is proposed in [7], and some classical deep learning-based object detectors such as Faster-RCNN, SSD, YOLOv3 are adopted in [8]–[10], which get the state-of-the-art performance. For the classical detectors, the framework modifications from generic object detection to signal detection mainly lie in adjusting the size of the prior anchors to fit the signals by statistical or clustering methods. However, the aspect ratios of signals in the spectrogram vary dramatically, even the well-designed prior anchors have trouble matching all signals. Moreover, most classical detectors locate an object by an internal point, whose receptive field in CNN cannot cover the signals with very long time length. Above defects make the performance reduction of classical detectors. So this letter aims to design a more reasonable framework specifically for spectrogram–based signal detection task.

The main contributions of this study are as follows: 1) We make a detailed analysis of the differences between signal detection and generic object detection, and the resulting defects of the classical deep learning-based object detectors. 2) We propose an anchor-free framework that utilizes all of the time point features at each frequency, rather than features at only one point, to predict the center frequency and the shape attributes of each signal. 3) The numeric experiments on simulation signal + real-world background demonstrates the distinct improvements of the proposed method both in accuracy and speed.

Manuscript received March 15, 2022; revised May 24, 2022; accepted May 26, 2022. Date of publication June 3, 2022; date of current version June 17, 2022. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Beibei Wang. *(Corresponding author: Weihao Li.)*

Weihao Li is with the College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China, and also with the National Key Laboratory of Science and Technology on Blind Signal Processing, Chengdu 610041, China (e-mail: liweihao21@nudt.edu.cn).

Keren Wang and Ling You are with the National Key Laboratory of Science and Technology on Blind Signal Processing, Chengdu 610041, China (e-mail: cfan662003@163.com; youlingu1971@sina.com).

Zhitao Huang is with the College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China (e-mail: huangzhitao@nudt.edu.cn).

Digital Object Identifier 10.1109/LSP.2022.3179958

1070-9908 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: University Town Library of Shenzhen. Downloaded on October 10,2025 at 12:30:21 UTC from IEEE Xplore. Restrictions apply.

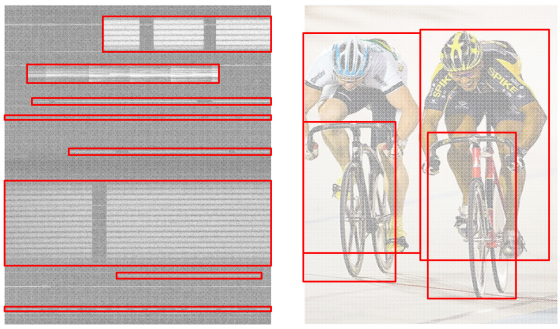

Fig. 1. Output of signal detection and generic object detection.

II. Problem Analysis

A. HF Signal Detection in Wideband Spectrogram

The bandwidth of HF signals varies and the gaps between signals are uneven. So, one of the advantages of wideband detection is that it avoids the parameter choices of channelization, and captures simultaneously multiple signals instead, with the predictions of time-frequency location. For the recognition of signal classes, some classes can be recognized directly in the wideband and some cannot, which could be filtered to the narrowband using the predicted bandwidth to be further identified by a third-party classifier. Another advantage of wideband detection is reducing the calculated amount, for example, splitting the 27 MHz HF band into 4 kHz narrowband or 125 kHz wideband both with 3 kHz overlap, the total processed frequency of narrowband is $[(27000-4)/1+1]\times 4 = 107988$kHz, and that of wideband is $[(27000-125)/122+1]\times 125 = 27661$kHz, which is a lot less.

The wideband detection is a prior screening of narrowband fine recognition and thus the richness of output information can be configured according to the engineering requirements. In this letter, we define the output of wideband detection including:

1) *Signal class:* For some classes that cannot be recognized in the spectrogram of the current resolution, they are lumped into one class to wait for further recognition in the narrowband.
2) *Start and end time:* Many HF signals transmits as sequential burst waves, and all burst waves can be detected as one signal in wideband detection, i.e., the start of the first wave and the end of the last wave are respectively the start and end of that signal.
3) *Signal bandwidth.*

B. Signal Detection vs. Generic Object Detection

We inherit the idea of object detection for signal detection in this letter. The task of signal detection is a specific application field of object detection, and thus the objects and background present unique characteristics different from generic objects and natural background, which can serve as prior information to design a more targeted and effective detector. Fig. 1 is a sketch map comparing the output of signal detection and generic object detection, and we conclude the characteristics of HF signal detection in wideband spectrogram as listed below:

1) *Objects are all signals but of different classes*: The outlines of signals appear to be slender rectangles with small aspect ratios and horizontal angles.
2) *Bounding boxes are arranged vertically*: There is only one signal transmitted at a frequency (multiple burst waves are treated as one signal). It is rare that two different classes of interested signals occupy the same center frequency during the time length of a spectrogram processed.
3) *Background scene is noise and interference signals*: The background of signal detection is simpler than that of generic object detection, but a lot of interference signals increase the probability of false alarm.

