

An Approach of Sea Clutter Suppression for SAR Images by Self-Supervised Complex-Valued Deep Learning

Qinglong Hua^{ID}, Yun Zhang, *Member, IEEE*, Huilin Mu^{ID}, Yicheng Jiang^{ID}, and Dan Xu

Abstract—Strong reflections from the marine surface reduce the contrast between the target of interest and the background in synthetic aperture radar (SAR) images and severely affect the interpretation of the image. This letter proposes a framework of SAR sea clutter suppression based on a new self-supervised training strategy referred to as Clutter2Clutter (C2C), which mines self-supervised information from a large number of unlabeled SAR patches for network training. This letter also proposes a complex-valued UNet++ (CV-UNet++) network model to make full use of both amplitude and phase information of the complex SAR image, and the C2C strategy is used to train the CV-UNet++ for sea clutter suppression. Experiments on GF-3 and TerraSAR-X SAR data show that the proposed method has a better effect on suppressing sea clutter and is able to preserve the target-of-interest energy well.

Index Terms—Clutter2Clutter (C2C), complex-valued deep learning, sea clutter suppression, synthetic aperture radar (SAR).

I. INTRODUCTION

SYNTHETIC aperture radar (SAR) has indisputable advantages in marine monitoring [1]. The application of using SAR technology to detect ship targets has a wide range of realistic prospects. Sea clutter caused by strong reflections from the marine surface is a typical problem for SAR image interpretation, which has complex characteristics and a strong time correlation. It leads to the degradation of SAR image quality and affects the performance of ship detection [2] and classification [3]. Several methods have been proposed to reduce sea clutter, including constant false alarm rate (CFAR) [4], wavelet [5], [6], and filtering methods [7], [8]. Zhao *et al.* [9] proposed a shedding irrelevant patterns (SIP) method by constructing a regression function that could suppress clutter and preserve the target patterns concurrently. Some techniques achieve clutter suppression by employing multiple channels, such as DPCA [10] and STAP [11]. Li *et al.* [12] proposed a subspace projection clutter suppression method, which shows better performance than STAP. The effect of traditional algorithms relies on the manual selection of parameters, and it is difficult to achieve a balance between preserving different image features and removing clutter.

Manuscript received 29 May 2022; accepted 12 June 2022. Date of publication 16 June 2022; date of current version 22 August 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61201308 and Grant 61671490 and in part by the Key Laboratory of Marine Environmental Monitoring and Information Processing, Ministry of Industry and Information Technology. (*Corresponding author:* Yun Zhang.)

Qinglong Hua, Yun Zhang, Yicheng Jiang, and Dan Xu are with the School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China (e-mail: huqinglong_hit@163.com; zhangyunhit@hit.edu.cn; jiangyc@hit.edu.cn; xudan_hit@163.com).

Huilin Mu is with the Air Defense and Antimissile School, Air Force Engineering University, Xi'an 710051, China (e-mail: muhl_afeu@163.com). Digital Object Identifier 10.1109/LGRS.2022.3183582

In parallel to the massive success of CNN in the computer vision area, several deep-learning-based methods [13]–[15] have been developed. Mu *et al.* [14] proposed a CV-GMTINet to discriminate moving targets from stationary objects and suppress background clutter. Most of these methods convert the problem of clutter suppression into a regression task and use CNN-based architecture for training in an end-to-end fashion. Meanwhile, these methods require image pairs of clutter and clutter-free for supervised training, and a big problem in supervised deep-learning-based techniques is the lack of ground-truth data.

In recent years, some literature has tried to explore self-supervised learning due to its potential to leverage the massive amount of unlabeled data. Lehtinen *et al.* [16] proved that the network is able to learn a clean representation as long as two noise images with independent and identical noise distributions are available and proposed a Noise2Noise (N2N) training strategy. In [17], an adversarial learning framework is utilized to generate SAR image pairs from the same scene, and the image pairs are trained using the N2N strategy. In [18], a registration network is designed to compensate for the motion between two adjacent frames in video SAR. Whether using adversarial learning or the registration network, the cost of constructing pairs of images with independent and identical distributions is still high. Moreover, the above-mentioned deep networks all belong to real-valued networks, which could only process the amplitude data of SAR images and lack the ability to perceive SAR phase information.

