Prior-Guided Data Augmentation for Infrared Small Target Detection

Ao Wang, Student Member, IEEE, Wei Li, Senior Member, IEEE, Zhanchao Huang, Graduate Student Member, IEEE, Xin Wu, Member, IEEE, Feiran Jie, and Ran Tao, Senior Member, IEEE

Abstract—Recently, a lot of deep learning (DL) methods have been proposed for infrared small target detection (ISTD). A DL-based model for the ISTD task needs large amounts of samples. However, the diversity of existing ISTD datasets is not sufficient to train a DL model with good generalization. To solve this issue, a data augmentation method called prior-guided data augmentation (PGDA) is proposed to expand the diversity of training samples indirectly without additional training data. Specifically, it decouples the target description and localization abilities by preserving the scale distribution and physical characteristics of targets. Furthermore, a multiscale infrared small target dataset (MSISTD) consisting of 1077 images with 1343 instances are constructed. The number of images and the number of instances in MSISTD are 2.4 times and 2.5 times than those of the existing largest real ISTD dataset single-frame infrared small target (SIRST) benchmark, respectively. Extensive experiments on the SIRST dataset and the constructed MSISTD dataset illustrate that the proposed PGDA improves the performance of existing DL-based ISTD methods without extra model complexity burdens. In comparison with SIRST, MSISTD has been evaluated as a more comprehensive and accurate benchmark for ISTD tasks.

Index Terms—Data augmentation, infrared images, infrared small target detection, multiscale infrared dataset.

I. INTRODUCTION

In the past few years, single-frame infrared small target detection has received more and more attention. Compared to other sensing devices (active imaging radar, visible imaging), infrared imaging has the following characteristics: anti-interference, excellent portability [1], strong cloud penetration, blind spot detection, etc. For these reasons, the small infrared target small detection (ISTD) has been widely used in military applications such as early warning, missile tracking systems [2], precision guidance, and maritime surveillance. However, infrared sensors usually start at a long distance from the target. So the target usually occupies only a small part of the image. In extreme cases, they are just Gaussian point-like targets. Other reasons, such as complex backgrounds, flash noise, atmospheric scattering, optical defocus, etc., also lead to ultralow signal-to-noise ratio (SNR) [3] and weak texture features in infrared images. In addition, the infrared imaging platform and the small target are moving at high speed, which makes it difficult to acquire a sufficient number of single-frame images. All of the above reasons greatly reduce the performance of temporal recognition.

Single-frame ISTD methods are roughly classified into two categories: data-driven and model-driven methods. For model-driven algorithms, they are divided into four categories including filter-based methods [4], [5], [6], low-contrast-based methods [7], [8], [9], [10], low-information-based methods [11], [12], [13], and data structure-based methods [14], [15], [16], [17]. The filter-based methods are divided into those for the spatial domain and the frequency domain. They are widely used because of the small number of hyperparameters. The data structure-based approach models the context and the target separately from different data structures. It mainly includes subspace structure-based, dictionary-based, and tensor representation-based approaches. However, the effectiveness of model-driven approaches is very dependent on the specific scenario and the manually adjusted hyperparameters. When the field of the dataset varies too much, the performance of the above algorithms drops dramatically.

The current approaches based on DL are divided into five categories: optimization approaches for specific criteria, GAN-based methods, transformer-based methods, global and local feature fusion methods, and feature fusion methods based on specific attention modules. McIntosh et al. [18] developed a target-to-clutter ratio detection network (TCRNet) by optimizing the TCR criterion to prioritize the representation of the infrared small object. GAN-based methods [2], [19] were proposed to focus on essential features of ISTD and predict the intensity of targets. Targets are identified and reconstructed by the generator.
according to the data distribution. Meanwhile, discriminators are constructed to improve the data-fitting ability of the generator. Vision transformer [20] and swin-transformer [21] have been applied in all downstream tasks [22] in the vision domain. Liu et al. [23] only used simple ViT architecture to model foreground and background. Global and local feature fusion-based methods usually consider extreme class-imbalance between the target and background pixels. They usually use the global attention module and the local attention module [24], [25] to extract global and local information, respectively. Coarse-grained features extracted by the large-scale branch and fine-grained features extracted by the local branches are fused through the developed bottom-up structure. Feature-fusion-based methods [26], [27], [28], [29], [30] usually design different kinds of feature enhancement modules manually. These plug-and-play components make the features of small targets more prominent while making the background effectively suppressed, but they greatly increase the complexity.

On the one hand, most of these models are evaluated on multiframe small target datasets and single-frame small target datasets with prominent targets. There is no comprehensive benchmark to use these datasets to evaluate the performance of individual models. And there is no apparent performance gap between model-driven methods and data-driven methods on these datasets. On the other hand, existing models improve the performance of certain evaluation metrics by increasing the parameters and complexity of the models. However, few of them notice the enhancement of models by data augmentation (DA). In addition, generic DA algorithms are designed for mainstream datasets and are not optimized and changed specifically for ISTD. Also, the existing DA algorithms change the scale distribution and grayscale intensity distribution of small targets. Representative DA methods such as mosaic augmentation [31], cut-mix [32], and mixup [33] mix the information contained in different images with appropriate changes to ground truth labels. In practice, these methods often leave the performance of the model unchanged or even degraded.

Therefore, a dataset is constructed to properly evaluate the gap between model-driven and data-driven methods. In addition, it allows for a more accurate assessment of the performance of existing DL methods. A DA strategy is proposed to address the limitations of generic DA algorithms. It improves the performance of DL-based methods without increasing the algorithmic complexity of the model and introducing additional data. To this end, the contributions of this work are as follows.

1) A prior-guided data augmentation (PGDA) method is proposed to generate more training samples for infrared small target detection. It introduces the prior scale distribution and physical representation of targets to keep the statistics of the original and generated samples consistent. In addition, the PGDA decouples the description ability of the target and the localization ability of the target, which enhances the DL methods performance.

2) A multiple scenes infrared small target dataset (MSISTD) is constructed to improve the diversity of the ISTD dataset and the DL methods of generalization. It extends the original 427 datasets (SIRST) to 1077 and adds high-quality annotations. It involves more diverse backgrounds, richer target shapes, and less prominent samples.

3) Extensive experiments on existing SIRST [30] and the constructed MSISTD illustrate that the proposed PGDA allows DL methods to obtain improvement in both IoU and nIoU metrics without extra model complexity burdens. The strategy outperforms other methods. In addition, it is sufficiently robust and effective for the ISTD problem when compared with generic data enhancement methods and small sample training.

The rest of this article is organized as follows. In Section II, we describe the framework of PGDA as a whole and its two components. In Section III, the MSISTD dataset is constructed and elaborated. In Section IV, extensive comparison experiments and ablation experiments are presented. Section V concludes this article.

