SAR Target Recognition based on Model Transfer and Hinge Loss with Limited Data

Qishan He[1], Lingjun Zhao[1(✉)], Gangyao Kuang[1], and Li Liu[2,3]

[1] State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System,College of Electronics Science and Technology,National University of Defense Technology,Changsha,China
[2] College of System Engineering, National University of Defense Technology, China
[3] CMVS, Univeristy of Oulu, Finland

Abstract. Convolutional neural networks have made great achievements in field of optical image classification during recent years. However, for Synthetic Aperture Radar automatic target recognition(SAR-ATR) tasks,the performance of deep learning networks is always degraded by the insufficient size of SAR images, which cause both severe over-fitting and low-capacity feature extraction model. On the other hand, models with high feature representation ability usually lose anti-overfitting capability to a certain extent, while enhancing the network's robustness leads to degradation in feature extraction capability. To balance above both problems, a network with model transfer using the GAN-WP and non-greedy loss is introduced in this paper. Firstly, inspired by the Support Vector Machine's mechanism, multi-hinge loss is used during training stage. Then, instead of directly training a deep neural network with the insufficient labeled SAR dataset, we pretrain the feature extraction network by an improved GAN, called Wasserstein GAN with gradient penalty and transfer the pre-trained layers to an all-convolutional network based on the fine-tune technique. Furthermore, experimental results on the MSTAR dataset illustrate the effectiveness of the proposed new method, which additional shows the classification accuracy can be improved more largely than other method in the case of sparse training dataset.

Keywords: SAR-ATR · Transfer learning · Generative adversial network.

1 Introduction

Synthetic aperture radar(SAR) is an active imaging radar system, which has the characteristics of a variety of polarization modes and the imaging conditions are not affected by weather conditions. With the development of deep learning in thefield of optical image classification, a large number of classification networks have been applied to SAR image processing. Deep learning in Synthetic aperture radar-Automatic Target Recognition(SAR-ATR) has made great achievements

in SAR image preprocessing due to its automaticity and high accuracy from feature extraction. However, data-driven method is overly dependent on the scale of labeled data, and it will cause serious overfitting due to the scarce SAR dataset and the high cost and difficulty of manual annotation SAR data compared with the optical images.

The mainstream methods to improve the robustness of deep learning models in SAR-ATR can be mainly based on two ideas: 1) Strengthening the feature extraction ability of target network by augmenting dataset, 2) Reforming network structure to improve generalization ability. Chen proposed the all-convolutional networks [1] that substituted all full connection layers by convolutional layers, which greatly reduced the number of trainable parameters and increased the accuracy of MSTAR dataset under SOC conditions to 99% for the first time. Hai combined knowledge distillation with network quantization strategies. This method [2] greatly compressed the parameters of ResNet-18 to a three-layer network and outperformed other method's model with the same parameter's quantity. Zhong used the idea of filter based model pruning and transfer learning, which improved the generalization ability in small network and accelerated the forward propagation process [3]. Although the above methods reduce the model size to prevent from getting caught up in overfitting, training suchs model still demands a certain amount of training examples, moreover methods based on compressing model inevitably degrade network's representational capacity.

Huang [4] proposed a CNN using model transfer learning for the first time. By pretraining feature extraction model in unlabeled SAR scene images and then migrating to SAR target images, the accuracy of MSTAR under SOC and scarce training examples conditions reached to 97%. Zhang [5] showed that generative adversarial network could extract more universal features than autoencoder, by pre-training target network from unlabeled data through info-GAN. Liu [6] used electromagnetic simulation software and 3-D CAD models to generate a large number of SAR vehicle data. Although the above methods based on transfer learning improves the test accuracy using generated simulation data or unlabeled scene images, how to generate robust models relying on existing limited dataset is still a research difficulty. Qin [7] proposed the CAE-HL, which introduced the autoencoder to offset the feature extraction ability deficit with hinge loss. Wanger [8] combined SVM and CNN to obtain a network model with stronger robustness.

