



Article

A Video SAR Multi-Target Tracking Algorithm Based on Re-Identification Features and Multi-Stage Data Association

Anxi Yu ^{1,*}, Boxu Wei ^{1,2}, Wenhao Tong ¹, Zhihua He ¹ and Zhen Dong ¹

¹ College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China; wbx@nudt.edu.cn (B.W.); twh10355@nudt.edu.cn (W.T.); zhihuahe@nudt.edu.cn (Z.H.); dongzhen@nudt.edu.cn (Z.D.)

² Chongqing Xingwang Network System Research Institute Co., Ltd., Chongqing 400015, China

* Correspondence: yu_anxi@nudt.edu.cn; Tel.: +86-137-8731-7450

Abstract: Video Synthetic Aperture Radar (ViSAR) operates by continuously monitoring regions of interest to produce sequences of SAR imagery. The detection and tracking of ground-moving targets, through the analysis of their radiation properties and temporal variations relative to the background environment, represents a significant area of focus and innovation within the SAR research community. In this study, some key challenges in ViSAR systems are addressed, including the abundance of low-confidence shadow detections, high error rates in multi-target data association, and the frequent fragmentation of tracking trajectories. A multi-target tracking algorithm for ViSAR that utilizes re-identification (ReID) features and a multi-stage data association process is proposed. The algorithm extracts high-dimensional ReID features using the Dense-Net121 network for enhanced shadow detection and calculates a cost matrix by integrating ReID feature cosine similarity with Intersection over Union similarity. A confidence-based multi-stage data association strategy is implemented to minimize missed detections and trajectory fragmentation. Kalman filtering is then employed to update trajectory states based on shadow detection. Both simulation experiments and actual data processing experiments have demonstrated that, in comparison to two traditional video multi-target tracking algorithms, DeepSORT and ByteTrack, the newly proposed algorithm exhibits superior performance in the realm of ViSAR multi-target tracking, yielding the highest MOTA and HOTA scores of 94.85% and 92.88%, respectively, on the simulated spaceborne ViSAR data, and the highest MOTA and HOTA scores of 82.94% and 69.74%, respectively, on airborne field data.

Keywords: video synthetic aperture radar; multi-target tracking; shadow detection; ReID features; multi-stage data association; cosine similarity; intersection over union similarity



Academic Editor: Andrzej Stateczny

Received: 12 February 2025

Revised: 27 February 2025

Accepted: 7 March 2025

Published: 8 March 2025

Citation: Yu, A.; Wei, B.; Tong, W.; He, Z.; Dong, Z. A Video SAR Multi-Target Tracking Algorithm Based on Re-Identification Features and Multi-Stage Data Association. *Remote Sens.* **2025**, *17*, 959. <https://doi.org/10.3390/rs17060959>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Video Synthetic Aperture Radar (ViSAR) can continuously observe imaging areas on the ground in all weather conditions to acquire echo data, thus constructing SAR image sequences. This enables the wide-ranging collection of information on the motion states of targets within the imaging area and the radiation characteristics of targets and background environments, playing a crucial role in tasks such as battlefield monitoring and cross-border reconnaissance [1–4].

ViSAR has the capability to directly detect and track ground moving targets at walking speeds in the video sequences obtained from imaging, surpassing the minimum detectable speed limitations of traditional Ground Moving Target (GMT) technologies [5]. In the ViSAR mode, there exists defocusing and shifting of the energy of moving targets, creating

clustered “shadows” at the actual positions of ground targets. By utilizing the information from moving target shadows and energy, it is possible to monitor, image, reposition moving targets, establish a battlefield situation in the observation area, and accurately grasp the deployment of critical targets on the battlefield [4].

In recent years, domestic and international scholars have conducted in-depth analyses of ViSAR imaging and the mechanisms behind the formation of moving target shadows [6–10]. They have quantitatively analyzed the performance of shadow detection, proposed a series of ViSAR moving target detection methods based on shadows, and validated the effectiveness of background differencing techniques [11,12] and machine learning methods [13–16]. Furthermore, to achieve high-performance ViSAR moving target tracking based on effective shadow detection, Chao Zhong [17] and Xu [18] proposed a motion target tracking framework based on kernel correlation filters, effectively combining the shadow features of targets in ViSAR image sequences with the energy in the range-Doppler spectrum. Wang et al. [11] studied the positive impact of target shadows on tracking low Radar Cross Section (RCS) targets in ViSAR sequences. Henke et al. [19] proposed a multi-target tracking method for ViSAR based on Kalman filtering. He Zhihua et al. [20] utilized the spatiotemporal information of ViSAR frame images, achieved stabilization through registration and denoising preprocessing to reduce the impact of speckle noise, and implemented moving target shadow detection and tracking in ViSAR image sequences based on a differential fusion method, effectively reducing false alarm rates while ensuring detection rates. Liu Yujie [21] matched and associated multiple targets with long intervals based on differences in ship image features, calculated the differences between ship target features and targets using different weight values based on constraints gates depending on speed and acceleration, and, finally, based on the principle of minimizing differences, achieved trajectory filtering and target tracking. In the realm of deep learning-based ViSAR moving target tracking, Liang Zihan et al. [22] applied the SiamMask algorithm to the task of dynamic target tracking in ViSAR images, achieving good tracking performance. Ding Jinshan et al. [23] realized robust ViSAR moving target detection and tracking through the joint detection of moving target energy and shadows in the signal domain and image domain, using Faster R-CNN and trajectory association schemes. Hu Yao [24] proposed an anchor-free multi-target tracking framework, optimizing the network structure with an attention mechanism to improve the tracking effect of ViSAR moving targets. Qiu Lei et al. [25] extracted deep features from ViSAR images using an improved lightweight EfficientDet network, predicted target detection boxes using a Bi-LSTM network, and combined data association strategies to track moving target shadows in ViSAR, also achieving good results.

However, due to factors such as speckle noise affecting ViSAR systems [20], there are often issues such as a high number of missed detections, a high number of false alarms, and fragmented tracking trajectories. Consequently, during the process of shadow-based multi-target tracking in ViSAR, there are frequent occurrences of high identity switches of moving target shadows and numerous trajectory interruptions. Therefore, accurately extracting moving target shadow features from ViSAR image sequences and improving target tracking success rates and trajectory filtering accuracy have always been the continuous goals pursued by scholars [23]. When targets are briefly occluded, the predicted results of the Kalman filter can typically be used as the current state of the occluded target; however, in cases of significant changes in target motion states or prolonged occlusion periods, the uncertainty of target motion significantly increases, rendering the predictions of the Kalman filter unreliable and leading to a significant decrease in the accuracy of multi-target data association. Robust descriptions of ViSAR target shadow depth features can effectively improve the quality of multi-target data association. Researchers have introduced popular

re-identification features from the field of computer vision, attempting to reassociate trajectories in a lost state with detection results through similar distance matching and Mahalanobis distance adjustment, thereby increasing the success rate of reassociating lost state trajectories and achieving success in the field of deep learning-based optical video multi-target tracking [26–28]. Bewley [29] first proposed SORT (simple online and realtime tracking) to achieve high-precision target tracking, while DeepSORT (Simple Online and Realtime Tracking with a Deep Association Metric) added the appearance embedding branch based on SORT for target matching in the data association part, enabling accurate matching of targets even after long-term occlusion and disappearance [30]. Most tracking algorithms choose to discard target detection boxes with confidence levels below a certain threshold during the object detection stage, which results in tracking interruptions for real targets with temporarily low confidence levels due to occlusion or scattering changes. The ByteTrack (Multi-object tracking by associating every detection box) algorithm [31] improves the data association method by recalling low score detection boxes for data association while maintaining the tracking of high score detection boxes. It effectively utilizes low score detection, reduces the number of ID switches and missed detections, and obtains better tracking results.

Inspired by this, this paper proposes a ViSAR multi-target tracking method based on re-identification (ReID) features and multi-stage data association. This method treats ViSAR moving target shadow detection as a preliminary task for multi-target tracking, using the detected candidate shadow image slices as input. It generates a multi-target data association cost matrix by combining ReID feature cosine similarity and Intersection over Union (IoU) similarity and adopts a multi-stage data association strategy based on differences in shadow detection confidence, effectively reducing the risk of missed detections in moving targets and target trajectory interruptions. Section 2 introduces the algorithm of this paper from the aspects of shadow ReID feature extraction, multi-stage data association, and trajectory filtering.