## II. RELATED WORK

### A. Data Augmentation for ISTD

The previous work [34] concentrates on GAN-based synthetic data augmentation. It achieves the image-to-image translation to produce synthetic background images from visible spectrum images. By means of the intensity modulation of GAN, targets are rendered on the synthetic images. They allow most DA methods to be improved. However, it requires a large number of visible and infrared image pairs to train the modal transfer network. In addition, the training of GAN networks causes a large amount of computational complexity and model burden. Moreover, the synthesized background images require manual selection and extensive hyperparameter settings.

In fact, there are large sets of IR background data that do not require synthesis. Modality transfer from visible images to infrared images is not necessary. Furthermore, if we remove the intensity modulation, it degenerates into the regular copy-paste type methods. Copy-paste type methods are simple ways to combine information from multiple images in an object-aware manner to copy instances of objects from one image and paste them onto another image. Depending on the different mainstream tasks, it differs from copying the exact pixels of the object in semantic segmentation to copying the pixels within the bounding box in target detection. Google copy-paste [35] did not model surrounding visual context rather than combining large-scale jittering and self-training. Chen et al. [36] introduced AdaResampling, which pretrained the semantic segmentation network and used median filtering to select the position of the object. Instaboost [37] simply introduced random jittering to objects, which could improve the performance of Mask R-CNN [38]. Cut-paste-and-learn [39] proposed a simple way to generate a large quantity of datasets. It utilizes cutting object instances and pastes them on actual backgrounds. And models were trained on the mixed dataset. Small target augmentation [40] used oversampling and enhancement techniques to solve the problem of sample imbalance.

Most of the abovementioned methods focus on the performance of the generic dataset and are not specifically designed for small IR targets. In practice, copy-paste augmentation works
terribly on ISTD problems because it is sensitive to the scale of information and physical representation. The proposed PGDA follows the above paradigm with some differences.

1) We dynamically adjust the grayscale intensity and area ratio value of the pasted instances after introducing prior knowledge of the training set.
2) We perform multiscale random cropping on the original image to obtain subgraphs. They replace the original image as the background where the instances are pasted.
3) We employ the nongrid transforms to make the shape of targets more diverse.

On the one hand, the proposed PGDA makes more detailed optimizations for the above problems so that copy-paste’s performance get sufficient improvement. On the other hand, the proposed method improves the performance of the existing DL methods performance without extra model complexity burdens and lots of visible-infrared image pairs, compared to the GAN-based methods.

B. Datasets for ISTD

Model-driven algorithms and most DL algorithms train their network on self-built datasets. These datasets are not sufficient for training a network with good generalization. Only a small minority of them are available such as SIRST [30] MFSIRST [19]. These datasets facilitate the ISTD task, where correlation algorithms can be compared. However, the following limitations prevent these from being used as a comprehensive and accurate benchmark.

The current ISTD datasets have few real images. Besides, most of them are synthetic. Except for the SIRST [30], the existing ISTD datasets use image synthesis methods to expand the number of samples. Zhang et al. [29] deflated the SIRST dataset to a fixed size. Then, they cropped out four subfigures on each image. Through affine transformation, a training set including 8525 images and a testing set with 545 images were finally generated. Wang et al. [19] collected the shape and texture information and background information to generate 10 000 figures. However, a great many generated images occurred pixel missing problems after the affine transformation. The physical representation of the small target already differs from the real ones. They have the problems of low-resolution and pixel-missing, which lead to application bottlenecks in many ISTD tasks.

III. PROPOSED PGDA STRATEGY

The proposed PGDA is shown in Fig. 2, which mainly consists of two parts: the target enhancement module and the subfigure-split module. In this section, we first introduce the overall framework. Then, we present the details of the subfigure-split module and the target enhancement block.

A. PGDA Framework

The proposed PGDA generally obeys the following paradigm: 1) selecting a set of instance images and background images; 2) cropping the instances and performing a series of data augmentation processes on them; 3) selecting the locations and background where the instances are pasted; 4) fusing the instances and the background to generate a new image.

In the proposed strategy, the target enhancement consists of Gaussian paste and signal-to-clutter ratio (SCR) perception. SCR perception is used after Gaussian weighting to make them match the physical representation of the actual targets. Subfigure-split is to perform multiple multiscale random crop of the background and adjust the resolution of the background according to the scale of the target. PGDA decouples the ability to describe the target and the ability to localize the target, which overcomes the limitations of the general-purpose algorithm for the ISTD problem.

Before starting the whole algorithm, we calculate the target scale distribution and the SCR distribution of the training set to integrate them into the algorithm as prior knowledge. The area ratio is used to measure the scale characteristics of the target. The proportional ratio between the total number of pixels occupied by a small target and the total number of pixels in the background is defined mathematically below

\[
\text{AreaRatio} = \frac{\sum \text{Pixel}_T}{\sum \text{Pixel}_B}
\]

where \(\sum \text{Pixel}_T\) represents the total number of pixels of the target, and \(\sum \text{Pixel}_B\) represents the total number of pixels of the whole background. We calculate the area ratio values for all small targets in the training set and count all values below 0.15%. This area ratio is set as \(A\).

The signal-to-clutter ratio (SCR) is used to describe how difficult it is to detect small targets in infrared images [8], [41], [42]. The larger the SCR value, the easier the target is to detect. SCR is mathematically defined as follows:

\[
\text{SCR} = \frac{|\mu_T - \mu_B|}{\sigma_B}
\]

where \(\mu\) represents the mean value of the target pixels, \(\mu_B\) and \(\sigma_B\) represent the average pixel value and standard deviation in the neighboring region, respectively. We set the radius \(d\) of the small target neighborhood to 20. In Fig. 1, the visualization of the SCR calculation process is shown. We calculate the SCR values for all small targets. To eliminate outlier values such as too large and too small, all the above SCR values are retained from 20% to 80% after using the data binning algorithm. We record this fraction of SCR values as \(S\).

After counting the above two distributions, we crop down all the small targets according to the labeled images. We define this set of small target images as \(T\). Then, We utilize the subfigure-split module to generate a candidate background set which is defined as \(B\). Next, we randomly select the targets, subfigure from \(T\) and \(B\), respectively. The locations of the targets are also randomly selected. Combining the SCR distribution \(S\) and scale distribution \(A\), PGDA is designed to dynamically adjust the relative grayscale intensity differences and scale differences between the targets and background image. Finally, we paste.
and among all the collections of images background images. Two parts are represented as into background images with the presence of targets and pure background to obtain a large number of background images. The set of subfigures in the training set require a subfigure segmentation algorithm random cropping on a single image to obtain multiple images. The second way is to perform multiple multiscale subfigures. The overlapping area between subfigures and the target to the selected position in the background image to achieve spatial reorganization.

**B. Subfigure-Split**

According to the convention, the size of targets less than 0.15% is defined as small targets in the ISTD field. However, the generic DA algorithms do not take into account the scale distribution properties of the small targets in ISTD. In addition, the contradiction between low signal-to-noise ratio of low-resolution images and small-scale targets hinders the effectiveness of the generic DA algorithms. Compared to the copy-paste [35], the existing data are insufficient to use a supervised model trained with PGDA to generate pseudolabels on unlabeled data. To enhance the diversity of the dataset, subfigure-split is designed to depend solely on the train set rather than introduce additional data and solve the above drawbacks.

