Pattern Recognition 143 (2023) 109788

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Infrared small target segmentation networks: A survey

Renke Kou[a], Chunping Wang[a], Zhenming Peng[b], Zhihe Zhao[c], Yaohong Chen[d], Jinhui Han[e], Fuyu Huang[a], Ying Yu[a], Qiang Fu[a,*]

[a] Shijiazhuang Campus, Army Engineering University, Shijiazhuang, China
[b] School of Information and Communication Engineering, University of Electronic Science and Technology, Chengdu, China
[c] Department of Craniofacial Plastic and Aesthetic Surgery, The Third Affiliated Hospital of Air Force Medical University, Xi'an, China
[d] Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China
[e] College of Physics and Telecommunication Engineering, Zhoukou Normal University, Zhoukou, China

ARTICLE INFO

Article history:
Received 14 December 2022
Revised 28 May 2023
Accepted 27 June 2023
Available online 28 June 2023

Keywords:
Infrared small target
Characteristic analysis
Segmentation network
Deep learning
Collaborative technology
Data-driven
False alarm
Missed detection

ABSTRACT

Fast and robust small target detection is one of the key technologies in the infrared (IR) search and tracking systems. With the development of deep learning, there are many data-driven IR small target segmentation algorithms, but they have not been extensively surveyed; we believe our proposed survey is the first to systematically survey them. Focusing on IR small target segmentation tasks, we summarized 7 characteristics of IR small targets, 3 feature extraction methods, 8 design strategies, 30 segmentation networks, 8 loss functions, and 13 evaluation indexes. Then, the accuracy, robustness, and computational complexities of 18 segmentation networks on 5 public datasets were compared and analyzed. Finally, we have discussed the existing problems and future trends in the field of IR small target detection. The proposed survey is a valuable reference for both beginners adapting to current trends in IR small target detection and researchers already experienced in this field.

1. Introduction

Infrared (IR) detection systems have certain advantages over active radar imaging systems, namely strong concealment, good portability, and the ability to detect blind areas; and they also have advantages over visible light imaging systems, namely a strong anti-interference ability and smoke penetrability, which is suitable for both day and night scenes [1]. With the development of stealth and camouflage technology, active radar imaging and visible light imaging systems are often unable to meet certain detection requirements, especially in dark environments with strong electromagnetic interference, whereas IR detection systems can ef-

Abbreviations: GCN, global convolutional network; ASPP, atrous spatial pyramid pooling; DFANet, deep feature aggregation network; STDC, short-term dense concatenate; DeRy, deep model reassembly; MAC, memory access cost; MDConv, mixed depthwise convolution; TAM, topological attribution map; GNN, graph neural network; TT, true target; NB, normal background; HB, high-brightness background; EB, edge background; PNHB, pixel-sized noises with high brightness; COCO, common objects in context; CIFAR, Canadian institute for advanced research; MDvsFA, miss detection-false alarm; SIRST, single-frame infrared small target; IRSTD-1k, infrared small target detection with 1000 images; SIATD, small infrared airborne targets dataset; IRSAT, infrared small aircraft targets; UAV, unmanned aerial vehicle; AGPCNet, attention-guided pyramid context network; ACM, asymmetric contextual modulation; AFFPN, attentional fusion feature pyramid network; SBAM, simplified bilinear interpolation attention module; EAA, enhanced asymmetric attention; BAA, bottom-up asymmetric attention; SA, shuffle attention; DNANet, dense nested attention network; SPSCNet, subpixel sampling cuneate network; MSFRF, multiscale feature reweighted fusion module; LSPM, local similarity pyramid module; MTUNet, multi-task UNet; LPNet, local patch network; MPANet, multi-patch attention network; AE, attention encoder; MSPB, multi-scale patch branch; ISNet, infrared shape network; TFD, Taylor fine difference; TOAA, two orientation attention aggregation; GAN, generative adversarial network; FNNet, feature neutralization network; FPNet, feature perturbation network; TBCNet, target-background-classification network; TEM, target extraction module; SCM, semantic constraint module; STNet, small target network; LW-IRSTNet, lightweight infrared small target segmentation network; RISTDnet, robust infrared small target detection network; LCL, local contrast learning; MLCL, multi-scale local contrast learning; ViT, vision transformer; MSA, multihead self-attention; FEM, feature enhancement module; MTU-Net, multi-level TransUNet; MVTM, multi-level ViT module; MFFM, multi-level feature fusion model; IAANet, internal attention-aware network; RPN, region proposal network; SG, semantic generator; BiConvLSTM, bidirectional convolutional long short-term memory; 3DConv, 3D convolutional structure; CM, concatenate matrix; MoCoPnet, motion-contrast prior driven network; CD-RG, central difference residual group; CD-RDB, central difference dense block; CD-Conv, central difference convolution; LSTA, local spatio-temporal attention module; MSE, mean square error; BCEloss, binary cross-entropy loss; TPR, true positive rate; FPR, false positive rate; JSC, Jaccard similarity coefficient; mIoU, mean intersection over union; PR curve, precision-recall curve; SGD, stochastic gradient descent.

* Corresponding author.
E-mail address: Fu_Qiang@aeu.edu.cn (Q. Fu).

https://doi.org/10.1016/j.patcog.2023.109788
0031-3203/

fectively supplement or even replace such traditional technologies [2,3]. Thus, IR detection systems have been widely used in various fields, including reconnaissance, maritime surveillance, precision guidance, and missile interception in military applications, as well as medical imaging, providing early warnings of fires, agricultural production, and leakage measurement in civil applications [4–8].

When a target is quite far (more than 10 km) away from an IR detector, it typically occupies a small area of the image field of view. Furthermore, under the influence of atmospheric scattering refraction and various noises, small targets in IR images have a low signal-to-noise ratio (SNR) and insufficient texture details. Therefore, solving the robustness problem of IR small target detection algorithms has always been a research challenge. Additionally, real-time inference is a pressing problem that is faced when an IR small target detection algorithm is deployed in a mobile terminal. Overall, precision and speed are the key indicators of the performance of IR small target detection systems. Therefore, designing an IR small target detection algorithm with high accuracy, high speed, and good robustness that can be deployed to mobile terminals is significantly important. Recently, various researchers have proposed IR small target detection algorithms mainly based on model-driven single-frame image detection and multi-frame (sequence) image detection, as shown in Fig. 1.

There are many surveys [1,9] on model-driven IR small target detection algorithms. With the development of deep learning, particularly after the introduction of Wang et al.'s [10] IR small target dataset, many data-driven IR small target segmentation networks have emerged, as shown in Fig. 2. Since the labels of IR small target public datasets are in the form of masks, there has been little research on the detection networks of IR small targets with or without anchors. However, IR small target segmentation networks

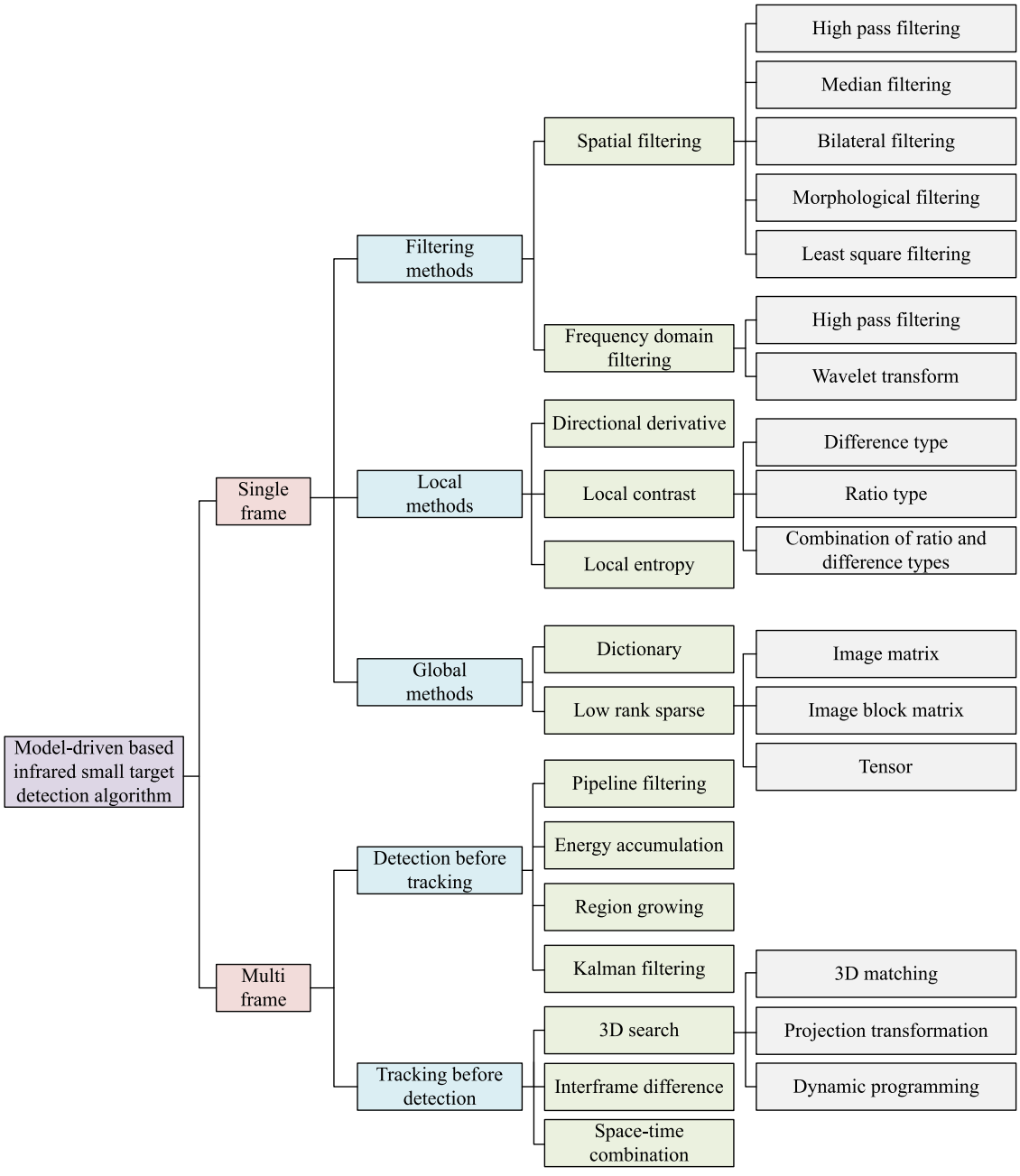

Fig. 1. Mind-mapping of IR small target detection algorithms based on model-driven approach.

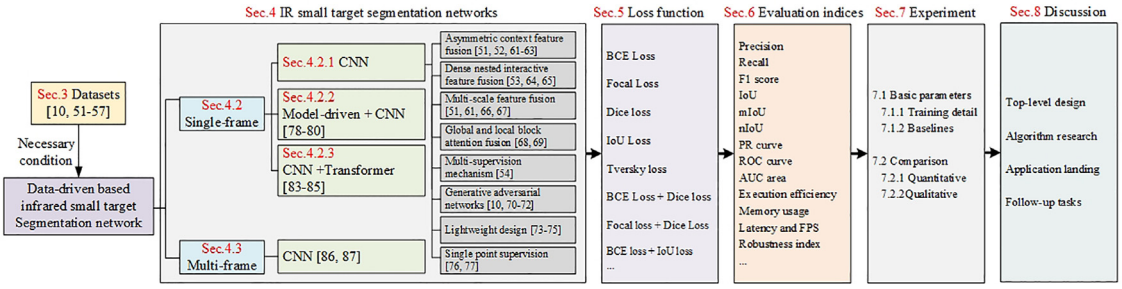

Fig. 2. Mind-mapping of IR small target segmentation networks based on data-driven approach.

supported by public datasets have achieved good detection results and have been widely studied.

In addition, it can be seen from Figs. 1 and 2 that whether it is model-driven or data-driven algorithms, IR small target detection is mainly divided into single-frame or multi-frame detection, with each type of algorithm having its own advantages and disadvantages. The multi-frame detection algorithm mainly utilizes spatial temporal information to detect IR small targets and predict their motion trajectories in sequence images. When there are certain interferences in the field of view that are very similar to the real target (such as shattered clouds), these methods are usually more effective than single-frame algorithms. However, multi-frame detection algorithms undoubtedly increase computational complexity, resulting in poor real-time performance. Conversely, single-frame detection algorithms have better real-time performance, but in complex background interference, detection accuracy needs to be improved. In summary, this survey will comprehensively summarize single-frame and multi-frame IR small target segmentation networks. The contributions of this study are as follows.

(1) We present a novel systematic summary of IR small target segmentation networks based on a deep learning framework. Additionally, we summarized 8 design strategies, 8 loss functions, and 13 evaluation indices of such networks.
(2) For the first time, 8 public datasets, among which 5 public single-frame datasets have been thoroughly analyzed, were introduced; and 7 characteristics of IR small targets were deeply mined. The labels of the 5 public single-frame datasets are all in the form of masks; we provide anchor box annotation for IR small targets in these datasets.
(3) Eighteen segmentation networks were reproduced. Training and testing were conducted using these 5 public datasets. The accuracy, robustness, and computational complexity of the different segmentation networks were compared and analyzed using 8 evaluation indices.
(4) In the survey, we have comprehensively discussed existing problems and future trends in the field of IR small target detection and proposed solutions to these problems.
(5) Relevant materials such as labels, comparative baseline codes, and trained models (. pkl) are available at kourenke/Review-IR-small target-segmentation-networks (github.com).

The remainder of the survey is organized as follows: Section 2 briefly summarizes the design strategies of classic segmentation networks and lightweight networks and discusses the relationship between these networks and the design of IR small target segmentation networks. Section 3 introduces the 8 public datasets considered in the survey and summarizes the seven characteristics of IR small targets using probability statistics. Sections 4–6 summarize IR small target segmentation networks, loss functions, and evaluation indexes, respectively. Section 7 provides quantitative and qualitative analyses of 18 different segmentation networks, and Section 8 discusses the existing problems and future trends in IR small target detection. Finally, Section 9 summarizes the study.

2. Related work

The design ideas of IR small target segmentation networks are largely derived from the classic segmentation networks and lightweight networks. Therefore, we have briefly summarized the innovative points of classic segmentation networks and lightweight networks below.

2.1. Design basis of classic segmentation networks

The encoder–decoder structure, feature fusion method, and innovative points of classic segmentation networks are summarized in Table 1. The design concepts of classic segmentation networks can be summarized as follows. 1) In the coding stage, classical classification networks, such as VGG [11], ResNet [12], GoogLeNet [13], and Xception [14] are primarily used as backbones for coding (downsampling and feature extraction); 2) In the decoding stage, deconvolution, depooling, and linear interpolation (upsampling) are mainly used; 3) The sum or concat methods, corresponding to encoding and decoding, respectively, are employed to conduct feature fusion on the same-scale feature map; 4) Inspired by the attention mechanism of SENet [15], spatial attention or channel attention mechanism modules [16] have been incorporated to extract target features more effectively. 5) The atrous spatial pyramid pooling (ASPP) module under DeepLabV2 [17] has been proposed to solve the problem of multiple target sizes. 6) Classical networks only focus on segmentation accuracy, have high hardware requirements owing to a large number of network parameters, and consume a significant amount of time.

2.2. Design basis of the lightweight networks

A complex network model usually has a better accuracy than a simple one does; however, owing to its high storage space and consumption of computing resources, it is difficult to effectively apply this network to various embedded platforms. Therefore, many scholars have focused on reducing the parameters and computational complexity as much as possible while ensuring that the segmentation accuracy is not too low. The innovative ideas of lightweight networks are summarized in Table 2.

As shown in Table 2, the following strategies were adopted to achieve a lightweight network: 1) reducing the depth of the backbone, number of channels, number of convolutional layers, and complexity in integration methods; 2) replacing the convolutional layer with a group/depth-separable convolution; 3) adding early data processing; 4) canceling the fully-connected layer; and 5) using the design concept of transfer learning to reorganize or split the network model. In addition, the Xinchao Wang team has

Table 1
Summary of classical segmentation networks.

Year	Models	Encoder	Decoder	Feature fusion methods	Innovations
2015	FCN [18]	AlexNet/VGG/ GoogLeNet	Deconvolution	Sum	One of the first networks designed for image segmentation.
2015	DeconvNet [19]	VGG-16	Depooling and deconvolution	–	A decoder is constructed by depooling + deconvolution.
2015	UNet [20]	UNet	Deconvolution	Concat	1. Completely symmetric U-shaped structure. 2. Suitable for medical image segmentation tasks.
2016	FusionNet [21]	ResNet	Deconvolution	Sum	The residual structure of this model is inspired by that of ResNet.
2017	SegNet [22]	VGG-16	Depooling	–	A depooling position index is introduced in the decoding process.
2017	GCN [23]	ResNet	Deconvolution	Sum	To reduce the number of parameters, asymmetric convolution is introduced.
2015 2017 2018	DeepLabV1–3, 3+ [17,24–26]	Xception	Bilinear	Concat	1. Atrous convolution and depthwise separable convolution are introduced. 2. A residual structure is introduced. 3. A multi-scale ASPP module is proposed.
2018	ExFuse [16]	ResNet	Deconvolution and bilinear	Attention mechanism	1. Spatial and channel attention mechanisms are introduced between the encoder and decoder. 2. Multi-scale supervision training is introduced.

Table 2
Summary of lightweight networks.

