

Article

A Generalized Zero-Shot Learning Framework for PolSAR Land Cover Classification

Rong Gui ¹ , Xin Xu ^{1,*}, Lei Wang ¹ , Rui Yang ^{1,2} and Fangling Pu ^{1,2}

¹ School of Electronic Information, Wuhan University, Wuhan 430072, China; ronggui2013@whu.edu.cn (R.G.); wanglei2016@whu.edu.cn (L.W.); ruiyang@whu.edu.cn (R.Y.); flpu@whu.edu.cn (F.P.)

² Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China

* Correspondence: xinxu@whu.edu.cn; Tel.: +86-27-6875-2836; Fax: +86-27-6875-6275

Received: 7 June 2018; Accepted: 17 August 2018; Published: 19 August 2018



Abstract: Most supervised classification methods for polarimetric synthetic aperture radar (PolSAR) data rely on abundant labeled samples, and cannot tackle the problem that categorizes or infers unseen land cover classes without training samples. Aiming to categorize instances from both seen and unseen classes simultaneously, a generalized zero-shot learning (GZSL)-based PolSAR land cover classification framework is proposed. The semantic attributes are first collected to describe characteristics of typical land cover types in PolSAR images, and semantic relevance between attributes is established to relate unseen and seen classes. Via latent embedding, the projection between mid-level polarimetric features and semantic attributes for each land cover class can be obtained during the training stage. The GZSL model for PolSAR data is constructed by mid-level polarimetric features, the projection relationship, and the semantic relevance. Finally, the labels of the test instances can be predicted, even for some unseen classes. Experiments on three real RadarSAT-2 PolSAR datasets show that the proposed framework can classify both seen and unseen land cover classes with limited kinds of training classes, which reduces the requirement for labeled samples. The classification accuracy of the unseen land cover class reaches about 73% if semantic relevance exists during the training stage.

Keywords: generalized zero-shot learning; semantic attributes; classification; polarimetric SAR; polarization feature

1. Introduction

As an important means of longtime earth surface monitoring and large-scale land cover information acquiring, synthetic aperture radar (SAR) has become increasingly important in land cover classification [1,2], natural disaster prevention [3], target recognition [4,5] and urban observation [6,7]. More and more SAR systems, e.g., RadarSAT-2, TerraSAR-X, Gaofen-3 [2,8] can acquire polarimetric SAR (PolSAR) images and provide more information about terrain targets and land cover [9] by emitting and receiving fully-polarized radar waves. Therefore, developing classification methods for PolSAR interpretation to identify land cover information has been widely studied. Generally, PolSAR land cover classification algorithms can be divided into three categories called statistical model-based algorithms [10,11], scattering mechanism analysis-based algorithms [7,12,13], and the algorithms combining machine-learning classifiers and polarimetric features [14,15] or PolSAR data [16]. Many works focus on the third category among abovementioned approaches; more and more complicated and efficient supervised learning [14–16] and deep learning algorithms [17–19] have been proposed to interpret the PolSAR land cover information.

Despite the fact that the some traditional supervised learning and deep learning algorithms for PolSAR classification can achieve promising results, they fail to recognize unseen classes which are not

included in the training data. That is, in the test stage, if there is a new land cover class which was never seen in the training stage, they do not possess the ability to learn or infer the new classes, which may limit these methods for practical applications [20]. On the other hand, remote sensing has traditionally enjoyed an abundance of data, but obtaining label information has always been an important bottleneck in classification studies [21,22]. Especially for PolSAR interpretation, correct and sufficient labeled samples are usually rare and labeling the samples always needs some professional knowledge. In short, labeling samples in PolSAR imagery is difficult, expensive, and time-consuming [23]. Therefore, it is especially necessary to develop methods for PolSAR interpretation to identify the unseen targets or land cover, since the massiveness of the PolSAR data, the urgent need for PolSAR interpretation tasks, and the labeled data are not always available.

Zero-shot learning (ZSL) aims to recognize objects whose instances have not been seen during training [24–26] by means of leveraging intermediate semantic information, which could significantly extend machine-learning abilities for handling practical problems [27]. With the capability of transferring semantic knowledge, ZSL can be regarded as a good complement to conventional supervised learning [28,29]. Therefore, ZSL has received much attention recently in computer vision research [27] and has obtained promising results for identifying unseen samples in several standard datasets [24,26,27], including the Animals with Attributes dataset (AwA) [28], CUB-200-2011 Birds dataset (CUB) [27], aPascal and aYahoo datasets [26], ImageNet [24], and so on. Typical intermediate semantic information, such as visual attributes [27,28,30] or Word2Vec vector representations [22,24,26] are shareable to both seen and unseen classes [28,30,31], thus, ZSL can learn how to recognize new unseen classes that have no training samples by relating them to seen classes that were previously learned.

However, despite ZSL achieving promising results in several standard datasets, few articles on ZSL application in the remote sensing interpretation have been reported. Li et al. [22] applied ZSL for high spatial resolution remote sensing images classification and the average accuracy was about 58% for the UC Merced land use dataset with the unseen/seen ratio equal to 5/16 (16 kinds of training samples and the testing samples were the other five kinds). Song et al. [32] employed ZSL to SAR target recognition demonstrated on the MSTAR data set, where seven targets samples were used in training and the eighth target type was used for testing. Sumbul et al. [21] studied the ZSL problem for fine-grained street tree recognition in aerial data and achieved a 14.3% recognition accuracy with an unseen/seen ratio of 16/24. Moreover, ZSL methods have been studied in an unrealistic setting where the test data are assumed to come from unseen classes only [33]. Generally, the seen classes are more common than the unseen ones, therefore, it is unrealistic to assume that we will never encounter seen classes during the test stage. In a real scenario, though there are limited types of labeled training classes, the test samples always contain both the seen samples and unseen samples. This problem is known as generalized zero-shot learning (GZSL) [24,33] and is considered a more challenging problem setting [33,34]. GZSL relaxes the unrealistic assumption in conventional ZSL that test data belong only to unseen novel classes. In GZSL, test data may also come from seen classes, and the labeling space is the union of both the unseen and seen classes [33].

As mentioned above, to get correct and sufficient labeled samples in PolSAR imagery is always difficult, expensive, and time consuming. For some obvious land cover classes such as urban areas, forest lands, water areas, etc., it is easy to obtain labeled samples in PolSAR imagery, but the labeled samples of rural areas, wetland, and grasslands are relatively difficult to obtain. Furthermore, a common situation for PolSAR land cover classification is that there are only samples of urban areas and water areas, but the actual test data contain more abundant land cover categories and need to get more detailed category information. On the other hand, there is rich scene information that can be described by semantic relationships in the PolSAR imagery, for example, both forest lands and grasslands exhibit certain surface scattering characteristics. Therefore, GZSL can provide more practical solutions for PolSAR interpretation to recognize some new land cover categories without labeled samples, while conventional supervised learning and common deep learning algorithms

always fail to categorize unseen instances. This paper focuses on the GZSL for PolSAR land cover classification since we hold the view that the GZSL in PolSAR land cover classification is more practical than ZSL.

The main challenge for GZSL applied in PolSAR land cover classification comes from the intermediate semantic information representation of PolSAR land cover classes and the inference of whether a testing sample belongs to a seen or unseen class. Attributes often refer to the well-known common semantic characteristics of objects and can be acquired by human annotation [21], or neural language models such as Word2Vec and GloVe [35]. The commonly used attributes in traditional ZSL are mainly for natural images or ground-level images. Whether some of the related annotated attributes and the neural language models are available for PolSAR land cover classes' semantic representation need to be further verified. The inference progresses of GZSL are often more complicated than ZSL, since the search space is limited to the unseen classes only in the ZSL inference, while the search space is the union of both seen and unseen classes in GZSL inference. Recent work [24,33] has shown that the accuracies of most ZSL approaches drop significantly in the GZSL setting in which the test samples always contain both the seen samples and unseen samples. To solve these two challenges of the GZSL framework applied in PolSAR classification the following steps have been designed and conducted: Firstly, the semantic attributes from the Word2Vec semantic vectors [35], SUN attributes [36], and the selected SUN attributes are collected and evaluated to describe the characteristics of the PolSAR typical classes and semantic relevance between attributes is obtained to relate unseen and seen classes. Then the projection relationship between the mid-level representation of PolSAR data samples and class attributes is established by a latent embedding model [37] during training stage. At last, for every test instance, through the GZSL model constructed by the polarization feature representation, the projection relationship, and semantic relevance, the labels of the test instances can be predicted. Semantic relevance is used to constraint and to amend the scores between the unseen and seen classes in this prediction process. Even though some test instances do not have training samples backing them, the inference progress in the proposed framework is determined by the polarization feature, projection relationship, and semantic relevance. The proposed method has the following contributions and advantages:

- The adaption of the available semantic attributes for PolSAR land cover class description has been evaluated, including the Word2Vec semantic vectors, SUN attributes, and the selected SUN attributes.
- By utilizing the rich polarization features and semantic information in the PolSAR imagery, the proposed GZSL framework can provide a more practical solution for PolSAR interpretation to classify some new land cover categories without labeled samples, which can reduce the requirement for sample labeling and make the framework has the ability to identify the new types in PolSAR land cover classification.

To the best of our knowledge, the GZSL framework for PolSAR land cover classification has not been studied in the remote sensing literature even though it is a highly probable scenario where new land cover categories can be introduced after the training stage or when no training examples exist for several rare classes that are still of interest. The remainder of the paper is structured as follows: in Section 2, the related work about the ZSL and GZSL framework, and the intermediate semantic information are briefly introduced. In Section 3, we present the workflow and the implementation of the proposed GZSL for the PolSAR land cover classification method in detail. In Section 4, we present and evaluate the results obtained by applying the proposed framework for PolSAR land cover classification. The capabilities and limitations are discussed in Section 5. Finally, the conclusions are presented in Section 6.

2. Related Work

In this section, a brief overview of the previous studies related to the ZSL and GZSL framework is provided, followed by a short introduction to intermediate semantic information.