Subfigure-split is employed in two ways. The first way is to use the subfigure-cutting algorithm in the remote sensing field to obtain different images by controlling the resolution of subfigures and the overlapping area between subfigures and figures. The second way is to perform multiple multiscale random cropping on a single image to obtain multiple images. In Fig. 2, the second way is illustrated. All the images and masks in the training set require a subfigure segmentation algorithm to obtain a large number of background images. The set of background images is defined as $\mathcal{B}$, where $\mathcal{B}$ is subdivided into background images with the presence of targets and pure background images. Two parts are represented as $\mathcal{B}_T$ and $\mathcal{B}_O$, respectively. In the subsequent processing, we increase the resolution of all subfigures using interpolation for the later steps of pasting. After selecting a pair of targets and background, we calculate the corresponding area ratio. Next, we compare whether the value is in the scale distribution. If it is or is not in this range, we randomly assign a new area ratio ($\beta$) value according to $\mathcal{A}$. We scale this area ratio by changing the object size $\pm 10\%$. According to this value, we fix the resolution of the background to change to another resolution.

In order to verify whether the algorithm can preserve the scale distribution, the subfigure-split is employed to generate different synthesis rations of small target images. In Fig. 3, the distribution of the histogram clearly indicates that the scale distribution of the generated images after processing by the augmentation algorithm is consistent with the trend of the distribution of the original dataset. Also, the distribution of the number of targets per image is constant since the small targets are pasted onto the subfigures. To achieve this outcome, two components play a key role. In the subfigure-split module, we give preference to pure background images $\mathcal{B}_O$ among all the collections of images cropped down. Because random cropping actually enlarges the subregions of the image. If there are already small targets in the background, this leads to a change in the scale distribution of the generated dataset. Second, the area ratio of each target is dynamically allocated to ensure the approximate invariance of the scale distribution before fusing the target and background.

Subfigure-split greatly improves the attention of subregions by multiscale random cropping. Moreover, the decoupling and reorganization process of target and background space enhances the diversity of data and effectively alleviates the problem of overfitting DL methods.

**C. Targets Enhancement**

The ISTD problem is different from the target detection of common objects. It has two important properties. The first is that target is characterized as Gaussian point-like or has a more pronounced Gaussian edge attenuation. Second, due to the infrared imaging mechanism, the temperature of the target is greater than the temperature of the surrounding environment. That is why ISTD is widely used in areas such as missile warning, camouflage identification, and ultralong range target tracking. However, most of the existing methods for fusing background and infrared small targets ignore considering the characteristics of ISTD. The representations of the generated images are different from those of the real images. In Fig. 4, we list two typical problems in current small target image fusion algorithms.

The first problem is the incorrect Gaussian grayscale representation of the target. Due to the low-resolution characteristics of the infrared target, some of the images do not clearly characterize the full Gaussian characteristics. Also, the cropping algorithm loses a portion of the target’s edges, where the Gaussian attenuation is most apparent. For these two reasons, the generated target will have only one grayscale intensity. In Fig. 4, the specific phenomenon is also illustrated.

The second problem is the background migration of the target. Since the infrared imaging mechanism is based on the temperature difference of different targets, the gray intensity of the cropped-down target is determined by the current background. When it is pasted onto another image, the grayscale intensity of the target may be weaker than that of the background. In Fig. 5, we have listed two typical examples. The first one is that the overall intensity of the target is much lower than the background intensity, which results in exactly opposite temperature representations of the target and the background. In the second example, the central gray intensity of the target matches the gray intensity of the background, but the edge attenuation of the Gaussian makes the outermost circle of the small target lower.
Fig. 2. Overview of the proposed PGDA. We crop all the instances in the training set and selected several small targets randomly. Then, we perform a series of pre-processing operations including Gaussian paste and SCR perception on them. Also, we randomly select the subfigure and place the enhanced targets into the selected positions on it. Finally, targets and background images are fused.

Fig. 3. Synthesis ratio represents the number of generated images compared to the original number of images.

Fig. 4. Background migration and Gaussian grayscale representation.
fused. Otherwise, the foreground is enhanced and diminished by randomly assigning random SCR values in random $S$ to solve backward for the difference in the average gray intensity of the foreground according to (2).

Gaussian paste overcomes the problems caused by low-resolution and cropping algorithms, which makes the generated target obey Gaussian distribution forcibly to satisfy the first property of ISTD. SCR perception calculates the intensity difference and stability difference before the target and background fusion. By introducing prior knowledge of the training set and random intensity variations, the small target also has some flexibility while maintaining the SCR distribution of the original dataset. More importantly, it ensures that the gray intensity of the target is always higher than that of the background to satisfy the second property of ISTD. In Fig. 6, we list some representative generated images and visualize the corresponding heatmaps to observe the Gaussian representations and grayscale intensities of small targets.

IV. MULTIPLE SCENE INFRARED SMALL TARGET DATASET

A. Image Collection and Overview

The constructed MSISDT dataset has 1077 images, which includes a total of 1343 instances. It expands the number of images of the original SIRST by 2.4 times and the number of instances by 2.5 times in the Table I. In addition, it greatly expands the context of single-frame small target detection from being limited to the detection of aerial targets only to the detection of multiscene, multiscale, and lower SCR targets. It includes pure sky background, dense cloud sky background, low signal-to-noise ratio sky background, sea-land interface background, woodland sky background, urban sky background, etc.

Algorithm 1: Target Enhancement.

**Input:** Small target image data set $T$; SCR distribution of the training set $S$; Pure background images $B_o$; Number of small targets per image $N_t$.

**Output:** Synthesis Image $O$

1. Randomly select a background image $B_o$ in $B_o$.
2. for $i$ in $N_t$ do
   3. Randomly select a small target $T_i$ in $T$.
   4. Randomly generate location center coordinates $(x_o, y_o)$.
   5. $T_i = g(t(T_i)); g(t())$ denotes geometric transformation.
   6. $T_i = n(t(T_i)); n(t())$ denotes no-rigid transformation.
   7. $T_i = g(p(T_i)); g(p())$ denotes Gaussian paste.
   8. Define location neighborhood as $R(B_o(x_o, y_o))$ and calculate the region $scr = f(x_o, y_o)$ according to (2) and radius $d$.
   9. if $min(S) < f(x_o, y_o) < max(S)$ then
      10. Continue;
   11. end
   12. else
      13. random choice $S_i$ in $S$.
      14. $Diff = S_i \times std(R(B_o(x_o, y_o))) - avg(T_i)$
      15. $+ avg(R(B_o(x_o, y_o)))$;
      16. $T_i = Colorjitter(T_i, Diff)$;
   17. end
18. end
19. return $T_i$;
20. Note: Paste indicates blending the target with the background.
TABLE I

