

Lira-YOLO: a lightweight model for ship detection in radar images

ZHOU Long^{1,2,*}, WEI Suyuan², CUI Zhongma¹, FANG Jiaqi¹,
YANG Xiaoting¹, and DING Wei^{1,3}

1. Beijing Institute of Remote Sensing Equipment, Beijing 100854, China;

2. Xi'an High-tech Research Institute, Xi'an 710025, China;

3. School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

Abstract: For the detection of marine ship objects in radar images, large-scale networks based on deep learning are difficult to be deployed on existing radar-equipped devices. This paper proposes a lightweight convolutional neural network, LiraNet, which combines the idea of dense connections, residual connections and group convolution, including stem blocks and extractor modules. The designed stem block uses a series of small convolutions to extract the input image features, and the extractor network adopts the designed two-way dense connection module, which further reduces the network operation complexity. Mounting LiraNet on the object detection framework Darknet, this paper proposes Lira-you only look once (Lira-YOLO), a lightweight model for ship detection in radar images, which can easily be deployed on the mobile devices. Lira-YOLO's prediction module uses a two-layer YOLO prediction layer and adds a residual module for better feature delivery. At the same time, in order to fully verify the performance of the model, mini-RD, a lightweight distance Doppler domain radar images dataset, is constructed. Experiments show that the network complexity of Lira-YOLO is low, being only 2.980 Bflops, and the parameter quantity is smaller, which is only 4.3 MB. The mean average precision (mAP) indicators on the mini-RD and SAR ship detection dataset (SSDD) reach 83.21% and 85.46%, respectively, which is comparable to the tiny-YOLOv3. Lira-YOLO has achieved a good detection accuracy with less memory and computational cost.

Keywords: lightweight, radar images, ship detection, you only look once (YOLO).

DOI: 10.23919/JSEE.2020.000063

1. Introduction

In the field of military reconnaissance and object strike, the detection of marine ships in radar images has been exten-

sively studied. Traditional detection algorithms, for example, the constant false-alarm rate (CFAR) [1–5], always have the problem of low detection accuracy. In recent years, due to the significant advantages of deep learning in the field of object detection [6–12], the application of deep learning to the detection of marine objects in radar images has become a research hotspot. For example, Li et al. [13] applied fast region convolutional neural network (R-CNN) [14] to ship detection in synthetic aperture radar (SAR) image and proposed a new data set named SAR ship detection dataset (SSDD). In [15], Qu et al. used the improved OTSU method to segment the SAR image and selected the target through the trained convolutional neural network (CNN). The algorithm improves the detection speed while reducing false positives.

The above algorithms have outstanding performance in the detection accuracy, much higher than that of traditional methods. However, aircraft equipment with radar sensors has limited storage space and computing power, so it is difficult for large object detection models to deploy then. Designing a lightweight network detection model is an effective way to solve this problem [16–19]. The lightweight network detection model aims to achieve a better detection accuracy through efficient network design, using less memory and computing power. Researchers have proposed some models for the lightweight network [20–22]. In the field of object detection, a common method is to load some lightweight CNN into object detection models. At present, it is mainly mounted in single shot multi-box detector (SSD) [23] to form a lightweight target detection network. Typical networks, such as MobileNet [24], use deep separable convolution to design the network. The ShuffleNet [25] reduces the amount of parameters through applying group convolution and solves the problem of poor information between groups through the channel shuffle operation.

Manuscript received May 13, 2019.

*Corresponding author.

This work was supported by the Joint Fund of Equipment Pre-Research and Aerospace Science and Industry (6141B07090102).

Based on the Pelee [26] model, this paper designs a new CNN LiraNet and puts the network on the you only look once (YOLO) [27–29] model. Then, Lira-YOLO, a lightweight model for ship detection in radar images is proposed. At the same time, in order to solve the redundancy problem of the radar images data set used in [30], this paper reconstructs the previous data set, then proposes a new small dataset named mini-RD. Finally, the good performance of Lira-YOLO is verified by experiments on the mini-RD and SSDD datasets.