2.1. From Zero-Shot Learning to Generalized Zero-Shot Learning

Zero-shot learning is an attractive new task that has recently aroused increasing attention [22,30,35]. ZSL has made it possible to recognize new category without acquiring training examples beforehand via leveraging semantic information. Typically, there are three main parts in a conventional ZSL framework, including the image feature extraction, intermediate semantic information of the training and test classes, and semantic embedding \mathbf{W} [22,31,38], as illustrated in Figure 1a. We denote the label space of the seen classes in the training stage as $C_S = \{1, 2, \dots, p\}$; $C_U = \{p + 1, p + 2, \dots, p + q\}$ means the label space of the unseen classes in test stage, in ZSL framework where $C_S \cap C_U = \emptyset$. The ZSL framework proposed by Lampert et al. [39], first links the attribute annotations that are human prior knowledge with low-level features extracted from images. Typically, the taxonomy of ZSL framework could be distinguished as the approaches based on independent semantics and the approaches based on semantic embedding [29,39]. For ZSL approaches based on independent semantics, it consists of learning an independent classifier per semantic [29,31]. Due to its simplicity, ZSL approaches based on independent semantics became widely popular, including direct attribute prediction [39], semantic auto-encoder [26], and latent embedding [37]. While, for ZSL approaches based on semantic embedding, these were accomplished using a label embedding function φ , to map each class C_i into a vector $\varphi(C_i)$ in the space of attributes. Semantic embedding based methods such as indirect attribute prediction [39], label embedding [38], semantic similarity embedding [40], and so on, learn single multi-class classifiers that optimally discriminate between all seen classes, and the predicted probabilities and attribute annotations of the seen classes are then used to estimate object attribute values [27,29]. As illustrated in Figure 1a, the label inference progress of a ZSL usually employs the 1-nearest neighbor (NN) strategy [25,26,31].

Although the aforementioned ZSL models have shown considerable promise on some standard datasets, a key limitation in most of these models is that, at the test stage, they are highly biased towards predicting the seen classes [24,33,34]. This is because the ZSL model can only learn from the seen classes in the training stage, while the test examples only come from the unseen classes and the search space is limited to the unseen classes only. The more challenging setting where the training and test classes are not disjoint is known as generalized zero-shot learning (GZSL), and is considered a more formidable problem setting [33,34].

The biggest difference between GZSL and ZSL is the setting of test classes, as illustrated in Figure 1a,b. The training and test classes are assumed to be strictly disjoint in the ZSL framework, while in GZSL, the test data are from both seen and unseen classes. Recent work [33] has shown that the accuracies of most ZSL approaches drop significantly in GZSL settings. Thus, the inference progress should more complicated than ZSL. With the ability to identify both the unseen and seen classes via leveraging semantic information, GZSL can provide a new way for PolSAR image land cover classification by utilizing the rich polarization features and semantic information in PolSAR imagery. To a certain extent, GZSL can reduce the requirement for sample labeling and make the framework has the ability to identify new types in PolSAR land cover classification.

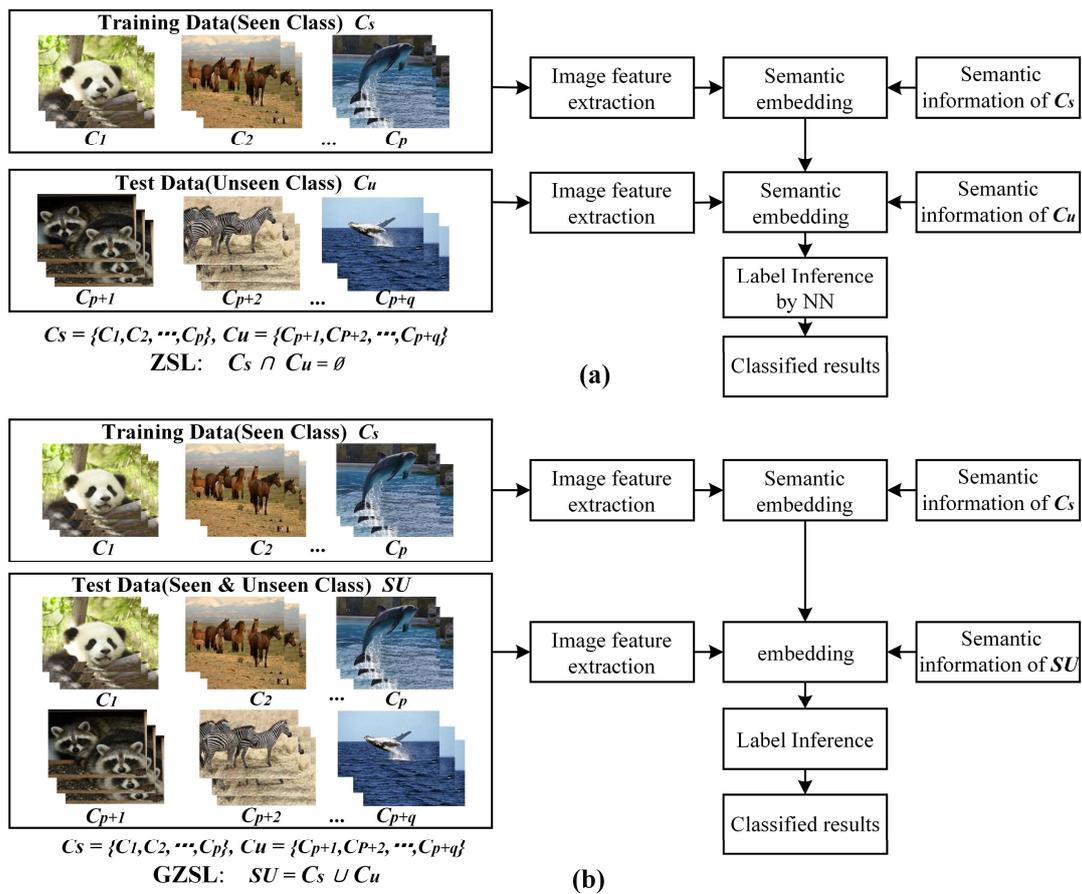


Figure 1. The comparative illustration of the conventional ZSL framework and GZSL framework: (a) the conventional ZSL framework; and (b) the GZSL framework.

2.2. Intermediate Semantic Information

Since no training instances are available for some unseen test categories in GZSL, image features alone are not sufficient to form the association between the unseen and seen classes. Thus, intermediate semantic information (auxiliary information [21] or side information [20]) can act as an intermediate layer for building this association between the unseen and seen classes. As typical intermediate semantic information, attributes [20,21] acquired by human annotation have been successfully used in ZSL tasks for the identification of different animal, bird, or dog species or indoor and outdoor scene categories in computer vision. Another main type of intermediate semantic information available is the word vectors from a linguistic corpus with neural language models such as the Word2Vec model [41]. The word vectors are vector representations of words learned in a large-scale language corpus which suggests that words frequently appearing in a common context would result in a closer distance [28], which can be used to map names of scene classes (both seen and unseen) to semantic vectors [22,26,37,40].

Remote sensing images, especially PolSAR images, contain rich scene information, including the land cover surface properties and spatial properties [42]. The SUN attribute database [43,44] is the first large-scale scene attribute database, build on top of the diverse SUN categorical database that spans more than 700 categories and 14,000 images. Additionally, the SUN attributes are related to materials, surface properties, functions, and spatial envelope properties. The adaption of Word2Vec semantic vectors, SUN attributes, and the selected SUN attributes for PolSAR land cover class description will be evaluated in Section 4.

3. Methodology

The applied GZSL workflow (Figure 2) of PolSAR land cover classification is organized by three parts including the polarization feature representation, semantic representation about PolSAR land cover classes, and the generalized zero-shot learning with semantic relevance. The test instances are from both the seen and unseen classes, and we need to classify them into the joint labeling space of both types of seen and unseen classes. The detailed description of the method works as follows.

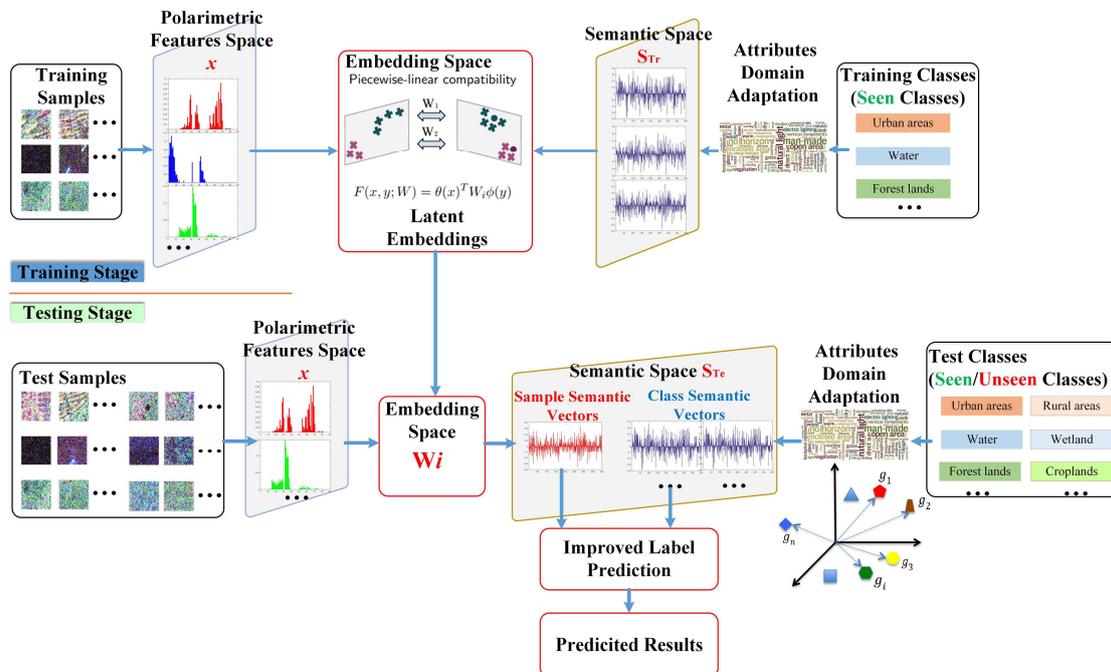


Figure 2. The flowchart of the proposed method.

