

Article

Nonlinear Manifold Learning Integrated with Fully Convolutional Networks for PolSAR Image Classification

Chu He ^{1,2,*}, **Mingxia Tu** ¹, **Dehui Xiong** ¹ and **Mingsheng Liao** ^{2,3}¹ Electronic and Information School, Wuhan University, Wuhan 430072, China; claire0409@whu.edu.cn (M.T.); dhuixiong@whu.edu.cn (D.X.)² State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China; liao@whu.edu.cn³ Collaborative Innovation Center for Geospatial Technology, 129 Luoyu Road, Wuhan 430079, China

* Correspondence: chuhe@whu.edu.cn; Tel.: +86-27-6875-4367

Received: 18 January 2020; Accepted: 11 February 2020; Published: 17 February 2020



Abstract: Synthetic Aperture Radar (SAR) provides rich ground information for remote sensing survey and can be used all time and in all weather conditions. Polarimetric SAR (PolSAR) can further reveal surface scattering difference and improve radar's application ability. Most existing classification methods for PolSAR imagery are based on manual features, such methods with fixed pattern has poor data adaptability and low feature utilization, if directly input to the classifier. Therefore, combining PolSAR data characteristics and deep network with auto-feature learning ability forms a new breakthrough direction. In fact, feature learning of deep network is to realize function approximation from data to label, through multi-layer accumulation, but finite layers limit the network's mapping ability. According to manifold hypothesis, high-dimensional data exists in potential low-dimensional manifold and different types of data locates in different manifolds. Manifold learning can model core variables of the target, and separate different data's manifold as much as possible, so as to complete data classification better. Therefore, taking manifold hypothesis as a starting point, nonlinear manifold learning integrated with fully convolutional networks for PolSAR image classification method is proposed in this paper. Firstly, high-dimensional polarized features are extracted based on scattering matrix and coherence matrix of original PolSAR data, whose compact representation is mined by manifold learning. Meanwhile, drawing on transfer learning, pre-trained Fully Convolutional Networks (FCN) model is utilized to learn deep spatial features of PolSAR imagery. Considering complementary advantages, weighted strategy is adopted to embed manifold representation into deep spatial features, which are input into support vector machine (SVM) classifier for final classification. A series of experiments on three PolSAR datasets have verified effectiveness and superiority of the proposed classification algorithm.

Keywords: PolSAR image classification; deep network; manifold learning

1. Introduction

Synthetic Aperture Radar (SAR) is a typical representative of remote sensing technology, and it has both range and azimuth resolution. Compared with common remote sensing data, Polarimetric SAR (PolSAR) data stores ground objects' scattering echo in sinclair or scattering matrix, which can describe land cover more effectively. PolSAR image classification is an important subject in remote sensing image interpretation system, and its results can be widely applied in many fields, such as earth resource survey, vegetation identification and so forth [1].

1.1. SAR Imagery Feature Extraction

In many classification methods, feature extraction and proper classifier are two important components. For PolSAR data, there are mainly statistical features and polarization features of target decomposition based on original data.

Early statistical characteristics are generally a simple transformation of original polarization scattering matrix, coherence matrix and covariance matrix, such as polarization ratio, phase difference [2]. In 1988, based on complex gaussian distribution, Kong et al. [3] carried out statistical analysis on scattering matrix to achieve maximum likelihood classification. Considering that different types of data have similar canonical covariance matrix, Novak et al. [4] designed a filter classifier based on whitening transformation. Another main manual feature is based on target decomposition, whose purpose is to analyze target's scattering mechanism with appropriate physical constraints, such as Pauli decomposition [5], Freeman decomposition [6]. Different target decompositions have their own emphases and advantages, and are widely used in ground object classification.

After the initial extraction of underlying features, selecting high-level features that represents original data better, and completing final classification with an appropriate classifier are also quite important. Manual feature engineering can filter features through pearson coefficient or L1 regularization. Traditional statistical features are not robust in the face of complex terrain scene, while target decomposition methods have problems like rough scattering boundary. With the resolution improvement, spectral complexity in SAR imagery becomes more obvious, making feature selection much difficult. Therefore, feature engineering with fixed pattern has poor data adaptability and low feature utilization. However, feature learning method represented by deep learning provides a good idea to solve this problem.

Neural network based on back-propagation can automatically adapt to the data and specific probability model or priori distribution hypothesis are not necessary anymore. Beyond manual feature-based traditional methods in many aspects, classification framework based on Convolutional Neural Network (CNN) has been gradually applied in PolSAR image classification [7]. For example, Ronny Hansch [8] used complex neural network to realize PolSAR data learning and classification. He [9] achieved hierarchical terrain classification based on bayesian network and conditional random field in 2017. However, there are essential differences between the imaging mechanism of optical image and SAR imagery. Overfit and weak generalization are likely to occur, if directly transfer deep models that perform well on optical images to SAR imagery. How to solve these problems adaptively becomes a difficulty for SAR image interpretation. By selecting the most relevant spectral bands in hyperspectral image, Reference [10] tackles the problem of dimensionality curse and the limited number of training samples. In Reference [11], 3D convolution kernel is applied, to extract spectral and spatial features for hyperspectral imagery simultaneously, which retains spectral information to enhance classification.

1.2. Nonlinear Learning for PolSAR Data

As an imaging system based on the coherence principle, there is lots of noise-like speckle in PolSAR image. In 1975, Goodman [12] carried out statistical modeling on speckle, and found its amplitude and intensity obeyed Nakagami and gamma distribution [13] respectively. Method above is suitable for describing even areas such as river and grassland. For non-uniform areas like forest and urban areas in high-resolution SAR image, better nonlinear modeling should be introduced for detailed description.

