GlobalSIP
IEEE Signal Processing Society

GAN-NL: UNSUPERVISED REPRESENTATION LEARNING FOR REMOTE SENSING IMAGE CLASSIFICATION

Yiping Duan, Xiaoming Tao, Mai Xu, Chaoyi Han, Jianhua Lu
Tsinghua National Laboratory for Information Science and Technology
Department of Electrical Engineering, Tsinghua University , Beijing, 10084, P.R.China
School of Electronica and information Engineering, Beihang University, Beijing 100191, P. R. China
{yipingduan, taoxm, lhh-dee}@tsinghua.edu.cn, maixu@buaa.edu.cn, hancy16@mails.tsinghua.edu.cn

Abstract

In this paper, we propose a generative adversarial network with non-local spatial information (GAN-NL) for remote sensing image classification. Specifically, a non-local layer is incorporated into a generative adversarial network for unsupervised representation learning. Then, a classification network is designed to infer the labels of the images.

Introduction

Recently, the deep learning methods always need a large number of labeled samples for training and meet a bottleneck when processing the remote sensing images. Just as important the local spatial information, the non-local spatial information represent the self similarity of the image in a global view. Therefore, we incorporate the non-local information into GAN for remote sensing image classification. The proposed approach is shown in Fig. 1.

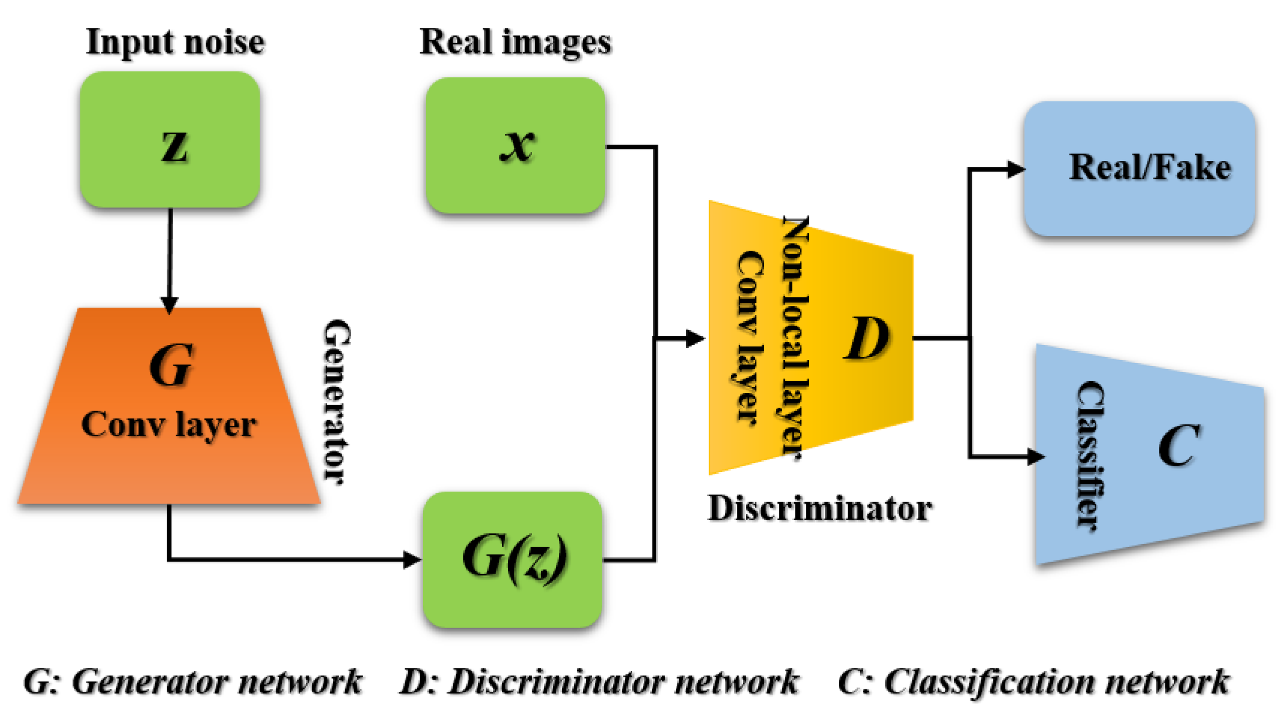

Fig. 1: Overview of the proposed approach, including the generator, discriminator and the classification networks.