To improve both the anti-overfitting and feature extraction ability in the case of scarce training data, a network, called WGAN-HL-Convnet, consisting of an improved GAN, Wasserstein GAN with gradient penalty and multi-classification hinge loss, is proposed in this paper. Firstly, multi-classification hinge loss is introduced into the training stage of network model to adjust the decision of boundary determination. whose optimization of the loss function is similar to the optimization problem under the constraint condition of SVM. When training samples are far away from the boundary margin, the network no longer pays attention to their contribution to the loss, which is the essential difference between the hinge loss and cross-entropy loss. Then, in order to make up for the

degradation of feature extraction ability caused by the loss function, the pre-training model based on WGAN-GP is migrated, and the full connection layer in the discriminator of GAN is replaced with the convolution layer, which greatly reduces the number of model parameters and further improves the generalization ability of the network.

The remainder of this paper is organized as follows. Section2 introduces the implementation of our method. Section3 discusses the performance of the experimental results. Section4 gives the conclusion.

2 Method

2.1 Multi-class Hinge Loss for SAR-ATR

In order to illustrate the applicability of hinge loss to the case of insufficient data, binary classification problem is discussed because of its convenience of feature visualization. Most loss function respond to all training data so as to extract information from existing data as much as possible and improve model's representational ability. While training sample reduces and the probability distribution of training samples and test samples has a certain deviation, as depicted in Fig. 1(a), however, such loss functions fail to obtain an appropriate judgement applicable to test sets, and the cause lies in that all of data characteristics generate a certain loss. In Fig. 1(b), it is showed not all data points are necessary to participate in the formation of decision line, and if only the feature points closed to the hyperplane are focused, the decision line obtained by limited training examples is more likely to be practical in the test samples.

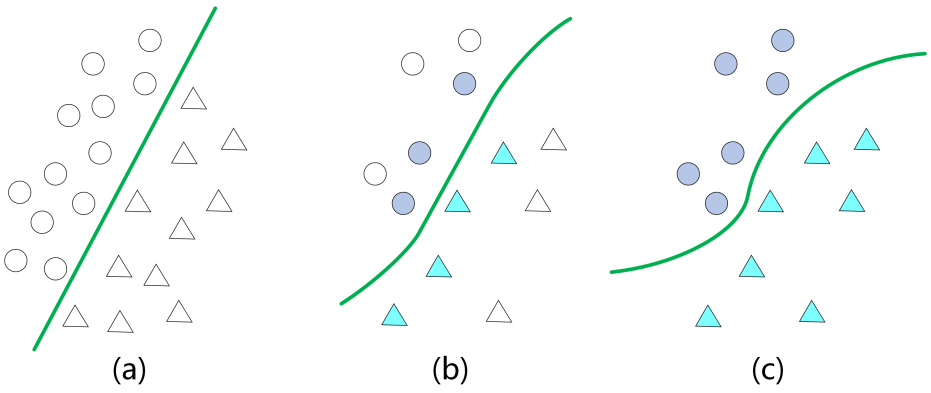

Fig. 1. Different decision lines under sufficient or unsufficient training examples with different loss functions.(a)under sufficient trainset;(b)under insufficient trainset with hinge loss;(c)under insufficient trainset with other loss

Given an input image x_i and its corresponding label $y_i \in \{+1, -1\}$ whose hinge loss can be expressed as $L(x_i, y_i) = \max(0, 1 - y_i f(x_i))$. When $y_i f(x_i) > 1$, the point is judged correctly and far from the judgement plane, so the loss is 0, that means it contributes nothing to the updating of model parameter. Other loss functions, such exponentially loss or cross-entropy loss, remain positive to all data points, as shown in Fig. 2, regardless of whether their predictions are right or wrong. This can lead the decision line in Fig. 1(c) to magnify its deviation degree after minimize the loss of all training data. Therefore, training network with Hinge loss rather than commonly used cross-entropy loss is more likely to improve the generalization ability and robustness in the case of insufficient SAR target images.

For multi-class classification problem, the multi-class hinge loss is used in this paper. Given an input image x_i, the corresponding loss function can be expressed as:

$$L(x_i, y_i) = \sum_{j \neq y_i} \max(0, f(x_i)_j - f(x_i)_{y_i} + \Delta) \tag{1}$$

where $f(x)$ is the output of network, $f(x)_j$ denotes each category score of the output, $j \in \{1, 2, ..., C\}$, and Δ denotes the threshold value.