In addition, to describe polarization information, covariance matrix and coherence matrix are commonly used expressions. They are Hermitian positive definite matrix, which forms Riemannian manifold instead of Euclidean space [14]. Therefore, every pixel in PolSAR image can be visualized as a point on a low-dimensional manifold, which indicates that manifold learning is more reasonable for nonlinear modeling of PolSAR data. As shown in Reference [15], nonlinear dimensionality reduction

method based on manifold can mine the compact structure embedded in the original high-dimensional space, for better PolSAR imagery classification.

Recently, a large number of nonlinear manifold methods have been proposed to learn feature space with good separability for high-dimensional data. For example, Ainsworth used Local Linear Embedding(LLE) [16] and Isometric Mapping(Isomap) [17] in References [18,19] respectively, to learn the polarization covariance matrix directly. In order to retain data's local characteristics, Tu et al. [15] processed multidimensional PolSAR data based on nonlinear Laplacian Eigenmaps (LE) [20], demonstrating that the separability of eigenfeature mined by LE was better than that of the original features. To integrate label information to enhance feature discrimination, Shi et al. [21] proposed a supervised graph embedding model for dimension reduction of polarized features. In the corresponding low-dimensional subspace, discrimination information from training samples is well preserved.

1.3. Problems and Motivation

Generally, the essence of PolSAR image classification is a nonlinear mapping problem, which can be regarded as function mapping in essence [22]. Through nonlinear structure in deep network, feature transformation layer by layer can realize approximation of any substantive continuous function, thus making classification or prediction easier.

Inspired by Colah [23], Figure 1 shows an example of how the input is transformed in a CNN. After convolution layer, the input image is rotated and scaled, but relative position between two lines does not change. Such linear transformation cannot separate lines by a linear plane. However, after tanh's nonlinear activation, the original two lines undergo essential transformation and can be linearly separated, as shown in Figure 1c.

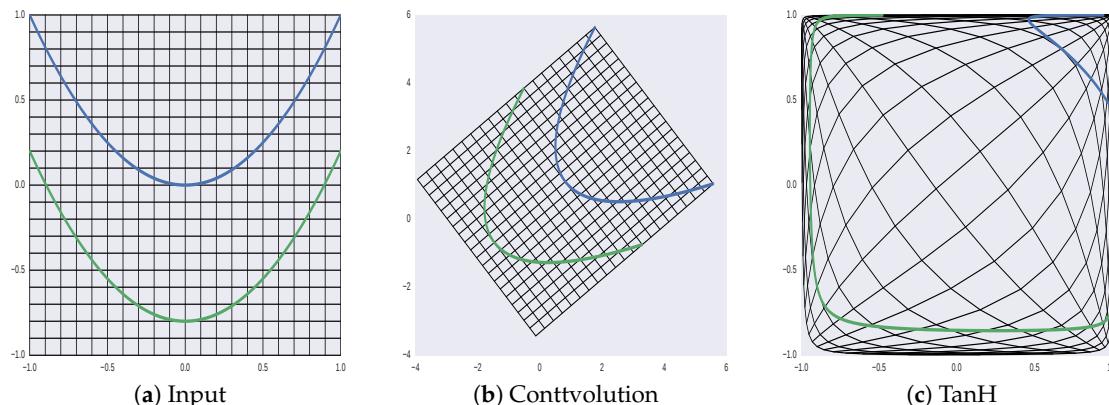


Figure 1. Behaviour of convolution and TanH layers. Here, two curves (a), which cannot be linearly separated, are fed into a standard CNN; the input undergoes different types of transformations in convolution and TanH layers. (b) and (c) display the output of convolution and TanH layers, respectively.

In deep network, convolution realizes linear transformation and then activation function forms nonlinearity. If there is enough data and network layers are unlimited, then the mapping and superposition of multi-layer networks can achieve approximation of any substantive continuous function. However, with layer number increases, gradient disappearance occurs, greatly limiting the network's mapping ability. In addition to better modeling of high-dimensional polarized data, nonlinear manifold learning can also accelerate the approximation of nonlinear classification plane. If function in manifold method is complex enough to resolve any curved surface, it can even replace multi-layer fitting in deep network. Though function ability is limited, we still hope to combine the advantages to build a unified linear-nonlinear framework for PolSAR image classification.

On the other hand, attribute information in PolSAR imagery is quite precise and complex, manual features can not automatically adapt to data itself. Lots of work have confirmed that feature learning is superior to traditional methods for PolSAR image interpretation [9,24]. Feature learning methods complete underlying feature abstraction through network layers iteratively, so as to learn data's essential information. At present, as one of the best methods of semantic segmentation, Fully Convolutional Networks (FCNs) [25] is considered to be applied to PolSAR image classification, to obtain abstracted high-level representation. However, for limited PolSAR data, training a deep network from scratch is difficult to converge. Meanwhile, considering that finite layers in deep network have limited mapping ability, this paper introduces nonlinear manifold learning to model the core variables of PolSAR image and tries to reveal the essential structure of high-dimensional data in low-dimensional subspace, to achieve more efficient PolSAR image classification.

