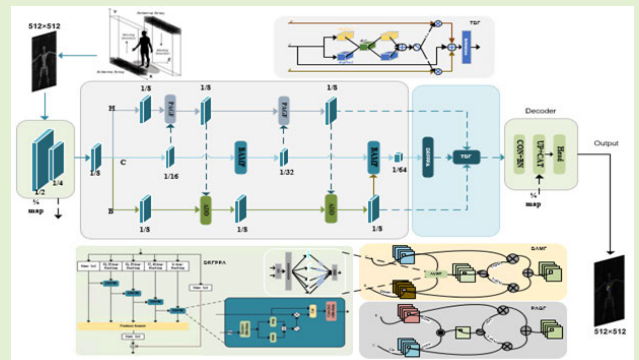


A Boundary-Enhanced Multibranch Network for Improved Object Detection in Low-Resolution Millimeter-Wave Images

Chuizheng Zeng^{ID}, Junhua Ding^{ID}, Minghui Yuan^{ID}, and Yinwei Li^{ID}, *Member, IEEE*

Abstract—With the rapid development of millimeter-wave technology in security inspection, its application in detecting contraband has demonstrated significant advantages. However, the low resolution, low pixel density, and unclear boundaries of terahertz images generated by millimeter-wave imaging significantly limit the accurate detection of contraband. This article proposes a boundary-enhanced multi-branch network (BEMBNet) model based on a three-branch architecture. By designing a dedicated boundary branch and dynamic fusion modules, the model optimizes the extraction of target boundary features and the representation of global semantic information. The model efficiently integrates multiscale features using dynamic reinforced feature pyramid pooling module (DRFPPM) and boundary-aware multiscale fusion (BaMF). Experiments demonstrate that BEMBNet outperforms existing advanced models across multiple metrics scenarios, exhibiting exceptional segmentation accuracy and an innovative solution for millimeter-wave image processing and technologies.

Index Terms—Boundary-enhanced multibranch network (BEMBNet), contraband detection, deep learning, millimeter wave.



I. INTRODUCTION

WITH the continuous advancement of millimeter-wave technology, millimeter-wave imaging has achieved significant advancements in contraband detection applications. Compared to traditional metal detectors and X-ray systems, millimeter-wave imaging offers distinct advantages [1], [2]. Metal detectors, while effective for identifying metallic items, cannot detect plastic or other nonmetallic contraband, limiting

Received 8 March 2025; revised 16 April 2025; accepted 23 April 2025. Date of publication 2 May 2025; date of current version 2 July 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 61988102, in part by the Artificial Intelligence Promotes Scientific Research Paradigm Reform and Empowers Discipline Advancement Plan, in part by the National Key Research and Development Program of China under Grant 2023YFF0719200, and in part by the Natural Science Foundation of Shanghai under Grant 21ZR1444300. The associate editor coordinating the review of this article and approving it for publication was Prof. Fangqing Wen. (Corresponding author: Minghui Yuan.)

The authors are with the Terahertz Technology Innovation Research Institute, School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China (e-mail: 1920512447zcxxj@gmail.com; 223330889@st.usst.edu.cn; yuanminghui@usst.edu.cn; livw@usst.edu.cn).

Digital Object Identifier 10.1109/JSEN.2025.3564729

their utility in modern security checks. X-ray systems, though capable of revealing detailed object structures, pose health risks due to ionizing radiation. This is particularly concerning for airport staff and travelers frequently exposed to security screenings, as prolonged exposure may lead to significant health issues.

In contrast, millimeter-wave imaging technology employs nonionizing radiation, making it virtually harmless to humans while enabling the detection of a wide range of materials, including both metallic and nonmetallic contraband, thereby offering more comprehensive security screening capabilities. Specifically, passive millimeter-wave imaging detects the weak natural radiation emitted by objects, ensuring high safety and privacy [3], [4], [5], [6]. For instance, this technique can effectively identify concealed metallic and nonmetallic items beneath clothing, such as plastic weapons or explosives. Active millimeter-wave imaging, on the other hand, generates images by emitting millimeter waves and capturing the reflected signals, excelling in long-range detection and maintaining high stability under adverse conditions such as smoke, fog, or rain. By integrating advanced data analysis, it can rapidly identify various contraband items, including firearms, knives, and both solid and liquid explosives.

Moreover, millimeter-wave imaging provides a safer, faster, and more efficient alternative to traditional X-ray or gamma-ray technologies, addressing the growing demand for security checks. This technological evolution not only drives innovation in public safety but also supports more complex security screening scenarios. Integrating millimeter-wave imaging with multiple-input multiple-output (MIMO) and synthetic aperture radar (SAR) technologies significantly enhance imaging resolution and speed [7], [8]. MIMO, leveraging multiple antennas for signal transmission and reception, enhances signal processing and image quality. SAR technology simulates a larger antenna array to enhance image detail, enabling more precise imaging. This synergy not only boosts screening efficiency but also guarantees stable operation.

However, due to the unique mechanisms of millimeter-wave and terahertz imaging, the resulting images often suffer from low resolution and insufficient signal-to-noise ratio (SNR) [9], [10]. These issues severely degrade the quality of captured images, making it difficult to detect images fine details and boundaries. Common issues such as blurred targets edges and loss of high-frequency details pose significant challenges for the accurate detection and localization of contraband. Current imaging methods often struggle to effectively recover meaningful features under such challenging conditions. The existing methods mainly rely on global features and neglect the extraction of boundary details, which leads to the limited accuracy of contraband detection. To address these issues, this article proposes a novel multibranch multiscale semantic segmentation network, the boundary-enhanced multibranch network (BEMBNet). The network uses semantic segmentation to detect contraband. Building on traditional dual-branch networks, incorporates a dedicated boundary branch to capture edge details and boundary features of target objects. We used the MIMO-SAR imaging system independently developed in our laboratory for data collection. Experimental results on the dataset show that this method achieves better segmentation results compared to existing semantic segmentation methods.

The main contributions of this article are as follows.

- 1) A novel three-branch multiscale semantic segmentation network (BEMBNet) designed to address the challenges of low resolution and blurred boundaries in SAR image detection.
- 2) A specialized boundary-processing branch has been introduced. At the cost of increasing the number of parameters, the method efficiently addresses the common issue of blurred or poorly defined boundaries in SAR images. Additionally, an appropriate boundary-aware multiscale fusion modulation (BamFM) is incorporated to facilitate the fusion of this branch with the main branch, effectively enhancing the networks both global and local perception of boundary structures.
- 3) A dynamic reinforced feature pyramid pooling module (DRFPPM) is proposed to implement a fusion scheme based on a channel attention mechanism. Dynamic weighting is applied within channels to more accurately identify salient regions, improving the model's perceptual accuracy.

- 4) A three-branch fusion module (TBFM) is proposed to facilitate the final fusion of the multibranch network, module efficiently balancing detailed spatial information, boundary clarity, and global semantic consistency.

II. RELATED WORK

In recent years, the rapid growth of artificial intelligence has driven significant advancements in deep convolutional neural networks (CNNs), particularly in image classification [11], object detection [12], [13], and image segmentation [14], [15], [16]. A variety of advanced deep learning techniques have been employed to detect targets in millimeter-wave imagery [17], [18], [19].

