ShipDeNet-20: An Only 20 Convolution Layers and <1-MB Lightweight SAR Ship Detector

Tianwen Zhang, *Member, IEEE*, and Xiaoling Zhang, *Member, IEEE*

***Abstract*— Existing most deep learning-based synthetic aperture radar (SAR) ship detectors have huge network scale and big model size. Thus, to solve these defects, we propose a lightweight SAR ship detector "ShipDeNet-20" with 20 convolution layers and <1 MB (0.82 MB) model size. We use fewer layers and kernels, and depthwise separable convolution (DS-Conv) to ensure ShipDeNet-20's lightweight attribute. Moreover, we also propose a feature fusion module (FF-Module), a feature enhance module (FE-Module), and a scale share feature pyramid module (SSFP-Module) to compensate for the raw ShipDeNet-20's accuracy loss. Experimental results on the open SAR ship detection data set (SSDD) reveal that the accuracy and speed of ShipDeNet-20 are both superior to the other nine state-of-the-art object detectors. Finally, detection results on another two wide-region SAR images show ShipDeNet-20's strong migration ability. ShipDeNet-20 is a novel SAR ship detector, built from scratch, lighter than others by tens even hundreds of times, helpful for real-time SAR application and future hardware transplantation.**

***Index Terms*— Deep learning, lightweight network, ship detection, synthetic aperture radar (SAR).**

I. Introduction

DEEP learning has received much attention in the synthetic aperture radar (SAR) field, e.g., urban interpretation [1], [2], scattering extraction [3], ship detection [4]–[9], etc., for its advantages of simplicity and high-efficiency compared to traditional methods.

So far, many superb research results have emerged for SAR ship detection. Lin *et al.* [4] proposed a squeeze and excitation rank faster regions-convolutional neural network (Faster R-CNN) to improve accuracy and speed. Cui *et al.* [5] proposed a Dense Attention Pyramid Network for multiscale SAR ship detection. Chang *et al.* [6] proposed a simplified You Only Look Once v2 (YOLOv2) to enhance speed by high-performance computing. Zhao *et al.* [7], [8] proposed a coupled-convolutional neural network (CNN) and a cascade coupled-CNN for small and densely clustered ship detection. Deng *et al.* [9] proposed a multiscale object proposal network and an accurate object detection network for ship detection in multimodal remote sensing images. Deng *et al.* [10] also proposed a learning deep SAR ship detector from scratch.

However, the above methods all have huge network scale and large model size, causing more decrease in detection speed and more challenges in hardware transplantation (A <10 MB model can be transplanted on field-programmable gate arrays (FPGAs) or DSPs [11].). Thence, the topic of lightweight networks is now a research hotspot [12]–[16].

Manuscript received January 17, 2020; revised February 23, 2020; accepted April 22, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61571099 and in part by the National Key Research and Development Program of China under Grant 2017YFB0502700. *(Corresponding author: Xiaoling Zhang.)*

The authors are with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: twzhang@std.uestc.edu.cn; xlzhang@uestc.edu.cn).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2020.2993899

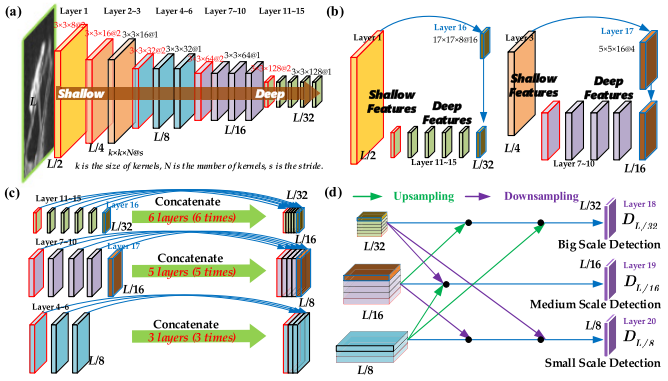

Fig. 1. ShipDeNet-20. (a) Backbone. (b) FF-Module. (c) FE-Module. (d) SSFP-Module.

