mathematics

MDPI

Article

Improving the Performance of RODNet for MMW Radar Target Detection in Dense Pedestrian Scene

Yang Li [1], Zhuang Li [1], Yanping Wang [1,*], Guangda Xie [2], Yun Lin [1], Wenjie Shen [1] and Wen Jiang [1]

[1] College of Information, North China University of Technology, Beijing 100144, China
[2] College of Electrical and Control Engineering, North China University of Technology, Beijing 100144, China
[*] Correspondence: wangyp@ncut.edu.cn

Abstract: In the field of autonomous driving, millimeter-wave (MMW) radar is often used as a supplement sensor of other types of sensors, such as optics, in severe weather conditions to provide target-detection services for autonomous driving. RODNet (A Real-Time Radar Object-Detection Network) is one of the most widely used MMW radar range–azimuth (RA) image sequence target-detection algorithms based on Convolutional Neural Networks (CNNs). However, RODNet adopts an object-location similarity (OLS) detection method that is independent of the number of targets to obtain the final target detections from the predicted confidence map. Therefore, it gives a poor performance on missed detection ratio in dense pedestrian scenes. Based on the analysis of the predicted confidence map distribution characteristics, we propose a new generative model-based target-location detection algorithm to improve the performance of RODNet in dense pedestrian scenes. The confidence value and space distribution predicted by RODNet are analyzed in this paper. It shows that the space distribution is more robust than the value distribution for clustering. This is useful in selecting a clustering method to estimate the clustering centers of multiple targets in close range under the effects of distributed target and radar measurement variance and multipath scattering. Another key idea of this algorithm is the derivation of a Gaussian Mixture Model with target number (GMM-TN) for generating the likelihood probability distributions of different target number assumptions. Furthermore, a minimum Kullback–Leibler (KL) divergence target number estimation scheme is proposed combined with K-means clustering and a GMM-TN model. Through the CRUW dataset, the target-detection experiment on a dense pedestrian scene is carried out, and the confidence distribution under typical hidden variable conditions is analyzed. The effectiveness of the improved algorithm is verified: the Average Precision (AP) is improved by 29% and the Average Recall (AR) is improved by 36%.

Keywords: RODNet; target detection; Gaussian mixture model; KL divergence; maximum likelihood

MSC: 68T01

check for updates

Citation: Li, Y.; Li, Z.; Wang, Y.; Xie, G.; Lin, Y.; Shen, W.; Jiang, W. Improving the Performance of RODNet for MMW Radar Target Detection in Dense Pedestrian Scene. *Mathematics* **2023**, *11*, 361. https://doi.org/10.3390/math11020361

Academic Editors: Fan Zhang, Songhe Feng, Yongsheng Zhou and Junlin Hu

Received: 11 December 2022
Revised: 6 January 2023
Accepted: 8 January 2023
Published: 10 January 2023

1. Introduction

Millimeter-wave radar is one of the main sensors for vehicle and pedestrian detection in the field of intelligent transportation. Compared with other sensors such as optical cameras and lidars, millimeter-wave radars are more robustly resistant to environmental changes such as different weather and lighting conditions. However, due to the complex scattering effects of clutter, distributed target, and multipath propagation with limited target resolution [1–3], it is difficult for millimeter-wave radar data to describe the semantic features of the scene, including the shape of the target, the relationship between targets, etc. The key target-detection problem of millimeter-wave radar is optimizing both false alarm and missed detection [4].

The existing radar data processing flow is mainly as follows. First, the radar raw data are subjected to range and azimuth (RA) Fast Fourier Transform (FFT), and then a

Constant False Alarm Rate Detection (CFAR) is performed in the range–azimuth domain data to obtain target point cloud, velocity, and RCS information [5], as shown in Figure 1. Unfortunately, the target-scattering response in nature scenes is more complex than the simplified clutter background model used in CFAR. The local signature of target-scattering response is destroyed after CFAR, which leads to less priori information such as target size and radar measurement accuracy being used to optimize both false alarm and missed detection. Sparser point cloud makes it worse [6].

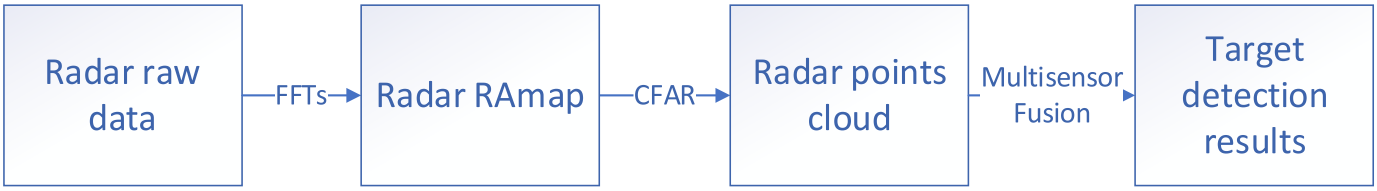


Figure 1. CFAR-based radar target-detection flow chart.

Recent development of machine-learning techniques, such as CNNs [7–14], enable new methods for radar target detection used in the microwave and optical high-resolution remote-sensing community [15–24]. Refs. [7–11] highlighted that over recent years, a lot of studies have been published pursuing multi-sensor fusion with various deep convolutional neural networks and obtaining a state-of-the art performance in object detection and recognition. DCNNs were first introduced by K. Fukushima [12], using the concept of a hierarchical representation of receptive fields from the visual cortex, as presented by Hubel and Wiesel. Afterward, Weibel et al. Ref. [13] proposed convolutional neural networks (CNNs) that share weights with temporal receptive fields and Backpropagation training methods. Later, Y. LeCun [14] presented the first CNN architecture for document recognition. The DCNN models typically accept 2D images or sequential data as the input. Based on radar image, Ref. [19] designed an aggregation module to merge the data from all chirps in the same frame to make full use of radar data in object detection and various Gaussian noises with different parameters are employed to increase data diversity and reduce over-fitting based on the analysis of training data. Ref. [20] proposed a novel cross-modality deep-learning framework for the radar object detection task using the Squeeze-and-Excitation network. A novel noise detection approach is also explored in the study to increase the model's ability to handle noise. Ref. [21] proposed novel scene-aware sequence mix augmentation (SceneMix) and scene-specific post-processing to generate more robust detection results. Ref. [22] proposed a dimension apart network (DANet) for a radar object detection task to be lightweight and capable of extracting temporal–spatial information from the RAMap sequences. Ref. [23] also proposed a lightweight, computationally efficient, and effective network architecture to conquer the question of the trade-off between computational efficiency and performance of radar object-detection tasks. The radar cube is introduced in [24] to maximize the use of radar input data. The aforementioned literature is mainly on the use of data and network lightweight research, but no one has studied improvement methods based on target number.

Spatial bias is a type of inductive bias in CNNs that assumes a certain type of spatial structure present in the data, which can be used to learn the local signature of target-scattering response in the RA image sequence. However, the use of CNNs in low-resolution data such as a Multiple-Input Multiple-Output (MIMO) radar image in the automatic driving scenario presents certain challenges. RODNet firstly introduces CNNs into MIMO radar image feature expression [25]. It uses a chirp-merging module (M-Net), and a temporal deformable convolution (TDC) operation directly learns and encodes the target features from the range–azimuth images to improve the limitation of low resolution MIMO radar image. The M-Net is designed to enhance the SNR of input data using multi-chirp RA images. In addition, the TDC uses the deformable convolution network (DCN) [26] to accomplish the 3D CNN [8,9,27,28] to handle the radar object dynamic motion within the input RA images per frame.

After encoding the multi-chirp RA images for each frame, the RODNet will generate a confidence map (ConfMap) of predicated target locations and classes from the merged image signature codes. The value of ConfMap is assumed to be a Gaussian distribution representing the possibility of point target-detection results occurring at a range–azimuth location. The final detection result is then obtained by the position-processing method. The sorting process is shown in Figure 2. However, RODNet is a cross-supervised training scheme that is characterized by a lack of bounding box. This makes it difficult for RODNet to predict the boundary of the target in the RA coordinate system based on ConfMap. The object-location similarity (OLS) metric, which is like an intersection over union (IoU) [29–36], is proposed to estimate the position of the point target. As it depends only on the sparse high ConfMap value information, the unknown effects on the ConfMap from multi-target, multipath scattering and poor resolution make it difficult to balance the false alarm and missed detection in the dense scenes, as shown in Figure 3. RODNet detects the two proximal pedestrian targets as one because their ConfMaps are overlapped.

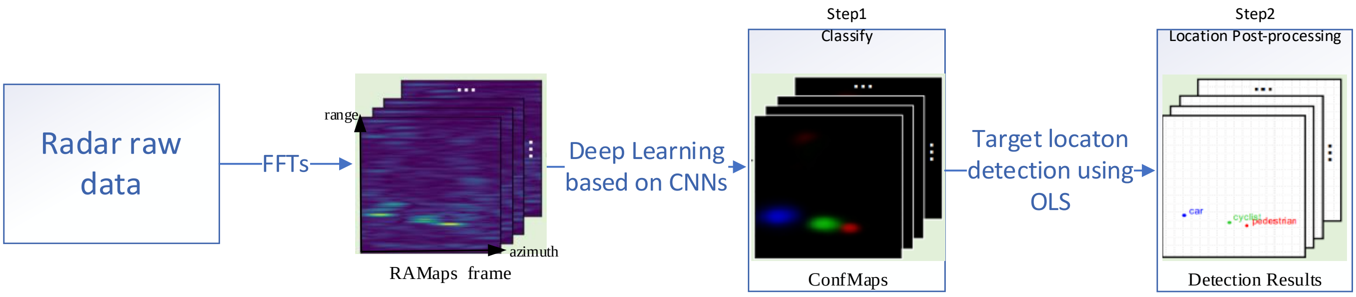


Figure 2. RODNet radar data target-detection flow chart [25]. For radar RA image and the ConfMap predicated by RODNet, their horizontal and vertical coordinates are radar azimuth and range in meter, and every image size is 128 × 128. Each coordinate corresponds to a grid.

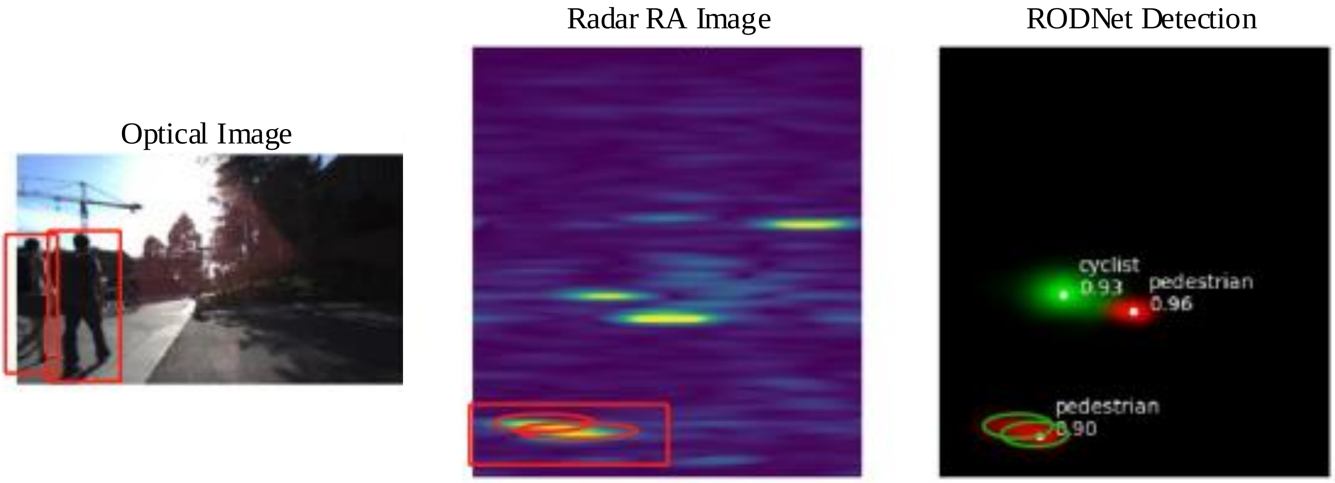


Figure 3. RODNet radar data target-detection results from left to right are optical image, radar RA map (Input data, two walking persons in red box), and radar detection results.

The ConfMap of a certain class predicated by CNNs-based RODNet is believed to indicate the probability of target in an RA grid. However, the OLS method ignores the fact that the number of targets is a hidden variable of the target location. This makes the complex scattering of dense scenes affect the distribution of ConfMap values, which in turn affects the accuracy of target number estimation and ultimately leads to errors in target detection. In order to avoid this problem, this paper intends to independently estimate the target number using ConfMap value distribution and estimate target location using the occupied grid spatial distribution of ConfMap. This new algorithm can improve the detection accuracy by correcting the wrong target estimation number.

Overall, our main contributions are the following:

- Based on the actual data, we analyze the characteristics of ConfMap predicated by RODNet and the limitations of the OLS target-location detection method. The relationship among ConfMap value distribution, occupied grid spatial distribution, and target number is analyzed.
- GMM-TN, a target-state likelihood model with ConfMap value for observation is introduced for simulating the conditional ConfMap value and occupied grid spatial distribution with target number as condition.
- The KL distance measure between the predicated ConfMap of RODNet and the simulated ConfMap under the condition of the given target number is derived, and the maximum posteriori target number estimation is constructed. The CRUW dataset is used to verify the improved missed detection and false alarm of the method.

This paper is organized as follows: the relationship between the value of ConfMap, the spatial distribution of occupied grids, and the number of targets is analyzed by actual data in Section 2. Section 3 describes the method of establishing the GMM-TN model and constructs the maximum for a posteriori target number estimation method. In Section 4, the experiment applied to the CRUW dataset scenarios is introduced. Finally, we summarize our algorithm and experiment results.

2. RODNet ConfMap Characteristics and Limitations

2.1. ConfMap Characteristic Analysis

The RODNet is implemented based on a 3D CNN with an autoencoder architecture. The network architecture first uses some basic three-dimensional convolution structures, then adds a skip connection layer [37] and implements a three-dimensional convolution neural network based on the Hourglass (HG) architecture to ensure that each new convolution operation is not inferior to the previous one. In order to better apply the information contained in all chirps, the author proposes a chirp-merging module (M-NET). Given that the relative motion between the radar and the target causes the radar reflection mode to change with time and the multi-level features are fused, a temporal deformable convolutional (TDC) layer is added to the network [26]. Finally, the network achieved a better classification result. Its encoder and decoder schematic are shown in Figures 4 and 5.

In the decoder shown in Figure 5, we perform feature restoration through deconvolution operations to expand the size of the feature map. At the same time, a convolution operation is performed after each layer of deconvolution, which can effectively extract features while expanding the size of the feature map.

The value of the predicated ConfMap of RODNet can be equivalent to the probability $P_{cls}(r, a)$ that a specific class cls of target will appear on the grid (r, a). RODNet can detect three classes of targets, namely cars, cyclists, and pedestrians. The three classes of target ConfMaps can be obtained from the three channels of RODNet output after normalized processing. The network framework of RODNet is built based on CNNs. Due to the characteristics of CNNs, the predicated ConfMap of RODNet will have local relevance. Figure 6 is a visual representation of a ConfMap.