1.4. Contributions and Structure

In view of the two main problems aforementioned, that is, how to mine deep features with strong adaptability and high utilization and nonlinear learning for high-dimensional polarized feature, this paper proposed a new method, namely nonlinear manifold learning integrated with fully convolutional networks for PolSAR image classification. First of all, due to different imaging mechanism, classification performance will be degraded when directly transplanting deep network trained on optical images to PolSAR data. Therefore, transfer learning is adopted for reference, the pre-trained FCN model is utilized to learn deep spatial features for PolSAR image automatically. At the same time, high-dimensional polarized feature is extracted through target decompositions based on PolSAR data's scattering matrix and coherence matrix. It is then mapped to the deep core feature space which is more representative of the original data through nonlinear manifold learning. From the perspective of deep learning, manifold method is actually a shallow learning method that lacks of spatial constraints. Considering that FCN is not good at capturing details, this paper integrated deep spatial features from FCN with subspace representation learnt by nonlinear manifold method, to obtain complementary advantages. Finally, the fused feature is fed into a discriminant module for classification. Overall, the main contributions of this paper are listed as follows:

- To automatically learn effective features, deep network is employed and it breaks through the limitation of manual features. In this paper, FCN model, pre-trained on optical images, is transferred to learn nonlinear deep multi-scale spatial information of PolSAR image. Specifically, “RGB” pseudo color maps serve as the input of FCN, to fit well-trained parameters, the high-level semantic information adaptively learned by FCN can greatly promote classification.
- To nonlinear model non-uniform areas in PolSAR image, a manifold-based nonlinear learning method is employed to capture the most essential structure of high-dimensional polarized data. Nonlinear manifold modeling can supplement the mapping ability of deep network with finite layers. In addition to removing redundant information in high-dimensional polarized data, nonlinear manifold method can also explore intrinsic representation in low-dimensional subspace, to improve feature's distinguishing ability.
- The shallow manifold subspace representation is embedded into deep spatial features learned by FCN in a weighted way, which makes their advantages complementary. The final fused features contain multiple types of information, from local to global, polarization to space, enhancing representation ability for classification.

The content of this paper is as follows: Section 2 introduces the high-dimensional polarized space and basic structure of FCN. Details of the proposed scheme: nonlinear manifold learning integrated with fully convolutional networks for PolSAR image classification is illustrated in Section 3. The following Section 4 describes our experiments on three PolSAR datasets and presents detailed analysis. Finally, the conclusion of this paper is drawn in Section 5.

2. Related Work

2.1. Multi-Dimensional Polarization Data Space

The electromagnetic scattering process of radar target is a linear transformation. Polarized SAR system can characterize ground objects by transmitting and receiving polarized radar waves, to measure the medium's related properties. According to incident waves, the polarization variable of the target can be represented by a complex two-dimensional matrix (Sinclair scattering matrix [1]):

$$E^{re} = [S]E^{tr} = \begin{bmatrix} E_H^{re} \\ E_V^{re} \end{bmatrix} = \frac{e^{ik_0 r}}{r} \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix} \begin{bmatrix} E_H^{tr} \\ E_V^{tr} \end{bmatrix}. \quad (1)$$

The superscript *re* represents incident wave emitted by antenna and *tr* is the received scattering wave. k_0 denotes wave number of the electromagnetic wave, r means the distance between the scattering target and the receiving antenna, S is polarized scattering matrix. Among them, S_{HH} and S_{VV} are co-polarization components while S_{HV} and S_{VH} are cross-polarization components. Under the condition of transceiver-receiver co-location, cross-polarization components are equal, that is $S_{HV} = S_{VH}$, according to reciprocity definition. Therefore, matrix S can be expressed as:

$$[S] = e^{i\phi_0} \begin{bmatrix} |S_{HH}| & |S_{HV}| e^{i(\phi_z - \phi_0)} \\ |S_{HV}| e^{i(\phi_x - \phi_0)} & |S_{VV}| \end{bmatrix}. \quad (2)$$

The matrix element $S_{ij}, ij \in (HH, HV, VH, VV)$ represents complex scattering characteristics (scattering coefficient or amplitude) obtained by transmitting and receiving radar waves in direction i and j respectively. Scattering matrix S can describe coherent scatterers effectively, but not well for distributed scatterers in natural scenes. In order to reduce the influence of speckle noise, it is necessary to utilize statistical methods to analyze scatterers by second-order descriptors, such as coherence matrix T , shown in the formula below:

$$T = \langle k \cdot k^H \rangle = \begin{bmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{bmatrix}, \quad (3)$$

where k is the target vector of polarized scattering matrix S under Pauli basis, T_{ij} is the element of polarized coherence matrix T .

Considering the complexity of scattering mechanism, it is difficult to analyze coherence matrix T directly. Therefore, many target decomposition methods based on coherence matrix are proposed. This kind of decomposition method can be described by the following formula:

$$[T] = \sum_i^k q_i [T]_i \quad (4)$$

where $[T]_i$ and q_i represent the i th component's scattering model and coefficient, respectively.

In practical applications, objects in real scenes are often complex and changeable, a single polarized decomposition method can not describe different objects well. For example, Huyuen decomposition can be used to analyze man-made targets with high-density and it tends to treat scatterers in natural scenes as noise terms; Freeman-Durden three-component scattering model is suitable for targets satisfying reflective symmetry, such as forest canopy. But in urban areas, Yamaguchi decomposition [26] is introduced since symmetry is no longer satisfied. Considering that different target decomposition methods prefer different types of ground objects, a reasonable and effective way is to stack polarized features from different target decompositions, to describe PolSAR data jointly. In view of this, this paper adopts three different types of feature (52-dimension in total) to represent PolSAR image,

including statistics of scattering coherence matrix T ; polarized coherence decompositions based on scattering matrix S ; incoherence decompositions based on coherence matrix T .