Thus, to solve these defects, we propose a lightweight SAR ship detector "ShipDeNet-20" (20 convolution layers) with <1 MB (0.82 MB) model size. We use fewer layers and kernels and depthwise separable convolution (DS-Conv) [12]–[15] to ensure ShipDeNet-20's lightweight attribute. Furthermore, we also propose a feature fusion module (FF-Module), a feature enhance module (FE-Module), and a scale share feature pyramid module (SSFP-Module) in order to compensate for the raw ShipDeNet-20's loss in accuracy. Experimental results on the open SAR ship detection data set (SSDD) [17] reveal that ShipDeNet-20's accuracy and speed are both superior to the other nine state-of-art detectors. Finally, results on another two wide-region images can show ShipDeNet-20's strong migration capability. ShipDeNet-20 is a one-stage SAR ship detector on the basis of the grid division principle of YOLO, which is built from scratch and largely lighter than others. It is helpful for real-time SAR application for its faster detection speed and future hardware transplantation for its smaller model size.

The main contributions of this letter are as follows:

1) ShipDeNet-20 is proposed, lighter than the other nine state-of-the-art detectors by tens even hundreds of times.
2) FF-Module, FE-Module, and SSFP-Module are proposed to make up the accuracy loss of the raw ShipDeNet-20.
3) ShipDeNet-20's SAR ship detection accuracy and speed are both superior to the other nine state-of-the-art detectors.

II. ShipDeNet-20

Fig. 1 shows ShipDeNet-20's network architecture. Fig. 1(a) is its backbone (raw ShipDeNet-20), Fig. 1(b) is the FF-Module, Fig. 1(c) is the FE-Module, and Fig. 1(d) is the SSFP-Module.

In Fig. 1, there are 20 convolution layers in all in ShipDeNet-20. Thereinto, 15 layers are in its backbone, 2 ones are in the FF-Module, and 3 ones are in the SSFP-Module. Moreover, the number of layers in ShipDeNet-20 is

universally fewer than the other detectors [18]–[24], bringing its lightweight model.

In ShipDeNet-20, all convolution layers adopt DS-Conv that decouples channel and spatial correlation [15]. Thus, DS-Conv has fewer network parameters compared to the conventional convolution (C-Conv) [25] (see their computation complexity contrast in [12]). Finally, ShipDeNet-20 becomes more lightweight by using DS-Conv.

A. Backbone

From Fig. 1(a), the backbone is a pipeline structure. In Fig. 1(a), $L/2$, $L/4$, $L/8$, $L/16$, and $L/32$ denote the feature maps' size in different layers (L is the image size, set as 160 same as [12]).

In Fig. 1(a), as the network deepening, the feature maps' size becomes smaller ($L/2 \rightarrow L/32$), possibly losing ship features, so we set more kernels in the network's backend ($8 \rightarrow 128$). Moreover, the total number of kernels is also universally fewer than the other detectors [18]–[24], bringing a more lightweight model.

B. FF-Module

In Fig. 1(b), the features of Layers 1,3 are shallow and that of Layers 7–10,11–15 are deep. If we fuse features in different depths, the fused features will become more contextual and semantic [26], which can improve detection accuracy.

In Fig. 1(b), we enforce: (1) the deepest layers ($L/32$, Layers 11–15) to fuse the shallowest layers ($L/2$, Layer 1) by Layer 16; (2) the second deepest layers ($L/16$, Layers 7–10) to fuse the second shallowest layers ($L/4$, Layer 3) by Layer 17, in order to further strengthen the benefits of such feature fusion [26].

Moreover, the kernel dimension of Layer 16 and Layer 17 is set, respectively, as $17 \times 17 \times 8$ at 16 and $5 \times 5 \times 16$ at 4, to keep their feature map's size same as Layers 11–15 and Layers 7–10.

C. FE-Module

From Fig. 1(c), we concatenate the same size feature maps to enhance the ship features' representativeness [12], [26]. Before this feature enhancement, only one layer's features are utilized. However, after this feature enhancement, more layers' features are utilized. In this way, the detection accuracy can be improved.

For example, the features are enhanced by six times (six layers superposition) in $L/32$, five times (five layers superposition) in $L/16$, and three times (three layers superposition) in $L/8$.