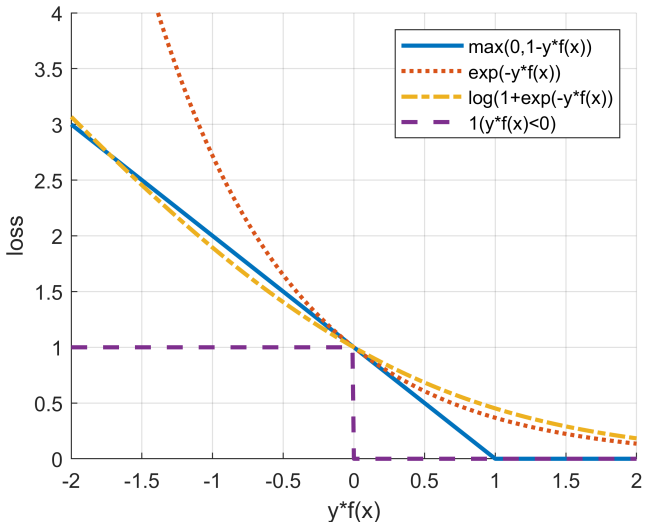

Fig. 2. Hinge loss and other loss functions

2.2 WGAN-HL-Convnet

Overflow of WGAN-HL-Convnet Fig. 3 demonstrates the overall architecture of the method. According to reference [9], although CNN based on hinge

loss can avoid extracting redundant features under the condition of insufficient samples, the number of effective features extracted is much sparser than the CNN based on ordinary cross-entropy loss. To enhance the representability of the target network, the unsupervised generative adversarial network and transfer learning technique are introduced to the proposed method. The classifier layers in the target classification are redesigned to make it capable of the SAR recognition tasks in the case of sparse training data.

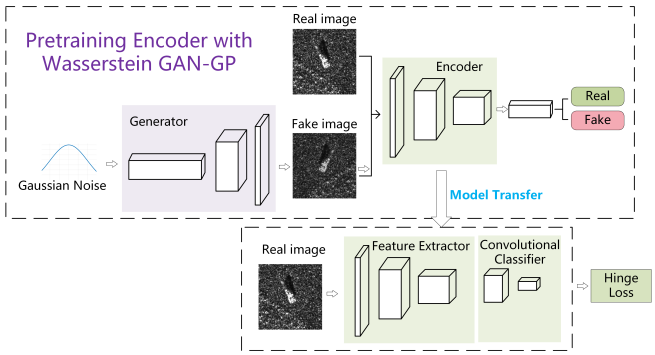

Fig. 3. Processing flow of the proposed method

According to the theory of transfer learning, the training stage can be divided into a pretraining phase and a fine-tuning phase. The typical GAN is known as an unsupervised learning framework to counterfeit images that visually looks like a real image. By means of adversarial training, the discriminator will have the representational ability to distinguish the authenticity of the input images. Meanwhile, the convolutional layers in the discriminator map the original input to the hidden feature space. Therefore, a GAN will be trained to learn universal features from the limited data as far as possible, and then based on the model transfer learning idea, a convolutional classifier will be added to create a SAR recognition model using the hinge loss. Lastly, the whole network will be finetuned to convergence. By combining pretraining technique using GAN and hinge loss, the model extracts a complete feature representation to compensate for the degradation in feature extraction ability, which reduces the over-fitting and owe-fitting risk.

Pretraining Encoder using WGAN with gradient penalty An unsupervised learning method based on Wasserstein GAN is adopted to improve feature extraction capability. Traditional GAN plays a 'minmax' game through alternately optimizing the following adversarial function:

$$\min_{G}\max_{D}\{V(D,G) = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1 - D(G(z))]\} \quad (2)$$

where x denotes the sample from real probability distributionand $p_{data}(x)$ and z denotes the noise vector, of which the elements are produced randomly by a Gaussian distribution.

SAR target images have obvious background speckle noise and irregular scattering light spots, which make it hard for generator to generate visually similar SAR images. Since only a fraction of the background clutter and target contour images can be generated, it is difficult for discriminator to converge to ideal state. What's more, typical GAN usually uses Sigmod function, $f_{sig\ \mathrm{mod}}(x) = \frac{1}{1+e^{-x}}$, as the activation function of the last layer, and the derivative $f'_{sig\ \mathrm{mod}}(x) = f(x)(1-f(x)) \leq \frac{1}{4}$ which causes the gradient disappearing to zero due to the multiplicative effect of the gradient back propagation.