A. Terahertz Grayscale Images

In recent years, various deep learning algorithms have been applied to target detection in millimeter-wave images. However, most of these methods are tailored for RGB and high-resolution images, which differ greatly from the low-resolution and grayscale nature of millimeter-wave images [20], [21], [22]. Moreover, millimeter-wave detectors and imaging algorithms are limited, resulting in a significantly lower SNR compared to conventional optical images [23]. Millimeter-wave signals are widely used not only in imaging but also in wireless communications. In complex environments, they are prone to severe multipath reflections, which cause signal delay and distortion—further exacerbating the SNR problem. To address these challenges in the communication domain, MIMO-based channel estimation (MIMO CE) [24] techniques have introduced noise weighting mechanisms to suppress interference and enhance estimation accuracy. These methods highlight the importance of robust noise handling in millimeter-wave signal processing, a concern equally critical in imaging scenarios. These characteristics lead to low image contrast, resulting in unclear overall target descriptions and inaccurate boundary segmentation.

Addressing these challenges, target localization and segmentation in millimeter-wave images with SAR imaging characteristics face significant technical obstacles. Several deep learning methods have been proposed to improve detection performance in these scenarios. Liu et al. [25] proposed a method to enhance feature map resolution by applying dilated convolutions. Their approach introduced context embedding into the object detection network and employed dilated convolutions with varying dilation factors on the final feature map. While this method demonstrated the ability to detect contraband, its performance was relatively average. Additionally, its reliance on high-resolution feature extraction resulted in a large model size and increased computational overhead. Pang et al. [26] introduced a multilevel segmentation analysis method for detecting concealed objects. Their approach employed techniques such as vector quantization, expectation maximization, and Bayesian decision-making to cluster pixels using a Gaussian mixture model. This method enabled rapid detection and segmentation of moving individuals carrying concealed weapons. However, it exhibited limitations in identifying object categories inside the human body, unable to achieve precise segmentation. Ding et al. [27] The feature extraction network adopts a dual-branch parallel output

structure, improving the detection of small and blurry targets. However, it struggles with complete boundary segmentation, making it difficult to precisely outline the contours and details of complex objects, which may affect the accuracy and reliability of detection results.

B. Boundary Features

In specific applications such as terahertz image segmentation, local details and edge information play an irreplaceable role. Due to the characteristics of terahertz imaging, the contrast between targets and backgrounds is often weak, making high-frequency edge details critical for distinguishing different regions [28], [29], [30], [31], [32]. Furthermore, the low resolution and contrast of terahertz images make it challenging for traditional segmentation models relying solely on global semantic features to effectively handle local complex features, limiting their performance.

To meet the demands of detecting edge regions and local details: FDGNet [33] introduces frequency disentanglement to enhance domain generalization in hyperspectral image classification. This approach isolates frequency components to improve model robustness and more accurate boundary detection. SFBINet [32] innovatively handles boundary ambiguities by incorporating spatial attention mechanisms and adaptive features, improving accuracy in detecting subtle changes at image boundaries. PIDNet [34] proposes the concept of PID controller, introducing independent boundary branches to assist in the fusion of the outputs of the other two branch networks, in order to enhance the recognition ability of edge regions; DSNet [35] maintains high resolution through both branches, enhancing the extraction and fusion of high-resolution features. Coupled with an efficient feature fusion module, it significantly improves the model's performance in detail processing. The application of these techniques has markedly improved segmentation accuracy in scenarios requiring a high focus on edges and local details.

C. Summary of Related Work

Millimeter-wave image object detection faces challenges due to low resolution, grayscale properties, and poor SNR, which result in low contrast and unclear boundaries. Traditional methods often rely on high-resolution features, increasing computational cost and failing to handle weak contrast effectively. Our model addresses these issues with a multibranch architecture and a dedicated boundary branch for better edge detection. It also employs advanced feature fusion techniques to improve handling of local details and robustness under low SNR conditions, leading to more accurate detection and segmentation compared to existing methods.

III. METHODS

A. Boundary-Enhanced Multibranch Network

This article proposes a semantic segmentation model for millimeter-wave SAR images, called BEMBNet, targeting the detection of contraband. The model addresses challenges posed by low resolution and blurred boundaries in complex scenarios, incorporating deep design and optimization for feature extraction and boundary information utilization. In the

feature extraction stage, a three-branch parallel feature extraction network is adopted. The collaboration between different branches effectively tackles various challenges, achieving a balance between accuracy and computational efficiency. The network framework is shown in Fig. 1. In FDGNet [33], feature disentangling effectively separates and integrates features at different scales, which helps the model better address the low-contrast issues of small objects. However, when facing low SNRs, especially when the signal is obscured by noise, feature disentangling may lead to the loss of important features. In SFBINet [32], bidirectional interaction and multiscale feature fusion enhance small object detection performance, ensuring that small objects are appropriately represented across multiple scales. However, in real-time scenarios that require fast responses, this approach may result in reduced model efficiency. Compared to PIDNet [34], this study introduces a more efficient boundary feature fusion strategy, achieving high-quality fusion of multibranch features, thereby improving segmentation accuracy in edge regions. Similarly, compared to DSNet [35], this model employs a low-resolution context branch as a hub for feature extraction. This design significantly reduces the computational burden caused by direct interaction of high-resolution features, lowering the number of floating-point operations (FLOPs) and greatly improving computational efficiency while maintaining segmentation accuracy. As a result, the model demonstrates significant advantages in computational resource consumption and inference speed.

In the feature extraction stage, the three-branch parallel architecture consists of the boundary branch (B), context branch (C), and high-resolution branch (H), which work collaboratively to extract and fuse features from different aspects.

The boundary branch (B) leverages multiscale feature extraction to capture high-frequency information in the image, enhancing the perception of target edges. The context branch (C) integrates global contextual information through layer-by-layer convolution operations and uses the boundary-aware multiscale fusion (BamF) module to fuse the high-frequency features provided by the boundary branch (B), enabling more precise feature maps and ensuring better multi-level feature fusion. The high-resolution branch (H) maintains high resolution to preserve detailed image information while utilizing down-sampling operations from the context branch to extract rich global information. For the fusion strategy among the three branches, this article introduces a novel TBFM fusion module, which ensures dynamic fusion of the three branches' features at the same resolution. This approach enables efficient integration of multiple feature maps. Through this multibranch collaboration and fusion strategy, BEMBNet simultaneously retains detailed information and global contextual information, significantly improving the model's segmentation performance in edge detection and complex scenarios.