D. SSFP-Module

In Fig. 1(d), we use the feature pyramid [27] to detect multiscale ships. The big-scale detection ($D_{L/32}$) is used to detect big ships, the medium-scale one ($D_{L/16}$) is used for medium ships, and the small-scale one ($D_{L/8}$) is used for small ships. Moreover, we also perform scale share (the features in different detection scales are shared by down-sampling or up-sampling), that is

$$D'_{L/32} = D_{L/32} + \text{DownSa}^{\times 2}(D_{L/16}) + \text{DownSa}^{\times 4}(D_{L/8}) \quad (1)$$

$$D'_{L/16} = D_{L/16} + \text{UpSa}^{\times 2}(D_{L/32}) + \text{DownSa}^{\times 2}(D_{L/8}) \quad (2)$$

$$D'_{L/8} = D_{L/8} + \text{UpSa}^{\times 2}(D_{L/16}) + \text{UpSa}^{\times 4}(D_{L/32}) \quad (3)$$

where $\text{DownSa}^{\times n}(\cdot)$ and $\text{UpSa}^{\times n}(\cdot)$ refer n times down-sampling and up-sampling. In this way, the feature pyramid will become more robust [28] (SSFP-Module is also a detection network).

III. Experiments

Our hardware platform is a personal computer with NVIDIA RTX2080Ti GPU and Intel i9-9900k CPU based on Keras.

A. Data Set

SSDD is used to verify the correctness of ShipDeNet-20 and confirm the effectiveness of FF-Module, FE-Module, and SSFP-Module. SSDD is proposed and annotated by Li *et al.* [17] that has been used by many other scholars [5], [6], [12], [29], [30].

SSDD contains 1160 SAR images with 2358 ships, obtained from Sentinel-1, RadarSat-2, and TerraSAR-X. SAR images in SSDD have various polarization modes, various resolutions, and abundant ship scenes. Thus, it is a rather reasonable data set for its sample diversity that can evaluate ship detectors' robustness.

B. Training Strategies

We randomly divide SSDD into a training set, a validation set, and a test set with 7:2:1 ratio [17]. Our loss function comes from [12]. We initialize the network parameters by using the Glorot uniform distribution [31], and then train ShipDeNet-20 from scratch for 2000 epochs in total via the Adam algorithm [32] with batch size as 32 [12]. The "Poly" learning rate policy [33] is adopted with its initial learning rate as 0.001.

C. Evaluation Index

Recall, Precision and mean average precision (mAP) is, respectively, defined by

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \quad (4)$$

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \quad (5)$$

$$\text{mAP} = \int_0^1 \text{P(R)} d\,R \quad (6)$$

where TP is true positive, FN is false negative, FP is false positive, P is precision, R is recall, P(R) is Precision–Recall (P-R) curve, and mAP is mean average precision.

In this letter, we use mAP to represent detection accuracy, and use frames per second (FPS) to represent detection speed.

IV. Results

A. Results on SSDD

Fig. 2 shows the detection results of ShipDeNet-20 on SSDD (Intersection over Union (IoU) ≥ 0.5 [12]). From Fig. 2, most ships with multiscales and multiscenes can be detected. A densely distributed ship and an inshore ship are missed. Two false-alarms appear for duplicate detection of the same ship and islands' erroneous judgment.

B. Discussion on DS-Conv and Three Modules

Table I shows ShipDeNet-20's evaluation indexes on SSDD. From Table I, the following conclusions can be obtained:

1) DS-Conv is indeed lighter than C-Conv ($0.57 <$ 3.31 MB) but it exists loss in accuracy (85.23% mAP $<$ 96.01% mAP).
2) FF-Module can improve accuracy (90.91% mAP $>$ 85.23% mAP) probably given that the fused features can become more contextual and semantic, making detection more robust.
3) FE-Module can improve accuracy (91.08% mAP $>$ 85.23% mAP) probably given that more layers' features

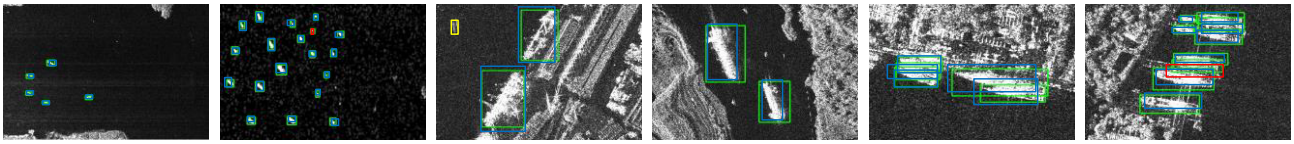

Fig. 2. SAR ship detection results. Green means ground truths, blue means correct detections, red means missed detections, and yellow means false-alarms.