| Datasets       | Image type | Background scene             | Number | Label type   | Target type            | Public |
|----------------|------------|-------------------------------|--------|--------------|------------------------|--------|
| SIRST [30]     | Real       | Cloud/City/Sea               | 427    | Segmentation | Point/Spot/Extended   | True   |
| MFIRST [19]    | Synthetic  | Cloud/City/River/Road        | 10000  | Segmentation | Point/Spot            | True   |
| NUDT-SIRST [26]| Synthetic  | Cloud/City/Sea/Field/Highlight| 1327   | Segmentation | Point/Spot/Extended   | False  |
| CQU-SIRST [11] | Synthetic  | Cloud/City/Sea               | 1676   | Point        | Point                 | False  |
| AUG-SIRST [29] | Synthetic  | Cloud/City/Sea               | 9050   | Segmentation | Point/Spot            | True   |
| MSISTD         | Real       | Cloud/City/Sea/River/Sky/Low-SNR/Field/Forest | 1077   | Segmentation | Point/Spot/Extended   | True   |

Fig. 7. Representative images of some scenes in the MSISTD dataset.

urban building background, forest-to-ground background, inland river scene, and other scenes. Several representative figures are shown in Fig. 7. The dataset includes 427 images from the original SIRST dataset [30], 100 single-frame real datasets from the dataset [19], and 650 high-resolution single-frame images with high-quality annotations.

The MSISTD dataset is collected by acquiring a large number of images of infrared sequences, and then screening those that meet the definition of small targets for annotation and processing. For this purpose, we process over 1200 sequences of video data [46], [47], [48]. There is also a big portion of them, obtained by actually filming. In addition, the small targets usually keep a long distance away from the infrared camera, as well as several limitations of infrared imaging. Small targets are characterized by two important properties, extremely low scale, and extremely low SCR. As a result, we can not perceive and distinguish the types of targets by humans, which leaves the MSISTD dataset with only two categories: foreground and background. We provide annotation forms for the semantic segmentation of datasets, and transformation scripts for the annotation of the bounding box and key point. The MSISTD is currently publicly available at https://github.com/Crescent-Ao/MSISTD.
Fig. 8. Number of small targets which their area ratio above 0.15% in the SIRST and MSISTD datasets is too small in the distribution. Therefore, the maximum value of the x-axis coordinate of the histogram is limited to 0.15%.

B. Dataset Statistics and Comparison

We perform the corresponding visualization analysis for the scale and SCR characteristics of the dataset in Fig. 8. The average area ratio indicator of the SIRST is 0.048%, while that of the proposed MSISDT is 0.069%. The constructed MSISDT contains more small targets with different shapes to increase the diversity of the dataset, which leads to an increase in the average area ratio. In the two distribution plots, the MSISDT scale distribution trend is in line with the original SIRST distribution trend, while the overall number of small IR targets has increased substantially. In terms of SCR metrics, the mean value of SCR in the MSISDT dataset is 8.94, while the corresponding value in the original dataset is 12.03. This can increase the detection ability of DL methods under extreme conditions. From the visualization of the distribution, the MSISDT has greatly increased the number of small targets with low SCR values while maintaining the original data scale distribution. This also indicates the diversity of scenarios in the dataset, which greatly increases the detection difficulty. We count the distribution of the resolution of the dataset similarly. In the MSISDT dataset, the average length of images is 377 pixels, and the average width of images is 286, while those of the images in the SIRST dataset are 303 and 220, respectively. The higher resolution allows DL methods to extract effective semantic information from datasets.

In addition, the current ISTD datasets have less diversity of scenarios and targets. They mostly focus on two scenes including dense clouds and different sky backgrounds. The pixel intensities in these scenes are smoothly transformed without flash noises. DL methods are easily overfitted on these datasets. In addition, these datasets pay less attention to the targets, which are performed as Gaussian point-like targets. Monotonous backgrounds and salient targets make DL methods easily overfitting. When scenes and target types are switched, DL methods perform worse than model-driven methods. In this article, a multiple scenes infrared small target dataset (MSISTD) is constructed to improve the diversity and quantity of the ISTD dataset. It covers lots of different scenes and backgrounds and contains different types of targets including drones, people, vessels, helicopters, and so on.

In summary, the MSISTD dataset has significantly increased the number of instances while maintaining the original dataset scale distribution. Moreover, the SCR distribution of the dataset tends to be more toward low SCR, because of its background diversity, and greater detection difficulty. The MSISTD dataset properly evaluates the gap between model-driven and data-driven approaches. In addition, it allows for a more accurate assessment of the performance of existing deep learning. More importantly, the overall image resolution of the dataset is substantially improved, which allows it to be widely used in deep learning algorithms.

V. EXPERIMENTAL PERFORMANCE AND DISCUSSION

A. Dataset and Evaluation Metrics

The attempt is toward DA methods for DL methods and model-driven methods, so we evaluate the proposed strategy on two datasets, i.e., (SIRST dataset and MSISTD dataset) using different separate models. Since too few test images may cause low confidence in the model predictions, we adjust the ratio of training samples and testing samples on the SIRST dataset for experiments. All single-frame small target datasets are split into 50% for training and 50% for testing. Note that, since MSISTD includes SIRST, most of the comparison experiments are conducted in the former.

The MSISTD uses ten metrics to evaluate deep learning models. We use intersection over union (IoU), and normalized intersection over union (nIoU) to evaluate shape description ability. In addition, we also use three common metrics including Precision, Recall, and F1-measure to evaluate the overall performance of the model. We utilize the AUC metric instead of the ROC curve in traditional algorithms since most deep learning algorithms tend to have lower threshold points and quantified less variance in the ROC curve. At last, $P_d$ and $F_a$ are used to evaluate localization ability

$$nIoU = \frac{1}{N} \sum_{i}^{N} \frac{TP_i}{T_i + P_i - TP_i}$$  (3)
where \( T_i, P_i, TP_i, N \) stand for label, predict result, true positive pixel numbers, and the number of samples

\[
F_{\text{measure}} = \frac{2 \times \text{Prec} \times \text{Recall}}{\text{Prec} + \text{Recall}}
\]

(4)

\[
Pd = \frac{T_{\text{correct}}}{T_{\text{All}}}, \quad Fa = \frac{P_{\text{false}}}{P_{\text{All}}}
\]

(5)

where \( T_{\text{correct}} \) and \( T_{\text{All}} \) represent the numbers of correctly predicted targets and all targets. Correspondingly, \( P_{\text{false}} \) and \( P_{\text{All}} \) represent the numbers of falsely predicted pixels and all image pixels. If the centroid derivation of the target is less than three pixels, the targets are considered as correctly predicted. Otherwise, they are defined as false alarms.

B. Implementation Details and Comparison Methods

The models chosen for the experiments mainly include DL models. Comparative experiments and ablation experiments are conducted on the SIRST and MSISTD datasets. Furthermore, small sample training experiments are also conducted. The former is to verify the effectiveness and robustness of the proposed DA strategy, and the latter is to demonstrate that the PGDA also works in extreme case.

