Editorial

Remote Sensing of Target Object Detection and Identification II

Paolo Tripicchio (ID)

Institute of Mechanical Intelligence, Department of Excellence in Robotics and AI, Scuola Superiore Sant'Anna, 56124 Pisa, Italy; paolo.tripicchio@santannapisa.it

The ability to detect and identify target objects from remote images and acquisitions is paramount in remote sensing systems for the proper analysis of territories. The field of applying such a technology spans environmental [1] and urban [2] monitoring, hazard and disaster management, and defense and military applications. The existing literature has taken advantage of the large amounts of data acquired by sensors mounted on satellite, airborne, and unmanned aerial vehicle (UAV) platforms. While satellite imaging is still the foremost source of data, as also confirmed by the contributions collected in this Special Issue, UAV platforms have had exponential growth in recent years [3] and, given their low-cost effectiveness, this has allowed, and will allow in future, the acquisition and coverage of a wide range of environments exploiting customized setups and coverage algorithms [4]. Research applications exploit different phenomena and technologies, which include synthetic aperture radar (SAR) [5] imaging, multispectral and hyperspectral imaging, and images (or videos) acquired in the visible and near-infrared (VNIR) wavelength ranges. With the recent improvements in the sensing technologies regarding their spatial resolution and spectral content, and with the rapid development of artificial intelligence techniques that exploit convolutional neural networks (CNNs) or deep neural networks (DNNs), the results that novel approaches will achieve in the near future are promising.

The articles belonging to this Special Issue provide a comprehensive overview of the advancements, challenges, and future trends in object detection and tracking, with a particular focus on remote sensing applications. They discuss a wide range of topics, including different types of targets (e.g., ships, small targets), imaging modalities (e.g., optical, SAR, infrared), image processing techniques, and deep learning algorithms.

This editorial attempts to summarize the novelties and drawbacks of the methods and studies presented by the contributors in the context of current research trends, and also considering future developments.

A group of articles discusses different aspects of ship detection in remote sensing images, including challenges, advancements, and datasets. These sources specifically focus on ship detection in SAR images, which poses unique challenges due to the presence of speckle noise and the need for robust algorithms that can handle different ship sizes and orientations. Another group addresses the problem of detecting small targets in infrared images, which is a complex task due to the small size of the targets, low contrast with the background, and the presence of noise and clutter. A third group focuses on target tracking in image sequences, which involves estimating the trajectory of a target over time. This is particularly useful in applications such as surveillance and navigation. All contributions refer to the use of various image processing techniques to either enhance the quality of images or extract meaningful information. Examples of these techniques include background suppression, edge enhancement, and Hough transformation [6]. Many of the sources discuss the use of deep learning algorithms, particularly convolutional neural networks (CNNs) and Transformers [7], for object detection and tracking tasks, and several of them highlight the importance of evaluating the performance of detection and tracking algorithms using appropriate metrics and datasets.

A detailed summary of the novelty and drawbacks of each contribution, whose list can be found at the end of this editorial, is provided to introduce and discuss the current challenges and future development in remote sensing object detection and tracking.

The first contribution is a survey that offers a comprehensive overview of the technologies, challenges, and prospects of ship detection in optical images obtained through remote sensing. The article examines various ship detection technologies in chronological order, dividing them into traditional methods, methods based on convolutional neural networks (CNNs), and methods based on Transformers. The advantages and disadvantages of each category of method are analyzed. The article particularly focuses on the challenges that arise from ship detection in optical images obtained from remote sensing, which are mainly: complex marine environments, in which images can be influenced by factors such as light, weather, and the presence of objects other than ships, making identification difficult; the presence of insufficient discriminating features, since ships often occupy very small areas in images, and this makes it difficult to extract distinctive features; the problems of large scale variations, density distribution, and rotation, since ships can have very different sizes in images, can be very close to each other, and are probably oriented in different directions; the presence of large aspect ratios, since ships often have an elongated shape; and the problem of imbalance between positive and negative samples, since images usually contain many more background areas than ships. For each challenge, the article reviews and analyzes the solutions proposed in the literature, mainly based on CNNs, and highlights their advantages and disadvantages. In addition, the article presents a collection of public datasets of optical images for ship detection, offering detailed information on their data distribution, such as the number of ships and their size in pixels. The performance of different detection models on these datasets is compared, and the effects of the different optimization strategies to address the challenges of ship detection are analyzed. Finally, the article explores the application of Transformers in ship detection, comparing their feature extraction capability with that of CNNs. The results show that Transformers, thanks to their ability to model long-range dependencies, have great potential in this field.