B. Interbranch Fusion Modules

1) *Path Aggregation With Feature Modulation*: The path aggregation with feature modulation (PagFM) module is used to handle the fusion between the context branch (C) and the

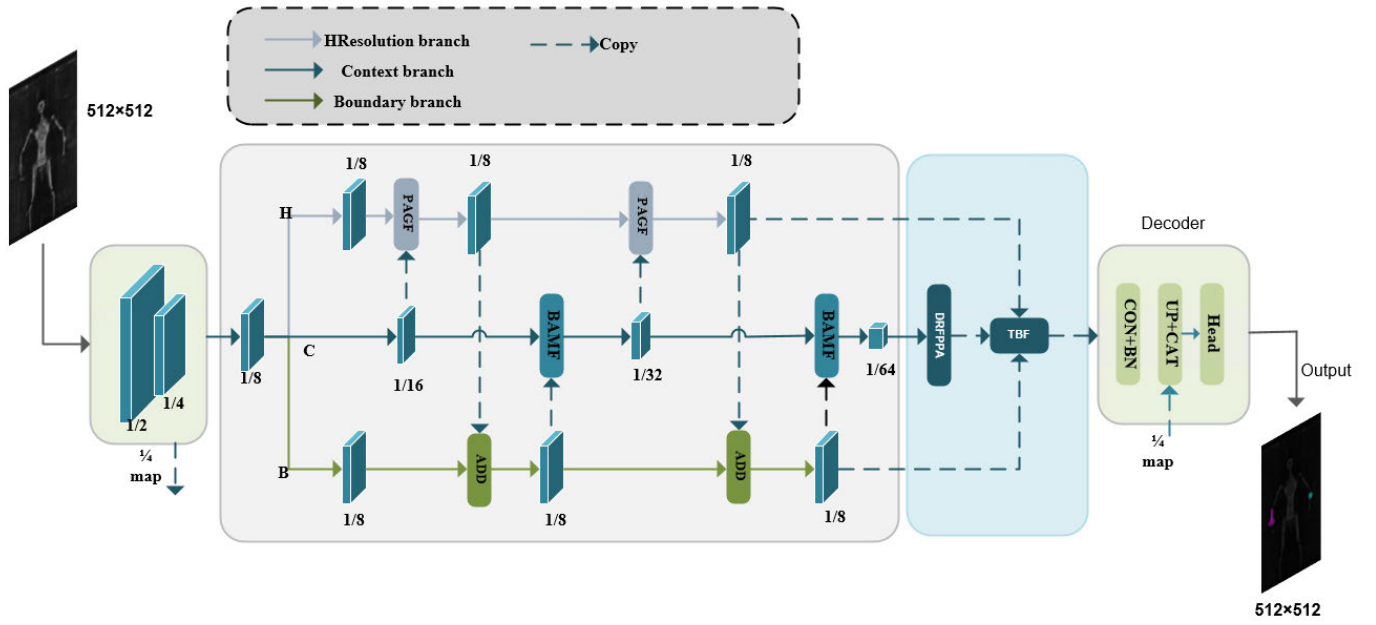


Fig. 1. Network structure diagram. PagF and BamF denotes the dual-branch fusion scheme. DRFPPA and TBF represent the feature pyramid and TBfMs.

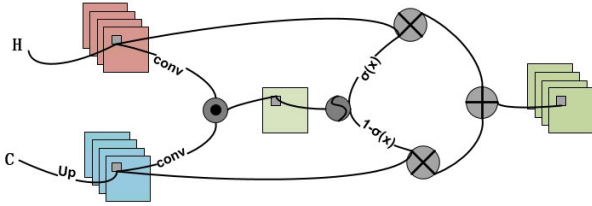


Fig. 2. Pag fusion module.

high-resolution branch (H). Its purpose is to enhance feature representation through feature modulation and adaptive fusion, as illustrated in Fig. 2.

In this module, the processed feature maps from the two branches are first used to compute similarity pixel by pixel. A one-dimensional similarity map is obtained by performing a dot product and summing the results. The dot product, as a measure of similarity, is an effective method that helps weight the capture of target features in local regions. This similarity map is then passed through a sigmoid activation function to generate a weight map. The sigmoid function acts as a non-linear activation, preventing information distortion caused by excessive weight allocation, while providing greater flexibility in the fusion process, enabling the network to automatically adapt to feature variations across different regions. Finally, the similarity-based weights are used to perform a weighted fusion with the vectors $\vec{X}_h$ and $\vec{X}_c$ corresponding to the pixels in the feature maps, producing the weighted output feature map. During the fusion process, the high-resolution branch selectively learns useful semantic features from the context branch, thereby minimizing information loss and ensuring effective feature integration. The formula for the PagFM module is as follows:

$$\sigma = \text{Sigmoid}(f_h(\vec{X}_h) \cdot f_c(\vec{X}_c)) \quad (1)$$

$$\text{Out} = (\sigma \vec{X}_h + (1 - \sigma) \vec{X}_c). \quad (2)$$

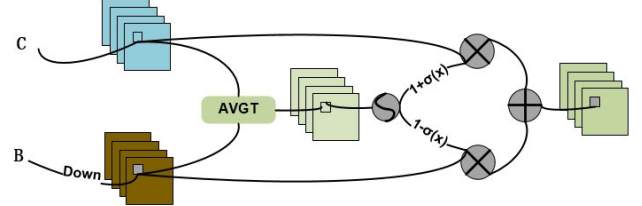


Fig. 3. Bam fusion module.

2) BaMF Modulation: To achieve dynamic fusion of features between the context branch (C) and the boundary branch (B) across different scales, this article proposes the BamFM. The BamFM is designed to enhance the context branch's ability to integrate boundary details and global semantic information by leveraging dynamic attention mechanisms and multiscale context information extraction, thereby improving segmentation performance. The overall structure of BamFM is shown in Fig. 3. The vectors corresponding to the pixels in the feature maps of C and B are $\vec{X}_c$ and $\vec{X}_b$. First, the feature maps from the context branch (C) and the boundary branch (B) are combined via feature stacking to generate a fused feature map. This fused feature map is then fed into a multiscale feature average pooling and gradient-based feature extraction module (AVGT), depicted in Fig. 4. The local attention module employs lightweight convolutions to extract pixel-level details. The multiscale context module uses adaptive pooling to generate feature maps of sizes 4×4 , 8×8 , and 16×16 , encoding context information at different scales. To ensure spatial consistency among multiscale features, the contextual features are upsampled back to the original resolution, preventing excessive detail loss during fusion due to scale differences. The global average pooling operation further extracts global semantic features, ensuring the completeness of the overall

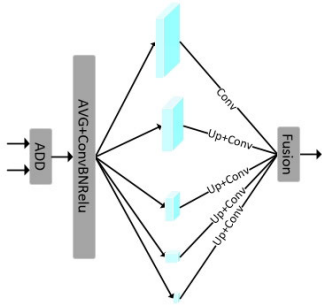


Fig. 4. AVGT.

feature distribution. Through upsampling, convolution, and multiscale feature fusion strategies, the approach effectively enhances the processing of boundary and detail regions in the image, particularly excelling in restoring boundary details and strengthening feature representation.

Finally, local features, context features, and global features are combined through weighted summation to generate the fused features. A sigmoid function is applied to produce dynamic attention weights. To preserve richer contextual information and high-level semantic features in the context branch (C), a larger weight is assigned to X_c during the weighted fusion process. The formula for the BamFM is as follows:

$$\vec{F} = \sum_{i=1,4,8,16,N} f(\text{AvgPool}_i(\vec{X}_c + \vec{X}_b)) \quad (3)$$

$$\sigma = \text{Sigmoid}(\vec{F}) \quad (4)$$

$$\text{Out} = ((1 + \sigma)\vec{X}_c + (1 - \sigma)\vec{X}_b). \quad (5)$$

This mechanism achieves the dynamic reorganization of feature branches, enabling the context branch (C) to simultaneously possess boundary awareness and global representation capabilities. It effectively enhances the model's ability to perceive targets at different scales and improves the segmentation accuracy of boundary details.

C. Dynamic Reinforced Feature Pyramid Pooling Module

To further extract contextual information from feature maps, we propose a new module called the DRFPPM. This module builds on the DAPPM [36] by replacing the original simple addition operation with a channel attention-based dynamic fusion mechanism. During the feature processing stage, dynamic weighting is applied within channels to more accurately identify salient regions. The introduction of an activation function enables adaptive adjustment of salient region features, further enhancing their representation capabilities, and improving the model's perceptual accuracy.