Year	Models	Innovations
2015	Distillation [27]	A knowledge distillation method wherein the small model is trained with useful information from the large model, and the compressed model provides a similar performance is proposed.
2015	InceptionV2 [28]	A high-resolution feature map descent strategy is proposed.
2015	Slimming [29]	A sparse network channel strategy is proposed to reduce the model size and memory footprint.
2016	ENet [30]	As a representative model for real-time semantic segmentation, it can improve the inference speed by reducing the number of channels, network depth, and shortcut connection.
2017	LinkNet [31]	1. Resnet18 is adopted as the encoder. 2. Feature fusion is carried out through the connection mode of the sum to further improve the accuracy.
2017	SqueezeNet [32]	1. 3×3 convolutions are replaced with 1×1 convolutions. 2. The number of input channels is decreased to 3×3 convolutions. 3. Downsampling is performed late in the network so that convolutional layers have large activation maps.
2017	MobileNetV1 [33]	1. Regular convolution is replaced with depthwise separable convolution to reduce the number of model parameters. 2. Two hyperoarameters (width and resolution multiplier) are introduced to adjust the number of model parameters and calculations. 3. Batch normalization is introduced to accelerate model convergence.
2018	MobileNetV2 [34]	1. A linear bottleneck is introduced to reduce the number of model parameters and calculations. 2. An inverted residuals module is introduced to reduce the amount of model parameters and improve feature expression.
2019	MobileNetV3 [35]	1. Part of the rectified linear unit (ReLU) activation functions is replaced with h-swish activation functions to improve the model accuracy. 2. The Squeeze-and-Excitation (SE) module is introduced to improve the model accuracy.
2017	ShuffleNetV1 [36]	Regular convolution is replaced with group convolution to reduce the number of model parameters.
2018	ShuffleNetV2 [37]	1. Equal channel width minimizes the memory access cost (MAC). 2. Excessive group convolution increases the MAC. 3. Network fragmentation reduces the degree of parallelism. 4. Element-wise operations are non-negligible.
2018	BiSeNet [38]	A bidirectional semantic segmentation network is proposed, and a feature fusion module is introduced to achieve a balance between speed and segmentation performance.
2019	DFANet [39]	Based on multi-scale feature propagation, DFANet substantially reduces the number of parameters while still obtaining a sufficient receptive field and enhancing the model learning ability, striking a balance between speed and segmentation performance.
2019	MixNet [40]	A new mixed depthwise convolution (MDConv) is proposed, in which different kernel sizes are mixed in a convolution operation for feature fusion.
2020	GhostNet [41]	A ghost module is proposed to further reduce the amount of computation.
2021	STDC [42]	The multi-path structure in BiSeNet is improved to reduce the model computation while extracting low-level details.
2021	AmalgamateGNN [43]	A dedicated slimmable graph convolutional operation as well as a novel topological attribution map (TAM) are proposed.
2022	KnowledgeFactor [44]	A large model is split into a series of sub-models; each shares some capabilities of the large model. It can handle a single task with minimal parameters.
2022	DeRy [45]	Different types of deep learning models are assembled like building blocks according to downstream tasks to achieve a high-performance reasoning ability.

proposed many solutions for the design of lightweight networks in the past three years, such as meta aggregation scheme [46], dataset distillation via factorization [47], dynamic sparse transformer scheme [48], and diffusion probabilistic model made slim [49]. If readers are interested in lightweight network design, you can have a deeper understanding.

2.3. Internal relationship between IR small target segmentation networks, classic segmentation networks, and lightweight networks

Some strategies proposed for classic segmentation networks that are shown in Table 1 can be used as a reference. However, IR small targets occupy fewer pixels in the image, as well as in-

sufficient texture and color information and blurred edges, so the network also needs to be designed according to its unique properties. Furthermore, the real-time requirement of IR small target detection is high since it is mainly used in the military. Therefore, the lightweight strategies in Table 2 can facilitate the design of IR small target segmentation networks.

3. Characteristics analysis of IR small targets

As shown in Fig. 3, an IR single-frame image containing a small target consists of three parts: target, background, and noise. These three components can be represented by an additive model as follows:

$$f_I(x,y) = f_T(x,y) + f_B(x,y) + f_N(x,y). \quad (1)$$

where (x, y) denote the coordinates of each pixel in the image; f_I is the raw image; and f_T, f_B, and f_N are the target, background, and noise components, respectively.

A real IR image that can intuitively explain the differences between a true target (TT) and various interference sources is shown in Fig. 3(a). Five typical components were locally magnified: TT, normal background (NB), high-brightness background (HB), edge background (EB), and pixel-sized noises with high brightness (PNHB), as shown in Fig. 3(b)–(f).

Fig. 3 shows the thermal radiation of an imaging scene finally forming an IR image through the complex process of atmospheric transmission, optical systems, and sensor photoelectric conversion. Generally, IR images have the following characteristics [50]. (1) An IR image is a grayscale image that represents the target and background temperature distributions lacking color information. (2) The spatial resolution of a thermal imaging system is generally lower than that of a visible image, resulting in small targets having relatively fuzzy edges, low contrast, and indistinct shapes. (3) An IR image cannot present the specific texture details of a target in a scene.

Overall, the task of IR small target detection is separating a true small target from a complex background. To further analyze the characteristics of IR small targets, we conducted a statistical analysis based on public datasets of IR small targets.

3.1. Public datasets of IR small targets

Datasets are necessary foundations for deep learning. After thoroughly analyzing existing literature, 8 public IR small target datasets have been summarized, as shown in Table 3. The currently published IR small target datasets shown in Table 3 have the following problems. (1) Compared with the number of samples in Common Objects in Context (COCO), ImageNet, Canadian Institute for Advanced Research (CIFAR), and other public datasets, the number in IR small target datasets is too small. (2) These datasets do not contain sufficiently rich IR image background and target information. (3) Dataset labels are pixel-level mask annotations, and their accuracy needs to be improved. (4) There are obvious traces of artificial synthesis in some of the samples. (5) The targets and backgrounds in the datasets have not been sorted and are instead presented in a hodgepodge form.

To promote the research on the detection networks of IR small targets with or without anchors. we have marked the anchor box of 5 public IR small target datasets (datasets 1–5 in Table 3). The forms of the anchor boxes are shown in Table 4, which include the number of targets, centroid coordinates, anchor box, and target pixel size.

3.2. Statistical analysis of characteristics of IR small targets

In the following statistical analysis, the training samples in datasets 1 and 2 and the training + test samples in datasets 3–5 were taken as statistical objects. The number distribution, standard deviation, position distribution, pixel size, aspect ratio, fill ratio, and local contrast of the IR small targets in the image were analyzed.

3.2.1. Distribution of IR small targets

Labeling the centroid of the IR small target in the dataset as the center point, the position distribution of the targets in the image was counted, as shown in Fig. 4. The IR small targets were randomly distributed in the image field of view, adhering to real-world scenarios.

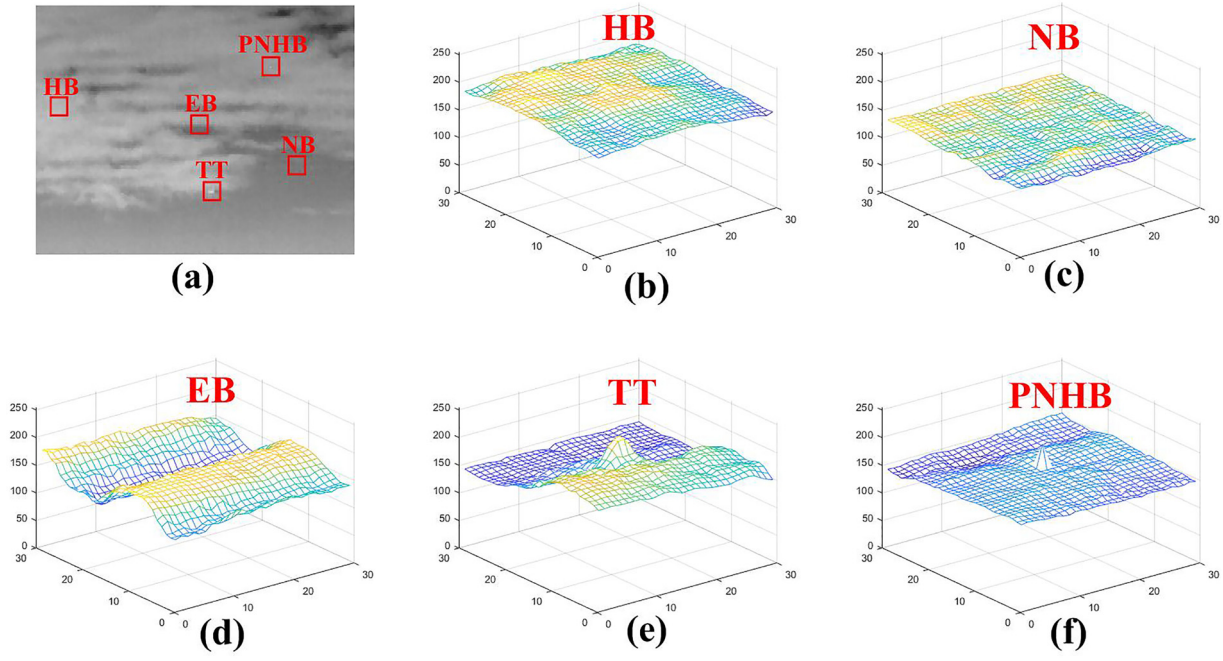

Fig. 3. (a) IR images containing small targets; (b)–(f) 3D mesh of different positions.

Table 3
Introduction to IR small target public datasets.

Datasets		Number of samples		Description of datasets
		Training	Test	
Single frame	MDvsFA [10] (Dataset 1)	9978	100	This is the first published IR small target dataset wherein part of the IR small target and the background are composite images.
	SIRST-Aug [51] (Dataset 2)	8525	545	Based on dataset 3, data enhancement (clipping, inversion, displacement, etc.) is carried out.
	SIRST [52] (Dataset 3)	341	86	This is the one of the first real IR small target dataset with high-quality images and labels.
	NUDT-SIRST [53] (Dataset 4)	1061	265	This dataset contains multiple target types, rich target sizes, and different clutter backgrounds.
	Dataset fusion survey of datasets 1–4	19,905	996	This survey fuses datasets 1–4 to further enrich training and test samples.
	IRSTD-1k [54] (Dataset 5)	800	201	This dataset contains targets of different shapes and sizes, rich clutter backgrounds, and accurate pixel level annotations.
Sequence	SIATD [55] (Dataset 6)	350 sequences (150,185 frames)		This is a semisynthetic dataset of IR small targets in a cluttered background.
	IRSAT [56] (Dataset 7)	22 sequences (16,177 frames)		This dataset contains frames of images of the sky, ground, and other backgrounds as well as a variety of scenes.
Video	Anti-UAV [57] (Dataset 8)	>300 videos		This is the unmanned aerial vehicle (UAV) small target detection dataset under different background conditions, which includes both IR and red, green, and blue (RGB) video sequences.

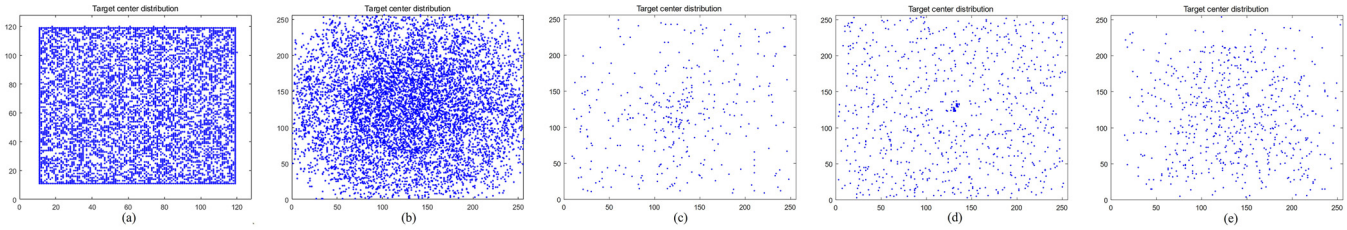

Fig. 4. (a)–(e) Centroid distribution of IR small targets in datasets 1–5.

Table 4
Format of labels.

Targets	Centroid coordinates	Anchor box (X,Y,H,W)	Pixels
2	(185.00, 70.00)	(184, 69, 3, 3)	9
	(201.50, 221.00)	(200, 219, 4, 5)	18

3.2.2. Probability distribution of target number

A group of eight connected pixels in a label are regarded as independent targets; the number of IR small targets was counted, as shown in Fig. 5(a). The IR images in datasets 1–5 were mainly of single targets and contained a few multiple targets, adhering to real-world scenarios. These datasets provide data support for verifying the multi-target detection capabilities of different algorithms.

3.2.3. Probability distribution of target standard deviation

Owing to the long imaging distance and influence of attenuation on the atmospheric transmission process, the image of a small target can present a Gaussian bright spot on the image plane. Its mathematical model can be approximated as a two-dimensional Gaussian function [4]:

$$f_T(x,y) = I_{max} * \exp\left(-\frac{1}{2}\left(\left(\frac{x}{\sigma_x}\right)^2 + \left(\frac{y}{\sigma_y}\right)^2\right)\right). \quad (2)$$

where I_{max} is the intensity of the target peak; and σ_x and σ_y are the diffusion parameters in the horizontal and vertical directions, respectively.

Noteworthily, Eq. (2) only represents an imaging model for ideal small targets. Practically, small targets have different shapes and gray distribution forms; therefore, it is difficult to use a unified mathematical model to describe them. Thus, the probability distribution of the standard deviation of the IR small targets can be statistically analyzed using datasets 1–5, as shown in Fig. 5(b). The peak value of the standard variance σ of the IR small targets on the datasets 1–4 was concentrated between 10 and 25, while that of the IR small targets on the dataset 5 was concentrated between 8 and 12.

3.2.4. Probability distribution of target size and aspect ratio

Regarding deep-learning algorithms, the network depth and convolution kernel size determine the receptive fields of these algorithms. Therefore, the target pixel size and aspect ratio in public datasets 1–5 was counted, as shown in Fig. 5(c) and (d). In Fig. 5(c), the targets in datasets 1–5 are basically less than 100 pixels in size, and dataset 5 belongs to tiny IR targets, which can be called difficult samples. Therefore, in the algorithm design process, downsampling should not be excessively large; otherwise, target information can be lost, and the algorithm cannot be trained. In Fig. 5(d), the peak probability of the aspect ratio of the IR small targets in datasets 1–5 is 1, which conforms to the characteristics of the Gaussian distribution of remote IR small targets.

3.2.5. Probability distribution of target fill ratio

In this survey, the filling ratio was defined as the ratio between the total gray value of the K pixels, including the target center, and the total gray value of the target pixels. In statistics, K takes 1/3 of the total number of target pixels; the statistical results are shown in Fig. 5(e). The energy contained in the target center and K neighborhoods accounted for 35–60% of the entire target energy.

3.2.6. Probability distribution of target local contrast

In IR images, the real target, which may not necessarily be the brightest, is slightly brighter than the surrounding neighborhood background. Although the background may have a high brightness,

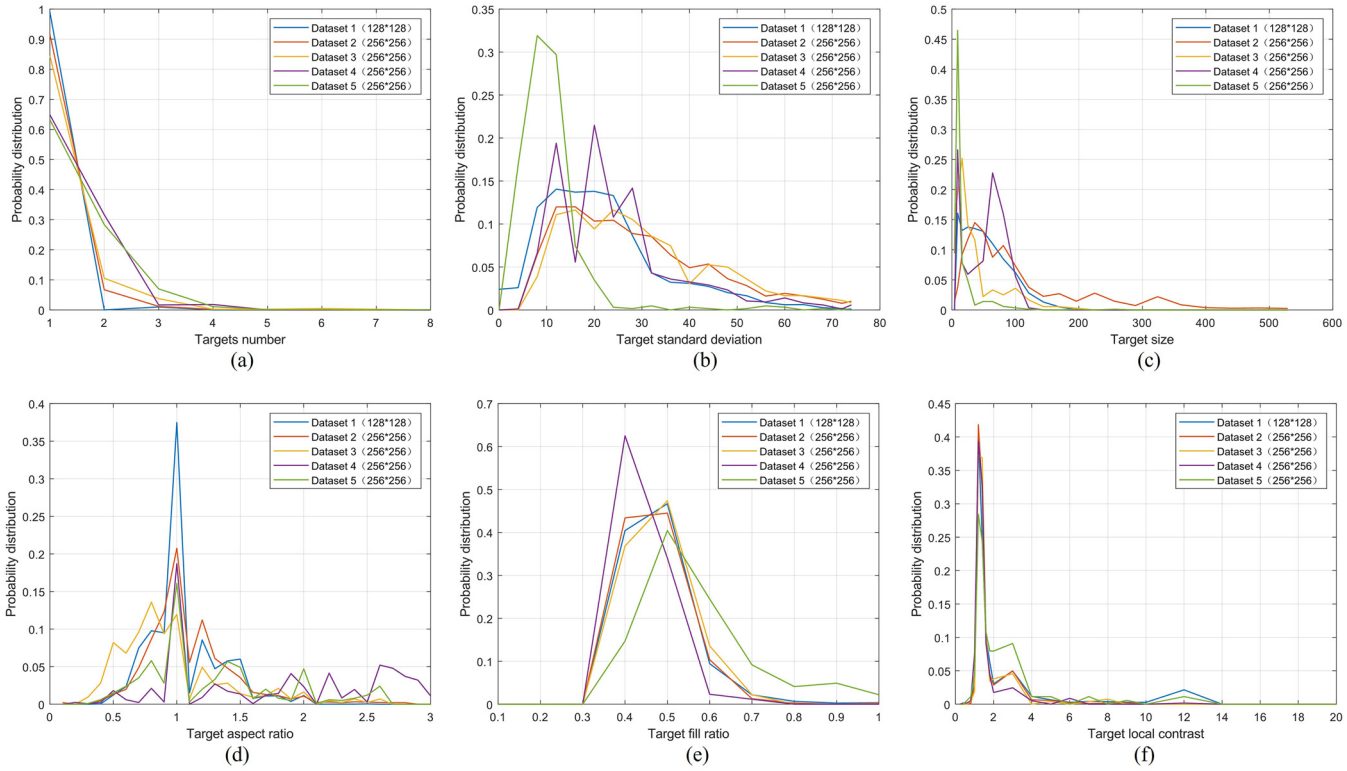

Fig. 5. (a)–(f) Probability distribution of the target's number, standard deviation, size, aspect ratio, fill ratio, and local contrast.

it is usually distributed gently over a large area, and its internal contrast is not prominent.

Following the calculation idea for local contrast presented in existing literature [58], the local contrast in this survey is defined as:

$$C_T = \min_i \frac{I_{mean_0}}{I_{mean_i}}, i = 1, 2 \ldots, 8. \tag{3}$$

where C_T is the local contrast between the target and the neighborhood background, I_{mean_0} is the average gray value of the target, and I_{mean_i} is the mean value of the ith sub-block of the background.

The statistical results for the local contrast obtained through Eq. (3) are shown in Fig. 5(f). The probability of the target's local contrast being greater than and less than 1 accounted for 95.41 and only 4.59%, respectively. The peak probability was concentrated at 1.2–2.2. Owing to the weak local contrast of IR small targets, many scholars have conducted modeling analyses and published a series of local contrast detection algorithms. However, when the small target is on the dark side of the background edge, and the average gray value of the bright-edge background is greater than that of the target, the contrast algorithm is invalid, and resultantly, its robustness is poor. In summary, the seven characteristics of the IR small targets have similar distribution rules in datasets 1–5. Therefore, these 5 datasets can be fused to form a richer IR small target dataset that provides richer data support for the data-driven IR small target segmentation algorithm.