2.2. FCN Structure

Since 2012, CNN has made great achievements in image classification and has been widely applied in many fields. The strength of CNN lies in its multi-layer structure, which can learn deep features automatically. However, due to large storage cost, redundant computation and difficulties in determining patch size, classification performance of traditional CNN-based methods is largely limited. As the pioneer of semantic segmentation, FCN [25] attempts to assign label for each pixel, that is, from image-level classification to pixel-level classification. FCN is fine-tuned on the pre-trained VGG-16 classification model, and can achieve end-to-end training. On the basis of VGG-16 network, FCN replaces all fully-connected layers with convolutional layers, to ensure that images of any size can be fed into the network. In order to achieve dense prediction, FCN adopts a coding and decoding architecture. There are de-convolution layers behind the full convolutional layers, to restore the output to the same size as the input image. FCN have three different structures: FCN-32s with single branch, FCN-16s with double branches and FCN-8s with three branches. In order to further improve the prediction performance, FCN adopts skip architecture to fuse information at different depths. Specifically, as shown in Figure 2, the feature map from conv7 layer after 2 times sampling, is fused with the fourth pooling layer's output, then the result of FCN-16s is obtained after 16 times up-sampling. Similarly, the output of FCN-8s with three branches, is also the fusion of feature maps with multiple scales, from the third, the fourth and the fifth pooling layers.

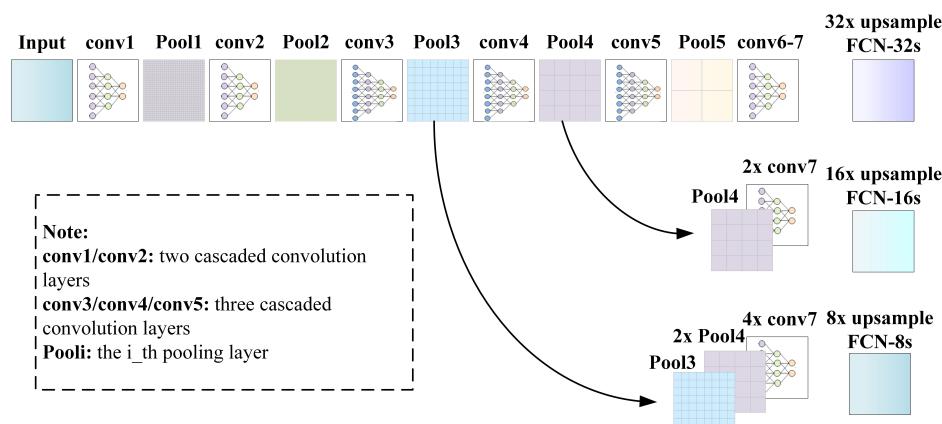


Figure 2. Architectures of Fully Convolutional Networks.

Among those three structures of FCN, it has been proved that FCN-8s has the best semantic segmentation performance both in theory and practice. The shallower layer can extract linear local features, while the deeper layer can learn high-level global structure. And the skip architecture in FCN-8s integrates information from three layers with different depths, thus the final feature maps include both rough information and fine features, and have stronger expressive ability. Therefore, this paper employs FCN-8s to learn the multi-scale deep spatial features of PolSAR images.

2.3. Classical Manifold Methods

Manifold is a general term of geometric objects including curves and surfaces of various dimensions. According to manifold hypothesis, data in high-dimensional space actually locates in a low-dimensional manifold, in which data invariance and coordinates correspond to core data variables are determined. The purpose of manifold learning is to seek the internal mapping between the original high-dimensional data and a low-dimensional manifold structure.

Various methods have different requirements for manifold properties. Under manifold hypothesis, manifold learning methods can be roughly divided into two categories: global method and local method. Proposed by Hinton [27], t-distributed Stochastic Neighbor Embedding (TSNE) is special and can control focus on global or local structure by parameter adjustment. Table 1 shows typical examples of various manifold methods:

Table 1. Typical manifold methods classification.

| Category | Method | Optimization |
|------------------------------|-------------------------------|--|
| Global Manifold | Multi-Dimensional Scaling | $\arg \min_y \sum_{i,j} (\ x_i - x_j\ - \ y_i - y_j\)^2$ |
| | Isometric Mapping | $\arg \min_y \sum_{i,j} (d_M(x_i, x_j) - \ y_i - y_j\)^2$ |
| Local Manifold | Locally Linear Embedding | $\arg \min_y \sum_i \ y_i - \sum_j w_{ij} y_j\ ^2$ |
| | Local Tangent Space Alignment | $\arg \min \sum_i \sum_j (I - Q_i Q_i^T) (x_{ij} - \bar{x}_i)^2$ |
| | Laplacian Eigenmap | $\arg \min_y \sum_{ij} (y_i - y_j)^2 w_{ij}$ |
| Global and local equilibrium | TSNE | $\arg \min \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}$ |

Global manifold methods consider that structural relationship between all data pairs is equally important for embedding determination. In global methods, the metric matrix based on distance of high-dimensional data points is constructed, which is then transformed into inner product matrix, followed by feature decomposition to obtain the manifold embedding. Local manifold methods regard that only local structural information plays a key role, thus accurate modeling of local structure is the focus, which reduces computation to a certain extent. As a global and local equilibrium, TSNE has been proved to have better mapping effect.