3.1. Polarization Feature Representation

PolSAR data can be represented by a scattering matrix **S**, covariance matrix **C**, or coherency matrix **T**, which can provide more scattering information for land cover classification [45]. The Yamaguchi decomposition [13,42,46] extended the three-component scattering model [46] by adding the helix scattering mechanism as the fourth component to deal with the observed actual phenomenon. The model can be expressed as:

$$T = f_s T_S + f_d T_D + f_v T_V + f_c T_H \tag{1}$$

T is the measured polarimetric coherency matrix. **T_S**, **T_D**, **T_V**, and **T_H** correspond to the coherency matrix for surface scattering, double-bounce scattering, volume scattering, and helix scattering, respectively, and $f_s, f_d, f_v,$ and f_c are the corresponding coefficients. Furthermore, the orientation of buildings areas with respect to the radar illumination also affects their polarimetric properties [7]; this is possible to cause confusion between those buildings areas and vegetation. A rotation of the coherency matrix (namely, deorientation) [13] can be adopted for a more accurate decomposition.

Based on the aforementioned scattering powers obtained from Yamaguchi four-component decomposition with deorientation, we employ our previous work [5,46,47] to obtain the mid-level components, called intermediates, which are unsupervised statistical patterns learned from PolSAR images. The mid-level polarization feature representation has proven to have a good classification ability, especially for the differentiation of building areas and vegetation [5,46], and the differentiation of building density [47] in PolSAR data. The flowchart of the applied mid-level polarization feature

(scattering mechanism based statistical feature) representation algorithm based on four-component decomposition with deorientation is given in Figure 3. The PolSAR images analyzed in proposed framework are mainly the 8-m resolution full polarimetric SAR imagery, thus, the sample size of 50×50 pixels can hold the composition of different land cover types and capture sufficient context information. The dimension of the applied mid-level polarization feature is 80. Typically, the volume scattering power and surface scattering power are much larger than the power of double-bounce scattering and helix scattering, for vegetation areas, water areas and some low- and medium-density building areas with special orientations in PolSAR imagery [7,46,47]. In order to keep enough characteristics of the volume scattering and surface scattering, and avoid too much of the zero value in double-bounce scattering and helix scattering characteristics after the merging, the merging number of the four scattering components are different.

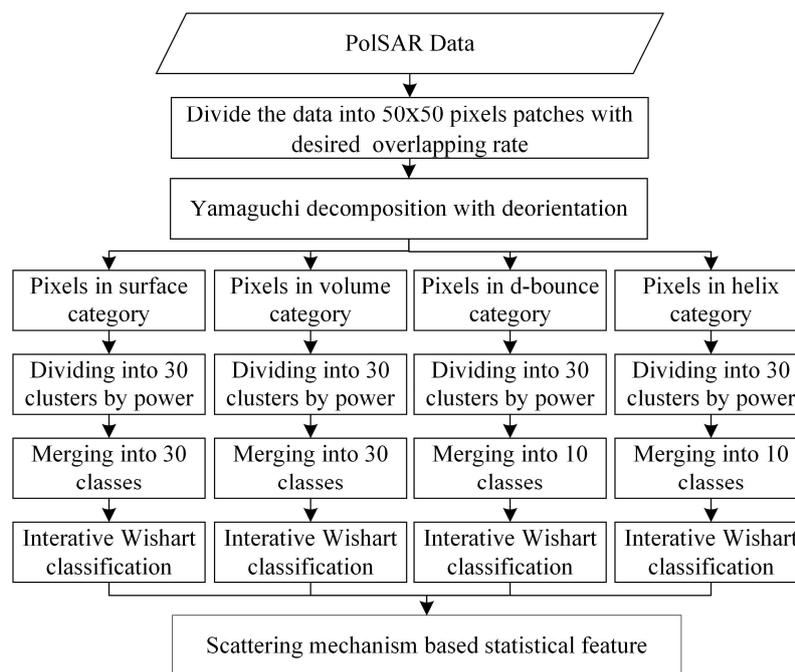


Figure 3. The applied mid-level polarization feature (scattering mechanism based statistical feature) representation algorithm.

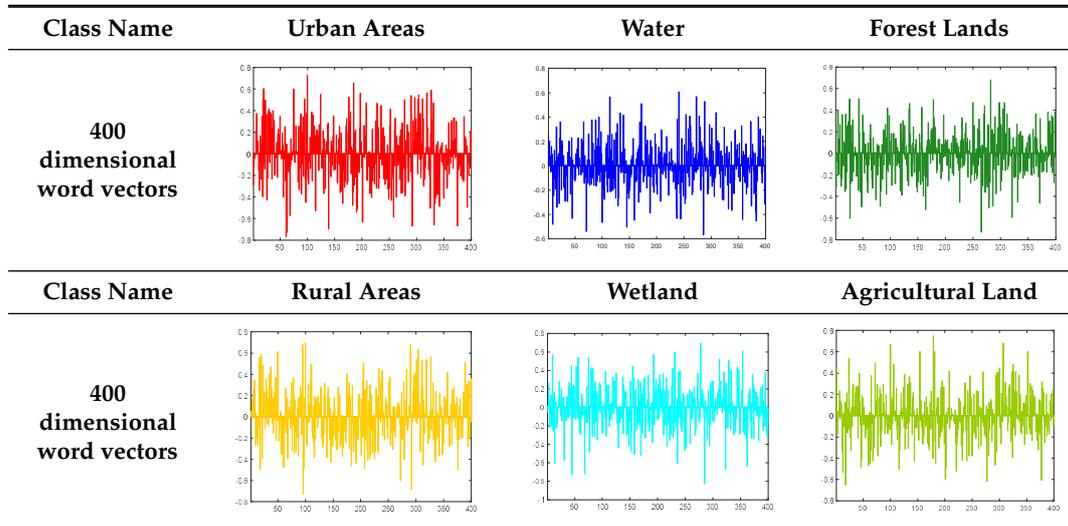
3.2. Semantic Representation Of PolSAR Land Cover Classes

The typical land cover classes in 8-m resolution PolSAR images include urban areas (c1), rural areas (c2), water (c3), forest lands (c4), croplands (c5), wetland (c6), and agricultural land (c7). The abovementioned land cover classes are all regional objects in PolSAR images under 8-m resolution and they possess certain surface properties and spatial properties. By employing the Word2Vec model [41], these land cover classes can all be mapped to semantic word vectors; the setting is the Skip-Gram model and the vector is of 400 dimensions. Some examples of the typical land cover classes and corresponding word vectors have been illustrated in Table 1.

On the other hand, we also employ the SUN attributes [43,44] since the 102-dimensional attribute vector, including some typical surface and spatial properties. These may correspond to the rich scene information in PolSAR images. So, some attributes of the corresponding scene categories are selected, the 'city' scene category represents the 'urban areas' class, the 'lake natural' scene category represents the 'water' class, the 'woodland' scene category represents the 'forest lands' class, the 'cornfield' scene category represents the 'croplands' class, the 'village' and 'factory outdoor' scene category represents the 'rural areas' class, and so on. As the same scene category in the SUN attributes has more

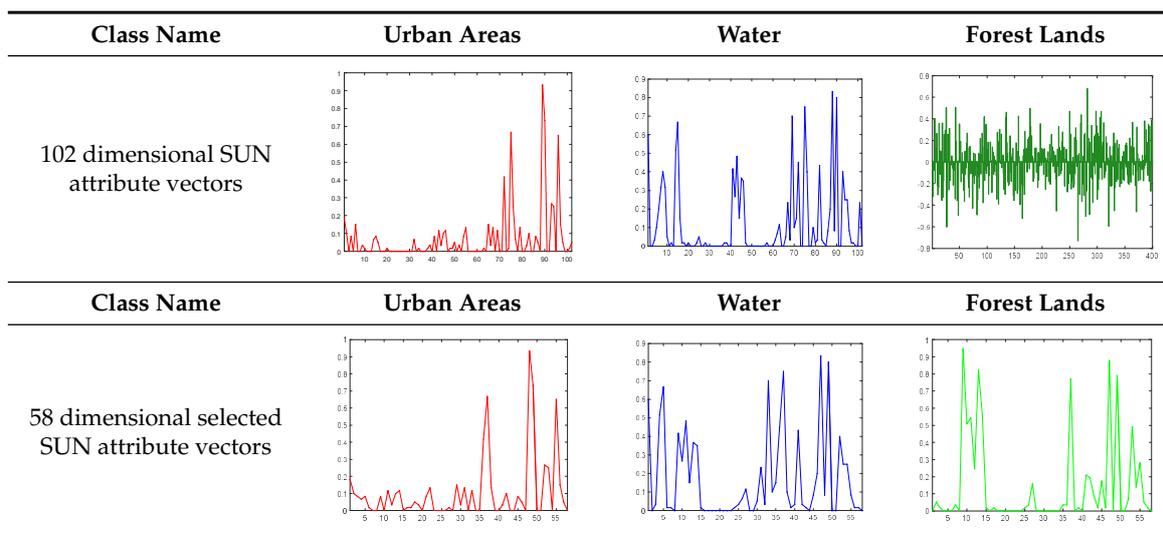
than one vector, for the same scene, we use the mean of all vectors to represent the scene category attribute vectors.

Table 1. The examples of the land cover class and the corresponding word vectors.



In detail, some attributes in the 102-dimensional attribute vectors, including ‘natural, man-made, glossy, matte, sterile,’ etc., are desirable attributes for the land cover classes’ descriptions. However, the attributes including ‘fire, sunny, rusty, warm, scary,’ etc., are too different in the actual properties of the land cover classes. Thus, we subjectively selected 58-dimensional attribute vectors, which are more suitable for describing the actual properties of the land cover classes from the 102-dimensional original SUN attribute vectors. Some examples of the typical land cover classes and corresponding SUN and selected SUN attribute vectors have been illustrated in Table 2. In Tables 1 and 2, the applied semantic information represented by word vectors and SUN attribute vectors are both quantified, which can represent typical word-based semantic vectors and attribute-based semantic vectors, respectively.

Table 2. The examples of the land cover class and corresponding SUN/selected SUN attribute vectors.