4. IR small target segmentation networks

After the initial proposal of the fully convolutional network (FCN) [18] for an image segmentation task, a series of classic segmentation networks (UNet [20], SegNet [22], and DeepLab [24]) and lightweight segmentation networks (ENet [30], linkNet [31], and BiSeNet [38]) were derived and subsequently applied in various fields, such as medical image segmentation, intelligent driving, and industrial detection. Several data-driven IR small target segmentation algorithms have emerged in the field of IR small target detection. This section summarizes the design strategies for different IR small target segmentation networks.

4.1. Essence of semantic segmentation

Semantic segmentation is a pixel-level classification task. The essence of a semantic segmentation network is an encoder–decoder structure, as shown in Fig. 6. In the coding phase, CNN or Transformer as a backbone is performed to extract high-dimensional features containing semantic information. In the decoding phase, the high-dimensional feature map is upsampled (through deconvolution, depooling, or bilinear interpolation), and the high-dimensional feature vector is generated into a semantic segmentation mask, which is mapped back to the original image, completing the segmentation process. In addition, to fully extract the features and semantic information of the target, many strategies such as attention mechanisms, multi-scale feature fusion, and "sum or concat" connections are nested in various types of networks. The structure of the encoder–decoder is constant in all types of segmentation networks, whether the network is a classical segmentation network, lightweight segmentation network, or segmentation network specially designed for different tasks.

4.2. Single-frame IR small target segmentation networks

Based on the 7 characteristics of IR small targets discussed in Section 3, it can be inferred that an IR small target segmentation task is essentially a binary semantic segmentation task with extremely unbalanced positive and negative samples [59]. This section summarizes different feature extraction methods (CNN, "CNN + Model-driven", and "CNN + Transformer") to support the comparison of the design ideas of different segmentation networks.

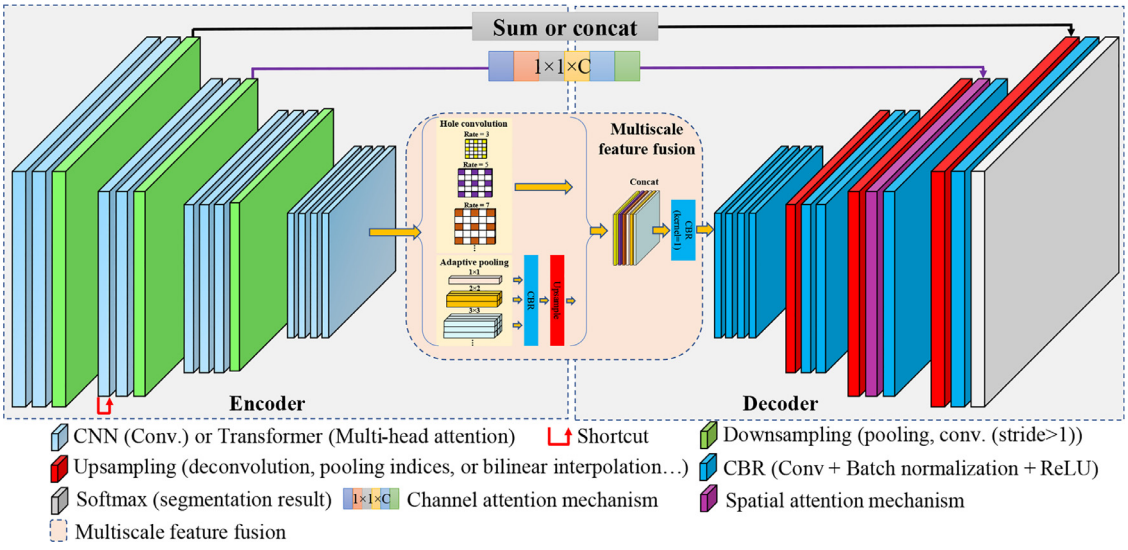

Fig. 6. Structure of encoding and decoding.

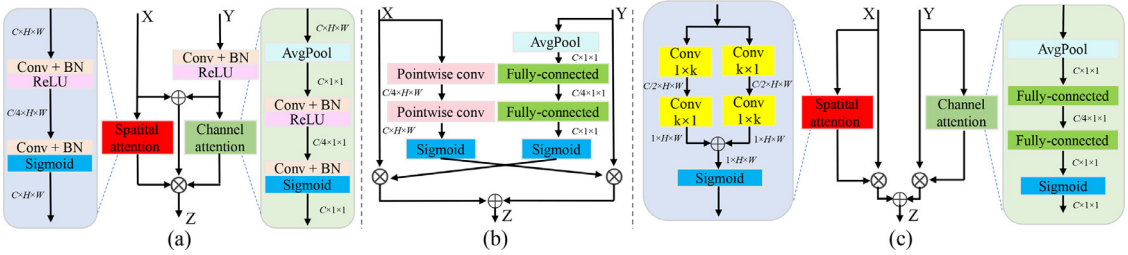

Fig. 7. (a)–(c) Structure of context feature fusion modules for AGPCNet, ACM, and AFFPN ([51] (Fig. 5), [52] (Fig. 5), [61] (Fig. 4)).

4.2.1. IR small target segmentation networks based on CNN method

Many scholars have received valuable information from the design strategies of classic segmentation networks and lightweight networks (Section 2) and combine the characteristics of IR small targets discussed in Section 3 to design IR small target segmentation networks based on CNN feature extraction. The eight strategies under the CNN framework are presented here.

4.2.1.1. Asymmetric context feature fusion strategy. Inspired by the use of ExFuse [16], DFN [60], and SENet [15] in extracting IR small target features effectively, networks such as attention-guided pyramid context network (AGPCNet) [51], asymmetric contextual modulation(ACM) [52], and attentional fusion feature pyramid network (AFFPN) [61] use a top-down channel attention mechanism to extract high-level semantic information and a bottom-up spatial attention mechanism to extract low-level feature information, as shown in Fig. 7(a)–(c); X and Y represent a shallow feature map and deep feature map, respectively. To maximally place the focus of a network on a certain position in space, a pixel-level spatial attention mechanism was adopted. The channel-attention mechanism of global average pooling was adopted to assign as many different weights to as many different channels as possible, and asymmetric context-feature fusion was employed to capture as many small target features as possible.

Yu et al. [62] constructed a simplified bilinear interpolation attention module (SBAM) for hierarchical feature map fusion. This module has a high inference speed and can focus on the characteristics of a target even in the absence of context, as shown in Fig. 8(a). On the one hand, SBAM uses point-by-point convolution to reduce parameters, and on the other hand, it uses an attention mechanism and bilinear interpolation to extract the target features and reduce the error caused by pixel offset. Tong et al. [63] proposed an efficient and powerful enhanced asymmetric attention (EAA) module that uses same-layer feature information exchange and cross-layer feature fusion to improve the feature extraction ability of IR small targets, as shown in Fig. 8(b). This module primarily consists of two components: a bottom-up asymmetric attention (BAA) block and a shuffle attention (SA) block, as shown in Fig. 8(b1) and (b2). The BAA block enables cross-layer feature fusion and highlights the fine details of a target, while the SA block focuses on spatial and channel feature information within the layers through channel shuffling.

4.2.1.2. Dense nested interactive feature fusion strategy. Li et al. [53], He et al. [64], and Liu et al. [65] designed the dense nested attention network (DNANet), subpixel sampling cuneate network (SPSC-Net), and Image Enhancement Net, respectively, to achieve progressive interactions between high- and low-level features and effectively integrate multi-scale feature maps containing low-level detailed features and high-level semantic information. Through repeated fusion and enhancement, the contextual information of the small targets was combined and fully utilized. The structures of these networks are shown in Fig. 9(a)–(c). In addition, we have designed various nested network structures, as shown in Fig. 9(d)–(i). Therefore, we can determine a more suitable nested structure through experiments.

4.2.1.3. Multi-scale feature fusion strategy. Although IR small targets occupy only a few pixels in an image, they undergo changes at different scales. To enable the network to learn the characteristics of IR small targets of different scales as much as possible, based on the ASPP module in DeeplabV2 [17], local similarity pyramid module (LSPM) [66], Attention-Guided Pyramid Context Network (AGPCNet) [51], AFFPN [61], and multi-task UNet (MTUNet) [67],

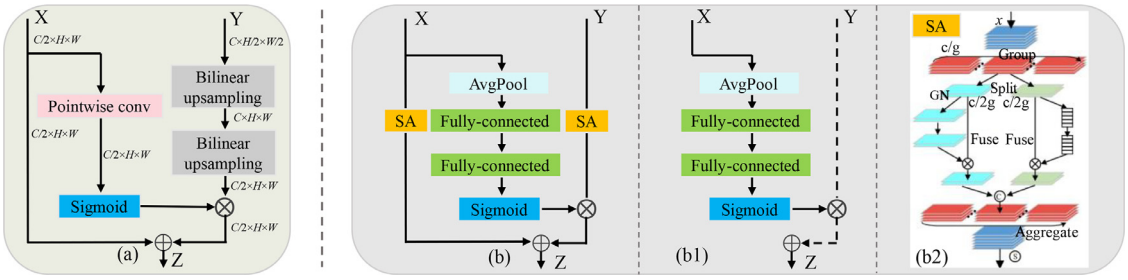

Fig. 8. (a) SBAM context feature fusion module. (b) EAA context feature fusion module. (b1) BAA Subblock in EAA Module. (b2) SA Subblock in EAA Module ([62] (Fig. 3(d)), [63] (Fig. 3)).

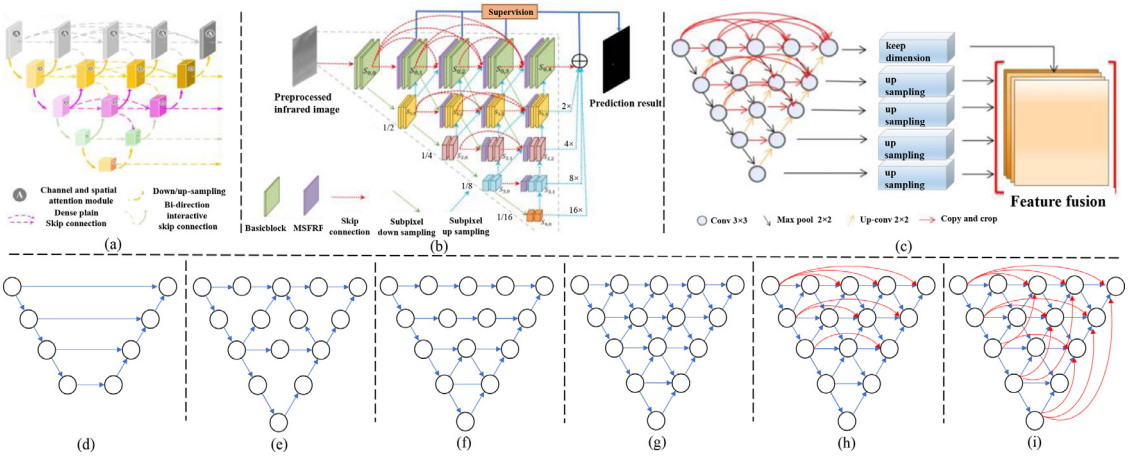

Fig. 9. (a)–(c) Structures of dense nested networks of DNA-Net, SPSCNet, and Image Enhancement Net ([53] (Fig. 3), [64] (Fig. 1), [65] (Fig. 5)). (d)–(i) Structures of inverted trapezoid (UNet), inverted triangle, inverted triangle + trapezoid, dense inverted triangle, dense inverted triangle + horizontal shortcut, and dense inverted triangle + horizontal and vertical shortcut.

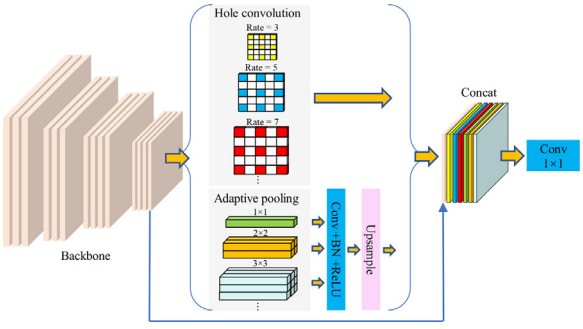

Fig. 10. Network structure based on multi-scale feature fusion strategy.

multi-scale feature maps can be constructed on high-level feature maps using hole convolution and adaptive global average pooling. Then, splicing is performed using the concat operation. Finally, feature fusion is performed using a 1 × 1 convolution, as shown in Fig. 10.

4.2.1.4. Global and local block attention fusion strategy. The extremely unbalanced positive and negative samples of an IR small target lead to a low training efficiency. Chen et al. [68] proposed a local patch network (LPNet) with global attention by combining the global and local characteristics of IR small target images. A supervised attention module trained using a small target spread map was proposed from a global perspective to suppress most background pixels irrelevant to small target features, as shown in Fig. 11(a). From a local perspective, the local patches are separated from the global features and share the same convolution weights in the patch network, as shown in Fig. 11(b). In PatchNet, an inception module composed of multiple k-Conv layers with different kernel sizes is used to extract multi-scale features from each local patch p_n. The subnet is used to capture small target features and select patches containing small targets. Then, patch fusion is used to fuse all the patches to obtain confidence maps; in patch fusion, the pixel value in patch C is the average of the corresponding pixels in all the patches (e.g., patches A and B in Fig. 11(b)) that cover this pixel. Exploiting global and local characteristics, multi-scale features of IR small targets are fused.

Wang et al. [69] proposed a multi-patch attention network (MPANet) based on an axial attention encoder (AE) and a multi-scale patch branch (MSPB) structure. The axial AE module highlights the effective features of small targets and suppresses background noise, as shown in Fig. 12(a). The MSPB module slices the original image into local blocks of different scales, and the axial attention module performs feature extraction. Then, the coarse- and fine-grained features of different semantic scales are integrated, as shown in Fig. 12(b). The overall design idea of the MPANet is as follows. An input is divided into three non-overlapping patches with different scales, where the branch with the original image resolution is regarded as a global branch. Position-sensitive attention is used as the basic module of the encoder, and the remaining branches are considered local. Correspondingly, the non-local block becomes the basic module of the local branches.

4.2.1.5. Multi-supervision mechanism strategy. To accurately capture the shape information of IR small targets, Zhang et al. [54] proposed a new IR shape network (ISNet) that mainly comprises a Taylor fine difference (TFD) block and two orientation attention aggregation (TOAA) blocks, as shown in Fig. 13. The TFD block gathers and enhances comprehensive edge information from different levels to improve the contrast between the target and back-

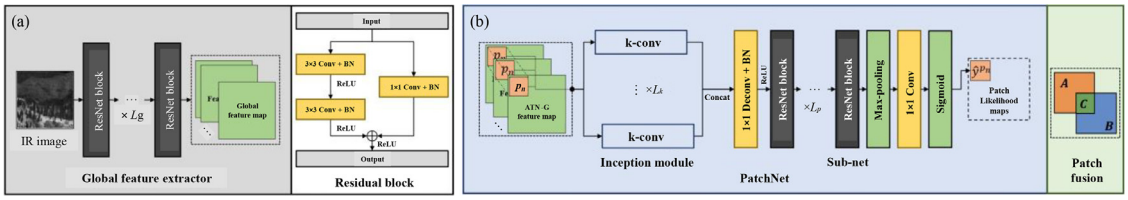

Fig. 11. Overall structure of LPNet based on global and local block attention fusion strategy. (a) Global feature extraction module, which mainly extracts features through the residual structure. (b) Structure of the PatchNet. ([68] (Figs. 3 and 5)). ATN-G: Attention-density map.

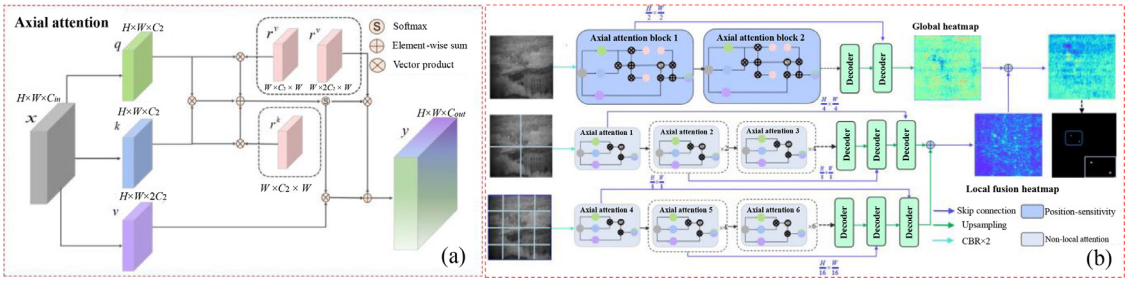

Fig. 12. (a) Axial-Attention module. (b) Overall structure of MPANet. ([69] (Figs. 1 and 2)).

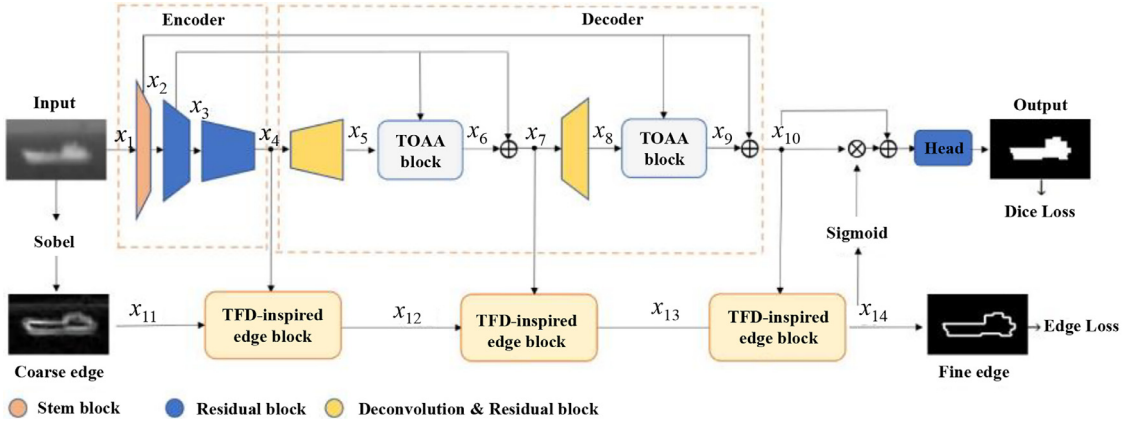

Fig. 13. Overall structure of ISNet, which has a U-Net structure with TOAA blocks and TFD-inspired edge blocks. ([54] (Fig. 1)).

ground. The TOAA block uses an attention mechanism to calculate low-level information in the row and column directions and fuses it with high-level information to capture target shape features and suppress noise for feature fusion. Finally, multiple supervision training sessions are conducted using dice loss and edge loss.