In order to train the discriminator fully, GAN based on Wasserstein distance with gradient penalty is used. Wasserstein distance, namely the bulldozer distance, can maintain smoothness even where there is no overlap between two distributions. To regression the Wasserstein distance in loss function, the sigmod activation function of the last layer in the discriminator is removed. Loss functions for the generator and the discriminator can be expressed as $-E_{z\sim P_z(z)}[D(G(z))]$ and $E_{z\sim p_z(z)}[D(G(z))] - E_{x\sim p_{data}(x)}[D(x)]$. In order to impose the Lipschitz constraint, the gradient penalty term focuses on generating the sample concentration region, and the real sample concentration region and the transition region between them are added to the discriminator's loss function, expressed as follows:

$$L(D) = E_{z\sim p_z(z)}[D(G(z))] - E_{x\sim p_{data}(x)}[D(x)] + \lambda E_{\widehat{x}\sim\widehat{X}}[\left\|\nabla_{\widehat{x}} D(\widehat{x})\right\|_p - 1]^2 \quad (3)$$

where $\widehat{x} = \varepsilon x + (1-\varepsilon)G(z)$, ε is a random variable that obeys 0-1 uniform distribution, and λ is a proportion controlling hyperparameter.

All-Convolutional network for Model transfer An all-convolutional network without fully connected layers is used in the proposed SAR target recognition model. The first five layers are transferred from the encoder of GAN, and the last two layers used to classify extracted features are initialized randomly. All connection layers adopt sparse connection instead of full connection, which effectively reduce the number of free parameters and avoid the severe overfitting due to limited training examples. The forward propagation of each convolutional kernel is express as:

$$O_j(w,h) = \sum_{i=1}^{N} \sum_{u,v=0}^{K-1} W_j^{(l)}(u,v) F_i^{(l)}(w-u, h-v) \quad (4)$$

where $O_j(w,h)$ denotes the output of the jth kernel, $F_i^{(l)}(w-u,h-v)$ refer to the pixel at the position $(w-u,h-v)$ of the ith feature map of the lth layer. And $W_j^{(l)}(u,v)$ is the trainable convolutional kernel with the size of K.

The overall flow architecture of the network is depicted in Fig. 4. For the feature extraction module, each convolutional block is followed by a leakRelu

function to avoid the activation value falling in the interval where the gradient is zero, which is expressed as, $LeakyRELU(x) = \max(0.2x, x)$. The stride of each convolutional operation is set to 2 to compress the size of feature image. The classification module is consist of two convolutional block. The size of the first convolutional block is 4, resulting in feature maps of size 1 × 1. The second one is a point-wise convolution block ensuring the final output size to be 1 × 1 × C, where C is the number of categories.

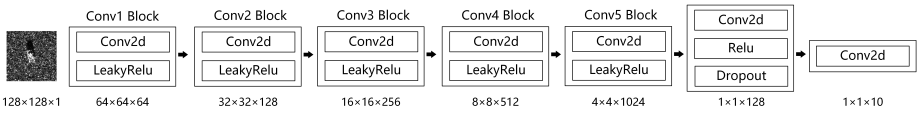

Fig. 4. The architecture of our model

3 Experiments and Results

3.1 Experimental data sets

The accuracy test is performed using the airborne Moving and Stationary Target Acquisition and Recognition(MSTAR) system. The dataset has released publicly 6 different categories of ground targets(armored personnel carrier, tank, rocket launcher, air defense unit, truck, bulldozer). To comprehensively assess the performance, the algorithm is tested both under standard operating conditions(SOC). SOC refers to that the serial numbers and target configurations in the test set are the same with those in the training set, but with different depression angles. The dataset under SOC consists of ten different of car targets, of which the optical and corresponding SAR images are shown in Fig. 5.