2. Related work

2.1 Why YOLO, not SSD?

SSD is an end-to-end real-time object detection model proposed by Liu in 2016 [23]. It has a good network portability and is one of the most widely used models. The existing lightweight object detection networks basically use SSD models, such as deeply supervised object detector (DSOD) [31] and Pelee. The original YOLO is based on the Darknet framework and is developed by the author alone. Its model mobility is not as good as SSD, and therefore it is less used. We proposed the YOLOv3 network in 2018. YOLOv3 uses a three-layer predictive structure and incorporates multiple layers of feature information. In comparison, although the SSD has feature fusion on the five convolution kernel scales, there is only one prediction layer, and the use of features does not necessarily have excellent YOLOv3 performance. Based on the above analysis, this paper attempts to use the original YOLO model based on the Darknet framework as the underlying module to build LiraNet. Through subsequent experiments, the good generalization performance of the YOLO model is verified. You can access the relevant code in this paper from <https://github.com/longlongZ/Lira-YOLO>.

2.2 Pelee

Pelee was published on International Conference on Learning Representations 2018 and a variant version of DenseNet named PeleeNet was proposed, which is a lightweight network for mobile devices [26]. It is deployed in the SSD model for object detection. PeleeNet follows the DenseNet’s innovative connectivity model and some key design principles, applying group convolution, and a two-way dense layer for different receptive fields was proposed. One of the dense layers uses a 3×3 smaller convolution kernel, which is a good way to capture small-scale objects. The other uses two 3×3 convolution kernels to learn the visual characteristics of large-scale objects. The stem block designed can effectively extract features with minimal computational resources. When PeleeNet is loaded

into the SSD, the 38×38 feature map is discarded to reduce the computational cost. At the same time, before each feature map is added to the prediction layer, a ResBlock is passed for better features transfer. Finally, on the visual object classes (VOC) dataset, Pelee surpasses the object detection performance of YOLOv2 [28] at a lower cost.

2.3 Dataset preparation

The mini-RD is a small dataset of radar ship objects in the range-Doppler (R-D) domain constructed in this paper. The dataset covers radar images of different sizes such as 320×32 , 320×64 , 320×128 , 800×64 , and $1\ 200 \times 128$. The final mini-RD contains more than 1 500 pictures, including ships and two types of jammers, and the ratio of the training set to the test set is about 8:2.

The SSDD dataset is the ship object dataset of the SAR images proposed in [13], which contains 1 260 pictures, including only a class of ships, and the sizes of the radar images are between 300×200 and 550×450 . This paper divides the training set and the test set roughly by 8:2.

In the process of making datasets, in order to increase the difficulty of training, this paper manually adds a lot of hard examples in the test set, including background noise, occlusion between objects, and a large number of small objects, as shown in Table 1.

Table 1 Datasets used in this paper

Dataset	Class	Training set	Test set
Mini-RD	3	1 280	267
SSDD	1	878	282

3. Lira-YOLO

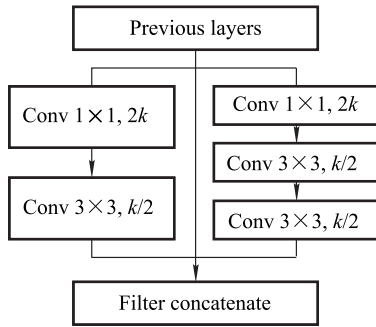
Based on PeleeNet, this paper proposes a new lightweight CNN, LiraNet, then mounts it on YOLO, and proposes an object detection model for radar images detection, Lira-YOLO.