3. Proposed Method

3.1. Transfer Learning Based on FCN-8s

As illustrated in Section 2.2, the skip architecture in FCN-8s consists of shallow and fine layers with receptive fields of various scales, it can learn multi-scale spatial characteristics from local to global. Therefore, this paper adopts FCN-8s trained on Pascal VOC 2011 for deep feature extraction.

At present, PolSAR image is really expensive and there are few accurate labeled samples. Serious over fitting occurs when FCN is directly adopted for PolSAR image classification. Therefore, reasonable transfer learning strategy [28] is quite necessary. Generally, new dataset size and its similarity with source dataset determine the specific migration learning scheme. For feature extraction in CNN, as shown in Figure 3, the underlying feature is the general description (like edges) of the target, while deeper layer pays more attention to abstraction of the input data. Therefore, when the new dataset is small and not similar to the source, a suitable scheme is to use the pre trained CNN model as the feature extractor.

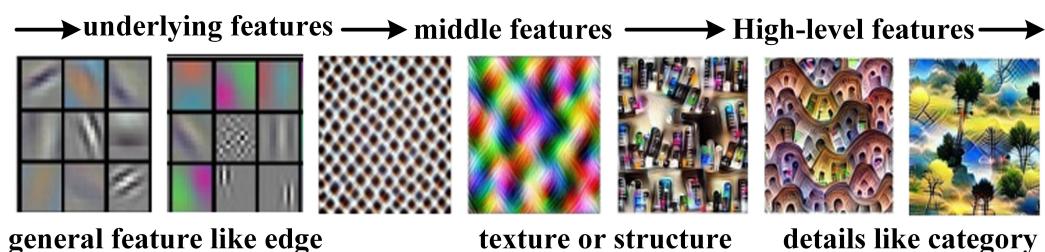


Figure 3. Feature extraction in CNN.

For PolSAR imagery, training samples available is few and it has essentially different imaging mechanism with optical image. Therefore, this paper employs FCN-8s model pre-trained on Pascal VOC dataset as a feature extractor, to obtain deep spatial features of PolSAR image. Specifically, the output of the network layer ‘score’ is extracted. If convolution layers are regarded as various filters, feature maps from ‘score’ layer is 21-d spatial features from 21 different filters, containing lots of nonlinear multi-scale spatial information.

3.2. Manifold Mapping Based on TSNE

TSNE algorithm is developed on the basis of Random Neighborhood Embedding(RNE). RNE emphasizes the similarity between data points rather than their specific location. It uses Euclidean distance to express this similarity, and further converts distance relationship between point χ_j and χ_i to conditional probability $p_{j|i}$, that is, χ_i selects χ_j as its adjacent point with conditional probability $p_{j|i}$. However, there are two main problems in RNE, one is asymmetry, $p_{i|j} \neq p_{j|i}$. The other is crowding problem, since there are differences between data’s low-dimensional and high-dimensional distributions.

To solve the problems above, student t-distribution is introduced in TSNE, in which data similarity in original high-dimensional space and embedding space are represented as Gaussian joint probability and student t distribution probability, respectively. To ensure that the joint probability in two spaces is equal, distance between close data points in high-dimensional space will be retained in low-dimensional manifold as much as possible. That is, data points in the same category will be closer after embedding mapping, and data points in different categories will be more distant. TSNE considers both local and global structure and is more conducive to subsequent classification. The concrete algorithm mainly includes three steps:

(a) Distance similarity calculation in high dimensional space: Formula (5) is employed to obtain conditional probability of Gaussian distribution in high dimensional space, then joint probability of Gaussian distribution is derived from Equation (6).

$$p_{j|i} = \frac{\exp\left(\frac{-\|x_i - x_j\|^2}{\sigma^2}\right)}{\sum_{k \neq i} \exp\left(\frac{-\|x_i - x_k\|^2}{\sigma^2}\right)} \quad (5)$$

$$p_{ij} = \frac{p_{j|i}}{p_{i|j}} \quad (6)$$

where σ_i represents variance of Gaussian distribution with x_i as the center.

(b) Distance similarity calculation in low dimensional space: the joint probability of Student t distribution in low dimensional manifold is shown in Formula (7), with one freedom degree in t distribution.

$$q_{ij} = \frac{\left(1 + \|y_i - y_j\|^2\right)^{-1}}{\sum_{k \neq l} \left(1 + \|y_k - y_l\|^2\right)^{-1}} \quad (7)$$

(c) Manifold mapping: the optimal mapping is obtained by minimizing Kullback-Leibler(KL) divergence in two spaces through gradient descent method, expressed in Formula (8).

$$C = \sum_i KL(P_i \| Q_i) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (8)$$

To explore the mapping effects of different manifold methods, this paper adopts 5 classic manifold methods: Multi-Dimensional Scaling (MDS), Isomap, LLE, Spectral Embedding (SE) and TSNE for experiments on MNIST dataset. The dataset includes 10 handwritten numbers, from 0 to 9 and the

image size is 28×28 . In experiments, the subspace of manifold is set as three-dimension. For clear visualization, the first six categories, namely number 0 to 5, are visualized in Figure 4:

It can be seen from Figure 4b that MDS not only clusters data points of the same classes, but also disperses different categories. But the dispersion is not very obvious, different types are still mixed together. The dispersion of Isomap is slightly stronger than that of MDS, such as number 0. LLE shows good clustering ability, while performs badly in dividing different categories. The property of focusing on structural information makes SE perform well in handwritten datasets, as shown in Figure 4e. By comparing Figure 4f with the rest, it can be concluded that TSNE achieves the best performance on both aggregation and dispersion, greatly reducing difficulties for subsequent classification. For some special datasets, features of TSNE can be classified directly, even without feature space transformation.