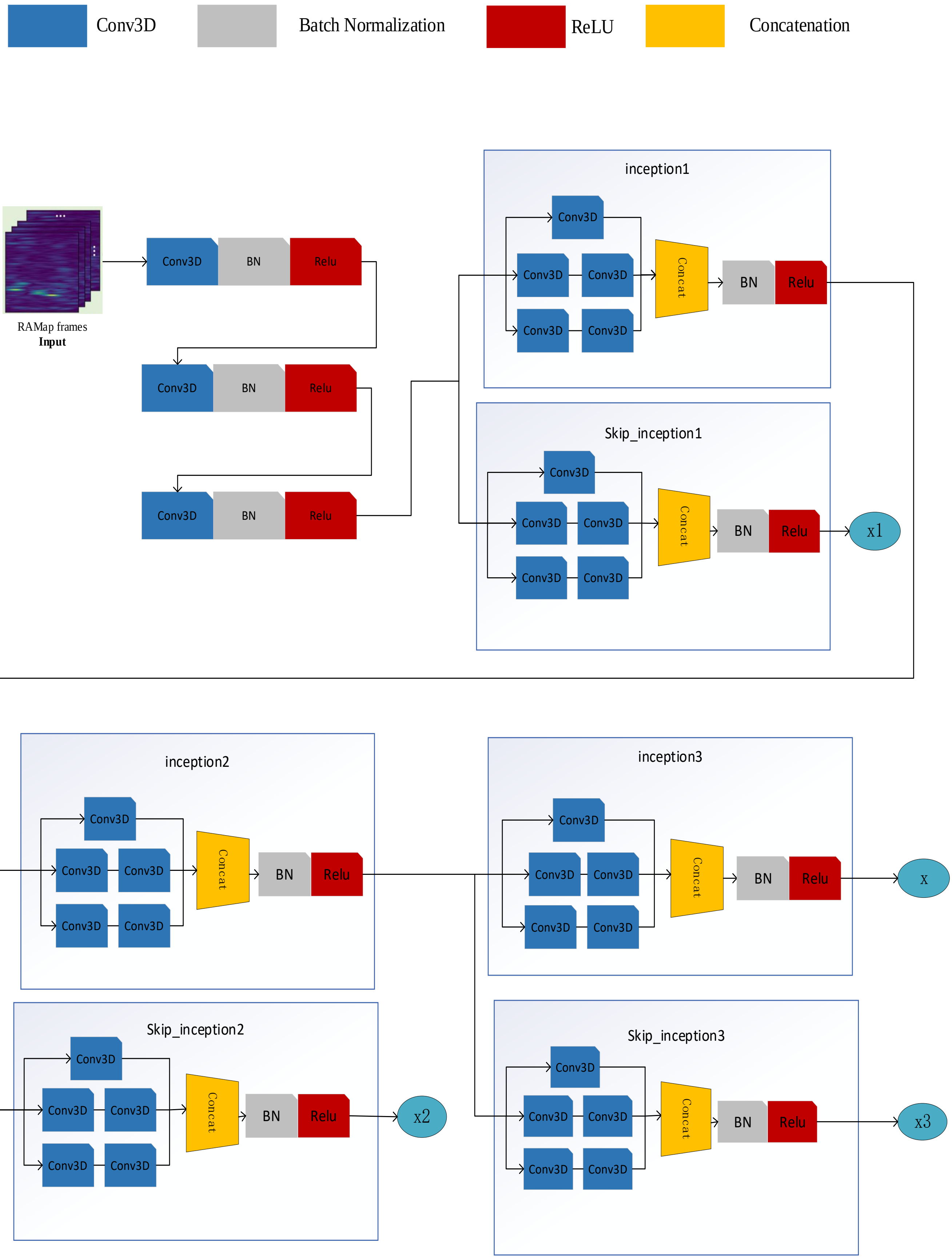


Figure 4. Hourglass three-dimensional convolutional neural network encoder with temporal inception convolution layers.

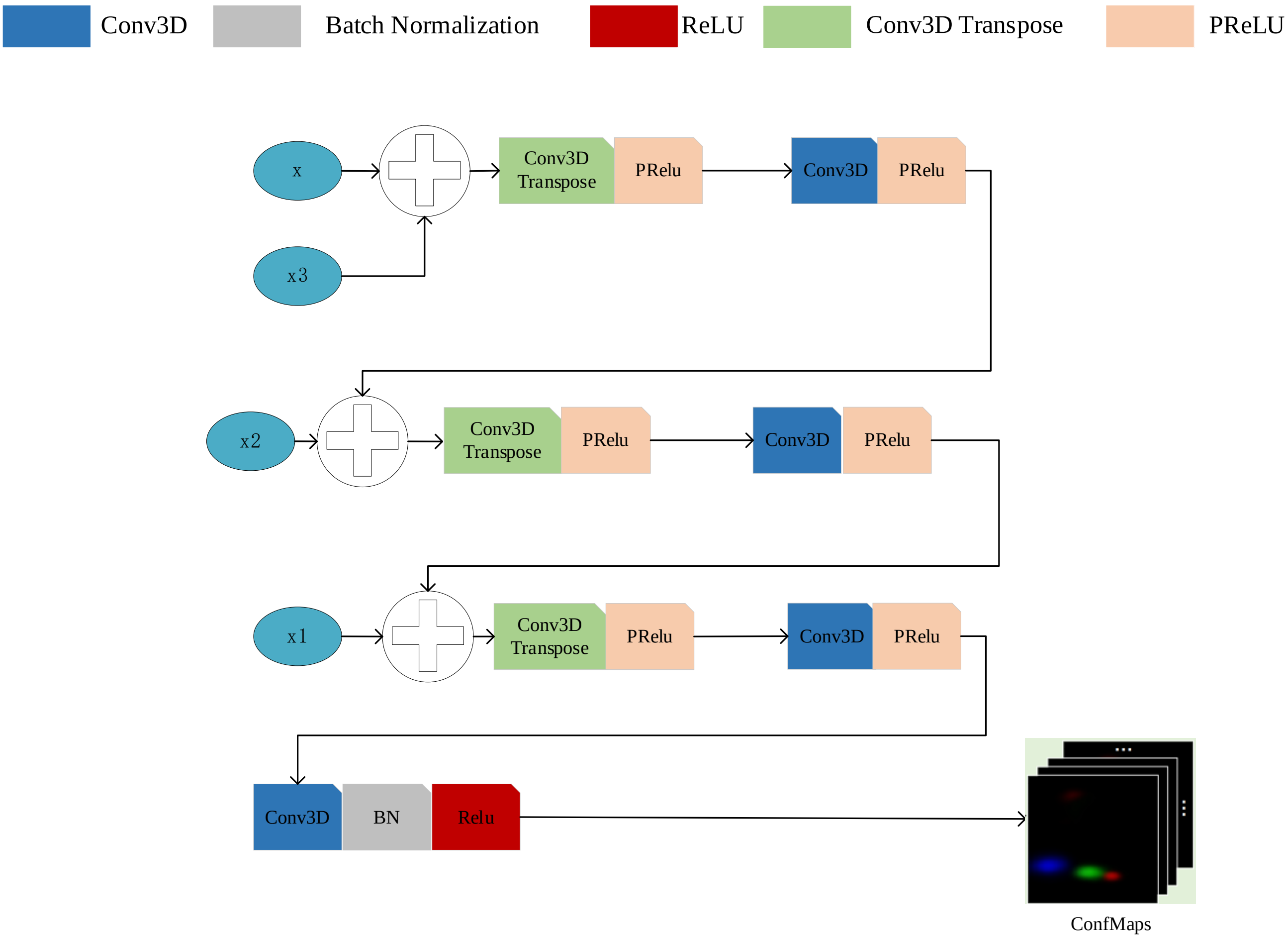
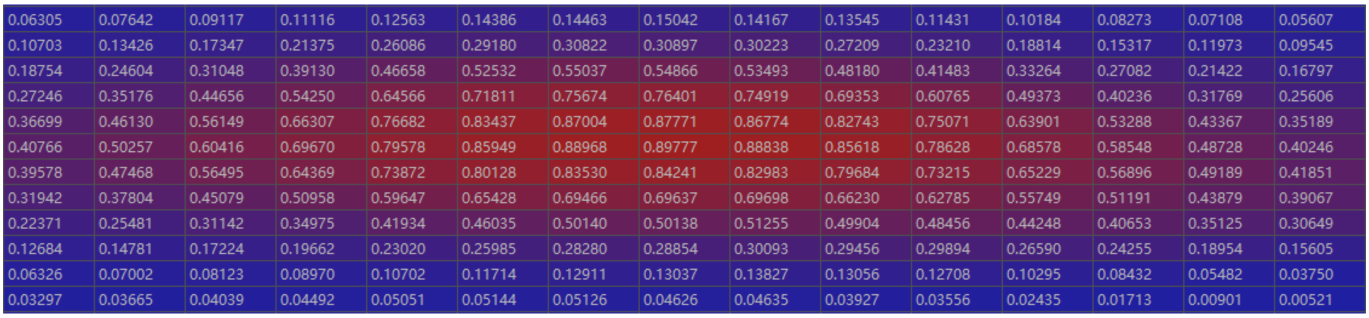


Figure 5. Hourglass three-dimensional convolutional neural network decoder with temporal inception convolution layers.

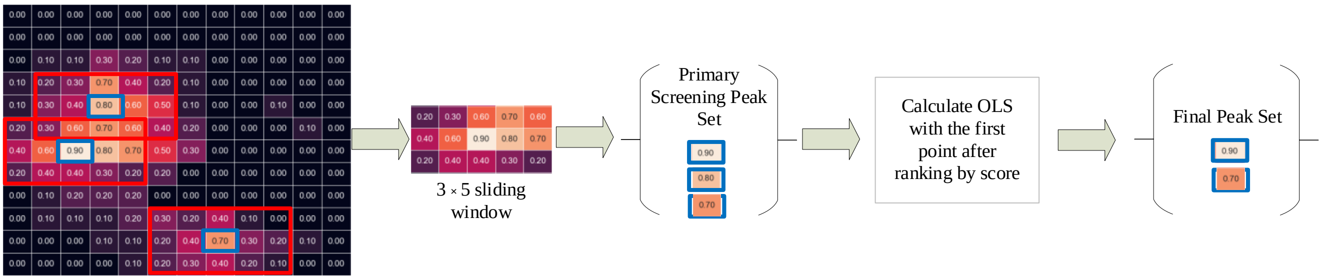


0.06305	0.07642	0.09117	0.11116	0.12563	0.14386	0.14463	0.15042	0.14167	0.13545	0.11431	0.10184	0.08273	0.07108	0.05607
0.10703	0.13426	0.17347	0.21375	0.26086	0.29180	0.30822	0.30897	0.30223	0.27209	0.23210	0.18814	0.15317	0.11973	0.09545
0.18754	0.24604	0.31048	0.39130	0.46658	0.52532	0.55037	0.54866	0.53493	0.48180	0.41483	0.33264	0.27082	0.21422	0.16797
0.27246	0.35176	0.44656	0.54250	0.64566	0.71811	0.75674	0.76401	0.74919	0.69353	0.60765	0.49373	0.40236	0.31769	0.25606
0.36699	0.46130	0.56149	0.66307	0.76682	0.83437	0.87004	0.87771	0.86774	0.82743	0.75071	0.63901	0.53288	0.43367	0.35189
0.40766	0.50257	0.60416	0.69670	0.79578	0.85949	0.88968	0.89777	0.88838	0.85618	0.78628	0.68578	0.58548	0.48728	0.40246
0.39578	0.47468	0.56495	0.64369	0.73872	0.80128	0.83530	0.84241	0.82983	0.79684	0.73215	0.65229	0.56896	0.49189	0.41851
0.31942	0.37804	0.45079	0.50958	0.59647	0.65428	0.69466	0.69637	0.69698	0.66230	0.62785	0.55749	0.51191	0.43879	0.39067
0.22371	0.25481	0.31142	0.34975	0.41934	0.46035	0.50140	0.50138	0.51255	0.49904	0.48456	0.44248	0.40653	0.35125	0.30649
0.12684	0.14781	0.17224	0.19662	0.23020	0.25985	0.28280	0.28854	0.30093	0.29456	0.29894	0.26590	0.24255	0.18954	0.15605
0.06326	0.07002	0.08123	0.08970	0.10702	0.11714	0.12911	0.13037	0.13827	0.13056	0.12708	0.10295	0.08432	0.05482	0.03750
0.03297	0.03665	0.04039	0.04492	0.05051	0.05144	0.05126	0.04626	0.04635	0.03927	0.03556	0.02435	0.01713	0.00901	0.00521

Figure 6. ConfMap diagram of a single pedestrian. The value for each grid in the figure corresponds to the probability that the target is a pedestrian. Because the channel where the pedestrian is located corresponds to the first channel in three-primary colors (RGB), the higher the probability value, the deeper the corresponding grid red.

2.2. *Object-Location Similarity (OLS) Limitation Analysis*

In the process of target-location processing [25], the author proposes a peak detection algorithm. The algorithm traverses ConfMap through a 3 × 5 sliding window and obtains the final location result by filtering the peak and recursively calculating OLS. The flow chart for target-location detection using OLS is shown in Figure 7.



0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.10	0.10	0.30	0.20	0.10	0.10	0.00	0.00	0.00	0.00	0.00
0.10	0.20	0.30	0.70	0.40	0.20	0.10	0.00	0.00	0.00	0.00	0.00
0.10	0.30	0.40	0.80	0.60	0.50	0.10	0.00	0.00	0.10	0.00	0.00
0.20	0.30	0.60	0.70	0.60	0.40	0.20	0.00	0.00	0.00	0.10	0.00
0.40	0.60	0.90	0.80	0.70	0.50	0.30	0.00	0.00	0.00	0.00	0.00
0.20	0.40	0.40	0.30	0.20	0.20	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.10	0.20	0.20	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.10	0.10	0.10	0.20	0.30	0.20	0.40	0.10	0.00	0.00	0.00
0.00	0.00	0.00	0.10	0.10	0.20	0.40	0.70	0.30	0.20	0.10	0.00
0.00	0.00	0.00	0.00	0.10	0.20	0.30	0.40	0.20	0.10	0.00	0.00

0.20	0.30	0.60	0.70	0.60
0.40	0.60	0.90	0.80	0.70
0.20	0.40	0.40	0.30	0.20

3 × 5 sliding window
Primary Screening Peak Set
0.90
0.80
0.70
Calculate OLS with the first point after ranking by score
Final Peak Set
0.90
0.70

Figure 7. OLS target-location detection flow chart.

The formula of OLS proposed in [25] is as follows:

$$OLS = \exp\left\{\frac{-d^2}{2(s \times k_{cls})^2}\right\} \tag{1}$$

where d is the distance (in meters) between any two points in a radar RA image; s is the object distance from the radar sensor, representing object scale information; and κ_{cls} is a per-class constant that represents the error tolerance for class cls, which can be determined by the object average size of the corresponding class. κ_{cls} is a priori setting such that OLS can be reasonably distributed between 0 and 1.

The OLS can be interpreted as a Gaussian distribution, with distance d as the bias and $(s \times \kappa_{cls})^2$ as the variance. Therefore, OLS is a similarity measure which also considers object size and distance, so it is more reasonable than other traditional distance measures (such as Euclidean distance, Mahalanobis distance, etc). This OLS metric is also used to match detections and ground truth for evaluation purposes.