4.2.1.6. Generative adversarial strategy. Owing to the positive and negative samples of IR small targets being extremely unbalanced, Wang et al. [10] introduced a generative adversarial network (GAN) for IR small target detection to balance the problem of missing detections and false alarms; a proposed conditional generation adversarial network (MDvsFA) consisting of two generators and a discriminator effectively balanced this problem. Regarding the generators, G1 and G2 reduce the missed detection and false alarm rates, respectively, as shown in Fig. 14(a). Finally, the discriminator network is used in adversarial training for effectively balancing the problem, as shown in Fig. 14(b). In the inference phase, the average value of the results generated by the two generators is used as the final segmentation result.

Zhao et al. [70] reported that IR small targets can be considered a special type of noise that can be predicted from an input image according to the data distribution and hierarchical characteristics learned by a GAN. The network uses U-Net as the backbone to generate a false image containing only targets and distinguishes it from the real target image to improve the detection capability of the network, as shown in Fig. 15.

Zhou et al. [71] proposed an anticompetitive game framework (PixelGame). This framework includes the feature neutralization network (FNNet) and feature perturbation network (FPNet), which are controlled by different players whose goal is minimizing their utility functions. The two parameters are optimized through competition, as shown in Fig. 16.

Given the scarcity of IR small target public datasets, Kim et al. [72] proposed a method for generating IR small target data. They used the GAN framework to generate synthetic background images and IR small targets using two independent processes. In the first stage, an IR image is synthesized through the conversion of a visible image into an IR image. In the second stage, the target mask is implanted in the transformed image, as shown in Fig. 17.

4.2.1.7. Lightweight design strategy. Improving the detection accuracy and speed of IR small targets is key to the effectiveness of IR detection systems in satisfying the battlefield requirements of beyond-visual-range operations, pre-enemy detection, and timely warning. Therefore, designing an IR small target detection algorithm with high precision and speed that can be deployed in embedded devices is of great practical significance.

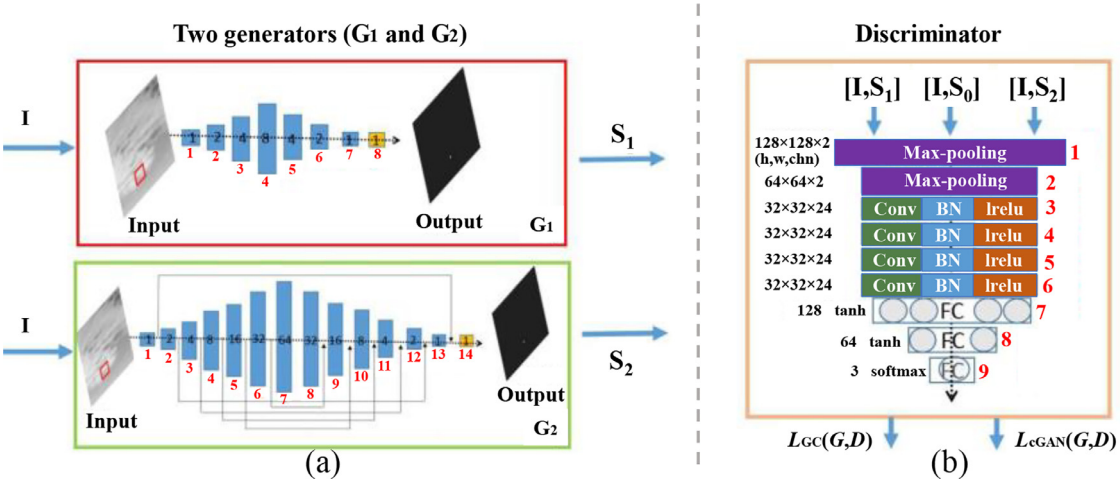

Fig. 14. Overall structure of MDvsFA. (a) Two generator networks G1, and G2. (b) Discriminator network ([10] (Fig. 2)).

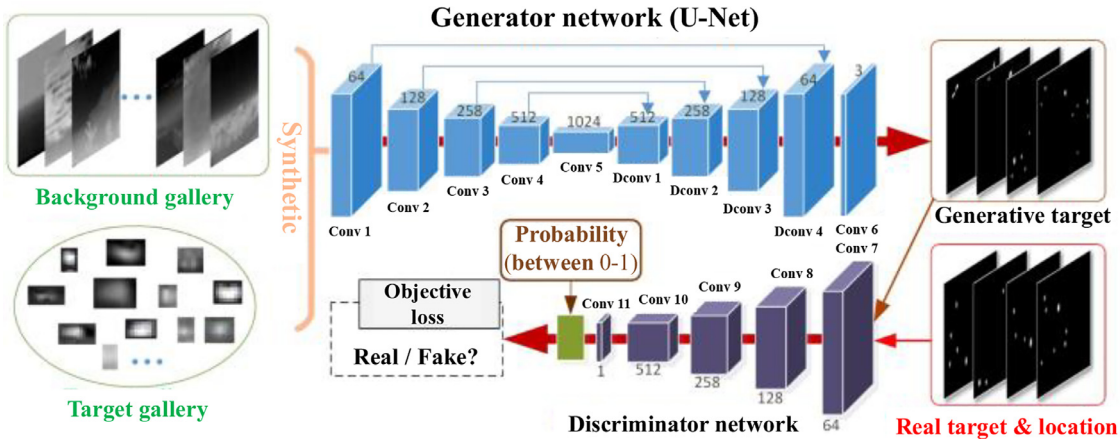

Fig. 15. Structure of generative adversarial network (GAN) ([70] (Fig. 2)).

Zhao et al. [73] proposed a novel lightweight joint loss function network includes target extraction loss, background suppression loss, and classification loss (TBCNet) to overcome the shortcomings of CNN in learning small target features. The TBCNet consists of a target extraction module (TEM) and a semantic constraint module (SCM), as shown in Fig. 18(a). The TEM uses the 2D convolutional layer and the MaxPooling layer to form the downsampling module, and uses the nearest neighbor interpolation and 2D convolutional layer to form the upsampling module. The SCM imposes a semantic constraint on TEM by combining the high-level classification task and solve the problem of the difficulty to learn features caused by class imbalance problem. Hu et al. [74] designed a lightweight small target network (STNet) with 13 convolutional layers that use residual connections in downsampling and upsampling, as shown in Fig. 18(b).

Kou et al. [75] believed that various multi-scale feature fusion modules and channel/spatial attention mechanism fusion modules did not significantly improve the segmentation accuracy of IR small targets but caused a sharp increase in parameters and floating-point operations (FLOPs) instead. Therefore, they designed a lightweight IR small target segmentation network model (LW-IRSTNet). This model compresses parameters to 0.16 M and FLOPs to 303 M by exploiting depth-separable convolution, hole convolution, and asymmetric convolution and simultaneously achieves a good detection accuracy. At the same time, Kou et al. [76], in combination with LW-IRSTNet, plan to design a general IR small target tracking algorithm for plug and play, as shown in Fig. 19.

4.2.1.8. Single point supervision strategy. The existing deep learning-based IR small target segmentation networks all rely on fully supervised training with pixel level annotations. However, pixel level label annotation requires a lot of labor costs, and the edge shape of infrared small targets is very fuzzy, making it difficult to achieve accurate annotation. To address this issue, Ying et al. [77] first proposed a novel framework for the problem of weakly supervised single-frame IR small target detection, dubbed label evolution with single point supervision (LESPS). Specifically, LESPS leverages the intermediate network predictions in the training phase to update the current labels, which serve as supervision until the next label update. Through iterative label update and network training, the network predictions can finally approximate the updated pseudo mask labels, and the network can be simultaneously trained to achieve pixel-level SIRST detection in an end-to-end manner. Li et al. [78] recovered the per-pixel mask of each target from the given single point label by using clustering approaches. They introduced randomness to the clustering process by adding noise to the input images, and then obtain much more reliable pseudo masks by averaging the clustered results. Thanks to this"Monte Carlo" clustering approach, Their method can accurately recover pseudo masks and thus turn arbitrary fully supervised SIRST detection networks into weakly supervised ones with only single point annotation.

In addition, considering that the proportion of IR small targets in images is small, if the down-sampling ratio becomes 32 times or higher, the advanced semantic information of small targets will

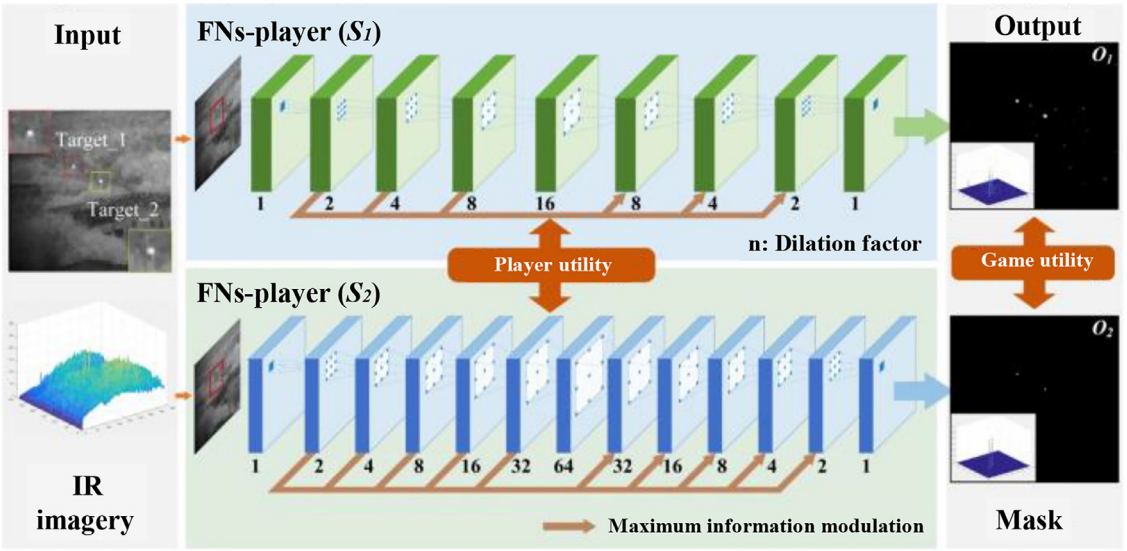

Fig. 16. Overall structure of PixelGame ([71] (Fig. 5)).

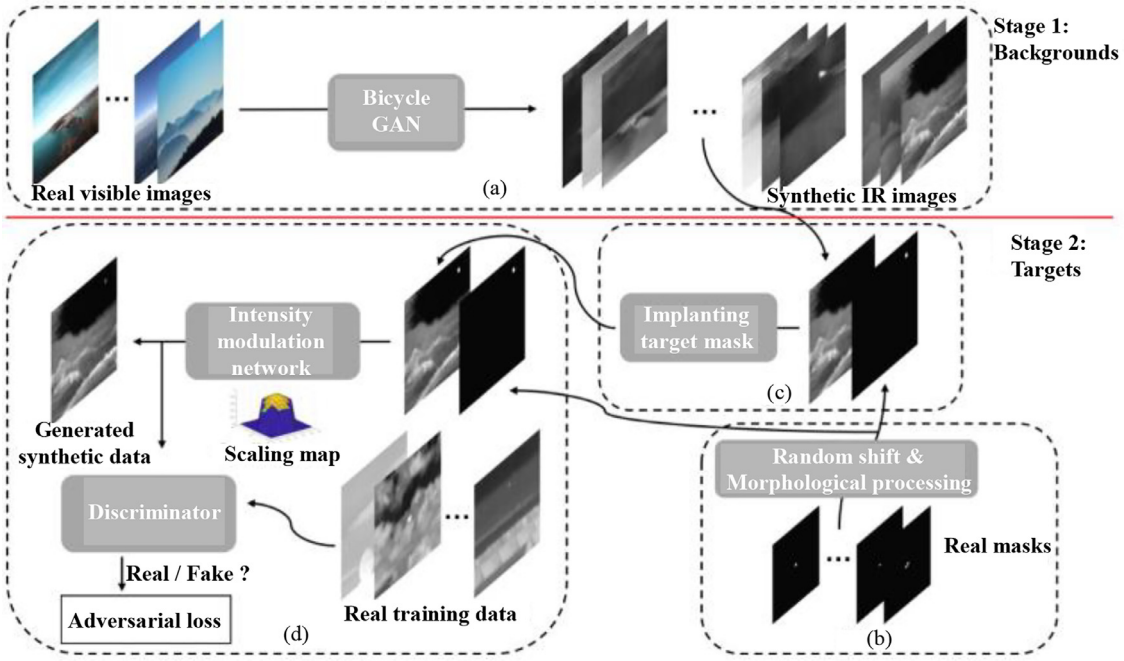

Fig. 17. Overall structure of the IR small target generation network ([72] (Fig. 1)).

be lost. Therefore, the down-sampling times of either 8 or 16 are adopted in the segmentation networks mentioned in the above strategies.

4.2.2. IR small target segmentation algorithm based on model-driven + CNN method

Considering that a IR small target may not be the brightest point in an image, a difference in local contrast is necessary. Hou et al. [79] proposed a robust IR small target detection network (RISTDnet). This network contains a feature extraction framework that combines handcrafted feature methods and CNNs. Five fixed-weight convolution kernels of weights 3×3, 5×5, 7×7, 9×9, and 11×11 were used to capture targets of different sizes in the original IR small target image. Subsequently, the encoder–decoder structure was used to segment the IR small target, as shown in Fig. 20(a). Based on the characteristics of the local contrast of IR small targets, Fan et al. [80] proposed a segmentation method for IR small targets that includes region proposal and employing a CNN. First, the intensity of the IR small target is enhanced using a traditional filtering method. Subsequently, potential target regions are found using corner detection. Finally, these regions are fed into a classifier based on a CNN to eliminate non-target regions and effectively suppress complex background clutter, as shown in Fig. 20(b).

Chuang et al. [81] proposed a new multi-scale local contrast learning network (MLCLNet). Local contrastive learning (LCL) is introduced into the network, as shown in Fig. 21(a). The LCL module adopts the concept of local contrast based on the model-driven approach. In network training, a combination of convolution and dilated convolution is used to learn the local contrast characteristics of IR small targets. First, an average filter is used to smooth each region block and obtain the average value. Then, the metrics between the target region and the eight adjacent regions are calculated. Finally, the calculation results and thresholds are used to

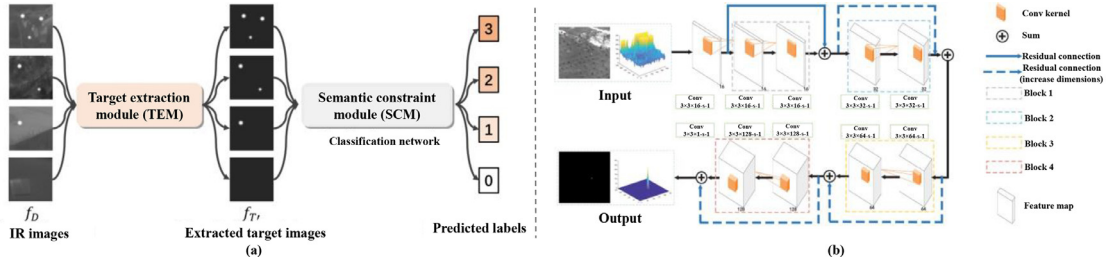

Fig. 18. (a) Overall structure of TBCNet ([73] (Fig. 3)). (b) Overall structure of STNet ([74] (Fig. 1)).

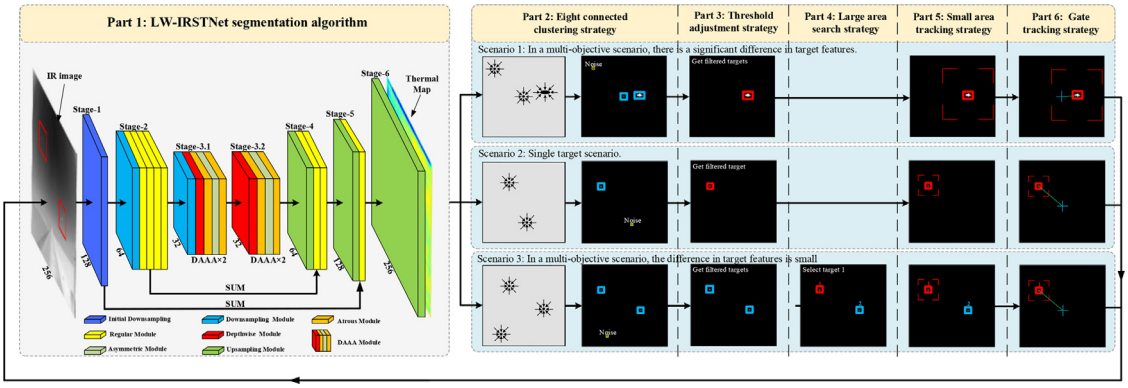

Fig. 19. Overall structure of LW-IRSTNet and IR Small Target Tracking Algorithm ([76] (Fig. 1)).

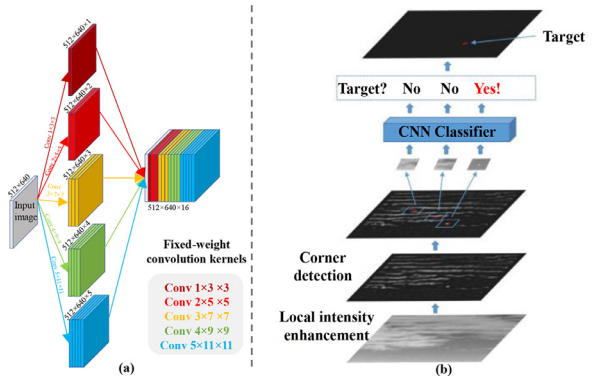

Fig. 20. (a) Feature extraction module with fixed weights in robust IR small target detection network (RISTDnet). ([79] (Fig. 1)). (b) Structure of region suggestion and CNN ([80] (Fig. 1)).