2. Methods

2.1. Algorithm Framework

The high-quality data association between the ViSAR moving target shadow detection results and the current multi-target trajectory predictions of the tracking filter is a crucial link in ViSAR multi-target tracking. Because ViSAR moving target tracking algorithms rely on target shadow detection results based on weak features rather than the target bodies themselves, traditional video multi-target data association and tracking algorithms often lead to a high number of trajectory fragments, low tracking success rates, and trajectory continuity. One reason for this is the significant random variability in the characteristics of ViSAR moving target shadows. In ViSAR images, the size of moving target shadows and the clutter-to-noise ratio of shadows vary with the speed of target motion. Some detections may receive low similarity scores due to significant variations in the size of detection boxes between adjacent frames, leading traditional algorithms to inaccurately match detection results with target trajectories based on IoU similarity of detection boxes. This results in significantly increased rates of missed associations, erroneous associations, and an increased number of trajectory interruptions, manifesting as the generation of numerous trajectory fragments. Another reason is that traditional algorithms directly discard low-confidence target shadows, making trajectories prone to breaking. However, many factors, such as variable ground vehicle target motion, complex time-space variable target electromagnetic scattering characteristics, synthetic aperture energy accumulation effect of SAR imaging, and the formation of speckle noise, determine that the shadow

features of targets in ViSAR images are variable, and real moving targets in ViSAR images are often evaluated as low confidence shadow detection results because of weak features.

Therefore, it is necessary to selectively and scientifically use low confidence detection to participate in the process of multi-target data association and tracking filtering and give full play to the advantages of the data processing framework of the integration of detection and tracking, so as to improve the accuracy and continuity of trajectory tracking. To address the aforementioned issues, this algorithm incorporates the following two optimization designs:

1. Introducing ReID features to improve the success rate of multi-target data association based on IoU similarity. Since ViSAR typically employs overlapped imaging, the characteristics of adjacent frame target shadows change relatively minimally. Additionally, clutter scatters from objects in shadow detection slices can provide feature information. By extracting the appearance information of the shadow characteristic region of the moving target in the ViSAR, the ReID feature can effectively distinguish the similar-looking targets and maintain a stable identity association through cosine similarity calculation, which significantly improves the robustness of the data association stage. Hence, combining ReID feature similarity with IoU similarity for multi-target data association can effectively mitigate the issue of false associations caused by the variability in characteristics of ViSAR moving targets.
2. Designing a multi-stage data association strategy based on shadow detection confidence levels to fully exploit valuable information contained in low-confidence detection results, addressing the problem of trajectory interruptions due to a high proportion of discarded low-confidence detection results. Drawing inspiration from the ByteTrack algorithm, by recalling some low-confidence shadow detections for multi-stage data association, categorizing high and low-confidence shadow detection results, classifying target trajectories into four categories, initial, confirmed, lost, and deleted trajectories, associating multi-target shadow detections and trajectories in multiple stages according to a certain priority, more valuable information is obtained. This approach effectively reduces the number of target identity switches, missed alarms, and the risk of target trajectory interruptions, thus enhancing trajectory continuity.

The overall algorithm framework is illustrated in Figure 1. The ViSAR moving target shadow detection network utilizes a deep learning network based on an enhanced YOLOv5s, which captures ViSAR moving target shadow detection image slices frame by frame along with their confidence, serving as input for the multi-target tracking algorithm. The ViSAR multi-target tracking algorithm based on ReID features and multi-stage data association comprises three main parts. Firstly, it involves confidence judgment for shadow detection, categorizing ViSAR shadow detections with confidence greater than or equal to 0.6 as high-score detections and those with confidence less than 0.6 but greater than 0.1 as low-score detections. The video multi-target tracking filter uses the classical Kalman filter to filter and predict target motion states and classifies target trajectories into four categories: initial, confirmed, lost, and deleted trajectories. The core of the algorithm lies in achieving multi-target data association between different shadow detection slices of the current frame and various target trajectories except for deleted trajectories. A multi-stage data association strategy is employed, sequentially matching high-score and low-score detections in four stages with different trajectory states with a certain priority. The similarity measure used in each stage of association is determined based on the characteristics of the matching objects, selecting ReID feature similarity or IoU similarity to construct the similarity cost matrix for the Hungarian matching algorithm, aiming to enhance the accuracy of multi-target data association, reduce false trajectory generation, lower the missed detection rate,

mitigate the risk of trajectory interruptions, and comprehensively improve multi-target tracking continuity.

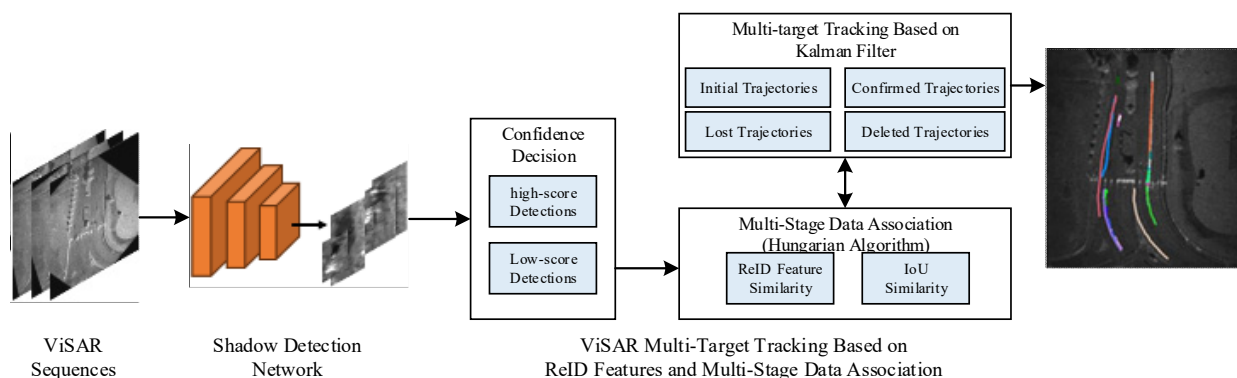


Figure 1. Algorithm framework of this paper.

2.2. Enhanced YOLOv5s Moving Target Shadow Detection Network

The features of ViSAR moving target shadows are weak and challenging to distinguish from the background. In recent years, both domestic and international scholars have extensively investigated ViSAR moving target detection and tracking methods, employing traditional background subtraction image processing techniques and deep learning-based intelligent approaches [11,16,20,22,32]. Despite these efforts, further theoretical and technical advancements are necessary to enhance algorithm accuracy, robustness, and timeliness. Several factors, including the scarcity of scattering characteristics of moving targets, the diversity of target motion states, complex cluttered backgrounds, static target shadows, and speckle noise, contribute to the overall weak features, wide scale range, and strong clutter interference exhibited by moving target shadows in ViSAR images [7,9,20]. These challenges significantly impede high-quality detection of moving target shadows.

Convolutional neural networks (CNNs) can capture grayscale features, edge details, and spatial positions of moving target shadows in ViSAR, enhancing detection capabilities compared to traditional methods [33]. However, two-stage detection networks such as CNN, Fast-RCNN, and Faster-RCNN, while accurate, have complex structures, numerous parameters, and high computational demands, making them impractical for real-time video processing and hardware deployment [34]. To balance the efficiency of ViSAR moving target shadow detection algorithms, single-stage network detection performance is often selected. In contrast, the single-stage YOLO algorithm [35], introduced in 2016, frames object detection as a regression problem, predicting bounding box parameters and performing classification directly on the output feature map. This approach notably accelerates network training and inference. Balancing model accuracy and network complexity, in this section, this paper opts for the YOLOv5s model [36] within the practical YOLOv5 framework, featuring a shallower depth, for optimization design.

To tackle issues like strong background interference, subtle moving target shadows, and multi-scale shadow characteristics in ViSAR images, a lightweight detection network for moving target shadows is proposed, built upon the enhanced YOLOv5s framework (Figure 2). The network enhancements encompass: (1) A Simplified Generalized Feature Pyramid Network (GFPN-S) that lightens the feature pyramid network, enriching semantic information in ViSAR via cross-layer connections. This enhancement bolsters the network's ability to extract target shadow features while significantly cutting down parameters by selectively pruning inter-module connections, without compromising detection performance. (2) An Improved Simplified Fast Spatial Pyramid Pooling (SPPFS) module that minimizes information loss during pooling, thereby enhancing detection accuracy. (3) The

incorporation of a decoupled head design to mitigate the impact of discrepancies between classification and position prediction tasks on network detection performance.

