

Article

Sample Generation with Self-Attention Generative Adversarial Adaptation Network (SaGAAN) for Hyperspectral Image Classification

Wenzhi Zhao ^{1,2}, Xi Chen ^{1,2,3}, Jiage Chen ^{4,*} and Yang Qu ^{1,2,3}

¹ State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

² Beijing Engineering Research Center for Global Land Remote Sensing Products, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

³ School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454003, China

⁴ National Geomatics Center of China, Beijing 100830, China; wenzhi.zhao@bnu.edu.cn (W.Z.); 211804020027@home.hpu.edu.cn (X.C.); 211804020029@home.hpu.edu.cn (Y.Q.)

* Correspondence: jiagechen@ngcc.cn

Received: 26 January 2020; Accepted: 3 March 2020; Published: 5 March 2020



Abstract: Hyperspectral image analysis plays an important role in agriculture, mineral industry, and for military purposes. However, it is quite challenging when classifying high-dimensional hyperspectral data with few labeled samples. Currently, generative adversarial networks (GANs) have been widely used for sample generation, but it is difficult to acquire high-quality samples with unwanted noises and uncontrolled divergences. To generate high-quality hyperspectral samples, a self-attention generative adversarial adaptation network (SaGAAN) is proposed in this work. It aims to increase the number and quality of training samples to avoid the impact of over-fitting. Compared to the traditional GANs, the proposed method has two contributions: (1) it includes a domain adaptation term to constrain generated samples to be more realistic to the original ones; and (2) it uses the self-attention mechanism to capture the long-range dependencies across the spectral bands and further improve the quality of generated samples. To demonstrate the effectiveness of the proposed SaGAAN, we tested it on two well-known hyperspectral datasets: Pavia University and Indian Pines. The experiment results illustrate that the proposed method can greatly improve the classification accuracy, even with a small number of initial labeled samples.

Keywords: hyperspectral image classification; sample generation; GAN; domain adaptation; self-attention

1. Introduction

With the fast development of remote sensing technology, hyperspectral sensors are now able to capture high spatial resolution images with hundreds of spectral bands, such as those on the recently launched satellites Zhuhai and Gaofen-5. With narrow and contiguous spectral bands, it is now possible to identify land cover targets at high accuracy. Therefore, hyperspectral images have been widely used in crop monitoring, mineral exploration, and urban planning. To achieve such applications, the primary task for hyperspectral data application is image classification. Due to the high-dimensionality of hyperspectral data, it is difficult to find representative features to discriminate between different classes. To explore robust features in the spectral domain, the principal component analysis (PCA), locally linear embedding, and neighborhood-preserving embedding have been widely used for efficient unsupervised feature extraction [1,2]. Meanwhile, the supervised dimension

reduction strategies also have been intensively studied to find discrimination features. For instance, non-parametric weighted feature extraction (NWFE), linear discriminant analysis (LDA), and local discriminant embedding (LDE) are efficient in discriminative feature exploration. In contrast to unsupervised feature learning methods, the supervised dimension reduction strategies can explore class-dependent features that could be used for image classification. However, instead of using features in the spectral domain, the spatial features also play an important role in hyperspectral image classification. Therefore, series of spectral-spatial classification methods were proposed, such as the 3D Gabor filters and the Extended Morphological Profiles (EMPs). Although the elaborated spectral-spatial features have demonstrated their capability in hyperspectral image classification, it is still difficult to capture the most effective features while considering the variety of intra-class data. Instead of relying on hand-crafted image features, deep learning has shown its great power in feature learning and image classification. Deep learning frameworks can generate robust and representative features automatically by using hierarchical structures. Consequently, deep learning-based methods have been widely used for hyperspectral image classification. For example, the deep belief network (DBN) and stacked auto-encoder (SAE) are investigated to extract non-linear invariant features. Complementary, the convolutional neural network (CNN) uses receptive fields to explore effective features from both spectral and spatial domains. To enhance this capability, derivative deep models such as ResNet, VGG, FCN, and U-Net also have successfully applied in hyperspectral image classification. However, deep learning frameworks require a large number of training samples, in order to efficiently classify hyperspectral images.

However, the labeled data are quite scarce in hyperspectral datasets, since the label collection involves expensive and time-consuming field investigation. Thus, labeled data shortage is one of the biggest challenges for the task of hyperspectral image classification. According to the Hughes effect, when the dimensionality of hyperspectral data is high, the limited number of labeled samples will result in low classification accuracy. Besides, deep learning frameworks also face the over-fitting problem when feeding with insufficient training samples. To compensate for the effects of labeled sample shortage, the semi-supervised learning and domain adaptation techniques are developed to increase the number of samples. In particular, the semi-supervised learning considers both the unlabeled samples and labeled ones to be integrated for model training. Furthermore, the domain adaptation aims to transfer existing labeled data to be used in new classification tasks. Meanwhile, some works were devoted to generating high-quality samples based on the standard spectral database by considering the correlations between spectral bands [3,4]. However, due to the impact of the atmosphere, bidirectional reflectance distribution function (BRDF) effect, and even the intra-class variation, it is difficult to generate realistic samples from the standard spectral library by considering such corrupted conditions when referring to physical models, such as the radiative transfer model [5].