In this letter, a practical suppression method of SAR sea clutter is introduced. Our contributions are summarized as follows. First, this letter proposes a framework for SAR sea clutter suppression based on a new self-supervised training strategy referred to as Clutter2Clutter (C2C). The framework does not require clean SAR images as ground truth. Second, the complex-valued UNet++ (CV-UNet++) network is proposed for sea clutter suppression, which makes full use of both amplitude and phase information of SAR images. Third, visual and quantitative experiments on GF-3 and TerraSAR-X SAR data show that the proposed method could significantly suppress sea clutter and preserve the target-of-interest energy well, which outperforms several traditional methods. The proposed method could be used for the suppression of heterogeneous and strong sea clutter before target detection and sea suppression of moving ship targets after detection.

II. THEORY FOR C2C

The implementation details of the proposed method are shown in Fig. 1, which consists of three parts, and the three parts will be introduced in this section.

1558-0571 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See <https://www.ieee.org/publications/rights/index.html> for more information.

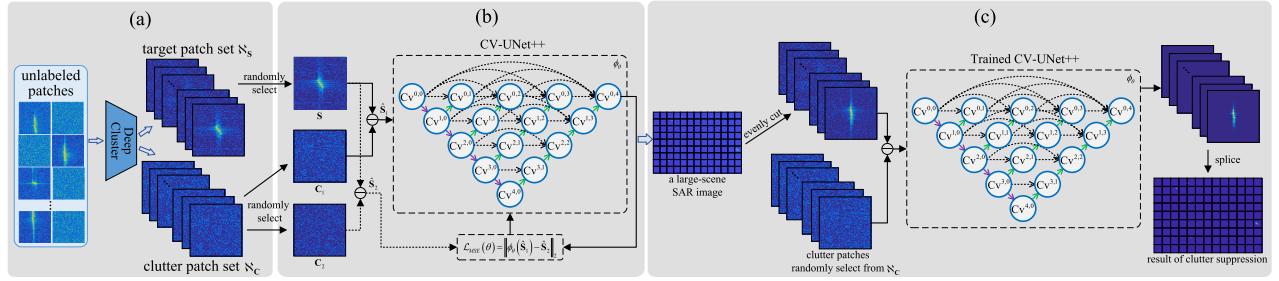


Fig. 1. Algorithm of clutter suppression based on C2C strategy and CV-UNet++. (a) Preparation before training. (b) C2C clutter suppression strategy. (c) Large-scene SAR sea clutter suppression.

A. SAR Sea Clutter Model

Let $\mathbf{S} \in \mathbb{C}^{W \times H}$ be the complex SAR image with a size of $W \times H$, $\mathbf{T} \in \mathbb{C}^{W \times H}$ be the clean image without any clutter, $\mathbf{C} \in \mathbb{C}^{W \times H}$ be the sea clutter image, and $\mathbf{N} \in \mathbb{C}^{W \times H}$ be the noise. Then, \mathbf{S} could be obtained by the following additive model [14]:

$$\mathbf{S} = \mathbf{T} + \mathbf{C} + \mathbf{N}. \quad (1)$$

Assume that the amplitude of \mathbf{C} obeys the Weibull distribution [19], and the phase obeys the uniform distribution [20]. The probability density function of \mathbf{C} could be written as

$$p(\|\mathbf{C}\|) = \frac{k}{\lambda} \left(\frac{\|\mathbf{C}\|}{\lambda} \right)^{k-1} \exp \left(- \left(\frac{\|\mathbf{C}\|}{\lambda} \right)^k \right) \quad (2)$$

$$p(\arg(\mathbf{C})) \sim U(-\pi, \pi) \quad (3)$$

where $\lambda > 0$ is the scale parameter. $k > 0$ is the shape parameter.

Then the expectation of \mathbf{C} is

$$\mathbb{E}(\mathbf{C}) = \lambda \Gamma \left(1 + \frac{1}{k} \right) \quad (4)$$

where \mathbb{E} denotes the expectation operator. $\Gamma(\cdot)$ denotes the gamma function. The aim of sea clutter suppression is to estimate \mathbf{T} from the complex SAR image \mathbf{S} . Based on the SAR sea clutter model in (1)–(4), the proposed method in Section II-C could be trained without using any ground-truth images.