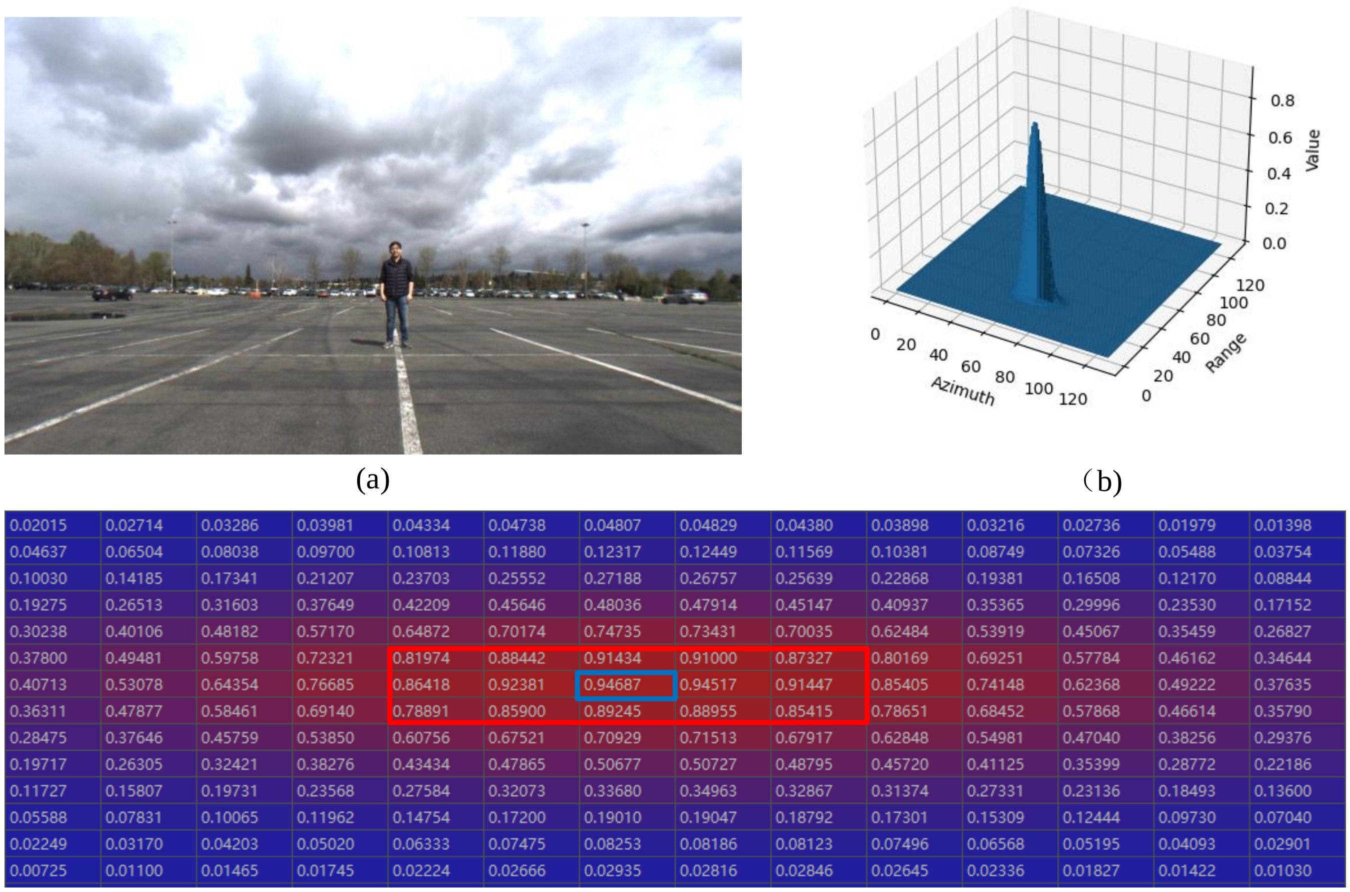
According to the OLS target-location detection method, we can infer that the author makes three assumptions as follows:

- Point target hypothesis;
- No point target overlap will occur in ConfMap;
- The value distribution of ConfMap is related to the corresponding probability distribution of point target location and class;
- Considering the above assumptions, there are several typical cases, which can be analyzed as below.

Case (1): Single target in ConfMap

There is only one target in the ConfMap area, but multiple grids with high probability values appear. In this case, we first need to analyze the reasons for this situation. Because the input of RODNet is radar RA map frames, that is, the original radar data are obtained by FFTs transforms, each frame of the radar RA map is obtained by multiple chirp combinations, and each chirp acquisition is affected by the randomness of radar positioning, that is, the radar accuracy itself is a random quantity, which may cause the process of extracting features in multiple chirp combinations to be extracted into multiple high probability points. In addition, because the target collected by the radar itself should be a distributed target during the acquisition process, the point target assumption is not satisfied, which will result in multiple strong points in the data and eventually be extracted into multiple high probability grids. Finally, the adjacent target may produce a multipath effect, which will also affect the point target hypothesis. The above results in the situation of case (1).

For a single pedestrian target, because such targets are small, the initial screening process can directly obtain the peak value, so the method can obtain good results. Figure 8 shows this case.



(a)

（b)

0.02015	0.02714	0.03286	0.03981	0.04334	0.04738	0.04807	0.04829	0.04380	0.03898	0.03216	0.02736	0.01979	0.01398
0.04637	0.06504	0.08038	0.09700	0.10813	0.11880	0.12317	0.12449	0.11569	0.10381	0.08749	0.07326	0.05488	0.03754
0.10030	0.14185	0.17341	0.21207	0.23703	0.25552	0.27188	0.26757	0.25639	0.22868	0.19381	0.16508	0.12170	0.08844
0.19275	0.26513	0.31603	0.37649	0.42209	0.45646	0.48036	0.47914	0.45147	0.40937	0.35365	0.29996	0.23530	0.17152
0.30238	0.40106	0.48182	0.57170	0.64872	0.70174	0.74735	0.73431	0.70035	0.62484	0.53919	0.45067	0.35459	0.26827
0.37800	0.49481	0.59758	0.72321	0.81974	0.88442	0.91434	0.91000	0.87327	0.80169	0.69251	0.57784	0.46162	0.34644
0.40713	0.53078	0.64354	0.76685	0.86418	0.92381	0.94687	0.94517	0.91447	0.85405	0.74148	0.62368	0.49222	0.37635
0.36311	0.47877	0.58461	0.69140	0.78891	0.85900	0.89245	0.88955	0.85415	0.78651	0.68452	0.57868	0.46614	0.35790
0.28475	0.37646	0.45759	0.53850	0.60756	0.67521	0.70929	0.71513	0.67917	0.62848	0.54981	0.47040	0.38256	0.29376
0.19717	0.26305	0.32421	0.38276	0.43434	0.47865	0.50677	0.50727	0.48795	0.45720	0.41125	0.35399	0.28772	0.22186
0.11727	0.15807	0.19731	0.23568	0.27584	0.32073	0.33680	0.34963	0.32867	0.31374	0.27331	0.23136	0.18493	0.13600
0.05588	0.07831	0.10065	0.11962	0.14754	0.17200	0.19010	0.19047	0.18792	0.17301	0.15309	0.12444	0.09730	0.07040
0.02249	0.03170	0.04203	0.05020	0.06333	0.07475	0.08253	0.08186	0.08123	0.07496	0.06568	0.05195	0.04093	0.02901
0.00725	0.01100	0.01465	0.01745	0.02224	0.02666	0.02935	0.02816	0.02846	0.02645	0.02336	0.01827	0.01422	0.01030

(c)

Figure 8. Case (1) data. (**a**) Optical photo of case (1). (**b**) A 2D ConfMap of case (1). (**c**) A 3D ConfMap of case (1). The red box in (**c**) represents a 3 × 5 slider, and the blue box represents the peak that is finally filtered.

Case (2): No overlapped multiple targets in ConfMap

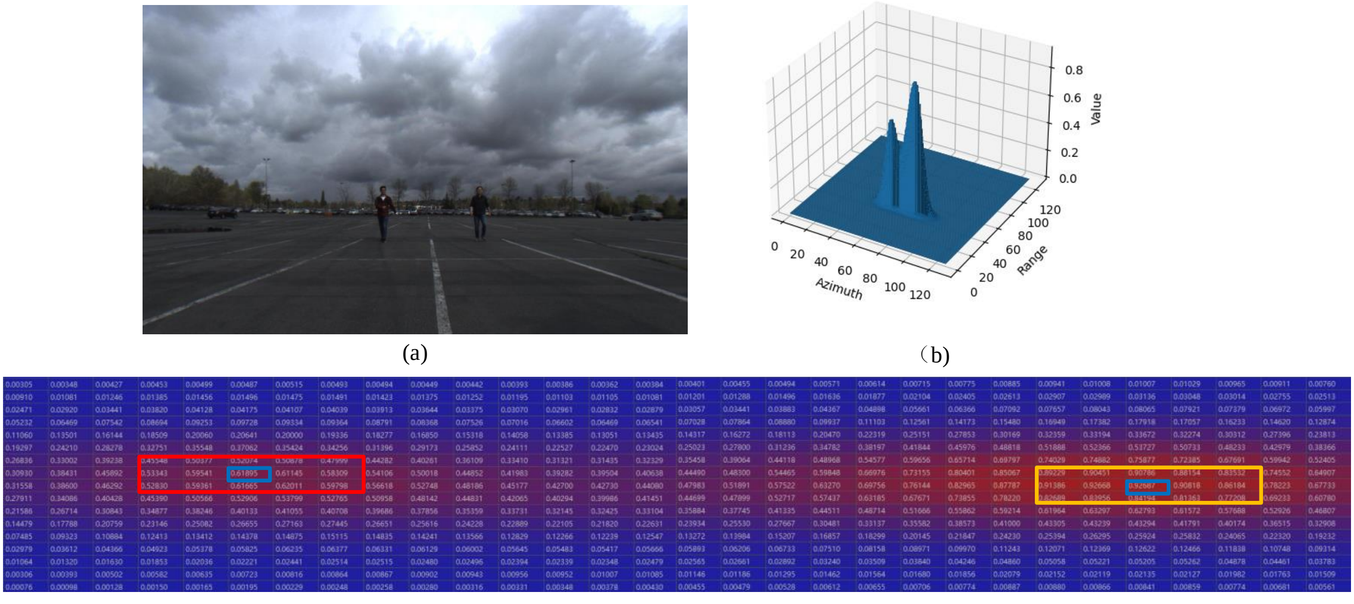
There are two targets in the region of the ConfMap; they are far apart, and each of the two targets has a peak. The schematic diagram of this case is shown in Figure 9. When the initial screening is completed, two peaks can be obtained. In the process of calculating OLS, the larger the d is, the smaller the exponential part is, and the OLS is closer to 0. This target-location processing method can also give accurate results.

Case (3): Type 1 overlapped multiple targets in ConfMap

There are two targets in the region of the ConfMap, but there is only one grid with a high probability value. Case (3) is shown in Figure 10. Because only one peak point is retained during the initial screening process, the result is missed.

Case (4): Type 2 overlapped multiple targets in ConfMap

There are two targets in the region of the ConfMap, but the high probability grid points are close, or even overlap. Case (4) is shown in Figure 11. When two grids with high probability appear at the same time at close range, d will become smaller in OLS calculation, the index part will approach 0, and OLS will become larger. When the threshold is exceeded, another strong point will be suppressed. In this case, OLS will also result in missed detection.



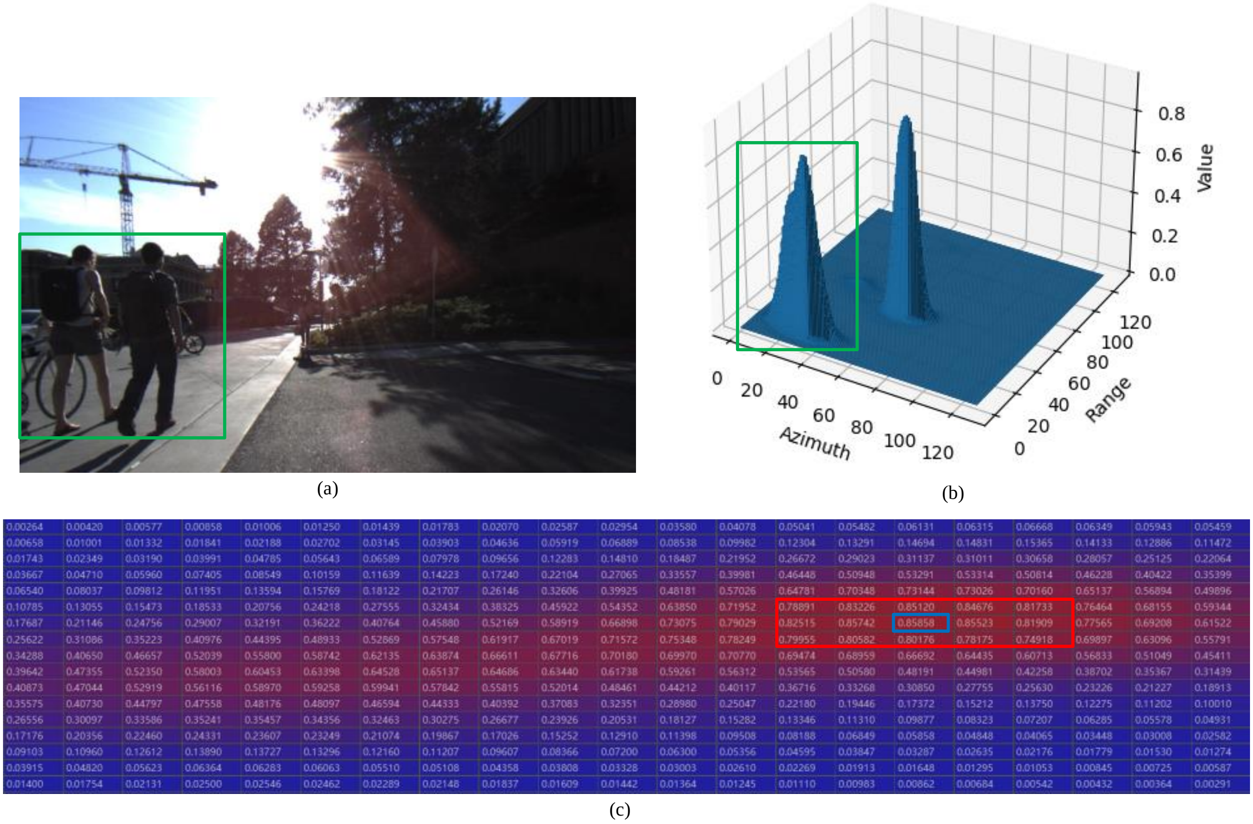
(a)

（b)