Xia et al. (Contribution 2) present a novel method for differentiating corner reflector arrays from ships in anti-ship scenarios. This distinction is crucial in naval warfare, as corner reflector arrays are often deployed as decoys to confuse enemy radar systems. Their method leverages the distinct scattering characteristics exhibited by corner reflector arrays and ships. These characteristics become more pronounced when processed using a mismatched filter with an adjusted frequency modulation slope. By modifying the frequency modulation slope of the LFM signal within the filter, the main lobe of the signal output is broadened. This broadening reduces the compression level, as compared to a matched filter, thereby accentuating the differences in scattering characteristics between ships and corner reflector arrays. Two key features are extracted from the two-dimensional range-Doppler image obtained after applying the mismatched filter. These are the variance of the width, and the variance of the intervals of regions with normalized amplitude within a specific range. These features effectively capture the differences in the spatial distribution of scattering points between the two target types, aiding in their discrimination. The extracted features are then used to train a Support Vector Machine (SVM) classifier with a Gaussian kernel. This classifier demonstrates high efficacy in distinguishing between ships and corner reflector arrays. It has to be noted, however, that the method's performance is susceptible to degradation in the presence of noise, particularly at low signal-to-noise ratios (SNRs). This vulnerability arises because the extracted features rely on the distribution of scattering points, which noise can distort. The effectiveness of the method also hinges on the selection of specific parameters, such as the Doppler factor range and the point extraction range. While the method has shown resilience to minor variations in these parameters, optimizing their selection is paramount for achieving optimal performance. The authors propose potential solutions to mitigate these limitations, including the implementation of noise suppression techniques and the refinement of parameter selections. They also suggest future research directions, such as validating the method using real-world data in operational scenarios and integrating information from other domains, such as polarization, to enhance identification accuracy.

The third contribution describes a method for detecting oriented ships in Synthetic Aperture Radar (SAR) images based on RepPoints, which are representation points that capture the object's shape and orientation, which is based on an anchor-free detection architecture consisting of two main components: Scattering-Point-Guided Adaptive Sample Selection(SPG-ASS) and SPG learning. The improved sample selection method integrates the scattering point location information to select higher-quality samples during training, thus preventing model degradation caused by low-quality samples. The SPG learning mechanism improves the quality of RepPoints in the initialization stage, enabling the network to learn more refined representations of ships' electromagnetic characteristics, while reducing land clutter interference in complex nearshore environments. The method has shown good generalization and reliability across different datasets with varying characteristics, suggesting its adaptability to practical application scenarios. Furthermore, ablation experiments demonstrate the effectiveness of the individual components, namely SPG-ASS and SPG learning, in improving detection performance. However, both the adaptive sample selection scheme and the adaptive learning part rely on extracting the scattering points from the target. If the area occupied by ships is limited, or if the scattering from the ships is weak, resulting in fewer or no corner points being extracted, the method might fail. The authors have suggested that, in the future, they will explore redesigning the scattering point extraction part and introducing more efficient and advanced network structures for scattering feature extraction and fusion.

Tian et al. (contribution 4) present a new lightweight model (LMSD-Net) for ship detection, specifically small ships, in optical remote sensing imagery. The model is designed to address challenges posed by variations in ship size, background clutter, and the limited capabilities of embedded systems. A fog simulation method is used to augment the training dataset with more foggy images. This method simulates the effect of fog on scene radiance, improving the model's robustness in adverse weather conditions. A new feature extraction module, called Efficient Layer Aggregation of C3 (ELA-C3), is introduced for more efficient information aggregation. ELA-C3 enhances feature learning without significantly increasing the number of model parameters. A feature fusion method is proposed to fuse features extracted at different scales. It utilizes learnable weights for channels during bidirectional fusion, allowing the model to focus on the most relevant information and reducing the number of parameters, as compared to the original architecture it is based on. A Contextual Transformer (CoT) block is added to the detection head to improve its localization accuracy. The CoT block combines the global relationship modeling capability of transformers with the computational efficiency of convolutional neural networks. Finally, an improved version of the CIoU loss function, called V-CIoU, is proposed to address the issue of slow convergence when the aspect ratios of the ground truth box and the predicted box are similar. V-CIoU introduces a penalty term based on the variance of aspect ratios, improving detection performance for small ships.