Therefore, we can obtain the word vector matrix S_{wv} , SUN attributes vector matrix S_{sa} , and the selected SUN attribute vector matrix S_{ssa} about the typical land cover classes in the 8-m resolution PolSAR images. The size of S_{wv} is $n \times 400$; n is the class number in the training and testing of PolSAR

data samples, including the seen and unseen classes. The size of \mathbf{S}_{sa} is $n \times 102$, the size of \mathbf{S}_{ssa} is $n \times 58$. Then, we further construct semantic relevance \mathbf{G} between the attributes, \mathbf{G}_{wv} for \mathbf{S}_{wv} , \mathbf{G}_{sa} for \mathbf{S}_{sa} , \mathbf{G}_{ssa} for \mathbf{S}_{ssa} , correspondingly. For example:

$$\mathbf{S} = [s_1; s_2; \cdots; s_i; \cdots; s_n] \quad (2)$$

$$s'_i = s_i / \|s_i\|_2, \quad \mathbf{S}_N = [s'_1; s'_2; \cdots; s'_i; \cdots; s'_n] \quad (3)$$

$$\mathbf{G}_N = \mathbf{S}_N^T \mathbf{S}_N, \quad \mathbf{G}_N = [g_1; g_2; \cdots; g_i; \cdots; g_n] \quad (4)$$

$$g'_i = g_i / \|g_i\|_2, \quad \mathbf{G} = [g'_1; g'_2; \cdots; g'_i; \cdots; g'_n] \quad (5)$$

In Equation (2), s_i is the attributes vector of a land cover class. From Equations (2)–(5), \mathbf{G}_{wv} , \mathbf{G}_{sa} , and \mathbf{G}_{ssa} can be obtained. The size of \mathbf{G} is $n \times n$ and it represents the semantic relationship between land cover. The semantic relevance \mathbf{G} will be applied to take part in determining whether the test sample is a seen or unseen class.

3.3. Generalized Zero-Shot Learning with Semantic Relevance

As the above-mentioned ZSL and GZSL framework in Section 2.1, how to represent the projection relationship \mathbf{W} between attributes with image features is a key for ZSL and GZSL. Here we employed the latent embedding model [37,48,49] to obtain \mathbf{W} ; latent embedding is a non-trivial extension of structured joint embedding [50] (SJE, the objective used for learning \mathbf{W} in SJE is similar to that proposed for the structured SVM parameter learning). Instead of learning a single mapping transformation in SJE, latent embedding learns a piecewise linear compatibility function of K parameter matrices \mathbf{W}_i ($i = 1, \dots, K, K \geq 2$). Latent embedding ZSL applies a ranking based objective to learn the model using an efficient and scalable stochastic gradient descent based solver [37].

For a typical ZSL task, the training data $D = \{(x_n, y_n)\}_{n=1}^N$ with the labels y_n from the label space of the seen classes $C_S = \{1, 2, \dots, p\}$, x_n represents the image features of the sample. We denote $C_U = \{p+1, p+2, \dots, p+q\}$, the label space of unseen the classes ($C_S \cap C_U = \emptyset$). The main goal of ZSL is to classify the test data into the unseen classes, assuming the absence of the seen classes in the test stage. In other words, each test data is assumed to come from, and will be assigned to, one of the labels in C_U . Given a test instance x (image features), conventional latent embedding ZSL will be labeled as the class whose semantic representation maximizes the following Equations (6) and (7). The inference in ZSL is usually based on the nearest neighbor strategy. It has been reported that latent embedding was state-of-the-art in ZSL for the benchmark datasets in computer vision [37,48,49].

$$F_c(x) = \max_{1 \leq i \leq K} x^T \mathbf{W}_i \mathbf{S}_{Te} \quad (6)$$

$$\hat{y} = \operatorname{argmax}_{c \in C_u} F_c(x) \quad (7)$$

In Equation (6), \mathbf{S}_{Te} is the semantic vector matrix of the test classes, only the unseen classes in ZSL. $F_c(x)$ means the discriminant scoring function. \hat{y} represents the predicted label. However, it has been proved that if the above ZSL classification prediction process in Equations (6) and (7) has been simply employed in the GZSL setting, the accuracies of most ZSL approaches drop significantly in this setting [33]. That is, nearly all test data from the unseen classes are misclassified into the seen classes when the ZSL approaches are applied with the test samples containing both seen and unseen classes. Thus, to realize GZSL in the proposed framework, the label prediction process must be improved.

There has been very little work on generalized zero-shot learning (GZSL) [24,33,34], but GZSL is more attractive and practical in the PolSAR interpretation application. GZSL can provide a new way for PolSAR image land cover classification by utilizing the rich polarization features and semantic information in the PolSAR imagery to identify both the unseen and seen classes via leveraging semantic information. In the existing GZSL model proposed in [33], the main idea was to introduce a calibration

factor to calibrate the classifiers for both seen and unseen classes. That model has been tested on benchmark datasets, including the AWA and CUB. This means that the inference stage is important for GZSL.

Here the semantic relevance obtained from Section 3.2 has been employed to the inference progress of the proposed GZSL framework. After the projection relationship \mathbf{W} between the semantic attributes with polarimetric features is established by means of the latent embedding in the training stage. GZSL test samples include not only the unseen classes, but also the seen classes. In the proposed GZSL framework for PolSAR land cover classification, \mathbf{S}_{Te} includes the seen and unseen classes' semantic vectors. The above semantic relevance \mathbf{G} was introduced to constraint and amend the GZSL classification prediction process. The union of the seen classes C_S and unseen classes C_U are represented by $SU = C_S \cup C_U$.

$$F_c = \{f_1, f_2, \dots, f_p, f_{p+1}, \dots, f_{p+q}\}, c \in SU \quad (8)$$

$$M_c = \underset{c \in SU}{\operatorname{argmax}} \{f_i - g_{ij}, f_j\}, \text{ s.t., } \begin{cases} f_i = \max\{f_1, f_2, \dots, f_p, f_{p+1}, \dots, f_{p+q}\}, i \in \{1, 2, \dots, p\} \\ f_j = \max\{f_{p+1}, f_{p+2}, \dots, f_{p+q}\}, j \in \{p+1, p+2, \dots, p+q\} \\ f_i \geq f_j \\ g_{ij} = \mathbf{G}(i, j) \end{cases} \quad (9)$$

$$\hat{y} = \begin{cases} M_c \\ \underset{c \in SU}{\operatorname{argmax}} F_c, & \text{otherwise} \end{cases} \quad (10)$$

Given a test instance x which represents the mid-level polarization feature in the proposed framework, we can get a preliminary discriminant score from Equation (6). For every test instance, the preliminary discriminant score vector has $p + q$ elements, as shown in Equation (8). Then, the final label of the test sample can be obtained from the classification rule in Equations (9) and (10). The corresponding semantic relevance g_{ij} is an amendment to reduce the scores for the seen classes.

3.4. GZSL For PolSAR Land Cover Classification

This paper is aimed at recognizing both seen and unseen instances in PolSAR images by applying the semantic information, as illustrated in Figure 2. The GZSL is constructed by means of land cover class semantic attribute descriptions, the relationship between semantic attributes, and the projection between the image feature layer, intermediate semantic information layer, and the class label layer. In Section 3.1, the effective mid-level polarization feature is first extracted. Then, as in Section 3.2, the semantic attributes from the Word2Vec semantic vectors, SUN attributes, and the selected SUN attributes are collected and analyzed to describe the characteristics of the PolSAR typical classes, and the semantic relevance \mathbf{G} between the attributes is obtained. Moreover, the projection relationship \mathbf{W} between the mid-level representation of the PolSAR data samples and class attributes are established by latent embedding during the training stage. Finally, for every test instance, through the GZSL model constructed by polarization feature representation, projection relationship \mathbf{W} and semantic relevance \mathbf{G} , the labels of the test instances can be predicted, even though some test classes do not have training samples. By utilizing the rich polarization features and semantic information in the PolSAR imagery, the proposed GZSL can provide a more practical solution for PolSAR interpretation to recognize some new land cover categories without labeled samples, while the conventional supervised approaches always fail to categorize the unseen instances. To a certain extent, GZSL can reduce the requirement for sample labeling and make the framework has the ability to identify the new types in PolSAR land cover classification.

It should be noted that the selected unseen class should correspond with some semantically related seen samples, e.g., when the unseen class is croplands, there should be some vegetation classes in the seen samples; when the unseen class is building areas, there should be some man-made classes in the seen samples. The semantic relation is the base of the ZSL and GZSL ability, and these can

avoid the projection domain shift problem [26,51] to some degree. This phenomenon will be further demonstrated by experiments in Sections 4 and 5.

4. Experimental Results

In this section, we first introduce the experimental data and then present the classification results obtained by applying the proposed method from the RadarSAT-2 fully-polarimetric SAR imagery. Moreover, the evaluation of the experimental results has been shown.

4.1. Experimental Data and the Settings

The effectiveness of the proposed method has been tested on three RadarSAT-2 PolSAR datasets (C-band at fine quad-pol mode, with a resolution of 8 m). The selected data contain more types of land cover classes and the orientations of the building areas are also more complicated; higher requirements are put forward for the classification algorithm. In total, the experimental data contain seven kinds of typical land cover classes. The basic information and the GZSL unseen/seen ratio of the selected experimental data are shown in Table 3. All the samples from PolSAR data are 50×50 pixels patches, and the training samples are selected randomly.

Table 3. The data description of the employed PolSAR data.

Data	Flevoland Data	Wuhan Data1	Wuhan Data2
Imaging time	2008	2011–12	2011–12
Image sizes (pixels)	1400×1200	5500×2400	5500×3500
Land cover classes	4 classes: c1, c3, c4, c5 *	4 classes: c1, c2, c3, c4	6 classes: c1, c2, c3, c4, c6, c7,
Ground truth	Available (4 classes) Seen: c1, c3, c4/ c1, c3, c5/ c1, c4, c5/ c3, c4, c5, unseen: c5/ c4/ c3/ c1	No ground truth	No ground truth
Seen/unseen		Seen: c1, c3, c4, unseen: c2	Seen: c1, c3, c4, unseen: c2, c6, c7,
Training samples	300	300	300
Test samples	15,776	20,805	30,441

* Urban areas (c1), rural areas (c2), water (c3), forest lands (c4), croplands (c5), wetland (c6), and agricultural land (c7).