0.00305	0.00348	0.00427	0.00453	0.00499	0.00487	0.00515	0.00493	0.00494	0.00449	0.00442	0.00393	0.00386	0.00362	0.00384	0.00401	0.00455	0.00494	0.00571	0.00614	0.00715	0.00775	0.00885	0.00941	0.01008	0.01007	0.01029	0.00965	0.00911	0.00760
0.00910	0.01081	0.01246	0.01385	0.01456	0.01496	0.01475	0.01491	0.01423	0.01375	0.01252	0.01195	0.01103	0.01105	0.01081	0.01201	0.01288	0.01496	0.01636	0.01877	0.02104	0.02405	0.02613	0.02907	0.02989	0.03136	0.03048	0.03014	0.02755	0.02513
0.02471	0.02920	0.03441	0.03820	0.04128	0.04175	0.04107	0.04039	0.03913	0.03644	0.03375	0.03070	0.02961	0.02832	0.02879	0.03057	0.03441	0.03883	0.04367	0.04898	0.05661	0.06366	0.07092	0.07657	0.08043	0.08065	0.07921	0.07379	0.06972	0.05997
0.05232	0.06469	0.07542	0.08694	0.09253	0.09728	0.09334	0.09364	0.08791	0.08368	0.07526	0.07016	0.06602	0.06469	0.06541	0.07028	0.07864	0.08880	0.09937	0.11103	0.12561	0.14173	0.15480	0.16949	0.17382	0.17918	0.17057	0.16233	0.14620	0.12874
0.11060	0.13501	0.16144	0.18509	0.20060	0.20641	0.20000	0.19336	0.18277	0.16850	0.15318	0.14058	0.13385	0.13051	0.13435	0.14317	0.16272	0.18113	0.20470	0.22319	0.25151	0.27853	0.30169	0.32359	0.33194	0.33672	0.32274	0.30312	0.27396	0.23813
0.19297	0.24210	0.28278	0.32751	0.35548	0.37062	0.35424	0.34256	0.31396	0.29173	0.25852	0.24111	0.22527	0.22470	0.23024	0.25023	0.27800	0.31236	0.34782	0.38197	0.41844	0.45976	0.48818	0.51888	0.52366	0.53727	0.50733	0.48233	0.42979	0.38366
0.26836	0.33002	0.39238	0.45548	0.50377	0.52074	0.50878	0.47999	0.44282	0.40261	0.36109	0.33410	0.31321	0.31435	0.32329	0.35458	0.39064	0.44118	0.48968	0.54577	0.59656	0.65714	0.69797	0.74029	0.74882	0.75877	0.72385	0.67691	0.59942	0.52405
0.30930	0.38431	0.45892	0.53343	0.59541	0.61895	0.61145	0.58309	0.54106	0.50018	0.44852	0.41983	0.39282	0.39504	0.40638	0.44490	0.48300	0.54465	0.59848	0.66976	0.73155	0.80401	0.85067	0.89229	0.90451	0.90786	0.88154	0.83532	0.74552	0.64907
0.31558	0.38600	0.46292	0.52830	0.59361	0.61665	0.62011	0.59798	0.56618	0.52748	0.48186	0.45177	0.42700	0.42730	0.44080	0.47983	0.51891	0.57522	0.63270	0.69756	0.76144	0.82965	0.87787	0.91386	0.92668	0.92687	0.90818	0.86184	0.78223	0.67733
0.27911	0.34086	0.40428	0.45390	0.50566	0.52906	0.53799	0.52765	0.50958	0.48142	0.44831	0.42065	0.40294	0.39986	0.41451	0.44699	0.47899	0.52717	0.57437	0.63185	0.67671	0.73855	0.78220	0.82689	0.83956	0.84194	0.81363	0.77208	0.69233	0.60780
0.21586	0.26714	0.30843	0.34877	0.38246	0.40133	0.41055	0.40708	0.39686	0.37856	0.35359	0.33731	0.32145	0.32425	0.33104	0.35884	0.37745	0.41335	0.44511	0.48714	0.51666	0.55862	0.59214	0.61964	0.63297	0.62793	0.61572	0.57688	0.52926	0.46807
0.14479	0.17788	0.20759	0.23146	0.25082	0.26655	0.27163	0.27445	0.26651	0.25616	0.24228	0.22889	0.22105	0.21820	0.22631	0.23934	0.25530	0.27667	0.30481	0.33137	0.35582	0.38573	0.41000	0.43305	0.43239	0.43294	0.41791	0.40174	0.36515	0.32908
0.07485	0.09323	0.10884	0.12413	0.13412	0.14378	0.14875	0.15115	0.14835	0.14241	0.13566	0.12829	0.12266	0.12239	0.12547	0.13272	0.13984	0.15207	0.16857	0.18299	0.20145	0.21847	0.24230	0.25394	0.26295	0.25924	0.25832	0.24065	0.22320	0.19232
0.02979	0.03612	0.04366	0.04923	0.05378	0.05825	0.06235	0.06377	0.06331	0.06129	0.06002	0.05645	0.05483	0.05417	0.05666	0.05893	0.06206	0.06733	0.07510	0.08158	0.08971	0.09970	0.11243	0.12071	0.12369	0.12622	0.12466	0.11838	0.10748	0.09314
0.01064	0.01320	0.01630	0.01853	0.02036	0.02221	0.02441	0.02514	0.02515	0.02480	0.02496	0.02394	0.02339	0.02348	0.02479	0.02565	0.02661	0.02892	0.03240	0.03509	0.03840	0.04246	0.04860	0.05058	0.05221	0.05205	0.05262	0.04878	0.04461	0.03783
0.00306	0.00393	0.00502	0.00582	0.00635	0.00723	0.00816	0.00864	0.00867	0.00902	0.00943	0.00956	0.00952	0.01007	0.01085	0.01146	0.01186	0.01295	0.01462	0.01564	0.01680	0.01856	0.02079	0.02152	0.02119	0.02135	0.02127	0.01982	0.01763	0.01509
0.00076	0.00098	0.00128	0.00150	0.00165	0.00195	0.00229	0.00248	0.00258	0.00280	0.00316	0.00331	0.00348	0.00378	0.00430	0.00455	0.00479	0.00528	0.00612	0.00655	0.00706	0.00774	0.00887	0.00880	0.00866	0.00841	0.00859	0.00774	0.00681	0.00561

(c)

Figure 9. Case (2) data. (**a**) Optical photo of case (2). (**b**) A 2D ConfMap of case (2). (**c**) A 3D ConfMap of case (2). The red and yellow box in (**c**) represents a 3 × 5 slider, and the blue box represents the result of the primary screening peak.



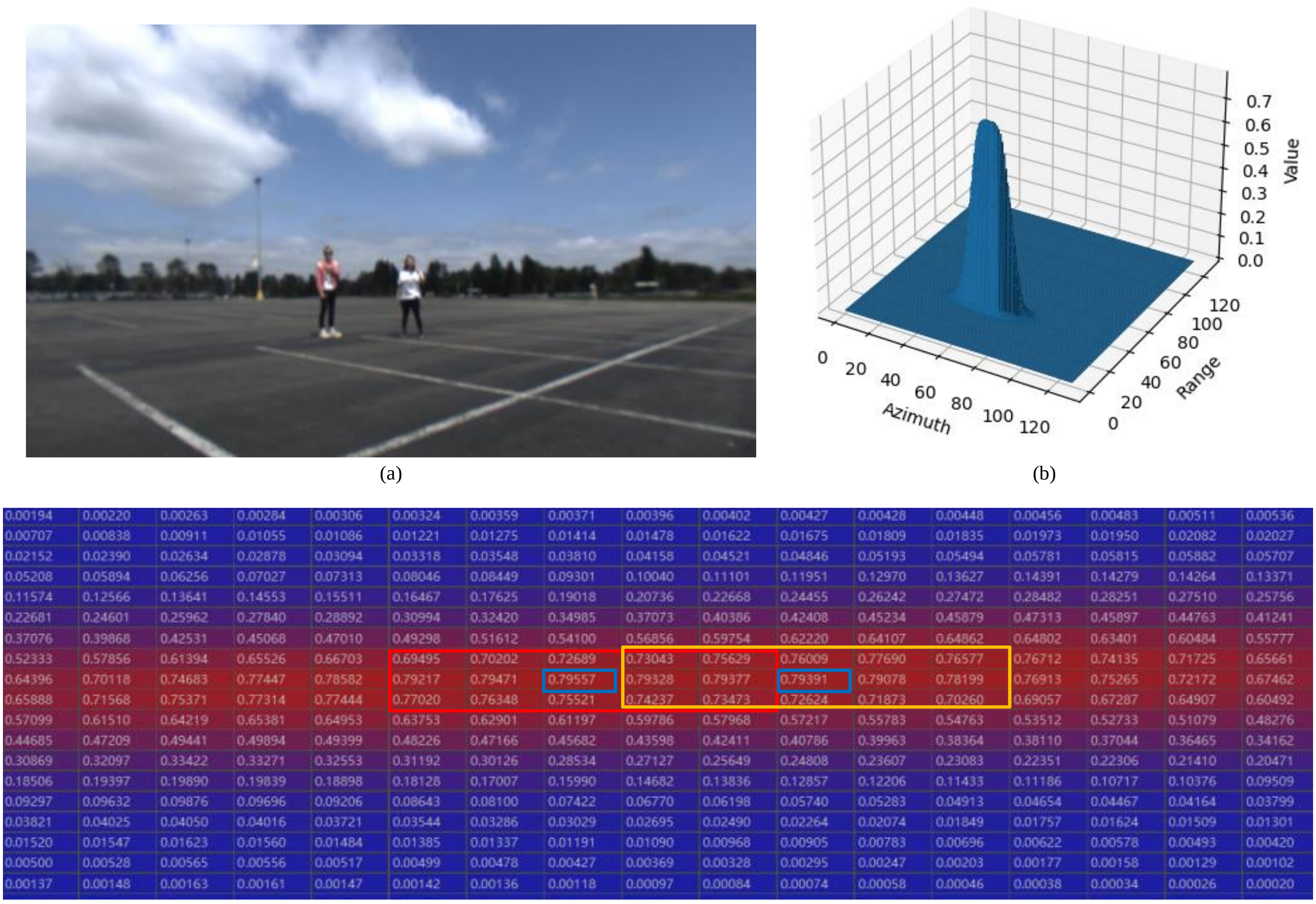
(a)

(b)

0.00264	0.00420	0.00577	0.00858	0.01006	0.01250	0.01439	0.01783	0.02070	0.02587	0.02954	0.03580	0.04078	0.05041	0.05482	0.06131	0.06315	0.06668	0.06349	0.05943	0.05459
0.00658	0.01001	0.01332	0.01841	0.02188	0.02702	0.03145	0.03903	0.04636	0.05919	0.06889	0.08538	0.09982	0.12304	0.13291	0.14694	0.14831	0.15365	0.14133	0.12886	0.11472
0.01743	0.02349	0.03190	0.03991	0.04785	0.05643	0.06589	0.07978	0.09656	0.12283	0.14810	0.18487	0.21952	0.26672	0.29023	0.31137	0.31011	0.30658	0.28057	0.25125	0.22064
0.03667	0.04710	0.05960	0.07405	0.08549	0.10159	0.11639	0.14223	0.17240	0.22104	0.27065	0.33557	0.39981	0.46448	0.50948	0.53291	0.53314	0.50814	0.46228	0.40422	0.35399
0.06540	0.08037	0.09812	0.11951	0.13594	0.15769	0.18122	0.21707	0.26146	0.32606	0.39925	0.48181	0.57026	0.64781	0.70348	0.73144	0.73026	0.70160	0.65137	0.56894	0.49896
0.10785	0.13055	0.15473	0.18533	0.20756	0.24218	0.27555	0.32434	0.38325	0.45922	0.54352	0.63850	0.71952	0.78891	0.83226	0.85120	0.84676	0.81733	0.76464	0.68155	0.59344
0.17687	0.21146	0.24756	0.29007	0.32191	0.36222	0.40764	0.45880	0.52169	0.58919	0.66898	0.73075	0.79029	0.82515	0.85742	0.85858	0.85523	0.81909	0.77565	0.69208	0.61522
0.25622	0.31086	0.35223	0.40976	0.44395	0.48933	0.52869	0.57548	0.61917	0.67019	0.71572	0.75348	0.78249	0.79955	0.80582	0.80176	0.78175	0.74918	0.69897	0.63096	0.55791
0.34288	0.40650	0.46657	0.52039	0.55800	0.58742	0.62135	0.63874	0.66611	0.67716	0.70180	0.69970	0.70770	0.69474	0.68959	0.66692	0.64435	0.60713	0.56833	0.51049	0.45411
0.39642	0.47355	0.52350	0.58003	0.60453	0.63398	0.64528	0.65137	0.64686	0.63440	0.61738	0.59261	0.56312	0.53565	0.50580	0.48191	0.44981	0.42258	0.38702	0.35367	0.31439
0.40873	0.47044	0.52919	0.56116	0.58970	0.59258	0.59941	0.57842	0.55815	0.52014	0.48461	0.44212	0.40117	0.36716	0.33268	0.30850	0.27755	0.25630	0.23226	0.21227	0.18913
0.35575	0.40730	0.44797	0.47558	0.48176	0.48097	0.46594	0.44333	0.40392	0.37083	0.32351	0.28980	0.25047	0.22180	0.19446	0.17372	0.15212	0.13750	0.12275	0.11202	0.10010
0.26556	0.30097	0.33586	0.35241	0.35457	0.34356	0.32463	0.30275	0.26677	0.23926	0.20531	0.18127	0.15282	0.13346	0.11310	0.09877	0.08323	0.07207	0.06285	0.05578	0.04931
0.17176	0.20356	0.22460	0.24331	0.23607	0.23249	0.21074	0.19867	0.17026	0.15252	0.12910	0.11398	0.09508	0.08188	0.06849	0.05858	0.04848	0.04065	0.03448	0.03008	0.02582
0.09103	0.10960	0.12612	0.13890	0.13727	0.13296	0.12160	0.11207	0.09607	0.08366	0.07200	0.06300	0.05356	0.04595	0.03847	0.03287	0.02635	0.02176	0.01779	0.01530	0.01274
0.03915	0.04820	0.05623	0.06364	0.06283	0.06063	0.05510	0.05108	0.04358	0.03808	0.03328	0.03003	0.02610	0.02269	0.01913	0.01648	0.01295	0.01053	0.00845	0.00725	0.00587
0.01400	0.01754	0.02131	0.02500	0.02546	0.02462	0.02289	0.02148	0.01837	0.01609	0.01442	0.01364	0.01245	0.01110	0.00983	0.00862	0.00684	0.00542	0.00432	0.00364	0.00291

(c)

Figure 10. Case (3) data. (**a**) Optical photo of case (3). (**b**) A 2D ConfMap of case (3). (**c**) A 3D ConfMap of case (3). The red box in (**c**) represents a 3 × 5 slider, and the blue box represents the result of the primary screening peak. The area corresponding to (**c**) is the position indicated by the green box in (**a**,**b**).