The classical deep learning-based detectors used in generic object detection have not considered the above characteristics and thus cause a reduction in performance. We sum up two typical defects of classical detectors for signal detection:

1) *Limited receptive field*: Each point in the feature map has a receptive field after multiple CNNs. The classical detectors usually use one or a few points to make predictions whose receptive field is hard to cover the signal with a long transverse length, which results in an incomplete bounding box.
2) *Prior anchor mismatch*: Many classical detectors design prior anchors in advance, but they are hard to match signals during training because the aspect ratios of signals are small and vary dramatically, which causes effectiveness degradation. In addition, the design of prior anchors is a complex process.

In addition, the large model size of classical detectors is unnecessary here since signal detection is a target-specific field compared to the generic object detection.

III. Proposed Deep Learning Framework

Considering the HF signals have slender shapes and are arranged vertically in the spectrogram, we propose to locate the signal by its center frequency, i.e., shrinking the feature map to only frequency and channel dimensions, and predicting center frequency and box attributes using features at each frequency. Each frequency feature merges the whole time features, whose receptive field could cover the entire signal, and the center frequency-based regression does not need the prior anchors. The structure of the proposed framework is shown in Fig. 2.

A. Feature Extraction

A feature pyramid network (FPN) is utilized to extract the feature map from the input spectrogram. Here we chose the ResNet22 (ResNet18 plus one more convolution stage of four CNNs) combined with four transposed CNNs as the FPN structure. The purpose of ResNet18 adds one more convolution stage to form the ResNet22 is to increase the receptive field of CNN to cover the signals with very wide bandwidth–in the next section, the proposed method will be tested on signals with bandwidth up to 48 kHz which needs the addition of one convolution stage through our tuning.

Each transposed CNN doubles the size of the feature map that is then added with the feature map of the same size in ResNet22. From the above operation, the FPN could effectively blend the subtle features in the large feature maps and the high-dimensional features in the small feature maps. After the feature extraction, a feature map with 1/4 of the input size and 16 channels is obtained.

Authorized licensed use limited to: University Town Library of Shenzhen. Downloaded on October 10,2025 at 12:30:21 UTC from IEEE Xplore. Restrictions apply.

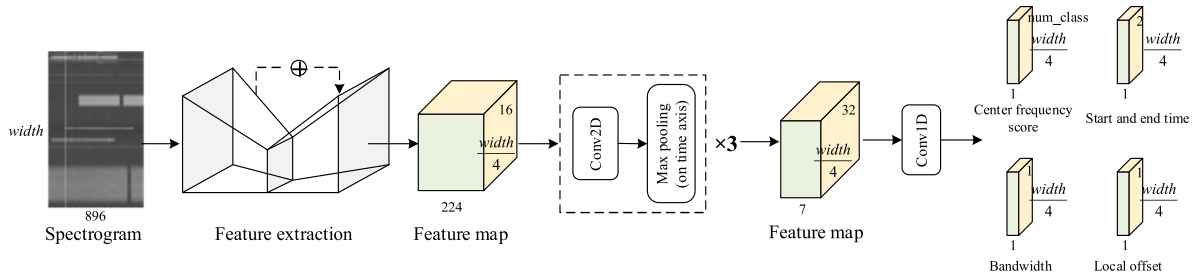

Fig. 2. The proposed deep learning framework.

B. Regression on the Feature Map

Absorbing the characteristics of signal detection, we propose to utilize features along the time axis at each frequency bin to predict the center frequency and the shape attributes.