TABLE I

EVALUATION INDEXES OF THE DETECTION RESULTS OF SHIPDENET-20

Type	FF-Module	FE-Module	SSFP-Module	Recall	Precision	mAP	FPS	Parameter	FLOPs	Model Size
C-Conv	-	-	-	97.28%	88.18%	96.01%	223	828,534	1,656,096	3.31 MB
DS-Conv	×	×	×	87.50%	90.96%	85.23%	**284**	108,665	216,376	0.57 MB
DS-Conv	✓	×	×	93.48%	86.87%	90.91%	272	112,273	233,592	0.60 MB
DS-Conv	×	✓	×	94.02%	86.50%	91.08%	268	123,257	245,560	0.64 MB
DS-Conv	×	×	✓	93.48%	**92.47%**	91.88%	270	117,177	233,406	0.61 MB
DS-Conv	✓	✓	×	94.57%	80.18%	92.16%	266	126,865	252,776	0.66 MB
DS-Conv	✓	×	✓	94.57%	92.06%	93.67%	252	121,697	242,446	0.65 MB
DS-Conv	×	✓	✓	94.57%	91.10%	93.08%	247	160,953	320,958	0.79 MB
DS-Conv	✓	✓	✓	**98.37%**	87.02%	**97.07%**	233	165,473	329,998	0.82 MB

TABLE II

EVALUATION INDEXES OF DIFFERENT NUMBERS OF KERNELS IN BACKBONE

Number of kernels in backbone	Recall	Precision	mAP	FPS	Parameter	FLOPs	Model Size
2 → 4 → 8 → 16 → 32	73.91%	93.79%	72.44%	250	24,305	48,454	0.28 MB
4 → 8 → 16 → 32 → 64	91.30%	87.05%	89.20%	240	59,945	119,470	0.42 MB
8 → 16 → 32 → 64 → 128	**98.37%**	87.02%	**97.07%**	233	165,473	329,998	0.82 MB
16 → 32 → 64 → 128 → 256	96.74%	86.41%	95.10%	221	513,521	1,025,038	2.15 MB
32 → 64 → 128 → 256 → 512	97.28%	**95.21%**	96.39%	217	1,757,585	3,511,054	6.89 MB
64 → 128 → 256 → 512 → 1024	97.83%	94.74%	96.40%	202	6,437,585	12,866,830	24.75 MB

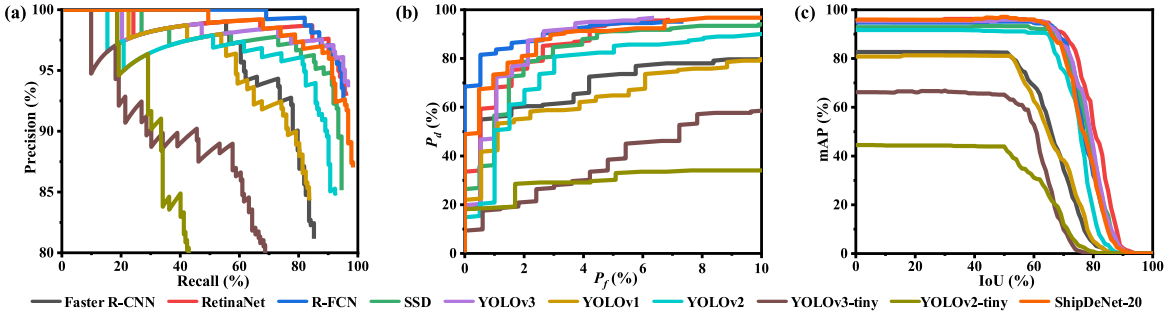

Fig. 3. Results of different methods on SSDD. (a) P(R) curves. (b) Pd-Pf curves. (c) mAP-IoU curves. (d) mAP (accuracy)-FPS (speed) bar graph.

are fully used after enhancement, making detection more efficient.