3.1 LiraNet

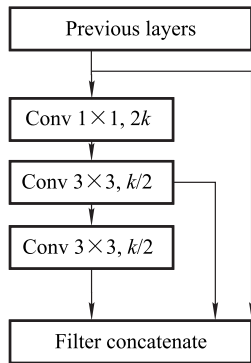
(i) A new dense layer

In the original two-way dense layer of the PeleeNet, the 1×1 and 3×3 convolutions of the previous layers are repeatedly used for group convolution. We try to reuse these two levels of convolution operations. This is because the multiplexing of convolution kernels does not change the perception of densely connected layers on different scales of targets. After multiplexing, the dense layers are still two branches, corresponding to large-scale and small-scale targets. At the same time, LiraNet uses a total of 21 dense layers. The multiplexing of 1×1 and 3×3 convolutional layers can effectively reduce the total computational overhead of the network. Through the experimental verification in the SSDD and mini-RD datasets, after sharing, the network reduces the complexity of 1.3 Bflops, and has little

effect on the accuracy. Fig. 1 shows the dense layers of PeleeNet and the proposed one.



(a) Two-way dense layer of PeleeNet [26]



(b) The proposed dense layer

Fig. 1 Dense layer

(ii) A new stem block

Motivated by tiny-DSOD [6], we design an efficient stem block. The structure is shown in Fig. 2. The stem block first extracts features from the input images by a convolution kernel of the size 3×3 , then concatenates the 1×1 and 3×3 convolution operations, applies a smaller convolution kernel depth, and finally is downsampled with a 2×2 maximum pooling operation.

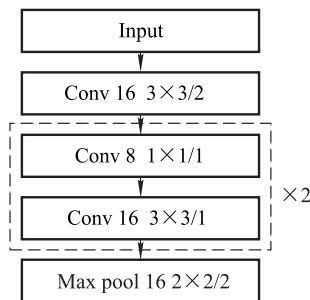


Fig. 2 The proposed stem block

Compared with the stem block in PeleeNet, the new stem block designed is more streamlined, and the channels number of the convolution kernel is smaller, which is more

in line with the lightweight design of the network. The advantage of this stem block is that it not only extracts the features of the input images well, but also consumes less computational complexity.

(iii) Other

The design of the transition layer and the selection of the number of the dense block layers follow the practice of PeleeNet. The transition block is shown in Fig. 3.

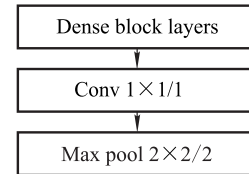


Fig. 3 Transition block

Table 2 shows the network structure of LiraNet.

Table 2 Network structure of LiraNet

Stage	Layer	Output size
Input size		$416 \times 416 \times 3$
Stem block		$104 \times 104 \times 16$
Extractor	Dense layer $\times 3+$ Transition layer	$52 \times 52 \times 128$
	Dense layer $\times 4+$ Transition layer	$26 \times 26 \times 128$
	Dense layer $\times 8+$ Transition layer	$13 \times 13 \times 128$
	Dense layer $\times 6+$ Conv $1 \times 1/1$	$13 \times 13 \times 128$

3.2 Lira-YOLO

To meet the needs of the object detection task, we equip LiraNet to the YOLO model. Through experiments, the two-layer YOLO prediction layer can better utilize features under a small amount of computational power, but the 3×3 convolution operation before the second prediction layer of Tiny-YOLOv3 consumes a large amount of computation and memory [32]. At the same time, this operation has a great impact on the accuracy of the SSD dataset. We refer to the ResBlock operation before accessing the SSD prediction layer in Pelee [33], but we finally use a convolution kernel with a depth of 128 to better apply to object detection tasks with a small number of classes. The Res-Block module is shown in Fig. 4. H represents the height of the convolution kernel, W represents the width of the convolution kernel, D represents the depth of the convolution kernel, and F represents the depth after processing.