(a)

(b)

0.00194	0.00220	0.00263	0.00284	0.00306	0.00324	0.00359	0.00371	0.00396	0.00402	0.00427	0.00428	0.00448	0.00456	0.00483	0.00511	0.00536
0.00707	0.00838	0.00911	0.01055	0.01086	0.01221	0.01275	0.01414	0.01478	0.01622	0.01675	0.01809	0.01835	0.01973	0.01950	0.02082	0.02027
0.02152	0.02390	0.02634	0.02878	0.03094	0.03318	0.03548	0.03810	0.04158	0.04521	0.04846	0.05193	0.05494	0.05781	0.05815	0.05882	0.05707
0.05208	0.05894	0.06256	0.07027	0.07313	0.08046	0.08449	0.09301	0.10040	0.11101	0.11951	0.12970	0.13627	0.14391	0.14279	0.14264	0.13371
0.11574	0.12566	0.13641	0.14553	0.15511	0.16467	0.17625	0.19018	0.20736	0.22668	0.24455	0.26242	0.27472	0.28482	0.28251	0.27510	0.25756
0.22681	0.24601	0.25962	0.27840	0.28892	0.30994	0.32420	0.34985	0.37073	0.40386	0.42408	0.45234	0.45879	0.47313	0.45897	0.44763	0.41241
0.37076	0.39868	0.42531	0.45068	0.47010	0.49298	0.51612	0.54100	0.56856	0.59754	0.62220	0.64107	0.64862	0.64802	0.63401	0.60484	0.55777
0.52333	0.57856	0.61394	0.65526	0.66703	0.69495	0.70202	0.72689	0.73043	0.75629	0.76009	0.77690	0.76577	0.76712	0.74135	0.71725	0.65661
0.64396	0.70118	0.74683	0.77447	0.78582	0.79217	0.79471	0.79557	0.79328	0.79377	0.79391	0.79078	0.78199	0.76913	0.75265	0.72172	0.67462
0.65888	0.71568	0.75371	0.77314	0.77444	0.77020	0.76348	0.75521	0.74237	0.73473	0.72624	0.71873	0.70260	0.69057	0.67287	0.64907	0.60492
0.57099	0.61510	0.64219	0.65381	0.64953	0.63753	0.62901	0.61197	0.59786	0.57968	0.57217	0.55783	0.54763	0.53512	0.52733	0.51079	0.48276
0.44685	0.47209	0.49441	0.49894	0.49399	0.48226	0.47166	0.45682	0.43598	0.42411	0.40786	0.39963	0.38364	0.38110	0.37044	0.36465	0.34162
0.30869	0.32097	0.33422	0.33271	0.32553	0.31192	0.30126	0.28534	0.27127	0.25649	0.24808	0.23607	0.23083	0.22351	0.22306	0.21410	0.20471
0.18506	0.19397	0.19890	0.19839	0.18898	0.18128	0.17007	0.15990	0.14682	0.13836	0.12857	0.12206	0.11433	0.11186	0.10717	0.10376	0.09509
0.09297	0.09632	0.09876	0.09696	0.09206	0.08643	0.08100	0.07422	0.06770	0.06198	0.05740	0.05283	0.04913	0.04654	0.04467	0.04164	0.03799
0.03821	0.04025	0.04050	0.04016	0.03721	0.03544	0.03286	0.03029	0.02695	0.02490	0.02264	0.02074	0.01849	0.01757	0.01624	0.01509	0.01301
0.01520	0.01547	0.01623	0.01560	0.01484	0.01385	0.01337	0.01191	0.01090	0.00968	0.00905	0.00783	0.00696	0.00622	0.00578	0.00493	0.00420
0.00500	0.00528	0.00565	0.00556	0.00517	0.00499	0.00478	0.00427	0.00369	0.00328	0.00295	0.00247	0.00203	0.00177	0.00158	0.00129	0.00102
0.00137	0.00148	0.00163	0.00161	0.00147	0.00142	0.00136	0.00118	0.00097	0.00084	0.00074	0.00058	0.00046	0.00038	0.00034	0.00026	0.00020

(c)

Figure 11. Case (4) data. (**a**) Optical photo of case (4). (**b**) A 2D ConfMap of case (4). (**c**) A 3D ConfMap of case (4). The red and yellow box in (**c**) represents a 3 × 5 slider, and the two blue box represents the primary screening peak. One of them will be suppressed.

In summary, in cases (1) and (2), the OLS-based location detection method performs well. However, in cases (3) and (4), the method has missed detection and has been verified by measured radar data, confirming the limitations of the method. The reason for case (3) is the missed detection of the sliding window. Furthermore, the reason for case (4) is that the d in Equation (1) is too small, which causes the value of OLS to become larger than the threshold, so one point is wrongly suppressed. We have considered increasing the OLS threshold or modifying the size of the sliding window. Although the former can ensure the adjacent peak point is retained, it may also lead to an error that should be suppressed. The latter situation may cause the peak point to be missed in dense cases (or overlapping). Simple adjustments cannot directly solve the problem, so a new method needs to be proposed. In the next chapter, we will introduce our new method in detail.

3. Dense Target-Location Detection Method Based on Maximum Likelihood Estimation

To solve the problem of missed detection of RODNet, we developed the Gaussian Mixture Model with target number (GMM-TN). It can be considered that this is also a clustering algorithm that automatically detects the best number of overlapping clusters. The K-means algorithm is the part of it which is used only to find the centroids of the assumed cluster number. Then, the Gaussian Mixture Model is used to simulate the ConfMap under this assumption. For each cluster number assumption, the KL distance between the simulated ConfMap and the ConfMap generated by RODNet is calculated.

If the KL distance is the shortest, the corresponding cluster number assumption will be considered as the best number of clusters.

3.1. Gaussian Mixture Model with Target Number

The assumptions used in the GMM-TN model are listed as follows:

- Point target hypothesis;
- Point target overlap may occur in ConfMap;
- Both the value distribution and occupied grid spatial distribution of ConfMap are related to the corresponding probability distribution of point target location and class.

The simulated ConfMap $P_{cls}(a,r)$ can be equivalent to the full probability of occurrence of class *cls* point targets at grid (a, r) in a 128 × 128 grid image (refer to the radar RA map in Figure 3) is shown in Equation (2).

$$P_{cls}(a,r) = \frac{1}{B}\sum_{n=0}^{N} P_{n,cls}(a,r) \tag{2}$$

where a and r are the column index and row index of radar RA map, respectively; B is the normalized coefficient of probability; and N is the number of targets in a single radar RA map. $P_{n,cls}(a,r)$ is the probability of occurrence of nth point target of class *cls* at grid (a, r), which is related with several hidden variable conditions and can be rewritten as follows

$$\begin{aligned} P_{n,cls}(a,r) &= P((a,r)|\sigma_{cls}, a_0(n), r_0(n)) \\ &= \frac{1}{2\pi}\exp\left\{-\frac{[(r-r_0(n))\times 2]^2 + (a-a_0(n))^2]}{2(\sigma_{cls})^2}\right\} \end{aligned} \tag{3}$$

$$\sigma_{cls} = 2\arctan(\frac{l_{cls}}{2r}) \times c_{cls} \tag{4}$$

where $r_0(n)$ and $a_0(n)$ are the row and column grid index of nth point target, respectively. The condition hidden variable σ_{cls}, as shown in Equation (4), represents the standard deviation of the size of the target in the ConfMap which is approximate to a Gaussian distribution, where l_{cls} is a class-related prior information, a pedestrian is 1, a cyclist is 2, and a car is 3. c_{cls} is also priori information. The value of pedestrians is 15, the value of cyclists is 20, and the value of cars is 30. The denominator is multiplied by 2 because the reflection pattern in the radar RA image is usually elliptical.

Considering that the key operation for dense target detection depends mainly on the azimuth resolution ability, we also can use the marginal probability distribution in azimuth direction. The two-dimensional probability density Equation (2) becomes:

$$P_{cls}(a) = \frac{1}{C}\sum_{r=1}^{128} P_{cls}(a,r) \tag{5}$$

where C is the normalized coefficient of probability.

The overall flow chart is shown in Figure 12. In the classification stage, we make use of exactly the same feature extraction part of RODNet—the network and weights of original RODNet to generate a ConfMap, as shown in Figure 12a. Because the problem we are dealing with is the location processing, our improvements to RODNet are shown in Figure 12b.

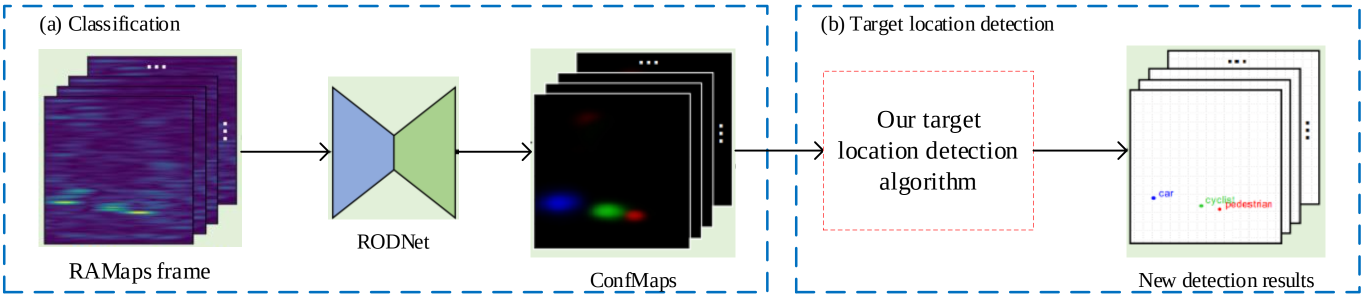


Figure 12. A flow chart of improved RODNet algorithm proposed in this paper.

3.2. Target Number Estimation Method

Based on the simulation method mentioned in Section 3.1, we combine clustering and KL distance to construct a set of target number determination methods. We first perform the following operations on the output results with ConfMap:

(1) Sort the scores of the results;
(2) Calculate the distance between each coordinate and the highest score coordinate in turn;
(3) If the threshold is exceeded, the grid coordinate is saved to another list; otherwise, the coordinates and the surrounding area greater than 0.3 enter our algorithm;
(4) Loop steps (2) and (3) until the last one is over and repeat steps (2) and (3) in the 'another list'.

Our target-location detection algorithm process is shown in the Figure 13.

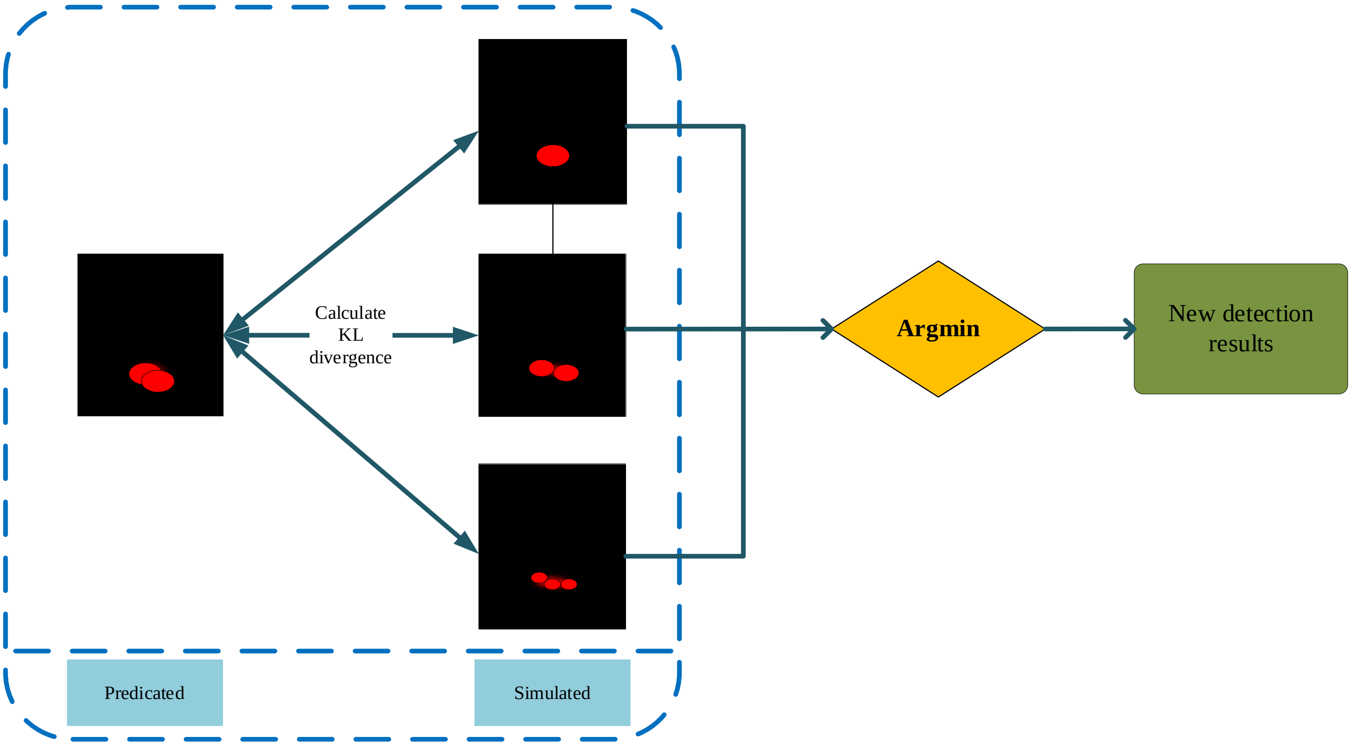


Figure 13. Flow chart of target number estimation method.

The overall process of the method is as follows. First, the ConfMaps output from the RODNet network is extracted. Because the number of targets and the location of point targets are unknown in the ConfMap with overlap, the Maximum A Posteriori parameter estimation framework is constructed by combining the GMM-TN and the predicted ConfMap as a condition to determine the number and location of point targets. Because we aim at the part of the network after the completion of the classification of the ConfMap, no matter how the network structure is designed, this method can ensure that after classification, our target-location detection will improve the accuracy of the results. The subsequent experimental section will involve further discussion of the results.

In order to obtain $r_0(n)$ and $a_0(n)$ in Equation (3), we assume each clustering center of a ConfMap occupied grid spatial distribution as the nth point target center with target number as a predefined condition. There are four categories of clustering method: the partition clustering method, the density-based clustering method, the hierarchical clustering method, and the new method. K-means, which is used to perform this operation, is one of the most widely used partition clustering algorithms [38,39]. It should be noted that not only the K-means method but also any clustering algorithm with a pre-defined number of clusters and based on the geometric distance between sample points can be applied to the GMM-TN algorithm to find the clustering centroids.

After determining the target center and the simulation method, it is necessary to measure the difference between two kinds of distribution: one is the normalized ConfMaps predicated by RODNet, expressed by $P_{cls}^{RODNet}(a,r)$. Another distribution is the simulated ConfMap $P_{cls}(a,r)$ defined in Equation (2).

KL divergence is a widely used measure of the similarity between two probability distributions. In this paper, the two-dimensional KL divergence is defined as:

$$KL_{2D} = D(P_{cls}^{RODNet}(a,r)||P_{cls}(a,r)) = \sum_{a=1}^{128}\sum_{r=1}^{128} P_{cls}^{RODNet}(a,r)\log_2\frac{P_{cls}^{RODNet}(a,r)}{P_{cls}(a,r)} \tag{6}$$

As the KL divergence is asymmetric, we can modify Equations (6) to (7):

$$\begin{aligned} KL_{2D} &= D(P_{cls}^{RODNet}(a,r)||P_{cls}(a,r)) + D(P_{cls}(a,r)||P_{cls}^{RODNet}(a,r)) \\ &= \sum_{a=1}^{128}\sum_{r=1}^{128} P_{cls}^{RODNet}(a,r)\log_2\tfrac{P_{cls}^{RODNet}(a,r)}{P_{cls}(a,r)} + \sum_{a=1}^{128}\sum_{r=1}^{128} P_{cls}(a,r)\log_2\tfrac{P_{cls}(a,r)}{P_{cls}^{RODNet}(a,r)} \end{aligned} \tag{7}$$

According to Equation (7), the smaller the KL_{2D} is, the smaller two-dimensional probability distribution difference is, and vice versa.