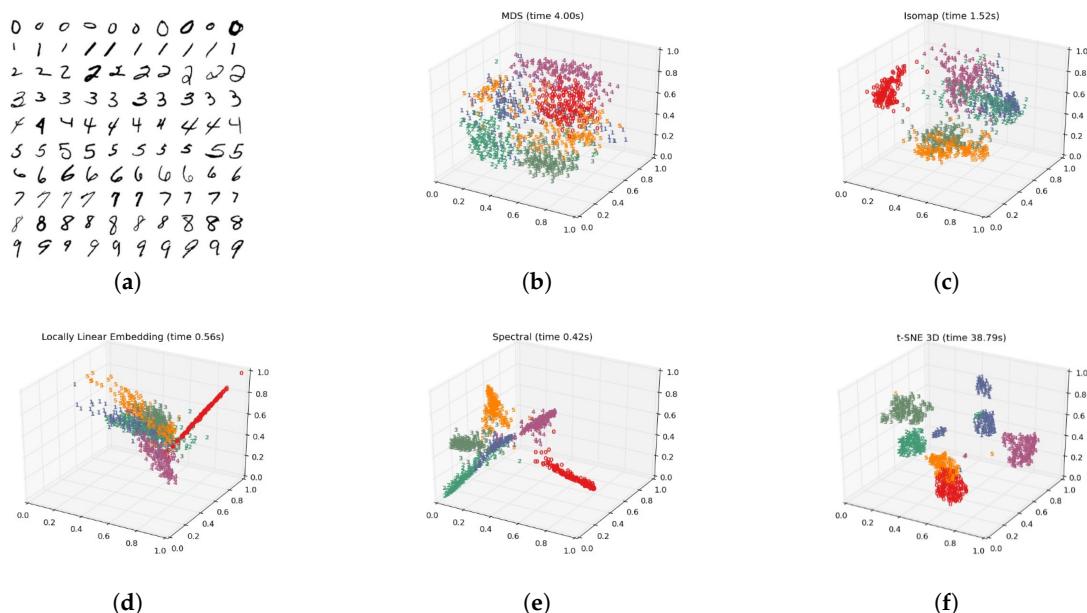


Figure 4. Different manifold visualization on Mnist dataset. (a) Original image sample. (b) MDS. (c) Isomap. (d) LLE. (e) SE. (f) TSNE.

3.3. The Proposed Algorithm Framework

In the proposed algorithm, in order to get features with adaptability, FCN model and migration learning strategy are adopted, to learn deep spatial features for PolSAR imagery. At the same time, for improving mapping efficiency of high-dimensional polarized data, nonlinear manifold learning is employed to mine the essential structure of PolSAR data.

When deep spatial features of FCN and compact manifold representation are obtained respectively, this paper utilizes weighted strategy to fuse the two features, shown in Figure 5, making them complementary so as to enhance feature discrimination. $F_{FCN} \in R^{K_1 \times M}$, represents multi-scale deep features by M samples with K_1 dimension, here $K_1 = 21$. $F_{TSNE} \in R^{K_2 \times M}$ stands for representation of nonlinear manifold learning. Then the fused feature can be denoted as:

$$F = [w_1^* \phi(F_{FCN}) \quad w_2 * \phi(F_{TSNE})] \text{ s.t. } w_1 + w_2 = 1 \quad (9)$$

where w_1, w_2 are weighting factors and are set to 0.75, 0.25 according to previous experience.

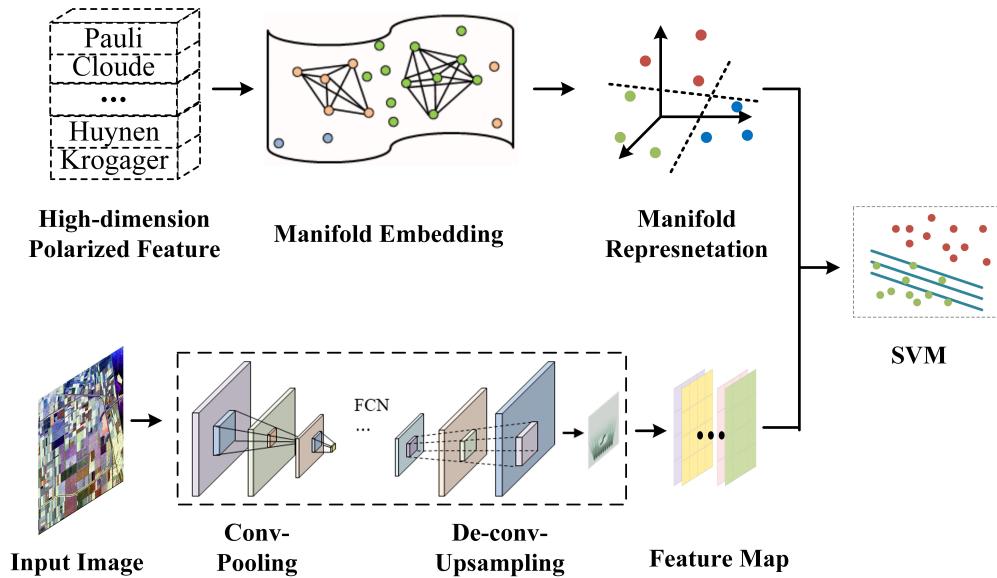


Figure 5. The framework of the proposed algorithm.