Different public models have different optimizers and learning rate tuning strategies involving too many hyperparameters. We adapt the corresponding settings in these papers. The models chosen for the experiments mainly include data-driven models and model-driven models. The former is to verify the effectiveness of the proposed data PGDA, and the latter is to demonstrate that the PGDA also works in extreme case.

In model-driven methods, we compare several methods including multiscale grey difference weighted image entropy (MGDWIE) [49], weighted strengthened local contrast measure model (WSLCM) [10], fast adaptive masking and scaling with iterative segmentation model (FAMISIS) [16], partial sum of tensor nuclear norm (PSTNN) [12], reweighted infrared patch tensor (RIPT) [13], novel local contrast descriptor (NLCD) [51], nonconvex optimization with LP-norm constraint (NOLC) [50] and self-regularized weighted sparse model (SRWS) [17]. We compared several deep learning models: asymmetric contextual modulation (ACM) [30], attention-guided pyramid context networks (AGPCNet) [29], dense nested attention network (DNANet) [26], multipatch attention network (MPANet) [53], HRNet [52], patch merging attention [53], and full scale dynamic head network (FSDNet) [26]. All of the above seven open-source models are successfully reproduced. In addition, MDvsFAGAN [19] has a large uncertainty due to its weaker performance and different output results of the three branches. All the networks used in our experiments are implemented with PyTorch 1.10 and run on an NVIDIA Geforce RTX 3090.

C. Experimental Results

We conduct comparisons on the SIRST and MSISTD datasets, respectively. Our metrics on SIRST are consistent with the original papers. Only two metrics (IoU and nIoU) are selected. On the SIRST in the Table II, all models have corresponding improvements in their respective metrics. As for the IoU metric, all models show an average increase of 1.53 and a maximum increase of 2.60 after using PGDA. On the nIoU metric, all models show an average increase of 1.14 and a maximum increase of 2.05 after using PGDA. On the MSISTD dataset, the PGDA strategy increases the IoU metric by an average of 1.68 and the nIoU metric by an average of 1.60 for all models. Moreover, it enables the model IoU metric to improve by 2.97 and the nIoU metric to get the max improvement of 6.21. The visualization is shown in Fig. 9. The results on two datasets successfully confirm the effectiveness and robustness of the PGDA strategy.

Since the MSISTD dataset is twice as large as the SIRST, more evaluation metrics are used to fully evaluate the PGDA. In Table III, we present the results of different models in detail, which indicate the PGDA performance. Due to the forced increase in the number of small targets, the distribution of the training dataset is slightly different from the distribution of the original dataset. Therefore, some of the models are lower in the Prec metric. However, due to the high performance of the strategy in recall boosting, the F1 metric of all models has improved accordingly.

The experimental assertions are likewise verified for both Pd and Fa metrics. The subfigure-split algorithm enhances the attention of the model to the subfigure. Due to the randomness of the target paste position, the model can cope with various cases of pictures. The above two aspects make the model’s localization ability improve substantially. We conduct comparison experiments with each of the seven different models. On the existing SIRST, all algorithms obtain performance improvements in both metrics. The PGDA strategy results in improvement performance improvements in most of the individual models when conducted on the MSISTD. Although there is a reduction in some of the metrics, we split IoU/nIoU, Prec/Recall, and Pd/Fa into three groups. All three groups
Fig. 9. Detection results are predicted by data-driven methods and model-driven methods on the constructed MSISTD. The red box represents omission detection. The orange box represents a false alarm. The blue box represents correct detection.

TABLE III

COMPARATIVE EXPERIMENTAL RESULTS ON THE SIRST DATASET

| Methods   | IoU(%) | nIoU(%) | AUC(%) | Pd(%) | Fa(%) | Prec(%) | Recall(%) | F1(%) |
|-----------|--------|---------|--------|-------|-------|---------|-----------|-------|
| MGDWIE [49] | 86.91  | 89.73   | 3462.3 | 0.25  | 89.73 | 0.51    |           |       |
| WSLCM [10]  | 51.51  | 299.1   | 7.77   | 0.75  | 3.36  | 1.22    |           |       |
| FAMSIS [16] | 22.47  | 69.25   | 80.20  | 4740.0| 0.69  | 43.23   | 1.35      |       |
| RIFT [13]   | 3.00   | 20.14   | 70.22  | 85.48 | 972.7 | 3.13    | 41.66     | 5.83  |
| PSTNN [12]  | 5.79   | 27.91   | 73.4   | 76.97 | 538.4 | 6.24    | 48.21     | 11.05 |
| NOLC [50]   | 3.19   | 21.42   | 70.12  | 84.17 | 907.4 | 3.34    | 41.41     | 6.18  |
| NLCD [51]   | 3.91   | 21.46   | 70.20  | 84.14 | 723.2 | 4.14    | 41.33     | 7.54  |
| SRWS [17]   | 4.3    | 28.25   | 64.24  | 412.6 | 86.20 | 3.08    | 29.00     | 8.62  |
| ACM [30]    | 56.20  | 59.99   | 88.32  | 88.12 | 14.22 | 72.87   | 21.06     | 71.59 |
| w/PGDNet [29] | 53.09  | 52.77   | 70.79  | 85.19 | 10.21 | 65.04   | 74.29     | 69.36 |
| w/PGDA [29] | 54.82+1.72 | 52.82+0.05 | 73.39+2.6 | 87.48+2.29 | 7.02+3.09 | 64.09+0.95 | 70.73+4.84 | 70.82+1.46 |
| AGPCNet [29] | 58.31  | 58.43   | 89.76  | 92.81 | 11.69 | 69.04   | 78.96     | 75.67 |
| w/PGDNet [52] | 59.57+1.26 | 60.33+1.90 | 88.25+1.51 | 94.28+1.47 | 17.7+0.08 | 70.97+1.93 | 78.76+0.20 | 74.68+1.01 |
| w/PGDA [52] | 61.48  | 64.21   | 96.70  | 89.29 | 5.67  | 78.14   | 74.25     | 76.14 |
| PMANet [53] | 60.79  | 63.38   | 87.78  | 90.17 | 5.41  | 76.52   | 74.73     | 75.61 |
| w/PGDA [53] | 62.60+1.12 | 64.99+0.78 | 97.54+0.84 | 89.58+0.59 | 6.30+0.70 | 72.51+2.05 | 86.14+0.07 | 78.74+2.71 |
| FSDNet [26] | 61.69  | 64.76   | 96.69  | 89.85 | 5.34  | 79.64   | 73.24     | 76.31 |
| w/PGDNet [26] | 61.82+1.03 | 64.08+0.70 | 88.40+0.62 | 91.34+1.17 | 6.08+0.67 | 78.54+0.40 | 75.51+1.26 | 76.99+0.85 |
| w/PGDA [26] | 62.46+0.77 | 64.64+0.12 | 97.54+0.85 | 90.76+0.91 | 7.03+1.69 | 80.18+0.54 | 73.87+0.63 | 76.90+0.59 |
| DNA [54]    | 62.14  | 60.66   | 92.27  | 90.33 | 3.45  | 73.60   | 83.80     | 78.23 |
| w/PGDNet [54] | 65.08+2.94 | 66.87+6.21 | 90.18+2.09 | 93.40+3.07 | 5.15+1.7 | 78.21+4.61 | 79.49+4.01 | 78.84+0.60 |