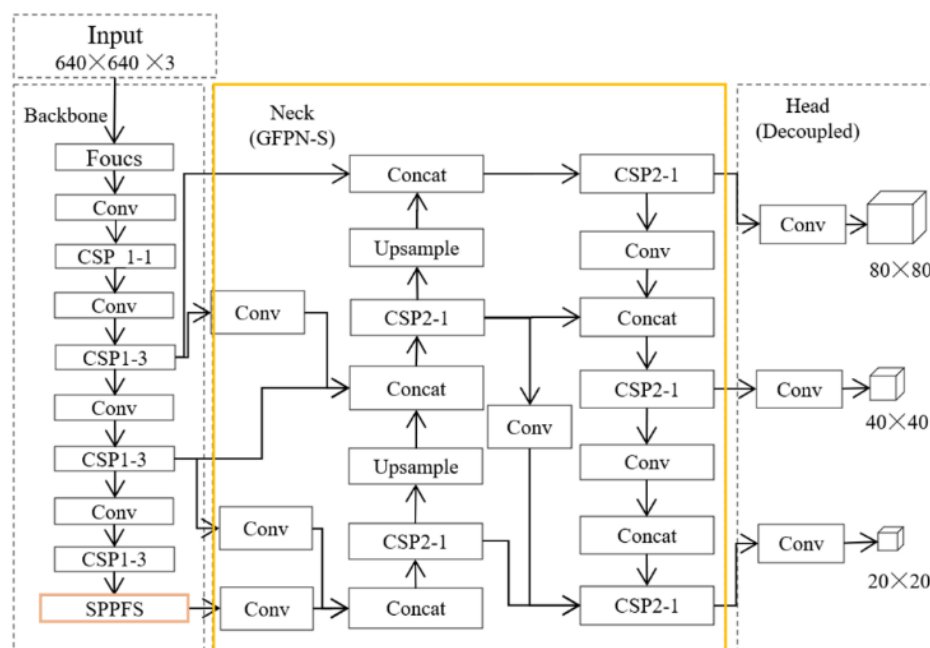


Figure 2. Enhanced YOLOv5s ViSAR moving target shadow detection network.

Airborne ViSAR data from Sandia National Laboratories [37] are used to verify and analyze the shadow detection results of moving targets in ViSAR. Figure 3 showcases the feature maps output by the network both before and after the enhancement. The results clearly demonstrate that, following the network improvement, the distinguishing features of target shadows are more pronounced, leading to a significant improvement in their differentiation from the background.

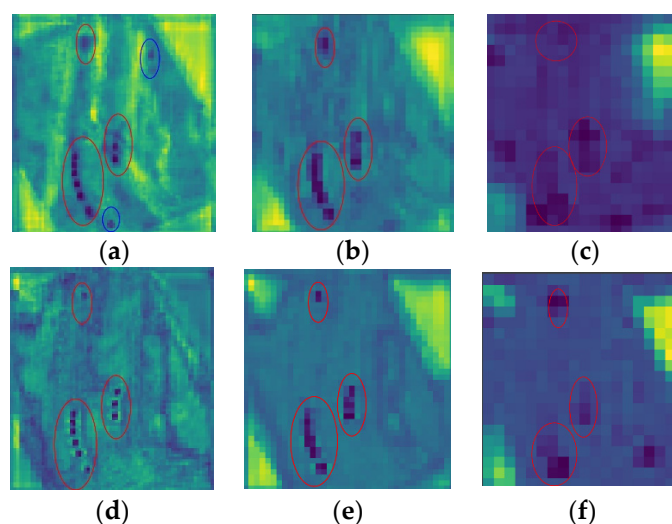


Figure 3. Feature maps output of the network before and after the enhancement (Frame 200). (a) Small-scale prediction layer. (b) Medium-scale prediction layer. (c) Large-scale prediction layer. (d) Small-scale prediction layer (enhanced). (e) Medium-scale prediction layer (enhanced). (f) Large-scale prediction layer (enhanced).

The shadow detection results, as depicted in Figure 4, provide a clear comparison of the performance of the enhanced network versus the original YOLOv5s network. In

Figure 4, red indicates detection results, blue boxes mark false alarms, yellow boxes mark missed detections, and green boxes represent low-confidence detections with confidence levels below 0.5. The enhanced network notably reduces the number of false alarms and missed detections while enhancing the overall confidence of the detection results.

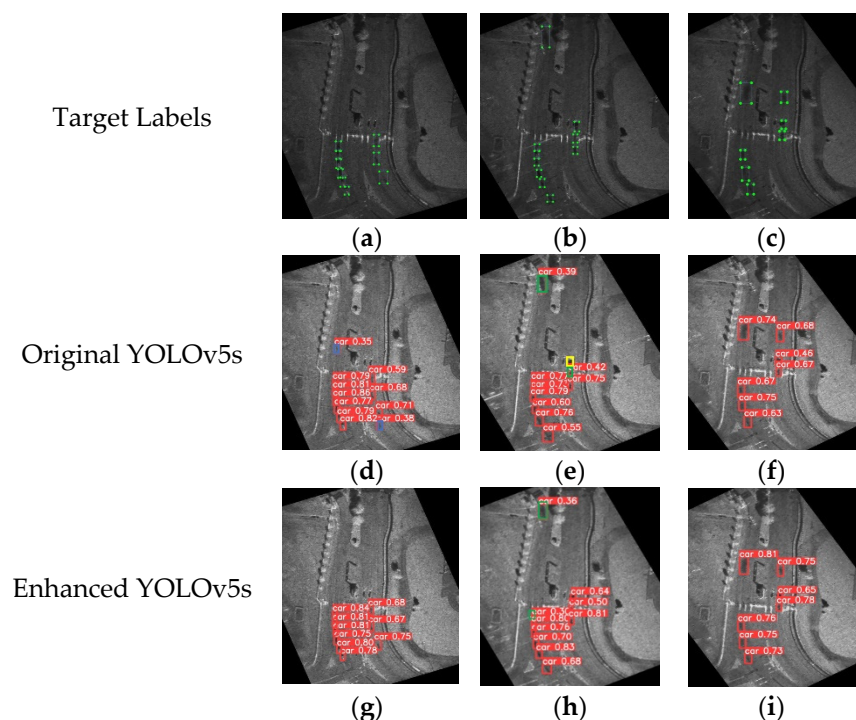


Figure 4. Shadow detection results of airborne ViSAR data. (a) frame151. (b) frame200. (c) frame250. (d) frame151. (e) frame200. (f) frame250. (g) frame151. (h) frame200. (i) frame250.

Table 1 shows the ablation results of the enhanced YOLOv5s network. The enhanced ViSAR moving target shadow detection network exhibits a series of good detections metrics, including a precision (R) of 93.4%, a recall rate (R) of 94.1%, a mean Average Precision (mAP) of 95.8%, and a Frame Per Second (FPS) rate of 47.3.

Table 1. Ablation experiment of enhanced YOLOv5s network.

GFPN-S	-	✓	✓	✓	✓
SPPFS	-	-	✓	-	✓
Decoupled-Head	-	-	-	✓	✓
P	85.8%	88.0%	93.5%	90.4%	93.4%
R	88.7%	91.5%	92.1%	92.9%	94.1%
mAP	76.1%	94.1%	94.1%	94.6%	95.8%
FPS	53.4	51.5	49.4	50.2	47.3

It can be seen that, although methods like contrast enhancement and network optimization have been used to enhance shadow feature extraction, there are still many low-confidence shadows in the detection results. The distribution of 1163 correctly detected moving target shadow results in the validation set at different detection confidence levels are analyzed, as shown in Figure 5. It is evident that there are a significant number of correct detections within the confidence interval of [0~0.6], accounting for approximately 30% of the total.

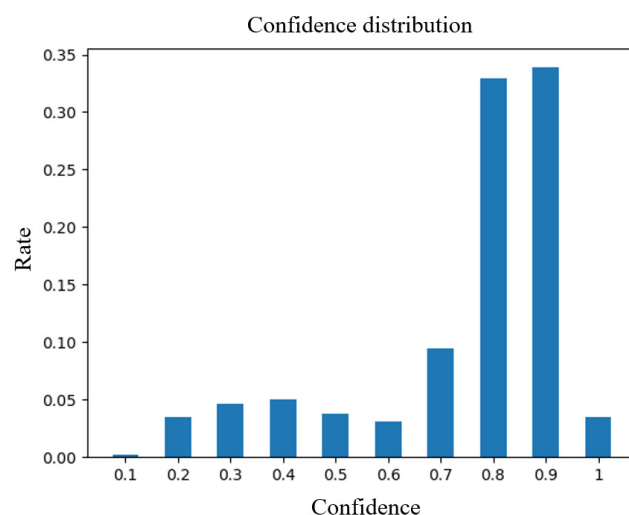


Figure 5. Confidence distribution of correctly detected moving target shadow results.

To minimize the adverse effects of false alarms on multi-target data association and trajectory filtering, traditional multi-target tracking algorithms like SORT and DeepSORT typically discard detections with low confidence (<0.5). Only detections with confidence above a certain threshold are retained for data association with target trajectories, inevitably leading to a high rate of missed alarms and trajectory interruptions and subsequently impacting the performance of the tracking algorithm.

2.3. ReID Feature Extraction

The results of ViSAR target shadow detection can be forward inferred to obtain embedded feature prediction values through a re-identification feature extraction network [38]. Training the embedding features is treated as a classification problem, where targets with the same ID are considered as one class. Target features are classified through a fully connected layer followed by SoftMax, providing the likelihood of each ID corresponding to the detection results.