Based on Equation (5), it is reasonable to consider the azimuth-only case; therefore, the one-dimensional KL divergence can be defined as:

$$\begin{aligned} KL_{1D} &= D(P_{cls}^{RODNet}(a)||P_{cls}(a)) + D(P_{cls}(a)||P_{cls}^{RODNet}(a)) \\ &= \sum_{a=1}^{128} P_{cls}^{RODNet}(a)\log_2\tfrac{P_{cls}^{RODNet}(a)}{P_{cls}(a)} + \sum_{a=1}^{128} P_{cls}(a)\log_2\tfrac{P_{cls}(a)}{P_{cls}^{RODNet}(a)} \end{aligned} \tag{8}$$

Two maximum likelihood target-number estimation methods are proposed based on two-dimensional and one-dimensional symmetric KL divergence Equations (7) and (8), which are named the 2DMLKL and 1DMLKL method, respectively.

At last, the estimated number of targets with the smallest one-dimensional or two-dimensional symmetric KL divergence can be calculated by Equation (9) or (10):

$$\hat{N} = \underset{N\in\{0,1,\cdots,N_{\max}\}}{\text{Argmin}} \; KL_{1D} \tag{9}$$

$$\hat{N} = \underset{N\in\{0,1,\cdots,N_{\max}\}}{\text{Argmin}} \; KL_{2D} \tag{10}$$

where $N_{\max}$ is the predefined maximum possible number of targets in a radar RA image.

The final target number and the corresponding cluster center result are taken as the output, and the probability value corresponding to the cluster center in the predicted ConfMap is taken as the final score.

4. Experiments and Analysis

4.1. Experimental Data and Processing Steps

The dataset used in this paper is part of an open-source dataset called Camera-Radar of the University of Washington (CRUW) [40], which uses the format of radar RA images. The sensor platform includes a pair of stereo cameras and a 77 GHz millimeter-wave FMCW radar. The assembled and mounted sensors are well calibrated and synchronized. The stereo camera setup is designed to provide ground truth values for the dataset. The dataset contains over 3 h of 30 FPS (approximately 400 K frames) camera radar data for different driving scenarios, including campus roads, city streets, highways, parking lots, etc. In addition, it also includes several visual failure scenes with very poor image quality, namely darkness, strong light, blur, etc.

For the content of this paper, we reconstruct the CRUW dataset and divide it into two parts: Dense Pedestrian scene (for short DP scene) and Non-Dense Pedestrian scene (for short NDP scene). It should be mentioned that in the original CRUW dataset, the data containing pedestrians accounts for about 50% of the total CRUW. In the data containing pedestrians, the scene with dense pedestrians accounts for about 8%.

The specific experimental arrangements are as follows. First, we trained RODNet. The encoding–decoding structure we use is shown in Figures 4 and 5. RODNet is based on PyTorch [41]. The models are optimized with an Adam optimizer. The network is trained with a batch size of 2 on NVIDIA RTX2080Ti GPU. The initial learning rate was 1×10^{-5}, and the learning rate step was 5. The optimization is set to stop after 100 epochs. Then we select the best model in the training effect and test with target-location detection using OLS, 2DMLKL and 1DMLKL, respectively. Finally, we will show the four typical cases mentioned in Section 2 with typical scene data. The result graph will show the ConfMap results under the assumption of different target numbers simulated by GMM-TN and K-means clustering, and it can be visually compared with the ConfMap predicted by RODNet.

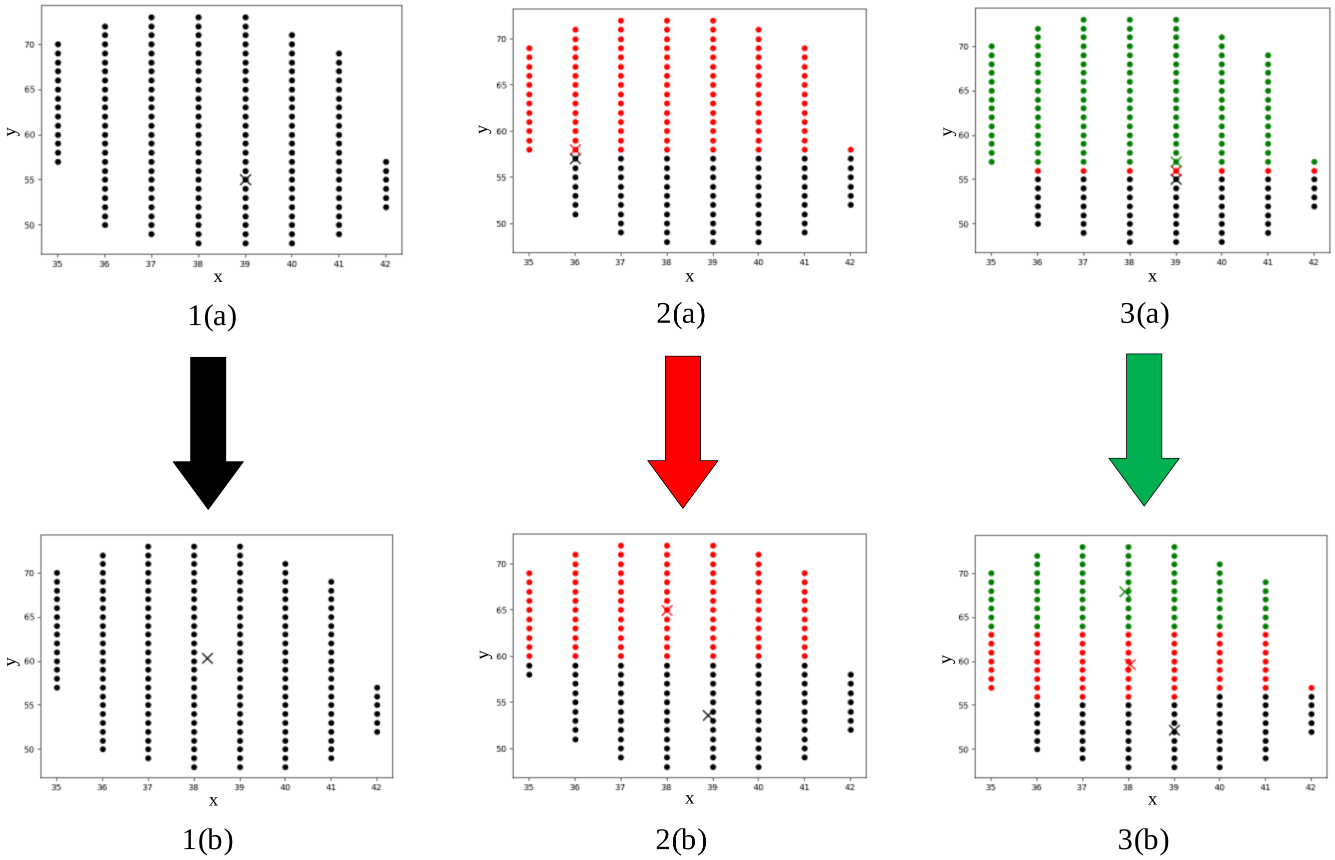
Later, we will introduce the process of our method in detail and show the comparison of the results of target-location detection using OLS and target-location detection using 2DMLKL and 1DMLKL, respectively, for typical scene data. Then, we will compare and verify the difference between the 1DMLKL and 2DMLKL methods and finally give the overall evaluation.

4.2. Typical Sample Results and Analysis

4.2.1. GMM-TN Model Based ConfMap Simulation Results

The process is shown in Figure 13. We take the predicated ConfMap of RODNet as input. For better simulation, we set a threshold of 0.3. The grids with values exceeding the threshold are retained, and those less than the threshold are no longer considered, and the values in these grids less than the threshold are changed to 0. Finally, we normalize the ConfMap after threshold filtering.

Because the value of the ConfMap grid is the probability value that the grid has a target, instead of the probability value that the grid belongs to a certain cluster, it cannot be used as the input of the cluster. Therefore, in the process of clustering, we only consider occupying the grid and use the K-means method to determine the target center. The evolution of clustering iteration is shown in Figure 14.



1(a) 2(a) 3(a)

1(b) 2(b) 3(b)

Figure 14. Evolutionary Graph of Clustering Iteration. The horizontal and vertical coordinates in the figure represent range grid coordinates and azimuth grid coordinates, respectively. (**1a**–**3a**) are processed predicated ConfMap of RODNet. We use the assumed number of targets as the input of clustering, randomly select the grid corresponding to the number of targets as the initial clustering center, and then cluster according to the principle of K-means. (**1b**–**3b**) are the clustering result graphs corresponding to the three assumed target numbers.

Using the GNN-TN method mentioned above, we simulate several hypothetical ConfMaps, as shown in Figure 15.

4.2.2. Target Number Estimation Compared with the Other Clustering Methods

The ConfMaps simulated by GMM-TN model are compared with the ConfMap of RODNet by using the KL divergence. It is worth mentioning that because KL divergence > 0, we need to normalize the ConfMaps. Then we are going to calculate the 2DKL divergence using Equation (7) and the 1DKL divergence using Equation (8). After obtaining the corresponding KL divergence value, we use Equations (9) and (10), respectively, to obtain the final number of targets. Then, we output the corresponding position results and scores.

This algorithm is similar to the other clustering algorithms which can find the best number of clusters automatically, such as Mean-shift [42], DBSCAN [39,43], OPTICS [44,45], and BIRCH [46]. Therefore, it is necessary to compare their performance with the GMM-TN algorithm. We randomly take a group of single and double pedestrian data from CRUW for analysis. For the typical single-pedestrian case, the automatic target number estimation results of K-means-based GMM-TN, Gaussian-mixtures-based GMM-TN, Mean-shift, DBSCAN, OP-TICS, and BIRCH are shown in Figure 16. For the double pedestrian case, their responses in ConfMap are overlapped, and their results are shown in Figure 17.

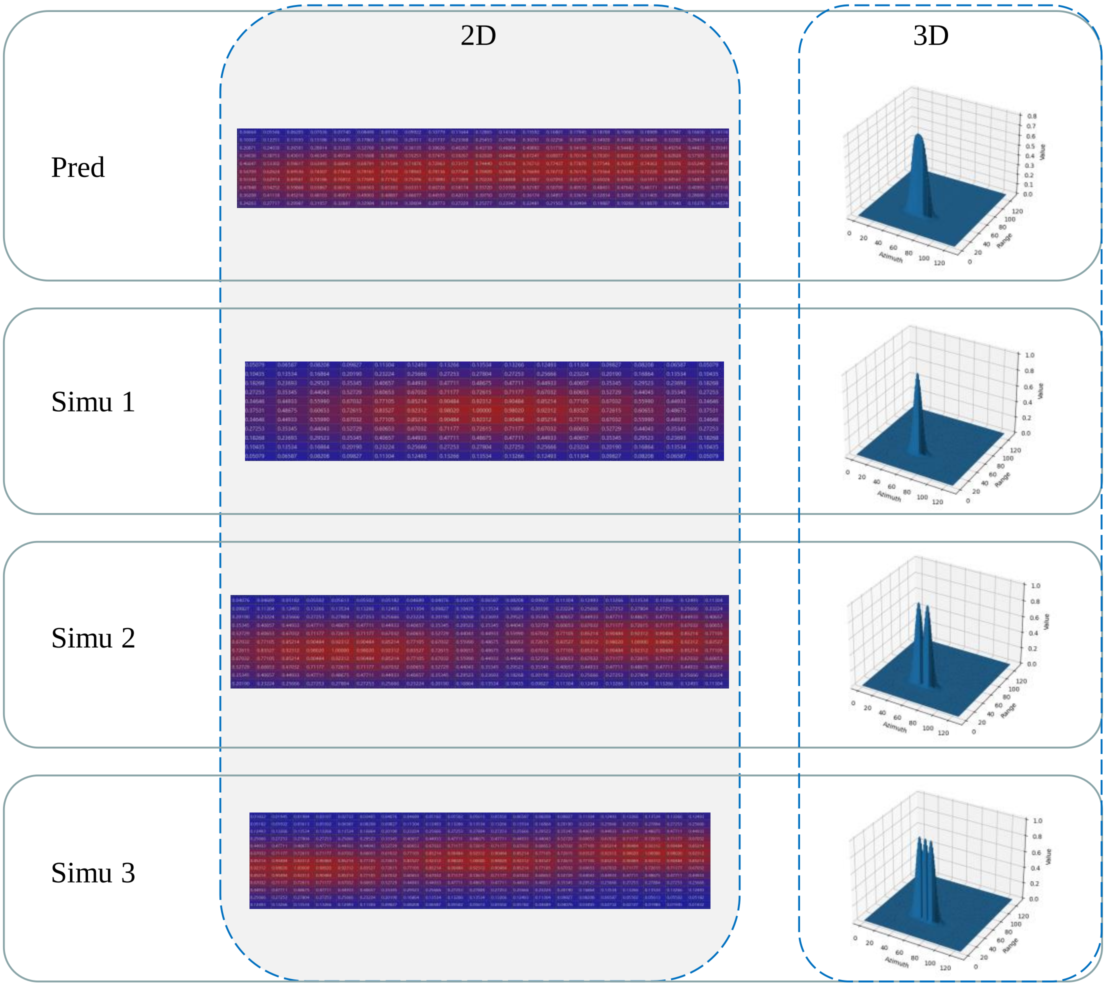


Figure 15. GMM-TN instance implementation diagram. In order to show the spatial distribution more intuitively, we also show the 3D schematic diagram.

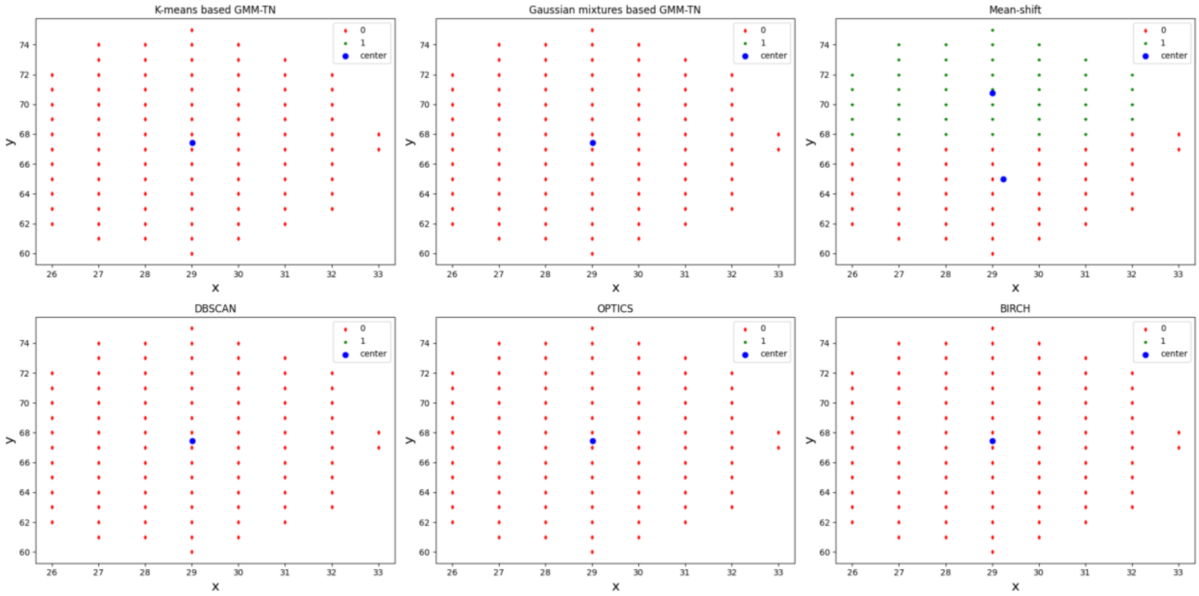


Figure 16. Automatic target number estimation results (single target) using K-means-based GMM-TN, Gaussian-mixtures-based GMM-TN, Mean-shift, DBSCAN, OPTICS, and BIRCH.