4) SSFP-Module can improve accuracy (91.88% mAP > 85.23% mAP) probably given that multiscale detection performance can be improved and scale share features can make feature pyramid more effective and robust.
5) The accuracy of the use of two modules is superior to that of the use of only one module probably given the boost of feature representativeness from different types of modules.
6) The detection accuracy of the use of all three modules is the best probably given that the combination of all three modules can give full play to the best performance of the network.

C. Discussion on Backbone

Table II shows the performance change of different numbers of kernels in the backbone. In Table II, n_1, n_2, n_3, n_4, n_5 denote the number of kernel, respectively, in layers with a feature map size of $L/2$ (n_1), $L/4$ (n_2), $L/8$ (n_3), $L/16$ (n_4), and $L/32$ (n_5) in Fig. 1(a). Moreover, generally, n is the power of 2.

From Table II, the following conclusions can be obtained:

1) If $n_1 \leq 8$, the accuracy increases as n_1 increases coming from network learning ability enhancement; meanwhile, the speed decreases coming from more computational cost [FLoating point of OPerations (FLOPs)].
2) If $n_1 = 8$, the accuracy reaches the best. If $n_1 > 8$, the accuracy fluctuates within a stable range, showing that the network has learned enough features and more kernels are redundant.

D. Some Critical Thinking

Our critical thinking for ShipDeNet-20 is as follows:

1) The raw ShipDeNet-20 exists loss in accuracy given that the network has poor learning ability due to limited parameters.
2) More parameters do not mean better performance (exist an upper-limit of accuracy). We should make a specific analysis for specific problems considering detection task's difficulty.
3) We find that over-fitting occurs in a huge model if training 2000 epochs. The pretraining may address such problem but it is bound to lead to reduced training efficiency.
4) The eventual well fine-tuned ShipDeNet-20 containing well-designed backbone and specific types of three modules after careful optimization can maximize the network performance.

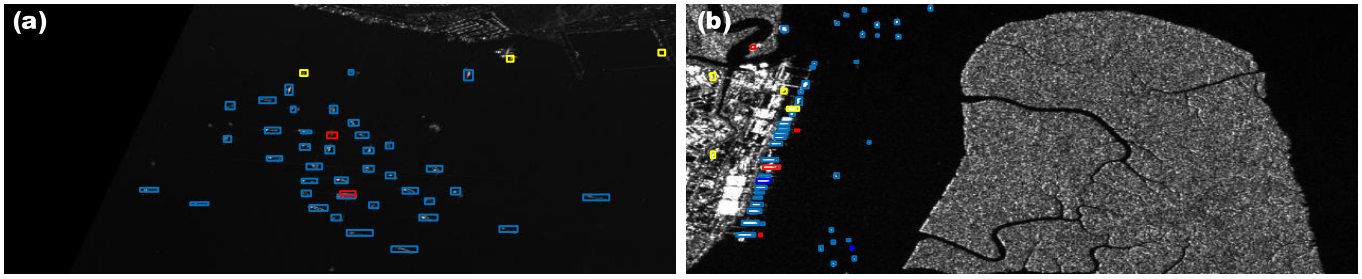

Fig. 4. Results on another two images. (a) Image 1. (b) Image 2. Blue means correct detections, red means missed-detections, and yellow means false-alarms.