Common ReID backbone network models include the DenseNet series, ResNet series, MobileNetV2, among others. This paper utilized ViSAR data of vehicle targets released by Sandia National Laboratories. Through network training and validation, the optimal ReID feature extraction network model was established. Video sequences were fed into a dynamic target shadow detection network based on YOLOv5s to obtain a set of ViSAR vehicle target shadow detection image slices. A total of 178 images were selected from 10 vehicle target shadows, with 160 images for training and 18 images for validation, maintaining a training-to-validation ratio of approximately 9:1. Figure 6 displays partial image slices of four targets. The sizes of the images varied, but, during training, all images were resized to 224×224 before being input into the ReID network. The training process involved training in two phases by freezing the backbone network, with 15 epochs before and after freezing. Experimental results of ViSAR dynamic target shadow classification using different ReID feature extraction network models are shown in Table 2. The number of parameters reflects the model's computational complexity, while mAP and Rank-k indicate the model's classification accuracy. The Rank-5 metrics for all models reached 100%.

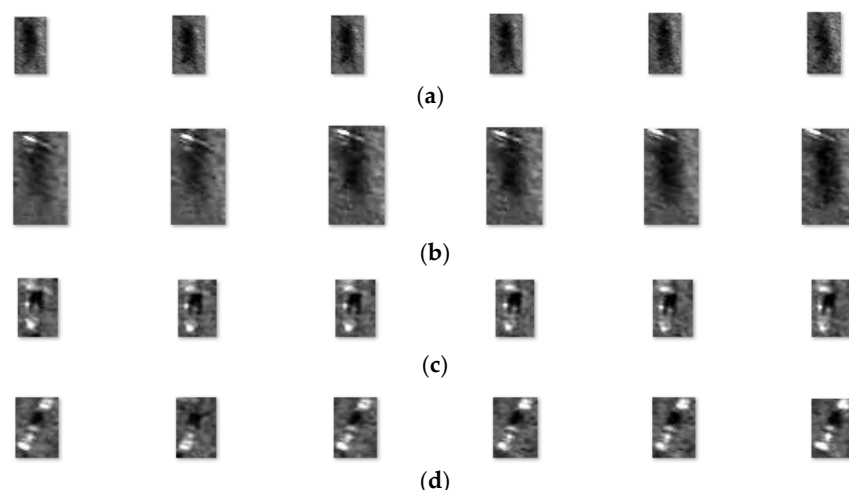


Figure 6. Partial targets of the ReID dataset. (a) ID01 target shadow illustration. (b) ID02 target shadow illustration. (c) ID03 target shadow illustration. (d) ID05 target shadow illustration.

Table 2. Experimental results of ViSAR dynamic target shadow classification using different ReID feature extraction network models.

Model	Parameters	mAP	Rank-1	Rank-5
MobileNetV2	3.5 M	72%	74.43%	100%
EfficientV2B0	5.3 M	88%	89.66%	100%
EfficientV2B1	8.2 M	87%	86.21%	100%
DenseNet121	8.1 M	96%	96.55%	100%
ResNet50	25.6 M	84%	82.76%	100%
InceptionV3	23.9 M	83%	86.21%	100%
VGG16	138.4 M	97%	96.55%	100%

Further analysis indicates that the MobileNetV2 model, which has the fewest parameters, has the lowest accuracy. The ResNet50 and InceptionV3 models show limited accuracy improvement with relatively large parameter sizes. Although the VGG16 model has the highest accuracy, its parameter count is several times or even tens of times larger than other models, resulting in lower computational efficiency. The DenseNet121 (Densely Connected Convolutional Networks 121) model exhibits significantly higher classification accuracy than the two EfficientV2 models, approaching the accuracy of VGG16, while maintaining a smaller parameter count. Therefore, due to its dual advantages of high classification accuracy and efficiency, this paper selects DenseNet121 as the ReID feature extraction network model.

2.4. Multi-Stage Data Association Based on Detection Confidence

2.4.1. Similarity Cost Matrix

In the ViSAR multi-target tracking algorithm, to ensure target identity consistency, it is necessary to match multiple shadow detection results of the current frame with all existing multi-target trajectories based on similarity. The key to improving matching accuracy lies in constructing an appropriate similarity cost matrix, whether based on generative and discriminative models in traditional target tracking algorithms or implementing multi-target tracking using deep learning. Essentially, the goal is to determine the similarity between target detection results and trajectories. This paper selects the classic IoU similarity and ReID feature cosine similarity to generate the similarity cost matrix for ViSAR multi-target data association.

(1) IoU Similarity

The calculation formula for *IoU* is shown in Equation (1).

$$IoU = \frac{Inter}{Union} \quad (1)$$

In the equation provided, *Inter* represents the intersection between the trajectory prediction box obtained through Kalman filtering and the shadow detection box output by the dynamic target shadow detection algorithm, while *Union* represents their union. IoU calculates the proportion of the common area between the prediction box and the detection box to the total area covered by both boxes. It effectively reflects the consistency of their positions in a two-dimensional space and the distribution of pixels. By computing the IoU of each shadow detection box with all trajectory prediction boxes, the IoU similarity cost matrix for multi-target data association can be obtained.

(2) ReID Feature Cosine Similarity

Let $\vec{x}$ and $\vec{y}$ represent the current ReID feature vectors of the dynamic target shadow trajectory and the ReID feature vector of the ViSAR dynamic target shadow detection, respectively. The ReID features cosine similarity, denoted as *cos_similarity*, measures the similarity between the two ReID feature vectors by calculating the cosine value of the angle between them. The formula for calculating the cosine similarity is as follows:

$$\cos_similarity(\vec{x}, \vec{y}) = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\| \|\vec{y}\|} \quad (2)$$

where $\|\vec{x}\|$ and $\|\vec{y}\|$ represent the L2 norm of two ReID feature vectors $\vec{x}$ and $\vec{y}$, and $\cdot$ represents the dot product operation. The *cos_similarity* value ranges from -1 to 1 , where a value closer to 1 indicates stronger similarity.

If the multi-target data association in the current video frame needs to handle m dynamic target shadow detections and n target trajectories, with their ReID feature vectors denoted as $\vec{x}_i (i = 1, 2, \dots, m)$ and $\vec{y}_j (j = 1, 2, \dots, n)$, respectively, the *cos_matrix* C for ReID features can be computed as follows:

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mn} \end{pmatrix} \quad (3)$$

$$c_{ij} = \cos_similarity(\vec{x}_i, \vec{y}_j) = \frac{\vec{x}_i \cdot \vec{y}_j}{\|\vec{x}_i\| \|\vec{y}_j\|} (i = 1, 2, \dots, m; j = 1, 2, \dots, n) \quad (4)$$

This paper utilizes the Hungarian algorithm to complete the multi-target data association in ViSAR based on the similarity cost matrix generated by IoU similarity and ReID feature cosine similarity. The Hungarian algorithm [39] is a combinatorial optimization algorithm that solves task assignment problems in polynomial time. It is mainly used to address the bipartite graph maximum matching problem in the field of operations research. Following the “minimum cost” criterion, it gradually constructs the maximum matching by finding augmenting paths. When solving data association problems in multi-target tracking, it has advantages such as simple algorithm implementation and low complexity.

2.4.2. Multi-Stage Data Association Strategy

Based on the differences in tracking states, the trajectories of ViSAR dynamic target shadows are divided into four categories: initial trajectory, confirmed trajectory, lost trajectory, and deleted trajectory. The initial trajectory is created by the initially detected high-confidence dynamic target shadow and remains as an initial trajectory even after successfully associating in the second frame. The confirmed trajectory is a trajectory currently maintained stably by the Kalman filter, usually transitioning from an initial trajectory or lost trajectory after successfully tracking for three consecutive frames. The lost trajectory refers to a trajectory that failed to be successfully associated with detection results, with its trajectory state being the prediction value of the Kalman filter. If it fails to associate for N consecutive frames, the lost trajectory is converted to a deleted trajectory.

To maximize the utilization of shadow detection results with different confidence levels, this paper designs a four-stage data association strategy based on shadow detection confidence. The algorithm flow is shown in Figure 7, and the corresponding operational steps are as follows:

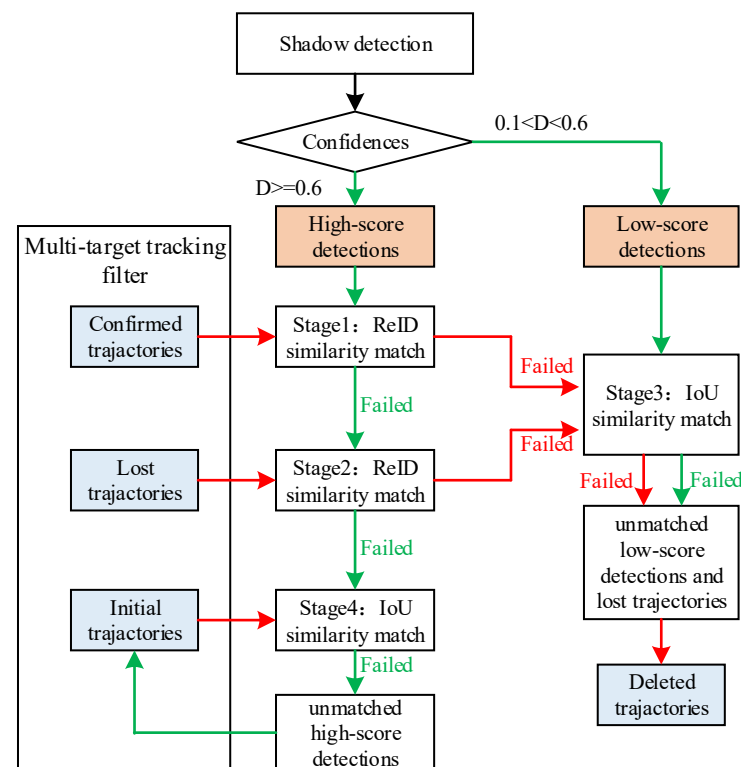


Figure 7. Multi-stage data association algorithm flowchart. The green paths represent the target detection result, while the red paths represent the trajectory.