Following a different strategy, the generative adversarial networks (GANs) aims to mimic and produce high-quality realistic data to increase the number of training samples. Standard GANs consist of two adversarial modules: a generator that captures the original data distribution and a discriminator that tries to make a discrimination between the generated data and the original ones [6]. To enrich the training samples for hyperspectral image classification, an unsupervised 1D GAN was proposed to capture the spectral distribution [7]. It is trained by feeding unlabeled samples, which are further transformed as a classifier in a semisupervised manner. Thus, the generator cannot learn class-specific features during the training process. To consider the label information, modified GANs have included the label information, such as conditional GAN (CGAN) [8], InfoGAN [9], deep convolutional GAN (DCGAN) [10], auxiliary GAN (AXGAN), and categorical GAN (CatGAN). Consequently, the conditional GANs have been widely used in remote sensing image processing [11–14]. For example, the conditional GAN has been used for data fill in cloud masked area. Meanwhile, for high-resolution remote sensing imagery, a DCGAN-based model was proposed to classify image scenes. In addition, the GAN-based semisupervised model has been utilized for hyperspectral image classification by exploring the information from unlabeled samples [15,16]. However, the training of GANs can easily

collapse due to the contradictory nature of two-player games. To improve the stability of GANs, a triple GAN was proposed to achieve better performances in discriminative ability [17]. However, to generate additional labeled spectral profiles, the current GANs are sensitive to noises and neglect the relationships between spectral bands. Besides, the generated samples are often alienated from the original ones which, inevitably fail in boosting classification results.

To solve the above problems, in this paper, we propose the self-attention generative adversarial adaptation network (SaGAAN) to generate high quality labeled samples in the spectral domain for hyperspectral image classification. In general, two modifications have been made in this framework: the self-attention mechanism is included to formulate long-range dependencies [18] and reduce unintentional noises to stabilize GAN models and the cross-domain loss term is added to increase the similarity between generated samples and the original ones. Therefore, the SaGAAN is able to generate high-quality realistic samples by considering band dependencies and cross-domain loss. Based on the generated samples, better classification results can be achieved.

The rest of this paper is constructed as follows. Section 2 describes the background of relevant studies. Section 3 gives the detailed information about the proposed SaGAAN. Section 4 details the experimental results and comparisons with other methods. Finally, the conclusion is given in Section 5.

2. Related Work

2.1. Generative Adversarial Networks (GAN)

Different from the discriminative models, GAN is one of the representative models in the field of generative modeling. Instead of exploring discriminative features, the generative model aims to estimate the distribution from unknown data p_{data} . In the scope of discriminative modeling, GANs use the framework of the deep neural network to formulate data distribution. Traditionally, a GAN consists of two adversarial players: a generator and a discriminator. The generator aims to generate realistic samples to fool the discriminator while the discriminator also constantly upgrades itself to make a better identification of fake or real samples.

Mathematically, the generator can be represented by G with the parameters θ_G . Similarly, the discriminator is denoted as D with the parameters of θ_D . The standard loss function of GAN models is

$$\min_G \max_D (E_x p_d [\log D(x)] + (E_z p_z [\log(1 - D(G(z)))])) \quad (1)$$

where x is the the sample data from unknown distribution p_{data} and z is the noise space to initialize the generator. GAN has achieved great success in image generation, information restore and data fusion [19–21]. Recently, some improved GANs can perform image classification by adding class-specific terms to the discriminator (e.g., [7,14,15]). However, the power of the generator and additional samples derived from the generator remain unexplored. Therefore, it is necessary to analyze the quality of generated samples and its improvements in hyperspectral image classification.

2.2. Domain Adaptation

Although GAN has the ability to generate realistic samples to enrich the training dataset. It is difficult to stabilize the GAN during its training process, especially for high-dimension data generation. Moreover, the generated samples are often alienated from the original ones (e.g., due to spectral shifts), which fail in high-quality sample generation. To improve the ability of sample generation, the domain adaptation term is proven to be useful. In general, there are several categories in domain adaptation methods, e.g. representation matching, transferable feature selection, and selective sampling [22].

To compensate for the effects of data shifts, domain adaptation aims to make samples transferable across different datasets. Suppose there are two domains called source domain and target domain, which are two data acquisitions of different times or regions. To formulate the classification problem, the joint probability distribution of class labels and its observations from source and target domain are

$J^s(X^s, L)$ and $J^t(X^t, L)$, respectively. X is the input data (e.g., spectral bands) and L is the output label. Domain adaptation methods can transfer the classifier trained on the source domain to predict class labels in the target domain. In this scope, multidimensional histogram matching [23], principle component analysis (PCA) based data alignment, and kernel PCA (KPCA) have been used for domain adaptation [24]. Similarly, the maximum mean discrepancy (MMD) has been used to minimize the sample distances between the source and target domain [25]. Meanwhile, the semisupervised domain adaptation methods have also been intensively studied. For instance, the maximum-likelihood (ML) classifier has been extended with Bayesian rules for the problem of domain adaptation. Based on the assumption of Gaussian distribution, cross-domain information can be effectively captured. To explore domain invariant features, deep learning algorithms such as CNN and GANs also could be used to reduce domain shifts [26,27]. Therefore, domain adaptation is one of the most effective strategies to reduce the discrepancy between two separate domains.