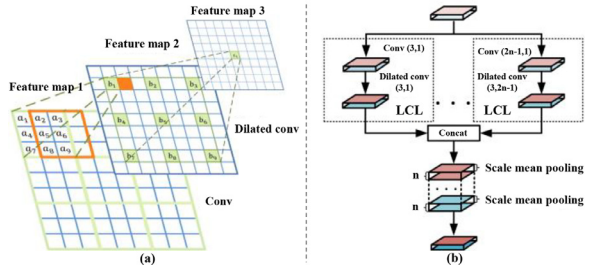

Fig. 21. (a) Structure of the local contrast learning (LCL) module (b) Structure of the multi-scale local constrast learning (MLCL) module. ([81] (Figs. 2 and 3)).

determine whether there is a target in the area. This method can reduce dependence on dataset samples, thereby improving the detection accuracy of datasets with fewer samples. Additionally, considering the difference in target size, an MLCL module was constructed based on the LCL module. The feature information of IR small targets can be fully mined by fusing the local contrast information at different scales, as shown in Fig. 21(b).

4.2.3. IR small target segmentation algorithm based on CNN + transformer method

Dosovitskiy et al. [82] proposed a Vision Transformer (ViT). Through this, they proved for the first time that a transformer could completely replace a CNN and be directly applied to the classification and prediction of image block sequences. The overall structure of a ViT is illustrated in Fig. 22(a). Subsequently, the Swine Transformer proposed by Liu et al. [83] achieved a remarkable performance in semantic and instance segmentation tasks. It employs an image block (window) division scheme that differs from that of ViT, as shown in Fig. 22(b). The self-attention for each window is calculated, and a new window is generated by moving the window partition with cyclic and reverse cyclic displacements. Owing to the satisfactory detection and segmentation results achieved using the Swine Transformer, the transformer structure has been introduced in medical image processing, industrial detection, and other fields.

Liu et al. [84] were the first to use a transformer for IR small target segmentation. They believed that existing deep-learning-based methods were limited by the locality of CNNs, weakening their ability to capture large-range dependencies. Additionally, IR targets usually go undetected by detection models owing to their weak appearances. Therefore, we propose an IR small target segmentation method based on the CNN + Transformer method. The self-attention mechanism of a transformer was adopted to learn the interactive information of image features over a larger range. Furthermore, a feature enhancement module was designed to learn the discriminative features of small targets to avoid missing detections. The network structure is illustrated in Fig. 23; the network contains three parts: a feature-embedding module, a compound encoder with k encoder layers, and a decoder. The feature-embedding module is used to obtain a compact feature representation. In the compound encoder, each encoder layer has two main parts: a multi-head self-attention (MSA) mechanism and a feature

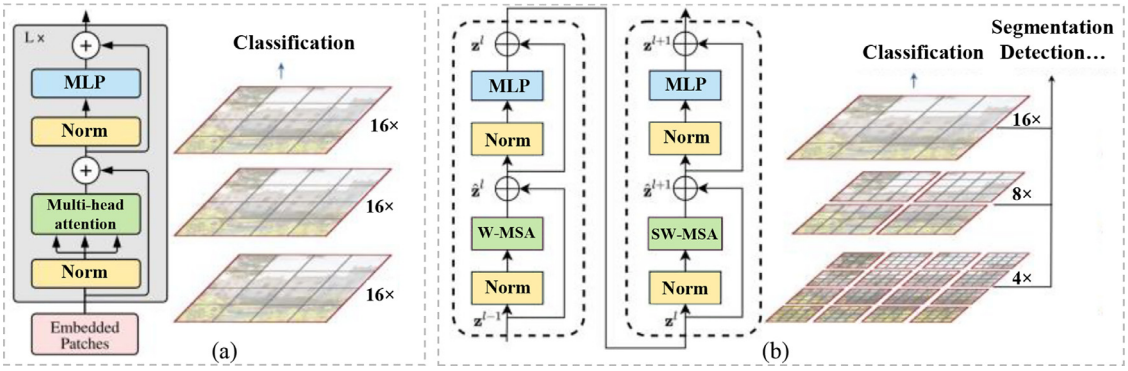

Fig. 22. Structures of ViT Vs Swin Transformer ([82] (Fig. 1), ([83] (Figs. 1 and 3)). MLP is multi-layer perceptron. W-MSA and SW-MSA are multi-head self-attention modules with regular and shifted windowing configurations, respectively.

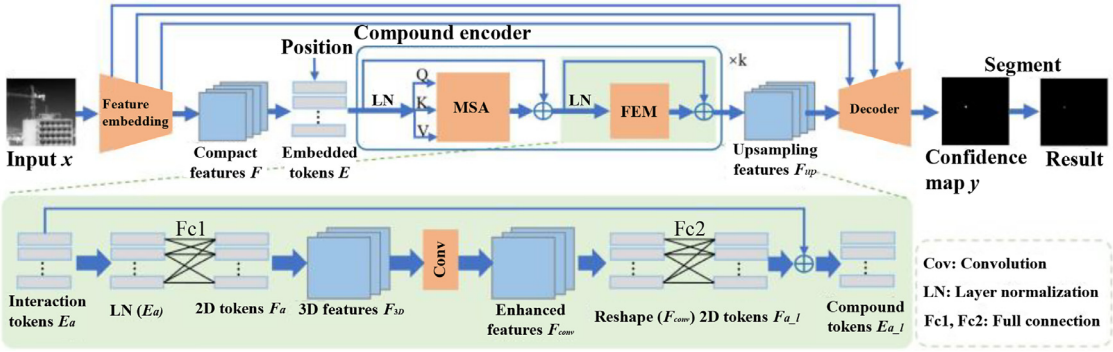

Fig. 23. Overall structure of Convolution + Transformer method ([84] (Fig. 2)).

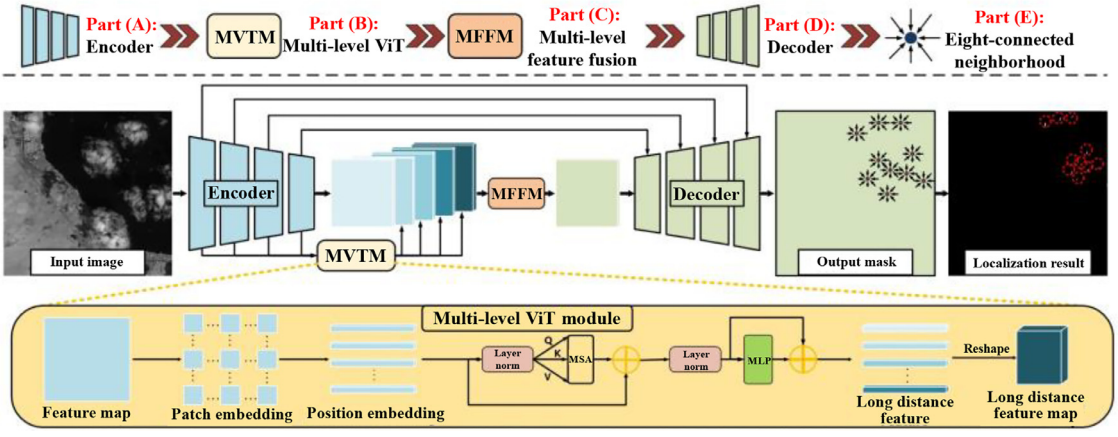

Fig. 24. Overall structure of MTU-Net ([85] (Fig. 3)).

enhancement module (FEM); MSA is used to learn the interaction information between all the embedded tokens, and the FEM can learn the more inherent features of small targets with a higher resolution. Using the decoder, a confidence map of the small target can be obtained. Finally, the detection results are obtained using adaptive threshold segmentation.

Wu et al. [85] proposed the multi-level TransUNet (MTU-Net), which uses a hybrid ViT encoder and CNN to extract multi-level features. The local feature map is extracted using several convolutional layers and then fed to the multi-level feature extraction module (MVTM) to capture long-distance correlations, as shown in Fig. 24. The MTUNet contains five parts: an encoder, a multi-level ViT module (MVTM), a multi-level feature fusion module (MFFM), a decoder, and eight connected neighborhood clustering modules. First, the input image is fed into the CNN encoder to coarsely extract multi-scale features. Then, features of different levels are extracted from the long-distance features using an MVTM, and multi-level features are fed into the MFFM, where they are concatenated and fused to incorporate long-distance information. Following this, features with multi-level long-distance information are fed into a U-shaped decoder and fused at the nodes of the skip connection to generate the final predicted probability map. Finally, the predicted probability map is clustered, and the centroid of each target region is determined.

Wang et al. [86] proposed a coarse-to-fine internal attention-aware network (IAANet) comprising a region proposal network (RPN), semantic generator (SG), and AE for small target IR detection. First, the RPN proposes coarse local regions, and the SG generates a semantic feature map. Then, the transformer models the attention between the pixels in the coarse target regions, outputting

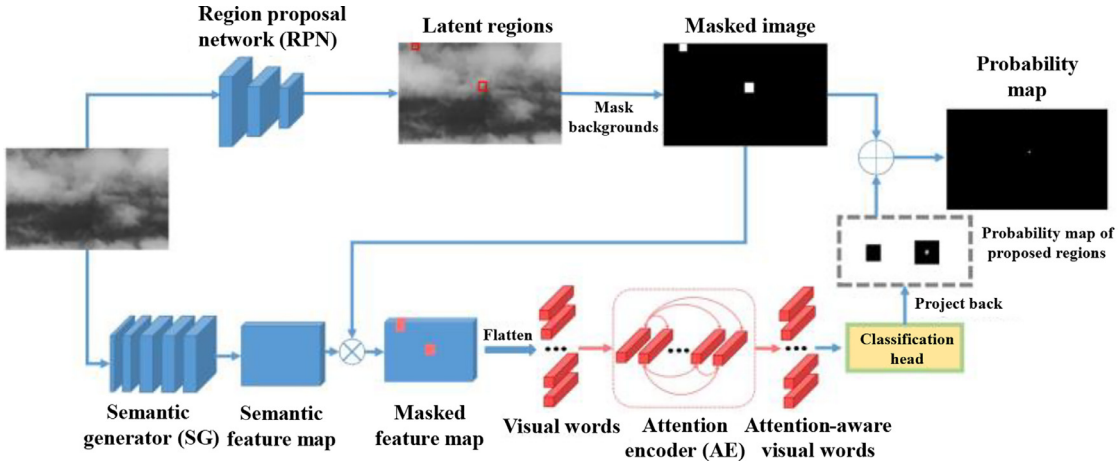

Fig. 25. Overall structure of IAANet. ([86] (Fig. 2)).

attention-aware features. Finally, predictions are made by feeding attention-aware features to the classification head. The network structure of the IAANet is illustrated in Fig. 25.

4.3. Multi-frame IR small target segmentation and super-resolution enhancement networks

Although single-frame IR small target detection has a low computational complexity and provides a good real-time performance, it does not consider the relationship between frames, resulting in missed detections or false alarms in complex environments. Therefore, some scholars have considered the relationship between image frames and accordingly proposed segmentation algorithms for IR small targets in image sequences based on a deep-learning framework.

Liu et al. [87] proposed an IR video sequence encoding and decoding model based on a Bidirectional Convolutional Long Short-Term Memory structure (BiConvLSTM) and 3D convolutional structure (3D-Conv), solving the problems of high similarity and dynamic changes in parameters. The overall structure of this model is illustrated in Fig. 26(a). To solve the problem of these dynamic changes, the BiConvLSTM structure learns the motion model of the target, as shown in Fig. 26(b), and to solve the problem of low local contrast, 3D-Conv expands the receptive field in the time dimension, as shown in Fig. 26(c). The BiConvLSTM structure is first used to process video sequence inputs. Then, the outputs at each time point are concatenated as the input of the 3D-Conv structure. The concatenated data are inputted to the concatenate matrix (CM). After the convolution of the 3D-Conv structure pooling operation, the CM is squeezed into a high-level feature vector that contains the location of the object and its probability. Then, the decoding part decodes the target information through a two-channel fully-connected structure to obtain the target position and probability.

Considering that the resolution of IR small target images is low, and the local contrast of targets are weak, Ying et al. [88] proposed an IR small target super-resolution enhancement network based on image sequences called motion-contrast prior driven network (MoCoPnet), as shown in Fig. 27(a). Specifically, a low-resolution (LR)

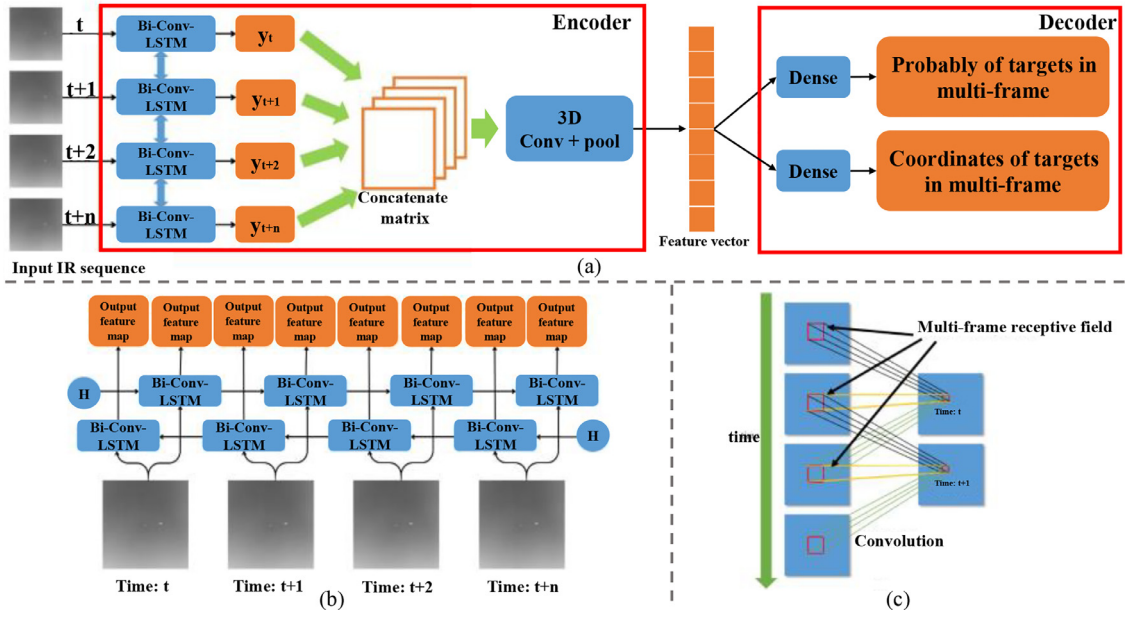

Fig. 26. (a) Overall structure of Multi-Frame Sequence detection. (b) Structure of BiConvLSTM block (c) Diagram of 3D Convolutional receptive field structure ([87] (Figs. 2-4)).

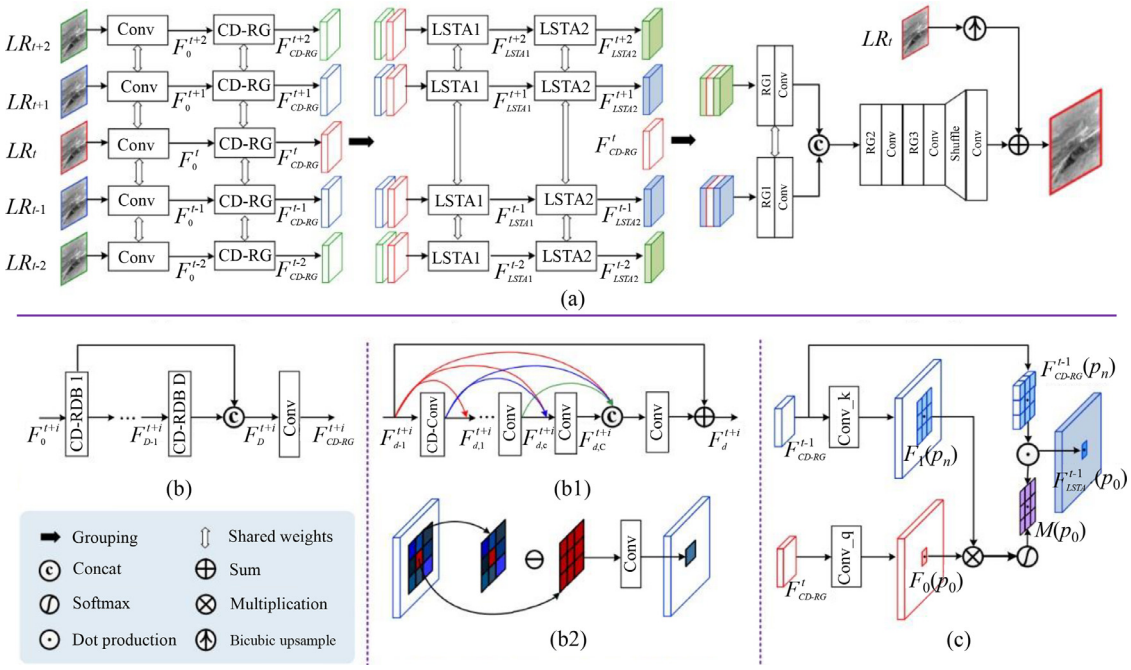

Fig. 27. (a) Overall structure of MoCoPnet. (b) represents the central difference residual group (CD-RG); and (b1) and (b2) represent its sub-modules central difference dense block (CD-RDB) and central difference convolution (CD-Conv), respectively. (c) represents the local spatio-temporal attention (LSTA) module with kernel size 3 and dilation rate 1. ([88] (Fig. 1)).

image sequence with 5 frames LR_{t+i} is first sent to a convolutional layer to generate the initial features (F_0^{t+i}), which are then sent to the central difference residual group (CD-RG) to generate features (F_{CD-RG}^{t+i}), as shown in Fig. 27(b). Then, the F_{CD-RG}^{t+i} is paired with the reference feature (F_{CD-RG}^{t}) and sent to two local spatiotemporal attention (LSTA) modules, as shown in Fig. 27(c). Next, the F_{CD-RG}^{t} is concatenated with compensated neighborhood frames (F_{LSTA2}^{t+i}) and then sent to a residual group (RG) and a convolution layer for coarse fusion. Afterwards, the two fused features are concatenated and sent to an RG and a convolution for fine fusion. Finally, the super-resolution (SR) image is obtained by adding the LR image.