B. Preparation Before Self-Supervised Training

Before self-supervised training, it is necessary to prepare a target patch set \mathcal{N}_s and a clutter patch set \mathcal{N}_c , which are constructed from a large number of unlabeled SAR patches by the unsupervised clustering method DeepCluster [21]. The clutter patch set \mathcal{N}_c contains clutter patches with different distributions and different intensities. The target patch is an image block of size $W \times H$ intercepted from SAR images, and the target patch contains both the target of interest and the background clutter. The size of the clutter patch is the same as the target patch, but the clutter patch contains only background clutter. In the subsequent training and test stages, clutter patches need to be randomly selected from the clutter patch set \mathcal{N}_c . Compared with constructing a large number of clutter-free ground truths in supervised learning, the preparation here is faster, and the cost is meager. Based on the target patch set and clutter patch set, a self-supervised strategy is constructed in Section II-C.

C. C2C Strategy

A target patch is randomly selected from the target patch set \mathcal{N}_s . Because the target patch contains the target-of-interest \mathbf{T} , background clutter \mathbf{C} , and noise \mathbf{N} , the selected target patch could be represented by \mathbf{s} like in (1). Then two clutter patches are randomly selected from the clutter patch set \mathcal{N}_c . Because the clutter patch contains background clutter and noise, the two clutter patches could be denoted by $\mathbf{C}_1 + \mathbf{N}_1$ and $\mathbf{C}_2 + \mathbf{N}_2$, respectively. SAR image pairs $(\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2)$ from the same scene could be obtained as follows:

$$(\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2) = (\mathbf{s} - \mathbf{C}_1, \mathbf{s} - \mathbf{C}_2). \quad (5)$$

After obtaining the SAR image pairs $(\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2)$, the clutter suppression model ϕ_θ is trained without clean ground-truth images. Mean squared error (mse) loss is employed to optimize ϕ_θ as follows:

$$\mathcal{L}_{\text{MSE}}(\theta) = \|\phi_\theta(\hat{\mathbf{s}}_1) - \hat{\mathbf{s}}_2\|_2 \quad (6)$$

where θ denotes the learnable parameters of ϕ_θ .

The expectation of $\hat{\mathbf{s}}_1$ could be written as

$$\begin{aligned} \mathbb{E}(\hat{\mathbf{s}}_1) &= \mathbb{E}(\mathbf{s} - \mathbf{C}_1 - \mathbf{N}_1) \\ &= \mathbb{E}(\mathbf{T} + \mathbf{C} + \mathbf{N} - \mathbf{C}_1 - \mathbf{N}_1) \\ &= \mathbb{E}(\mathbf{T} + \Delta\mathbf{C}_1 + \Delta\mathbf{N}_1) \\ &= \mathbb{E}(\mathbf{T}) + \mathbb{E}(\Delta\mathbf{C}_1) + \mathbb{E}(\Delta\mathbf{N}_1). \end{aligned} \quad (7)$$

As the clutter patch is randomly selected from \mathcal{N}_c , \mathbf{C} and \mathbf{C}_1 could be considered to have the same expectations from a statistical point of view, so $\mathbb{E}(\Delta\mathbf{C}_1) \approx 0$. Similarly, $\mathbb{E}(\Delta\mathbf{N}_1) \approx 0$. Hence, (7) could be written as

$$\mathbb{E}(\hat{\mathbf{s}}_1) \approx \mathbb{E}(\mathbf{T}). \quad (8)$$

Similarly, $\mathbb{E}(\hat{\mathbf{s}}_2) \approx \mathbb{E}(\mathbf{T})$. Therefore, the expectations of $\hat{\mathbf{s}}_1$ and $\hat{\mathbf{s}}_2$ are the same as the expectations of the clutter-free image \mathbf{T} . According to N2N strategy, the clutter suppression model ϕ_θ could minimize $\mathcal{L}_{\text{MSE}}(\theta)$ and generate a clutter-free image without using any clean ground-truth images. This C2C strategy mines self-supervised information from a large number of unlabeled SAR patches for network training, which belongs to the category of self-supervised learning.