GAN-NL for image classification

Non-local layer: The generic non-local layer in deep neural network is defined as,

$$y_i = \frac{1}{Z(x)} \sum_{\forall j} f(x_i, x_j)\, g(x_j)$$

where i and j represent the sites of the output spaces (such as the feature maps). $g(x_j)$ is the representation of the input data. $f(x_i, x_j)$ is the similarity of x_i and x_j in the representation space. $Z(x)$ is the normalized constant and written as

$$Z(x) = \sum_{\forall j} f(x_i, x_j)$$

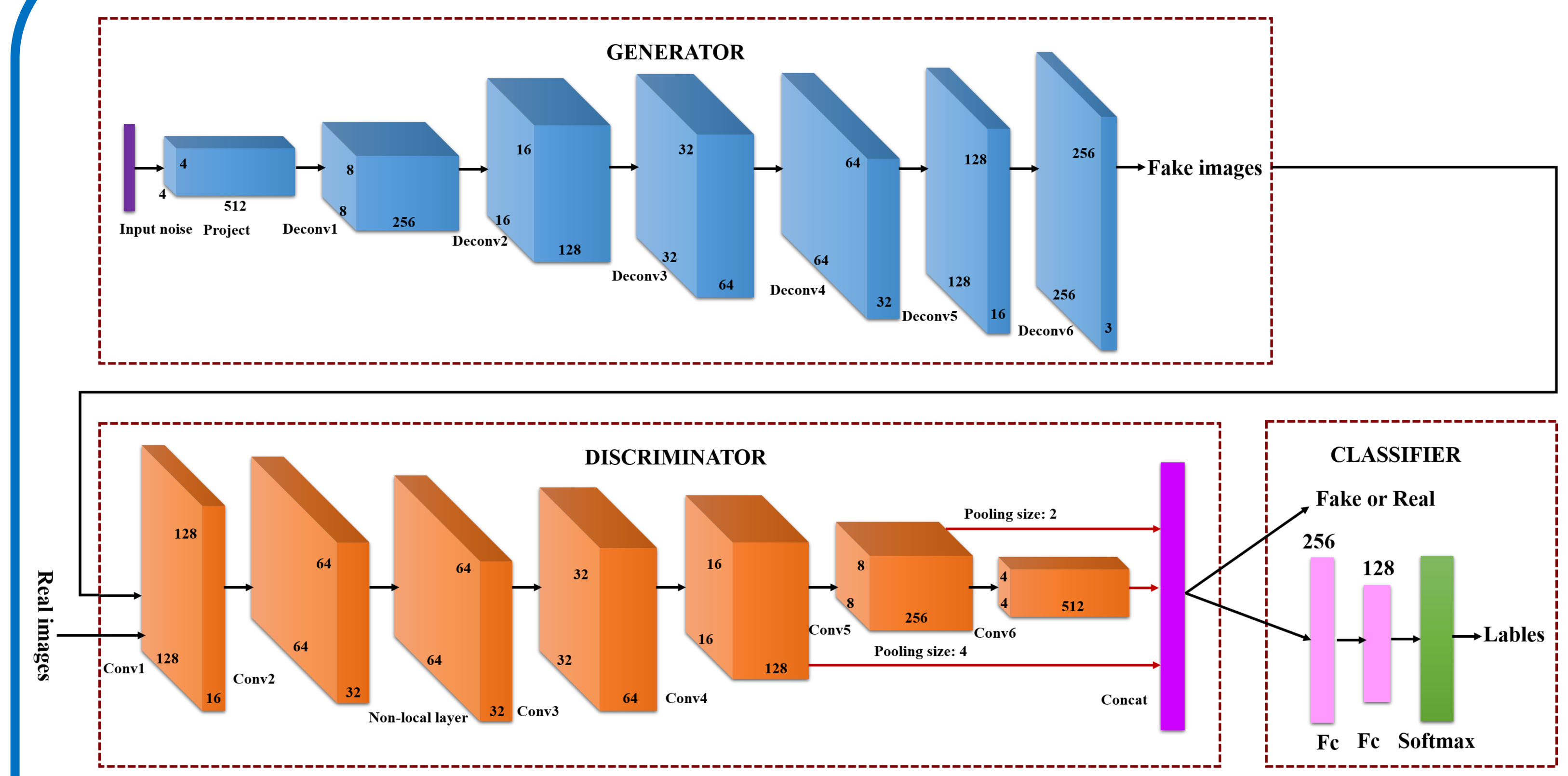

Fig. 2: The network structures of GAN-NL. It includes the generator, discriminator and classifier. The non-local layer is incorporated into the discriminator to improve the representation ability of the model

The similarity $f(x_i, x_j)$ is defined as

$$f(x_i, x_j) = e^{\theta(x_i)^T \phi(x_j)}$$

where $\theta(x_i)$ and $\phi(x_j)$ are two embeddings.

Architectures: The network architectures of GAN-NL are shown in Fig. 2. It includes the generator G, the discriminator D and the classifier C. The generator includes 8 layers. The discriminator has 9 layers. The classification module includes two stages. The first stage is to train GAN-NL and the second stage is to train the classifier.

Discriminator: To train the discriminator, the loss function is to maximize

$$D_{loss} = \mathrm{E}_{\mu\sim pdata(\mu)} log D(x) + \mathrm{E}_{z\sim pz(z)} \left[log\left(1 - D(G(z))\right) \right]$$

where z is the input noise, and $G(z)$ is the output of the generator. x is the real image. $D(G(z))$ and $D(x)$ are the outputs of the discriminator.