Fig. 5. SAR image examples and their correspondingoptical images of ten types of targets in the MSTAR database

3.2 Training details

The Adam optimizer is utilized for training GAN since it could adjust the learning rate dynamically. The batch size is 32, and initial learning rate of discriminator and generator is 0.003. The input noise of the generator is a 128-dimensional random vector of which the elements are produced by as Gaussian distribution . When training the classification, the initial learning rate is set to 0.001. Each sample in the MSTAR dataset is resized to 128×128 and no image augmentation and preprocessing algorithm is applied to the SAR images. All experiments are conducted on a Linux computer with a NVIDIA 3090 GPU card and 32 GB of memory. The used neural network framework is Pytorch.

3.3 Experiment results

Generated images with learned features A common method to verify the feasibility of features extracted from GAN is to see whether the data can be generated from the features derived from real images through discriminator. We randomly enter some real images into encoder to get the feature vectors and then feed the feature vectors into the generator to contrust the fake images at different training stages of WGAN-gp. The generated images are given in Fig. 6(b)-(c). It is suggested that the generated image turn to be identical to the real image increasingly as the training epoches increases, which demonstrates the effectiveness of the feature extraction through the discriminator.

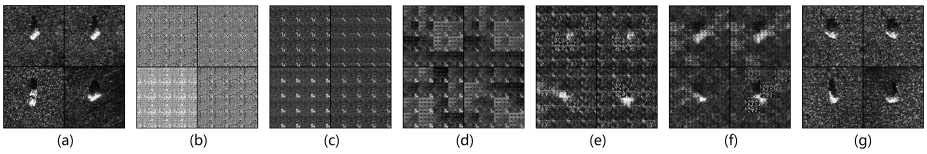

Fig. 6. Generated SAR images with WGAN-gp, where (a) is the real images.(b)-(g) present the fake images at different training epoches

Testing Accuracy This method mainly focuses in SAR target recognition research on limited number of training samples. Therefore, based on the MSTAR SOC dataset, training sets with smaller number of samples are constructed, we contruct four subsets with 200, 500, 1000, 2000 samples have been established by randomly extracted 20, 50, 100, 200 samples respectively from each class of training dataset, and the testset is unchanged. We use subset-200, subset-500, subset-1000, subset-2000 to denote the four training sets

To illustrate the effectiveness of the proposed network under limited number of training samples, three control trails, i.e. baseline-CNN, baseline-Convnet, HL-Convnet, WGAN-HL-Convnet are set up. The specific implementation are

shown as follows: 1) baseline-CNN directily training the random initialized target work of full connection layers in the classification modules with cross-entropy loss function, 2) baseline-Convnet: directily training the random initialized target work of convolutonal layers in the classification modules with cross-entropy function, 3) HL-Convnet: training the baseline-Convnet with hinge loss instead of cross-entropy loss, 4) WGAN-HL-Convnet: training the HL-Convnet which is initialized by the pretrained feature extraction module with WGAN-GP.

Table 1. Test accuracies of the trained models on SOC dataset.

Training set	baseline-CNN	baseline-Convnet	HL-Convnet	WGAN-HL-Convnet
subset-2000	95.63	96.86	96.37	**98.35**
subset-1000	91.22	93.73	94.15	**97.64**
subset-500	81.22	85.77	86.43	**93.64**
subset-200	64.55	65.07	72.20	**86.43**

Table 1 records the recognition accuracy of four networks trained on the four subsets. We repeat every experiment five times and choose the average value to mitigate the impact of fluctuation. The second row shows when the training data is abundant, the test accuracies of baseline-CNN, baseline-Convnet, HL-Convnet, WGAN-HL-Convnet are 95.63%, 95.75%, 94.06%, 97.19%. The performances of four methods are very close to each other, but our method still rank the first accuracy. The accuracy of baseline-Convnet surpasses that of HL-Convnet, which verifies cross-entropy loss accelerates in extracting useful features better than hinge loss under sufficient training data condition. As the number of training samples decreases, our method outperforms the other method significantly, exceeding the second highest accuracy over 5.49%, 7.21% and 14.23% under subset-1000, subset-500, subset-200 conditions respectively. Note that, in such three conditions the HL-Convnet behaves better than baseline-Convnet and our method behave better than HL-Convnet. The experiment demonstrates that in the case of scarce training samples, hinge loss is useful in improving the robustness of the classification network, and pretraining network through WGAN can further greatly improve the performance by enhancing the capability of feature extraction.

Table 2. Detailed accuracies of each categories on subset-2000.