Stage 1: ReID feature similarity data association is performed between high-score detections and confirmed trajectories. The confirmed trajectories with successful association are maintained as confirmed ones, the confirmed trajectories with failed association are transferred to Stage 3, and the high-score detections with failed association are transferred to Stage 2.

Stage 2: ReID feature similarity data association is performed between unmatched high-score detections from Stage 1 and lost trajectories. The lost trajectories with successful association are marked as the confirmed trajectories, the lost trajectories with failed association are transferred to Stage 3, and the high-score detections with still failed association are transferred to Stage 4.

Stage 3: IoU similarity data association is performed between low-score detections and unmatched confirmed or lost trajectories from Stage 1 and Stage 2. The trajectories with successful association are marked as confirmed trajectories, the trajectories with failed association are marked as lost trajectories, and the low-score detections with failed association are discarded. At the same time, the lost trajectories that remain unmatched for 30 consecutive frames are marked as deleted trajectories.

Stage 4: IoU similarity data association is performed between unmatched high-score detections from Stage 2 and initial trajectories. Initial trajectories are marked as confirmed trajectories after three consecutive successful associations or as deleted trajectories if the association fails. Some remaining unmatched high-score detections establish their own initial trajectories.

The multi-stage data association strategy above has the following characteristics:

1. High-score detections are sequentially associated with confirmed trajectories, lost trajectories, and initial trajectories, reflecting the algorithm's emphasis on maximizing the protection of the continuity of confirmed and lost trajectories.
2. Low-score detections are retained but only associated with unmatched trajectories at a lower priority level, reflecting the algorithm's priority to protect high-score detections. Low-score detections are considered as potentially beneficial supplements to reduce the risk of missed detections and dynamic target shadow trajectory breaks.
3. In Stages 1 and 2, when high-score detections are associated with confirmed trajectories and lost trajectories, the ReID features of these targets are significant. Due to the relatively ambiguous IoU similarity matrix for multiple target trajectories, ReID features are used for Hungarian matching processing.
4. In Stages 3 and 4, traditional IoU similarity is used for Hungarian matching between low-score detections, initial trajectories, unmatched high-score detections, and trajectories. This is because the ReID feature description capability of these targets is relatively weak, and IoU similarity can provide more reliable representation in terms of target position and spatial distribution.
5. Unmatched low-score detections are not used to establish initial trajectories, which helps control the false trajectory rate.

2.5. Multi-Target Tracking Filters

Kalman filtering [40] overcomes the limitations of Wiener filtering, which can only be used in stationary environments. It has a strict recursive relationship, good prediction accuracy, and lower computational costs, making it widely used in tracking algorithms. In ViSAR images, Kalman filtering can effectively predict dynamic systems in the presence of various uncertainties, demonstrating strong adaptability to false target interference.

Establish the state vector of the Kalman filter for ViSAR moving target shadows, as shown in Equation (5).

$$S_k = \left[x(k), y(k), w(k), h(k), \tilde{x}(k), \tilde{y}(k), \tilde{w}(k), \tilde{h}(k) \right]^T \quad (5)$$

In this equation, where k represents the frame number of the video, x , y , $\tilde{x}$, and $\tilde{y}$ are used to represent the two-dimensional position and velocity of the target center. w , h , $\tilde{w}$, and $\tilde{h}$ are used to represent the width and height of the target bounding box.

This model effectively reflects the changes in the position, size, and motion state of the dynamic target shadow.

3. Results

3.1. Tracking Performance Evaluation Metrics

For the results of ViSAR dynamic target tracking, calculate the IoU similarity of the tracking trajectory for each frame with all actual trajectories, construct a similarity cost matrix, and use the Hungarian algorithm to complete the pairing of tracking trajectories with actual trajectories. If the IoU similarity of a pairing is greater than a specified threshold, the target is considered correctly tracked once. If a tracking trajectory is not successfully paired or the IoU similarity is less than the specified threshold, the tracking trajectory is considered a false trajectory. If an actual trajectory is not successfully paired or the IoU similarity is less than the specified threshold, the actual trajectory is considered missed tracking. If the ID of the actual trajectory matched by a tracking trajectory in the current frame changes from the ID matched in the previous frame, an identity switch is noted. Based on this, various performance metrics of the tracking results can be summarized as follows.

(1) Individual Performance Metrics

- *TP*: Number of correct tracks, the total number of target trajectories correctly tracked in each frame.
- *FP*: Number of false alarms, the total number of false tracking trajectories in each frame.
- *FN*: Number of missed detections, the total number of target trajectories not correctly tracked in each frame.
- *IDSW*: Number of identity switches, the total number of trajectory identity switches that occur in each frame.

Typically, an ideal multi-target tracking algorithm should have a larger value for *TP* and smaller values for *FP*, *FN*, and *IDSW* [41].

(2) Comprehensive Performance Metrics

- *MT* (Mostly Tracked): Number of trajectories that are correctly tracked in more than 80% of frames. [42]
- *ML* (Mostly Lost): Number of trajectories that are not correctly tracked in more than 80% of frames. [43]
- *MOTA* (Multiple Object Tracking Accuracy): The calculation method is shown in Equation (6), where t is the frame number and GT_t is the number of actual trajectories in frame t . This metric reflects the total proportion of various tracking errors such as false alarms, missed detections, and identity switches in all ground truths. A higher *MOTA* value indicates higher tracking accuracy of the algorithm.

$$MOTA = 1 - \frac{\sum_t (FN_t + FP_t + IDSW_t)}{\sum_t GT_t} \quad (6)$$

- *MOTP* (Multiple Object Tracking Precision): The precision of multi-object tracking calculated as shown in Equation (7). Here, t represents the frame number, $d_{t,i}$ represents the overlap between the detection box of the tracking trajectory and the actual trajectory, which is the i -th correctly tracked actual trajectory in frame t , and c_t is the number of correctly tracked actual trajectories in frame t . *MOTP* provides the average overlap of the detection boxes between all correctly tracked actual trajectories and tracking trajectories. A higher value indicates a higher tracking precision of the algorithm.

$$MOTP = \frac{\sum_{t,i} d_{t,i}}{\sum_t c_t} \quad (7)$$

- **HOTA (Higher Order Tracking Accuracy):** This parameter is a comprehensive metric obtained by integrating the tracking evaluation into three subtasks: detection, association, and localization. The specific calculation method is calculated as shown in Equation (8) and detailed in reference [43]. Among them, $DetA_\alpha$ is used to measure the detection accuracy, and $AssA_\alpha$ is used to measure the association accuracy. The calculations of both rely on the threshold α of the Hungarian matching. Compared to $MOTA$ and $MOTP$, this metric can provide a more comprehensive evaluation of the performance of ViSAR tracking algorithms. Generally, the higher the $HOTA$, the better the overall performance of the tracking algorithm.

$$HOTA = \int_{0 < \alpha \leq 1} \sqrt{DetA_\alpha + AssA_\alpha} \quad (8)$$

3.2. ViSAR Moving Target Shadow Tracking Experiment Results

The experiment uses two sets of ViSAR data. The first set of experimental data is simulated spaceborne ViSAR data, and the second set is actual airborne ViSAR data released by Sandia National Laboratories [37]. Two mainstream video multi-object tracking algorithms, DeepSORT and ByteTrack, are selected for comparison with the algorithm proposed in this paper. A ViSAR dynamic target shadow detection network based on YOLOv5s generates shadow detection slice images for each frame of the video, serving as input for the tracking algorithm.

(1) Simulation Experiment Results of Spaceborne ViSAR

The spaceborne ViSAR simulation data consisted of Ka-band ViSAR image data produced using a spaceborne SAR simulation platform on three land vehicle targets. The satellite was positioned at an orbit height of 500 km, with an incidence angle of 30 degrees, a signal transmission bandwidth of 480 MHz, a resolution of 0.5 m, and a video frame rate of 10 Hz. The simulation depicted a natural land scene model, as illustrated in Figure 8, which included backgrounds such as grassland, roads, lakes, and three vehicle targets moving on the road. Target T1 traveled along route A-B-C-D, target T2 moved along the route D-C-B-A, and target T3 followed the circular road route B-E-C. In Figure 8b, the red area represents the target's location, and the green area represents the simulated initial calculation of the target shadow's location, including the shadow formed by the target's position and the shadow formed by the geometric projection of the spaceborne SAR. The moving target shadows in each frame image are manually labeled as ground truth data.