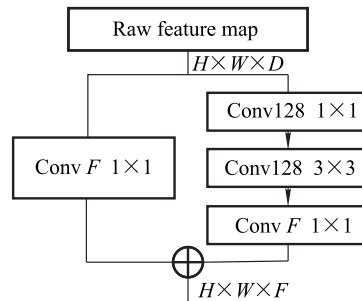


Fig. 4 ResBlock module ($F=(\text{classes}+5) \times 3$)

Finally, we provide two different prediction modules: the v1 module uses the original YOLO prediction layer, whose precision is higher on the SSDD dataset; the v2 module adds the ResBlock module before prediction, and

the advantage of v2 is that the calculation amount and the occupied memory are smaller. In summary, the model structure of Lira-YOLO is shown in Fig. 5.

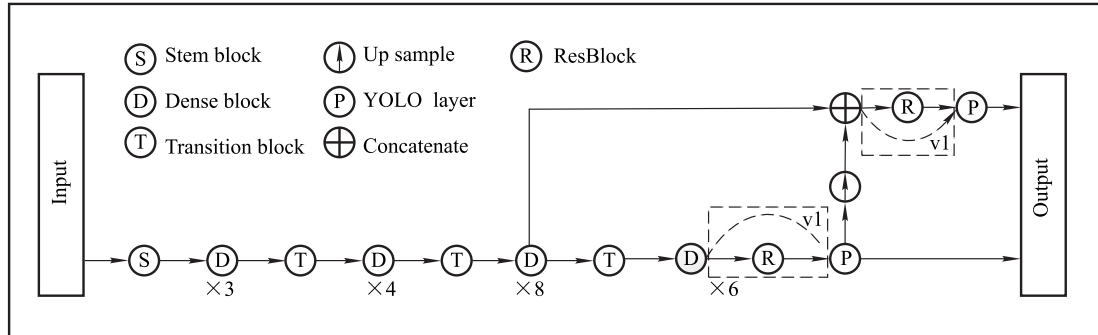


Fig. 5 Lira-YOLO

4. Experiment

4.1 Experimental environment

The experimental environment configuration in this paper is shown in Table 3.

Table 3 Experimental environment configuration

System	Ubuntu18.04
CPU	Intel Core i5-8400 CPU @2.80 GHz×6
GPU	GeForce GTX 1070
Hardware acceleration	CUDA9.0; CUDNN7.5

4.2 Contrast experiment with other models

For an objective assessment, we compile some excellent lightweight networks based on the Darknet framework.

We experiment with stem block modules of several networks on DSOD backbone networks, which are based on the SSDD dataset firstly. The specific experimental results are shown in Table 4, where mAP refers to mean average precision.

Table 4 Stem block experiments of group A based DSOD

Parameter	DSOD	#A1	#A2	#A3	The proposed model	PeleeNet
Weight/MB	13.6	13.6	13.6	14.4	13.5	10.1
Computation/Bflops	13.716	14.181	13.282	17.973	13.028	6.105
mAP/%	88.16	87.86	87.70	88.05	87.93	86.93

Experiments of group A validate the performance of the proposed stem block. #A1–#A3 use the backbone network of DSOD, and the stem block uses the corresponding parts of tiny-DSOD, PeleeNet and Inception-v4 respectively. The designed stem block consumes less computation. However, the backbone network of DSOD consumes more computing cost and memory space than PeleeNet.

Later, we conduct experiments based on PeleeNet and get better results. The results are shown in Table 5.

Table 5 Experiments of group B based PeleeNet

Parameter	PeleeNet	#B1	#B2	Lira-YOLO.v1
Weight/MB	10.1	7.6	7.6	7.6
Computation/Bflops	6.105	4.287	4.720	4.232
mAP/%	86.93	86.39	86.30	88.04

As can be seen from Table 5, #B1 uses the dense layer designed on the basis of PeleeNet, which has obvious advantages in terms of calculation amount and memory space. Compared with the original PeleeNet, it reduces the operation amount of 1.818 Bflops and the memory usage of 2.5 MB.