Compared to traditional feature concatenation methods, DRFPPM handles low-resolution images more effectively, especially in complex scenarios with blurred target boundaries. Unlike conventional feature concatenation, the dynamic reinforced feature module (DRFM) in DRFPPM exhibits superior performance in capturing salient regions and modeling global contextual information, making it particularly suitable for the low-resolution and boundary-blurred scenarios we address.

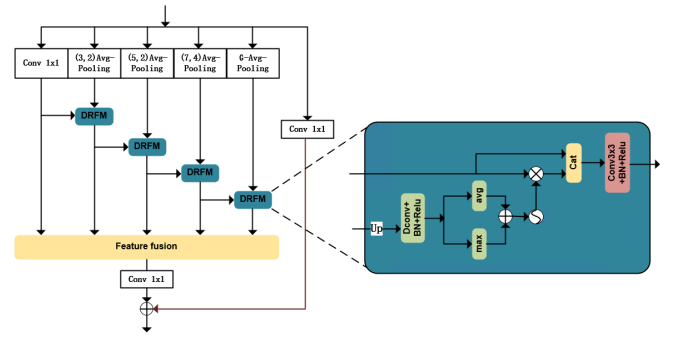


Fig. 5. DRFPPM.

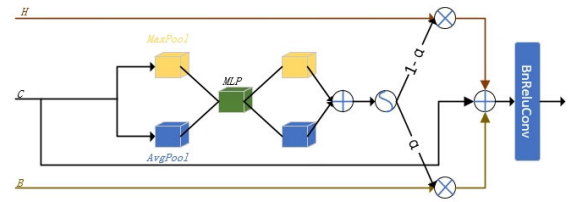


Fig. 6. TBF process.

The architecture of DRFPPM and its dynamic feature fusion module, DRFPPM, is illustrated in Fig. 5.

D. Three-Branch Fusion Module

The TBFM is one of the key components proposed in this article, designed to integrate information from the high-resolution branch (H), context branch (C), and boundary branch (B). The goal is to enhance the model's boundary awareness and global representation capabilities for semantic segmentation tasks. A key challenge in the fusion lies in the allocation of their respective weight ratios. The initial distribution of these weights is particularly critical. Once a base allocation of the three branch weights is established, a channel attention mechanism is employed to further refine and dynamically adjust these weight ratios. By employing the dynamic fusion strategy, TBFM effectively combines the strengths of each branch, achieving efficient integration of semantic features, boundary details, and local spatial information, as illustrated in Fig. 6.

In TBFM, dynamic attention weights are first calculated using the context branch (C). A lightweight attention mechanism based on adaptive pooling is employed to encode global contextual information, generating pixel-level dynamic boundary weights. These weights adjust the contribution ratios of the high-resolution branch (H) and boundary branch (B), allocating appropriate feature weights according to the requirements of specific regions. Under the condition of ensuring real-time task execution, channel attention provides a computationally efficient solution that enhances accuracy. Small object targets appear as low-contrast and simple structures in low resolution and noise. Channel attention enhances the features related to these small objects and reduces background interference by enhancing the most important spectral channels, thereby improving detection accuracy.

The boundary branch (B) captures target edge information, focusing on improving the clarity of boundary regions and the model's adaptability to complex edge scenarios. For semantically salient regions, the module relies heavily on the high-resolution features from the high-resolution branch (H) to ensure rich details. For boundary regions, more reliance is placed on the boundary features from the boundary branch (B), making the edges clearer. The context branch (C), serving as the baseline for global features, is directly incorporated into the fused features to ensure the semantic consistency of the final output. The vectors corresponding to the pixels in the feature maps input to the three branches H , C , and B are $\vec{X}_h$, $\vec{X}_c$, and $\vec{X}_b$, respectively. First, the feature map $\vec{X}_c$ undergoes the channel attention module. W_0 and W_1 are weight matrices used in the model. $\vec{F}_{avg}^c$ and $\vec{F}_{max}^c$ represent the feature vectors obtained from average pooling and max pooling, respectively. These feature vectors are then passed through a shared multilayer perceptron (MLP) network to generate the final channel attention feature map. Subsequently, the channel attention weights are computed using the Sigmoid function to reweight different input features, which are then fused through weighted summation and convolution to produce the final output. The formula for the three-branch feature fusion is as follows:

$$\vec{F} = \text{MLP}(f_{\text{AvgPool}}(\vec{X}_c)) + \text{MLP}(f_{\text{MaxPool}}(\vec{X}_c)) \\ = W_1(W_0(\vec{F}_{avg}^c)) + W_1(W_0(\vec{F}_{max}^c)) \quad (6)$$

$$\sigma = \text{Sigmoid}(\vec{F}) \quad (7)$$

$$\text{Out} = \text{Conv}(\sigma \vec{X}_h + (1 - \sigma) \vec{X}_b + \vec{X}_c). \quad (8)$$

This dynamic fusion mechanism ensures that TBF module efficiently balances detailed spatial information, boundary clarity and global semantic consistency, resulting in robust segmentation performance in complex scenarios.

IV. EXPERIMENTS

A. Dataset

The dataset originates from real-world imaging conducted using an active millimeter-wave body security imaging system based on an MIMO-SAR architecture developed in the laboratory. The system's structure is shown in Fig. 7. [27] This planar scanning system produces JPEG images with a resolution of 200×400 pixels. [37] Examples of the imaging results are illustrated in Fig. 8. Based on near-field radar imaging technology, the system leverages significant differences in the reflection characteristics between materials such as metals, ceramics, and body tissue. These differences in reflection intensity create prominent high-brightness regions in the imaging process. However, when contraband items are in close proximity to the body surface, parts of the contraband extending beyond the body's surface result in localized blurring. Additionally, due to distance variations, the scanning device's imaging can suffer from angular deviations, causing distortion in certain areas. Due to the fixed focal length of the system, scenarios such as those depicted in right side of Fig. 8 arise, where the imaging becomes more blurred, further complicating the task of distinguishing the edges of contraband from the body's contours.

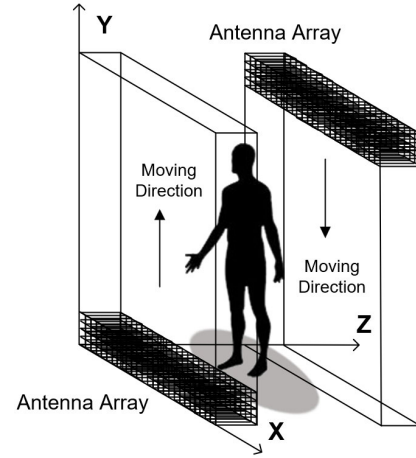


Fig. 7. MIMO-SAR security inspection system structure diagram.

Moreover, the system's resolution limitations make contraband detection even more challenging, particularly for small objects with blurred edges. In low-resolution images, the difficulty of accurately detecting contraband is significantly increased. Imaging noise and blurring effects become key factors that affect detection accuracy, further limiting the effective identification and localization of contraband items.