Note: The FSDNet and DNA employ the same structure. The difference is that the UNet++ structure is replaced with UNet++. In the decoding process, a dynamic head is utilized for feature enhancement between five different semantic differences. The FSDNet is available at https://github.com/KaleDemo/FSDNet. Model-driven methods do not use the proposed PGDA.
Fig. 10. Problem of positive and negative sample imbalance is solved to some extent by FSDNet due to its focal loss. Therefore, the performance of the FSDNet gets significant improvement.

We compare some of the current mainstream DA algorithms including oversampling [40], mixup [33], and cut-paste-learn [39]. Due to the specificity of the ISTD problem, Mosaic causes the scale distribution of the target to change, so we do not compare this algorithm. The oversampling method causes the ratio of positive and negative samples to increase, but destroys the ratio of the number of samples per image, even when the original dataset is retained. In the Table IV, the indicator decline phenomenon is also consistent with the performance. Mixup is also quite an effective method for DA algorithm. It does not destroy the original scale distribution, and makes the detection difficulty substantially higher by fusing two images, which improves the generalization of the algorithm to some extent. However, due to the SCR characteristics of the small infrared target, it is extremely difficult to detect. Also, image fusion makes small targets drowned in noise. To summarize, the former destroys the scale and number distribution of small targets, and the latter destroys the SCR distribution of small targets. The two components of PGDA perform excellently in preserving the scale and SCR distribution of the dataset. It is more applicable to the ISTD problem than the generic DA algorithm. We also explore the performance of the strategy for a very low number of images in Fig. 10. In this experiment, we use two models (ACM and FSDNet), and the experimental settings are the same as those in Table III, but the maximum number of training epochs is reduced to 100. Under extreme conditions, the DA algorithm still results in different degrees of improvement for both models. At a training set proportion of 0.1 (53 images), ACM increases the IoU metric by 6.2 and the nIoU metric by 2.7. In the other three proportions, the ACM model also has an average increase of 2.62 points and 2.20 points on IoU and nIoU metrics, respectively. The PGDA strategy does not perform as well on the FSDNet model as it does on the ACM model, because its focal loss [54] can address the part of the sample imbalance problem. But it also gets a large improvement with four different training set proportions. In summary, the PGDA strategy still recombines context and target to increase the diversity of the extreme small sample data, which enhances the performance of DL methods.

D. Ablation Study

We set up four groups of ablation experiments to demonstrate the effectiveness of the proposed PGDA in Table V.

### TABLE IV

| Methods                | IoU(%) | nIoU(%) |
|------------------------|--------|---------|
| Baseline [26]          | 62.14  | 60.66   |
| Oversampling [40]      | 59.48  | 61.53   |
| Cut-paste-learn [39]   | 62.65  | 62.21   |
| Geometric transform    | 63.25  | 62.97   |
| Mixup [33]             | 63.92  | 64.24   |
| PGDA                   | 65.08  | 66.87   |

Note: The baseline method is DNANet. The algorithm does not add two modules default to the original data. If we follow an algorithm similar to the Copy-Paste paradigm for data augmentation, we refer to the results of the copy-paste strategy in Table IV. Bold indicates the best results.

### TABLE V

| Modules | Gaussian blur | SCR perception | Subfigure Split | IoU(%) | nIoU(%) |
|---------|---------------|----------------|----------------|--------|---------|
| ACM     | ✓             | ✓              | ✓              | 59.48  | 61.53   |
| FSDNet  | ✓             | ✓              | ✓              | 60.61  | 61.81   |
| PGDA    | ✓             | ✓              | ✓              | 63.89  | 65.13   |
|         | ✓             | ✓              | ✓              | 65.08  | 66.87   |

Note: The baseline method is DNANet. The algorithm does not add two modules default to the original data. If we follow an algorithm similar to the Copy-Paste paradigm for data augmentation, we refer to the results of the copy-paste strategy in Table IV. Bold indicates the best results.
In order to illustrate the difference between PGDA and copy-paste algorithms, we split out the Gaussian Blur module in the target enhancement module. When only the Gaussian Blur module works, the proposed PGDA algorithm degenerates into the copy-paste [40] algorithm. When the SCR perception module dynamically assigns grayscale values, the synthesized image increases the descriptive power of the target by conforming to the physical representation of the target. This makes the copy-paste algorithm improve the IoU indicator by one point. When only the subfigure split module works, the original background images are replaced with multiscale cropped subfigures. The algorithm has achieved a great improvement in the two indicators of IoU and nIoU. The multiscale subgraph not only realizes the spatial reorganization of the target and background but also realizes the DA of random cropping. When the two components act simultaneously, the strategy aligns the background but also realizes the DA of random cropping. When only the subfigure split module works, the original background images are replaced with multiscale cropped subfigures. The algorithm has achieved a great improvement in the two indicators of IoU and nIoU. The multiscale subgraph not only realizes the spatial reorganization of the target and background but also realizes the DA of random cropping.

VI. CONCLUSION

In this work, a PGDA strategy has been proposed to expand the diversity of training samples indirectly without introducing additional data. The two components of PGDA (target enhancement component and subfigure split module) successfully decouple target description capability and target localization capability by introducing prior knowledge of the training set. The strategy effectively preserves the scale distribution of the dataset and the actual representation of small targets. A new MSISTD dataset is constructed. The number of images and the number of instances in MSISTD is 2.4 times and 2.5 times than those of the existing largest ISTD dataset SIRST, respectively. It consists of 1077 images with 1343 instances. Moreover, it measures the performance gap between current data-driven and model-driven methods, while accurately evaluating DL methods. Ablation experiments, small-sample training, and comparison experiments on the MSISTD dataset and SIRST dataset demonstrate that the proposed PGDA is sufficiently robust and effective for the number of datasets and types of models.

Despite the demonstrated benefits, the existing algorithms can only locate but not identify the target. The strategies to improve the infrared image quality and the algorithms to extract target classification features in ISTD are still to be studied in the future.