Based on the designed dense layer, #B1, #B2 and Lira-YOLO.v1 use the stem blocks designed by PeleeNet, DSOD and this paper respectively. The final experiments show that Lira-YOLO.v1 has good performance in various indicators.

All the above experiments use the prediction layer of the v1 version. Then we experiment with the v1 and v2 versions on the mini-RD and SSDD datasets. What is encouraging is that on the mini-RD dataset, the version with the added ResBlock operation has better performance than the original YOLO prediction layer. We add the spatial pyramid pooling (SPP) [34] module to the YOLO prediction layer, but the experimental results are not ideal. The comparison results of the model are shown in Table 6.

In the end, compared to YOLOv3 and tiny-YOLOv3, Lira-YOLO achieves a good detection accuracy with less memory storage space and network complexity. YOLOv3 is not as accurate as tiny-YOLOv3 on the mini-RD dataset. We think that in a small dataset, an overly complex net-

work may overfit the training set, so the generalization performance on the test set is not good enough.

Table 6 Performance comparison of models

Parameter	YOLOv3	Tiny-YOLOv3	Lira-YOLO.v1	Lira-YOLO.v2
Weight/MB	246.3	34.7	7.6	4.3
Computation/Bflops	65.290	5.448	4.232	2.980
mAP/%	SSDD	89.72	88.69	86.30
	Mini-RD	83.96	84.14	82.06

Fig. 6 and Fig. 7 show the loss curve of Lira-YOLO on the SSDD and mini-RD datasets, respectively.

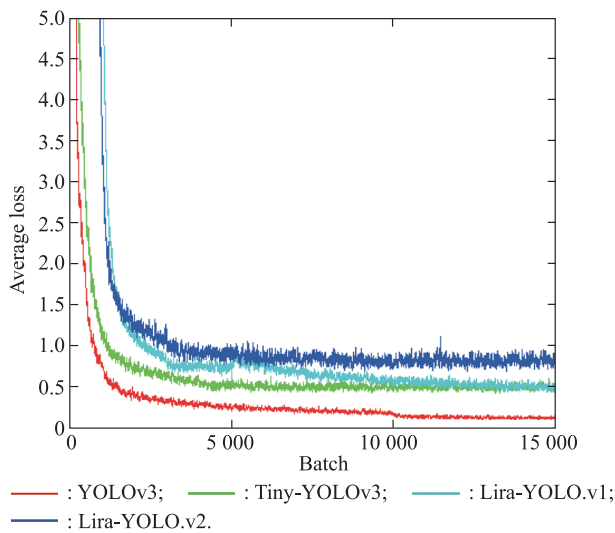


Fig. 6 Loss of SSDD

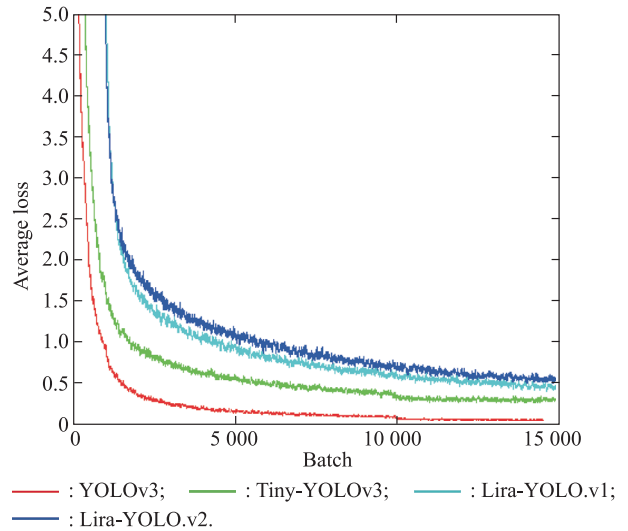


Fig. 7 Loss of mini-RD

We can see that Lira-YOLO has a good convergence and a strong feature learning ability.