In fused feature F , $F \in R^{(K_1+K_2)*M}$, that is, each sample is represented by a $(K_1 + K_2)$ dimension feature vector, which contains a variety of information, from local to global, polarization to spatial structure. Considering that the sample size of PolSAR data is quite small, the fusion features are finally fed into Support Vector Machine(SVM) for subsequent classification.

4. Experiments and Analysis

This chapter consists of experiment data, evaluation standards and detailed results analysis. For fully demonstration of the proposed method and other contrastive schemes, experiments on a series of PolSAR datasets have been carried out. Visualized classification results and evaluation criteria are displayed to make a sufficient comparison of various algorithms.

4.1. Experiment Datasets and Evaluation Standards

4.1.1. Flevoland Dataset

The first experiment data is from Flevoland region in Netherlands, collected by AIRSAR system of NASA laboratory in 1989, belonging to L-band fully polarized image. The scene size is 750×1024 pixels, and the resolution of range and azimuth are 6.7 m and 12.1 m respectively. In addition to background, there are 11 different types of ground objects, including rapeseed, grassland, forest, pea, alfalfa, wheat, beet, bare land, stem bean, water and potato. As a classic farmland PolSAR data, this image contains rich ground types and multi-plantings have similar scattering mechanism, thus it is widely adopted by researchers to verify different PolSAR image classification algorithms. Its Pauli RGB image, Ground Truth (GT) and corresponding labels are shown in Figure 6.

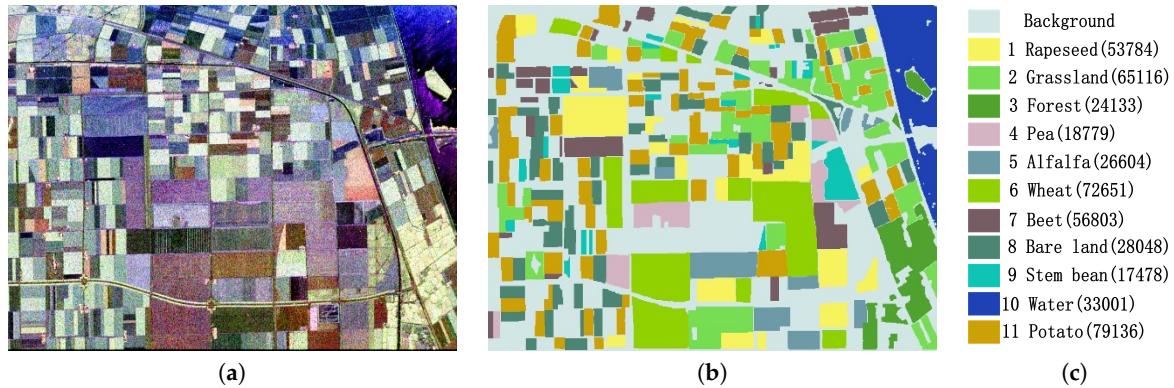


Figure 6. Flevoland dataset. (a) Pauli RGB. (b) Ground Truth. (c) labels.

4.1.2. Foulum Dataset

The second PolSAR data was collected by EMISAR system in Foulum area of Denmark on April 17, 1998. The selected image size is 1024×747 pixels, and the range and azimuth resolution are 0.75 m and 1.5 m respectively. As a L-band fully PolSAR image, this scene contains 5 kinds of ground objects (except the background): broad-leaved crop, fine-stem crop, bare land, town and forest. Figure 7 shows the Pauli RGB image and the corresponding GT with category labels.

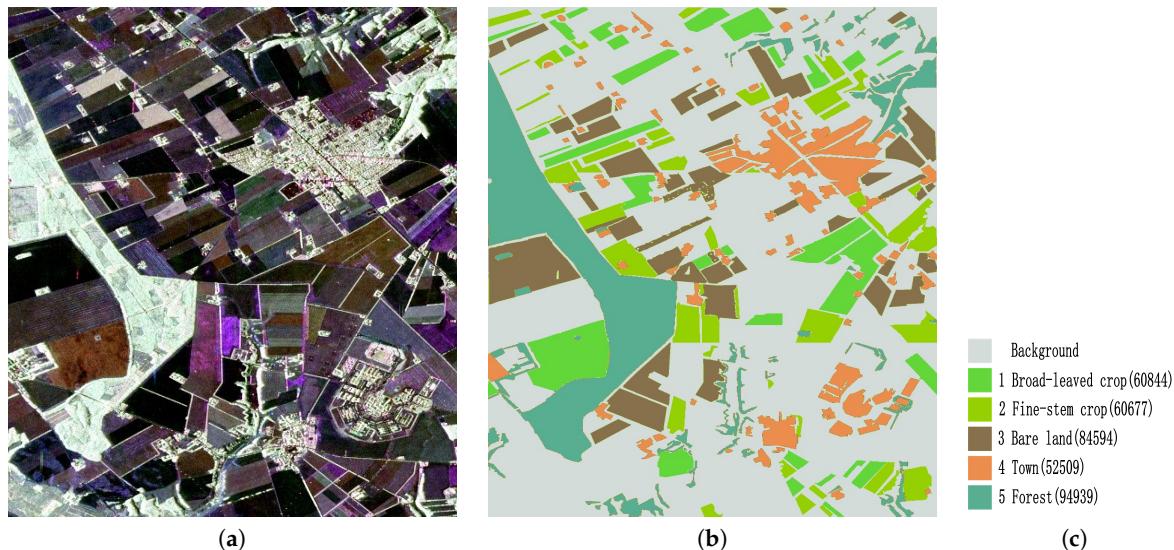


Figure 7. Foulum dataset. (a) Pauli RGB. (b) Ground Truth. (c) labels.