Generator: The generator is used to generate the data $G(z)$ as real as possible which can fool the discriminator D. The loss function of the generator is to minimize

$$G_{loss} = loss_{perceptual} + loss_{feature}$$

where $loss_{perceptual}$ is the perceptual loss and $loss_{feature}$ is the feature match loss.

Experimental results

NWPU-RESISC45 dataset is a new and challenging dataset for remote sensing image classification. The classification accuracy of each class is shown in Table I. The highest improvement is 16% in mountain class. Chaparral class achieves the highest accuracy of 0.9853. Moreover, we compare the proposed approach with the latest approaches. The corresponding results are shown in Table II. The parameters of our model is

Table1. The classification accuracy (%) of each class. The different labels represent different scenes. 1: airplane, 2: airport, 3: baseball_diamond, 4: basketball_court, 5: beach, 6: bridge, 7: chaparral, 8: church, 9: circular_farmland, 10: cloud, 11: commercial_area, 12: dense_residential, 13: desert, 14: forest, 15: freeway, 16: golf_course, 17: ground_track_field, 18: harbor, 19: industrial_area, 20: intersection, 21: island, 22: lake, 23: meadow, 24: medium_residential, 25: mobile_home_park, 26: mountain, 27: overpass, 28: palace, 29: parking_lot, 30: railway, 31: railway_station, 32: rectangular_farmland, 33: river, 34: roundabout, 35: runway, 36: sea_ice, 37: ship, 38: snowberg, 39: sparse_residential, 40: stadium, 41: storage_tank, 42: tennis_court, 43: terrace, 44: thermal_power_station, 45: wetland.