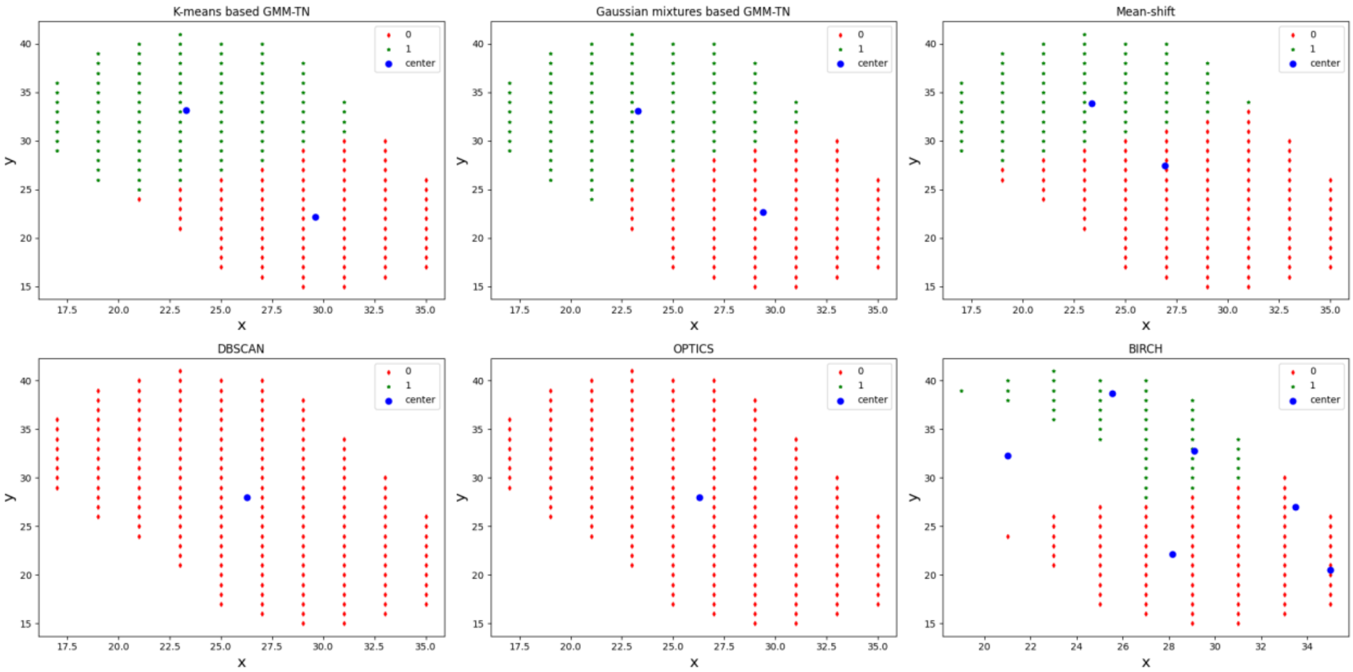


Figure 17. Automatic target number estimation results (double overlapped targets) using K-means-based GMM-TN, Gaussian-mixtures-based GMM-TN, Mean-shift, DBSCAN, OPTICS, and BIRCH.

According to the two cases results, we find the GMM-TN algorithm developed in this paper can provide accurate automatic target numbers, either using the K-means or the Gaussian mixtures clustering. Their centroids are somewhat different. However, the Mean-shift and BIRCH algorithms tend to cluster single target into multiple targets, which will lead to a false detection problem. In contrast, the DBSCAN and OPTICS algorithms tend to cluster double targets into single target, which will lead to a missed detection problem.

From the graph signatures of spatial distribution shape, the single target ConfMap generated by RODNet usually satisfies the elliptic hypothesis. For two adjacent targets, their spatial distribution of ConfMap can be modeled as two ellipses overlapping together. We believe that this is related to the FFTs operation used in radar RA map processing and the CNN-based RODNet. This kind of spatial distribution shape feature facilitates clustering methods such as K-means and Gaussian mixtures, which separate samples in n groups of equal variances, minimizing a criterion known as the inertia or within-cluster sum-of-squares. In addition, the KL distance between the simulated ConfMap and the ConfMap generated by RODNet is sensitive to different target number assumptions. However, other clustering algorithms are not directly related to the number of targets but to cluster density or bandwidth.

4.2.3. Detection Results Comparison for Dense and Non-Dense Pedestrian Scenes

We experimented with our methods in various typical scenarios and compared it with the RODNet method. The results are shown in Figure 18. The experiment selected typical results for the four cases mentioned above.

In order to ensure that the method mentioned in this paper will not have a negative impact on non-dense target situation, we also conducted experiments on NDP scenes, such as Case (1) and Case (2).

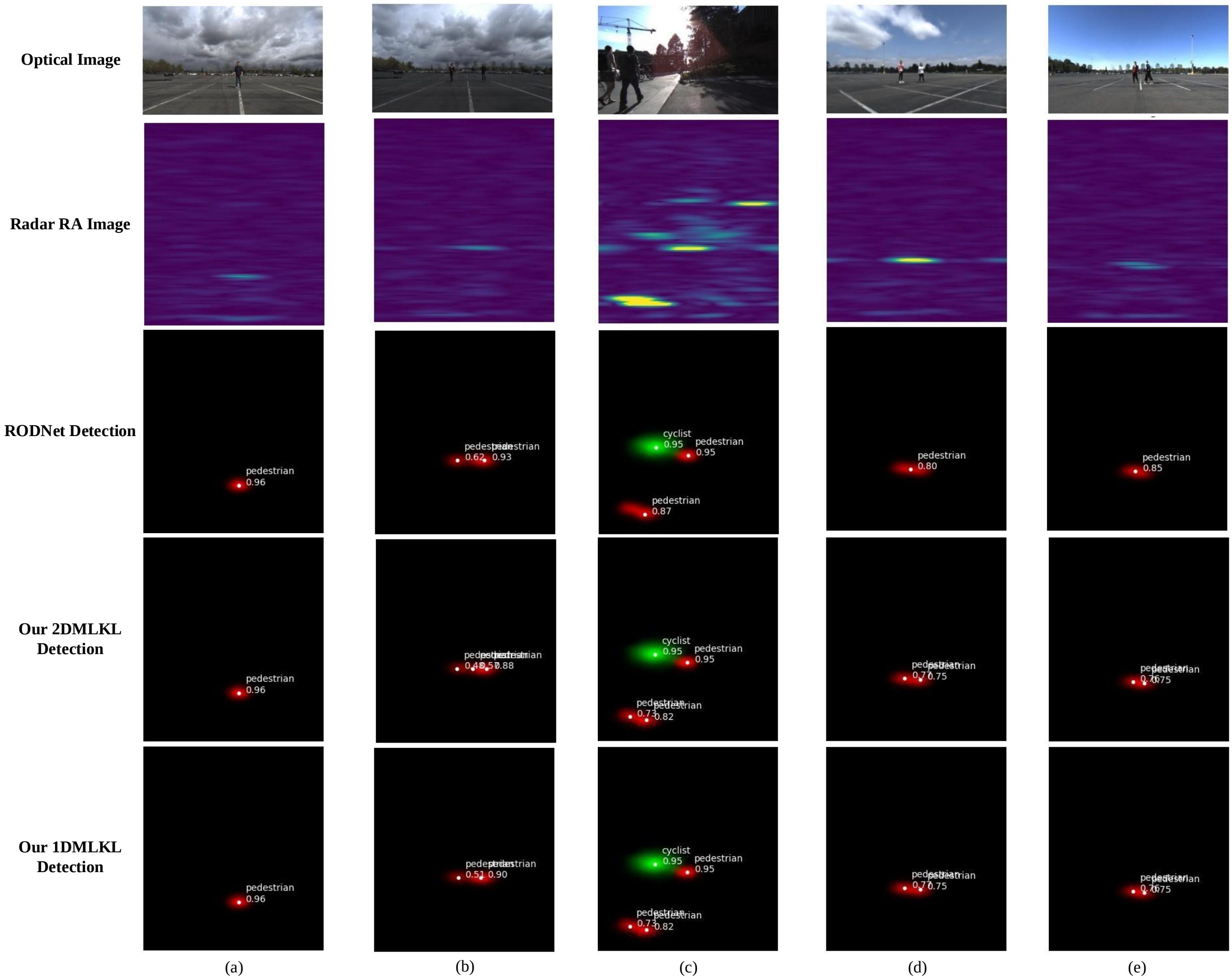


Figure 18. Comparison of experimental results, where (**a**) corresponds to Case (1), (**b**) corresponds to Case (2), (**c**) corresponds to Case (3), and (**d**) and (**e**) correspond to Case (4).

From the diagram, it can be found that for Case (1), both the target-location detection using OLS and using the 2DMLKL (Equation (7)) and 1DMLKL (Equation (8)) methods we proposed perform well. In Case (2), we found that there was a false detection in the detection of 2DMLKL in some frames. The correct number of targets is 2, whereas the number of detected targets is 3. Moreover, the 1DMLKL method accurately detected two targets. For Case (3) and Case (4), both of our methods can detect two dense targets well, whereas the target-location detection using OLS gave an unsatisfactory result.

According to Equations (7) and (8), the larger the assumed target number is, the stronger the corresponding randomness is, which is also confirmed from the results. In addition, because the main influencing factor that makes dense pedestrians unable to be accurately detected is the azimuth axis, the range axis has little effect on the results, so the effect has been improved after using 1DMLKL. In the following paper, we will further use the evaluation indicators to evaluate and analyze the experimental results.

4.3. Statistical Evaluation of Large Amounts of Data

First of all, it is worth mentioning that the author of [25] performed inference and evaluation on the human-annotated data. The quantitative results are shown in Table 1. The RODNet results are compared with the following baselines that also use radar-only inputs: (1) a decision tree uses some handcrafted features from radar data [6]; (2) CFAR detection is first implemented, and a radar object classification network with ResNet backbone [47] is appended; and (3) similar to (2), a radar object classification network with VGG-16 backbone based on CFAR detections is mentioned in [6]. Therefore, we make use of exactly the same network and weights of original RODNet to generate ConfMap.

Table 1. Radar object detection performance evaluated on CRUW dataset.

Methods	AP	AR
Decision Tree [6]	4.70	44.26
CFAR+ResNet [47]	40.49	60.56
CFAR+VGG-16 [6]	40.73	72.88
RODNet [25]	85.98	87.86

Because this article is derived from the optimization of the dense pedestrian situation of RODNet, we use the original evaluation method, that is, OLS as an indicator to calculate the average precision (AP) and average recall (AR). The formula of OLS is Equation (1). AP and AR correspond to Equations (11) and (12). Here, true positive (TP) represents correctly located and classified instances, false positive (FP) represents the false alarm, and false negative (FN) represents the missed detection and/or incorrectly classified instance.

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

We first calculate OLS between each detection result and ground truth in every frame. Then, we use different thresholds from 0.5 to 0.9 with a step of 0.05 for OLS and calculate the AP and AR for different OLS thresholds, which represent different localization error tolerance for the detection results. Here, we use AP and AR to represent the average values among all different OLS thresholds from 0.5 to 0.9 and use $\mathrm{AP}^{\mathrm{OLS}}$ and $\mathrm{AR}^{\mathrm{OLS}}$ to represent the values at a certain OLS threshold. Overall, we use AP and AR as our main evaluation metrics for the radar object detection task, which is the same as the original.

Here are some comparison results in Table 2.

Table 2. Comparison of the results of the method proposed in this paper with those in the original paper.

Scene	Evaluation Index	RODNet	ROD-2DMLKL	ROD-1DMLKL
Dense pedestrian	AP	55.70%	80.62%	84.59%
	AR	52.27%	83.33%	88.28%
Non-dense pedestrian	AP	93.43%	87.98%	92.68%
	AR	94.72%	88.47%	93.32%

Through Table 2, we can find that under the new position processing method proposed in this paper, the detection effect of RODNet in the case of the DP scene has been significantly improved, but for the NDP scene, there are some declines. After the distance direction is compressed into one dimension, the two scenes are further improved, and the scene for NDP is also similar to the original method.

We analyze the reason why our method causes the decrease of AP and AR in NDP scenes. As shown in Figure 19, in the radar RA images, the target position in the label is in the upper right part of the 'ellipse', as shown by the star in the Figure 19, and the target center we obtain by clustering is in the positive center of the 'ellipse'. In the CRUW dataset, the author only introduces the CRF (camera–radar fusion) method, which is annotated by the fusion of optical detection projection and radar peak point, without revealing more details. The label may have an error with the actual position. In addition, OLS is also applied to the evaluation process. Target number estimation and position estimation give the evaluation results at the same time and are not carried out separately. Therefore, we consider adding 'target number accuracy' (for short TNA) as another evaluation.

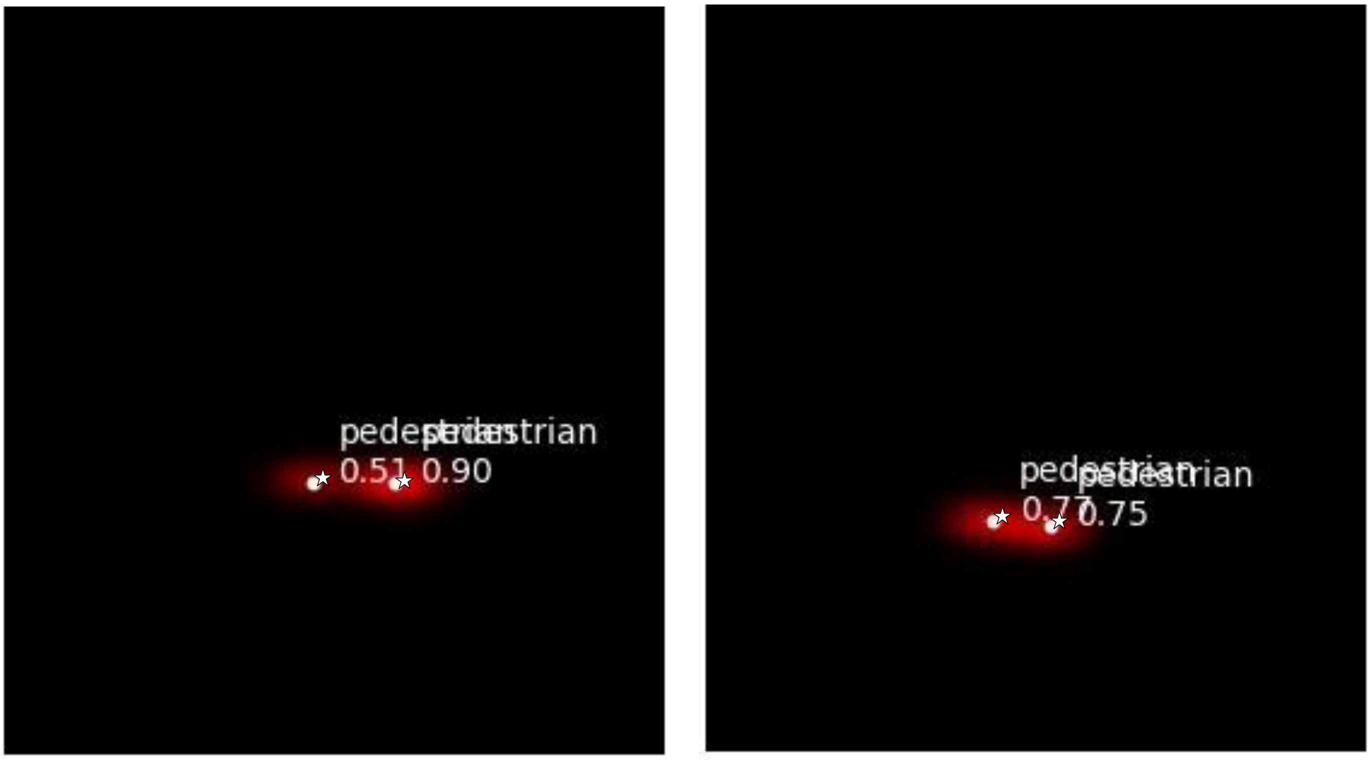


Figure 19. Detection results with labels. Circles represent the test result, and stars represent the true value of the ground.