Contribution 5 introduces a new method for detecting small objects in remote sensing images. The method, called Multi-Vision Transformer (MVT), is based on a Transformer-like neural network, and proposes the first remote sensing dataset based on event cameras, called the Event Object Detection Dataset (EOD Dataset). This dataset consists of over 5000 event streams, and includes six object categories: cars, buses, pedestrians, bicycles, boats, and ships. MVT consists of three modules: a downsampling Module, a Channel Spatial Attention Module (CSA), and a Global Spatial Attention Module (GSA). The CSA focuses on short-range dependencies within feature maps, improving the representation of channel- and spatial location-level features. The GSA module, consisting of Window-Attention and Grid-Attention, considers long-range dependencies in the feature maps, capturing the global information and long-distance connections in a single operation. Finally, a novel cross-scale attention mechanism (Cross Deformable Attention (CDA)) that progressively merges high-level features with low-level features is proposed, reducing the computational complexity of the Transformer encoder and the entire network while preserving the original performance. The authors suggest that, in order to improve possible

loss of details due to event camera captures, a possible solution could be to combine data from the event cameras with those from traditional cameras to exploit the advantages of both technologies.

The Multiscale Feature Extraction U-Net (MFEU-Net), presented in the sixth contribution, is a convolutional neural network designed for infrared small and dim target detection (ISDT). The network's architecture is based on the U-Net structure, which enables the fusion of multiscale information through skip connections. This allows the network to have different receptive fields at different levels, improving its ability to detect targets of varying sizes. In particular, MFEU-Net utilizes a combination of Residual U-block (RSU) blocks and Inception modules to extract multiscale information. Moreover, it incorporates a multidimensional attention mechanism, which operates on both channels and space. This mechanism enables the network to focus on the important areas of the image, enhancing detection in complex scenarios and reducing false alarms. The results show superior performance compared to other ISDT detection algorithms, achieving higher detection rates, lower false alarm rates, and higher IoU values on various datasets.

Contribution 7 presents another method aimed at ISDT detection, called the Group Regularized Principle Component Pursuit (GPCP), which is a group-regularized low-rank and sparsity decomposition model. This method addresses the limitations of traditional patch-based models, such as Infrared Patch Image (IPI), which are often sensitive to strong edges and background clutter due to their failure to consider the diversity of data structure. Unlike traditional methods that utilize a single low-rank constraint for the entire background component, GPCP employs a group low-rank constraint for background estimation. This approach allows for the use of different singular value thresholds for the low-rank decomposition of image groups corresponding to different complexities. Consequently, GPCP can better explore the local structure of the image and achieve a more accurate decomposition result. By dividing image data into groups based on brightness and clutter level, GPCP can more effectively suppress background clutter, particularly in areas with strong edges. This capability is demonstrated by experimental results on various detection scenes, where the GPCP achieves higher background suppression factors, as compared to other methods. Although GPCP utilizes Singular Value Decomposition (SVD) [8], in the same way as other patch-based models, its computational complexity is lower than its baseline model, IPI. This is attributed to the grouping strategy that divides image data into smaller groups, reducing the overall computational cost of SVD decomposition. By integrating group low-rank regularization with the sparsity constraint for background and target separation, GPCP improves the detection accuracy and overcomes the limitations of traditional decomposition-based methods. However, further research is needed to optimize the grouping criteria, further reduce the computational complexity, and explore more efficient background modeling methods.

On the same topic, the Background-Suppression Proximal Gradient (BSPG) method (contribution 8) enhances detection accuracy and computational efficiency in patch-based methods, particularly in suppressing strong background edges. It incorporates a novel continuation strategy during the alternating update of low-rank and sparse components, so as to suppress strong edges that are often mistaken as targets. This strategy retains more components during the low-rank matrix update while reducing the sparse matrix's update speed, enabling the model to mitigate the influence of strong edges. Approximate Partial SVD (APSVD) is employed to expedite the resolution of the low-rank sparse decomposition problem. This approach is more efficient than the full SVD because it leverages the fact that the soft-thresholding operation utilizes only a portion of the singular values. To further enhance processing speed, BSPG employs GPU multi-thread parallelism strategies to accelerate the construction and reconstruction of patch images, which can be divided into repetitive and independent subtasks. While efforts have been made to reduce computational complexity and exploit computation parallelism, further research may enhance the time performance and possible limitations due to data dependency.

Chen et al. (contribution 9) present a novel approach for anomaly detection in hyperspectral images (HSI), named the Multi-Dimensional Low-Rank (MDLR) method. Unlike previous tensor-based methods that mainly focused on the low dimensionality of the spectral dimension, MDLR considers low dimensionality along all dimensions of HSI: width, height, and spectrum. This three-dimensional analysis allows for more comprehensive background information extraction, improving the separation between background and anomalies. To impose low-rank constraints on the background tensor, MDLR utilizes Weighted Schatten p-norm Minimization (WSNM) on the slices of the f-diagonal tensor obtained through t-SVD decomposition. This approach allows for better preservation of the low-rank structure of the background, as compared to traditional nuclear norm minimization. MDLR utilizes a norm that penalizes the anomaly tensor, promoting joint sparsity in both the spectral and spatial domains. This takes into account the fact that anomalies tend to be spatially localized and exhibit low spectral density. Finally, the authors suggest that dimensionality reduction techniques could be integrated in the future to mitigate the computational complexity due to the t-SVD.

