



Article Hyperspectral Image Classification with Feature-Oriented Adversarial Active Learning

Guangxing Wang 🕩 and Peng Ren *🕩

College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao 266580, China; s18050019@s.upc.edu.cn

* Correspondence: pengren@upc.edu.cn

Received: 26 October 2020; Accepted: 25 November 2020; Published: 26 November 2020



Abstract: Deep learning classifiers exhibit remarkable performance for hyperspectral image classification given sufficient labeled samples but show deficiency in the situation of learning with limited labeled samples. Active learning endows deep learning classifiers with the ability to alleviate this deficiency. However, existing active deep learning methods tend to underestimate the feature variability of hyperspectral images when querying informative unlabeled samples subject to certain acquisition heuristics. A major reason for this bias is that the acquisition heuristics are normally derived based on the output of a deep learning classifier, in which representational power is bounded by the number of labeled training samples at hand. To address this limitation, we developed a feature-oriented adversarial active learning (FAAL) strategy, which exploits the high-level features from one intermediate layer of a deep learning classifier for establishing an acquisition heuristic based on a generative adversarial network (GAN). Specifically, we developed a feature generator for generating fake high-level features and a feature discriminator for discriminating between the real high-level features and the fake ones. Trained with both the real and the fake high-level features, the feature discriminator comprehensively captures the feature variability of hyperspectral images and yields a powerful and generalized discriminative capability. We leverage the well-trained feature discriminator as the acquisition heuristic to measure the informativeness of unlabeled samples. Experimental results validate the effectiveness of both (i) the full FAAL framework and (ii) the adversarially learned acquisition heuristic, for the task of classifying hyperspectral images with limited labeled samples.

Keywords: hyperspectral image classification; active learning; generative adversarial networks

1. Introduction

Hyperspectral imaging is characterized by simultaneously capturing the radiance of the earth's surface at several hundreds of contiguous wavelength bands [1]. Despite the usual coarse spatial resolution, the acquired hyperspectral images record abundant spectral information of imaging areas [2], valuable for various remote sensing applications, such as monitoring and management of environmental changes, agricultural land-use, mineral exploitation, etc. [3–5]. Hyperspectral images are in the form of 3D cubes with two spatial dimensions and a spectral dimension (i.e., number of bands) [1,6]. Spatial pixels in a hyperspectral image correspond to the reflectance of the materials on the earth's surface. Hyperspectral image classification necessitates a classifier that learns to predict the class label of each pixel given a fraction of labeled pixels for training [7–9].

In recent literature, the design of deep learning classifiers [10–13] has been at the forefront of efforts, leading to a dramatic improvement in terms of classification accuracy. Deep learning classifiers directly extract representative features from labeled samples and specifies a parameterized mapping from data space to label space. In essence, the performance of deep learning classifiers relies heavily on

the volume of labeled samples for training [14,15]. A deep learning classifier would achieve remarkable performance for hyperspectral image classification given sufficient labeled samples but normally shows deficiency in the situation of learning with limited labeled samples [16]. Classifying hyperspectral images with limited labeled samples is a major demand as collecting and labeling hyperspectral images are prohibitively labor- and material-consuming compared with doing similar operations for natural images [17–20]. Specifically, collecting hyperspectral images requires deploying specialized imaging spectrometers. In addition, coarse spatial resolution and extensive bands bring about difficulties for labeling. Under the circumstances, it turns to be ill-posed to learn an effective deep learning classifier with limited labeled hyperspectral samples.

Active learning provides a potential means for alleviating this deficiency. Active learning does not change the internal structure of a deep learning classifier but behaves like an efficient labeling strategy. The fundamental principle of the active learning protocol lies in building the training set of labeled samples iteratively [21], by querying informative unlabeled samples and assigning them supplementary labels within multiple loops. Pool-based active learning and active learning by synthesis are two representatives [22]. We concentrate on the pool-based one, which is indeed the scheme employed by almost all the active learning-based hyperspectral image classification methods [17,18,23–28]. Given a pool (i.e., a set) of labeled hyperspectral samples and a pool of unlabeled hyperspectral samples, active learning algorithms strategically query a fixed number (referred to as 'budget' in active learning) of most informative ones from the unlabeled pool [22]. The queried samples are labeled additionally and put into the labeled pool to facilitate model improvement [29].

Which samples are queried depends on certain criteria which are referred to as acquisition heuristics. There is no generic criterion and each sophisticated acquisition heuristic is designed elaborately by active learning practitioners. Existing active learning-based hyperspectral classifiers normally employ off-the-shelf uncertainty-based algorithms [30,31], such as the least confidence [32], the entropy sampling [32], the bayesian active learning disagreement (BALD) [33], etc. Indeed, the above algorithms query unlabeled samples by evaluating the uncertainty based on the output of classifiers, in which representational power is bounded by the number of labeled samples at hand. In this scenario, these active deep learning methods tend to underestimate the feature variability of hyperspectral images spanning from spectral domain to spatial domain.

To address the limitation and comprehensively capture the feature variability of hyperspectral images, we propose a feature-oriented adversarial active learning (FAAL) strategy in this article. We exploit the high-level features obtained from one intermediate layer of a deep learning classifier rather than its output. We employ a deep learning classifier that combines 3D convolutional layers, 2D convolutional layers, and dense layers (i.e., fully connected layers) [11]. The use of 3D convolutional layers accords with the 3D nature of hyperspectral images and facilitates feature extraction [34,35] from both spectral domain and spatial domain. Moreover, we simply divide the classifier into (i) a convolutional module, which learns the high-level features of hyperspectral samples, and (ii) a dense module, which learns to perform predictions. Further, we developed an acquisition heuristic based on adversarial learning with the high-level features. We arrange a generative adversarial network (GAN) in addition to the deep learning classifier for deriving the acquisition heuristic. The GAN comprises two subnetworks: (i) a feature generator that generates fake high-level features, and (ii) a feature discriminator that discriminates between the real high-level features extracted from the convolutional module and the fake high-level features generated by the feature generator. The two subnetworks co-evolve during adversarial learning. Trained with both the real and the fake high-level features, the feature discriminator comprehensively captures the feature variability of hyperspectral images and yields a powerful and generalized discriminative capability. We leverage a well-trained feature discriminator, a purely parameterized yet simple neural network, as the acquisition heuristic to query informative unlabeled samples. The informativeness of them is measured by the estimated probabilities of the well-trained feature discriminator.

The divided deep learning classifier and the feature-oriented GAN form the full feature-oriented adversarial active learning (FAAL) framework. Multiple active learning loops, where both the classifier and the GAN are trained with ever-increasing labeled data, render the classifier robust and generalized for hyperspectral image classification. Overall, the GAN undertakes a pretext task [36] for the mainstream classification task of the classifier.