4.2. Results and Evaluation of the Flevoland Data

For the RadarSAT-2 Flevoland data, as shown in Figure 4, the training samples include urban areas, water, and forest lands; the testing samples include urban areas, water, forest lands, and croplands; that is, the unseen class is croplands. The PolSAR data is first divided into 50×50 pixel patches with an 80% overlapping rate; that is, the number test samples of the Flevoland data is 15,776. Additionally, the sample number for each training class is 100 and all the three kinds of labeled samples are selected randomly. The four kinds of test class results obtained by the proposed GZSL method have been shown in Figure 4, the accuracy has been evaluated in Table 4, and the adapted semantic attributes from the Word2Vec semantic vectors, SUN attributes, and the selected SUN attributes have been evaluated, too.

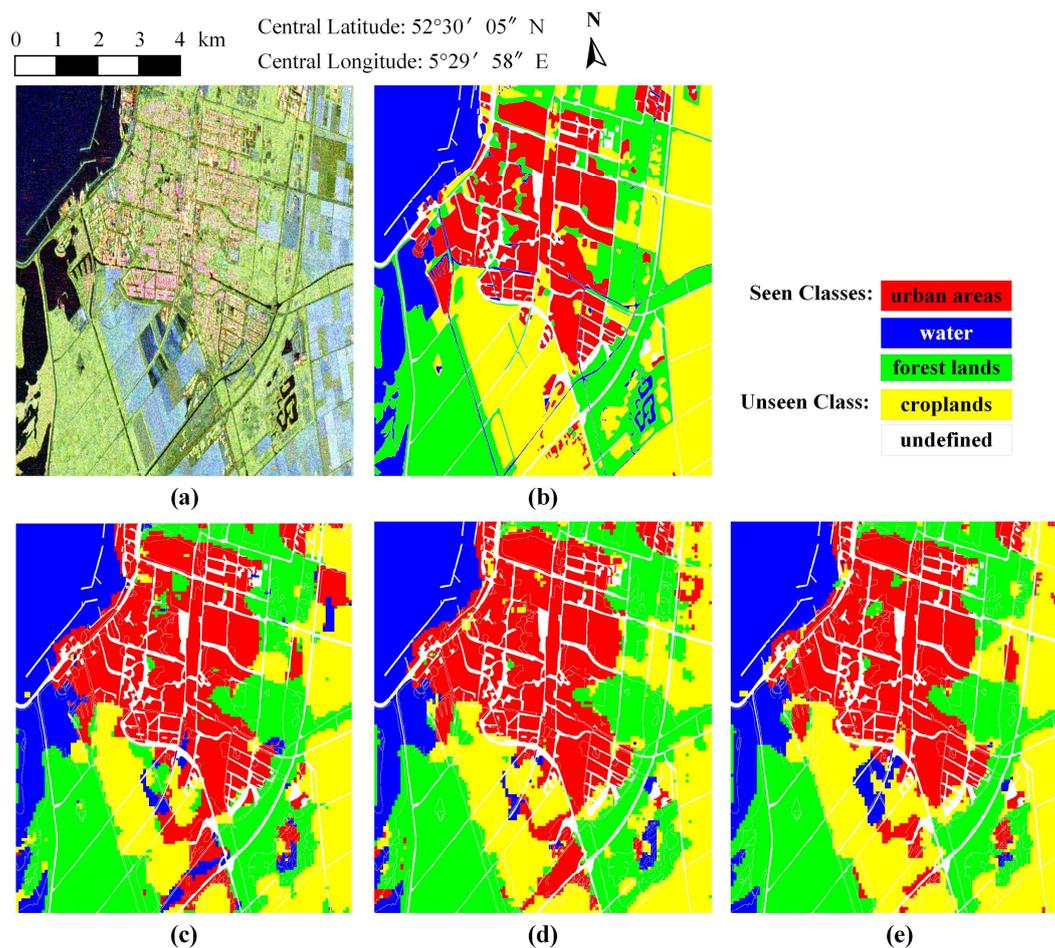


Figure 4. The GZSL classification results of the Flevoland data (the unseen class is croplands). (a) Pauli RGB image of the Flevoland data; (b) the ground truth map of the Flevoland data; (c) the results of the Word2Vec attributes applied; (d) the results with SUN attributes applied; and (e) the results with the selected SUN attributes applied.

In Figure 4, when the training classes are urban areas, water, forest lands, and the test classes are urban areas, water, forest lands, and croplands, the GZSL results applied Word2Vec attributes, SUN attributes, and the selected SUN attributes are shown in Figure 4c–e, correspondingly. Additionally, the proposed method can recognize the unseen class with an accuracy of about 60.76%, 68.72%, and 73.93% through the applied Word2Vec attributes, SUN attributes, and the selected SUN attributes, accordingly, without the training cropland samples. Furthermore, the overall accuracies of the proposed GZSL framework in the Flevoland data are all above 74.5% when the unseen/seen ratio is 1/3; that is, the training/testing classes' ratio is 3/4. Among the employed attributes, the results in Figure 4 and Table 4 show the best effect are the selected SUN attributes, with an overall accuracy of about 78%. We repeated experiments in Figure 4e ten times, the mean and standard deviation of the unseen class accuracy were 73.07% and 0.015, and the mean and standard deviation of the overall accuracy were 77.56% and 0.0075.

For the experimental settings in Figure 4, the semantic relevance between the unseen class (croplands in Figure 4) and the forest lands in the seen classes is of great semantic similarity. In order to further verify the semantic relevance and domain shift problem described in Section 3, the following experiments set the urban areas, water, and forest lands as the unseen class, one by one, and then used the other three classes of samples to obtain the information of the four land cover classes' information

and to make corresponding quantitative evaluations. All the GZSL results in Figure 5 applied the selected SUN attributes.

Table 4. The evaluation of GZSL classification results in the Flevoland data (%) (the unseen class is cropland).

		Seen			Unseen
		urban areas	water	forest lands	croplands
Word2Vec attributes	urban areas	87.09	0.87	7.95	4.09
	water	8.12	82.43	7.25	2.20
	forest lands	8.89	5.30	74.25	11.57
	croplands	18.86	4.87	15.51	60.76
Overall accuracy (OA): 74.52					
		urban areas	water	forest lands	croplands
SUN attributes	urban areas	90.01	0.63	4.11	5.26
	water	8.75	79.52	8.72	3.01
	forest lands	9.54	4.55	76.14	9.77
	croplands	14.82	1.23	15.22	68.72
Overall accuracy (OA): 77.59					
		urban areas	water	forest lands	croplands
Selected SUN attributes	urban areas	87.13	0.65	6.16	6.06
	water	6.48	79.14	10.76	3.36
	forest lands	8.07	3.66	74.71	13.56
	croplands	11.52	3.75	10.79	73.93
Overall accuracy (OA): 78.04					

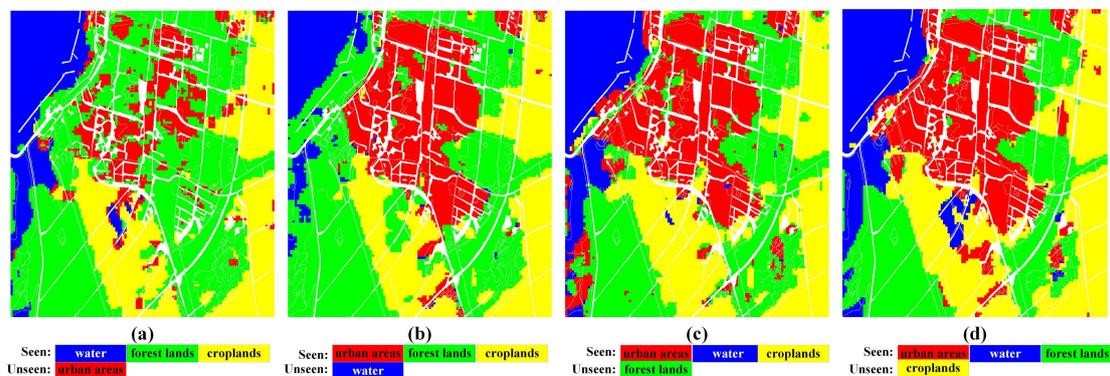


Figure 5. The GZSL classification results of the Flevoland data with different unseen classes. (a) The results of the unseen class with urban areas; (b) the results of the unseen class with water; (c) the results of the unseen class with forest lands; and (d) the results of the unseen class with croplands.

From Figure 5a–d, the classification results have been obtained with the unseen class of urban areas, water, forest lands, and croplands, correspondingly. As illustrated in Figure 5a, the training classes are water, forest lands, and croplands, and the test classes are urban areas, water, forest lands, and croplands. The accuracy has been evaluated in Table 5. As shown in Figure 5 and Table 5, it can be found that when the unseen classes are urban areas and water, the overall classification accuracies are low: 68.53% and 72.65%, respectively. Additionally, when the unseen classes are forest lands and croplands, the overall classification accuracies are higher: 77.42% and 78.04%, respectively. These are due to the high semantic relevance between forest lands and croplands, which can be distinguished based on the distance relationships described in Section 3.4; while the semantic relevance between urban areas with other land cover classes in the Flevoland data is poor, so are those of the

water with other land cover classes. Since ZSL and GZSL depend on the semantic relevance to identify unseen classes, if the semantic relevance between the seen classes and unseen classes are poor, the above-mentioned domain shift problem [26,51] tends to appear, making it difficult to identify the unseen classes with poor semantic relevance accurately.

Table 5. The evaluation of GZSL classification results in Flevoland data with different unseen class (%).