The authors of contribution 10 describe a tracking algorithm for sonar detection, called the Improved Multi-Kernel Correlation Filter (IMKCF), which is designed to detect and track weak underwater targets in complex marine environments. This problem is particularly challenging in environments with a low signal-to-reverberation ratio, where reverberation interference can make it difficult to distinguish the target from the background noise. Although the kernel correlation filter algorithm has been successful in visual tracking, it has not previously been applied to underwater target tracking. Using weighted information from historical samples to solve the coefficients of multiple nonlinear kernels adaptively, this method addresses the problem of limited robustness in tracking single features, taking full advantage of multiple complementary features. In particular, when a tracking result is deemed unreliable, a redetection module uses the historical reliability tracking results to drive a Kalman filter, which predicts the location of the target candidate. The use of multiple features, an adaptive update of the kernel coefficients, and the inclusion of a redetection module improve the method's performance over traditional tracking algorithms.

Contribution 11 proposes a method for suppressing faint/dim background stars in infrared, based on recursive moving target indication, to enhance the detection of space targets in optical image sequences. The suppression of stars with a low signal-to-noise ratio (SNR) has been largely ignored by previous research, but can negatively impact accuracy and real-time performance, particularly for time-before-space (TBS) detection methods. Unlike other TBS methods, which are closely tied to their corresponding target detection methods, the proposed method is versatile, and can be utilized as an efficient pre-processing step for most target detection and tracking methods. Additionally, a multi-frame adaptive threshold segmentation method is put forward to create an accurate star mask, enabling the real-time suppression of dim stars.

Contribution 12 presents a multi-stage joint detection and tracking model (MJDTM) for the real-time detection of space targets, such as space debris and satellites, using optical image sequences. The authors argue that although space target surveillance is critical for aerospace safety, it is becoming increasingly difficult, due to increasingly complex space environments. This article addresses the limitations of existing approaches that struggle to suppress background noise and mostly focus on single tasks, such as detection or tracking. The model uses an improved local contrast method to extract potential small space targets in optical image sequences. It uses a star target suppression method that exploits the differences in motion relative to Earth and real-time satellite attitude data to distinguish between space and star targets. The model is implemented on a specialized heterogeneous multi-core processing platform based on FPGA and DSP to meet real-time processing requirements. The authors note that although the proposed model shows improvements in detection accuracy while maintaining real-time processing speed, it may not perform well

for targets with a low SNR. Additionally, the model relies on real-time satellite attitude data, which could be a limitation.

The SE-RRACycleGAN algorithm (contribution 13) introduces several improvements for single image deraining in an unsupervised manner. It contains an innovative Recurrent Rain-Attentive Module, designed to enhance the detection of rain-related information by concurrently considering both rainy and clean images, by not only incorporating spatial and channel attention blocks, but also employing an LSTM unit to capture spatiotemporal dependencies within images, facilitating the modeling of complex rain streak patterns that interact with the scene. The addition of Squeeze-and-Excitation (SE) blocks to the generator enables the model to learn discriminative features, facilitating the capture of intricate rain patterns and the representation of the underlying image structure. This capability is particularly significant for deraining tasks requiring both local and global features. Finally, to enhance the visual similarity between the generated image and the input image, the algorithm's loss function includes the content loss. These improvements allow SE-RRACycleGAN to surpass most state-of-the-art unsupervised methods, particularly on the Rain12 dataset and real rainy images, making it highly competitive compared to supervised techniques.

Conflicts of Interest: The author declares no conflicts of interest.