In order to make full use of the amplitude and phase information of the complex SAR image, CV-UNet++ is used as the clutter suppression model ϕ_θ , as shown in Fig. 2. CV-UNet++ is an extension of UNet++ [22] to the complex domain. The convolutional layer, pooling layer, and activation function in UNet++ are all extended to the complex domain.

In the convolutional layer, the input feature mapping \mathbf{a}^{l-1} is convolved with the weight matrix \mathbf{W}^l of the l th convolutional

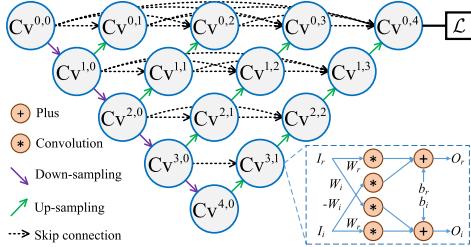


Fig. 2. CV-UNet++ architecture.

layer, and the bias \mathbf{b}^l is added to obtain \mathbf{z}^l . \mathbf{z}^l is activated by the complex-valued activation function $\sigma_{\Re-\Im}(\cdot)$ to obtain the output \mathbf{a}^l of the convolutional layer. The calculation process could be expressed as follows:

$$\begin{aligned}\mathbf{a}^l &= \sigma_{\Re-\Im}(\mathbf{z}^l) \\ &= \sigma(\Re(\mathbf{W}^l) * \Re(\mathbf{a}^{l-1}) - \Im(\mathbf{W}^l) * \Im(\mathbf{a}^{l-1})) + \Re(\mathbf{b}^l) \\ &\quad + j\sigma(\Re(\mathbf{W}^l) * \Im(\mathbf{a}^{l-1}) + \Im(\mathbf{W}^l) * \Re(\mathbf{a}^{l-1}) + \Im(\mathbf{b}^l))\end{aligned}\quad (9)$$

where j is the imaginary unit. \Re and \Im denote the real and imaginary parts of a complex number, respectively. Character $*$ is the convolution operation. The complex-valued activation function $\sigma_{\Re-\Im}(\cdot)$ activates the real and imaginary parts separately. For the complex variable z , the activation process of $\sigma_{\Re-\Im}(\cdot)$ is as follows:

$$\sigma_{\Re-\Im}(z) = \sigma(\Re(z)) + j\sigma(\Im(z)) \quad (10)$$

where $\sigma(\cdot)$ represents the real-valued activation function. Note that the $\sigma(\cdot)$ activation function of the CV-UNet++ is the leaky-ReLU function.

Assume the error term of the l th convolutional layer is δ^l , then the error term of the $l-1$ th layer could be derived as

$$\begin{aligned}\delta^{l-1} &= \left(\Re(\delta^l) * (\Re(\mathbf{W}^l))^\Theta \right. \\ &\quad \left. + \Im(\delta^l) * (\Im(\mathbf{W}^l))^\Theta \right) \odot \sigma'(\Re(\mathbf{z}^{l-1})) \\ &\quad - j \left(\Re(\delta^l) * (\Im(\mathbf{W}^l))^\Theta \right. \\ &\quad \left. - \Im(\delta^l) * (\Re(\mathbf{W}^l))^\Theta \right) \odot \sigma'(\Im(\mathbf{z}^{l-1}))\end{aligned}\quad (11)$$

where $\sigma'(\cdot)$ is the derivative of the activation function $\sigma(\cdot)$. $(\cdot)^\Theta$ denotes the matrix rotates by 180° . The gradients of l th convolutional layer's weight and bias are

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \mathbf{W}^l} &= \Re(\mathbf{a}^{l-1}) * \Re(\delta^l) + \Im(\mathbf{a}^{l-1}) * \Im(\delta^l) \\ &\quad + j(\Re(\mathbf{a}^{l-1}) * \Im(\delta^l) - \Im(\mathbf{a}^{l-1}) * \Re(\delta^l)) \\ \frac{\partial \mathcal{L}}{\partial \mathbf{b}^l} &= \sum \delta^l.\end{aligned}\quad (12)$$

Then the update formula of weight and bias in the l th convolutional layer are

$$\mathbf{W}^l \leftarrow \mathbf{W}^l + \eta \frac{\partial \mathcal{L}}{\partial \mathbf{W}^l} \quad (13)$$

$$\mathbf{b}^l \leftarrow \mathbf{b}^l + \eta \frac{\partial \mathcal{L}}{\partial \mathbf{b}^l} \quad (14)$$

where η is the learning rate.