4.3 Results and analysis

Fig. 8 and Fig. 9 show the results of Lira-YOLO on the SSDD and Mini-RD datasets. The results are satisfactory. Lira-YOLO has a good detection ability of small objects, and good detection results for radar images of different resolutions.

The Lira-YOLO model has a good performance. Through analysis, we think the main reasons are as follows.

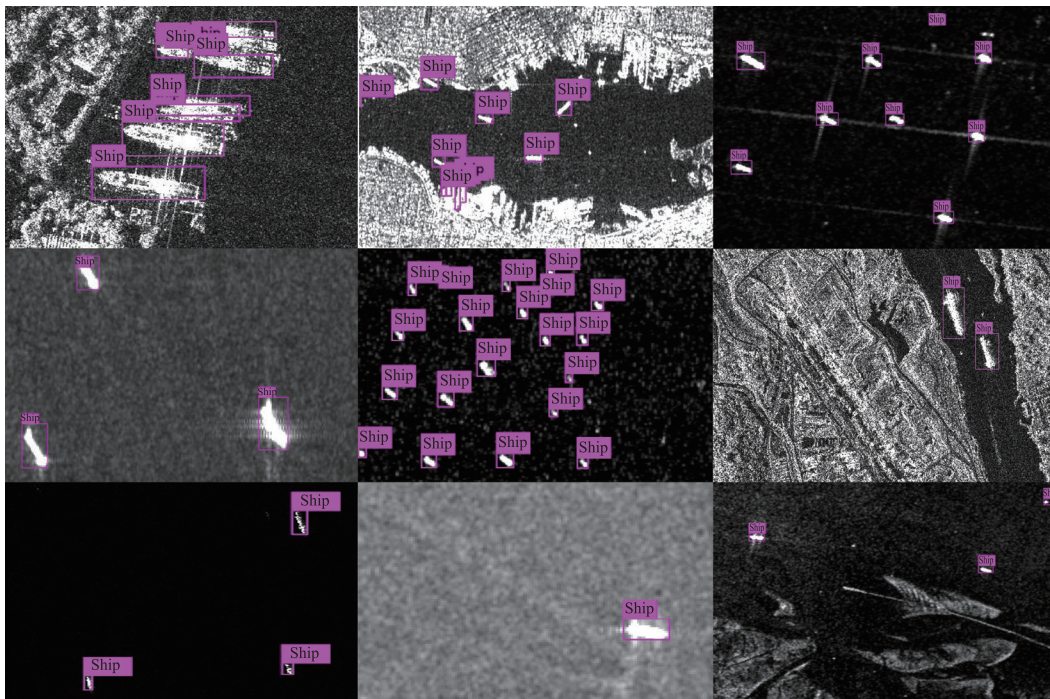


Fig. 8 Test results of SSDD

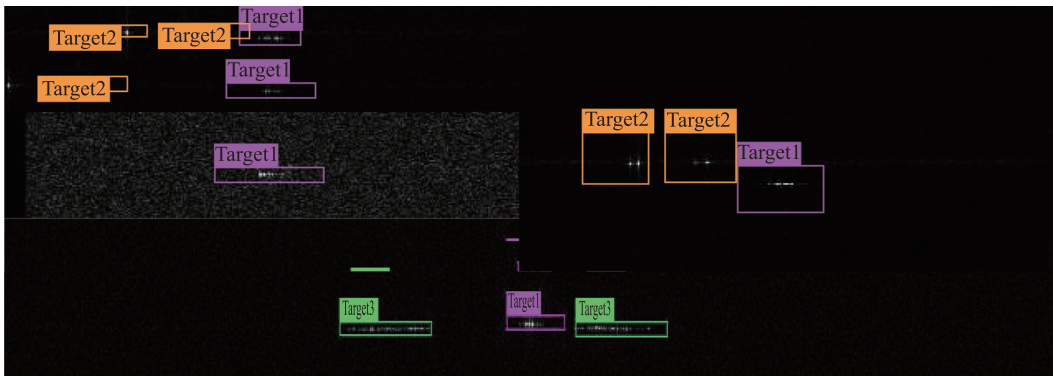


Fig. 9 Test results of mini-RD

(i) The design of LiraNet is simple and efficient. By applying group convolution, dense connection, bottleneck design, etc., LiraNet not only extracts features efficiently, but also reduces the amount of parameters and calculations.