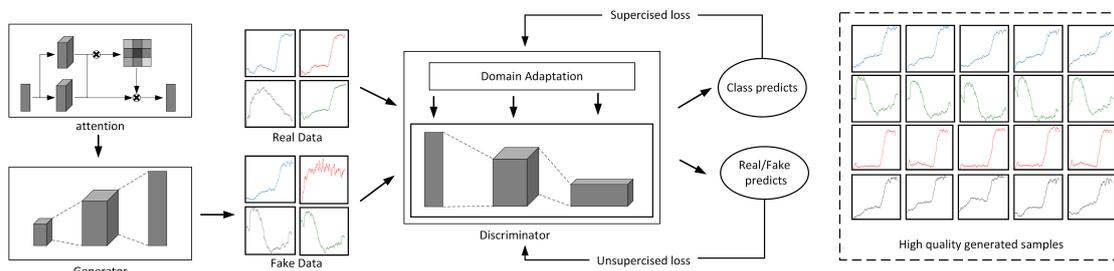


Figure 1. The workflow of Self-attention Generative Adversarial Adaptation Network (SaGAAN).

2.3. Attention Models

The generative models can directly estimate the data distribution from real samples. Compared to nature image generation, hyperspectral samples have an abundant number of spectral bands, which makes it hard for GANs to capture the dependencies between spectral bands. The attention mechanism has the ability to capture the global contextual information and model long-range dependencies. For instance, self-attention [28] calculates the response for a specific position inside a sequence by attending to all positions within the same sequence. Self-attention has proven to be useful in terms of machine translation models [29]. In addition, the combination of self-attention and deep learning algorithms can significantly improve the precision of the image classification, image generation, and spatial-temporal pattern recognition [30,31]. To formulate the conventional self-attention models, we have

$$\alpha_{i,j} = \frac{\exp(Q(x_i)^T K(x_j))}{\sum_{i=1}^N \exp(Q(x_i)^T K(x_j))} \tag{2}$$

where $\alpha_{i,j}$ indicates the response when attending to the location i, j over the entire sequence. The output of self-attention layer is

$$Attention(Q, K, V) = (QK^T)V \tag{3}$$

where $Q = W_f x, K = W_g x$ and $V = W_v x$. $W \in R^{C \times C}$ are the convolution weights with the kernel sizes of 1×1 . The final output of the attention layer has the form of $g_i = \gamma Attention + x_i$.

3. Self-Attention Generative Adversarial Adaptation Network

To improve the stability of the traditional GANs and increase the quality of generated samples, we propose the Self-attention Generative Adversarial Adaptation Network (SaGAAN) for hyperspectral sample generation and classification, as shown in Fig. 1. SaGAAN considers both self-attention and domain adaptation to improve the quality of generated samples. Specifically, it is difficult to stabilize the traditional GANs during the training process. Furthermore, the generated samples are often alienated from the original ones. Therefore, to ensure that the generated samples

are similar to the input original ones, we introduce the domain adaptation technique to constrain the similarity between generated and original samples. Suppose the generated samples are $G(z)$ and the original ones are O . To construct the domain adaptation term, for a N -layer discriminator D , we have

$$L_{domain}(G(z), O) = \sum_{n=1}^N ||D_n(O) - D_n(G(z))|| \quad (4)$$

where $D_n(O)$ represents the deep features from the discriminator by middle layer activation. To better measure the divergence between generated samples and the reference ones, the maximum mean discrepancy (MMD) loss function applied in this study measures the distances between two probability distributions. The MMD attains its minimum zero if the original data and generated samples are equal.

Suppose the original hyperspectral profiles are $o \in O$ with the data distribution P_O to be learned. For SaGAAN, the generator G learns to map a variant z from latent space to the original data space $G(z) \in X$ with the distribution of P_G and conditional label $y \in Y$. Then, the discriminator evaluates the sample whether from the original distribution or generated ones. Different from the minimax loss or hinge loss, the MMD loss uses kernel k to map the discrepancy between two samples. Given two distributions P_G and P_O , the square MMD distance have the following formulation

$$D_k^2(P_G, P_O) = ||\mu_{P_G} - \mu_{P_O}||^2 = \mathbb{E}_{g, g'}(k(g, g')) + \mathbb{E}_{o, o'}(k(o, o')) - 2\mathbb{E}_{o, g}(k(o, g)) \quad (5)$$

where g, g' are two samples from the generator and o, o' are two samples from the original dataset. The kernel $k(o, g)$ measures the similarity between two samples. When the generated samples have a distribution that is equal to the original one P_O , $D_k^2(P_G, P_O)$ is zero.

Instead of using MMD as the loss function for adversarial network optimization, SaGAAN calculates the MMD term for domain adaption. Thus, the discriminator D has the ability to measure the discrepancies between two samples. The objective function for discriminator can be formulated as

$$\max_D L_{D_{ada}} = \mathbb{E}_{g, g'}(k_D(g, g')) + \mathbb{E}_{o, o'}(k_D(o, o')) - 2\mathbb{E}_{o, g}(k_D(o, g)) \quad (6)$$

To maximize the loss function, the discriminator aims to reduces $\mathbb{E}_{o, g}(k_D(o, g))$ that forces generated samples away from the original ones. Meanwhile, the discriminator minimizes the intra-class variance by implementing $\mathbb{E}_{g, g'}$ and $\mathbb{E}_{o, o'}(k_D(o, o'))$. Similarly, the loss function for the generator is

$$\min_G L_{G_{ada}} = \mathbb{E}_{g, g'}(k_D(g, g')) + \mathbb{E}_{o, o'}(k_D(o, o')) - 2\mathbb{E}_{o, g}(k_D(o, g)) \quad (7)$$

The discrepancy between generated samples and the original ones can be reduced by implementing the MMD-based domain adaptation term. However, the generator usually introduces noises from latent distribution and neglected long range dependencies. Therefore, it is still important to consider the band dependencies for hyperspectral sample generation. For SaGAAN, the self-attention mechanism is integrated to improve the quality of generated samples.