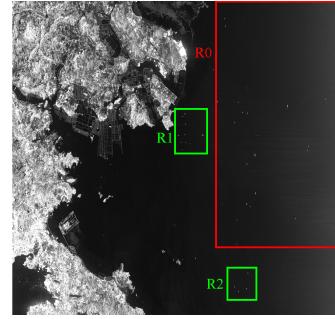


Fig. 3. GF-3 SAR data for experiment. R0 is the training region. R1 and R2 are the test regions.

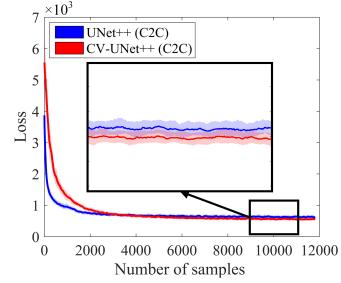


Fig. 4. Loss function converges of UNet++ and CV-UNet++.

The pooling layer could be regarded as sampling the feature maps. Similar to the definition of $\sigma_{\Re-\Im}(\cdot)$, the pooling layer in CV-UNet++ could be represented as

$$\text{ave}_{\Re-\Im}(z) = \text{ave}(\Re(z)) + j\text{ave}(\Im(z)) \quad (15)$$

where $\text{ave}(\cdot)$ denotes the average pooling operation.

With the SAR image pairs $(\hat{\mathbf{S}}_1, \hat{\mathbf{S}}_2)$, an effective clutter suppression model ϕ_θ could be trained by the C2C strategy and CV-UNet++ without ground truth.

D. Sea Clutter Suppression for Large-Scene SAR Images

After training CV-UNet++ by the C2C strategy, sea clutter suppression for large-scene SAR images is performed. First, the large-scene SAR images is cropped according to the size of $W \times H$ to form a patch set. Note that the cropping of the large-scene SAR here is not a sliding window, and there is no overlap between the patches. The processing for each patch \mathbf{S} is as follows: 1) randomly select I clutter patches $(\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_I)$ in the clutter patch set \mathbf{N}_C ; 2) the patch \mathbf{S} is subtracted from the I clutter patches $(\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_I)$ to obtain $(\hat{\mathbf{S}}_1, \hat{\mathbf{S}}_2, \dots, \hat{\mathbf{S}}_I)$ according to the C2C strategy; and 3) $(\hat{\mathbf{S}}_1, \hat{\mathbf{S}}_2, \dots, \hat{\mathbf{S}}_I)$ are input into the trained CV-UNet++ model to obtain I clutter suppression patches $(\mathbf{T}'_1, \mathbf{T}'_2, \dots, \mathbf{T}'_I)$, and $\mathbf{T}' = (1/I) \sum_{i=1}^I \mathbf{T}'_i$ is took as the final clutter suppression image of the patch \mathbf{S} . Finally, the above processing is performed on all the patches in the patch set and then the clutter-suppressed patches are spliced into the final output.

III. EXPERIMENTAL RESULTS

A. Model Training

On the one-scene GF-3 SAR data, as shown in Fig. 3, one training region R0 and two test regions, R1 and R2, are selected. According to the proposed method of Section II-B, R0 is used to construct the target patch set \mathbf{N}_S and the