Our contributions are summarized as follows.

- We develop an active deep learning framework, referred to as feature-oriented adversarial active learning (FAAL), for classifying hyperspectral images with limited labeled samples. The FAAL framework integrates a deep learning classifier with an active learning strategy. This improves the learning ability of the deep learning classifier for classifying hyperspectral images with limited labeled samples.
- To the best of our knowledge, neither the focus on high-level features nor the adversarial learning methodology has been explored for active learning-based hyperspectral image classification. In contrast, the active learning within our FAAL framework is characterized by an acquisition heuristic which is established via high-level feature-oriented adversarial learning. Such exploration enables our FAAL framework to comprehensively capture the feature variability of hyperspectral images and thus yield an effective hyperspectral image classification scheme.
- Our FAAL framework achieves state-of-the-art performance on two public hyperspectral image datasets for classifying hyperspectral images with limited labeled samples. The effectiveness of both the full FAAL framework and the adversarially learned acquisition heuristic is validated by rigorous experimental evaluations.

The rest of this article is structured as follows. Section 2 gives some preliminary knowledge of our work. Section 3 describes our FAAL framework in detail. Section 4 provides experimental evaluations and discussion. Finally, Section 5 concludes this article with several directions for future research.

2. Preliminaries

2.1. Active Learning

Active learning benefits learning with limited labeled samples via building the training set iteratively [21]. We refer readers to Reference [37] for a systematic review. We follow the pool-based scheme. Given a small pool of labeled samples and a large pool of unlabeled samples, active learning algorithms query a fixed number (referred to as 'budget' in active learning) of samples from the unlabeled pool after the model trained on current labeled samples converges [22]. Specifically, they query unlabeled samples that, if labeled, would produce a considerable improvement in classification [38]. A specific acquisition heuristic (i.e., a criterion) based commonly on model uncertainty determines which samples are queried [39]. The queried samples are assigned with supplementary labels and incorporated into the labeled pool. Meanwhile, they are removed from the unlabeled pool. Samples in the newly built labeled pool are used to update the model in the next active learning loop. The process iterates multiple times (referred to as 'loops' in active learning) until a given threshold is approached. The performance of the model is improved progressively as trained with ever-increasing labeled samples.

The investigation of acquisition heuristics lies at the core of an active learning method [33]. Unlike common uncertainty-based algorithms that depend on the output of a classifier, we propose to cope with the intermediate high-level features. This is one means of capturing the feature variability of hyperspectral images within the active learning paradigm.

Generative adversarial networks (GANs) have promoted the progress of adversarial learning in the deep learning era [40–43]. Vanilla GAN has two subnetworks: a generator that maps noise drawn typically from a Gaussian distribution into realistic images, and a discriminator that distinguishes real images from the generated ones. To be precise, the discriminator estimates the probability that an image comes from the training data rather than being generated [40]. Not only does the generative ability of the generator improves, but also the discriminative capability of the discriminator advances during adversarial learning where the two subnetworks are alternately optimized [44]. The potential

promising for applications spanning from computer vision to remote sensing [8,45–48]. Recent literature has seen several semi-supervised learning methods using GANs to augment (generate) unlabeled data [45,47,49,50] developed for hyperspectral image classification. Despite the prevailing use of the generative ability, the discriminative capability is somewhat overlooked in the field of hyperspectral image analysis [46]. In this article, we leverage the discriminative capability of the discriminator to establish a core unit of active learning, i.e., acquisition heuristic. The discriminator behaves as a critical part of our active deep learning framework and forms another means of capturing the feature variability of hyperspectral images within the active learning paradigm.

of either the generative ability of the generator or the discriminative capability of the discriminator is

3. Feature-Oriented Adversarial Active Learning

Our feature-oriented adversarial active learning (FAAL) framework performs hyperspectral image classification in a spatial-spectral manner [51,52]. We commence by conducting dimensionality reduction (e.g., with 30 spectral bands retained) [53,54] to preprocess hyperspectral images. Following, we exert a spatial window (e.g., sized 25) on them and obtain a group of hyperspectral image cubes (e.g., sized $25 \times 25 \times 30$). Our FAAL framework receives as input such hyperspectral image cubes that comprise object pixels (i.e., spectra to be classified) and their spatial neighbor pixels with respect to the hyperspectral images after dimensionality reduction. Without loss of generality, we refer to such a cube as a sample.

Let (x_i, y_i) , i = 1, ..., L be the *i*th sample-label pair in the labeled pool with size *L* and \overline{x}_j , j = 1, ..., U be the *j*th sample in the unlabeled pool with size *U*, separately. We omit the subscripts of notations to make generalizations unless otherwise specified. Active learning-based hyperspectral classifiers delve into querying the most informative hyperspectral samples to be labeled from the unlabeled pool according to a specific acquisition heuristic. Our FAAL framework, composed of a deep learning classifier and a generative adversarial network (GAN), achieves the active query based on adversarial learning with high-level features. The adversarial learning renders the discriminator subnetwork in the GAN increasingly powerful and generalized in discriminative capability, making it an acquisition heuristic in nature. Figure 1 illustrates the adversarial learning with high-level features. Figure 2 shows the feature map changes within the GAN. Figure 3 displays the active query of unlabeled samples outside training. More details are given in the following subsections.



Figure 1. Adversarial learning with high-level features. The process involves two distinctive components: (i) a classifier that is divided into a convolutional module and a dense module, and (ii) a generative adversarial network (GAN) with two subnetworks: a feature generator *G* and a feature discriminator *D*. The convolutional module learns representative high-level features *f* from labeled hyperspectral samples *x*. The dense module transforms *f* into final predictions \hat{y} . The GAN provides adversarial learning with the high-level features. Specifically, *G* maps noise *z* into high-level feature space to generate fake high-level features \tilde{f} . *D* learns to distinguish *f* (treated as 'Real') from \tilde{f} (treated as 'Fake'). Trained with both the real and the fake high-level features, *D* captures the feature variability of hyperspectral images and yields a powerful and generalized discriminative capability. We leverage well-trained *D* as the acquisition heuristic for active learning to measure whether an unlabeled sample is worth querying or not.



Figure 2. Feature map changes of the feature generator *G* (**left**) and the feature discriminator *D* (**right**). *G* starts by a dense layer to expand the input low dimensional (e.g., 100×1) noise to an appropriate size (8192 × 1) to be reshaped to a piece of small feature maps ($4 \times 4 \times 512$). 2D transposed convolutional layers (Transposed Conv2D) are applied to up-scaling feature maps. Among the procedure, we crop the 16×16 feature maps to 15×15 by abandoning the last row and the last column to make them up-sampled smoothly to match the target size ($17 \times 17 \times 64$). Flattening feature maps with the size yields the generated fake high-level features. *D* comprises simply of three dense layers and transforms the input real/fake high-level features into real value probabilities.