B. Implementation Details

The SARB-X SARB-Y1 and SARB-Y2 datasets were developed using this equipment, with segmentation experiments conducted. The SARB-X dataset includes six interference categories, containing over 2000 images, with 90% used for training and 10% for validation. Among these, over 1400 images depict standard postures without distortion, while more than 600 images feature nonstandard postures, distortion in the target's position, or overlapping targets, as shown in right side of Fig. 8. The SARB-Y1 dataset covers four categories and contains over 1100 images, with 90% used for training and 10% for validation. To evaluate the model's robustness under more complex conditions, we took into account real-world scenarios where irregular body postures during security checks may result in issues such as occlusion or angle differences of prohibited items. As a result, the SARB-Y2 dataset was designed, consisting of 500 images, primarily featuring irregular body postures, target overlap, and incomplete displays. The entire dataset is intended for testing purposes. During the experiments and tests, all images were resized to 512×512 pixels. The three datasets were annotated using Labelme, with targets hidden along the body's edge during the data collection process. In the SARB-X dataset, the contraband items identified include scissors, wrenches, hammers, handguns, knives, and metal lighters. The SARB-Y1 and SARB-Y2 dataset excludes scissors and lighters, focusing on imaging of prohibited items under specific conditions.

C. Computational Platform

The computational platform used for this study is equipped with a single RTX 3090 GPU, and the environment includes PyTorch 1.12.0, CUDA 11.6, and cuDNN 8.3. We employed

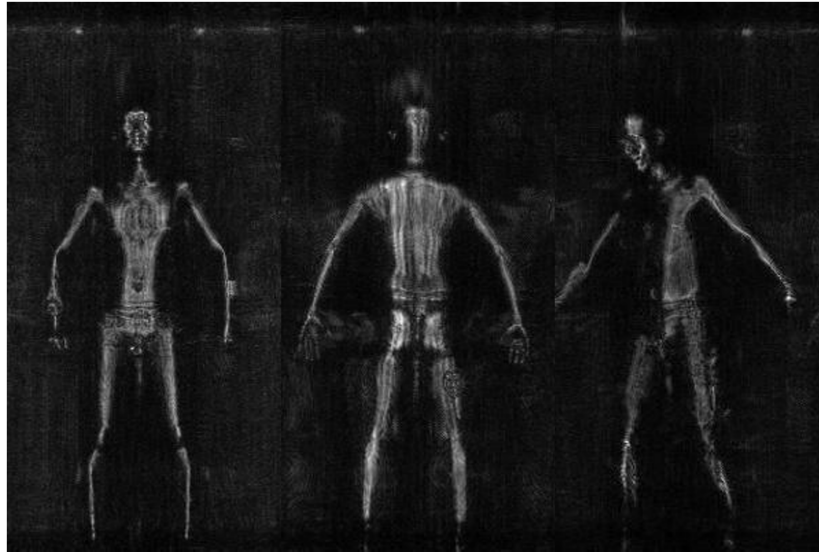


Fig. 8. MIMO-SAR security images.

TABLE I
EFFICIENCY OF DRFPPM

Model	PAPPM	DAPPM	DRFPPM(our)	Miou	FPS	Params
	✓			72.76	27.89	32.95
BEMBNNet		✓		73.21	25.65	32.95
			✓	74.02	25.59	32.3

PAPPM is proposed by PIDNet and DAPPM is proposed by DDRNet.

stochastic gradient descent (SGD) as the optimizer, with a momentum of 0.9. No pretraining or freezing experiments were used for model training. Both the training and validation phases were conducted on this platform. During training, the batch size was set to 16, and the training duration was 300 epochs. For measuring inference speed, we normalized the batch size into the convolutional layers and set the batch size to 1 for inference speed testing.

D. Ablation Experiments

1) *Efficiency of DRFPPM*: To verify the effectiveness of our DRFPPM, we compared it with the DAPPM [36] and PAPPM [34] through a simple ablation study. As shown in Table I, we managed to reduce the number of parameters while achieving an accuracy of 74.02%. This demonstrates that, under the same training conditions, our feature pyramid module performs better.

2) *Efficiency of Three-Branch Architecture*: To validate the effectiveness of the three-branch architecture and the TBFM, we conducted an experiment where we removed the boundary branch (*B*) and integrated its feature extraction into the high-resolution branch (*H*), thus reducing the network to a two-branch architecture with *H* and *C* branches performing parallel feature extraction. The results showed that our three-branch module significantly improved segmentation performance while causing only a slight reduction in frames per second (FPS). Additionally, we compared our three-branch fusion strategy with two alternatives: the simple ADD module and the bag module from the PIDNet [34], [38] architecture.

TABLE II
EFFICIENCY OF FUSION AND THIRD BRANCH

Model	Fusion Scheme			Third Branch	Miou	FPS
	ADD	Bag	TBF(our)			
	✓				69.41	26.84
BEMBNNet	✓			✓	72.94	24.26
		✓		✓	73.31	27.96
			✓	✓	74.02	25.59

Where ADD refers to the addition strategy, Bag refers to the fusion strategy used in PIDNet, and Third branch indicates whether the third branch is used.

As seen in Table II, our TBFM achieved superior results, further proving the effectiveness of our proposed approach.

E. Comparison

To ensure fairness, we conducted comparison experiments on our platform with identical training parameters across two datasets. The selected state-of-the-art models for comparison include Deeplabv3+ [29], HRNetv2-w32 [38], DDRNet-23 [36], PIDNet-L, PIDNet-M [34], DSNet-256, and DSNet-Base [35]. These are all highly representative models, and for each network, we selected the most notable version for the comparison. As shown in Table III, the results of all models trained and tested on the SARB-X dataset are presented. The experimental results indicate that our method achieves a good balance between inference speed and accuracy. BEMBNNet outperforms other models in segmentation performance, with MPA and mIoU values of 74.02% and 83.24%, respectively. A higher mIoU indicates that the predicted masks are closer to the ground truth, while a higher MPA indicates more accurate localization of the targets. Specifically, BEMBNNet achieves higher accuracy than DSNet-base, PIDNet-L, and DDRNet-23, with improvements of 0.65%, 3.71%, and 1.63%, respectively. Moreover, BEMBNNet achieves better performance than the second-best DSNet-base in terms of mIoU, while having only 1/5 of the FLOPs required by DSNet-base. This demonstrates the efficiency and effectiveness of our model.

TABLE III
COMPARISON OF SEGMENTATION PERFORMANCE OF
EACH MODEL IN THE SARB-X

Method	Miou	MAP	GFLOPs	Params	FPS
Deeplabv3+	63.23	70.18	54.71	170.68	21
HRNetv2	64.46	72.33	29.54	90.97	11.1
DDRNet-23	72.39	81.89	20.15	36.0	50.55
PIDNet-M	69.78	78.96	28.5	44.58	33.2
PIDNet-L	70.31	79.34	36.94	69.01	27
DSNet-256	72.65	80.61	9.82	101.8	23.97
DSNet-base	73.37	82.56	37.46	328.3	25.73
BEMBNet(our)	74.02	83.24	32.4	59.23	25.59

TABLE IV
COMPARISON OF SEGMENTATION PERFORMANCE OF EACH
MODEL IN THE SARB-Y1 AND SARB-Y2

SARB	Method	Miou	MAP	hammer	pistol	wrench	knife
Y1	HRNetv2	67.78	75.78	56.59	62.78	65.92	53.79
	DDRNet-23	73.17	84.07	65.18	66.47	72.12	62.25
	PIDNet-L	72.74	82.48	64.04	66.71	71.13	61.82
	DSNet-base	75.91	85.2	70.14	70.39	76.25	62.93
	BEMBNet(our)	76.49	85.88	70.51	70.82	73.85	67.4
Y2	HRNetv2	60.61	74.88	47.44	54.68	58.32	42.63
	DDRNet-23	61.36	75.04	47.14	56.35	58.16	46.17
	PIDNet-L	60.84	75.03	48.35	55.77	59.06	41.17
	DSNet-base	63.11	76.65	51.79	56.65	59.71	47.52
	BEMBNet(our)	65.07	78.82	51.60	59.35	61.45	53.73

Additionally, BEMBNet performs well across all model parameters compared to other models.