Seen		Unseen		Overall Accuracy(OA)
water 72.08	forest lands 86.62	croplands 70.04	urban areas 39.42	
urban areas 79.63	forest lands 83.91	croplands 71.85	water 47.55	72.65
urban areas 82.39	water 73.64	croplands 78.26	forest lands 75.18	77.42
urban areas 87.13	water 79.14	forest lands 74.71	croplands 73.93	78.04

4.3. Results of the Wuhan Data1

For the RadarSAT-2 Wuhan Data1, as shown in Figure 6a, the training samples include urban areas, water, and forest lands; the testing samples include urban areas, water, forest lands, and rural areas. That is, the unseen class are the rural areas. The PolSAR data are first divided into 50×50 pixel patches with a 50% overlapping rate, the number of test samples in Wuhan Data1 is 20,805. Additionally, the sample number for each training class is 100 and all the three kinds of the labeled samples are selected randomly. The results obtained by the proposed GZSL method have been shown in Figure 6b–d, and the adapted semantic attributes from the Word2Vec semantic vectors, the SUN attributes, and the selected SUN attributes have been shown accordingly.

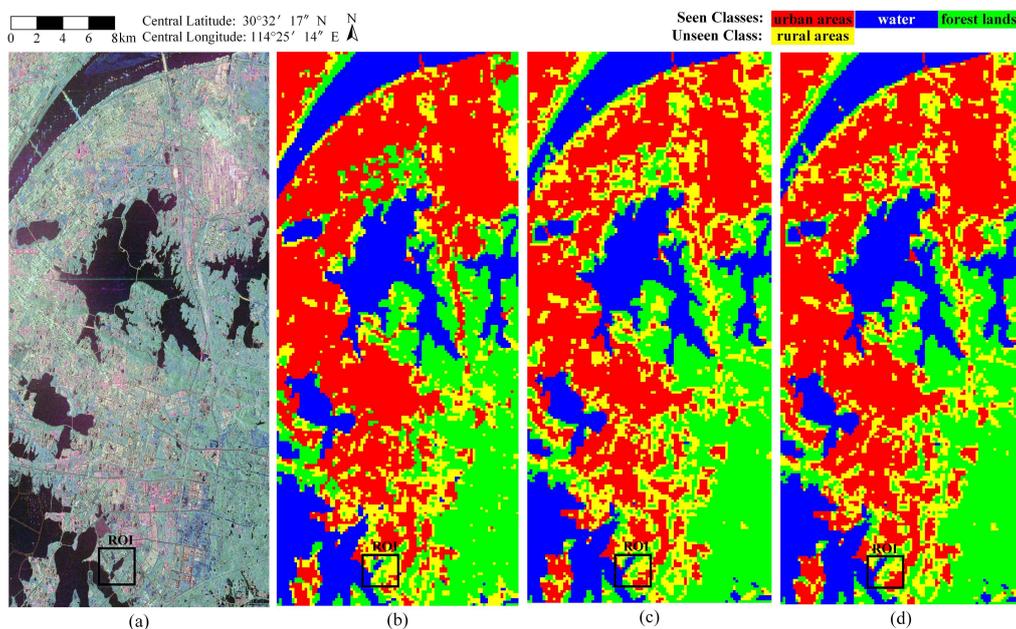


Figure 6. The Wuhan Data1 and the GZSL classification results. (a) The Pauli RGB image of the Wuhan Data1; (b) the results with the Word2Vec attributes applied; (c) the results with the SUN attributes applied; and (d) the results with the selected SUN attributes applied.

For the experiment with the Wuhan Data1, we set the unseen class as rural areas; that is, there are no training samples with rural areas. However, from the GZSL results in Figure 6b–d, it can be seen that the rural areas' information can also be classified. In detail, it is easy to find that the rural areas in Figure 6c,d are larger than the results in Figure 6b. This is mainly due to the semantic difference. The SUN attribute database does not have a category that strictly corresponds to the rural areas, so we use a combination of attributes from 'townhouse' and 'village' to represent the semantic information of the rural areas.

In order to illustrate the effectiveness of the classification of the rural areas' information in Figure 6, we chose an ROI for verification. This ROI correspond to Canglongdao in the Jiangxia district, a suburb of Wuhan. Additionally, the corresponding ROI optical images from Google Earth have been illustrated in Figure 7b. There are a large number of low and small buildings in this ROI, and the buildings are sparse, which are different from the urban areas. The results in Figure 7c–e reflect the semantic relevance between the urban areas and rural areas in the Word2Vec attributes space, the SUN attributes spaces, and the selected SUN attribute spaces.

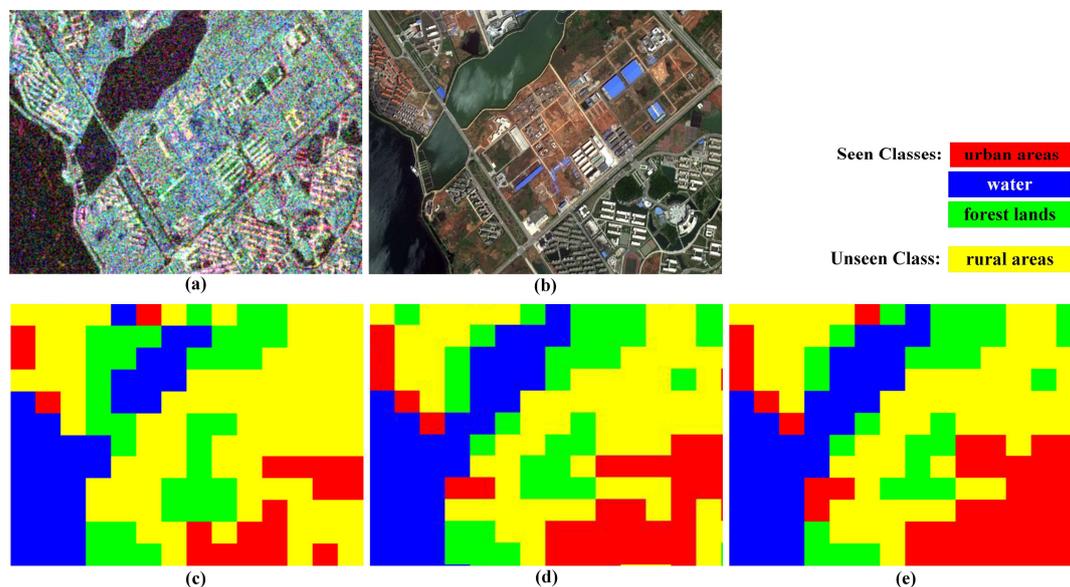


Figure 7. The ROI in Figure 6 and the GZSL classification results. (a) The Pauli image of ROI; (b) the corresponding Google Earth optical image of ROI; (c) the results with the Word2Vec attributes applied; (d) the results with the SUN attributes applied; and (e) the results with the selected SUN attributes applied.

The results in Sections 4.2 and 4.3 show that the proposed GZSL framework can classify some new land cover categories without labeled samples by using the semantic information between land cover categories and effective polarization features. The experimental results also illustrate the potential of semantic information for PolSAR land cover classification, and the GZSL mechanism can reduce the requirement for sample labeling to a certain extent. Moreover, the effectiveness of adapted semantic attributes about Word2Vec semantic vectors, SUN attributes, and the selected SUN attributes have been evaluated and compared. From the quantitative evaluation of Flevoland data in Table 4, the SUN attributes are slightly better than the Word2Vec semantic vectors under the same experimental conditions. This may be due to the fact that the SUN attributes contain more spatial and surface properties, which is consistent with the characteristics of remote sensing images. The results in Table 5 prove that the semantic relation is the base of the GZSL ability.

5. Discussion

Since the land cover classes in the Flevoland data and Wuhan Data1 are not very numerous, the unseen class in the above experiment is one class; that is, the ratio of the training classes and testing classes is 3/4. To further illustrate the effectiveness of the proposed GZSL method, we implemented the following experiment on Wuhan Data2, as shown in Figure 8a. There are more rich land cover classes in Wuhan Data2. The land cover classes in this data mainly consist of six types: urban areas, water, forest lands, rural areas, wetland, and agricultural land.

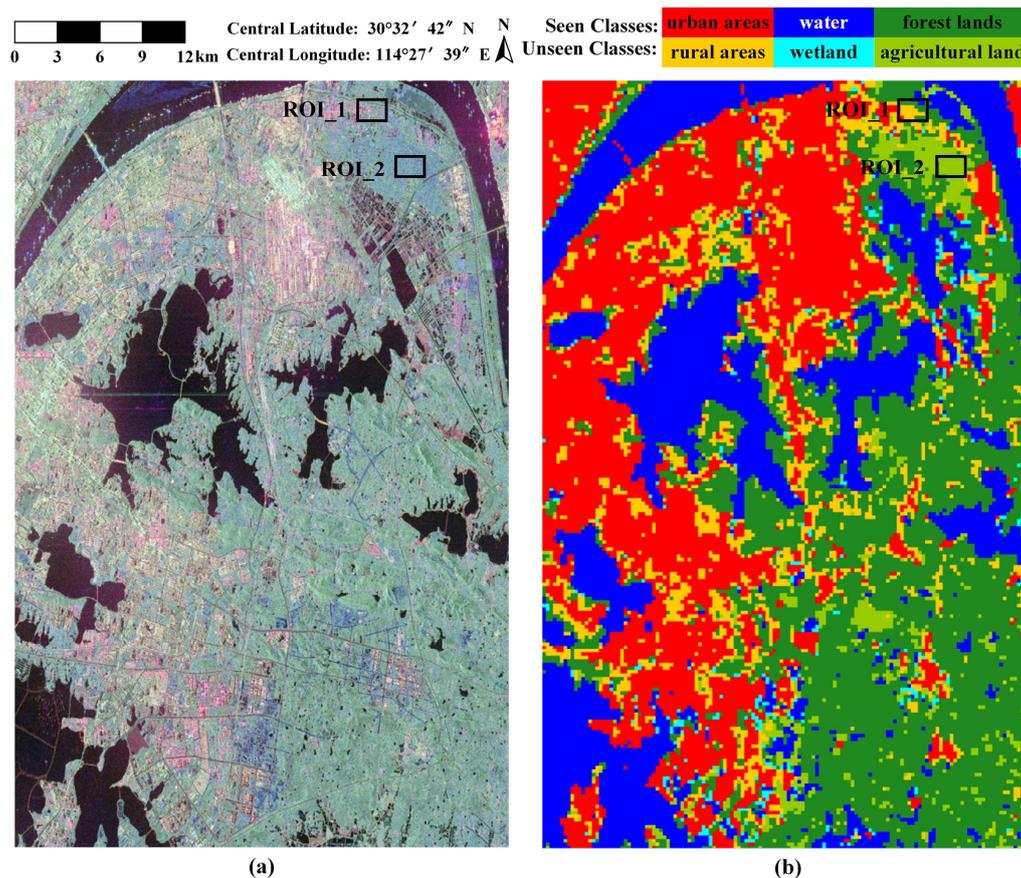


Figure 8. The Wuhan Data2 and the GZSL classification results. (a) Pauli RGB image of Wuhan Data2, and (b) the GZSL result with the Word2Vec attributes applied.