(ii) The improved feature fusion layer of this paper takes into account the lightweight design and feature expression capabilities of the model [33]. Firstly, we add the residual module to better relate the context features, then do multi-scale feature fusion, and finally use two sub-layers of the YOLO prediction layer. These designs and improvements have significantly improved the performance of the model.

(iii) The Darknet frame provides good support for model performance. The contributions of the Darknet include the calculation of the anchor point coordinates suitable for the dataset by the K-means algorithm, the loss function providing good guidance for model training, and the good mobility of the Darknet frame and its applicability to different scale target detection.

5. Conclusions

This paper proposes a lightweight model for ship detection in radar images named Lira-YOLO, which uses the idea of dense connection and group convolution. We design a CNN named LiraNet, which has a low complexity, few parameters and a strong feature expression ability. The network is loaded into the YOLO model and good experimental results are obtained. The model can effectively detect ship targets in different resolution radar images. The generalization performance of the YOLO model is verified.

References

- [1] NOVAK L, OWIRKA G J, BROWER W S, et al. The automatic target-recognition system in SAIP. *Lincoln Laboratory Journal*, 1997, 10(2): 187–202.
- [2] VACHON P W, CAMPBELL J W M, BJERKELUND C A, et al. Ship detection by the RADARSAT SAR validation of detection model predictions. *Canadian Journal of Remote Sensing*, 1997, 23(1): 48–59.
- [3] JIANG Q, AITNOURI E, WANG S, et al. Automatic detection for ship target in SAR imagery using PNN-model. *Canadian Journal of Remote Sensing*, 2000, 26(4): 297–305.
- [4] KRONAUGE M, ROHLING H. Fast two-dimensional CFAR procedure. *IEEE Trans. on Aerospace and Electronic Systems*, 2013, 49(3): 1817–1820.
- [5] HOU B, CHENG X Z, JIAO L C. Multilayer CFAR detection of ship resolution targets in very high resolution SAR images. *IEEE Geoscience and Remote Sensing Letters*, 2015, 12(4): 811–815.
- [6] NGAN I T H, ZHENG Y, ZHU C, et al. Multiple scale faster-RCNN approach to driver's cell-phone usage and hands on steering wheel detection. *Proc. of the Computer Vision & Pattern Recognition Workshops*, 2016: 46–53.
- [7] LI Y X, LI J W, LIN W Y, et al. Tiny-DSOD: lightweight object detection for resource-restricted usages. <http://arXiv.org/abs/1807.11013>.
- [8] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks. *Proc. of the European Conference on Computer Vision*, 2014: 818–833.
- [9] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition. <http://arXiv.org/abs/1409.1556>.
- [10] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904–1916.
- [11] FU C, LIU W, RANGA A, et al. DSSD: deconvolutional single shot detector. <http://arXiv.org/abs/1701.06659>.
- [12] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection. <http://arXiv.org/abs/1708.02002>.
- [13] LI J W, QU C W, PENG S J, et al. Ship detection in SAR images based on convolutional neural network. *Systems Engineering and Electronics*, 2018, 40(9): 1953–1959. (in Chinese)
- [14] GIRSHICK R. Fast R-CNN. *Proc. of the IEEE International Conference on Computer Vision*, 2015: 1440–1448.
- [15] QU C W, LIU C, ZHOU Q, et al. Ship targets detection method based on convolutional neural network for SAR image. *Fire Control & Command Control*, 2019, 44(1): 40–44.
- [16] HAN S, MAO H Z, DALLY W J. Deep compression: compressing deep neural networks with pruning, trained quantization and huffman coding. <http://arXiv.org/abs/1510.00149>.
- [17] YU X Y, LIU T L, WANG X C, et al. On compressing deep

models by low rank and sparse decomposition. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7370–7379.