Table IV shows the results of training and testing in the SARB-Y1 and SARB-Y2 dataset, where we selected the models that performed well on the SARB-X dataset for further comparison. The experiment demonstrates that, regardless of the quality of the SAR images, our model still maintains superior segmentation performance. BEMBNet stands out particularly in this task, outperforming other models across multiple evaluation metrics under the more complex conditions of the SARB-Y2 dataset. The robustness of the model has also been well validated.

SARB-Y2 was used as the test set, with SARB-Y1 serving as the training set. The segmentation results were tested using four representative images from the SARB-Y2 dataset. The experimental results of each model are shown in Fig. 9.

It can be observed that our method maintains a leading performance even under conditions of incomplete contraband imaging and varying human postures, with MPA and mIoU values of 65.07% and 78.82%, respectively. Only the BEMBNet successfully achieves precise segmentation of the contraband targets in all four images. Particularly, in the last image, where the human body is in a tilted posture and the pistol is not fully visible in the imaging, BEMBNet still delivers an excellent segmentation performance. The results demonstrate that our model exhibits strong robustness under complex conditions.

For the segmentation performance of various object categories in the SARB-X dataset, this article provides the results of each model in Table V, which shows the segmentation performance on different object types within the dataset. It can be seen that the model presented in this article achieves outstanding performance in segmenting these six types of contraband, with the IoU scores for several contraband items

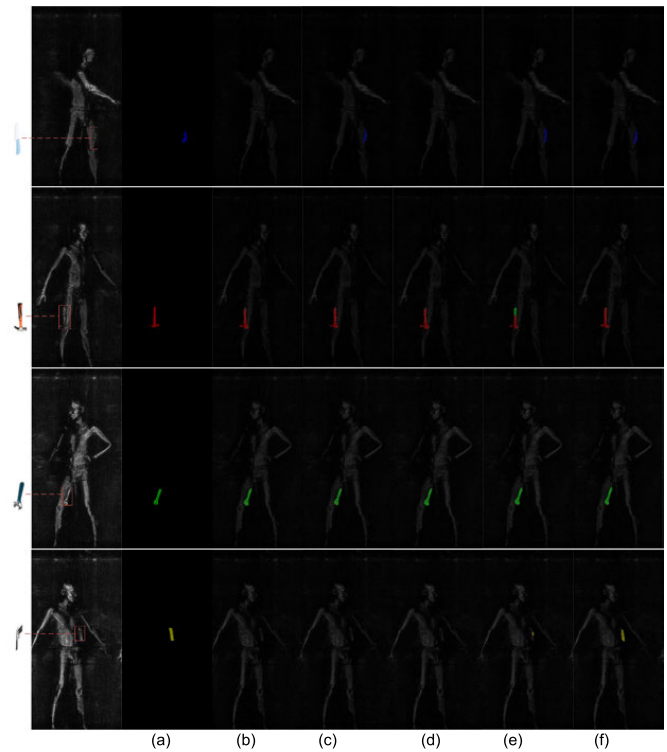


Fig. 9. The far-right column shows the original images along with the corresponding forbidden items, with the visualization results for five instances in the SARB-Y2 dataset. In the visualizations, background is represented in black, knife in blue, wrench in green, hammer in red, and pistol in yellow. The subscript represents the model used, and G.T. denotes the ground truth annotation of the original image. (a) G.T. (b) HR. (c) DDR. (d) PID-L. (e) DS-B. (f) OUR.

consistently reaching an excellent level. The segmentation results were tested using five representative images from the SARB-X dataset. The experimental results of each model are shown in Fig. 10.

The first three rows show the segmentation results for six types of contraband under normal conditions. The fourth row displays the segmentation performance on images with overlapping contraband objects, while the fifth row deals with images where the human posture is irregular and the photograph is distorted.

In the first row, which shows clear frontal images, all models successfully segmented the contraband in the images. However, only BEMBNet was able to segment some of the shape of the open end of the wrench head. In the second row, only DSNet and BEMBNet showed better identification of the hammer head. Although BEMBNet overall segmentation performance was not as good as DSNet, its segmentation details around the head were closer to the ground truth. In the third row, the segmentation results across the models were similar, with all models identifying the two contraband items well. However, for the segmentation of the target boundaries, BEMBNet was closer to the ground truth.

However, in the fourth row, where the test image contains overlapping targets, the segmentation task becomes more difficult. From the results, other models showed incorrect or incomplete segmentation, while BEMBNet achieved the segmentation most closely aligned with the ground truth. In the fifth row, where distorted images were tested, earlier models

TABLE V
COMPARISON OF TARGET SEGMENTATION PERFORMANCE OF EACH MODEL IN THE SARB-X DATASET

Method	Miou	hammer	pistol	wrench	knife	scissors	lighter
HRNetv2	64.46	48.73	65.55	58.69	44.98	71.76	61.72
DDRNet-23	72.39	59.6	74.15	70.87	60.08	74.66	67.56
PIDNet-M	69.78	57.85	66.62	68.97	57.1	71.18	66.82
PIDNet-L	70.31	61.08	71.68	68.19	59.02	73.97	58.37
DSNet-256	72.65	68.45	71.4	72.83	61.48	74.43	60.07
DSNet-base	73.37	68.23	72.3	71.5	61.58	75.15	65.27
BEMBNNet(our)	74.02	67.37	74.01	72.11	61.61	76.02	67.02