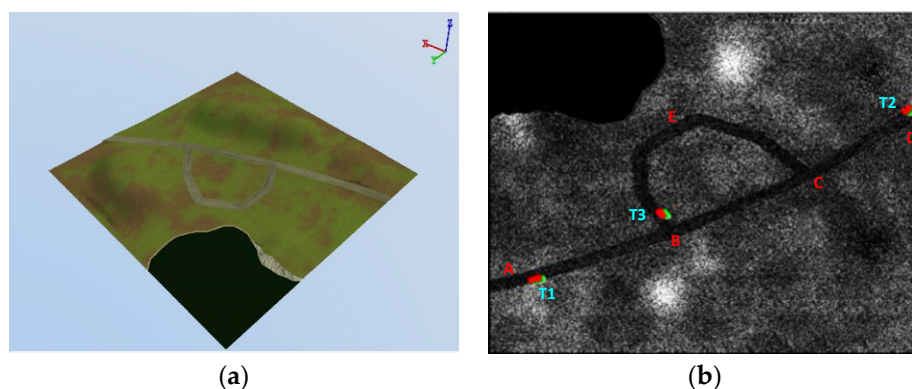


Figure 8. Simulation modeling of dynamic target shadows in spaceborne ViSAR. (a) Ground background simulation model. (b) Motion target and shadow simulation model.

The experimental results and tracking performance metric comparisons of the three tracking algorithms are shown in Figure 9 and Table 3, respectively. In the table, an upward arrow next to a metric indicates that the higher the value, the better the performance of the tracking algorithm, while a downward arrow indicates that the lower the value, the better the performance of the tracking algorithm.

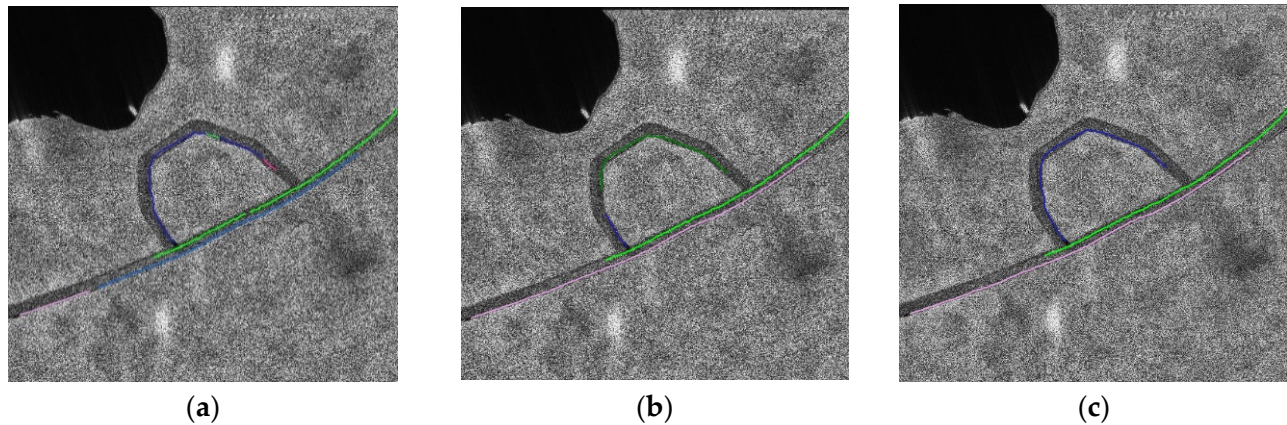


Figure 9. Experimental results comparison of different tracking algorithms (spaceborne ViSAR simulation data). (a) DeepSORT. (b) ByteTrack. (c) Proposed method.

Table 3. Comparison of tracking performance metrics of different algorithms (spaceborne ViSAR simulation data).

Methods	DeepSORT	ByteTrack	Proposed Method
TP ↑	520	517	560
FP ↓	9	12	7
FN ↓	63	66	23
IDSW ↓	3	1	0
MT ↑	3	3	3
ML ↓	0	0	0
MOTA ↑	87.14%	86.28%	94.85%
MOTP ↑	76.89%	76.58%	76.85%
HOTA ↑	80.45%	80.23%	92.88%

(2) Experiment Results with Actual Airborne ViSAR Data

For the Sandia National Laboratories ViSAR data [37], 200 consecutive frames of images (from the 150th frame to the 300th frame) were used. The moving target shadows in each frame image are manually labeled as ground truth data. The results of each tracking algorithm are shown in Figure 10. Figure 10a–c displays the multi-object trajectories generated by each algorithm, with each color corresponding to a trajectory ID. Figure 10d–f indicates the detection boxes of different target tracking trajectories in the 200th frame. Figure 10g–i shows the detection boxes of different target tracking trajectories in the 250th frame, while (j), (k), and (l) display the detection boxes of different target tracking trajectories in the 300th frame.

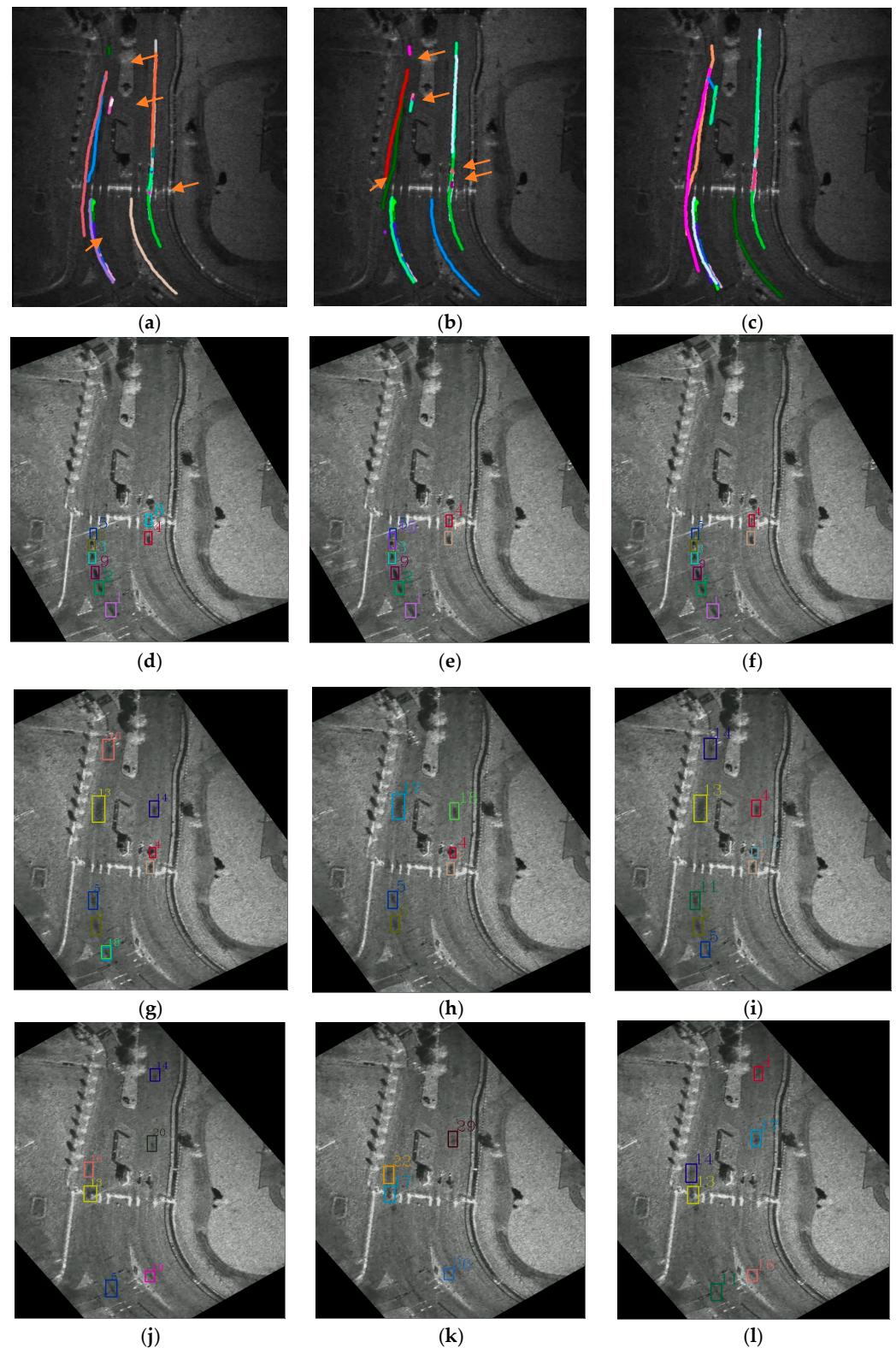


Figure 10. Experimental results comparison of different tracking algorithms (airborne ViSAR data). (a) DeepSORT (trajectories). (b) ByteTrack (trajectories). (c) Proposed method (trajectories). (d) DeepSORT (Frame 200). (e) ByteTrack (Frame 200). (f) Proposed method (Frame 200). (g) DeepSORT (Frame 250). (h) ByteTrack (Frame 250). (i) Proposed method (Frame 250). (j) DeepSORT (Frame 300). (k) ByteTrack (Frame 300). (l) Proposed method (Frame 300). The number represents the number of the target, and the number of the same target is the same in different frames, and orange arrows are used to identify some visible trajectory interruptions and their locations.