$$g_i = \gamma(QK^T)V + x_i \quad (8)$$

In general, the SaGAAN has two improvements compared to the traditional GAN model: the MMD-based domain adaptation for the discriminator and the self-attention mechanism for long-range dependency improvements. The final loss function can be formulated as

$$L_{total}(G, D) = \min_G \max_D (L_{D_{ada}} + L_{G_{ada}}) \quad (9)$$

For SaGAAN, the conditional adversarial network has been adopted for class-specific hyperspectral data generation. Once the loss function is optimized, SaGAAN can produce high-quality class-specific hyperspectral samples. Different from the traditional GAN, SaGAAN can effectively capture the band dependencies over the spectral domain and reduce noises. Moreover, the generated samples are closer to the original ones with the help of domain adaptation and MMD penalization. With the help of generated samples, it is now able to perform hyperspectral imagery classification without much additional training samples.

4. Experiments

4.1. Hyperspectral Datasets

To demonstrate the ability of the proposed SaGAAN, two well-known hyperspectral datasets were included. These two datasets were collected by the Reflective Optics System Imaging Spectrometer (ROSIS) and the AVIRIS sensor, respectively. Due to the high dimensionality of the above datasets and lack of training samples, it is difficult to interpret them efficiently. The detailed description of these two datasets are as follow.

4.1.1. Pavia University Dataset

The Pavia University dataset was acquired by the ROSIS sensor during a flight campaign over Pavia, northern Italy. The sizes of this dataset are 610×340 pixels, with the ground spatial resolution of 1.3 m. There are 103 spectral bands available after removing 12 noisy bands. The spectral bands range from 430 to 860 nm. Nine types of land cover targets were labeled for identification and 10% labeled samples were used for training and another 10% for testing. The pseudo-color composite image and the reference map are shown in Figure 5.

4.1.2. Indian Pines Dataset

The Indian Pines dataset was acquired by the AVIRIS sensor over the Indian Pines test site in northwestern Indiana. The size of the images in this dataset is 145×145 pixels, with high dimensionality in the spectral domain. The sensor system used in this case measured pixel response in 224 bands in the 400–2500 nm region of the visible and infrared spectrum. Due to atmospheric absorption, after removing noisy bands, 200 spectral bands are left for data analysis; 10% labeled samples were used for training and another 10% for testing. The pseudo-color composite image and the reference map are shown in Figure 6.

4.2. Configuration of SaGAN

To serve the purpose of hyperspectral sample generation, the SaGAAN framework is developed based on a 1D generator and a discriminator. To capture the data distribution over spectral bands, SaGAAN converts noises from latent space to realistic spectral profiles. The configuration of SaGAAN is illustrated in Table 1. Compared to the traditional GANs, SaGAAN uniquely pays attention to long-range dependencies and domain adaptation for high-quality sample generation. For the attention term, it is integrated inside of the generator to reduce noises and consider long-range dependencies during sample generation. Meanwhile, to ensure the generated samples are equilibrium to the original ones, the domain adaptation term is added inside of the discriminator. Due to the nature of deconvolution operation, we added an additional band to make sure the number of spectral bands is an odd number. Moreover, to better illustrate the effectiveness of SaGAAN, we included other sample generation approaches (the traditional GAN, self-attention GAN (SAGAN), and adaptation GAN (ADGAN)) for comparison.

Table 1. Detailed configuration of SaGAAN.

Name	Layer	Kernel	Stride	Features	Activation
G	1	1×13	1	128	ReLu
	2	1×2	2	256	ReLu
	3	1×2	1	64	ReLu
	4	1×2	2	1	tanh
	Att ^a	1×1	1	1	Non
D ^{b,c}	1	1×10	2	50	ReLu
	2	1×10	2	100	ReLu
	3	1×10	2	200	tanh
	4	1×3	1	50	tanh
	5	1×4	1	c	Non
	6	1×4	1	1	ReLu

^a Att is self-attention layer; ^b The first four layers of discriminator are used for domain adaptation; ^c The first five layers of discriminator are used for image classification. c represents the number of classes.

4.3. Effect of Domain Adaptation

Domain adaptation is one of the most important factors for high-quality sample generation. Due to the difficulty of adversarial network training, the generated samples are often alienated from the original ones. Therefore, how to reduce the discrepancies between generated samples and the original ones is the major challenge for successful adversarial network training. In SaGAAN, the discriminator contains an additional term to measure the feature distances between generated samples and the original spectral profiles. Specifically, the discriminator D as a 1D convolutional neural network (CNN) has L hidden layers. For each layer, the deep feature can be represented as $D_l(x), l \in L$, and the feature distance between generate samples and the original ones are $D_l(o_i) - D_l(g_i)$. Thus, SaGAAN is able to produce realistic samples based on the similarity measurement. To better illustrate the effect of domain adaptation term for SaGAAN, we developed two separate adversarial networks for hyperspectral sample generation with/without domain adaptation.

For convenience, we tested the domain adaptation term on the Pavia dataset by using the training dataset. Each network generated 0.2 million hyperspectral samples in total, which was about 22 thousand samples for each class. We mapped the generated samples into the lower dimension for better illustration, as shown in Figure 2. We can conclude that the projection map from domain adaptation samples has clear boundaries between different classes. Without domain adaptation, the generated sample often mixed together, which failed to guide supervised classification. Especially, for Classes 8 (Bitumen) and 3 (Self-Blocking Bricks), the inter-class similarity has been significantly reduced. Meanwhile, intra-class variation such as for Classes 1 (Asphalt) and 7 (Bitumen) also has been greatly suppressed. Therefore, domain adaptation is a major improvement in generative adversarial networks since it considers the mismatches between generated samples and the original ones.