REFERENCES

[1] M. Zhao, W. Li, L. Li, J. Hu, P. Ma, and R. Tao, “Single-frame infrared small-target detection: A survey,” IEEE Geosci. Remote Sens. Mag., vol. 10, no. 2, pp. 87–119, Jun. 2022.
[2] B. Zhao, C. Wang, Q. Fu, and Z. Han, “A novel pattern for infrared small target detection with generative adversarial network,” IEEE Trans. Geosci. Remote Sens., vol. 59, no. 5, pp. 4481–4492, May 2021.
[3] L. I. Voicu, R. Patton, and H. R. Myler, “Detection performance prediction on ir images assisted by evolutionary learning,” Proc. SPIE, vol. 3699, pp. 292–299, 1999.
[4] M. R. Azimi-Sadjadi and H. Pan, “Two-dimensional block diagonal LMS adaptive filtering,” IEEE Trans. Signal Process., vol. 42, no. 9, pp. 2420–2429, Sep. 1994.
[5] J. Peng and W. Zhou, “Infrared background suppression for segmenting and detecting small target,” ACTA Electronica Sinica, vol. 27, no. 12, pp. 47–51, 1999.
[6] V. T. Tom, T. Peli, M. Leung, and J. E. Bondaryk, “Morphology-based algorithm for point target detection in infrared backgrounds,” in Proc. Signal Process. Data Process. Small Targets, 1993, vol. 1954, pp. 2–11.
[7] C. P. Chen, H. Li, Y. Wei, T. Xia, and Y. Y. Tang, “A local contrast method for small infrared-target detection,” IEEE Trans. Geosci. Remote Sens., vol. 52, no. 1, pp. 574–581, Jan. 2014.
[8] X. Wang, Z. Peng, P. Zhang, and Y. He, “Infrared small target detection via nonnegativity-constrained variational mode decomposition,” IEEE Geosci. Remote Sens. Lett., vol. 14, no. 10, pp. 1700–1704, Oct. 2017.
[9] J. Han, S. Moradi, I. Faramarzi, C. Liu, H. Zhang, and Q. Zhao, “A local contrast method for infrared small-target detection utilizing a tri-layer window,” IEEE Geosci. Remote Sens. Lett., vol. 17, no. 10, pp. 1822–1826, Oct. 2020.
[10] J. Han et al., “Infrared small target detection based on the weighted strengthened local contrast measure,” IEEE Geosci. Remote Sens. Lett., vol. 18, no. 9, pp. 1670–1674, Sep. 2021.
[11] C. Gao, D. Meng, Y. Yang, Y. Wang, X. Zhou, and A. G. Hauptmann, “Infrared patch-image model for small target detection in a single image,” IEEE Trans. Image Process., vol. 22, no. 12, pp. 4996–5009, Dec. 2013.
[12] L. Zhang and Z. Peng, “Infrared small target detection based on partial sum of the tensor nuclear norm,” Remote Sens., vol. 11, no. 4, 2019, Art. No. 382, doi: 10.3390/rs11040382.
[13] Y. Dai and Y. Wu, “Reweighted infrared patch-mensor model with both non-local and local priors for single-frame small target detection,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 10, no. 8, pp. 3752–3767, Aug. 2017.
[14] M. Zhao, L. Li, W. Li, R. Tao, L. Li, and W. Zhang, “Infrared small-target detection based on multiple morphological profiles,” IEEE Trans. Geosci. Remote Sens., vol. 59, no. 7, pp. 6077–6091, Jul. 2021.
[15] M. Zhao, W. Li, L. Li, P. Ma, Z. Cai, and R. Tao, “Three-order tensor creation and Tucker decomposition for infrared small-target detection,” IEEE Trans. Geosci. Remote Sens., vol. 60, 2022, Art. No. 5000216.
[16] Y. Chen, G. Zhang, Y. Ma, J. U. Kang, and C. Kwan, “Infrared small target detection based on fast adaptive masking and scaling with iterative segmentation,” IEEE Geosci. Remote Sens. Lett., vol. 19, 2022, Art. No. 7000605.
[17] T. Zhang, Z. Peng, H. Wu, Y. He, C. Li, and C. Yang, “Infrared small target detection via self-regularized weighted sparse model,” Neurocomputing, vol. 420, pp. 124–148, 2021.
[18] B. McIntosh, S. Venkataramanan, and A. Mahalanobis, “Infrared target detection in cluttered environments by maximization of a target to clutter ratio (TCTR) metric using a convolutional neural network,” IEEE Trans. Aerosp. Electron. Syst., vol. 57, no. 1, pp. 485–496, Feb. 2021.
[19] H. Wang, L. Zhou, and L. Wang, “Miss detection vs. false alarm: Adversarial learning for small object segmentation in infrared images,” in Proc. IEEE/CVF Int. Conf. Comput. Vis., Seoul, South Korea, 2019, pp. 8508–8517.
[20] A. Desovitski et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” in Proc. 9th Int. Conf. Learn. Representations, 2020, pp. 1–21.
[21] Z. Liu et al., “Swin transformer: Hierarchical vision transformer using shifted windows,” in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2021, pp. 9992–10002.
[22] D. Hong et al., “Spectralformer: Rethinking hyperspectral image classification with transformers,” IEEE Trans. Geosci. Remote Sens., vol. 60, 2022, Art. no. 5518615.
[23] F. Liu, C. Gao, F. Chen, D. Meng, W. Zuo, and X. Gao, “Infrared small-target detection with transformer under complex backgrounds,” 2021, arXiv:2109.14379.
[24] F. Chen et al., “Local patch network with global attention for infrared small-target detection,” IEEE Trans. Aerosp. Electron. Syst., vol. 58, no. 5, pp. 3979–3991, Oct. 2022.
[25] X. Wu, D. Hong, J. Tian, J. Chanussot, W. Li, and R. Tao, “ORSIm detector: A novel object detection framework in optical remote sensing imagery using spatial-frequency channel features,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 7, pp. 5146–5158, Jul. 2019.
[26] B. Li et al., “Dense nested attention network for infrared small target detection,” 2021, arXiv:2106.00487.
[27] X. Tong, B. Sun, J. Wei, Z. Zuo, and S. Su, “Eauu-net: Enhanced asymmetric attention u-net for infrared small target detection,” Remote Sens., vol. 13, no. 16, 2021, Art. no. 3200, doi: 10.3390/rs13163200.

[28] Y. Dai, Y. Wu, F. Zhou, and K. Barnard, “Attentional local contrast networks for infrared small target detection,” IEEE Trans. Geosci. Remote Sens., vol. 59, no. 11, pp. 9813–9824, Nov. 2021.

[29] T. Zhang, S. Cao, T. Pu, and Z. Peng, “Apncnet: Attention-guided pyramid context networks for infrared small target detection,” 2021, arXiv:2111.03580.

[30] Y. Dai, Y. Wu, F. Zhou, and K. Barnard, “Asymmetric contextual modulation for infrared small target detection,” in Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis., 2021, pp. 950–959.

[31] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “Yolov4: Optimal speed and accuracy of object detection,” 2020, arXiv:2004.10934.

[32] S. Yun, D. Han, S. J. Oh, S. Chun, J. Che, and Y. Yoo, “Cutmix: Regularization strategy to train strong classifiers with localizable features,” in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2019, pp. 6023–6032.