In Figure 10a,b, for the vehicle target tracking trajectories of two traditional methods, DeepSORT and ByteTrack, orange arrows are used to identify some visible trajectory interruptions and their locations. In the figure, the track color change caused by identity switching due to track fracture is clearly visible. At the same time, due to the overlapping display of multiple target trajectories, the track color changes of some underlying trajectory interruptions and identity switching are covered. Nevertheless, it is still easy to observe that the tracking trajectories generated by the DeepSORT and ByteTrack algorithms have shorter durations, more trajectory breaks, and more identity switches. During a period when the speed of vehicle targets significantly changes before and after crossing a gate, the proposed algorithm achieved a higher tracking success rate. This indicates that the introduction of ReID features and the utilization of a multi-stage data association strategy can better adapt to the diversity of shadow features resulting from changes in target motion states, demonstrating the superior performance of the proposed algorithm. Compared with the traditional DeepSORT and ByteTrack, the proposed method can effectively reduce the incidence of trajectory fracture and significantly improve the accuracy of data association and trajectory continuity.

The detailed quantitative analysis results can be seen in the statistical results of nine tracking performance indicators in Table 4. Single-frame performance metrics and comprehensive performance metric calculations for the three different tracking algorithms are presented in the table.

Table 4. Comparison of tracking performance metrics of different algorithms (airborne ViSAR data).

Methods	DeepSORT	ByteTrack	Proposed Method			
			(C = 0.5)	(C = 0.6)	(C = 0.7)	(C = 0.6 + Only IoU)
TP ↑	1289	1298	1340	1419	1409	1332
FP ↓	30	25	74	62	62	85
FN ↓	335	326	284	205	215	292
IDSW ↓	12	12	5	9	10	8
MT ↑	9	8	9	11	11	9
ML ↓	3	3	1	1	1	1
MOTA ↑	76.79%	77.65%	79.22%	82.94%	82.33%	77.29%
MOTP ↑	59.64%	59.32%	58.22%	58.24%	58.35%	58.31%
HOTA ↑	68.62%	68.93%	68.67%	69.74%	68.93%	67.91%

4. Discussion

From the spaceborne ViSAR simulation experimental results in Figure 9 and Table 3, it can be observed that the performance of the three algorithms in terms of trajectory tracking precision (*MOTP*) is very close. However, the DeepSORT algorithm, due to discarding low-confidence detection results, experienced multiple incorrect associations and trajectory breaks. The ByteTrack algorithm, although only experiencing one identity switch, still had multiple trajectory interruptions for the moving target T3 on the curved path. Compared to the two traditional algorithms, the proposed method achieved the best tracking results. In various technical metric comparisons, metrics such as *TP*, *FP*, *FN*, *IDSW*, *MT*, *ML*, *MOTA*, *MOTP*, and *HOTA* all showed the best performance. There were no trajectory breaks or identity switches during the entire process. The two comprehensive tracking performance metrics, *MOTA* and *HOTA*, reached the highest values of 94.85% and 92.88%, respectively, reflecting the effectiveness of the proposed method in utilizing ReID feature descriptions and low-confidence shadow detections, and applying a multi-stage data association strategy.

Based on the actual airborne ViSAR data processing results in Figure 10 and Table 4, it is evident that, compared to the two traditional algorithms, the proposed algorithm demonstrates a clear advantage in various technical metric comparisons. Metrics such as TP, FN, IDSW, MT, ML, MOTA, and HOTA all show the best performance, with the two comprehensive tracking performance metrics MOTA and HOTA reaching the highest values of 82.94% and 69.74%, respectively. However, due to the algorithm's recall of a higher number of low-confidence shadow detections, while effectively improving the success rate of multi-target data association and trajectory continuity, it has somewhat increased the false alarm rate. Additionally, because the positioning accuracy of low-confidence shadow detections is typically lower, the MOTP metric of the proposed algorithm is slightly lower than the other two algorithms.

Compared with the data processing results of spaceborne ViSAR simulation in Table 3, all algorithms have a certain decrease in all metrics in Table 4 when processing the actual data of airborne video SAR. This is mainly due to the fact that a more rational moving target motion scene is set up in the spaceborne video SAR simulation experiment; two targets move on an approximately straight road, although there is still one target moving on a curve, and the movement rate of the three targets always remains stable, which makes the shadow features of the three moving targets always have good stability, and the challenge of moving target tracking is small. However, in the actual data of airborne video SAR, most of the vehicle targets do a "deceleration-stop-acceleration" variable speed movement near the checkpoint, and, when passing through the checkpoint, the building has a serious occlusion to the vehicle targets on the road, and these non-ideal factors make the shadow characteristics of the vehicle targets show a large change in the whole process, thereby reducing the accuracy of multi-target data association and the accuracy of track tracking.

All algorithms run on a deep learning server, equipped with an Intel i9-14900K @3.20 GHz CPU, NVIDIA RTX4090 GPU, 128 GB RAM, Window 10 operating system, and a Python 3.8 software development environment. Statistically, the processing time of the proposed tracking algorithm for all 200 frames of video data is 4.88 s, with an overall processing efficiency of approximately 41 Hz, meeting the real-time application requirements of the algorithm.

5. Conclusions

This study tackles the prevalent issues encountered in ViSAR dynamic target shadow tracking, including the abundance of low-confidence shadow detections, the high error rates associated with multi-target data association, and the fragility of tracking trajectories. A novel ViSAR dynamic target shadow tracking approach that leverages ReID feature extraction and a multi-stage data association strategy are proposed. Firstly, a lightweight enhanced YOLOv5s detection network reduces pooling loss, strengthens semantic information, and mitigates the impact of discrepancies between classification and position prediction tasks by applying GFPN-S, SPPFS, and a decoupling head design. Next, an optimal ReID feature extraction network model was established, and the DenseNet121 deep learning network is selected to extract robust, high-dimensional ReID features from the detected dynamic target shadows. Subsequently, ReID feature cosine similarity is integrated with IoU similarity to compute a cost matrix for multi-target data association between detections and trajectories. Moreover, a multi-stage data association tactic that prioritizes shadow detection confidence is developed, which effectively mitigates the risk of missed detections and tracking trajectory interruptions by revisiting some low-confidence detections. In the four-stage data association process, different track trajectories and high/low score detections are matched with different priorities for Hungarian matching, fully tapping into the valuable information in the low confidence shadow detection results. Finally, the

video multi-target tracking filter uses the classical Kalman filter to filter and predict target motion states.

The experiment uses simulated spaceborne ViSAR data and actual airborne ViSAR data released by Sandia National Laboratories. The experimental results illustrate that, compared to two mainstream video multi-object tracking algorithms, DeepSORT and ByteTrack, the proposed algorithm significantly enhances tracking performance, yielding the highest MOTA and HOTA scores of 94.85% and 92.88%, respectively, on the simulated spaceborne ViSAR data, and the highest MOTA and HOTA scores of 82.94% and 69.74%, respectively, on airborne field data. The new method enhances the accuracy of multi-target data association, reduces false trajectory generation, lowers the missed detection rate, mitigates the risk of trajectory interruptions, and comprehensively improves multi-target tracking continuity. The efficiency of the algorithm is well-suited for real-time applications, rendering it a promising solution with considerable practical utility.

It should be pointed out that the performance of the proposed method has only been verified on limited data, and the moving target shading characteristics of video SAR usually have significant differences and uniqueness for different video SAR systems, and the deep network model needs to be trained on more datasets to improve the generalization performance of the method. Looking forward to the future, with the continuous maturity of spaceborne and airborne video SAR technology and the increasingly rich datasets, more new algorithm models will be explored and discovered.