TABLE III
EVALUATION INDEXES OF DIFFERENT METHODS ON SSDD. SORT BY MODE SIZE FROM LARGE TO SMALL

Method	Recall	Precision	mAP	FPS	Parameter Quantity	Computational Cost (FLOPs)	Model Size
YOLOv1 [18]	84.07%	84.53%	81.24%	46	272,746,867	545,429,460	752.75 MB
YOLOv3 [19]	96.70%	**93.62%**	95.34%	45	61,576,342	307,592,895	235.44 MB
YOLOv2 [20]	92.86%	84.92%	90.09%	53	50,578,686	101,385,166	193.04 MB
R-FCN [21]	95.65%	92.63%	95.15%	6	47,663,806	95,040,404	181.24 MB
RetinaNet [22]	96.70%	93.12%	95.68%	3	36,382,957	72,545,184	139.25 MB
Faster R-CNN [23]	85.16%	81.15%	82.66%	3	28,342,195	46,981,897,900	108.54 MB
SSD [24]	94.51%	85.15%	92.67%	20	23,745,908	118,685,133	90.73 MB
YOLOv2-tiny [20]	47.80%	73.73%	44.40%	107	15,770,510	31,608,360	60.22 MB
YOLOv3-tiny [19]	70.33%	77.58%	64.64%	98	8,676,244	86,692,284	33.20 MB
Reference [12]	96.15%	87.94%	94.13%	111	3,299,862	8,869,714	38.05 MB
ShipDeNet-20 (This letter)	**98.37%**	87.02%	**97.07%**	**233**	**165,473**	**329,998**	**0.82 MB**

TABLE IV
DETAILED INFORMATION OF TWO GAOFEN-3 AND SENTINEL-1 SAR IMAGES

Number	Sensor	Resolution	Polarization	Image Size
Image 1	Gaofen-3	1 m	HH	3000 × 3000
Image 2	Sentinel-1	10 m	VV	1313 × 908

TABLE V
DETECTION RESULTS OF TWO GAOFEN-3 AND SENTINEL-1 SAR IMAGES

Number	Recall	Precision	mAP	GPU Time	CPU Time
Image 1	95.00%	92.68%	94.58%	278 ms	580 ms
Image 2	91.30%	91.30%	90.14%	259 ms	554 ms

E. Compared With State-of-the-Art

We compare ShipDeNet-20 to nine state-of-the-art detectors [18]–[24] to verify its good detection performance. To be clear, the initial specified 7:2:1 data set partition in Li *et al.* [17] cannot serve as a strict research baseline for its huge sample randomness. In the future, we will use the division criterion in Mao *et al.* [34] to guide our work, i.e., image indexes' suffix 1 and 9 are collected as test subset.

Fig. 3(a)–(d) shows their $P(R)$ curves, $P_d P_f$ curves [30], mAP-IoU curves, and mAP-FPS bar graph. Table II shows the evaluation indexes on SSDD.

From Fig. 3 and Table III, we can draw these conclusions:

1) ShipDeNet-20's speed is superior to others. The second fastest YOLOv2-tiny is still inferior to ShipDeNet-20.
2) ShipDeNet-20's accuracy is superior to others. The second most accurate RetinaNet is still inferior to ShipDeNet-20.
3) From Fig. 3(d) and Table III, ShipDeNet-20 can achieve both high-speed and high-accurate SAR ship detection (green bar and pink bar are both higher than others), while others cannot keep a good balance between accuracy and speed.
4) ShipDeNet-20 is lighter than others by tens even hundreds of times. The second lightest YOLOv3-tiny is still heavier than our ShipDeNet-20 by 40 times ($33.20 \gg 0.82$ MB).

TABLE VI
COMPARED WITH PREVIOUS WORK ON THE SAME DATA SET AND PLATFORM

Method	mAP	FPS	Model Size
Reference [29]	90.16%	100	58.30 MB
Reference [12]	94.13%	111	38.05 MB
Reference [30]	96.93%	115	23.17 MB
ShipDeNet-20 (This letter)	**97.07%**	**233**	**0.82 MB**

F. Migration Ability Verification

We also use ShipDeNet-20 to detect ships in another two wide-region SAR images to verify its strong migration ability.

Table IV shows the two images' detailed information. Based on sliding window means in [35], small subslices are generated and then they are inputted into ShipDeNet-20. Fig. 4 shows their detection results. Table IV shows their evaluation indexes.

From Fig. 4 and Table V, most ships can be successfully detected by ShipDeNet-20. Two ships are missed and three false-alarms appear in image 1 (94.58% mAP). Four ships are missed and four false-alarm appear in image 2 (90.14% mAP).