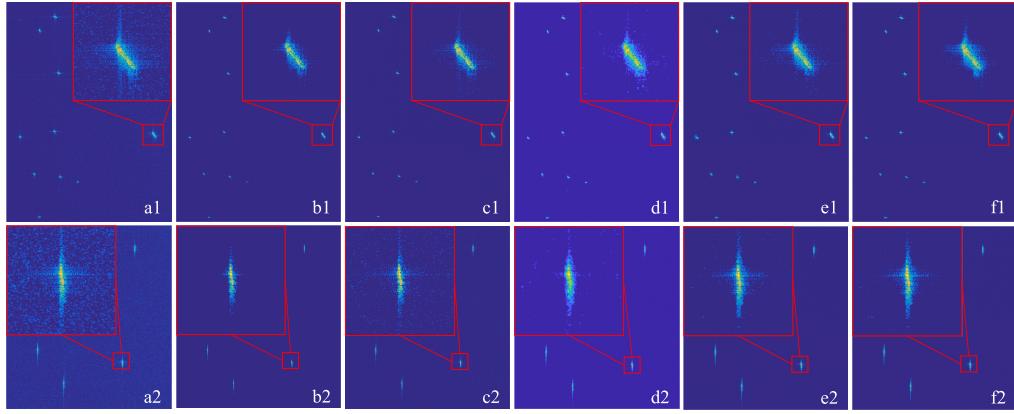


Fig. 5. Clutter suppression results. (a1) Original image of test region R1. (b1) Ground truth of R1. (c1)–(f1) Results of SIP, CV-GMTINet, UNet++, and CV-UNet++ trained by C2C, and CV-UNet++ trained by C2C on R1. (a2) Original image of R2. (b2) Ground truth of R2. (c2)–(f2) Results of SIP, CV-GMTINet, UNet++, and CV-UNet++ on R2.

clutter patch set \aleph_C . \aleph_S contains 20 000 target patches. \aleph_C contains 40 000 clutter patches. The patch size is $W \times H = 128 \times 128$. The number of epochs is ten, and the number of training samples in each batch is 16. The initial learning rate is 0.0001 and is reduced to 0.00001 for six epochs. The convolution kernel numbers of CV-UNet++ are 4, 8, 16, 32, and 64, respectively. For a fair comparison, UNet++ has the same degree of freedom (DoF) as CV-UNet++, so the convolution kernel numbers of UNet++ are 8, 16, 32, 64, and 128, respectively.

The loss function curves of UNet++ and CV-UNet++ trained by the C2C strategy are shown in Fig. 4. From Fig. 4, the convergence speed of UNet++ is faster than CV-UNet++ at the initial stage (one to three epochs). But the loss of UNet++ almost no longer declines after three epochs, while the loss of CV-UNet++ continues to decline and is smaller than UNet++. The better performance in the later training stage reflects the superiority of complex-valued deep learning.

B. Performance Comparison on SAR Data

SIP, CV-GMTINet, UNet++, and CV-UNet++ are used to suppress clutter in the test regions R1 and R2, respectively. The clutter suppression results are shown in Fig. 5. According to the qualitative analysis in Fig. 5, the traditional method of SIP has certain clutter suppression effects. By comparing the original image in Fig. 5(a2) and the clutter suppression result of CV-GMTINet in Fig. 5(d2), the CV-GMTINet with supervised training has a good effect on clutter suppression in the background area, but is not satisfactory in the target area. Both UNet++ and CV-UNet++ have good suppression effects on background clutter, and CV-UNet++ has a better effect than UNet++. The two deep-learning-based methods suppress the clutter energy while preserving the target energy, and even the defocused energy of moving targets is well preserved. UNet++ and CV-UNet++ retain the originality and completeness of the target of interest to the greatest extent compared with traditional methods. They hardly cause the loss of information or energy contained in the target, which is very beneficial for the subsequent estimation of target motion parameters and refocusing processing. This fully reflects the advantages of the C2C strategy and deep learning network for SAR sea clutter suppression.

The ratio of the radar signal to clutter and noise (SCNR) is used to compare the performance of the above clutter

TABLE I
SCNR (IN dB) PERFORMANCE COMPARISONS

	Origin	SIP	CV-GMTINet	UNet++ (C2C)	CV-UNet++ (C2C)
R1	12.34	16.62	20.66	19.58	21.02
R2	10.16	13.08	18.21	18.27	18.72
OA	9.95	13.36	18.55	19.11	19.72

suppression methods. The SCNR (in dB) performance comparisons and the overall average (OA) SCNR after clutter suppression for a dataset containing four other SAR large-scene images are shown in Table I. From the quantitative analysis of Table I, whether in the high SCNR region R1 or the low SCNR region R2, the SCNRs of the two test regions are still low after clutter suppression by the traditional method. Compared with the traditional method, the two test regions, R1 and R2, after clutter suppression by UNet++ and CV-UNet++, have a higher SCNR, and CV-UNet++ has the largest SCNR. Whether it is analyzed from a qualitative or quantitative perspective, CV-UNet++ trained by the C2C strategy has the best performance of clutter suppression.