- [18] LI Y X, LI J W, LIN W Y, et al. Tiny-DSOD: lightweight object detection for resource-restricted usages. <http://arXiv.org/abs/11013>.
- [19] WONG A, SHAFIEE M J, LI F, et al. Tiny SSD: a tiny single-shot detection deep convolutional neural network for real-time embedded object detection. Proc. of the 15th Conference on Computer and Robot Vision, 2018: 95–101.
- [20] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7794–7803.
- [21] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation. <http://arXiv.org/abs/1902.09212>.
- [22] CAO Y, XU J R, LIN S, et al. GCNet: non-local networks meet squeeze-excitation networks and beyond. <http://arXiv.org/abs/1904.11492>.
- [23] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector. <http://arXiv.org/abs/1512.02325>.
- [24] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications. <http://arXiv.org/abs/1704.04861>.
- [25] ZHANG X, ZHOU X, LIN M, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices. <http://arXiv.org/abs/1707.01083>.
- [26] WANG R J, LI X, LING C X, et al. Pelee: a real-time object detection system on mobile devices. Proc. of the 32nd International Conference on Neural Information Processing Systems, 2018: 1963–1992.
- [27] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779–788.
- [28] JOSEPH R, ALI F. YOLO9000: better, faster, stronger. <http://arXiv.org/abs/1612.08242>.
- [29] REDMON J, FARHADI A. YOLOv3: an incremental improvement. <http://arXiv.org/abs/1804.02767>.
- [30] ZHOU L, WEI S Y, CUI Z M, et al. Multi-objective detection of complex background radar image based on deep learning. Systems Engineering and Electronics, 2019, 41(6): 1258–1264. (in Chinese)
- [31] SHEN Z Q, LIU Z, LI J G, et al. DSOD: learning deeply supervised object detectors from scratch. Proc. of the IEEE International Conference on Computer Vision, 2017: 1937–1945.
- [32] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection. <http://arXiv.org/abs/1612.03144>.
- [33] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770–778.
- [34] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904–1916.

Biographies



ZHOU Long was born in 1995. He received his B.S. degrees from Xi'an University of Posts & Telecommunications in 2017. He is currently an M.S. degree candidate in Xi'an High-tech Research Institute. His current major research interests are lightweight convolutional neural network and object detection of radar images.
E-mail: 1792563879@qq.com



WEI Suyuan was born in 1971. She received her Ph.D. degree from Xidian University in 2013. Her major research interests are pattern recognition and computer network & communication.
E-mail: weisuyuan2002@136.com



CUI Zhongma was born in 1978. He received his B.S. degree from Anhui University in 2000, and M.S. degree from the Second Academy of China Aerospace Science and Industry Corporation Limited in 2003, respectively. His major research interests are radar system design and intelligent identification technology.
E-mail: czmsy@sina.com



FANG Jiaqi was born in 1984. He received his B.S., M.S. and Ph.D. degrees from Xidian University in 2007, 2011 and 2016, respectively. His major research interests are radar signal processing and intelligent identification technology.
E-mail: fangjiaqi123@hotmail.com



YANG Xiaoting was born in 1989. She received her B.S. and M.S. degrees from Hebei University of Science and Technology and Beijing Institute of Technology in 2012 and 2016, respectively. She is now an engineer and her major research interest is object detection based on radar signals.
E-mail: younger-bit@sina.com



DING Wei was born in 1992. He received his B.S. degree in North China University of Water Resources and Electric Power in 2017. He is currently an M.S. degree candidate in University of Electronic Science and Technology of China. His current major research interest is lane detection.
E-mail: screamdw@126.com