The specific operation is, as depicted in Fig. 2, to feed the feature map to three consecutive CNN and max pooling, and it should be emphasized that the max pooling is only implemented on the time axis, i.e., shrinking the time length of feature map. The output of CNN and max pooling sequences continues to be fed into four parallel 1D CNNs to further shrink the time axis to 1 and convert the channels to correspond to the attributes. Four attribute maps are obtained after the 1D CNN layer which refer to four attributes at each frequency. The 'center frequency score' refers to the probability of being a center frequency (y_{cf}) of each signal class; the 'start and end time' refers to the start abscissa (x_s) and end abscissa (x_e) of detected signal; the 'bandwidth' (BW) refers to the vertical width of detected signal in the spectrogram; the 'local offset' (LO) is to recover the vertical discretization error during the pooling of feature extraction. When obtaining above four attributes, the lower left and upper right points of a bounding box can be calculated as in (1), where R is the shrinking size during feature extraction:

$$R \cdot (x_s, y_{cf} + LO - BW/2, x_e, y_{cf} + LO + BW/2). \quad (1)$$

For training of 'center frequency score' heatmap, we firstly splat all ground truth keypoints $\widetilde{p}$ on a heatmap with a Gaussian kernel $Y_{yc} = \exp(-(y - y_{\widetilde{p}})^2/(2\sigma^2))$, where σ is an object size-adaptive standard deviation [11], and then a penalty-reduced pix-wise focal loss [12] is calculated as in (2), where hyper-parameters α and β are set to 2 and 4 in our experiments, and N is the number of keypoints. The remaining three attribute maps 'start and end time', 'bandwidth', and 'local offset' directly predict the value and thus we extend the L1 loss by adding a normalization term (except for 'local offset') for each keypoint, as in (3) to (5).

$$L_{cf} = -\frac{1}{N}\sum_{yc} \begin{cases} (1-\hat{Y}_{yc})^{\alpha}\log(\hat{Y}_{yc}) & \text{if } Y_{yc}=1 \\ (1-Y_{yc})^{\beta}(\hat{Y}_{yc})^{\alpha}\log(1-\hat{Y}_{yc}) & \text{otherwise} \end{cases}, \quad (2)$$

$$L_{se} = \frac{1}{N}\sum_{k=1}^{N}\frac{|\hat{X}_{sk} - X_{sk}| + |\hat{X}_{sk} - X_{ek}|}{2 \cdot |X_{ek} - X_{sk}|}, \quad (3)$$

$$L_{bw} = \frac{1}{N}\sum_{k=1}^{N}\frac{|\hat{BW}_k - BW_k|}{BW_k}, \quad (4)$$

$$L_{lo} = \frac{1}{N}\sum_{k=1}^{N}|\hat{LO}_k - LO_k|, \quad (5)$$

The total loss is formulated as the weighted sum of four attribute losses, and we set $\lambda_1 = 5, \lambda_2 = 1$ and $\lambda_3 = 1$ according to the principle of approximate loss values:

$$L = L_{cf} + \lambda_1 L_{se} + \lambda_2 L_{bw} + \lambda_3 L_{lo}. \quad (6)$$

The proposed method makes predictions at each frequency, instead of at each pixel, of the feature map using features along time, which fits well the shape characteristics of signals and the task requirements of wideband detection. Because the shape attributes are directly regressed at each keypoint, the prior anchors are abandoned totally. Without the tedious anchor design process, the framework becomes elegant and avoids the possible effect reduction brought by the mismatch.

IV. Numerical Studies

A. Dataset Generation

To simulate the complexity of the HF environment as much as possible, we adopt the dataset generation strategy of simulation signal + real-world background. Specifically, real HF wideband signals are collected in four cities at different times and then added with simulated six classes of signals in the time domain. To avoid error training, the background does not contain the interested signal classes. The spectrograms are calculated by STFT, and the details for signal and spectrogram are displayed in Table I. 5000 samples are utilized for training using background from three cities, while 1000 samples with background from the other city are employed for testing.

B. Parameter Choice and Implementation Details

The hyper-parameters of three CNN + max pooling sequences and four parallel 1D CNNs are shown in Table I. For training implementation, data augmentation including random scaling, random crop, and Gaussian noise is adopted. The network is optimized by an Adam optimizer with an exponentially decaying 2×10^{-4} learning rate, and it is trained with 32 batch size and 200 epochs to convergence. Some measures are utilized to inhibit overfitting such as dropout and L2-regularization. All of the experiments are performed on a Tesla P40 GPU and the Tensorflow deep learning tool.

The baselines for comparison are Energy + CNN [7], Faster-RCNN [8], SSD [9], and YOLOv3 [10] which have all been used for signal detection in the spectrogram in previous works. Energy + CNN is a typical traditional method that captures

Authorized licensed use limited to: University Town Library of Shenzhen. Downloaded on October 10,2025 at 12:30:21 UTC from IEEE Xplore. Restrictions apply.