According to Section II-D, by simply averaging the subtraction of target and clutter patch, that is, $\hat{S} = 1/I \sum_{i=1}^I \hat{S}_i$, a certain effect of clutter suppression would also be obtained. When the number of clutter patches I is five, this simply averaging method improves the SCNR by about 3 dB, and its effect is similar to the SIP method. As the increase of I , the SCNR improvement is more obvious, which confirms to a certain extent that when \aleph_C contains enough clutter patches, $E(\Delta C_1) \approx 0$ is satisfied statistically.

The calculation efficiency is statistically analyzed on a computer configured with Intel Xeon¹ Silver 4110 CPU and NVIDIA GeForce RTX 1080 graphics card. Count the processing time of each method on the test regions R1 and R2, and the results are shown in Table II. The OA in Table II represents the OA processing time for the dataset containing four other SAR large-scene images. From Table II, the processing time of UNet++ is the shortest, and the processing time of SIP is the longest. By using the graphics card for parallel computing, the processing time of the three deep learning methods is shorter than SIP. Because of the cross-multiplication of complex numbers, CV-UNet++ takes about twice as much computation

¹Registered trademark.

TABLE II
CALCULATION EFFICIENCY STATISTICS (S)

	SIP	CV-GMTINet	UNet++ (C2C)	CV-UNet++ (C2C)
R1	244.06	19.12	18.23	34.21
R2	108.02	8.46	8.11	13.14
OA	137.68	15.21	12.93	22.35

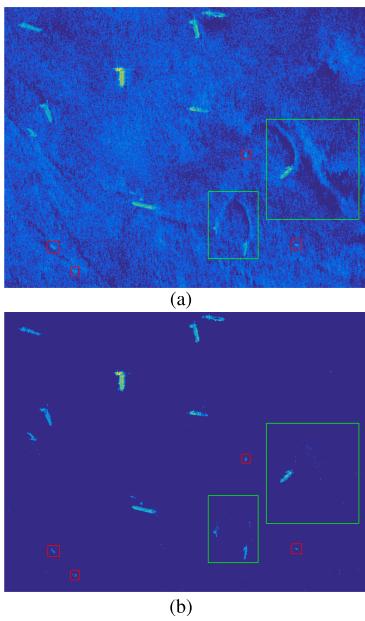


Fig. 6. Performance of the proposed method with TerraSAR-X data. (a) Original image. (b) Clutter suppression result of CV-UNet++. Weak targets are in the red rectangle. Heterogeneous and strong clutter is in the green rectangle.

as UNet++ for the same DoF. And the number of parameters in CV-UNet++ is more than CV-GMTINet. Therefore, it is reasonable that the time consumption of CV-UNet++ is bigger than UNet++ and CV-GMTINet.

Finally, an additional experiment is performed on a TerraSAR-X SAR image with heterogeneous and strong clutter to test the performance of the proposed methods at the clutter edge area and the complex background. From Fig. 6(a), the original image has a strong clutter background, and its SCNR is 9.74 dB. The image after clutter suppression by the proposed method is shown in Fig. 6(b). It could be seen from Fig. 6(b) that the sea clutter in Fig. 6(a) is significantly suppressed, and all ship targets, even weak targets in the red rectangle, are well preserved. Heterogeneous and strong clutter in the green rectangle is well suppressed. The proposed method performs well in the clutter edge area and the complex background. Moreover, the SCNR of the image after clutter suppression is 18.05 dB, and the clutter suppression effect is satisfactory.

IV. CONCLUSION

This letter has proposed a framework of SAR sea clutter suppression based on the C2C training strategy and CV-UNet++, which could achieve suppressing sea clutter without ground truth. Experiments show that the proposed method could improve the clutter suppression performance while fully retaining the information of the target of interest

compared with traditional methods. In the future study, we will try to use the proposed method to suppress SAR radio-frequency interference.