For the RadarSAT-2 Wuhan Data2, the training samples include urban areas, water, and forest lands, and the testing samples include all six types mentioned above; that is, the unseen classes are the rural areas, wetland, and agricultural land. The PolSAR data are first divided into 50×50 pixel patches with a 50% overlapping rate, the number of test samples in Wuhan Data2 is 30,441. Additionally, the sample number for each training class is 100 and all the three kinds of the labeled samples are selected randomly. These three kinds of samples in the Wuhan Data2 can be the same as the Wuhan Data1 samples. The six kinds of test classes' results obtained by the proposed GZSL method has been shown in Figure 8b. Additionally, the adapted semantic attributes mainly come from the Word2Vec model, the main reason was that there was no exact category with attributes in the SUN attributes database on 'wetland' and 'agricultural land', and the 'rural areas' attributes in the SUN attributes database were not exactly the same.

From Figure 8b, it can be seen that the rural areas and agricultural land were basically recognized simultaneously, although neither of these classes have training samples. ROI_1 and ROI_2 correspond to the rural areas and agricultural land (most of these lands contain some low

vegetation). The corresponding ROI optical images from Google Earth have been illustrated in Figure 9a,c. Typical field investigation ROI images in the selected ROIs have been illustrated in Figure 9b,d. Combined with the provided reference images in Figure 9, it can be seen that the ROI_1 and ROI_2 classification results in Figure 8b have a certain degree of effectiveness. Moreover, the ratio of the training classes and testing classes is 3/6, which also shows that the proposed method is effective and has certain extensibility. In the above experiments, there is a certain semantic relevance between the unseen classes and the seen classes, such as between the rural areas and the urban areas, between wetland and water. Thus, GZSL can classify some new land cover categories without the labeled samples, even though the ratio of the training classes and test classes is 3:6.

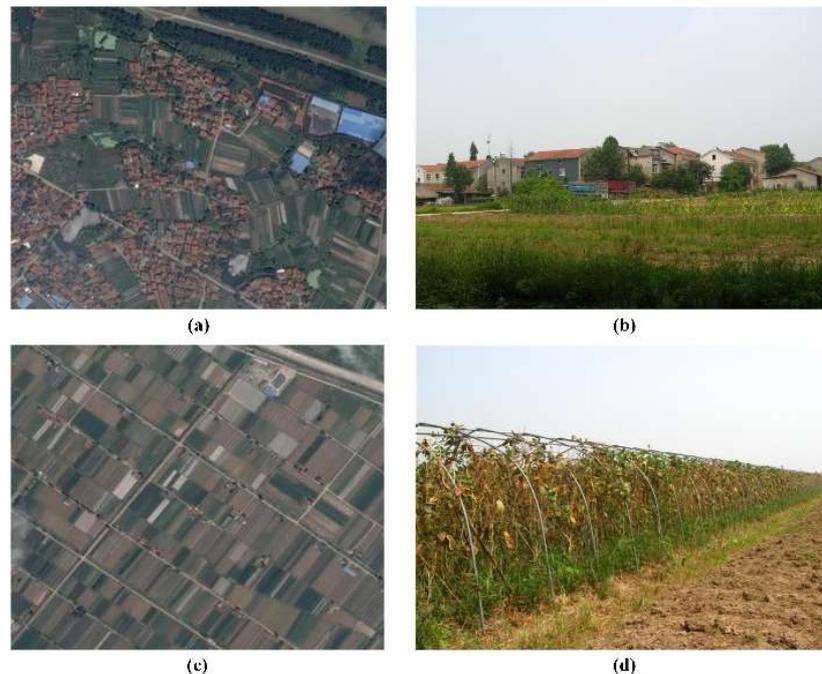


Figure 9. The ROIs in Figure 8 and field investigation images. (a) Corresponding Google Earth optical image of ROI_1; (b) the field investigation images in ROI_1 (latitude: $30^{\circ}39'48''$, longitude: $114^{\circ}29'13''$); (c) corresponding Google Earth optical image of ROI_2; and (d) the field investigation images in ROI_2 (latitude: $30^{\circ}39'06''$, longitude: $114^{\circ}30'54''$).

Experiments on the above three RadarSAT-2 PolSAR datasets show that the proposed method can classify 4–6 testing classes with only three training classes. That is, the ratio of unseen/seen classes can be 1/3–3/3. There are few works on ZSL application in the remote sensing interpretation, and GZSL for PolSAR land cover classification has not been studied in the remote sensing literature, to the best of our knowledge. We introduce the unseen/seen class ratio settings in other related works (ZSL applied in remote sensing interpretation) to make a comparative analysis. For the ZSL target recognition problem in the SAR image in [32], the ratio of unseen/seen classes is 1/7, that is, seven classes for training and only one class for testing. For the ZSL classification problem in the high spatial resolution image in [22], the average accuracy was about 58% with an unseen/seen ratio of 5/16. Another ZSL problem in [21] is the fine-grained street tree recognition in the aerial data; the recognition accuracy was about 14.3% with an unseen/seen ratio of 16/24. In our work, combining the actual basic requirements of the PolSAR interpretation, the unseen/seen ratio was 1/3–3/3, that is, the training/testing classes' ratio is 3/4–3/6. Due to the small amount of training samples selected in the experiments, less than 1.9% (300/15,996, 300/20,805, 300/30,441) of the samples were randomly selected as training samples, the accuracy of our experimental results is not very high at present. However, we think that our method is feasible and effective in the application of classifying some new land cover categories without labeled

samples, for PolSAR imagery interpretation with about 8-m resolution. The GZSL framework can reduce the requirement for sample labeling and give the framework the ability to identify new types in PolSAR land cover classification.

In subsequent research work, several topics need to be further studied. Firstly, a more professional semantic description about the land cover classes or targets in PolSAR imagery should be analyzed, including the scattering characteristic, resolution, polarization mode, incident angle, seasonal, and other information. Secondly, the semantic modeling methods and tools for the aforementioned semantic description of the PolSAR land cover classes or targets need to be further developed. A potential and promising semantic modeling method is the ontological semantic model [4,27], and this will be the focus of our research work in the next step. Thirdly, as an important topic in computer vision research, the ability of ZSL or GZSL has not yet been standardized; that is, there is no agreed upon ZSL or GZSL benchmark [24,33]. Thus, the potential application of ZSL and GZSL can be further explored for SAR image interpretation.

6. Conclusions

For PolSAR land cover classification, it is a highly probable scenario that new land cover categories can be introduced after the training stage, or that no training examples available for several rare and interesting classes. Inspired by generalized zero-shot learning (GZSL), which can categorize instances from both seen and previously unseen classes, this paper studies the problem of classifying both the unseen and seen land cover classes' information from the PolSAR image under a semantic expressed GZSL framework. By leveraging the rich semantic relevance between land cover attributes in the PolSAR imagery, the semantic relevance between attributes is first obtained to relate unseen and seen classes. Then, the projection relationship between the effective mid-level polarization features and class attributes is established by latent embedding during training. Finally, for every test instance, through the GZSL model constructed by the mid-level polarization feature, projection relationship, and semantic relevance, the labels of the test instances can be predicted, even though some test instances do not have training samples backing them. The quantitative and qualitative evaluation of experiments on the three RadarSAT-2 datasets have shown that the classification accuracy of an unseen class is about 73% if there are some semantically-related seen classes in the training stage. Additionally, the proposed method can classify 4–6 testing classes with only three training classes. This GZSL framework can reduce the requirement for sample labeling giving the framework the ability to identify new types in PolSAR land cover classifications. Moreover, three kinds of land cover class attributes, which include the Word2Vec semantic vectors, SUN attributes, and the selected SUN attributes, have been applied and evaluated.

Since the currently employed semantic attributes lack some polarimetric semantic expressions, including the resolution, scattering characteristic, incident angle, season information, and so on, the classification ability of the proposed GZSL method is still relatively conservative. In the future, we expect to continually improve and develop our research on semantic modeling of the land cover or targets in SAR images.

Author Contributions: All authors contributed to forming the general idea of the paper, and helped conceive and design the experiments. R.G. created the research design, performed the experiments, analyzed the data, and wrote the draft; X.X. conducted the coordination of the research activities and provided critical comments to improve the paper; L.W. helped edit the draft and contributed to developing the PolSAR feature representation algorithm; R.Y. contributed to the accuracy assessment and manuscript writing; and F.P. helped propose and develop the semantic information model.

Funding: This research was funded by the Technology Research and Development of the Major Project of High-Resolution Earth Observation System under grant 03-Y20A10-9001-15/16.

Acknowledgments: The authors would like to thank the reviewers and editors who provided valuable comments and suggestions for this article.