TABLE I
DATASET AND ALGORITHM PARAMETER SETTINGS

Simulation HF Signal	
signal classes	2FSK, 4FSK, PSK [a], GMSK, Morse, Speech
burst time length	300ms~5s
bandwidth	single frequency~48kHz [b]
CNR	0dB~20dB
HF channel	ITU-R HFMQ [13]
Spectrogram Settings	
size	831×4096 [c]
time length, time resolution	5s, 6.02ms
bandwidth, frequency resolution	125kHz, 30.52Hz
Number of contained simulation signals	2~10
Neural Network Hyper-parameters	
conv2d_1, _2, _3	(5,5,32) [d], (5,5,32), (3,3,32)
max pooling_1, _2, _3	(1,4), (1,4), (1,2)
conv1d_1, _2, _3, _4	(1,224,num_class), (1,224,2), (1,224,1), (1,224,1)

[a]The modulation order of PSK is hard to recognize in the spectrogram and thus is ignored.
[b]Different classes of signals have different frequency ranges.
[c]Length 831 is padded to 896 before input to network.
[d](55,32) means 5 × 5 kernel size and 32 channels.

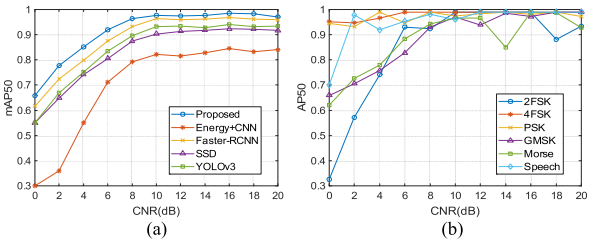

Fig. 3. Accuracy performance evaluation. (a) Comparison of different methods. (b) AP50 performance of the proposed method on different signals.

candidate regions in the spectrogram by energy detection then classifies signals by multiple CNNs. Faster-RCNN, SSD, and YOLOv3 are all classical detectors. We keep the structures of baselines the same as in their literature except for changing the anchor size of Faster-RCNN, SSD, and YOLOv3 by K-Means clustering on the size of ground truth boxes just like in [10].

C. Performance Evaluation

For accuracy performance evaluation, we consider the mean average precision with a 0.5 threshold (mAP50) metric [14]. In Fig. 3(a), the detection mAP50 of the proposed method and baselines under different carrier-to-noise ratio (CNR) is plotted. It is observed that the traditional energy-based method has relatively poor performance, probably because the chaotic HF environment reduces the universality of energy detection. The deep learning methods that have adaptive learning ability generally obtain better performance. The proposed method processes task-specific improvements which performs best in comparison, especially about 0.02 mAP50 higher than the two-stage detector Faster-RCNN. Fig. 3(b) shows the detection AP50 of the proposed method on different signals. The results indicate that the detection performance on different classes of signals is relatively balanced which are all detected well. In detail, signals with wide bandwidth or unique shape (4FSK, PSK, Speech) are detected better than that with narrow bandwidth or regular shape (2FSK, GMSK, Morse), considering the stability of AP50 curves under different CNR.

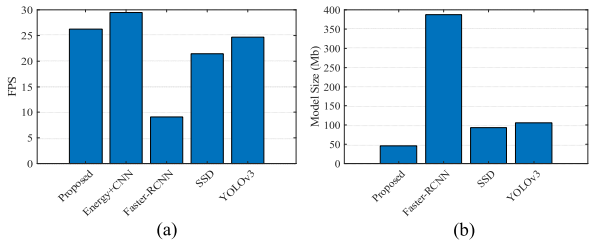

Fig. 4. Computational complexity evaluation. (a) FPS comparison of different methods. (b) Model size comparison of different methods.

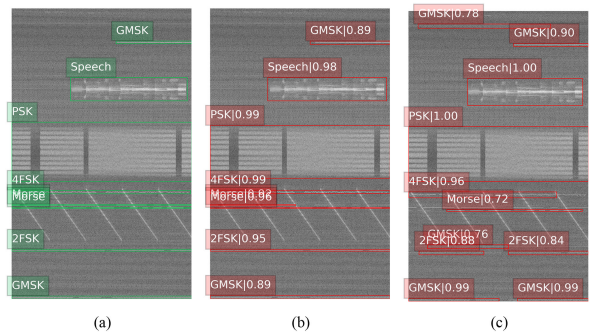

Fig. 5. Detection results. (a) Ground truth. (b) Proposed. (c) Faster-RCNN.