Method	2S1	BMP2	BRDM2	BTR70	BTR60	D7	T62	T72	ZIL131	ZSU234
WGAN-HL-Convnet	98.54	100.00	**92.70**	99.48	98.46	99.27	98.16	98.97	99.27	99.27
HL-Convnet	94.89	**91.28**	95.62	94.89	97.43	98.54	94.87	97.95	97.81	98.90
baseline-Convnet	**93.79**	96.92	**93.06**	97.95	97.43	97.08	94.50	100.00	99.63	99.27

To analyse the classification accuracies in each categories, we enter all the test images into the model in batches by category and calculate the accuracy score

Table 3. Detailed accuracies of each categories on subset-1000.

Method	2S1	BMP2	BRDM2	BTR70	BTR60	D7	T62	T72	ZIL131	ZSU234
WGAN-HL-Convnet	95.25	**92.82**	97.44	97.44	95.89	98.9	98.9	98.46	99.63	99.63
HL-Convnet	98.9	93.33	**84.30**	94.89	92.82	97.44	**84.61**	100.00	97.44	98.90
baseline-Convnet	95.98	96.41	**80.29**	93.36	**90.25**	97.81	**90.84**	94.89	98.17	98.9

on each type. The recall of each type is list in the Table 2-3. For WGAN-HL-Convnet, only the BRDM2 and BMP2 are relatively low, 92.7% and 92.82%, respectively in two tables. Other categories are all more than 95%. But for HL-Convnet, in Table 3, BRDM2, BTR60, T62 show lower scores than other types with accuracies of 84.30%, 84.61%. Similarly, baseline-Convnet has a poor performance accuracies in BRDM2, BTR60, T62 in Table 3. The results shows that the GAN-HL-Convnet not only has higher recognition accuracies than the HL-Convnet and baseline-Convnet in the case of scarce training dataset, but alse performs more balanced in each categories.

4 Conclusion

In the present paper, a method based on the hinge loss and model transfer using Wasserstein GAN is proposed to address the limited label difficulty in SAR-ATR. We transfer the feature extraction module pre-trained from GAN and finetune the whole network with newly added convolutional classification module. The transfer method produces a higher performance on MSTAR dataset than other methods. This superiority becomes more apparent as the training set becomes sparser. Its reveals us that combining hinge loss functions and GAN's pretraining through model transfer is a good way to improve recognition in the case of scarce training samples.

References

1. Chen S, Wang H, Xu F, et al.: Target Classification Using the Deep Convolutional Networks for SAR Images. IEEE Transactions on Geoscience and Remote Sensing **54**(8), 4806–4817 (2016)
2. Lan H, Cui Z, Cao Z, et al: SAR Target Recognition Via Micro Convolutional Neural Network. In: IEEE International Geoscience and Remote Sensing Symposium(2019)
3. Zhong C, Mu X, He X, et al.: SAR Target Image Classification Based on Transfer Learning and Model Compression. IEEE Transactions on Geoscience and Remote Sensing **16**(3), 412–416 (2019)
4. Huang Z, Pan Z, Lei B.: Transfer learning with deep convolutional neural network for SAR target classification with limited labeled data. Remote Sensing **9**(9), (2017)
5. Zhang W, Zhu Y, Fu Q: Deep Transfer Learning Based on Generative Adversarial Networks For SAR Target Recognition with label limitation. In: IEEE International Conference on Signal, Information and Data Processing(2019)

6. Liu L, Pan Z, Qiu X, et al: SAR Target Classification with CycleGAN Transferred Simulated Samples. In: IEEE International Geoscience and Remote Sensing Symposium(2018)
7. Qin R, Fu X, Dong J, et al: A semi-greedy neural network CAE-HL-CNN for SAR target recognition with limited training data. International Journal of Remote Sensing **41**(20), 7889–7911 (2020)
8. Wagner S A: SAR ATR by a combination of convolutional neural network and support vector machines. IEEE Transactions on Aerospace and Electronic Systems **52**(6), 2864–2872 (2016)
9. Zhang W, Zhu Y, Fu Q: Semi-Supervised Deep Transfer Learning-Based on Adversarial Feature Learning for Label Limited SAR Target Recognition. IEEE Access **7**, 152412–152420 (2019)