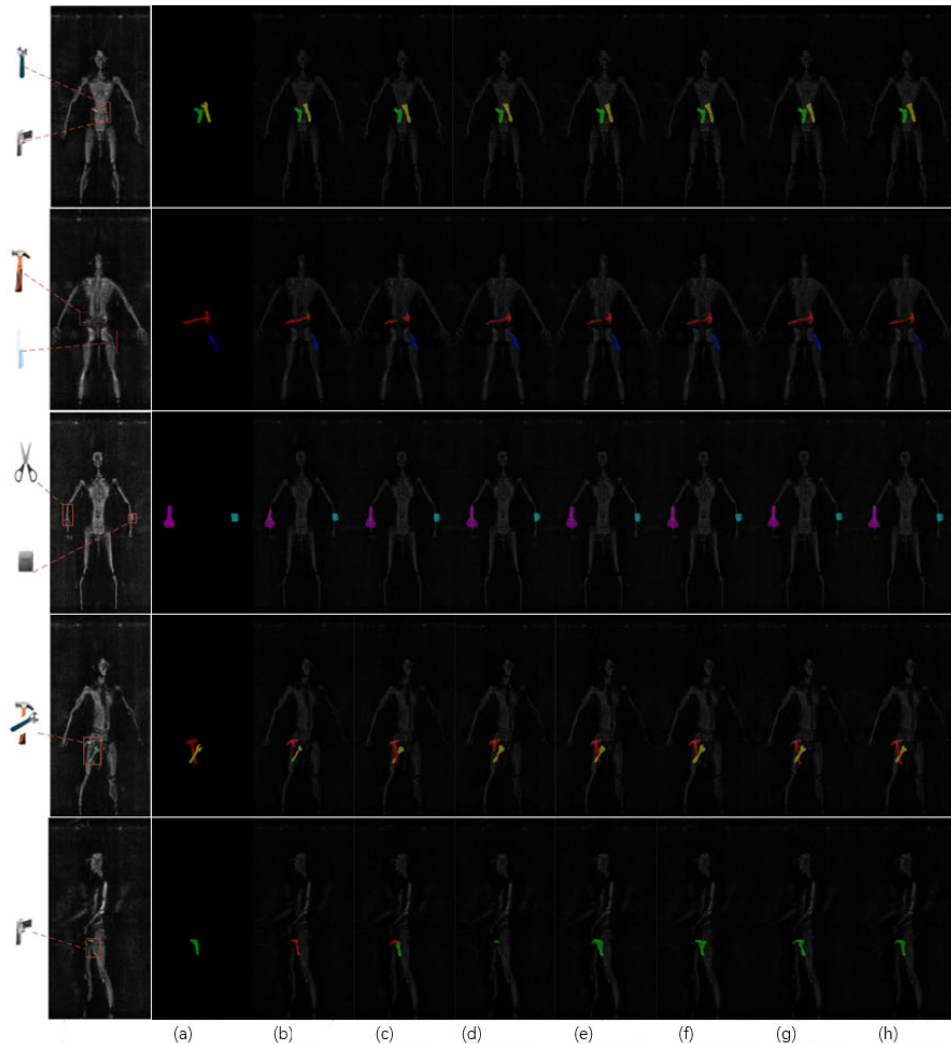


Fig. 10. Far-right column shows the original images along with the corresponding forbidden items, with the visualization results for seven instances in the SARB-X dataset. In the visualizations, background is represented in black, pistol in green, wrench in yellow, hammer in red, knife in blue, scissors in purple, and lighter in cyan. The subscript represents the model used, and G.T denotes the ground truth annotation of the original image. (a) G.T. (b) HR. (c) DDR. (d) PID-M. (e) PID-L. (f) DS-256. (g) DS-B. (h) OUR.

struggled to perform well, with BEMBNNet again being the closest to the ground truth.

These experimental results indicate that BEMBNNet has superior pixel classification accuracy and better capability in detecting small object boundaries, particularly in challenging and complex scenarios.

V. DISCUSSION

The BEMBNNet proposed in this article demonstrates superior performance in the task of semantic segmentation of

millimeter-wave SAR images. Experimental results show that BEMBNNet, through the design of a specialized boundary branch and dynamic fusion strategies, not only enhances the perception of boundary information but also significantly improves segmentation accuracy. Particularly in low-resolution and boundary-blurred scenarios, BEMBNNet is able to accurately capture both boundary features and global semantic information of the targets.

- 1) *Effectiveness of the multibranch architecture:* The experiments confirm that the joint design of the boundary

branch and context branch effectively improves the representation of both boundary and global features. Through a dynamic weighting mechanism, BEMBNet achieves precise segmentation of targets at different scales while reducing model complexity. This enables the model to maintain high segmentation accuracy in low-resolution and boundary-blurred images, while also improving computational efficiency. As a result, BEMBNet offers higher practical applicability, making it more efficient for real-world tasks.

- 2) *Comparison with existing techniques:* Compared to existing models such as HRNet and PIDNet, BEMBNet strikes a better balance between inference speed and segmentation accuracy. The experimental results show that BEMBNet outperforms these models in mIoU and MPA metrics, achieving higher accuracy, especially in low-resolution and boundary-blurred scenarios. This indicates that BEMBNet exhibits superior robustness and generalization ability, allowing it to handle various image qualities and perform better in different security inspection scenarios.
- 3) *Practical implications and limitations:* Although BEMBNet performs excellently on the experimental datasets, real-world security inspection environments still pose challenges such as image noise, lighting condition variations, and distortions, which may affect the model's performance. In particular, when dealing with poor imaging conditions, the accuracy of the model may degrade. Therefore, further optimization of BEMBNet is needed to better handle complex real-world scenarios. Future work could focus on improving the model's robustness by incorporating advanced image preprocessing techniques, noise reduction methods, and adversarial training, thereby enhancing its performance under a wider range of environmental conditions.

VI. CONCLUSION

This article proposes the BEMBNet model based on a multibranch architecture to address the challenges of low target resolution and blurred boundaries in millimeter-wave SAR images. By introducing a boundary-aware module and dynamic fusion strategies, the model efficiently integrates multiscale features, significantly enhancing image segmentation accuracy. Experimental results demonstrate that BEMBNet outperforms leading models across multiple metrics, particularly in low-resolution and complex background scenarios, showcasing its superior robustness. Ablation experiments further confirm that the proposed DRFPPM feature pyramid and TBF multibranch fusion strategies effectively improve the mIoU score. Moreover, BEMBNet shows promise for handling various real-world challenges, such as noisy environments and complex object textures, which are typically difficult for conventional models.

The design of BEMBNet provides a new direction for the development of future security inspection technologies and lays the foundation for related research in millimeter-wave image processing. Future work could further integrate deep learning with millimeter-wave imaging technologies to address

real-world challenges in more complex scenarios and large-scale datasets. Additionally, the versatility of BEMBNet can support a broad range of remote sensing applications, expanding its use beyond security to areas like environmental monitoring. Our method can be adapted for various other remote sensing tasks, including detecting vessels, classifying land types, and monitoring water bodies. Looking ahead, we plan to explore the use of instance segmentation for SAR image detection, aiming to extract as much valuable information as possible from these images. This approach could significantly contribute to the ongoing development of target recognition techniques in SAR imaging, offering new avenues for improving the accuracy and efficiency of automated detection systems in diverse applications.