Figure 3. Active query of unlabeled samples. Both the convolutional module of the classifier and the well-trained feature discriminator *D* are frozen. We let unlabeled hyperspectral samples \bar{x} input the convolutional module to obtain a pool of high-level features \bar{f} . Following, we put those features into the well-trained *D* and sort them in the light of estimated probabilities. We query the top *K* minimums, with which we trace back to the corresponding hyperspectral samples and label them additionally.

3.1. High-level Features from Classifier Division

A typical deep learning classifier extracts representative features layer by layer. We employ a classifier that combines 3D convolutional layers, 2D convolutional layers, and dense layers (i.e., fully connected layers). It is broadly adapted from the one developed by Roy et al. [11]. The use of 3D convolutional layers accords with the 3D nature of hyperspectral images, which facilitates spatial-spectral feature extraction for the downstream classification task. Assuming that there are *N* labeled samples for training, we use the widely used softmax cross entropy as the cost for the classifier, i.e., \mathcal{L}_{CLS} :

$$\mathcal{L}_{\text{CLS}} = -\sum_{i=1}^{N} y_i \log \operatorname{Softmax}(\hat{y}_i), \tag{1}$$

where y_i and \hat{y}_i denotes the groundtruth and the prediction of the *i*th sample, respectively. Softmax() indicates a softmax operation.

We divide the classifier into two modules by splitting two certain intermediate layers and consider the derived high-level feature space. Specifically, we adopt a simple division strategy that derives (i) a convolutional module including all 3D and 2D convolutional operations in the head of the classifier, and (ii) a dense module stacked by dense layers in the tail, as shown in the upper half of Figure 1. For simplicity, we use Conv3D, Conv2D, and Dense to represent 3D convolutional layers, 2D convolutional layers, and dense layers, respectively. Given the input labeled samples x, the convolutional module learns to extract representative high-level features f. The dense module transforms f into final predictions \hat{y} via multiple layer-wise non-linear transformations.

With such division, the mapping from data space to label space is interposed by high-level feature space and thus can be considered as a sequential combination of two mappings. The convolutional module accounts for mapping from data space to high-level feature space, and the dense module concludes the mapping from high-level feature space to label space. We implement an active query by coping with the high-level features in high-level feature space rather than the output of the classifier, i.e., data points in label space. The representational power of the classifier is bounded by the number of labeled samples at hand, making the neatly formed classifier output hardly reflects the feature abundance of hyperspectral images. Exploring the intermediate high-level features alleviates this bias to a certain extent. Besides, we leverage additional fake high-level features generated by a GAN to help capture the feature variability of hyperspectral images during the active query further, which will be introduced in the next subsection.

3.2. Adversarial Learning with High-Level Features

We train a GAN independent of the classifier, as shown in the lower half of Figure 1. The GAN is composed of a feature generator *G* and a feature discriminator *D*. Specifically, *G* maps noise *z* into high-level feature space to generate fake high-level features \tilde{f} . *D* treats f = E(x) (For unity, here, we use *E* to denote the convolutional module of classifier), the high-level features extracted from the convolutional module of the classifier as real and $\tilde{f} = G(z)$, those generated from noise as fake,

and learns to discriminate between them. To be precise, *D* receives high-level features (f or \tilde{f}) as input and estimates the probabilities that they are real. The output probabilities are real values between 0 and 1.

The configuration of neither *G* or *D* is elaborately designed. Figure 2 illustrates the feature map changes within *G* and *D* given a feature size of 18496×1 , separately. *G* begins with a dense layer to expand the input low dimensional (e.g., 100×1) noise *z* to a size (e.g., 8192×1) ready for being reshaped to a piece of small feature maps (e.g., $4 \times 4 \times 512$). We use 2D transposed convolutional layers (Transposed Conv2D) to up-scale feature maps. Particularly, we crop the 16×16 feature maps to 15×15 by abandoning the last row and the last column to make them up-sampled smoothly to match the target size (i.e., $17 \times 17 \times 64$). Flattening feature maps with that size yields the generated fake high-level features. To build *D*, we simply use three dense layers that transform the input real/fake high-level features into real value probabilities.

G and *D* co-evolve during adversarial learning where they are trained adversarially and alternately. *D* is commonly trained prior to *G*. The cost for the feature discriminator \mathcal{L}_D takes the form of standard binary cross entropy:

$$\mathcal{L}_{\mathrm{D}} = -\frac{1}{2} \mathbb{E}_{x \sim p(x)} \log D(E(x)) -\frac{1}{2} \mathbb{E}_{z \sim p(z)} \log(1 - D(G(z))),$$
(2)

where p(x) and p(z) are the empirical distribution of current labeled samples and an easy-to-sample prior distribution (e.g., Gaussian distribution) of noise, respectively. \mathbb{E} indicates an expectation operation.

To ensure that both *G* and *D* have strong gradients during training [55], the cost for the feature generator \mathcal{L}_{G} holds the form of cross entropy but changes into:

$$\mathcal{L}_{G} = -\frac{1}{2} \mathbb{E}_{z \sim p(z)} \log D(G(z)), \tag{3}$$

where p(z) is the prior distribution of noise.

In addition, we follow the disparity measurement [56] to facilitate a diverse feature synthesis, and extend \mathcal{L}_{G} with a regularization term \mathcal{L}_{REG} :

$$\mathcal{L}_{\text{REG}} = -\frac{|G(z_1) - G(z_2)|}{|z_1 - z_2|}.$$
(4)

Minimizing \mathcal{L}_{REG} explicitly maximizes the ratio of the distance between two generated features $G(z_1)$ and $G(z_2)$ with respect to that between two corresponding noise vectors z_1 and z_2 .

Trained with both the real and the fake high-level features, *D* captures the feature variability of hyperspectral images and yields a powerful and generalized discriminative capability. It estimates the probabilities that the input features are real rather than fake (generated). We freeze the training of a well-trained *D* and take its estimated probabilities as a criterion to measure whether an unlabeled sample is previously well-represented or not. The well-trained *D* is an acquisition heuristic for active learning in nature and is a purely parameterized yet simple neural network. In general, samples not previously well-represented yield high uncertainty [57]. The adversarially learned acquisition heuristic does not extricate from being uncertainty-based but does not explicitly measure uncertainty, either. Such implicit measurement by means of feature discrimination enhances capturing the feature variability of hyperspectral images. Besides, the adversarially learned acquisition heuristic is task-agnostic, so that we believe it scalable to other applications. Research on the property is beyond the scope of this article.