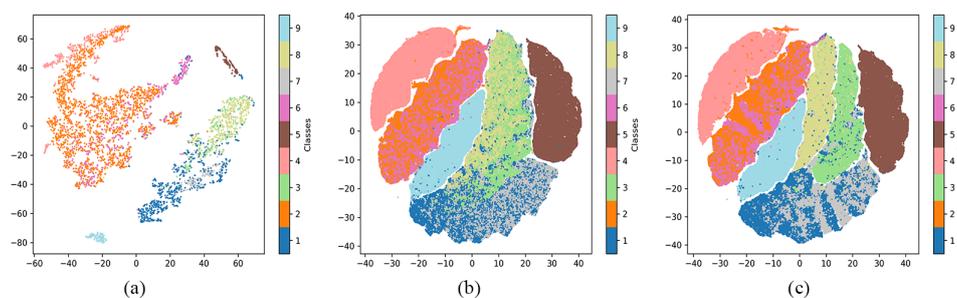


Figure 2. The 2D projection map based on generated samples before/after domain adaptation using SaGAAN: (a) the projection map of training samples; (b) the projection map of generated samples before domain adaptation; and (c) the projection map of generated samples after domain adaptation.

4.4. Effect of Self-Attention

Hyperspectral data contain hundreds of spectral bands that have long-range dependencies (e.g., vegetation has high reflectances in near infrared bands compared to the red band). However, the traditional GANs only pay attention to mimic spectral reflectances at local scales which neglected the relationships across spectral bands. Moreover, traditional GANs introduce noises that also impact high-quality sample generation. Different from domain adaptation, the self-attention mechanism focuses on capturing long-range dependencies between spectral bands. Meanwhile, the self-attention reduces unwanted noises and makes the curves of generated hyperspectral samples smoother.

To demonstrate the effectiveness of the self-attention mechanism, we compared the generated samples by using SaGAAN with/without self-attention constraint. To better understand the impact of self-attention, we chose Classes 5 and 6 in Pavia dataset for hyperspectral data generation. The generated samples are shown in Figure 3. The first two rows represent the spectral curves of painted metal sheets and the last two rows are bare soil reflectances. For the first two rows, we can conclude that much noise has been introduced, which resulted in spikes across the spectral bands, especially for the middle column. In addition, for the bare soil, the generated curves suffer from random noises when not using the self-attention constraint. However, the bare soil spectral profiles become much more similar to the original ones after adding the self-attention term. Moreover, long-range dependencies for spectral curves such as low points and high points have been well represented by the self-attention mechanism.

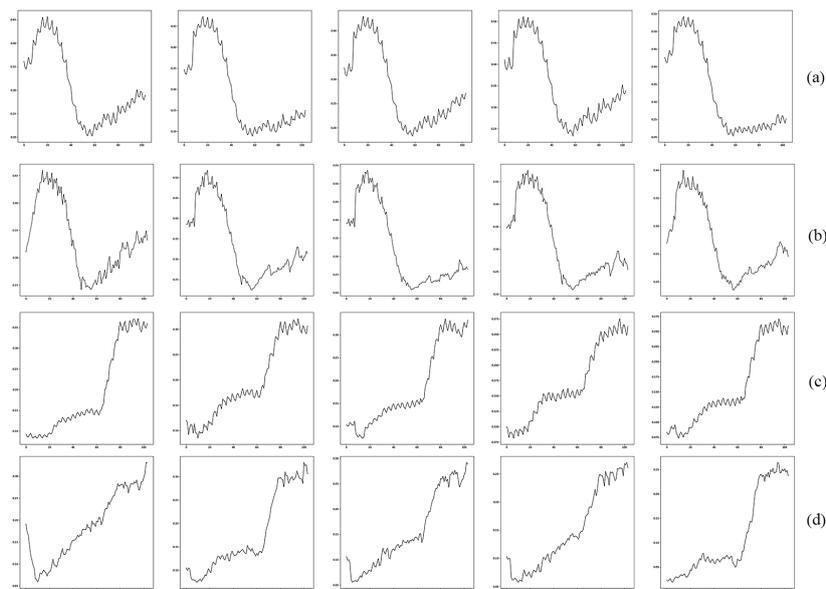


Figure 3. The spectral curves generated by SaGAAN with/without self-attention mechanism: (a) the Class 5 (Painted metal sheets) spectral curves generated without self-attention; (b) the Class 5 spectral curves generated with self-attention; (c) the Class 6 (Bare Soil) spectral curves generated without self-attention; and (d) the Class 6 spectral curves generated with self-attention.

4.5. Generated Sample Analysis

From the above, we can conclude that both domain adaptation and self-attention are crucial parts of high-quality spectral profile generation. In SaGAAN, we utilize MMD measurement to minimize the distances between the generated samples and the original ones. To calculate the MMD distance, the activation of hidden layers inside discriminator convert the generated samples and the original ones into deep feature representations. Then, the similarity of those features can be measured by implementing the MMD strategy. Meanwhile, the self-attention mechanism also enforces the generated samples to be aware of long-range dependencies across different spectral bands. In general, the domain

adaptation and self-attention will strongly stabilize SaGAAN during the training process and prevent potential gradient explosion. To illustrate the effectiveness of combining domain adaptation and self-attention, the loss function values for both generator and discriminator are shown in Figure 4. In this figure, the loss function values for generator and discriminator jitter at the beginning for SaGAAN without domain adaptation and self-attention. Moreover, the loss values can reach almost 6 and then raise again at Iteration 420 where the generator is not stable during the training process. When the domain adaptation and self-attention are involved, the loss values become much more stable through the entire training stage.