Class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
GAN-NL(our)	88.71	63.64	69.62	56.86	90.54	88.52	98.53	70.13	94.20	93.75	62.69	77.14	83.10	94.81	74.60
MARTA-GAN	85.48	53.03	77.22	62.75	93.24	88.52	94.12	59.74	94.20	98.75	70.15	78.57	87.32	90.91	58.73
Class	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
GAN-NL(our)	88.73	68.75	86.75	64.06	88.31	91.38	76.56	94.44	74.29	81.69	84.93	76.32	45.00	83.95	69.57
MARTA-GAN	91.55	59.38	86.75	54.69	89.61	82.76	85.94	94.44	65.71	78.87	68.49	67.11	41.67	88.89	72.46
Class	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**
GAN-NL(our)	73.24	65.38	68.06	73.91	80.00	87.93	75.41	87.50	69.23	66.22	57.14	44.30	82.81	71.21	78.69
MARTA-GAN	77.46	62.82	66.67	65.22	78.67	87,93	70.49	87.50	75.64	70.27	65.71	35.44	73.44	62.12	77.05

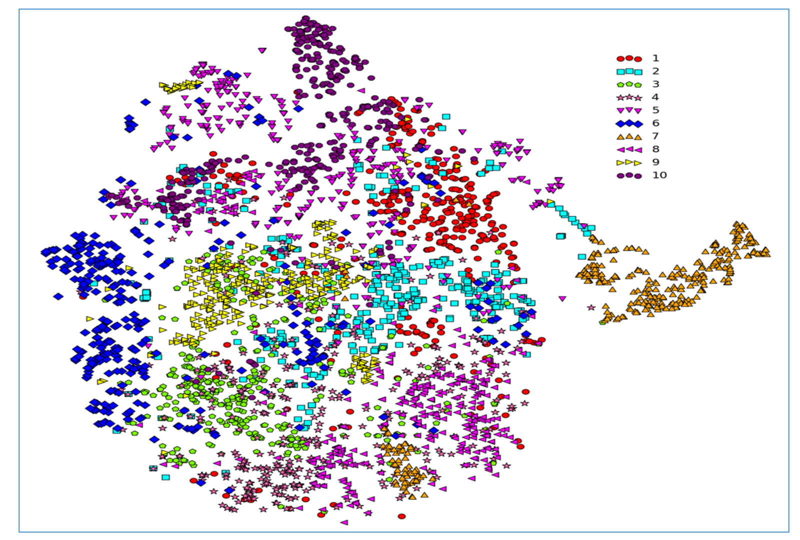

Fig. 3: The 2-D visualization of the generated features learned by GAN-NL in an unsupervised way. This figure shows the features of 10 classes, covering 3500 images.

Table2. The overall accuracy(%) of the reference approaches and our proposed approach.

Method	Learning way	Parameters	Accuracy
LLC [1]	Unsupervised	-	38.81
AlexNet[1]	Supervised	60M	76.69
VGGNet-16[1]	Supervised	138M	76.47
GoogleNet [1]	Supervised	7M	76.19
MARTA-GAN[2]	Unsupervised	2.8M	75.32
GAN-NL(our)	Unsupervised	4.4M	77.10

largely reduced and 4.4M. With the unsupervised way, our approach achieves about 35% improvement compared with the popular LLC (Locality-constrained Linear Coding) method.

We also show the 2D visualization of the features after the dimensionality reduction. Fig. 3 shows the visualization results from class 1 to class 10, covering 3500 images. We can see that these learned features are distinguishable.

Conclusions

This paper combines the generative and discriminative models for remote sensing classification. Moreover, the non-local layer is incorporated into the GAN to improve the ability of the model. The experimental results show that the proposed GAN-NL model improves the classification accuracy without any pre-training.

References

[1] Gong Cheng, Junwei Han, and Xiaoqiang Lu, "Remote sensing image scene classification: Benchmark and state of the art," Proceedings of the IEEE, vol. 105, no. 10, pp. 1865-1883, 2017.
[2] Daoyu Lin, Kun Fu, et al. "Marta gans: Unsupervised representation learning for remote sensing image," IEEE Geoscience & Remote Sensing letter, vol. 14, no. 11, pp.2092-2096, 2016.