REFERENCES

- [1] R. Pelich, N. Longepe, G. Mercier, G. Hajduch, and R. Garello, "AIS-based evaluation of target detectors and SAR sensors characteristics for maritime surveillance," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 8, pp. 3892–3901, Aug. 2015.
- [2] M. Tello, C. Lopez-Martinez, and J. J. Mallorqui, "A novel algorithm for ship detection in SAR imagery based on the wavelet transform," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 2, pp. 201–205, Apr. 2005.
- [3] H. Lang, S. Wu, and Y. Xu, "Ship classification in SAR images improved by AIS knowledge transfer," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 439–443, Mar. 2018.
- [4] D. J. Crisp, "The state-of-the-art in ship detection in synthetic aperture radar imagery," DSTO, Inf. Sci. Lab., Fairbairn, Canberra, ACT, Tech. Rep., May 2004.
- [5] H. Xie, L. E. Pierce, and F. T. Ulaby, "SAR speckle reduction using wavelet denoising and Markov random field modeling," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2196–2212, Oct. 2002.
- [6] F. Argenti and L. Alparone, "Speckle removal from SAR images in the undecimated wavelet domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2363–2374, Nov. 2002.
- [7] J.-S. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-2, no. 2, pp. 165–168, Feb. 1980.
- [8] V. S. Frost, J. A. Stiles, K. Shanmugan, and J. C. Holtzman, "A model for radar images and its application to adaptive digital filtering of multiplicative noise," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-4, no. 2, pp. 157–166, Feb. 1982.
- [9] X. Zhao, Y. Jiang, and W.-Q. Wang, "Efficient clutter suppression in SAR images with shedding irrelevant patterns," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 9, pp. 1828–1832, Sep. 2015.
- [10] R. K. Raney, "Synthetic aperture imaging radar and moving targets," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-7, no. 3, pp. 499–505, May 1971.
- [11] L. E. Brennan, J. Mallett, and I. S. Reed, "Adaptive arrays in airborne MTI radar," *IEEE Trans. Antennas Propag.*, vol. AP-24, no. 5, pp. 607–615, Sep. 1976.
- [12] B. Li, G. Sun, M. Xing, Y. Hu, L. Guo, and Z. Bao, "Clutter suppression via subspace projection for spaceborne HRWS multichannel SAR system," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 9, pp. 1538–1542, Sep. 2020.
- [13] S. Guo, Q. Zhang, Y. Shao, and W. Chen, "Sea clutter and target detection with deep neural networks," in *Proc. 2nd Int. Conf. Artif. Intell. Eng. Appl. (AIEA)*, Sep. 2017, pp. 316–326.
- [14] H. Mu, Y. Zhang, Y. Jiang, and C. Ding, "CV-GMTINet: GMTI using a deep complex-valued convolutional neural network for multichannel SAR-GMTI system," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2021.
- [15] X. Tang, L. Zhang, and X. Ding, "SAR image despeckling with a multilayer perceptron neural network," *Int. J. Digit. Earth*, vol. 12, no. 3, pp. 1–21, Mar. 2018.
- [16] J. Lehtinen *et al.*, "Noise2Noise: Learning image restoration without clean data," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 2965–2974.
- [17] Y. Yuan, J. Guan, P. Feng, and Y. Wu, "A practical solution for SAR despeckling with adversarial learning generated speckle-to-speckle images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [18] X. Huang, Z. Xu, and J. Ding, "Video SAR image despeckling by unsupervised learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10151–10160, Dec. 2021.
- [19] D. C. Schleher, "Radar detection in Weibull clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-12, no. 6, pp. 736–743, Nov. 1976.
- [20] K.-H. Ding, M. Rangaswamy, and L. Tsang, "Amplitude and phase distributions for bistatic scattering from Pierson-Moskowitz sea surfaces," in *Proc. IEEE Radar Conf.*, May 2008, pp. 1–6.
- [21] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, "Deep clustering for unsupervised learning of visual features," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 139–156.
- [22] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2018, pp. 3–11.