REFERENCES

- [1] X. Chen et al., "Terahertz (THz) biophotonics technology: Instrumentation, techniques, and biomedical applications," *Chem. Phys. Rev.*, vol. 3, no. 1, Mar. 2022, Art. no. 011311.
- [2] V. Camarchia, R. Quaglia, A. Piacibello, D. P. Nguyen, H. Wang, and A.-V. Pham, "A review of technologies and design techniques of millimeter-wave power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 7, pp. 2957–2983, Jul. 2020.
- [3] G. Valušis, A. Lisauskas, H. Yuan, W. Knap, and H. G. Roskos, "Roadmap of terahertz imaging 2021," *Sensors*, vol. 21, no. 12, p. 4092, Jun. 2021.
- [4] M. Gezimat and G. Singh, "Terahertz imaging and sensing for healthcare: Current status and future perspectives," *IEEE Access*, vol. 11, pp. 18590–18619, 2023.
- [5] M. S. Saadat, S. Sur, S. Nelakuditi, and P. Ramanathan, "MilliCam: Hand-held millimeter-wave imaging," in *Proc. IEEE 29th Int. Conf. Comput. Commun. Netw. (ICCCN)*, Aug. 2020, pp. 1–9.
- [6] T. Nozokido, M. Noto, and T. Murai, "Passive millimeter-wave microscopy," *IEEE Microw. Wireless Compon. Lett.*, vol. 19, no. 10, pp. 638–640, Sep. 2009.
- [7] R. Appleby and R. N. Anderton, "Millimeter-wave and submillimeter-wave imaging for security and surveillance," *Proc. IEEE*, vol. 95, no. 8, pp. 1683–1690, Aug. 2007.
- [8] X. Zang et al., "Metasurfaces for manipulating terahertz waves," *Light, Adv. Manuf.*, vol. 2, no. 2, p. 148, 2021.
- [9] Y. Peng et al., "Three-step one-way model in terahertz biomedical detection," *Photonix*, vol. 2, no. 1, pp. 1–18, Jul. 2021.
- [10] C. Wu, F. Zhang, B. Wang, and K. R. Liu, "mmTrack: Passive multi-person localization using commodity millimeter wave radio," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, Jul. 2020, pp. 2400–2409.
- [11] M. Kang, C.-M. Ting, F. F. Ting, and R. C.-W. Phan, "ASF-YOLO: A novel YOLO model with attentional scale sequence fusion for cell instance segmentation," *Image Vis. Comput.*, vol. 147, Jul. 2024, Art. no. 105057.
- [12] M. Q. Alkhatib et al., "Tri-CNN: A three branch model for hyperspectral image classification," *Remote Sens.*, vol. 15, no. 2, p. 316, Jan. 2023.
- [13] M. Yuan, Q. Zhang, Y. Li, Y. Yan, and Y. Zhu, "A suspicious multi-object detection and recognition method for millimeter wave SAR security inspection images based on multi-path extraction network," *Remote Sens.*, vol. 13, no. 24, p. 4978, Dec. 2021.
- [14] Z. Wei et al., "Stronger, fewer, & superior: Harnessing vision foundation models for domain generalized semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 28619–28630.
- [15] Z. Wu, X. Zhang, F. Li, S. Wang, L. Huang, and J. Li, "W-net: A boundary-enhanced segmentation network for stroke lesions," *Expert Syst. Appl.*, vol. 230, Nov. 2023, Art. no. 120637.
- [16] Y. He, L. Dai, and H. Zhang, "Multi-branch deep residual learning for clustering and beamforming in user-centric network," *IEEE Commun. Lett.*, vol. 24, no. 10, pp. 2221–2225, Oct. 2020.
- [17] B. Alexey and B. Evgeny, "Boundary loss for remote sensing imagery semantic segmentation," in *Proc. Int. Symp. Neural Netw. Cham, Switzerland: Springer*, 2019, pp. 388–401.
- [18] Y. Wu et al., "D-former: A U-shaped dilated transformer for 3D medical image segmentation," *Neural Comput. Appl.*, vol. 35, no. 2, pp. 1931–1944, Jan. 2023.

- [19] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 286–301.
- [20] X. Wang et al., "Self-paced feature attention fusion network for concealed object detection in millimeter-wave image," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 1, pp. 224–239, Jan. 2022.
- [21] P. Zhao et al., "Human tracking and identification through a millimeter wave radar," *Ad Hoc Netw.*, vol. 116, May 2021, Art. no. 102475.
- [22] S. Yeom et al., "Real-time outdoor concealed-object detection with passive millimeter wave imaging," *Opt. Exp.*, vol. 19, pp. 2530–2536, Jan. 2011.
- [23] P. Sun, T. Liu, X. Chen, S. Zhang, Y. Zhao, and S. Wei, "Multi-source aggregation transformer for concealed object detection in millimeter-wave images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 9, pp. 6148–6159, Sep. 2022.
- [24] H. Wang, P. Xiao, and X. Li, "Channel parameter estimation of mmWave MIMO system in urban traffic scene: A training channel-based method," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 1, pp. 754–762, 2024.
- [25] T. Liu, Y. Zhao, Y. Wei, Y. Zhao, and S. Wei, "Concealed object detection for activate millimeter wave image," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9909–9917, Dec. 2019.
- [26] L. Pang, H. Liu, Y. Chen, and J. Miao, "Real-time concealed object detection from passive millimeter wave images based on the YOLOv3 algorithm," *Sensors*, vol. 20, no. 6, p. 1678, Mar. 2020.
- [27] J. Ding, B. Su, and M. Yuan, "Object recognition for millimeter wave SAR images based on dual-branch multiscale fusion network," *IEEE Sensors J.*, vol. 24, no. 8, pp. 13465–13476, Apr. 2024.
- [28] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234–241.
- [29] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2017.
- [30] 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. 2016.
- [31] Q. Zhou, L. Wang, G. Gao, B. Kang, W. Ou, and H. Lu, "Boundary-guided lightweight semantic segmentation with multi-scale semantic context," *IEEE Trans. Multimedia*, vol. 26, pp. 7887–7900, 2024.
- [32] H. Sun, Y. Yao, L. Zhang, and D. Ren, "Spatial focused bitemporal interactive network for remote sensing image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5639115.
- [33] B. Qin, S. Feng, C. Zhao, B. Xi, W. Li, and R. Tao, "FDGNet: Frequency disentanglement and data geometry for domain generalization in cross-scene hyperspectral image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 2, no. 1, pp. 1–14, Aug. 2024.
- [34] J. Xu, Z. Xiong, and S. P. Bhattacharyya, "PIDNet: A real-time semantic segmentation network inspired by PID controllers," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2023, pp. 19529–19539.
- [35] Z. Guo, L. Bian, H. Wei, J. Li, H. Ni, and X. Huang, "DSNet: A novel way to use atrous convolutions in semantic segmentation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 35, no. 4, pp. 3679–3692, Apr. 2025.
- [36] Y. Hong, H. Pan, W. Sun, and Y. Jia, "Deep dual-resolution networks for real-time and accurate semantic segmentation of road scenes," 2021, *arXiv:2101.06085*.
- [37] B. Su and M. Yuan, "Object recognition for millimeter wave MIMO-SAR images based on high-resolution feature recursive alignment fusion network," *IEEE Sensors J.*, vol. 23, no. 14, pp. 16413–16427, Jul. 2023.
- [38] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 5693–5703.



Chuizheng Zeng received the bachelor's degree from Zhengzhou University of Light Industry, Zhengzhou, China, in 2022. He is currently pursuing the master's degree with the University of Shanghai for Science and Technology, Shanghai, China.

His research interests include deep learning, semantic segmentation, and boundary processing.



Junhua Ding received the B.S. degree from Shanghai Dianji University, Shanghai, China, in 2022. He is currently pursuing the M.S. degree with Shanghai University of Science and Technology, Shanghai.

His research interests include deep learning, object detection, and semantic segmentation.



Minghui Yuan received the Ph.D. degree from Southeast University, Nanjing, China, in 2006.

At present, he is an Associate Professor with the University of Shanghai for Science and Technology, Shanghai, China. His research interests include terahertz technology.



Yinwei Li (Member, IEEE) received the bachelor's degree in electronic information engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2009, and the Ph.D. degree in signal and information processing from the University of Chinese Academy of Sciences, Beijing, China, in 2014.

In 2019, he is with the Terahertz Technology Innovation Research Institute, Terahertz Spectrum and Imaging Technology Cooperative Innovation Center, University of Shanghai for Science and Technology, Shanghai, China. He is also with Shanghai Institute of Intelligent Science and Technology, Tongji University, Shanghai; and the School of Intelligent Emergency Management, University of Shanghai for Science and Technology. He is currently a Professor with the University of Shanghai for Science and Technology. His research interests include (inverse) synthetic aperture radar (SAR/ISAR) imaging, interferometric SAR system design and signal processing, and terahertz radar imaging.