Fig. 4 demonstrates the computational complexity of different methods, where FPS denotes the number of spectrogram samples processed per second. Fig. 4(a) shows that the proposed method is faster than the three classical detectors, especially more than 2 FPS higher than the one-stage detectors SSD and YOLOv3, but it is slightly slower than the Energy + CNN. Fig. 4(b) shows the storage footprint of the parameters of four deep learning models. The results indicate that the proposed method has an obviously smaller model size than other methods echoing the results of FPS, so it could be a preferable solution for engineering applications.

Some detection results are shown in Fig. 5, where the Faster-RCNN that performs best of baselines is compared. We can see the proposed method can detect most of the interested signals completely and not be interfered with background signals. Although the Faster-RCNN has also detected most signals, it could not capture the entire time length of signals with a very small aspect ratio, and it still has some false alarms and missing alarms. The results typically reveal the defects of classical detectors and the effectiveness of improvements in this letter.

V. CONCLUSION

Based on an elaboration on the characteristics of HF signal detection in wideband spectrogram, a task-specific detector is proposed in this letter that integrates all features along the time axis to predict boxes at each frequency. Numerical studies prove the improved performance and the portability of the proposed method, and we think it can be adjusted easily to apply to the spectrogram-based detection of other types of signals.

Authorized licensed use limited to: University Town Library of Shenzhen. Downloaded on October 10,2025 at 12:30:21 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] N. T. Nguyen and K. Lee, "Groupwise neighbor examination for tabu search detection in large MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 1136–1140, Jan. 2020.

[2] M. A. Albreem, M. Juntti, and S. Shahabuddin, "Massive MIMO detection techniques: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3109–3132, Oct.–Dec. 2019.

[3] S. Peng, A. Liu, X. Liu, K. Wang, and X. Liang, "MMSE turbo equalization and detection for multicarrier faster-than-Nyquist signaling," *IEEE Trans. Veh. Technol.,* vol. 67, no. 3, pp. 2267–2275, Mar. 2018.

[4] N. Samuel, T. Diskin, and A. Wiesel, "Learning to detect," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2554–2564, May 2019.

[5] M. Mohammadkarimi, M. Mehrabi, M. Ardakani, and Y. Jing, "Deep learning-based sphere decoding," *IEEE Trans. Wireless Commun.*, vol. 18, no. 9, pp. 4368–4378, Sep. 2019.

[6] Y. Arjoune and N. Kaabouch, "A comprehensive survey on spectrum sensing in cognitive radio networks: Recent advances, new challenges, and future research directions," *Sensors*, vol. 19, no. 1, Jan. 2019, Art. no. 126.

[7] Y. Yuan, Z. Sun, Z. Wei, and K. Jia, "DeepMorse: A deep convolutional learning method for blind morse signal detection in wideband wireless spectrum," *IEEE Access*, vol. 7, pp. 80577–80587, 2019.

[8] K. N. R. S. V. Prasad, K. B. D'souza, and V. K. Bhargava, "A downscaled faster-RCNN framework for signal detection and time-frequency localization in wideband RF systems," *IEEE Trans. Wireless Commun.,* vol. 19, no. 7, pp. 4847–4862, Jul. 2020.

[9] X. Zha, H. Peng, X. Qin, G. Li, and S. Yang, "A deep learning framework for signal detection and modulation classification," *Sensors*, vol. 19, no. 18, 2019, Art. no. 4042.

[10] R. Li, J. Hu, S. Li, S. Chen, and P. He, "Blind detection of communication signals based on improved YOLO3," in *Proc. 6th Int. Conf. Intell. Comput. Signal Process.*, Xi'an, China, 2021, pp. 424–429.

[11] H. Law and J. Deng, "Cornernet: Detecting objects as paired keypoints," in *Proc. Eur. Conf. Comp. Vis.*, Munich, Germany, 2018, pp. 734–750.

[12] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. Int. Conf. Comput. Vis.*, 2017, pp. 2999–3007.

[13] *Testing of HF Modems with Bandwidths of up to About 12 kHz Using Ionospheric Channel Simulators*, ITU Standard Rec. ITU-R F.1487, 2000.

[14] M. Everingham, S. Eslami, L. Van Goo, C. K. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes challenge: A retrospective," *Int. J. Comput. Vis.,* vol. 111, no. 1, pp. 98–136, Jun. 2015.

Authorized licensed use limited to: University Town Library of Shenzhen. Downloaded on October 10,2025 at 12:30:21 UTC from IEEE Xplore. Restrictions apply.