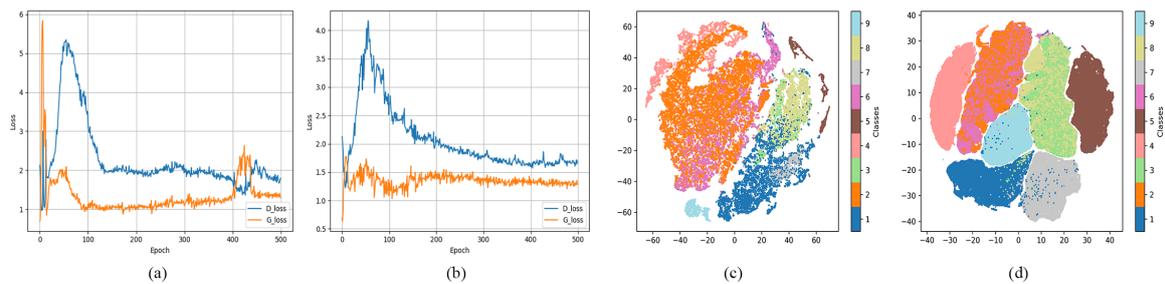


Figure 4. The training losses and sample 2D projection for SaGAAN with/without domain adaptation and self-attention: (a) the training loss values for SaGAAN without domain adaptation and self-attention; (b) the training loss values for SaGAAN with domain adaptation and self-attention; (c) the 2D projection of all available original hyperspectral samples; and (d) the 2D projection of generated hyperspectral samples using SaGAAN.

To measure the quality of generated samples, we mapped all available training samples in Pavia dataset to the two-dimension space, as shown in Figure 4c. The number of training samples is not evenly distributed, where Class 2 (Meadows) represents almost half of the total samples. In addition, samples are scattered in the feature space without significant class boundaries. Complementary, SaGAAN generated high-quality samples based on a small fraction (only 10%) of all available ones. In this experiment, SaGAAN generates 0.2 million samples and each class is equally distributed with 22 thousand samples, as shown in Figure 4d. With the help of domain adaptation and self-attention, SaGAAN generated high-quality samples that contain rich intra-class variation and clear boundaries between different classes. Based on high quality generated samples, better classification results can be achieved.

4.6. Hyperspectral Image Classification and Comparison

To demonstrate the effectiveness of the generated samples, we combined generated samples with the original dataset for the purpose of hyperspectral image classification. Specifically, for each dataset, we selected a specific number of generated samples that have the same sizes as the original training samples. For the purpose of image classification, the 1D CNN framework was applied for hyperspectral image classification. The configuration of 1D CNN is the same as the first five layers of the discriminator illustrated in Table 1. Finally, we tested the classification performances with or without using the additional generated samples.

4.6.1. Pavia University Dataset

In the experiment, we compared the SaGAAN-based hyperspectral image classification method with the three other image classification strategies. Specifically, the original training sample was directly fed into the 1D CNN framework for training and classification. Then, the domain adaptation-based sample generation strategy was applied to generate additional samples. Furthermore, the generated samples along with the original ones were fed into 1D CNN for training and classification. Meanwhile, the self-attention based sample generation also was applied for sample generation and 1D CNN

training. During the entire experiment, each method generated 4273 additional samples for the Pavia dataset, which is as same as the original training dataset.

The classification accuracies are reported in Table 2. For CNN with the original training dataset, the classification accuracy can reach 91.48%. However, due to the training sample shortage, the accuracy is relatively low for Class 7, where it is around 78%. With the help of sample generation strategy, the classification accuracies get higher with additional training samples. For the Ada-CNN, the domain adaptation has been adopted in the traditional GAN framework, and the generated samples along with the original ones were fed into CNN for classification. Therefore, the classification accuracy has increased to 92.08% with domain adaptation samples. Then, the self-attention based samples also increased the overall accuracy about 0.4%. Lastly, the SaGAAN generated high-quality samples were utilized to increase the classification accuracy. The classification maps of these four strategies are shown in Figure 5.

Table 2. Classification accuracies on Pavia University dataset.

Class	CNN	Ada-CNN	Att-CNN	SaGAAN
1	93.86	94.61	94.92	93.90
2	93.49	94.36	95.17	96.13
3	76.44	75.26	74.04	75.35
4	96.74	96.56	96.47	98.63
5	100	100	100	100
6	89.59	92.62	90.53	85.09
7	78.57	84.96	82.52	87.30
8	83.73	79.25	83.51	85.00
9	98.95	100	100	98.94
OA	91.48	92.08	92.48	92.53
AA	89.11	89.32	90.31	90.55
Kappa	88.63	89.44	90.00	90.10

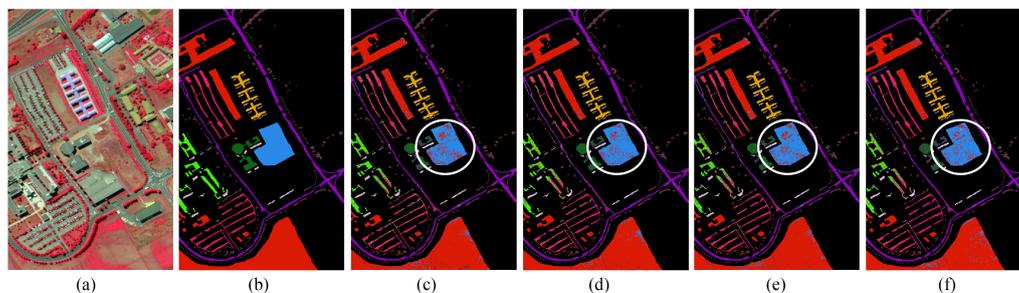


Figure 5. The classification maps of Pavia University dataset: (a) the original dataset; (b) the ground truth labels; (c) classification map with the original dataset; (d) classification map with domain adaptation term; (e) classification map with self-attention mechanism; and (f) classification map with SaGAAN.