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

References

1. Atwood, D.; Small, D.; Gens, R. Improving PolSAR Land Cover Classification with Radiometric Correction of the Coherency Matrix. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 848–856. [[CrossRef](#)]
2. Wang, L.; Xu, X.; Dong, H.; Gui, R.; Pu, F. Multi-pixel Simultaneous Classification of PolSAR Image Using Convolutional Neural Networks. *Sensors* **2018**, *18*, 769. [[CrossRef](#)] [[PubMed](#)]
3. Sato, M.; Chen, S.; Satake, M. Polarimetric SAR Analysis of Tsunami Damage Following the March 11, 2011 East Japan Earthquake. *Proc. IEEE* **2012**, *100*, 2861–2875. [[CrossRef](#)]
4. Gui, R.; Xu, X.; Dong, H.; Song, C.; Pu, F. Individual building extraction from TerraSAR-X images based on ontological semantic analysis. *Remote Sens.* **2016**, *8*, 708. [[CrossRef](#)]
5. Yang, W.; Yin, X.; Song, H.; Liu, Y.; Xu, X. Extraction of built-up areas from fully polarimetric SAR imagery via PU learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 1207–1216. [[CrossRef](#)]
6. Deng, L.; Yan, Y.; Sun, C. Use of Sub-Aperture Decomposition for Supervised PolSAR Classification in Urban Area. *Remote Sens.* **2015**, *7*, 1380–1396. [[CrossRef](#)]
7. Xiang, D.; Tang, T.; Ban, Y.; Su, Y.; Kuang, G. Unsupervised polarimetric SAR urban area classification based on model-based decomposition with cross scattering. *ISPRS J. Photogramm. Remote Sens.* **2016**, *116*, 86–100. [[CrossRef](#)]
8. Dong, H.; Xu, X.; Wang, L.; Pu, F. Gaofen-3 PolSAR Image Classification via XGBoost and Polarimetric Spatial Information. *Sensors* **2018**, *18*, 611. [[CrossRef](#)] [[PubMed](#)]
9. Peng, Y.; Chen, J.; Xu, X.; Pu, F. SAR Images Statistical Modeling and Classification based on the Mixture of Alpha-stable Distributions. *Remote Sens.* **2013**, *5*, 2145–2163. [[CrossRef](#)]
10. Lee, J.S.; Grunes, M.R.; Kwok, R. Classification of multi-look polarimetric SAR imagery based on complex Wishart distribution. *Int. J. Remote Sens.* **1994**, *15*, 2299–2311. [[CrossRef](#)]
11. Gao, W.; Yang, J.; Ma, W. Land cover classification for polarimetric SAR images based on mixture models. *Remote Sens.* **2014**, *6*, 3770–3790. [[CrossRef](#)]
12. Freeman, A.; Durden, S.L. A three-component scattering model to describe polarimetric SAR data. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 963–973. [[CrossRef](#)]
13. Yamaguchi, Y.; Sato, A.; Boerner, W.; Sato, R.; Yamada, H. Four-Component Scattering Power Decomposition with Rotation of Coherency Matrix. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 2251–2258. [[CrossRef](#)]
14. Tao, C.; Chen, S.; Li, Y.; Xiao, S. PolSAR land cover classification based on roll-invariant and selected hidden polarimetric features in the rotation domain. *Remote Sens.* **2017**, *9*, 660. [[CrossRef](#)]
15. Mahdianpari, M.; Salehi, B.; Mohammadimanesh, F.; Motagh, M. Random forest wetland classification using ALOS-2 L-band, RADARSAT-2 C-band, and TerraSAR-X imagery. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 13–31. [[CrossRef](#)]
16. Hänsch, R.; Hellwich, O. Skipping the real world: Classification of PolSAR images without explicit feature extraction. *ISPRS J. Photogramm. Remote Sens.* **2018**, *140*, 122–132. [[CrossRef](#)]
17. Zhao, Z.; Jiao, L.; Zhao, J.; Gu, J.; Zhao, J. Discriminant deep belief network for high-resolution SAR image classification. *Pattern Recognit.* **2017**, *61*, 686–701. [[CrossRef](#)]
18. Zhou, Y.; Wang, H.; Xu, F.; Jin, Y. Polarimetric SAR Image Classification Using Deep Convolutional Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1935–1939. [[CrossRef](#)]
19. De, S.; Bhattacharya, A. Urban classification using PolSAR data and deep learning. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; pp. 353–356.
20. Akata, Z. *Contributions to Large-Scale Learning for Image Classification*; Université De Grenoble: Grenoble, France, 2014.
21. Sumbul, G.; Cinbis, R.; Aksoy, S. Fine-Grained Object Recognition and Zero-Shot Learning in Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 770–779. [[CrossRef](#)]
22. Li, A.; Lu, Z.; Wang, L.; Xiang, T.; Wen, J. Zero-Shot Scene Classification for High Spatial Resolution Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4157–4167. [[CrossRef](#)]
23. Ding, Y.; Li, Y.; Yu, W. Learning from label proportions for SAR image classification. *EURASIP J. Adv. Signal Process.* **2017**, *41*, 1–12. [[CrossRef](#)]

24. Xian, Y.; Schiele, B.; Akata, Z. Zero-Shot Learning—The Good, the Bad and the Ugly. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 3077–3086.
25. Palatucci, M.; Pomerleau, D.; Hinton, G.; Mitchell, T. Zero-shot learning with semantic output codes. In Proceedings of the International Conference on Neural Information Processing Systems (NIPS), Bangkok, Thailand, 1–5 December 2009; pp. 1410–1418.
26. Kodirov, E.; Xiang, T.; Gong, S. Semantic Autoencoder for Zero-Shot Learning. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4447–4456.
27. Liu, B.; Yao, L.; Ding, Z.; Xu, J.; Wu, J. Combining Ontology and Reinforcement Learning for Zero-Shot Classification. *Knowl. Based Syst.* **2017**, *144*, 42–50. [[CrossRef](#)]
28. Luo, C.; Li, Z.; Huang, K.; Feng, J.; Wang, M. Zero-Shot Learning via Attribute Regression and Class Prototype Rectification. *IEEE Trans. Image Process.* **2018**, *27*, 637–648. [[CrossRef](#)] [[PubMed](#)]
29. Morgado, P.; Vasconcelos, N. Semantically Consistent Regularization for Zero-Shot Recognition. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 22–25 July 2017; pp. 2037–2046.
30. Long, Y.; Shao, L. Describing Unseen Classes by Exemplars: Zero-Shot Learning Using Grouped Simile Ensemble. In Proceedings of the 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, CA, USA, 2017; pp. 907–915.
31. Paredes, B.; Torr, P. An embarrassingly simple approach to zero-shot learning. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 2152–2161.
32. Song, Q.; Xu, F. Zero-Shot Learning of SAR Target Feature Space With Deep Generative Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2245–2249. [[CrossRef](#)]
33. Chao, W.; Changpinyo, S.; Gong, B.; Sha, F. An empirical study and analysis of generalized zero-shot learning for object recognition in the wild. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2016; pp. 52–68.
34. Arora, G.; Verma, V.; Mishra, A.; Piyush, R. Generalized Zero-Shot Learning via Synthesized Examples. *Mach. Learn.* **2017**, arXiv:1712.03878.
35. Yu, Y.; Ji, Z.; Guo, J.; Pang, Y. Zero-shot learning with regularized cross-modality ranking. *Neurocomputing* **2017**, *259*, 14–20. [[CrossRef](#)]
36. Lampert, C.; Nickisch, H.; Harmeling, S. Attribute-Based Classification for Zero-Shot Learning of Object Categories. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *36*, 453–465. [[CrossRef](#)] [[PubMed](#)]
37. Xian, Y.; Akata, Z.; Sharma, G.; Nguyen, Q.; Hein, M.; Schiele, B. Latent Embeddings for Zero-Shot Classification. In Proceedings of the IEEE Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 69–77.
38. Akata, Z.; Perronnin, F.; Harchaoui, Z.; Schmid, C. Attribute-Based Classification with Label-Embedding. In Proceedings of the NIPS 2013 Workshop on Output Representation Learning, Lake Tahoe, CA, USA, 9 December 2013.
39. Lampert, C.; Nickisch, H.; Harmeling, S. Zero Shot Deep Learning from Semantic Attributes Categorization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 453–465. [[CrossRef](#)] [[PubMed](#)]
40. Zhang, Z.; Saligrama, V. Zero-Shot Learning via Semantic Similarity Embedding. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 4166–4174.
41. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.; Dean, J. Distributed representations of words and phrases and their compositionality. In Proceedings of the 26th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–10 December 2013; pp. 3111–3119.
42. Zhang, F.; Ni, J.; Yin, Q.; Li, W.; Li, Z.; Liu, Y.; Hong, W. Nearest-regularized subspace classification for polsar imagery using polarimetric feature vector and spatial information. *Remote Sens.* **2017**, *9*, 1114. [[CrossRef](#)]
43. Patterson, G.; Hays, J. SUN attribute database: Discovering, annotating, and recognizing scene attributes. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 16–21 June 2012; pp. 2751–2758.
44. Patterson, G.; Xu, C.; Su, H.; Hays, J. The SUN attribute database: Beyond categories for deeper scene understanding. *Int. J. Comput. Vis.* **2014**, *108*, 59–81. [[CrossRef](#)]

45. Lee, J.; Pottier, E. Overview of polarimetric radar imaging. In *Polarimetric Radar Imaging: From Basics to Applications*; CRC Press: Boca Raton, FL, USA, 2009; pp. 1–4, ISBN 978-1-4200-5497-2.
46. Yang, W.; Liu, Y.; Xia, G.; Xu, X. Statistical mid-level features for building- up area extraction from high-resolution polsar imagery. *Prog. Electromagn. Res.* **2012**, *132*, 233–254. [[CrossRef](#)]
47. Gui, R.; Xu, X.; Dong, H.; Song, C. Urban Building Density Analysis from Polarimetric SAR Images. *Remote Sens. Technol. Appl.* **2016**, *31*, 267–274.
48. Zhang, Z.; Saligrama, V. Zero-Shot Learning via Joint Latent Similarity Embedding. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 6034–6042.
49. Wang, Q.; Chen, K. Zero-Shot Visual Recognition via Bidirectional Latent Embedding. *Int. J. Comput. Vis.* **2017**, *124*, 356–383. [[CrossRef](#)]
50. Akata, Z.; Reed, S.; Walter, D.; Lee, H. Evaluation of output embeddings for fine-grained image classification. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 2927–2936.
51. Fu, Y.; Hospedales, T.; Xiang, T.; Gong, S. Transductive multi-view zero-shot learning. *IEEE Tran. Pattern Anal. Mach. Intell.* **2015**, *37*, 2332–2345. [[CrossRef](#)] [[PubMed](#)]



© 2018 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/>).