List of Contributions

1.　Zhao, T.; Wang, Y.; Li, Z.; Gao, Y.; Chen, C.; Feng, H.; Zhao, Z. Ship Detection with Deep Learning in Optical Remote-Sensing Images: A Survey of Challenges and Advances. *Remote Sens.* **2024**, *16*, 1145.
2.　Xia, L.; Wang, F.; Pang, C.; Li, N.; Peng, R.; Song, Z.; Li, Y. An Identification Method of Corner Reflector Array Based on Mismatched Filter through Changing the Frequency Modulation Slope. *Remote Sens.* **2024**, *16*, 2114.
3.　Zhao, W.; Huang, L.; Liu, H.; Yan, C. Scattering-Point-Guided Oriented RepPoints for Ship Detection. *Remote Sens.* **2024**, *16*, 933.
4.　Tian, Y.; Wang, X.; Zhu, S.; Xu, F.; Liu, J. LMSD-Net: A Lightweight and High-Performance Ship Detection Network for Optical Remote Sensing Images. *Remote Sens.* **2023**, *15*, 4358.
5.　Jing, S.; Lv, H.; Zhao, Y.; Liu, H.; Sun, M. MVT: Multi-Vision Transformer for Event-Based Small Target Detection. *Remote Sens.* **2024**, *16*, 1641.
6.　Wang, X.; Han, C.; Li, J.; Nie, T.; Li, M.; Wang, X.; Huang, L. Multiscale Feature Extraction U-Net for Infrared Dim- and Small-Target Detection. *Remote Sens.* **2024**, *16*, 643.
7.　Li, M.; Wei, Y.; Dan, B.; Liu, D.; Zhang, J. Infrared Small Dim Target Detection Using Group Regularized Principle Component Pursuit. *Remote Sens.* **2024**, *16*, 16.
8.　Hao, X.; Liu, X.; Liu, Y.; Cui, Y.; Lei, T. Infrared Small-Target Detection Based on Background-Suppression Proximal Gradient and GPU Acceleration. *Remote Sens.* **2023**, *15*, 5424.
9.　Chen, X.; Wang, Z.; Wang, K.; Jia, H.; Han, Z.; Tang, Y. Multi-Dimensional Low-Rank with Weighted Schatten p-Norm Minimization for Hyperspectral Anomaly Detection. *Remote Sens.* **2024**, *16*, 74.
10.　Yue, W.; Xu, F.; Yang, J. Tracking-by-Detection Algorithm for Underwater Target Based on Improved Multi-Kernel Correlation Filter. *Remote Sens.* **2024**, *16*, 323.
11.　Zhang, L.; Rao, P.; Hong, Y.; Chen, X.; Jia, L. Infrared Dim Star Background Suppression Method Based on Recursive Moving Target Indication. *Remote Sens.* **2023**, *15*, 4152.
12.　Su, Y.; Chen, X.; Liu, G.; Cang, C.; Rao, P. Implementation of Real-Time Space Target Detection and Tracking Algorithm for Space-Based Surveillance. *Remote Sens.* **2023**, *15*, 3156.
13.　Wedajew, G.N.; Xu, S.S.-D. SE-RRACycleGAN: Unsupervised Single-Image Deraining Using Squeeze-and-Excitation-Based Recurrent Rain-Attentive CycleGAN. *Remote Sens.* **2024**, *16*, 2642.

References

1.　Yuan, Q.; Shen, H.; Li, T.; Li, Z.; Li, S.; Jiang, Y.; Xu, H.; Tan, W.; Yang, Q.; Wang, J.; et al. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sens. Environ.* **2020**, *241*, 111716. [CrossRef]
2.　Jat, M.K.; Garg, P.K.; Khare, D. Monitoring and modelling of urban sprawl using remote sensing and GIS techniques. *Int. J. Appl. Earth Obs. Geoinf.* **2008**, *10*, 26–43. [CrossRef]
3.　Pajares, G. Overview and current status of remote sensing applications based on unmanned aerial vehicles (UAVs). *Photogramm. Eng. Remote Sens.* **2015**, *81*, 281–330. [CrossRef]

4. Tripicchio, P.; Unetti, M.; D'Avella, S.; Avizzano, C.A. Smooth Coverage Path Planning for UAVs with Model Predictive Control Trajectory Tracking. *Electronics* **2023**, *12*, 2310. [CrossRef]
5. Liu, C.A.; Chen, Z.X.; Yun, S.H.A.O.; Chen, J.S.; Hasi, T.; Pan, H.Z. Research advances of SAR remote sensing for agriculture applications: A review. *J. Integr. Agric.* **2019**, *18*, 506–525. [CrossRef]
6. Illingworth, J.; Kittler, J. A survey of the Hough transform. *Comput. Vision, Graph. Image Process.* **1988**, *44*, 87–116. [CrossRef]
7. Aleissaee, A.A.; Kumar, A.; Anwer, R.M.; Khan, S.; Cholakkal, H.; Xia, G.S.; Khan, F.S. Transformers in remote sensing: A survey. *Remote Sens.* **2023**, *15*, 1860. [CrossRef]
8. Andrews, H.; Patterson, C.L.I.I.I. Singular value decomposition (SVD) image coding. *IEEE Trans. Commun.* **1976**, *24*, 425–432. [CrossRef]