Author Contributions: All the authors made significant contributions to the work. B.W., A.Y. and W.T. designed the research and analyzed the results. methodology, B.W., A.Y. and Z.H. performed the experiments. B.W. and A.Y. wrote the paper. Z.H., Z.D. and W.T. provided suggestions for the preparation and revision of the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: Author Boxu Wei was employed by the company Chongqing Xingwang Network System Research Institute Co. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Damini, A.; Balaji, B.; Parry, C.; Mantle, V. *Algorithms for Synthetic Aperture Radar Imagery XVII*; SPIE: Bellingham, DC, USA, 2010; Volume 7699.
2. Liu, B.; Zhang, X.; Tang, K.; Liu, M.; Liu, L. Spaceborne Video-SAR moving target surveillance system. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 2348–2351.
3. Zhang, Z.; Shen, W.; Xia, L.; Lin, Y.; Shang, S.; Hong, W. Video SAR Moving Target Shadow Detection Based on Intensity Information and Neighborhood Similarity. *Remote Sens.* **2023**, *15*, 1859. [[CrossRef](#)]
4. He, Z.; Chen, X.; Yi, T.; He, F.; Dong, Z.; Zhang, Y. Moving target shadow analysis and detection for ViSAR imagery. *Remote Sens.* **2021**, *13*, 3012. [[CrossRef](#)]
5. Bhamre, P.; Gupta, S. GMTI STAP performance under space-time impaired clutter environment. *IET Radar Sonar Navig.* **2019**, *13*, 1836–1841. [[CrossRef](#)]
6. Wei, B.; Yu, A.; Tong, W.; He, Z. Performance Analysis of Moving Target Shadow Detection in Video SAR Systems. *Remote Sens.* **2024**, *16*, 1825. [[CrossRef](#)]
7. Cheng, Y.; Ding, J.; Sun, Z.; Zhong, C. Processing of Airborne Video SAR Data Using the Modified Back Projection Algorithm. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5238013. [[CrossRef](#)]
8. Yang, T.; Zhang, X.; Xu, Q.; Zhang, S.; Wang, T. An embedded-gpu-based scheme for real-time imaging processing of unmanned aerial vehicle borne video synthetic aperture radar. *Remote Sens.* **2024**, *16*, 191. [[CrossRef](#)]

9. Yang, T.; Wang, T.; Zheng, N.; Zhang, S.; Meng, F.; Zhang, X.; Wu, Q. Miniaturization Design of High-Integration Unmanned Aerial Vehicle-Borne Video Synthetic Aperture Radar Real-Time Imaging Processing Component. *Remote Sens.* **2024**, *16*, 1273. [\[CrossRef\]](#)
10. Yin, Z.; Zheng, M.; Ren, Y. A ViSAR Shadow-Detection Algorithm Based on LRSD Combined Trajectory Region Extraction. *Remote Sens.* **2023**, *15*, 1542. [\[CrossRef\]](#)
11. Wang, H.; Chen, Z.; Zheng, S. Preliminary research of low-RCS moving target detection based on Ka-band video SAR. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 811–815. [\[CrossRef\]](#)
12. Zhang, Y.; Mao, X.; Yan, H.; Zhu, D.; Hu, X. A novel approach to moving targets shadow detection in VideoSAR imagery sequence. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), FortWorth, TX, USA, 23–28 July 2017; pp. 606–609.
13. Yan, S.; Zhang, F.; Fu, Y.; Zhang, W.; Yang, W.; Yu, R. A Deep Learning-Based Moving Target Detection Method by Combining Spatiotemporal Information for ViSAR. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 4014005. [\[CrossRef\]](#)
14. Wen, L.; Ding, J.; Loffeld, O. Video SAR Moving Target Detection Using Dual Faster R-CNN. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 2984–2994. [\[CrossRef\]](#)
15. Yan, H.; Huang, J.; Li, R.; Wang, X.; Zhang, J.; Zhu, D. Research on Video SAR Moving Target Detection Algorithm Based on Improved Faster Region-based CNN. *J. Electron. Inf. Technol.* **2021**, *43*, 615–622.
16. Bao, J.; Zhang, X.; Zhang, T.; Xu, X. ShadowDeNet: A Moving Target Shadow Detection Network for Video SAR. *Remote Sens.* **2022**, *14*, 320. [\[CrossRef\]](#)
17. Zhong, C.; Ding, J.; Zhang, Y. Video SAR Moving Target Tracking Using Joint Kernelized Correlation Filter. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 1481–1493. [\[CrossRef\]](#)
18. Xu, Z.; Zhang, Y.; Li, H.; Mu, H.; Zhuang, Y. A new shadow tracking method to locate the moving target in SAR imagery based on KCF. In Proceedings of the International Conference in Communications, Signal Processing and Systems, Harbin, China, 14–16 July 2017; Springer: Singapore, 2017; pp. 2661–2669.
19. Henke, D.; Magnard, C.; Frioud, M.; Small, D.; Meier, E.; Schaepman, M.E. Moving target tracking in single channel wide-beam SAR. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 4735–4747. [\[CrossRef\]](#)
20. He, Z.; Chen, X.; Yu, C.; Li, Z.; Yu, A.; Dong, Z. A Robust Moving Target Shadow Detection and Tracking Method for VideoSAR. *J. Electron. Inf. Technol.* **2022**, *44*, 3882–3890.
21. Liu, Y.J.; Qi, X.Y. A moving ship tracking method based on long time interval sequential SAR images. *J. Univ. Chin. Acad. Sci.* **2021**, *38*, 642–648.
22. Liang, Z.; Liang, C.; Zhang, Y.; Mu, H.; Li, G. Tracking of moving target based on SiamMask for video SAR system. In Proceedings of the IEEE International Conference on Signal, Information and Data Processing, Chongqing, China, 11–13 December 2019; pp. 1–4.
23. Ding, J.; Zhong, C.; Wen, L.; Xu, Z. Joint detection of moving target in video synthetic aperture radar. *J. Radars* **2022**, *11*, 313–323.
24. Hu, Y. *Research on Shadow-Based SAR Multi-Target Tracking Method*; University of Electronic Science and Technology of China: Chengdu, China, 2021.
25. Qiu, L.; Zhang, X.; Hao, D. VideoSAR moving target detection and tracking algorithm based on deep learning. *Remote Sens. Nat. Resour.* **2023**, *35*, 157–166.
26. Somers, V.; De Vleeschouwer, C.; Alahi, A. Body part-based representation learning for occluded person re-identification. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 2–7 January 2023; pp. 1613–1623.
27. Chen, X.; Fu, C.; Zhao, Y.; Zheng, F.; Song, J.; Ji, R.; Yang, Y. Saliency-guided cascaded suppression network for person re-identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 3300–3310.
28. Li, P.; Zhang, J.; Zhu, Z.; Li, Y.; Jiang, L.; Huang, G. State-aware re-identification feature for multi-target multi-camera tracking. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019.
29. Bewley, A.; Ge, Z.; Ott, L.; Ramos, F.; Upcroft, B. Simple online and realtime tracking. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 3464–3468.
30. Wojke, N.; Bewley, A.; Paulus, D. Simple online and realtime tracking with a deep association metric. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 3645–3649.
31. Zhang, Y.; Sun, P.; Jiang, Y.; Yu, D.; Weng, F.; Yuan, Z.; Luo, P.; Liu, W.; Wang, X. Bytetrack: Multi-object tracking by associating every detection box. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Cham, Switzerland, 2022; pp. 1–21.
32. Shi, H.Y.; Hou, Z.T.; Guo, X.H.; Li, J.W. Moving Targets Indication Method in Single High Resolution SAR Imagery Based on Shadow Detection. *Signal Process.* **2012**, *28*, 1706–1713.

33. Krizhevsky, A.; Sutskever, I.; Hinton, G. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems 25 (NIPS 2012), Lake Tahoe, NV, USA, 3–6 December 2012.
34. Yu, A.X.; Wei, B.X.; Tong, W.H.; He, Z.H.; Dong, Z. A ViSAR Moving Target Shadow Detection Network Based on Improved YOLOv5. In Proceedings of the 5th China International SAR Symposium (CISS2024), Xi'an, China, 27–29 November 2024.
35. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
36. Jiang, T.; Li, C.; Yang, M.; Wang, Z. An improved YOLOv5s algorithm for object detection with an attention mechanism. *Electronics* **2022**, *11*, 2494. [[CrossRef](#)]
37. Pathfinder Radar ISR Systems, Video SAR Real Data [EB/OL]. Available online: <https://www.sandia.gov/radar/video/index.html> (accessed on 30 December 2021).
38. He, L.; Liao, X.; Liu, W.; Liu, X.; Cheng, P.; Mei, T. Fastreid: A pytorch toolbox for real-world person re-identification. *arXiv* **2020**, arXiv:2006.02631.
39. Kuhn, H.W. The Hungarian method for the assignment problem. *Nav. Res. Logist. Q.* **1995**, *2*, 83–97. [[CrossRef](#)]
40. Kalman, R.E. A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng.* **1960**, *82*, 35–45. [[CrossRef](#)]
41. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
42. Bernardin, K.; Stiefelhagen, R. Evaluating multiple object tracking performance: The CLEAR MOT metrics. *EURASIP J. Image Video Process.* **2008**, *2008*, 246309. [[CrossRef](#)]
43. Luiten, J.; Osep, A.; Dendorfer, P.; Torr, P.; Geiger, A.; Leal-Taixé, L.; Leibe, B. Hota: A higher order metric for evaluating multi-object tracking. *Int. J. Comput. Vis.* **2021**, *129*, 548–578. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.