4.6.2. Indian Pines Dataset

For the Indian Pines dataset, we tested the SaGAAN with the three other classification strategies. The classification accuracies are illustrated in Table 3. From the results in the table, we concluded that the classification accuracy is quite low when performing the traditional CNN with a limited number of training samples. For Classes 1 and 4, the classification accuracies are 60% and 55.64%, respectively. The overall accuracy is 77.44% when only using the original samples. With the domain adaptation-based sample generation, the overall accuracy has increased to 80.58%, but still faces challenges in Classes 1 and 9 where the number of samples is relatively low. The self-attention mechanism has greatly improved the quality of generated samples; the overall accuracy is about

80.97%. However, the classification accuracies of each class are not in balance. SaGAAN considers both domain adaptation and self-attention mechanisms have significantly improved the quality of generated samples, and it improved the overall classification accuracy to 81.14%. In addition, detailed information about classification maps is shown in Figure 6.

Table 3. Classification accuracies on Indian Pines dataset.

Class	CNN	Ada-CNN	Att-CNN	SaGAAN
1	60.00	57.14	76.92	82.35
2	74.77	72.98	67.47	80.68
3	74.07	82.01	85.46	80.12
4	55.64	65.38	71.88	68.29
5	83.19	91.27	88.09	86.13
6	83.25	86.42	89.61	88.92
7	80.00	68.75	76.47	64.71
8	93.85	93.90	91.76	94.38
9	40.00	50.00	100	62.50
10	75.00	77.14	80.22	78.83
11	72.83	76.09	78.39	74.81
12	68.95	77.74	79.92	77.41
13	89.72	95.24	96.04	91.59
14	87.19	90.64	90.96	90.09
15	73.45	72.66	71.15	69.44
16	97.79	97.78	97.87	100
OA	77.44	80.58	80.97	81.14
AA	72.67	74.47	77.25	78.16
Kappa	74.14	77.72	78.19	78.38

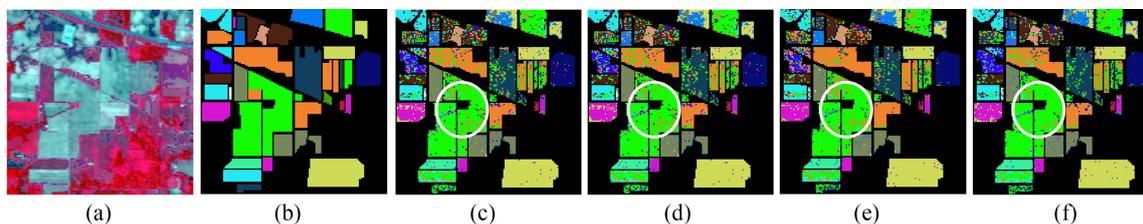


Figure 6. The classification maps of Indian Pines dataset: (a) the original dataset; (b) the ground truth labels; (c) classification map with the original dataset; (d) classification map with domain adaptation term; (e) classification map with self-attention mechanism; and (f) classification map with SaGAAN.

5. Conclusions

In this paper, to generate high-quality hyperspectral samples, we propose a self-attention generative adversarial adaptation network (SaGAAN) to generate realistic samples and improve the classification results of hyperspectral images. Specifically, we include the domain adaptation to increase the similarity between generated samples and the original ones. Meanwhile, to capture the long-range dependencies and reduce unwanted noises, the self-attention mechanism is also integrated with SaGAAN. The experimental results demonstrate that the SaGAAN has the ability to generate high-quality hyperspectral samples and boost the classification accuracy. In the future, we still need to focus on the spatial feature generation, which is also important for hyperspectral image classification.

Author Contributions: W.Z. conceived the idea of SaGAAN for hyperspectral sample generation; X.C. constructed the experimental design and tested on datasets; J.C. instructed us about experimental design; and Y.Q. helped with the writing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Key Research and Development Program of China (Grant No. 2018YFC1508903), the China Postdoctoral Science Foundation (Grant Nos. 2018M640087 and 2019T120063), and the Fundamental Research Funds for the Central Universities (Grant No. 2018NTST01).