3.3. Active Query of Unlabeled Samples

The above adversarial learning provides an acquisition heuristic for the subsequent active query. Overall, the GAN undertakes a pretext task [36] for the mainstream classification task. During the adversarial learning, D learns to output high probability values when receiving real high-level features as input and output low probability values given fake high-level features. When a frozen well-trained D generalizes to new input, it is intuitively sound that D would output low probability values if the new input is not well-represented by the pool of f, and vice versa.

Figure 3 illustrates the active query of unlabeled samples. Our acquisition heuristic performs an active query in the form of high-level features rather than computing using hyperspectral samples throughout. We let hyperspectral samples \overline{x} in the unlabeled pool input the frozen trained convolutional module of the classifier to obtain a pool of high-level features \overline{f} . We put those features into the frozen trained *D* as the aforementioned new input. Assume the budget of each active learning loop is *K*. We sort the output estimated probabilities by value and query the top *K* minimums. The output of the active query is a series of indices, with which we can easily trace back to the corresponding unlabeled samples. After being labeled, those queried samples are merged with the labeled samples at hand and removed from the unlabeled pool.

The GAN provides adversarial learning with high-level features, yielding a purely parameterized yet simple acquisition heuristic for actively querying unlabeled samples to be labeled. The feature variability of hyperspectral images is taken into consideration. Multiple active learning loops progressively improve the performance of the classifier for classifying hyperspectral images with limited labeled samples. We refer to the full strategy/framework as feature-oriented adversarial active learning (FAAL).

3.4. Workflow of Full Framework

The two distinctive components involved in our FAAL framework, i.e., a deep learning classifier and a GAN, are trained alternatively and separately. The classifier is trained prior to the GAN. Training one of them necessitates an ad hoc freezing of the training of the other.

The training process has two stages overall. Firstly, the two components are trained with initial labeled samples. Secondly, they are trained with ever-increasing labeled samples (or high-level features) within multiple active learning loops. Specifically, the GAN is unnecessary to be trained in the last active learning loop. Algorithm 1 summarizes the workflow of our FAAL framework.

Algorithm 1 Feature-oriented adversarial active learning.

1: repeat

2: Update classifier initially:

Minimize Equation (1) with initial labeled samples.

- 3: Update GAN initially:
 - a. Freeze classifier and obtain real high-level features *f* of current labeled samples.
 - b. Generate fake high-level features \tilde{f} from noise *z*.
 - c. Update *D* in terms of minimizing Equation (2).
 - d. Generate two groups of fake high-level features \tilde{f}_1 and \tilde{f}_2 from noise z_1 and z_1 , respectively.
 - e. Update *G* in terms of minimizing Equations (3) and (4).
- 4: **for** loop = 1, ..., 4 **do**

5: Active query of unlabeled samples:

- a. Freeze D.
- b. Query K unlabeled high-level features with minimum estimated probabilities.
- c. Trace back to unlabeled samples and label them.
- d. Merge newly labeled samples with previous ones.

e. Remove the queried samples from the unlabeled pool.

- 6: Update classifier using current labeled samples.
- 7: Update GAN using the high-level features of the current labeled samples.
- 8: end for
- 9: **for** loop = 5 **do**
- 10: Active query of unlabeled samples.
- 11: Update classifier using current labeled samples.
- 12: end for
- 13: **until** reaching the given threshold.

4. Experimental Results and Discussion

We rigorously evaluated our FAAL framework for the task of classifying hyperspectral images with limited labeled samples. We avoided using any data augmentation method. We adopted two public hyperspectral image datasets and organize three groups of experimental comparisons on them. All the quantitative comparisons were assessed using three common evaluation metrics: overall accuracy (AA), average accuracy (AA), and kappa coefficient (KAPPA). Larger values indicated better performance. All of the reported results were averaged over ten runs. In each run, initial labeled samples were randomly sampled without fixing random seeds. In all quantitative comparisons, we marked the best in bold.

4.1. Datasets

We adopted two public hyperspectral image datasets (the two datasets are available at http: //www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes), i.e., Indian Pines and Pavia University. The Indian Pines scene was acquired by the Airborne Visible Infrared Imaging Spectrometer (AVIRS) sensor, and the Pavia University scene was collected by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor. The ground sites of the two scenes are the Indian Pines test site, USA, and the Pavia University, Italy, separately.

Table 1 lists more basic information of the two datasets, including sensor, size, number of available bands, spectral range, ground sample distance (GSD, i.e., spatial resolution), and number of labeled classes. By comparison, the Indian Pines dataset exhibits relatively heavy class imbalance and the

Pavia University dataset has less labeled classes. Class information and respective labeled spectrum numbers of the two datasets are given in Tables 2 and 3, separately.

Dataset	Sensor	Size	No. Available Bands	Range	GSD	No. Classes
Indian Pines	AVIRS	145 imes 145	200	0.4–2.5 μm	20 m	16
Pavia University	ROSIS	610 imes 340	103	0.43–0.85 μm	1.3 m	9

Table 1. Basic information of Indian Pines and Pavia University.

щ	Color	Indian Pines					
#		Class	No. Labeled Pixels				
1		Alfalfa	46				
2		Corn-Notill	1428				
3		Corn-Mintill	830				
4		Corn	237				
5		Grass-Pasture	483				
6		Grass-Trees	730				
7		Grass-Pasture-Mowed	28				
8		Hay-Windrowed	478				
9		Oats	20				
10		Soybean-Notill	972				
11		Soybean-Mintill	2455				
12		Soybean-Clean	593				
13		Wheat	205				
14		Woods	1265				
15		Bldg-Grass-Trees-Drives	386				
16		Stones-Steel-Towers	93				
Total			10,249				

Table 2. Class information and labeled spectrum numbers of Indian Pines.

Table 3. Class information and labeled spectrum numbers of Pavia University.

#	Color -	Pavia University				
		Class	No. Labeled Pixels			
1		Asphalt	6631			
2		Meadows	18,649			
3		Gravel	2099			
4		Trees	3064			
5		Painted-metal-sheets	1345			
6		Bare-Soil	5029			
7		Bitumen	1330			
8		Self-Blocking-Bricks	3682			
9		Shadows	947			
Total			42,776			

4.2. Implementation Details

We implement our FAAL framework using Python in conjunction with the TensorFlow library. Our experimental environment contains 512 GB random access memory (RAM) and NVIDIA Tesla K80 graphic processing unit (GPU) computing accelerators (11GB memory). We retain 30 bands for each dataset after the dimensionality reduction. The spatial window size for spatial-spectral classification is 25. We set the dimensionality of noise to 100. The learning rates for the classifier and the GAN are 0.001 and 0.0002, respectively. Specifically, we set a minor decay rate of 0.000006 for

the training of the classifier. The number of active learning loops is five. We initialize the labeled pool by randomly selecting five samples per class. We set the unlabeled pool by randomly sampling another 1000 samples, leaving the rest for testing. By default, the budgets are set as 34 for Indian Pines, and 41 for Pavia University. After all the active learning loops finish, there are 250 samples in the labeled pool for each of the two datasets. The training epochs are 45 for both initial training and active training. The computational training time is about eleven minutes. Code is available at https://github.com/gxwangupc/FAAL.