5. Loss function

During the training process of semantic segmentation networks, the difference between the prediction results and labels needs to be calculated. The gradient descent method can then be used to minimize these differences to further study the associated rules and optimization directions. A loss function $L(\theta, d) \geq 0$ is usually used to represent the error between the true value θ and predicted value d. The more accurate the prediction, the smaller the loss value. Considering that an IR small target segmentation task is a binary task with extremely unbalanced positive and negative samples, commonly used loss functions such as Mean Square Error (MSE) loss, Mean Absolute Error (MAE) loss, and cross-entropy loss are not suitable for IR small target segmentation tasks. Considering the characteristics of IR small targets, the following loss functions can be used for training.

5.1. Binary cross-entropy loss

Binary cross-entropy loss (BCEloss) is applied to the binary task of positive and negative sample equalization; it is represented as:

$$L = -(y_t * \log(p_t) + (1 - y_t) * \log(1 - p_t)). \tag{4}$$

where p_t is the probability of the label y_t.

When the number of foreground pixels is considerably smaller than the number of background pixels, the training results are biased towards the background pixels, resulting in a poor effect. Huang et al. [66] used BCEloss to train IR small target segmentation tasks. However, the positive and negative samples in the IR small target segmentation task were extremely unbalanced, resulting in the training convergence being slow, and the training results not being good.

If a weight ω_t is added to Eq. (4), the problem of unbalanced positive and negative samples is solved, as follows:

$$L = -\omega_t(y_t * \log(p_t) + (1 - y_t) * \log(1 - p_t)). \tag{5}$$

where weight ω_t should be set depending on empirical judgments made for different tasks.

5.2. Focal loss

Lin et al. primarily proposed focal loss [89] to solve the balance problem between difficult and easy samples and between positive and negative samples. This function is represented as:

$$FL = -\alpha_t(1 - p_t)^{\gamma} \log(p_t). \tag{6}$$

where α_t and γ are weight factors that are used to solve the balance problem between difficult and easy samples and between positive and negative samples. This function has been used for training in many studies on IR small target segmentation tasks.

5.3. Dice loss and IoU loss

Diss loss is derived from the dice coefficient, which measures the similarity between prediction results and labels, and it is applied to binary segmentation of images with unbalanced positive and negative samples. It can be represented as:

$$Dice_{loss} = 1 - \frac{2|X \cap Y| + smooth}{|X||Y| + smooth}. \tag{7}$$

where X is the real target label pixel, and Y is the predicted target pixel. In practical applications, to prevent the denominator from

being zero, a value of typically 1 is often added to both the numerator and the denominator for smoothing.

The Intersection over Union (IoU) loss, like the dice loss, is also used to describe regional correlation; it is represented as follows:

$$IoU_{loss} = 1 - \frac{|X \cap Y| + smooth}{|X||Y| + smooth}. \tag{8}$$

Since IR small target segmentation tasks closely resembles medical image segmentation tasks, many studies [51,53,84] on IR small target segmentation tasks have adopted the dice loss or IoU loss.

5.4. Tversky loss

The Tversky loss is a generalized form of the dice loss, IoU loss, or Jaccard loss; it is represented as follows:

$$\text{Tversky}_{loss} = 1 - \frac{|X \cap Y|}{|X| \cap |Y| + \alpha|X - Y| + \beta|Y - X|}. \tag{9}$$

where α and β are adjustment coefficients. When α and β are 0.5, Eq. (9) represents the dice loss or IoU loss and when they are 1, it represents the Jaccard loss.

The Tversky loss is used for medical focus segmentation [90] and produces different results depending on the values of α and β. Accordingly, different α and β parameters can be used in the IR small target segmentation task to improve the segmentation accuracy.

5.5. Combined application of different loss functions

Zhang et al. [54] used a "BCE loss + dice loss" function in the ISNet network to achieve better shape prediction results for IR small targets. Chen et al. [67] used the "focal loss + dice loss" function for key point and pixel level prediction in the dual task of IR small target detection and segmentation. Ju et al. [91] used the "BCE loss + IoU loss" function to train an efficient IR small target detection network and achieved good results. In summary, for IR small target segmentation tasks, the effect of the imbalance between positive and negative samples should be considered when designing loss functions.

6. Evaluation indices

6.1. Accuracy indices

Accuracy indices, which are the most basic indices used to evaluate semantic segmentation networks, represent the degree of conformity between segmentation results and labels.

6.1.1. Precision, recall, and F1 score

The precision rate, recall rate, and F1 score are indices commonly used to measure the accuracy of models in binary tasks and can be defined using a confusion matrix, as shown in Fig. 28.

The precision rate represents the proportion of true positives (TP) in the predicted positive samples, as shown in Eq. (10). Since the precision rate is likely to reach 100% when the threshold value of a classification model is high, the precision rate is a one-sided evaluation index.

$$Precision = \frac{TP}{TP + FP}. \tag{10}$$

where TP is the true positive, and FP is the false positive.

The recall rate represents the ratio between the predicted true positives and the total number of labeled positive samples, as shown in Eq. (11). The recall rate is a measure of the ability of a model to classify positive samples; this rate it called the sensitivity

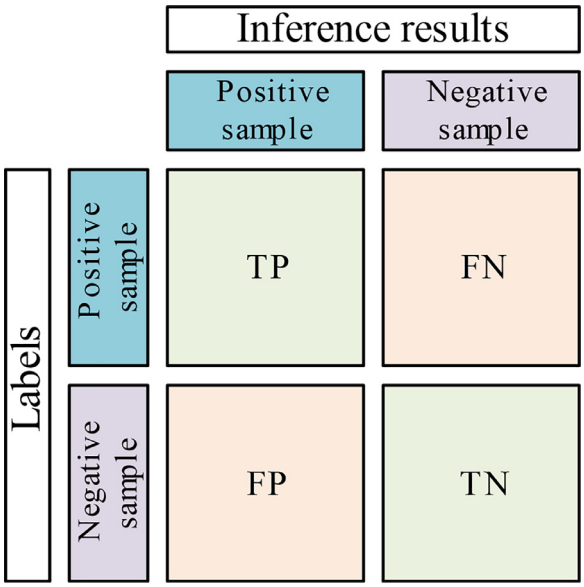

Fig. 28. Binary confusion matrix.

of the model or the true positive rate (TPR); the false-positive rate (FPR) is shown in Eq. (12). When the threshold value of the classification model is low, the recall rate may be 100%, making this evaluation indicator one-sided.

$$Recall = TPR = \frac{TP}{TP + FN}. \tag{11}$$

$$FPR = \frac{FP}{TN + FP}. \tag{12}$$

Given the shortcomings of the precision and recall rates, the F1 score was proposed to balance the two indices; it is represented in Eq. (13):

$$F_{1\ Score} = \frac{2 * Precision * Recall}{Precision + Recall}. \tag{13}$$

Eq. (13) shows that the F1 score will only be high when both the precision and recall rates are high. Therefore, the higher the F1 score, the more effective the model.

6.1.2. IoU and mIoU

Intersection over Union (IoU), also known as the Jaccard Similarity Coefficient (JSC), is the most commonly used index for semantic segmentation. In semantic segmentation, the IoU represents the overlap rate of the prediction mask and label pixels. The mean IoU (mIoU) is the arithmetic mean of the IoU values in each category and is used for the pixel overlap of the overall dataset; the IoU and mIoU are represented as follows:

$$IoU = \frac{|A \cap B|}{|AB|} = \frac{\sum_{i=0}^{k} p_{ii}}{\sum_{i=0}^{k}(\sum_{j=0}^{k} p_{ji} + \sum_{j=0}^{k} p_{ij} - p_{ii})}. \tag{14}$$

$$mIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}}. \tag{15}$$

where A and B denote the label and prediction segmentation areas, respectively; and p_{ji} is the number of pixels in class j predicted as class i. The IoU ranges from 0 to 1. The best result is obtained when the prediction and label areas overlap completely; here, the IoU value is 1.

6.1.3. PR curve, ROC curve, and AUC area

The PR curve represents the relationship between the precision and recall. This curve is sensitive to the sample proportion. The closer the curve to the upper-right corner, the better the classification performance of the algorithm. The receiver operating characteristic (ROC) curve represents the relationship between FPN and TPR. Thus curve is insensitive to the sample proportion; the closer it is to the upper-left corner, the better the algorithm performance. The Area Under the ROC Curve (AUC) represents the proportion of the total area under the ROC curve; the larger this area, the better the classification performance of the algorithm.

6.2. Calculation complexity indexes

6.2.1. Execution efficiency

Execution efficiency is an important index for measuring the performance of an algorithm. This index is evaluated using the FLOPs or running time of a code. Regarding FLOPs, they can estimate the occupancy of computing resources in a segmentation network, representing the computational cost of forward propagation in a CNN.

6.2.2. Memory usage

Memory usage represents the amount of memory space consumed during algorithm execution. In the deep-learning process, there are often many parameters; e.g., there are more than ten million parameters in the VGG16 network. Parameters represent the weight parameter ω and the offset parameter b that need to be learned during network training. In summary, the FLOPs and number of parameters represent the complexity of a network; the larger these values, the higher the network complexity.

6.2.3. Latency and FPS

The latency or frames per second (FPS), is the most direct index of an algorithm's performance. Latency represents the time required by a network to predict an image, excluding the time required for postprocessing while the FPS represents the number of frames a network can process per second; FPS = 1/latency.

6.2.4. Robustness index

Robustness measures the ability of a model to maintain stable performance indices under abnormal conditions, such as changes in data or algorithm parameters. To experimentally verify the robustness of different algorithms for IR small target segmentation tasks, we selected four different public datasets and fusion datasets, as described in Section 7. If the accuracy indices of an algorithm are high for different datasets, the robustness of that algorithm is relatively good.

7. Experimental analysis

7.1. Basic parameters

7.1.1. Training details

An Ubuntu 18.04 system with a memory of 32 GB was used. The central processing unit (CPU) was Intel® Xeon(R) E5–2630 v3@ 2.40 GHz x 32. An NVIDIA GeForce RTX 2080 Ti graphical processing unit (GPU) was used. The batch size was 64, the number of epochs of datasets 1 and 2 and fusion datasets 1–4 was 100, and the number of epochs for datasets 3 and 4 was 300. The optimizer was stochastic gradient descent (SGD), the momentum was 0.9, and the weight decay was 1e-4. The initial learning rate was set to 0.05, the Poly learning rate strategy was used, and the SoftIoULoss function was used [51].

Table 5
Computational complexity comparison analysis.

Year, Algorithms	FLOPs(M)	Params(M)	FPS
Classic segmentation networks			
2015, FCN [18]	22,313.4351	20.0978	5.6057
2015, UNet [20]	65,475.4447	34.5251	2.2231
2016, FusionNet [21]	123,722.0065	81.6691	1.4374
2017, SegNet [22]	40,013.6602	29.4277	2.6633
2017, GCN [23]	15,155.8881	58.1438	16.7688
2018, DeeplabV3+ [26]	20,755.5758	54.6075	2.0119
2018, Exfuse [16]	54,519.1878	126.4514	0.8159
2018, DFN [60]	10,274.8524	42.5063	2.4815
Lightweight segmentation networks			
2016, ENet [30]	535.1505	0.3491	12.2438
2017, LinkNet [31]	3020.65 025	11.5334	3.3754
2018, BiSeNet [38]	10,176.4395	23.0635	20.5268
2019, DFA [39]	541.7969	2.3662	3.6255
The networks designed for IR small targets			
2019, MDvsFA [10]	247,111.6145	3.7683	0.6416
2021, ACM-UNet [52]	503.7925	0.5198	36.0439
2021, LSPM [66]	61,706.9384	31.1408	0.0728
2022, DNANet [53]	14,279.0852	4.6968	2.7881
2023, AGPCNet [51]	43,179.4098	12.3520	2.6315
2023, LW-IRST [75]	303.5300	0.1632	30.0332

7.1.2. Comparison of baselines

Classic segmentation networks include FCN, U-Net, FusionNet, SegNet, GCN, DeeplabV3+, Exfuse, and DFN; lightweight segmentation networks include ENet, LinkNet, BiSeNet, and DFA; and networks designed for IR small targets include MDvsFA, ACM-UNet, LSPM, DNANet, AGPCNet, and LW-IRSTNet.

7.2. Comparison experiment

7.2.1. Quantitative comparative analysis

7.2.1.1. Computational complexity comparison analysis. In military applications, the real-time performance of IR small target detection algorithms is crucial. To compare the computational complexities of the different algorithms, FLOPs, params, and FPS were used for evaluation, as shown in Table 5 ([75] Table 3).

As shown in Table 5, the FLOPs and params of classical segmentation networks are relatively large, owing to the increased focus of these networks on segmentation accuracy over real-time. Although the FLOPs and params of lightweight networks are significantly compressed, the segmentation accuracy is inevitably affected. In segmentation networks specially designed for IR small target detection, some scholars have focused on improving segmentation accuracy while ignoring real-time issues, such as AGPCNet and DNANet. We proposed a network called LW-IRSTNet [75], mainly to solve the problem of the mobile terminal deployment of IR small target detection algorithms. The FLOPs and params of the proposed LW-IRSTNet were only 303 M and 0.16 M respectively, indicating that this network is lighter than other comparison baselines. Meanwhile, the accuracy of the LW-IRSTNet algorithm ranked second compared to other comparison baselines, as shown in Table 6.

7.2.1.2. Comparative analysis of algorithm accuracy and robustness. To verify the accuracy and robustness of the algorithm, we compared 18 segmentation networks on five datasets, as listed in Table 6 ([75] Table 3). As shown in Table 6, the LW-IRSTNet, AGPCNet, and DNANet exhibited the best detection effects for IR small target detection tasks. The classic segmentation networks FCN and U-Net and the lightweight segmentation network ENet also had good detection results for IR small target detection tasks. Regarding the five networks SegNet, GCN, Exfuse, DFN, and LinkNet, after 100 epochs of training, the loss function remained large, and there was no trend of convergence; therefore, we did not continue training.

Table 6
Comparative analysis of segmentation accuracy of different algorithms.

Algorithms	Date set 1		Date set 2		Date set 3		Date set 4		Fusion datasets 1–4	
	mIOU	F1	mIOU	F1	mIOU	F1	mIOU	F1	mIOU	F1
Classic segmentation networks										
FCN	0.3881	0.5592	0.7199	0.8372	0.5952	0.7463	0.6265	0.7704	0.6265	0.7704
U-Net	0.4243	0.5958	0.7059	0.8276	0.7092	0.8298	0.8917	0.9428	0.6471	0.7858
FusionNet	0.4018	0.5732	0.7019	0.8249	0.7200	0.8372	0.9072	0.9513	0.5981	0.7485
SegNet	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
GCN	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
DeeplabV3+	0.4075	0.5790	0.7255	0.8409	0.7278	0.8424	0.7087	0.8295	0.5615	0.7192
Exfuse	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
DFN	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Lightweight segmentation networks										
ENet	0.4505	0.6212	0.7136	0.8329	0.6936	0.8191	0.6578	0.7936	0.5913	0.7431
linkNet	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
BiSeNet	0.4437	0.6146	0.613	0.7601	0.5748	0.7300	0.4423	0.6133	0.5335	0.6958
DFA	0.4506	0.6212	0.6620	0.7966	0.5847	0.738	0.4796	0.6483	0.5230	0.6868
The networks designed for IR small targets										
MDvsFA	0.4540	0.5558	0.6278	0.7784	0.4117	0.6977	0.6437	0.7641	0.5119	0.6872
ACM-U-Net	0.4299	0.6013	0.6586	0.7942	0.6335	0.7757	0.6680	0.801	0.5334	0.6957
LSPM	0.4017	0.5732	0.6510	0.7886	0.6151	0.7617	0.6331	0.7753	0.5334	0.6957
DNANet	0.4036	0.5751	0.6811	0.8103	0.7529	0.8590	0.9374	0.9677	0.6542	0.7909
AGPCNet	0.4674	0.6370	0.7328	0.8427	0.7318	0.8451	0.8951	0.9447	0.6727	0.8043
LW-IRSTNet	0.4661	0.6359	0.7334	0.8462	0.7368	0.8485	0.7529	0.8590	0.6638	0.7979

Note: A value of 0.0000 indicates that the training did not converge after 100 epochs.

Table 7
Complex scenario test set.

No.	Scenario description		Number of targets	Target local contrast
	Background	Interference source		
1	Sky	Cumulus congestus	1	Very weak
2		Cumulus fractus	1	Very weak
3		Altocumulus	1	Very weak
4	Ground	Buildings and trees	1	Strong
5		Ground clutter	1	Strong
6		Jungle clutter	1	Very weak
7	Sky	Clean sky	2	Strong
8		Thin clouds and electric wires	2	Very weak

7.2.2. Qualitative comparative analysis

To compare and analyze the segmentation effects of 13 different algorithms (excluding five networks without convergence) on IR small targets in different complex air-ground scenes, we selected eight complex scenes (Table 7) for testing; the final visual detection effect is shown in Fig. 29. The network parameters of the 13 algorithms listed in Table 6 were trained using the fusion datasets 1–4. In Fig. 29, the red box indicates a correct detection, the yellow box indicates a missing detection, and the blue box indicates a false alarm.

In Fig. 29(a)–(c), under the sky background, both the energy of the IR small targets and local contrast are very low, resulting in small targets being submerged in cumulus congestus, cumulus fractus, or altocumulus clouds. The experimental results showed that only the LW-IRSTNet and DNA could effectively detect weak and IR small targets without false alarms or missed detections. In Fig. 29(d) and (e), under the ground background, the IR target has very strong energy and high local contrast. These results proved that all the 13 algorithms could detect IR small targets. However, because of the influence of jungles, buildings, vehicles, and other interference sources on the ground background, most algorithms generated many false alarms. In Fig. 29(f), under the ground background, both the energy of the IR small target and the local contrast are very low, and the target is submerged in the jungle background. In this scenario, only the LW-IRSTNet can effectively detect weak and IR small targets without false alarms or missed detections. The other comparison algorithms had false alarms or missed detections. In Fig. 29(g), under a clean-sky background, although there are two targets, the local contrast of these two targets is high, and there is no clutter interference. Therefore, the detection effects of the 13 algorithms were superior here. In Fig. 29(h), under the interference of thin clouds and electric wires, the detection effect of the 13 comparison algorithms on multiple targets is less than that in Fig. 29(g).