[33] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, “mixup: Beyond empirical risk minimization,” in Proc. 6th Int. Conf. Learn. Representations, 2018, pp. 1–13.

[34] J.-H. Kim and Y. Hwang, “GAN-based synthetic data augmentation for infrared small target detection,” IEEE Trans. Geosci. Remote Sens., vol. 60, no. 1, Feb. 2022, Art. no. 5002512.

[35] G. Ghiasi et al., “Simple copy-paste is a strong data augmentation method for instance segmentation,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2021, pp. 2917–2927.

[36] C. Chen, Y. Zhang, Q. Li, S. Wei, and J. Dong, “RNNet: A hybrid detector for object detection in drone-captured images,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop, 2019, pp. 100–108.

[37] H.-S. Fang, J. Sun, R. Wang, M. Gou, Y.-L. Li, and C. Lu, “InstaBoost: Boosting instance segmentation via probability map guided copy-pasting,” in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2019, pp. 682–691.

[38] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” in Proc. IEEE Int. Conf. Comput. Vis., 2017, pp. 2980–2988.

[39] D. Dwivedi, I. Misra, and M. Hebert, “Cut, paste and learn: Surprisingly easy synthesis for instance detection,” in Proc. IEEE Int. Conf. Comput. Vis., 2017, pp. 1310–1319.

[40] M. Kisantal, Z. Wojna, J. Murawski, J. Naruniec, and K. Cho, “Augmentation for small object detection,” in Proc. 9th Int. Conf. Adv. Comput. Inf. Tech., vol. 9, no. 17, pp. 1–15, Dec. 2019.

[41] H. Deng, X. Sun, M. Liu, C. Ye, and X. Zhou, “Small infrared target detection based on weighted local difference measure,” IEEE Trans. Geosci. Remote Sens., vol. 54, no. 7, pp. 4204–4214, Jul. 2016.

[42] L. Liangkui, W. Shaoyou, and T. Zhongxing, “Using deep learning to detect small targets in infrared oversampling images,” J. Syst. Eng. Electron., vol. 29, no. 5, pp. 947–952, Oct. 2018.

[43] N. Arad, “Grid-distortion on nonrectangular grids,” Comput. Aided Geom. Des., vol. 15, no. 5, pp. 475–493, 1998. Available: https://www.sciencedirect.com/science/article/pii/S016783969800003X.

[44] M. Gabrani and O. J. Tretiak, “Elastic transformations,” in Proc. Conf. Rec. 30th Asilomar Conf. Signals, Commun. and Image Anal., vol. 10, no. 3, pp. 465–483, 2006.

[45] F. Svanström, C. Englund, and F. Alonso-Fernandez, “Real-time drone detection and tracking with visible, thermal and acoustic sensors,” in Proc. IEEE 25th Int. Conf. Pattern Recognit., 2021, pp. 7265–7272.

[46] B. Hui et al., “A dataset for infrared detection and tracking of dim-small aircraft targets under ground/air background,” China Sci. Data, vol. 5, no. 3, pp. 5966–5978, 2020.

[47] X. Sun et al., “A dataset for small infrared moving target detection under clutter background,” China Sci. Data, to be published, doi: 10.11922/cs-data.2021.0015.zh.

[48] H. Deng, X. Sun, M. Liu, C. Ye, and X. Zhou, “Infrared small-target detection using multiscale gray difference weighted image entropy,” IEEE Trans. Aerosp. Electron. Syst., vol. 52, no. 1, pp. 60–72, Feb. 2016.

[49] T. Zhang, H. Wu, Y. Liu, L. Peng, C. Yang, and Z. Peng, “Infrared small target detection based on non-convex optimization with Lp-norm constraint,” Remote Sens., vol. 11, no. 5, 2019, Art. no. 559. [Online]. Available: https://www.mdpi.com/2072-4292/11/5/S559.

[50] Y. Qin, L. Bruzzone, C. Gao, and B. Li, “Infrared small target detection based on facet kernel and random walker,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 9, pp. 7104–7118, Sep. 2019.

[51] J. Wang et al., “Deep high-resolution representation learning for visual recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 10, pp. 3349–3364, Oct. 2021.

[52] H.-S. Fang, J. Sun, R. Wang, M. Gou, Y.-L. Li, and C. Lu, “InstaBoost: Boosting instance segmentation via probability map guided copy-pasting,” in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2019, pp. 682–691.

[53] A. Wang, W. Li, X. Wu, Z. Huang, and R. Tao, “Mplanet: Multi-patch attention for infrared small target object detection,” in Proc. IEEE Int. Geosci. Remote Sens. Symp., 2022, pp. 3095–3098.

[54] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2017, pp. 2980–2988.
Xin Wu (Member, IEEE) received the M.Sc. degree in computer science and technology from the College of Information Engineering, Qingdao University, Qingdao, China, in 2014, and the Ph.D. degree in information and communication engineering from the School of Information and Electronics, Beijing Institute of Technology (BIT), Beijing, China, in 2020. She is currently an Assistant Professor with the School of Computer Science, Beijing University of Posts and Telecommunications (BUPT), Beijing, China. Her research interests include signal/image processing, fractional Fourier transform, deep learning and their applications in biometrics, and geospatial object detection. Dr. Wu was the recipient of the Best Reviewer Award of the IEEE JSTARS in 2022 and the Jose Bioucas Dias award for recognizing the outstanding paper at the Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS) in 2021.

Feiran Jie is a Senior Engineer of science and technology with Electro-Optic Control Laboratory, Luoyang, China. His research interests include image processing and machine vision.

Ran Tao (Senior Member, IEEE) received the B.S. degree from the Electronic Engineering Institute of PLA, Hefei, China, in 1985, and the M.S. and Ph.D. degrees from the Harbin Institute of Technology, Harbin, China, in 1990 and 1993, respectively. He was a Senior Visiting Scholar with the University of Michigan, Ann Arbor, MI, USA, and the University of Delaware, Newark, DE, USA, in 2001 and 2016, respectively. He is a Professor with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China. He has authored or coauthored three books and more than 100 peer-reviewed journal articles. His research interests include fractional Fourier transform and its applications, theory, and technology for radar and communication systems. Dr. Tao is a Member of the Wireless Communication and Signal Processing Commission, URSI. He is a Fellow of the Institute of Engineering and Technology and the Chinese Institute of Electronics. He was the recipient of the National Science Foundation of China for Distinguished Young Scholars in 2006, the First Prize of Science and Technology Progress in 2006 and 2007 awarded by the Ministry of Education, the Distinguished Professor of Changjiang Scholars Program in 2009, and the First Prize of Natural Science in 2013 awarded by the Ministry of Education. He was a Chief-Professor of the Program for Changjiang Scholars and Innovative Research Team in University from 2010 to 2012 and has been a Chief-Professor of the Creative Research Groups with the National Natural Science Foundation of China since 2014. He is the Vice-Chair of the IEEE China Council and the International Union of Radio Science (URSI) China Council.