4.3. Analysis of the Naive Classifier

We start by analyzing our employed deep learning classifier (i.e., the basic classifier of our FAAL framework), which is configured with 3D convolutional layers, 2D convolutional layers, and dense layers. The configuration is generally adapted from Roy et al. [11].

We test two dimensionality reduction strategies. One is the traditional principal component analysis (PCA) [53]. The other is the superpixel-wise principal component analysis (SuperPCA) [54] that performs PCA on segmented homogenous regions [58,59]. To build a training set for the naive classifier, we randomly sample five samples per class and then randomly choose additional disjointed 170 samples for Indian Pines, and 205 for Pavia University. In total, there are 250 labeled samples for each dataset.

The naive classifier receives all the 250 labeled samples at once. However, these samples are selected in an absolutely random manner regardless of informativeness. An active classifier (naive classifier + active learning) takes into account informativeness during the process of iteratively adding queried samples. Its training starts with an extremely small number of labeled samples and runs multiple times, i.e., loops, during which samples are queried unrestricted by class. The model trained in the current loop can be thought of as a pre-trained model for resuming training in the next loop.

Table 4 assesses the naive classifier and our FAAL framework on the two datasets with the above two dimensionality reduction strategies, in terms of the OA, AA, and KAPPA. Results show that reducing dimensionality using SuperPCA delivers better performance than that using PCA. A sophisticated dimensionality reduction strategy plays a significant role in the downstream task. Hence, we preprocess hyperspectral data just using SuperPCA in the following evaluations. We observe that the naive classifier obtains plausible results in a mass. It suggests that the configuration of our employed deep learning classifier is considerably effective. A major source of the capability may be that the use of 3D convolutional layers accords with the 3D nature of hyperspectral images. Overall, our FAAL framework yields better performance than the naive classifier under the same dimensionality strategy. It highlights that an active learning strategy surpasses the naive setting, i.e., randomly selecting samples and training at once, in classifying hyperspectral images with limited labeled samples. Besides, this validates that the informativeness of samples carries a big weight in the scenario of learning with limited labeled samples.

Dataset		Indian Pines Pavia U			ivia Unive	rsity
Method/Metric	OA (%)	AA (%)	k (×100)	OA (%)	AA (%)	k (×100)
Classifier (PCA)	81.15	81.30	81.59	87.48	75.91	83.23
Classifier (SuperPCA)	87.39	89.16	85.60	88.93	82.14	85.23
FAAL (PCA)	84.71	88.10	82.54	92.39	85.51	89.83
FAAL (SuperPCA)	91.41	93.20	90.20	93.47	88.97	91.34

Table 4. Comparison of the naive classifier and feature-oriented adversarial active learning (FAAL) with two different dimensionality reduction strategies.

4.4. Comparison with Other Active Learning Classifiers

We selected three state-of-the-art methods for comparison. Each of them performs a spatial-spectral hyperspectral image classification based on active learning. Zhang et al. [26] combine active learning with a hierarchical segmentation responsible for extracting spatial information. Zhang et al. [25] incorporate active learning with an adaptive multi-view generation and an ensemble strategy. Cao et al. [17] integrate active learning and a convolutional neural network, followed by a Markov random field for finetuning.

AL-SV-HSeg (a single-view active learning framework with hierarchical segmentation) in Zhang et al. [26], AL-SV (a single-view active learning framework), AL-MV (a multi-view active learning framework), AL-MV-HSeg (a multi-view active learning framework with a hierarchical segmentation), and AL-MVE-HSeg (a multi-view active learning ensemble framework with a hierarchical segmentation) in Zhang et al. [25], and AL-CNN-MRF (an active deep learning framework with a Markov random field) in Cao et al. [17], are included as baselines. We compare our FAAL framework with their reported results directly. We use the same number of labeled samples as Zhang et al. [25,26]: five samples per class initially and 250 samples totally in the end.

Table 5 reports the comparison results on Indian Pines, in terms of the OA, AA, KAPPA, and accuracy of each class. In the case of using 250 labeled samples in total, our FAAL framework achieves the best performance measured by all the three general metrics. Specifically, FAAL with 250 labeled samples for training obtains higher AA compared with AL-CNN-MRF trained with 416 labeled samples in total. When trained with 300 labeled samples, our FAAL framework consolidates the gains and surpasses AL-CNN-MRF in terms of OA.

Table 6 gives the comparison results on Pavia University. Our FAAL framework trained with 250 labeled samples achieves higher OA and *k* than AL-SV-HSeg but lower AA. In comparison with the reported results of AL-CNN-MRF trained with 321 labeled samples in total, we change to initialize the labeled pool by randomly selecting ten samples per class and query 46 unlabeled samples in each active learning loop. 320 labeled samples are used in total. Despite the lower OA than AL-CNN-MRF, our FAAL framework exhibits effectiveness in terms of AA and surpasses AL-CNN-MRF.

Metric/Method	AL-SV	AL-SV-HSeg	AL-MV	AL-MV-HSeg	AL-MVE-HSeg	FAAL (250)	AL-CNN-MRF (416)	FAAL (300)
OA (%)	61.96	81.64	51.19	80.00	87.10	91.41	92.26	93.91
AA (%)	62.16	82.52	48.64	83.14	90.13	93.20	86.54	95.16
$k (\times 100)$	56.30	79.11	43.67	77.07	85.34	90.20	-	93.07
1	51.22	76.77	36.71	56.18	64.59	99.29	84.17	100
2	33.82	72.94	19.12	63.01	84.09	82.35	91.00	88.16
3	43.21	66.82	28.62	74.83	84.95	89.81	83.64	98.52
4	49.94	59.24	13.08	60.83	82.20	89.88	87.01	91.78
5	71.93	82.27	66.11	88.67	92.08	87.59	91.57	90.39
6	31.99	75.26	41.49	72.81	79.00	92.14	95.18	94.62
7	41.26	93.30	20.38	92.94	94.58	100	89.13	100
8	59.80	86.59	59.88	83.62	92.98	100	98.93	100
9	79.38	94.45	71.55	92.36	89.06	100	16.08	100
10	86.00	96.94	41.37	97.27	97.78	85.94	90.68	91.94
11	87.21	99.27	92.43	99.18	99.71	97.44	94.70	96.43
12	77.25	95.56	35.58	75.94	94.36	78.91	91.51	83.24
13	82.24	98.55	48.71	98.14	99.56	98.88	99.25	98.90
14	85.49	91.13	63.54	96.21	98.63	95.20	95.73	96.33
15	29.75	88.99	59.41	90.63	95.65	98.38	83.99	94.86
16	84.05	89.74	80.22	87.56	92.87	95.37	92.08	97.33

Table 5. Comparison of active classifiers and FAAL on Indian Pines.