Therefore, ShipDeNet-20 has a strong migration capability that can be applied to actual SAR ship detection.

G. Compared With Our Previous Work

ShipDeNet-20 is also compared to our previous work [12], [29], [30]. Table VI shows their contrast results under the same data set and platform. From Table VI, ShipDeNet-20 has higher accuracy (97.07% mAP), faster speed (233 FPS), and lighter model (0.82 MB) than our previous work.

V. CONCLUSION

A lightweight SAR ship detector ShipDeNet-20 is proposed in this letter. When combined with FF-Module, FE-Module, and SSFP-Module, ShipDeNet-20's accuracy and speed are both superior to the other nine state-of-the-art object detectors. Experimental results on the open SSDD verify ShipDeNet-20's effectiveness. Its good migration ability is confirmed on

another two SAR images. ShipDeNet-20 is a novel network that is built from scratch. It only has 20 convolution layers, and its model size is only 0.82 MB, lighter than other detectors by tens even hundreds of times, conducive to rapid maritime rescue and emergency military deployment, etc., due to its faster speed and future hardware transplantation due to its smaller model size.

ShipDeNet-20 can achieve higher accuracy, coming from:

1) FF-Module, FE-Module, and SSFP-Module proposed in this letter can effectively improve the detection accuracy.
2) SAR ship detection is easier than other tasks in the computer vision field, just binary classification (ship and background).
3) SSDD data set is more suited to lighter networks for its fewer samples. If not, heavy networks may occur over-fitting [30].
4) ShipDeNet-20 can be trained more efficiently given its faster training iteration speed due to its fewer parameters [30], [36].

Future works: Hardware transplantation will be studied in the future. Last but not least, the SSDD data set's publishers Li *et al.* [17] have contacted us to appeal and even command scholars of this field to use the new released data set division criterion in Mao *et al.* [34] to ensure a more fair comparison (i.e., image indexes' suffix 1 and 9 are set as the test set), so we will conduct our research under the guidance of such criterion in the future.

REFERENCES

[1] J. Zhao, Z. Zhang, W. Yao, M. Datcu, H. Xiong, and W. Yu, "OpenSARUrban: A Sentinel-1 SAR image dataset for urban interpretation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 187–203, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8952866, doi: 10.1109/JSTARS.2019.2954850.

[2] Z. Zhang, W. Guo, M. Li, and W. Yu, "GIS-supervised building extraction with label noise-adaptive fully convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, early access, Jan. 30, 2020, [Online]. Available: https://ieeexplore.ieee.org/document/8976267, doi: 10.1109/LGRS.2019.2963065.

[3] J. Zhao, M. Datcu, Z. Zhang, H. Xiong, and W. Yu, "Contrastive-regulated CNN in the complex domain: A method to learn physical scattering signatures from flexible PolSAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 10116–10135, Dec. 2019.

[4] Z. Lin, K. Ji, X. Leng, and G. Kuang, "Squeeze and excitation rank faster R-CNN for ship detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 751–755, May 2019.

[5] Z. Cui, Q. Li, Z. Cao, and N. Liu, "Dense attention pyramid networks for multi-scale ship detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8983–8997, Nov. 2019.

[6] Y.-L. Chang, A. Anagaw, L. Chang, Y. Wang, C.-Y. Hsiao, and W.-H. Lee, "Ship detection based on YOLOv2 for SAR imagery," *Remote Sens.*, vol. 11, no. 7, p. 786, 2019.

[7] J. Zhao *et al.*, "A Coupled Convolutional Neural Network for small and densely clustered ship detection in SAR images," *Sci. China Inf. Sci.*, vol. 62, no. 4, p. 042301, 2019.

[8] J. Zhao, Z. Zhang, W. Yu, and T.-K. Truong, "A cascade coupled convolutional neural network guided visual attention method for ship detection from SAR images," *IEEE Access*, vol. 6, pp. 50693–50708, 2018.

[9] Z. Deng, H. Sun, S. Zhou, J. Zhao, L. Lei, and H. Zou, "Multi-scale object detection in remote sensing imagery with convolutional neural networks," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 3–22, Nov. 2018.