We count the number of targets in the detection results of the three methods, and the results are shown in Table 3. The statistical method of TNA is as follows: the ratio of the number of frames correctly estimated by the target number to the number of frames in the total dataset, which is shown as Equation (13), where t is the number of true frames and a is the number of all frames.

$$TNA = \frac{t}{a} \tag{13}$$

Table 3. Comparison of the TNA of the method proposed in this paper with those in the original paper.

Scene	Evaluation Index	RODNet	ROD-2DMLKL	ROD-1DMLKL
Dense pedestrian	TNA	52.63%	89.73%	96.21%
Non-dense pedestrian		89.70%	90.62%	95.59%

From Table 3, we can conclude that the method proposed in this paper has significantly improved the estimation of the number of targets.

5. Conclusions

Aiming at the problem of missed detection of RODNet in dense pedestrian scenes, this paper proposes an improved method of a RODNet radar target-detection algorithm applied to dense pedestrian scenes. Firstly, we analyze ConfMap predicted by RODNet and the limitations of the OLS target-location detection method. The relationship between ConfMap value distribution, occupied grid spatial distribution, and target number is analyzed as well. Secondly, we propose a target state likelihood model called GMM-TN, which uses ConfMap value for observation to simulate the conditional ConfMap value and occupied grid spatial distribution with the target number as a condition. Finally, we propose a maximum posteriori target number estimation based on KL divergence to obtain a new number of targets. Then the CRUW dataset is used to verify the improved missed detection and false alarm of the method.

Because we focus on dense pedestrian scenes, we reconstruct the CRUW dataset into Dense Pedestrian (DP) scenes and Non-dense Pedestrian (NDP) scenes and design experiments. We used the evaluation indicators AP and AR in RODNet for evaluation and introduced a new indicator target-number accuracy rate to verify the method. The validity of the three main contributions in this paper is verified by analysis. The missed detection of DP scenes has been significantly improved, and our method also shows good robustness in NDP scenes.

However, the feature extraction part of RODNet has not been improved. The method proposed in this paper proves that ConfMap has the potential to provide more information. In the subsequent research, we consider trying a new architecture and using the subsequent information to optimize the previous feature extraction network. At the same time, this paper verifies that this method can obtain an accurate number of targets, but because the original dataset does not give a calibration method, the position estimation method cannot judge whether it is accurate or not. In the future, we will consider evaluating the performance of position detection results by projecting the detection results into optical images.

Author Contributions: Conceptualization, Y.L. (Yang Li) and Z.L.; methodology, Y.L. (Yang Li) and Z.L.; software, Z.L. and G.X.; validation, Z.L. and G.X.; formal analysis, Y.L. (Yang Li); investigation, Y.L. (Yang Li), Z.L. and G.X.; resources, Z.L. and G.X.; data curation, Z.L. and G.X.; writing—original draft preparation, Y.L. (Yang Li) and Z.L.; writing—review and editing, Y.L. (Yang Li) and Y.W.; visualization, Y.L. (Yang Li) and Z.L.; supervision, Y.L. (Yang Li) and Y.W.; project administration, Y.L. (Yang Li) and Z.L.; funding acquisition, Y.L. (Yang Li), Y.W., Y.L. (Yun Lin), W.S., and W.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China under grant 62131001 and 61971456, by the General Project of Science and Technology Plan of Beijing Municipal Commission of Education under grant KM202010009011 and by Yuyou Talent Training Program under grant 218051360020XN115/014 of the North China University of Technology.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: We used the open source dataset CRUW during our research, which can be found here: https://www.cruwdataset.org/.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

List of Acronyms: We sorted out the acronyms that appear in the article, as shown in the following:

Acronyms	**Full Name**
MMW	Millimeter-wave
RA	Range–azimuth
CNNs	Convolutional Neural Networks
OLS	Object-Location Similarity
GMM-TN	Gaussian Mixture Model with target number
KL divergence	Kullback–Leibler divergence
AP	Average Precision
AR	Average Recall
CFAR Detection	Constant False Alarm Rate Detection
FFT	Fast Fourier Transform
TDC	Temporal deformable convolution
MIMO	Multiple-Input Multiple-Output
DCN	Deformable convolution network
ConfMap	Confidence Map
IoU	Intersection over union
HG	Hourglass
2DMLKL	Two-dimensional symmetric KL divergence
1DMLKL	One-dimensional symmetric KL divergence
CRUW [40]	Camera-Radar of the University of Washington
DP scene	Dense Pedestrian scene
NDP scene	Non-Dense Pedestrian scene
CRF	Camera-radar fusion
TNA	Target Number Accuracy

References

1. de Ponte Müller, F. Survey on ranging sensors and cooperative techniques for relative positioning of vehicles. *Sensors* **2017**, *17*, 271. [CrossRef] [PubMed]
2. Yoneda, K.; Suganuma, N.; Yanase, R.; Aldibaja, M. Automated driving recognition technologies for adverse weather conditions. *Iatss Res.* **2019**, *43*, 253–262. [CrossRef]
3. Schneider, M. Automotive radar-status and trends. In Proceedings of the German Microwave Conference, Ulm, Germany, 5–7 April 2005; pp. 144–147.
4. Nabati, R.; Qi, H. CenterFusion: Center-based Radar and Camera Fusion for 3D Object Detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 5–9 January 2021; pp. 1527–1536.
5. Richards, M.A. *Fundamentals of Radar Signal Processing*; Tata McGraw-Hill Education: New Delhi, India, 2005.
6. Gao, X.; Xing, G.; Roy, S.; Liu, H. Experiments with mmwave automotive radar test-bed. In Proceedings of the 2019 53rd Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 3–6 November 2019.
7. Chen, X.; Ma, H.; Wan, J.; Li, B.; Xia, T. Multi-view 3d object detection network for autonomous driving. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1907–1915.
8. Du, X.; Ang, M.H.; Karaman, S.; Rus, D. A general pipeline for 3d detection of vehicles. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, Australia, 21–25 May 2018; pp. 3194–3200.
9. Liang, M.; Yang, B.; Wang, S.; Urtasun, R. Deep continuous fusion for multi-sensor 3d object detection. In Proceedings of the European Conference on Computer Vision (ECCV), Lecture Notes in Computer Science, Munich, Germany, 8–14 September 2018; Springer: Cham, Switzerland, 2018; Volume 11220, pp. 641–656.
10. Ku, J.; Mozifian, M.; Lee, J.; Harakeh, A.; Waslander, S.L. Joint 3d proposal generation and object detection from view aggregation. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 1–8.
11. Qi, C.; Liu, W.; Wu, C.; Su, H.; Guibas, L. Frustum PointNets for 3D Object Detection from RGB-D Data. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 918–927.
12. Fukushima, K. A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* **1980**, *36*, 193–202. [CrossRef] [PubMed]
13. Waibel, A.; Hanazawa, T.; Hinton, G.; Shikano, K.; Lang, K.J. Phoneme recognition using time-delay neural networks. *IEEE Trans. Acoust. Speech Signal Process.* **1989**, *37*, 328–339. [CrossRef]
14. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]

15. Danzer, A.; Griebel, T.; Bach, M.; Dietmayer, K. 2d car detection in radar data with pointnets. In Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, 27–30 October 2019; pp. 61–66.
16. Wang, L.; Tang, J.; Liao, Q. A study on radar target detection based on deep neural networks. *IEEE Sens. Lett.* **2019**, *3*, 1–4. [CrossRef]
17. Nabati, R.; Qi, H. Rrpn: Radar region proposal network for object detection in autonomous vehicles. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019; pp. 3093–3097.
18. John, V.; Nithilan, M.; Mita, S.; Tehrani, H.; Sudheesh, R.; Lalu, P. So-net: Joint semantic segmentation and obstacle detection using deep fusion of monocular camera and radar. In Proceedings of the Pacific-Rim Symposium on Image and Video Technology, Image and Video Technology, PSIVT 2019, Sydney, Australia, 18–22 September 2019; Springer: Cham, Switzerland, 2020; Volume 11994, pp. 138–148.
19. Yu, J.; Hao, X.; Gao, X.; Sun, Q.; Liu, Y.; Chang, P.; Zhang, Z.; Gao, F.; Shuang, F. Radar Object Detection Using Data Merging, Enhancement and Fusion. In Proceedings of the 2021 International Conference on Multimedia Retrieval (ICMR '21), Taipei, Taiwan, 21 August 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 566–572.
20. Sun, P.; Niu, X.; Sun, P.; Xu, K. Squeeze-and-Excitation network-Based Radar Object Detection With Weighted Location Fusion. In Proceedings of the 2021 International Conference on Multimedia Retrieval (ICMR '21), Taipei, Taiwan, 21 August 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 545–552.
21. Zheng, Z.; Yue, X.; Keutzer, K.; Vincentelli, A.S. Scene-aware Learning Network for Radar Object Detection. In Proceedings of the 2021 International Conference on Multimedia Retrieval (ICMR '21), Taipei, Taiwan, 21 August 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 573–579.
22. Ju, B.; Yang, W.; Jia, J.; Ye, X.; Chen, Q.; Tan, X.; Sun, H.; Shi, Y.; Ding, E. DANet: Dimension Apart Network for Radar Object Detection. In Proceedings of the 2021 International Conference on Multimedia Retrieval (ICMR '21), Taipei, Taiwan, 21 August 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 533–539.
23. Hsu, C.-C.; Lee, C.; Chen, L.; Hung, M.-K.; Lin, Y.-L.; Wang, X.-Y. Efficient-ROD: Efficient Radar Object Detection based on Densely Connected Residual Network. In Proceedings of the 2021 International Conference on Multimedia Retrieval (ICMR '21), Taipei, Taiwan, 21 August 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 526–532.
24. Gao, X.; Xing, G.; Roy, S.; Liu, H. RAMP-CNN: A Novel Neural Network for Enhanced Automotive Radar Object Recognition. *IEEE Sens. J.* **2021**, *21*, 5119–5132. [CrossRef]
25. Wang, Y.; Jiang, Z.; Gao, X.; Hwang, J.-N.; Xing, G.; Liu, H. RODNet: Object Detection under Severe Conditions Using Vision-Radio Cross-Modal Supervision. *arXiv* **2020**, arXiv:2003.01816.
26. Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 764–773.
27. Guo, W.; Wang, J.; Wang, S. Deep multimodal representation learning: A survey. *IEEE Access* **2019**, *7*, 63373–63394. [CrossRef]
28. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [CrossRef] [PubMed]
29. Urmson, C.; Anhalt, J.; Bagnell, D.; Baker, C.; Bittner, R.; Clark, M.; Dolan, J.; Duggins, D.; Galatali, T.; Geyer, C. Autonomous driving in urban environments: Boss and the urban challenge. *J. Field Robot.* **2008**, *25*, 425–466. [CrossRef]
30. Van Brummelen, J.; O'Brien, M.; Gruyer, D.; Najjaran, H. Autonomous vehicle perception: The technology of today and tomorrow. *Transp. Res. Part C Emerg. Technol.* **2018**, *89*, 384–406. [CrossRef]
31. Giacalone, J.-P.; Bourgeois, L.; Ancora, A. Challenges in aggregation of heterogeneous sensors for Autonomous Driving Systems. In Proceedings of the 2019 IEEE Sensors Applications Symposium (SAS), Sophia Antipolis, France, 11–13 March 2019; pp. 1–5.
32. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *52*, 436–444. [CrossRef] [PubMed]
33. Rashinkar, P.; Krushnasamy, V. An overview of data fusion techniques. In Proceedings of the 2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, India, 21–23 February 2017; pp. 694–697.
34. Fung, M.L.; Chen, M.Z.; Chen, Y.H. Sensor fusion: A review of methods and applications. In Proceedings of the 2017 29th Chinese Control and Decision Conference (CCDC), Chongqing, China, 28–30 May 2017; pp. 3853–3860.
35. Fayyad, J.; Jaradat, M.A.; Gruyer, D.; Najjaran, H. Deep learning sensor fusion for autonomous vehicle perception and localization: A review. *Sensors* **2020**, *20*, 4220. [CrossRef] [PubMed]
36. Bombini, L.; Cerri, P.; Medici, P.; Alessandretti, G. Radar-vision fusion for vehicle detection. In Proceedings of the International Workshop on Intelligent Transportation, Hamburg, Germany, 14–15 May 2006; pp. 65–70.
37. Newell, A.; Yang, K.; Deng, J. Stacked hourglass networks for human pose estimation. In Proceedings of the European Conference on Computer Vision, Venice, Italy, 22–29 October 2017; Springer: Berlin/Heidelberg, Germany, 2016; pp. 483–499.
38. MacQueen, J. Some methods for classification and analysis of multivariate observation. In Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Los Angelas, CA, USA, 18–21 July 1965; University of California: Los Angeles, CA, USA, 1967; pp. 281–297.
39. Ester, M.; Kriegel, H.P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, Oregon, 2–4 August 1996; pp. 226–231.

40. Wang, Y.; Wang, G.; Hsu, H.M.; Liu, H.; Hwang, J.N. Rethinking of Radar's Role: A Camera-Radar Dataset and Systematic Annotator via Coordinate Alignment. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021.
41. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L. Pytorch: An imperative style, high-performance deep learning library. *arXiv* **2019**, arXiv:1912.01703.
42. Comaniciu, D.; Meer, P. Mean Shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 603–619. [CrossRef]
43. Schubert, E.; Sander, J.; Ester, M.; Kriegel, H.P.; Xu, X. DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. *ACM Trans. Database Syst.* **2017**, *42*, 19. [CrossRef]
44. Ankerst, M.; Breunig, M.M.; Kriegel, H.P.; Sander, J. OPTICS: Ordering points to identify the clustering structure. *ACM Sigmod Rec.* **1999**, *28*, 49–60. [CrossRef]
45. Schubert, E.; Gertz, M. Improving the Cluster Structure Extracted from OPTICS Plots. In Proceedings of the LWDA 2018: Conference "Lernen, Wissen, Daten, Analysen", Mannheim, Germany, 22–24 August 2018; pp. 318–329.
46. Tian, Z.; Ramakrishnan, R. Miron Livny: BIRCH: An Efficient Data Clustering Method for Very Large Databases. *ACM Sigmod Rec.* **1999**, *25*, 103–114.
47. Angelov, A.; Robertson, A.; Murray-Smith, R.; Fioranelli, F. Practical classification of different moving targets using automotive radar and deep neural networks. *IET Radar Sonar Navig.* **2018**, *12*, 1082–1089. [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.