Metric/Method	AL-SV-HSeg	FAAL(250)	AL-CNN-MRF (321)	FAAL(320)
OA (%)	92.23	93.47	97.43	97.14
AA (%)	92.66	88.97	94.80	95.07
$k \; (imes 100)$	90.05	91.34	-	96.20
1	90.00	92.01	98.18	95.29
2	93.59	97.36	99.82	99.96
3	86.21	97.93	78.46	99.01
4	92.65	70.62	93.86	84.89
5	97.62	77.32	99.05	94.14
6	90.34	99.09	98.46	99.77
7	95.59	99.69	94.77	99.23
8	90.72	96.26	97.71	96.32
9	97.25	70.76	92.91	94.06

Table 6. Comparison of active classifiers and FAAL on Pavia University.

4.5. Study on Acquisition Heuristics

We finally compare the adversarially learned acquisition heuristic of our FAAL framework with off-the-shelf acquisition heuristics. We exert the random sampling (i.e., querying unlabeled samples randomly within each active learning loop), the least confidence [32], the entropy sampling [32], and BALD [33] onto our employed classifier, separately. The usage of labeled samples is as default.

Tables 7 and 8 list quantitative comparisons on the two datasets: Indian Pines and Pavia University, respectively. We observe that our FAAL framework achieves superior performance to the classifiers with other available acquisition heuristics. An intuitive reason is that our FAAL framework makes decisions for the active query relying on the high-level features instead of the output of the classifier. Adversarial learning with the high-level features makes our FAAL framework comprehensively capture the feature variability of hyperspectral images.

Metric/Heuristic	Random Sampling	Least Confidence	Entropy Sampling	BALD	FAAL
OA (%)	88.61	89.11	89.30	90.41	91.41
AA (%)	91.55	91.58	92.28	92.47	93.20
$k \ (imes 100)$	87.01	87.60	87.81	89.07	90.20
1	98.60	100	100	100	99.29
2	89.09	83.95	84.69	83.92	82.35
3	77.16	81.44	88.20	94.28	89.81
4	90.00	91.43	93.33	93.33	89.88
5	89.72	73.21	79.10	79.91	87.59
6	92.87	96.98	96.93	95.25	92.14
7	100	100	100	98.33	100
8	100	99.92	100	99.68	100
9	100	100	100	100	100
10	81.19	88.41	81.48	80.04	85.94
11	93.77	91.03	91.15	95.11	97.44
12	74.81	77.61	79.20	72.65	78.91
13	97.75	97.57	97.75	94.94	98.88
14	85.71	95.34	93.50	95.55	95.20
15	95.88	94.22	92.94	98.14	98.38
16	98.38	94.24	98.15	98.35	95.37

Table 7. Comparison of the classifier with different acquisition heuristics on Indian Pines.

9

51.63

Metric/Heuristic	Random Sampling	Least Confidence	Entropy Sampling	BALD	FAAL
OA (%)	90.16	90.36	91.08	92.09	93.47
AA (%)	83.69	85.28	86.35	88.24	88.97
k~(imes 100)	86.91	87.19	88.17	89.51	91.34
1	83.90	75.08	79.53	82.31	92.01
2	96.06	98.22	98.05	97.49	97.36
3	98.58	98.85	98.27	99.44	97.93
4	64.56	70.11	66.23	61.68	70.62
5	67.53	72.81	80.38	83.38	77.32
6	94.68	91.96	93.89	99.57	99.09
7	99.53	99.73	99.33	99.61	99.69
8	96.77	98.08	97.13	99.08	96.26

Table 8. Comparison of the classifier with different acquisition heuristics on Pavia University.

Figures 4 and 5 illustrate the qualitative results on Indian Pines and Pavia University, respectively. Groundtruth maps are provided firstly as references.

62.66

64.38

Classification maps predicted by the classifier with the random sampling, the least confidence, the entropy sampling, BALD, and our adversarially learned acquisition heuristic (i.e., our FAAL framework) are given separately. Overall, superior visual results are obtained with the adversarially learned acquisition heuristic of our FAAL framework.



Figure 4. Classification maps on Indian Pines. The first map is the groundtruth. The rest ones are predicted by our employed classifier with the random sampling, the least confidence, the entropy sampling, bayesian active learning disagreement (BALD), and our adversarially learned acquisition heuristic (i.e., our FAAL framework), separately. Table 2 lists what class each color represent. Specifically, the part in black means unlabeled.



Figure 5. Classification maps on Pavia University. The first map is the groundtruth. The rest ones are predicted by our employed classifier with the random sampling, the least confidence, the entropy sampling, BALD, and our adversarially learned acquisition heuristic (i.e., our FAAL framework), separately. Table 3 lists what class each color represent. Specifically, the part in black means unlabeled.

Figures 6 and 7 compare the classification performance after each active learning loop on Indian Pines and Pavia University, respectively. The comparisons are measured by OA, AA, and KAPPA,

70.76

71.74

separately. Not surprisingly, the varying curves achieved by the classifier with the random sampling are at the lowest almost all the time. The curves obtained with the adversarially learned acquisition heuristic of our FAAL framework tangle with those obtained with other available acquisition heuristics in the first two active learning loops, as well as edge ahead of them in the next three active learning loops.



Figure 6. Performance after each active learning loop on Indian Pines. The number of initial labeled samples is 80. There are 114, 148, 182, 216, and 250 labeled samples after the first, the second, the third, the fourth, and the fifth active learning loop, respectively. Three general metrics: OA (leftmost), AA (middle), and KAPPA (rightmost) are used for comparison.



Figure 7. Performance after each active learning loop on Pavia University. The number of initial labeled samples is 45. There are 86, 127, 168, 209, and 250 labeled samples after the first, the second, the third, the fourth, and the fifth active learning loop, respectively. Three general metrics: OA (leftmost), AA (middle), and KAPPA (rightmost) are used for comparison.

5. Conclusions

For this article, we developed an active deep learning strategy, i.e., FAAL, for classifying hyperspectral images with limited labeled samples. Our FAAL framework comprehensively captured the feature variability of hyperspectral images and included a purely parameterized yet simple acquisition heuristic. The acquisition heuristic was adversarially learned with high-level features, which stemmed from one intermediate layer of a deep learning classifier. Experimental evaluations on two public hyperspectral image datasets demonstrated that our FAAL framework achieves the state-of-the-art in classifying hyperspectral images with limited labeled samples.