[10] Z. Deng, H. Sun, S. Zhou, and J. Zhao, "Learning deep ship detector in SAR images from scratch," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 4021–4039, Jun. 2019.

[11] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size," 2016, *arXiv:1602.07360*. [Online]. Available: http://arxiv.org/abs/1602.07360

[12] Zhang, Zhang, Shi, and Wei, "Depthwise separable convolution neural network for high-speed SAR ship detection," *Remote Sens.*, vol. 11, no. 21, p. 2483, 2019.

[13] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*. [Online]. Available: http://arxiv.org/abs/1704.04861

[14] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," 2017, *arXiv:1707.01083*. [Online]. Available: http://arxiv.org/abs/1707.01083

[15] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," 2016, *arXiv:1610.02357*. [Online]. Available: http://arxiv.org/abs/1610.02357

[16] J. Bethge, C. Bartz, H. Yang, Y. Chen, and C. Meinel, "MeliusNet: Can binary neural networks achieve MobileNet-level accuracy?" 2020, *arXiv:2001.05936*. [Online]. Available: http://arxiv.org/abs/2001.05936

[17] J. Li, C. Qu, and J. Shao, "Ship detection in SAR images based on an improved faster R-CNN," in *Proc. SAR Big Data Era: Models, Methods Appl. (BIGSARDATA)*, Beijing, China, Nov. 2017, pp. 1–6.

[18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," 2015, *arXiv:1506.02640*. [Online]. Available: http://arxiv.org/abs/1506.02640

[19] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*. [Online]. Available: http://arxiv.org/abs/1804.02767

[20] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," 2016, *arXiv:1612.08242*. [Online]. Available: http://arxiv.org/abs/1612.08242

[21] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," 2016, *arXiv:1605.06409*. [Online]. Available: http://arxiv.org/abs/1605.06409

[22] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," 2017, *arXiv:1708.02002*. [Online]. Available: http://arxiv.org/abs/1708.02002

[23] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," 2015, *arXiv:1506.01497*. [Online]. Available: http://arxiv.org/abs/1506.01497

[24] W. Liu *et al.*, "SSD: Single shot MultiBox detector," 2015, *arXiv:1512.02325*. [Online]. Available: http://arxiv.org/abs/1512.02325

[25] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Oct. 1998.

[26] H. Lee and H. Kwon, "Going deeper with contextual CNN for hyperspectral image classification," 2016, *arXiv:1604.03519*. [Online]. Available: http://arxiv.org/abs/1604.03519

[27] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," 2016, *arXiv:1612.03144*. [Online]. Available: http://arxiv.org/abs/1612.03144

[28] Y. Zhao, R. Han, and Y. Rao, "A new feature pyramid network for object detection," in *Proc. Int. Conf. Virtual Reality Intell. Syst. (ICVRIS)*, Jishou, China, Sep. 2019, pp. 428–431.

[29] T. Zhang and X. Zhang, "High-speed ship detection in SAR images based on a grid convolutional neural network," *Remote Sens.*, vol. 11, no. 10, p. 1206, 2019.

[30] X. Zhang *et al.*, "High-speed and High-accurate SAR ship detection based on a depthwise separable convolution neural network," *J. Radars*, vol. 8, no. 6, pp. 841–851, 2019.

[31] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *J. Mach. Learn. Res.*, vol. 9, pp. 249–256, May 2010.

[32] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[33] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*. [Online]. Available: http://arxiv.org/abs/1706.05587

[34] Y. Mao, Y. Yang, Z. Ma, M. Li, H. Su, and J. Zhang, "Efficient low-cost ship detection for SAR imagery based on simplified U-Net," *IEEE Access*, vol. 8, pp. 69742–69753, 2020.

[35] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "A SAR dataset of ship detection for deep learning under complex backgrounds," *Remote Sens.*, vol. 11, no. 7, p. 765, 2019.

[36] T. Akiba, S. Suzuki, and K. Fukuda, "Extremely large minibatch SGD: Training ResNet-50 on ImageNet in 15 minutes," 2017, *arXiv:1711.04325*. [Online]. Available: http://arxiv.org/abs/1711.04325