References

1. Hang, R.; Liu, Q. Dimensionality reduction of hyperspectral image using spatial regularized local graph discriminant embedding. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 3262–3271.
2. Deng, Y.J.; Li, H.C.; Pan, L.; Shao, L.Y.; Du, Q.; Emery, W.J. Modified tensor locality preserving projection for dimensionality reduction of hyperspectral images. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 277–281.
3. Yu, L.; Xie, J.; Chen, S.; Zhu, L. Generating labeled samples for hyperspectral image classification using correlation of spectral bands. *Front. Comput. Sci.* **2016**, *10*, 292–301.
4. Bhatia, N.; Stein, A.; Reusen, I.; Tolpekin, V.A. An optimization approach to estimate and calibrate column water vapour for hyperspectral airborne data. *Int. J. Remote Sens.* **2018**, *39*, 2480–2505.
5. Cole, I.R.; Betts, T.R.; Gottschalg, R. Solar profiles and spectral modeling for CPV simulations. *IEEE J. Photovoltaics* **2011**, *2*, 62–67.
6. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In *Advances in Neural Information Processing Systems*; 2014, Montreal, Canada; pp. 2672–2680.
7. Zhan, Y.; Hu, D.; Wang, Y.; Yu, X. Semisupervised hyperspectral image classification based on generative adversarial networks. *IEEE Geosci. Remote Sens. Lett.* **2017**, *15*, 212–216.
8. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
9. Chen, X.; Duan, Y.; Houthoofd, R.; Schulman, J.; Sutskever, I.; Abbeel, P. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In *Advances in Neural Information Processing Systems*; 2016, Barcelona, Spain; pp. 2172–2180.
10. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434. .
11. Zhu, L.; Chen, Y.; Ghamisi, P.; Benediktsson, J.A. Generative adversarial networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5046–5063.
12. He, Z.; Liu, H.; Wang, Y.; Hu, J. Generative adversarial networks-based semi-supervised learning for hyperspectral image classification. *Remote Sens.* **2017**, *9*, 1042.
13. Rangnekar, A.; Mokashi, N.; Ientilucci, E.; Kanan, C.; Hoffman, M. Aerial spectral super-resolution using conditional adversarial networks. *arXiv* **2017**, arXiv:1712.08690.
14. Feng, J.; Yu, H.; Wang, L.; Cao, X.; Zhang, X.; Jiao, L. Classification of hyperspectral images based on multiclass spatial-spectral generative adversarial networks. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5329–5343.
15. Zhao, W.; Chen, X.; Bo, Y.; Chen, J. Semisupervised Hyperspectral Image Classification With Cluster-Based Conditional Generative Adversarial Net. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 539–543.
16. Zhang, M.; Gong, M.; Mao, Y.; Li, J.; Wu, Y. Unsupervised feature extraction in hyperspectral images based on wasserstein generative adversarial network. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 2669–2688.
17. Wang, X.; Tan, K.; Du, Q.; Chen, Y.; Du, P. Caps-TripleGAN: GAN-Assisted CapsNet for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 7232–7245.
18. Zhang, H.; Goodfellow, I.; Metaxas, D.; Odena, A. Self-attention generative adversarial networks. *arXiv* **2018**, arXiv:1805.08318.
19. Lin, D.; Fu, K.; Wang, Y.; Xu, G.; Sun, X. MARTA GANs: Unsupervised representation learning for remote sensing image classification. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2092–2096.
20. Singh, P.; Komodakis, N. Cloud-Gan: Cloud Removal for Sentinel-2 Imagery Using a Cyclic Consistent Generative Adversarial Networks. In Proceedings of the IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 1772–1775.
21. Palsson, F.; Sveinsson, J.R.; Ulfarsson, M.O. Single Sensor Image Fusion Using A Deep Convolutional Generative Adversarial Network. In Proceedings of the 2018 9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Amsterdam, The Netherlands, 23–26 September 2018; pp. 1–5.
22. Tuia, D.; Persello, C.; Bruzzone, L. Domain adaptation for the classification of remote sensing data: An overview of recent advances. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 41–57.

23. Inamdar, S.; Bovolo, F.; Bruzzone, L.; Chaudhuri, S. Multidimensional probability density function matching for preprocessing of multitemporal remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1243–1252.
24. Nielsen, A.A.; Canty, M.J. Kernel principal component and maximum autocorrelation factor analyses for change detection. *Image and signal processing for remote sensing XV. Int. Soc. Opt. Photonics* **2009**, 7477, 74770T.
25. Matasci, G.; Longbotham, N.; Pacifici, F.; Kanevski, M.; Tuia, D. Understanding angular effects in VHR imagery and their significance for urban land-cover model portability: A study of two multi-angle in-track image sequences. *ISPRS J. Photogramm. Remote Sens.* **2015**, *107*, 99–111.
26. Bashmal, L.; Bazi, Y.; AlHichri, H.; AlRahhal, M.; Ammour, N.; Alajlan, N. Siamese-gan: Learning invariant representations for aerial vehicle image categorization. *Remote Sens.* **2018**, *10*, 351.
27. Mao, X.; Wang, S.; Zheng, L.; Huang, Q. Semantic invariant cross-domain image generation with generative adversarial networks. *Neurocomputing* **2018**, *293*, 55–63.
28. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems*; 2017, Long Beach, CA; pp. 5998–6008.
29. Vaswani, A.; Bengio, S.; Brevdo, E.; Chollet, F.; Gomez, A.N.; Gouws, S.; Jones, L.; Kaiser, Ł.; Kalchbrenner, N.; Parmar, N.; et al. Tensor2tensor for neural machine translation. *arXiv* **2018**, arXiv:1803.07416.
30. Salazar, J.; Kirchhoff, K.; Huang, Z. Self-attention networks for connectionist temporal classification in speech recognition. In *Proceedings of the ICASSP 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, 17–17 May 2019; pp. 7115–7119.
31. Li, X.; Song, J.; Gao, L.; Liu, X.; Huang, W.; He, X.; Gan, C. Beyond rnns: Positional self-attention with co-attention for video question answering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Honolulu, HI, USA, 27 January 27–1 February 2019; Volume 33, pp. 8658–8665.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).