Our FAAL framework admits many possible extensions.

As either the GAN or the classifier can be separated from each other, imposing constraints on
each of them is feasible. In this scenario, is there an additional constraint capable of capturing the
feature variability of hyperspectral images further? As the active query is fully unsupervised and
unrestricted by class, is there an additional constraint responsible for learning to be class-balanced?

- High-level feature space is commonly low dimensional compared to data space [60–62]. To an extent, dealing with a low dimensional space would reduce the requirement of computational resources and ease the burden of network design. We believe that state-of-the-art feature extraction and band selection methods would come into effect in this direction. Besides, building a low dimensional latent space external to the classifier would be constructive despite the additional burden.
- The task-agnostic property of our acquisition heuristic makes it scalable to other applications, possibly spanning from computer vision to remote sensing. Research on this direction would further examine the effectiveness of the adversarially learned and purely parameterized yet simple acquisition heuristic.

Author Contributions: Conceptualization, G.W., and P.R.; methodology, G.W.; software, G.W.; writing–original draft preparation, G.W., and P.R.; writing–review and editing, G.W., and P.R.; supervision, P.R.; funding acquisition, P.R.. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Key Research and Development Program of China under Project 2019YFC1408400, and in part by the Innovative Research Team Program for Young Scholars at Universities in Shandong Province under Project 2020KJN010.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Audebert, N.; Le Saux, B.; Lefevre, S. Deep Learning for Classification of Hyperspectral Data: A Comparative Review. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 159–173. [CrossRef]
- 2. Wan, Y.; Ma, A.; Zhong, Y.; Hu, X.; Zhang, L. Multiobjective Hyperspectral Feature Selection Based on Discrete Sine Cosine Algorithm. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 3601–3618. [CrossRef]
- Ghamisi, P.; Yokoya, N.; Li, J.; Liao, W.; Liu, S.; Plaza, J.; Rasti, B.; Plaza, A. Advances in Hyperspectral Image and Signal Processing: A Comprehensive Overview of the State of the Art. *IEEE Geosci. Remote Sens. Mag.* 2017, 5, 37–78. [CrossRef]
- 4. Yokoya, N.; Grohnfeldt, C.; Chanussot, J. Hyperspectral and Multispectral Data Fusion: A Comparative Review of the Recent Literature. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 29–56. [CrossRef]
- 5. Luo, F.; Huang, H.; Duan, Y.; Liu, J.; Liao, Y. Local Geometric Structure Feature for Dimensionality Reduction of Hyperspectral Imagery. *Remote Sens.* **2017**, *9*, 790. [CrossRef]
- 6. Luo, F.; Zhang, L.; Zhou, X.; Guo, T.; Cheng, Y.; Yin, T. Sparse-Adaptive Hypergraph Discriminant Analysis for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 1082–1086. [CrossRef]
- 7. Ghamisi, P.; Plaza, J.; Chen, Y.; Li, J.; Plaza, A.J. Advanced Spectral Classifiers for Hyperspectral Images: A Review. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–32. [CrossRef]
- 8. Wang, G.; Ren, P. Delving Into Classifying Hyperspectral Images via Graphical Adversarial Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 2019–2031. [CrossRef]
- Luo, F.; Zhang, L.; Du, B.; Zhang, L. Dimensionality Reduction With Enhanced Hybrid-Graph Discriminant Learning for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* 2020, 58, 5336–5353. [CrossRef]
- 10. Ben Hamida, A.; Benoit, A.; Lambert, P.; Ben Amar, C. 3-D Deep Learning Approach for Remote Sensing Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4420–4434. [CrossRef]
- 11. Roy, S.K.; Krishna, G.; Dubey, S.R.; Chaudhuri, B.B. HybridSN: Exploring 3-D-2-D CNN Feature Hierarchy for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 277–281. [CrossRef]
- 12. Fang, B.; Li, Y.; Zhang, H.; Chan, J.C.W. Collaborative Learning of Lightweight Convolutional Neural Network and Deep Clustering for Hyperspectral Image Semi-Supervised Classification with Limited Training Samples. *ISPRS J. Photogramm. Remote Sens.* **2020**, *161*, 164–178. [CrossRef]
- 13. Li, X.; Ding, M.; Pižurica, A. Deep Feature Fusion via Two-Stream Convolutional Neural Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 2615–2629. [CrossRef]
- 14. Tran, T.; Do, T.T.; Reid, I.; Carneiro, G. Bayesian Generative Active Deep Learning. In Proceedings of the Internetional Conference Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 6295–6304.