We inferred the following conclusions from analyzing Fig. 29(a)–(f). 1) The detection effect of the sky background is better than that of the ground background. 2) The higher the target energy or local contrast, the better the detection effect. 3) The DNANet, AGPCNet, and LW-IRSTNet algorithms designed for IR small targets had a better detection effect than the other algorithms did in different complex backgrounds. 4) Our proposed LW-IRSTNet achieved good results in terms of detection accuracy, robustness, and computational complexity (from Tables 5, 6 and Fig. 29).

8. Discussion

Recently, IR small target detection technology has made significant progress; however, there is still room for improvement. We have reconsidered the task of IR small target detection and have discussed the existing problems and future trends of IR small target detection technology from four levels: motivation, algorithm, application, and follow-up tasks, as shown in Fig. 30.

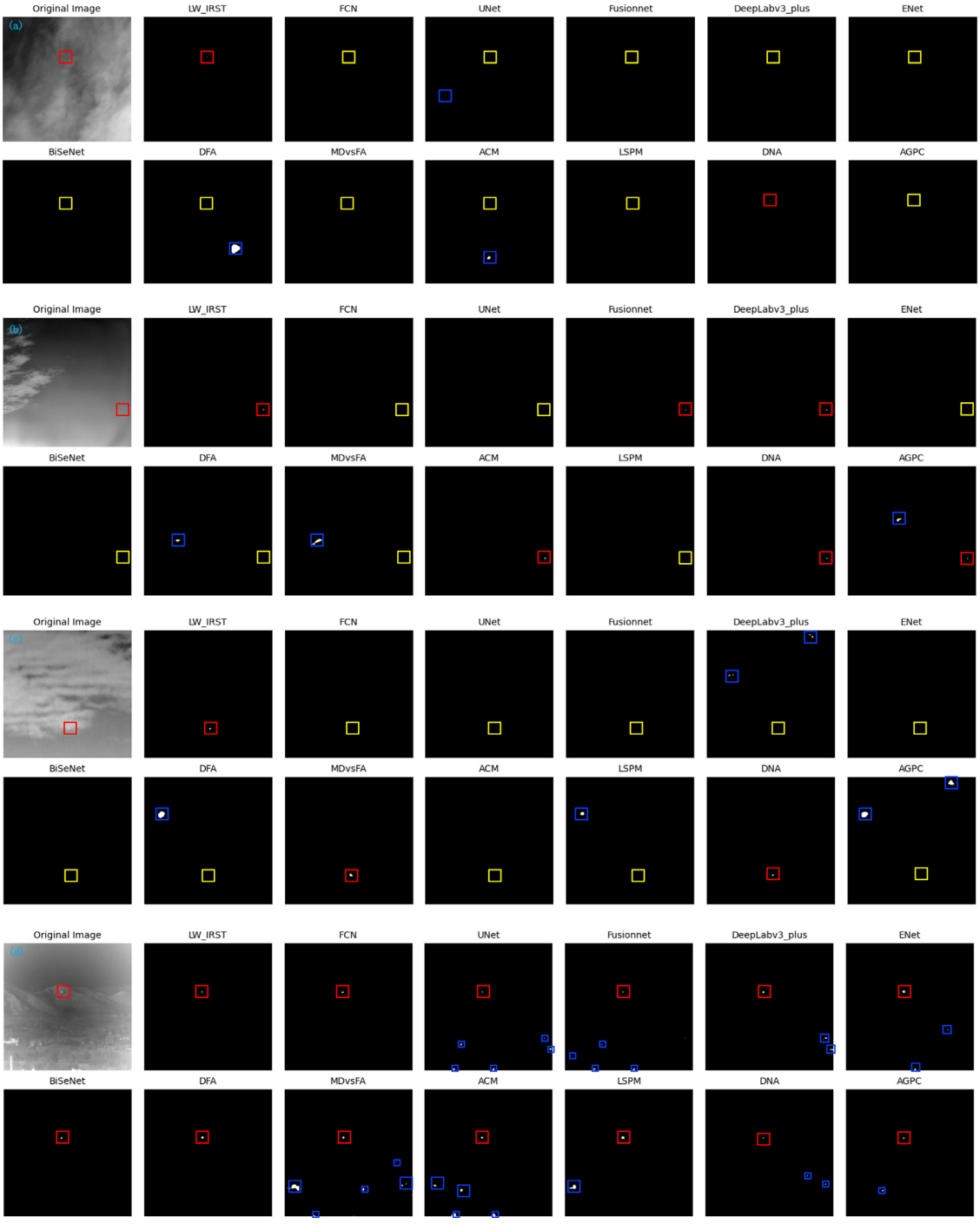

Fig. 29. IR small target segmentation results. (a) Cumulus congestus interference. (b) Cumulus fractus interference. (c) Altocumulus interference. (d) Buildings and trees interference. (e) Ground clutter interference. (f) Jungle clutter interference. (g) Clean sky background. (h) Thin clouds and electric wires interference.

8.1. Existing problems

8.1.1. Problems at the top design level

(1) *Research motivation unclear*. The types of targets and backgrounds in existing public datasets are not sufficiently specific and the definition of targets is not sufficiently clear, resulting in unclear task traction. Therefore, several studies were blinded.

(2) *Low research starting point*. With the support of existing public literature, datasets, and codes, the research threshold of IR small target detection technology is relatively low, and the range of researchers is relatively wide, resulting in an uneven quality of documents.

8.1.2. Problems at algorithm level

(1) *The datasets are not sufficiently rich*. Data are the basis of research on deep learning algorithms, while IR small target public datasets have low resolution, small quantity, and inaccurate mask labels that seriously affect the accuracy of segmentation algorithms.

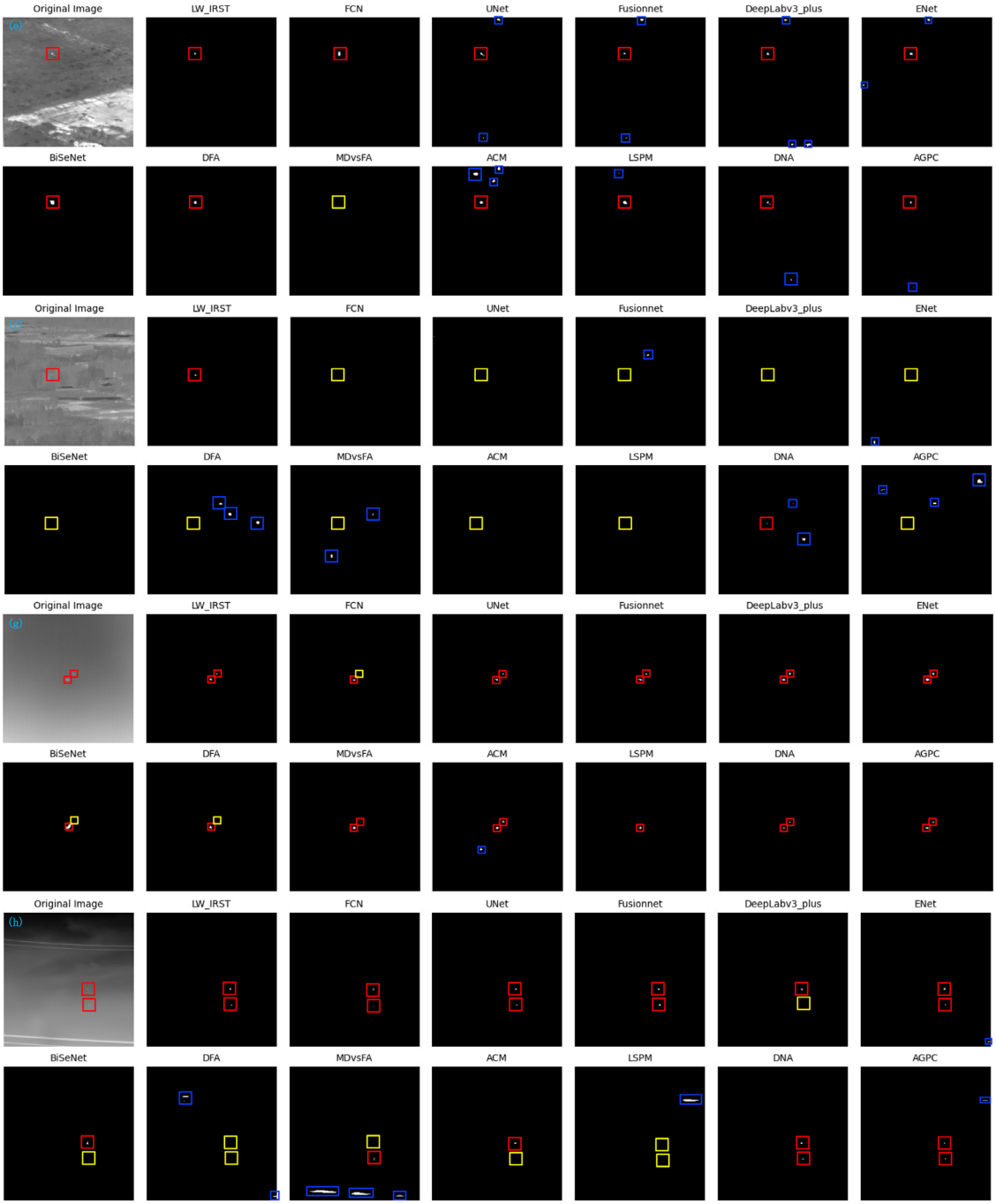

Fig. 29. Continued

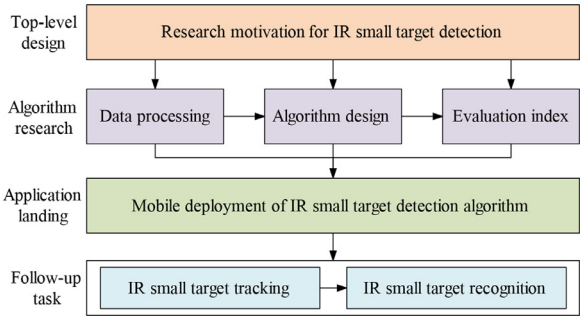

Fig. 30. Rethinking the problems and trends of IR small target detection technology.

(2) *The algorithm design needs to be optimized*. Fig. 29 clearly shows that the 13 different IR small target segmentation networks produced false positives or missed detections in cases with weak target energy, low local contrast, and serious background interference. Therefore, to extract the semantic information of IR small targets more effectively, more effort is needed in algorithm design.

(3) *Evaluation indices need to be discussed*. The existing evaluation indices of segmentation algorithms based on deep learning are too one-sided to fully reflect the performance of IR small target segmentation algorithms. Reasonable evaluation indices are necessary and warrant further discussion.

8.1.3. Problems at the deployment application level

Most scholars pay more attention to the detection accuracy of the algorithm while ignoring the importance of lightweight design, resulting in difficulties in mobile terminal deployment and balancing accuracy and computational complexity.

8.1.4. Problems at the follow-up task level

IR small target detection, tracking, recognition, and attacks are sequential task flows. Regarding these, detection technologies provide the necessary data support for follow-up tracking, identification, and attacks. However, owing to the difficulty in research on IR small target tracking and recognition technology and the lack of relevant literature, research in this field is not popular. Simultaneously, some scholars are studying methods to counter the detection, tracking, and recognition technology of IR small targets, such as using jamming bombs to simulate small targets and anti-IR radiation materials to reduce the IR radiation intensity of targets. These countermeasures increase the difficulty of IR small target detection, tracking, and recognition.

8.2. Feature trends

8.2.1. Algorithm design should be more innovative

(1) *Building rich datasets*. After clarifying the research objectives and background, existing public datasets were integrated to refine different scenarios and objectives. If the amount of data is insufficient, data collection and artificial synthesis can be targeted to build a richer dataset.
(2) *Producing high-quality images*. Multiband fusion technologies, such as IR and visible image fusion, IR and radar image fusion, and IR and low-light image fusion, can effectively supplement information, which is helpful in improving image quality and enhancing the semantic information of IR small targets. Simultaneously, attention should be paid to research on a material basis. The imaging quality can only be greatly increased by improving the resolution, response rate, SNR, and detection rate of the IR detector and by reducing the calibration temperature difference and noise power.
(3) *Paying attention to the fusion of data-driven and model-driven algorithms*. Model-driven algorithms are aimed at modeling specific tasks, objectives, and backgrounds. Therefore, these algorithms have good directionality, but poor robustness. However, data-driven algorithms are relatively robust owing to the support of strong data, computing power, and algorithms. Therefore, in algorithm design, more attention should be paid to the fusion of these two types of algorithms to improve the detection accuracy and robustness.
(4) *Strengthening the reverse thinking of algorithm design*. Typically, features and semantic information of IR small targets are extracted according to positive thinking and then segmented from the background. However, many false targets remain in complex contexts. To reduce the false alarm rate, we propose a "false alarm" and "target" collaborative modeling technology. The preliminary research idea is as follows. The IR small target and false alarm source are marked simultaneously in the dataset. Then, the IR small target and false alarm source are detected or segmented simultaneously through the collaborative modeling technology of model-driven and data-driven algorithms, and the authenticity of the target is determined.
(5) *Balancing accuracy and computational complexity of algorithm*. IR small target detection technology is mainly used in sophisticated military equipment, which requires a high real-time performance. Therefore, focus should be placed on a lightweight network design (Table 2 and Section 4.2.1 lightweight design strategy).

8.2.2. Deployment should be actively promoted

For the mobile deployment of IR small target detection technology, not only should lightweight algorithms be designed but model format conversion (converting Python/Tensorflow to frameworks such as onnx and ncnn), software and hardware adaptation, GPU acceleration, and CPU multi-thread resource utilization should also be considered.

8.2.3. Follow-up task research should be deepened

(1) Focus should be placed on research progress in the field of IR small target camouflaging and exploring new research directions for detection technologies.
(2) Research on IR small target tracking and recognition technology should be strengthened. Multimodal multi-source information fusion technology can be used to obtain rich information about IR small targets and thereby improve the tracking and recognition accuracy for these targets.

9. Conclusion

IR small target detection, as a key technology, has attracted considerable attention and made measurable progress over the past few decades. In actual scenes, when the target is more than 10 km or even dozens of kilometers from the infrared detector, coupled with an external influence, such as atmospheric scattering/refraction, lens pollution/distortion, optical defocusing, and various noise, IR detection systems can receive only a very weak target signal. Obviously, owing to weak targets and complex backgrounds, it is challenging to achieve a high detection rate, low false alarm rate, and high real-time performance. Therefore, we have systematically summarized IR small target segmentation algorithms based on a deep learning framework. The survey analyzes 7 characteristics, 8 algorithm design strategies, 8 loss functions, and 13 evaluation indices for IR small targets. Subsequently, the application effects of 18 different types of segmentation networks in IR small target detection tasks have been experimentally compared and analyzed. Finally, the existing problems and future trends in IR small target detection technology have been discussed. In addition, IR small target segmentation is very similar to the segmentation of small lesions in medical images, both of which belong to binary segmentation tasks with extremely unbalanced positive and negative samples. Therefore, the design strategies, loss functions, and evaluation indicators mentioned in this survey can also be used for reference in the field of medical image segmentation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant No. 62171467, Grant No. 61775030, Grant No. 61571096), Natural Science Foundation of Hebei Province of China (Grant No. F2021506004), and Natural Science Foundation of Sichuan Province of China (Grant No. 2022NSFSC40574).

Data availability

Data underlying the results presented in this survey are available in Dataset 1–4, Refs. [10,51–53,75].

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.patcog.2023.109788.

References

[1] M. Zhao, W. Li, L. Li, J. Hu, P. Ma, R. Tao, Single-frame IR small-target detection: a survey, IEEE Geosci. Remote Sens. Mag. 10 (2022) 87–119, doi:10.1109/MGRS.2022.3145502.

[2] R. Kou, C. Wang, Q. Fu, J. Zhang, F. Huang, Detection model and performance evaluation for the IR search and tracking system, Appl. Opt. 62 (2023) 398, doi:10.1364/AO.469807.

[3] R. Kou, H. Wang, Z. Zhao, F. Wang, Optimum selection of detection point and threshold noise ratio of airborne IR search and track systems, Appl. Opt. 56 (2017) 5268, doi:10.1364/AO.56.005268.

[4] R. Kou, C. Wang, Q. Fu, Y. Yu, D. Zhang, IR small target detection based on the improved density peak global search and human visual local contrast mechanism, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15 (2022) 6144–6157, doi:10.1109/JSTARS.2022.3193884.

[5] T. Liu, Q. Yin, J. Yang, Y. Wang, W. An, Combining deep denoiser and low-rank priors for IR small target detection, Pattern Recognit. 135 (2023) 109184, doi:10.1016/j.patcog.2022.109184.

[6] L. Deng, J. Zhang, G. Xu, H. Zhu, IR small target detection via adaptive M-estimator ring top-hat transformation, Pattern Recognit. 112 (2021) 107729, doi:10.1016/j.patcog.2020.107729.

[7] Y. Li, Y. Zhang, Robust IR small target detection using local steering kernel reconstruction, Pattern Recognit. 77 (2018) 113–125, doi:10.1016/j.patcog.2017.12.012.

[8] C. Gao, L. Wang, Y. Xiao, Q. Zhao, D. Meng, IR small-dim target detection based on Markov random field guided noise modeling, Pattern Recognit. 76 (2018) 463–475, doi:10.1016/j.patcog.2017.11.016.

[9] S.S. Rawat, S.K. Verma, Y. Kumar, Review on recent development in IR small target detection algorithms, Procedia Comput. Sci. 167 (2020) 2496–2505, doi:10.1016/j.procs.2020.03.302.

[10] H. Wang, L. Zhou, L. Wang, Miss detection vs. false alarm: adversarial learning for small object segmentation in IR images, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, doi:10.1109/ICCV.2019.00860.

[11] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: Proceedings of the International Conference on Learning Representations (ICLR), 2015, doi:10.48550/arXiv.1409.1556.

[12] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, doi:10.1109/CVPR.2016.90.

[13] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, doi:10.1109/CVPR.2015.7298594.

[14] F. Chollet, Xception: deep learning with depthwise separable convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, doi:10.1109/CVPR.2017.195.

[15] J. Hu, L. Shen, S. Albanie, G. Sun, E. Wu, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, doi:10.1109/CVPR.2018.00745.

[16] Z. Zhang, X. Zhang, C. Peng, D. Cheng, J. Sun, Exfuse: enhancing feature fusion for semantic segmentation, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, doi:10.1007/978-3-030-01249-6_17.

[17] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (2018) 834–848, doi:10.1109/TPAMI.2017.2699184.

[18] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 4 (2017) 640–651, doi:10.1109/TPAMI.2016.2572683.

[19] H. Noh, S. Hong, B. Han, Learning deconvolution network for semantic segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2015, doi:10.1109/ICCV.2015.178.

[20] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in: Proceedings of the Medical Image Computing Computer-Assisted Intervention (MICCAI), 2015, doi:10.1007/978-3-319-24574-4_28.

[21] T.M. Quan, D.G.C. Hildebrand, W.K. Jeong, FusionNet: a deep fully residual convolutional neural network for image segmentation in connectomics, Front. Comput. Sci. 3 (2021) 613981, doi:10.3389/fcomp.2021.613981.

[22] V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: a deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2017) 2481–2496, doi:10.1109/TPAMI.2016.2644615.

[23] C. Peng, X. Zhang, G. Yu, G. Luo, J. Sun, Large kernel matters — improve semantic segmentation by global convolutional network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, doi:10.1109/CVPR.2017.189.

[24] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Semantic image segmentation with deep convolutional nets and fully connected CRFs, in: Proceedings of the International Conference on Learning Representations (ICLR), 2015. doi: 10.48550/arXiv.1412.7062.

[25] L.C. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking atrous convolution for semantic image segmentation, arXiv preprint, 2017. doi:10.48550/arXiv.1706.05587.

[26] L.C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, doi:10.1007/978-3-030-01234-2_49.

[27] G. Hinton, O. Vinyals, J. Dean, Distilling the knowledge in a neural network, arXiv preprint, 2015. doi:10.48550/arXiv.1503.02531.

[28] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, doi:10.1109/CVPR.2016.308.

[29] Z. Liu, J. Li, Z. Shen, G. Huang, S. Yan, C. Zhang, Learning efficient convolutional networks through network slimming, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2017, doi:10.1109/ICCV.2017.298.

[30] A. Paszke, A. Chaurasia, S. Kim, E. Culurciello, ENet: a deep neural network architecture for real-time semantic segmentation, arXiv preprint, 2016. doi:10.48550/arXiv.1606.02147.

[31] A. Chaurasia, E. Culurciello, LinkNet: exploiting encoder representations for efficient semantic segmentation, in: Proceedings of the IEEE Visual Communications and Image Processing (VCIP), 2017, doi:10.1109/VCIP.2017.8305148.

[32] F.N. Iandola, S. Han, M.W. Moskewicz, K. Ashraf, W.J. Dally, K. Keutzer, SqueezeNet: alexNet-level accuracy with 50x fewer parameters and <0.5 MB model size, arXiv preprint, 2016. doi:10.48550/arXiv.1602.07360.

[33] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, MobileNets: efficient convolutional neural networks for mobile vision applications, arXiv preprint, 2017. doi:10.48550/arXiv.1704.04861.

[34] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.C. Chen, MobileNetV2: inverted residuals and linear bottlenecks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, doi:10.1109/CVPR.2018.00474.

[35] A. Howard, M. Sandler, G. Chu, L.C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q.V. Le, H. Adam, Searching for mobileNetV3, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, doi:10.1109/ICCV.2019.00140.

[36] X. Zhang, X. Zhou, M. Lin, J. Sun, ShuffleNet: an extremely efficient convolutional neural network for mobile devices, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2018, doi:10.1109/CVPR.2018.00716.

[37] N. Ma, X. Zhang, H.T. Zheng, J. Sun, ShuffleNet V.2.: practical guidelines for efficient CNN architecture design, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, doi:10.1007/978-3-030-01264-9_8.

[38] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, N. Sang, BiSeNet: bilateral segmentation network for real-time semantic segmentation, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, doi:10.1007/978-3-030-01261-8_20.

[39] H. Li, P. Xiong, H. Fan, J. Sun, DFANet: deep feature aggregation for real-time semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, doi:10.1109/CVPR.2019.00975.

[40] M. Tan, Q.V. Le, MixConv: Mixed depthwise convolutional kernels, arXiv preprint, 2019. doi:10.48550/arXiv.1907.09595.

[41] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, C. Xu, GhostNet: more features from cheap operations, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020, doi:10.1109/CVPR42600.2020.00165.

[42] M. Fan, S. Lai, J. Huang, X. Wei, Z. Chai, J. Luo, X. Wei, Rethinking BiSeNet for real-time semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2021, doi:10.1109/CVPR46437.2021.00959.

[43] Y. Jing, Y. Yang, X. Wang, M. Song, D. Tao, Amalgamating knowledge from heterogeneous graph neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2021, doi:10.1109/CVPR46437.2021.01545.

[44] X. Yang, J. Ye, X. Wang, Factorizing knowledge in neural networks, in: Proceedings of the European Conference on Computer Vision (ECCV), 2022, doi:10.1007/978-3-031-19830-4_5.

[45] X. Yang, D. Zhou, S. Liu, J. Ye, X. Wang, Deep model reassembly, arXiv preprint, 2022. doi:10.48550/arXiv.2210.17409.

[46] Y. Jing, Y. Yang, X. Wang, M. Song, D. Tao, Meta-aggregator: learning to aggregate for 1-bit graph neural networks, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, doi:10.1109/ICCV48922.2021.00525.

[47] S. Liu, K. Wang, X. Yang, J. Ye, X. Wang, Dataset distillation via factorization, in: Proceedings of the Neural Information Processing Systems (NeurIPS), 2022. doi: 10.48550/arXiv.2210.16774.

[48] S. Liu, J. Ye, S. Ren, X. Wang, DynaST: dynamic sparse transformer for exemplar-guided image generation, in: Proceedings of the European Conference on Computer Vision (ECCV), 2022, doi:10.1007/978-3-031-19787-1_5.

[49] X. Yang, D. Zhou, J. Feng, X. Wang, Diffusion probabilistic model made slim, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2023, doi:10.48550/arXiv.2211.17106.

[50] X. Guan, Research On Key Techniques of Small Target Detection For Airborne IR Search and Track System, University of Electronic Science and Technology of China, 2022 (Doctoral Dissertation).

[51] T. Zhang, L. Li, S. Cao, T. Pu, Z. Peng, Attention-guided pyramid context networks for detection IR small target under complex background, IEEE Trans. Aerosp. Electron. Syst. (2023) 1–13, doi:10.1109/TAES.2023.3238703.
[52] Y. Dai, Y. Wu, F. Zhou, K. Barnard, Asymmetric contextual modulation for IR small target detection, in: Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), 2021, doi:10.1109/WACV48630.2021.00099.
[53] B. Li, C. Xiao, L. Wang, Y. Wang, Z. Lin, M. Li, W. An, Y. Guo, Dense nested attention network for IR small target detection, IEEE Trans. Image Process 32 (2022) 1745–1758, doi:10.1109/TIP.2022.3199107.
[54] M. Zhang, R. Zhang, Y. Yang, H. Bai, J. Zhang, J. Guo, ISNet: shape matters for IR small target detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2022, doi:10.1109/CVPR52688.2022.00095.
[55] X. Sun, L. Guo, W. Zhang, Z. Wang, Q. Yu, Small aerial target detection for airborne IR detection systems using lightGBM and trajectory constraints, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14 (2021) 9959–9973, doi:10.1109/JSTARS.2021.3115637.
[56] B. Hui, Z. Song, H. Fan, P. Zhong, W. Hu, X. Zhang, J. Ling, H. Su, W. Jin, Y. Zhang, Y. Bai, A dataset for IR detection and tracking of dim-small aircraft targets under ground /air background, China Sci. Data 5 (2022), doi:10.11922/csdata.2019.0074.zh.
[57] N. Jiang, K. Wang, X. Peng, X. Yu, Q. Wang, J. Xing, G. Li, J. Zhao, G. Guo, Anti-UAV: A Large-Scale Benchmark for Vision-Based UAV Tracking," in, IEEE Transactions on Multimedia 25 (2021) 486–500, doi:10.1109/TMM.2021.3128047.
[58] C.L.P. Chen, H. Li, Y. Wei, T. Xia, Y.Y. Tang, A local contrast method for small IR target detection, IEEE Trans. Geosci. Remote Sens. 52 (2014) 574–581, doi:10.1109/TGRS.2013.2242477.
[59] H. Qin, R. Gong, X. Liu, X. Bai, J. Song, N. Sebe, Binary neural networks: a survey, Pattern Recognit. 105 (2020) 107281, doi:10.1016/j.patcog.2020.107281.
[60] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, N. Sang, Learning a discriminative feature network for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, doi:10.1109/CVPR.2018.00199.
[61] Z. Zuo, X. Tong, J. Wei, S. Su, P. Wu, R. Guo, B. Sun, AFFPN: attention fusion feature pyramid network for small IR target detection, Remote Sens 14 (2022) 3412 (Basel), doi:10.3390/rs14143412.
[62] C. Yu, Y. Liu, S. Wu, X. Xia, Z. Hu, D. Lan, X. Liu, Pay attention to local contrast learning networks for IR small target detection, IEEE Geosci. Remote Sens. Lett. 19 (2022) 1–5, doi:10.1109/LGRS.2022.3178984.
[63] X. Tong, B. Sun, J. Wei, Z. Zuo, S. Su, EAAU-Net: enhanced asymmetric attention U-Net for IR small target detection, Remote Sens 13 (2021) 3200 (Basel), doi:10.3390/rs13163200.
[64] X. He, Q. Ling, Y. Zhang, Z. Lin, S. Zhou, Detecting dim small target in IR images via subpixel sampling cuneate network, IEEE Geosci. Remote Sens. Lett. 19 (2022) 1–5, doi:10.1109/LGRS.2022.3189225.
[65] S. Liu, P. Chen, M. Woźniak, Image enhancement-based detection with small IR targets, Remote Sens 14 (2022) 3232 (Basel), doi:10.3390/rs14133232.
[66] L. Huang, S. Dai, T. Huang, X. Huang, H. Wang, IR small target segmentation with multiscale feature representation, Infrared Phys. Technol. 116 (2021) 103755, doi:10.1016/j.infrared.2021.103755.
[67] Y. Chen, L. Li, X. Liu, X. Su, A multi-task framework for IR small target detection and segmentation, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–9, doi:10.1109/TGRS.2022.3195740.
[68] F. Chen, C. Gao, F. Liu, Y. Zhao, Y. Zhou, D. Meng, W. Zuo, Local patch network with global attention for IR small target detection, IEEE Trans. Aerosp. Electron. Syst. 58 (2022) 3979–3991, doi:10.1109/TAES.2022.3159308.
[69] A. Wang, W. Li, X. Wu, Z. Huang, R. Tao, MPANet: multi-patch attention for IR small target object detection, in: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGRSS), 2022, doi:10.1109/IGARSS46834.2022.9884041.
[70] B. Zhao, C. Wang, Q. Fu, Z. Han, A novel pattern for IR small target detection with generative adversarial network, IEEE Trans. Geosci. Remote Sens. 59 (2021) 4481–4492, doi:10.1109/TGRS.2020.3012981.
[71] H. Zhou, C. Tian, Z. Zhang, C. Li, Y. Xie, Z. Li, PixelGame: IR small target segmentation as a Nash equilibrium, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15 (2022) 8010–8024, doi:10.1109/JSTARS.2022.3206062.
[72] J.H. Kim, Y. Hwang, GAN-based synthetic data augmentation for IR small target detection, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–12, doi:10.1109/TGRS.2022.3179891.
[73] M. Zhao, L. Cheng, X. Yang, P. Feng, L. Liu, N. Wu, TBC-Net: A real-time detector for IR small target detection using semantic constraint, arXiv preprint, 2019. doi:10.48550/arXiv.2001.05852.
[74] K. Hu, W. Sun, Z. Nie, R. Cheng, S. Chen, Y. Kang, Real-time IR small target detection network and accelerator design, Integration 87 (2022) 241–252, doi:10.1016/j.vlsi.2022.07.008.
[75] R. Kou, C. Wang, F. Huang, Y. Yu, Z. Peng, and Q. Fu, LW-IRSTNet: Lightweight IR Small Target Segmentation Network, TechRxiv preprint, 2023. doi:10.36227/techrxiv.22280995.
[76] R. Kou, C. Wang, Y. Yu, Z. Peng, F. Huang, Q. Fu, Infrared small target tracking algorithm via segmentation network and multi-strategy fusion, IEEE Trans. Geosci. Remote Sens. 61 (2023), doi:10.1109/TGRS.2023.3286836.
[77] X. Ying, L. Liu, Y. Wang, R. Li, N. Chen, Z. Lin, W. Sheng, S. Zhou, Mapping degeneration meets label evolution: learning infrared small target detection with single point supervision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2023. https://doi.org/10.48550/arXiv.2304.01484.
[78] B. Li, Y. Wang, L. Wang, F. Zhang, T. Liu, Z. Lin, W. An, Y. Guo, Monte Carlo linear clustering with single-point supervision is enough for infrared small target detection, arXiv preprint, 2023. doi:10.48550/arXiv.2304.04442.
[79] Q. Hou, Z. Wang, F. Tan, Y. Zhao, H. Zheng, W. Zhang, RISTDnet: robust IR small target detection network, IEEE Geosci. Remote Sens. Lett. 19 (2022) 1–5, doi:10.1109/LGRS.2021.3050828.
[80] M. Fan, S. Tian, K. Liu, J. Zhao, Y. Li, IR small target detection based on region proposal and CNN classifier, Signal Image Video Process. 15 (2021) 1927–1936, doi:10.1007/s11760-021-01936-z.
[81] C. Yu, Y. Liu, S. Wu, Z. Hu, X. Xia, D. Lan, X. Liu, IR small target detection based on multiscale local contrast learning networks, Infrared Phys. Technol. 123 (2022) 104107, doi:10.1016/j.infrared.2022.104107.
[82] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An Image is worth 16x16 words: Transformers for image recognition at scale, arXiv preprint, 2020. doi:10.48550/arXiv.2010.11929.
[83] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, doi:10.1109/ICCV48922.2021.00986.
[84] F. Liu, C. Gao, F. Chen, D. Meng, W. Zuo, X. Gao, IR small-dim target detection with Transformer under complex backgrounds. arXiv preprint, 2021. doi:10.48550/arXiv.2109.14379.
[85] T. Wu, B. Li, Y. Luo, Y. Wang, C. Xiao, T. Liu, J. Yang, W. An, Y. Guo, MTU-Net: multi-level transUNet for space-based IR tiny ship detection, IEEE Trans. Geosci. Remote Sens. 61 (2023) 5601015, doi:10.1109/TGRS.2023.3235002.
[86] K. Wang, S. Du, C. Liu, Z. Cao, Interior attention-aware network for IR small target detection, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–13, doi:10.1109/TGRS.2022.3163410.
[87] X. Liu, X. Li, L. Li, X. Su, F. Chen, Dim and small target detection in multi-frame sequence using Bi-Conv-LSTM and 3D-Conv structure, IEEE Access 9 (2021) 135845–135855, doi:10.1109/ACCESS.2021.3110395.
[88] X. Ying, Y. Wang, L. Wang, W. Sheng, L. Liu, Z. Lin, S. Zhou, MoCoPnet: Exploring local motion and contrast priors for IR small target super-resolution, arXiv preprint, 2022. doi:10.48550/arXiv.2201.01014.
[89] T.Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV, 2017, doi:10.1109/ICCV.2017.324.
[90] S.S.M. Salehi, D. Erdogmus, A. Gholipour, Tversky loss function for image segmentation using 3D fully convolutional deep networks, in: Proceedings of Machine Learning in Medical Imaging (MLMI), 2017, doi:10.1007/978-3-319-67389-9_44.
[91] M. Ju, J. Luo, G. Liu, H. Luo, ISTDet: an efficient end-to-end neural network for IR small target detection, Infrared Phys. Technol. 114 (2021) 103659, doi:10.1016/j.infrared.2021.103659.

Renke Kou received the M.E. degree in weapon science and technology from Air Force Engineering University, Xi 'an, China, in 2017. He is currently working toward the Ph.D. degree with the department of electronic and optical engineering, Army Engineering University, Shijiazhuang, China. His-research interests include image process and pattern recognition.

Chunping Wang received the Ph.D degree in electronic and optical engineering from Shijiazhuang Mechanical Engineering College, Shijiazhuang, China, in 1999. He is currently a Professor with Army Engineering University, Shijiazhaung, China. He obtained four national defense invention patents. His-research interest includes control theory and application, information processing.

Zhenming Peng received his Ph.D. degree in geodetection and information technology from the Chengdu University of Technology, Chengdu, China, in 2001. He is currently a Professor with the University of Electronic Science and Technology of China, Chengdu. His-research interests include image processing, signal processing, and target recognition and tracking. Prof. Peng is a member of Institute of Electrical and Electronics Engineers (IEEE), Optical Society of America (OSA), China Optical Engineering Society (COES), and Chinese Society of Astronautics (CSA).

Zhihe Zhao received the S.M.B degree in Xi 'an Medical University, Xi'an, China, in 2015. She is currently working toward the S.M.M degree in The Third Affiliated Hospital of Air Force Medical University, Xi 'an, China. Her research interests include infrared radiation therapy and Image processing.

Yaohong Chen received his Ph.D. degree from University of Chinese Academy of Sciences in 2022. He is currently an associate research fellow with the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences. His-research interests include the infrared imaging systems, infrared image processing and infrared imaging circuit.

Jinhui Han received the Ph.D. degree in electrical circuit and system from Huazhong University of Science and Technology, Wuhan, China, in 2016. He is currently a lecturer in the College of Physics and Telecommunication Engineering, Zhoukou Normal University, Zhoukou, China. His-current research interests are target detection and image processing.

Fuyu Huang received the Ph.D degrees from the Department of Optical and Electronic Engineering, Mechanical Engineering College, Shijiazhuang, China, in 2014. He is currently a Professor with Army Engineering University, Shijiazhaung, China. His-research interests include optical signal processing, infrared detection, and optical design.

Ying Yu received the M.S. degree in computer software and theory from Yunnan Normal University, Kunming, China, in 2014. She is currently working toward the Ph.D. degree with the department of electronic and optical engineering, Army Engineering University, Shijiazhuang, China. Her research interests include computer vision and pattern recognition.

Qiang Fu received the Ph.D degree in computer science and technology from Tsinghua University, Beijing, China, in 2017. He is a Lecturer with the Army Engineering University, Shijiazhuang, China. He has authored more than 50 technical papers. His-teaching and research interests include automatic control and image engineering.