- 15. Paoletti, M.; Haut, J.; Plaza, J.; Plaza, A. Deep Learning Classifiers for Hyperspectral Imaging: A Review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *158*, 279–317. [CrossRef]
- Yue, Z.; Gao, F.; Xiong, Q.; Wang, J.; Huang, T.; Yang, E.; Zhou, H. A Novel Semi-Supervised Convolutional Neural Network Method for Synthetic Aperture Radar Image Recognition. *Cogn. Comput.* 2019, 1–12. [CrossRef]
- 17. Cao, X.; Yao, J.; Xu, Z.; Meng, D. Hyperspectral Image Classification With Convolutional Neural Network and Active Learning. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 4604–4616. [CrossRef]
- 18. Samat, A.; Li, J.; Liu, S.; Du, P.; Miao, Z.; Luo, J. Improved Hyperspectral Image Classification by Active Learning Using Pre-Designed Mixed Pixels. *Pattern Recognit.* **2016**, *51*, 43–58. [CrossRef]
- 19. He, Z.; Liu, H.; Wang, Y.; Hu, J. Generative Adversarial Networks-Based Semi-Supervised Learning for Hyperspectral Image Classification. *Remote Sens.* **2017**, *9*, 1042. [CrossRef]
- Jia, S.; Zhuang, J.; Deng, L.; Zhu, J.; Xu, M.; Zhou, J.; Jia, X. 3-D Gaussian–Gabor Feature Extraction and Selection for Hyperspectral Imagery Classification. *IEEE Trans. Geosci. Remote Sens.* 2019, 57, 8813–8826. [CrossRef]
- 21. Ducoffe, M.; Precioso, F. Adversarial Active Learning for Deep Networks: A Margin Based Approach. *arXiv* **2018**, arXiv:1802.09841.
- 22. Zhu, J.; Jose, B. Generative Adversarial Active Learning. arXiv 2017, arXiv:1702.07956.
- 23. Liu, C.; He, L.; Li, Z.; Li, J. Feature-Driven Active Learning for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 341–354. [CrossRef]
- 24. Ni, D.; Ma, H. Active Learning for Hyperspectral Image Classification Using Sparse Code Histogram and Graph-Based Spatial Refinement. *Int. J. Remote Sens.* **2017**, *38*, 923–948. [CrossRef]
- Zhang, Z.; Pasolli, E.; Crawford, M.M. An Adaptive Multiview Active Learning Approach for Spectral–Spatial Classification of Hyperspectral Images. *IEEE Trans. Geosci. Remote Sens.* 2020, 58, 2557–2570. [CrossRef]
- Zhang, Z.; Pasolli, P.; Crawford, M.M.; Tilton, J.C. An Active Learning Framework for Hyperspectral Image Classification Using Hierarchical Segmentation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2016, 9, 640–654. [CrossRef]
- Haut, J.M.; Paoletti, M.E.; Plaza, J.; Li, J.; Plaza, A. Active Learning With Convolutional Neural Networks for Hyperspectral Image Classification Using A New Bayesian Approach. *IEEE Trans. Geosci. Remote Sens.* 2018, 56, 6440–6461. [CrossRef]
- 28. Liu, C.; Li, J.; He, L. Superpixel-Based Semisupervised Active Learning for Hyperspectral Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 357–370. [CrossRef]
- 29. Yoo, D.; Kweon, I.S. Learning Loss for Active Learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 93–102.
- 30. Jedoui, K.; Krishna, R.; Bernstein, M.S.; Fei-Fei, L. Deep Bayesian Active Learning for Multiple Correct Outputs. *arXiv* **2019**, arXiv:1912.01119.
- Kendall, A.; Gal, Y. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5574–5584.
- 32. Wang, D.; Shang, Y. A New Active Labeling Method for Deep Learning. In Proceedings of the 2014 International Joint Conference on Neural Networks (IJCNN), Beijing, China, 6–11 July 2014; pp. 112–119.
- 33. Gal, Y.; Islam, R.; Ghahramani, Z. Deep Bayesian Active Learning with Image Data. In Proceedings of the Internetional Conference Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 1183–1192.
- 34. Gao, F.; Ma, F.; Wang, J.; Sun, J.; Yang, E.; Zhou, H. Visual saliency modeling for river detection in high-resolution SAR imagery. *IEEE Access* **2018**, *6*, 1000–1014. [CrossRef]
- 35. Gao, F.; Huang, T.; Sun, J.; Wang, J.; Hussain, A.; Yang, E. A New Algorithm of SAR Image Target Recognition Based on Improved Deep Convolutional Neural Network. *Cogn. Comput.* **2019**, *11*, 809–824. [CrossRef]
- 36. Jing, L.; Tian, Y. Self-Supervised Visual Feature Learning with Deep Neural Networks: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**. [CrossRef] [PubMed]
- 37. Settles, B. *Active Learning Literature Survey;* Computer Sciences Technical Report 1648; University of Wisconsin–Madison: Madison, WI, USA, 2009.

- 38. Vondrick, C.; Ramanan, D. Video Annotation and Tracking with Active Learning. In Proceedings of the Advances in Neural Information Processing Systems, Granada, Spain, 12–17 December 2011; pp. 28–36.
- 39. Mottaghi, A.; Yeung, S. Adversarial Representation Active Learning. arXiv 2019, arXiv:1912.09720.
- 40. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.
- 41. Radford, A.; Metz, L.; Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv* **2015**, arXiv:1511.06434.
- 42. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein Generative Adversarial Networks. In Proceedings of the Internetional Conference Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 214–223.
- Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A.C. Improved Training of Wasserstein GANs. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5767–5777.
- 44. Creswell, A.; White, T.; Dumoulin, V.; Arulkumaran, K.; Sengupta, B.; Bharath, A.A. Generative Adversarial Networks: An Overview. *IEEE Signal Process. Mag.* **2018**, *35*, 53–65. [CrossRef]
- 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. [CrossRef]
- 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.* 2019, 57, 2669–2688. [CrossRef]
- 47. Zhong, Z.; Li, J.; Clausi, D.A.; Wong, A. Generative Adversarial Networks and Conditional Random Fields for Hyperspectral Image Classification. *IEEE Trans. Cybern.* **2020**, *50*, 3318–3329. [CrossRef]
- 48. 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. [CrossRef]
- 49. Hang, R.; Zhou, F.; Liu, Q.; Ghamisi, P. Classification of Hyperspectral Images via Multitask Generative Adversarial Networks. *IEEE Trans. Geosci. Remote Sens.* **2020**. [CrossRef]
- 50. 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. [CrossRef]
- 51. He, L.; Li, J.; Liu, C.; Li, S. Recent Advances on Spectral–Spatial Hyperspectral Image Classification: An Overview and New Guidelines. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 1579–1597. [CrossRef]
- Gao, Q.; Lim, S.; Jia, X. Spectral–Spatial Hyperspectral Image Classification Using A Multiscale Conservative Smoothing Scheme and Adaptive Sparse Representation. *IEEE Trans. Geosci. Remote Sens.* 2019, 57, 7718–7730. [CrossRef]
- 53. Tyo, J.S.; Konsolakis, A.; Diersen, D.I.; Olsen, R.C. Principal-Components-Based Display Strategy for Spectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 708–718. [CrossRef]
- Jiang, J.; Ma, J.; Chen, C.; Wang, Z.; Cai, Z.; Wang, L. SuperPCA: A Superpixelwise PCA Approach for Unsupervised Feature Extraction of Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.* 2018, 56, 4581–4593. [CrossRef]
- 55. Goodfellow, I. NIPS 2016 Tutorial: Generative Adversarial Networks. arXiv 2017, arXiv:1701.00160.
- Mao, Q.; Lee, H.Y.; Tseng, H.Y.; Ma, S.; Yang, M.H. Mode Seeking Generative Adversarial Networks for Diverse Image Synthesis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 1429–1437.
- 57. Sinha, S.; Ebrahimi, S.; Darrell, T. Variational Adversarial Active Learning. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 5971–5980.
- Jia, S.; Deng, X.; Zhu, J.; Xu, M.; Zhou, J.; Jia, X. Collaborative Representation-Based Multiscale Superpixel Fusion for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* 2019, 57, 7770–7784. [CrossRef]
- 59. Jia, S.; Deng, X.; Meng, X.; Zhou, J.; Jia, X. Superpixel-Level Weighted Label Propagation for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 5077–5091. [CrossRef]
- 60. Luo, F.; Du, B.; Zhang, L.; Zhang, L.; Tao, D. Feature Learning Using Spatial-Spectral Hypergraph Discriminant Analysis for Hyperspectral Image. *IEEE Trans. Cybern.* **2019**, *49*, 2406–2419. [CrossRef]

62. Chen, J.; Xie, Y.; Wang, K.; Zhang, C.; Vannan, M.A.; Wang, B.; Qian, Z. Active Image Synthesis for Efficient Labeling. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**. [CrossRef]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



 \odot 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/).