4.1.3. San Francisco Dataset

The last experiment data is a L-band PolSAR image, collected by AIRSAR system of NASA's Jet Propulsion Laboratory in the area of San Francisco Bay, in 1988. The image size is 900×1024 pixels, with spatial resolution of 10 m \times 10 m. This scene mainly includes five different ground objects, namely vegetation, ocean and urban area, which can be further divided into three types: high-density urban area, developed urban area and low-density urban area. The Pauli RGB image, GT and label of this data are presented in Figure 8.

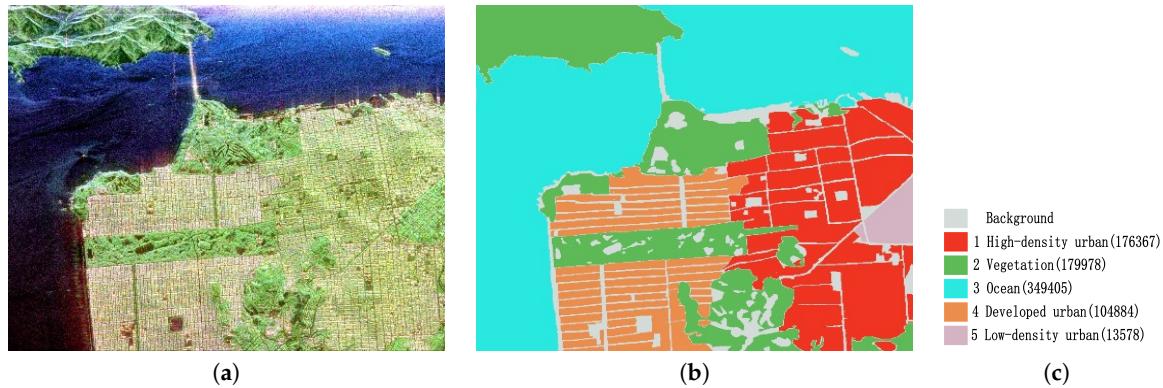


Figure 8. San Francisco dataset. (a) Pauli RGB. (b) Ground Truth. (c) labels.

4.1.4. Evaluation Standards

To evaluate the performance of different algorithms, overall accuracy (OA) of pixel classification, confusion matrix and kappa coefficient are used as evaluation indexes. Assuming that there are k classes in the original data, P_{ij} represents the total number of pixels that belong to class i but are predicted to be class j . Specifically, P_{ii} denotes true positives, P_{ij} means false positives and P_{ji} stands for false negatives, the formula of OA can be represented as:

$$OA = \frac{\sum_{i=1}^k p_{ii}}{\sum_{i=1}^k \sum_{j=1}^k p_{ij}} \quad (10)$$

Kappa coefficient is calculated based on confusion matrix, let P_{i+} and P_{+i} denote elements number in row i and column i in the matrix respectively. Then it can be expressed as:

$$\text{kappa} = \frac{\sum_{i=1}^k \sum_{j=1}^k p_{ij} * \sum_{i=1}^k p_{ii} - \sum_{i=1}^k (p_{i+} * p_{+i})}{\left(\sum_{i=1}^k \sum_{j=1}^k p_{ij} \right)^2 - \sum_{i=1}^k (p_{i+} * p_{+i})} \quad (11)$$

OA reflects the proportion of correctly classified pixels directly. However, some categories can not be recalled, especially for extremely unbalanced dataset. As an index for consistency test, the more unbalanced the confusion matrix is, the lower the kappa value is. Therefore, kappa coefficient can better measure whether the model prediction is consistent with the actual classification result.

4.2. Experiment Setup

4.2.1. Comparison Schemes

As illustrated in aforementioned section, the original 52-d features containing three kinds of polarimetric information, is used to represent PolSAR imagery. For different PolSAR datasets, it is necessary to determine the optimal subspace dimension through experiments. TSNE method adopted in this paper is an unsupervised manifold algorithm, which optimizes KL divergence to obtain the optimal data distribution in subspace.

When manifold representation is gained, this paper combines it with polarized deep features for joint classification. Drawing on transfer learning, Pauli RGB maps are fed into pretrained FCN-8s model to get 21-d feature maps from score layer, through network forward reasoning. In deep neural network, the original data undergoes different convolutional linear combination and nonlinear transformation of activation layers, the output sparse features can represent the spatial relationship better and have stronger discrimination ability for subsequent classification task.

After obtaining polarimetric manifold representation and deep space features respectively, they are fused by aforementioned weighting strategy. Finally, the SVM classifier with Gaussian kernel is adopted for classification, and the evaluation indexes are all calculated based on test samples.

The proposed algorithm focuses on constructing efficient fused features. To fully explore the performance of different features, variable-control method is employed for evaluation. After extracting different features from the following comparison methods, they are fed into SVM classifier to carry out a series of verification experiments.

(a) T3 features from polarized coherence matrix (T3): 3-d features composed of diagonal elements in original coherence matrix. The T3 features are directly adopted to represent the PolSAR image. Since this T3 feature only includes partial polarization information, it may not be able to fully describe the original PolSAR data.

(b) Manifold features located in optimal subspace (TSNE): considering that there is much redundancy in 52-d features extracted from the original imagery, manifold method is utilized to seek for compact intrinsic representation. Compared with deep features from neural network, manifold features are shallow and do not have high-level abstract meaning.

(c) Deep features from network (FCN): with transfer learning for reference, Pauli RGB images serves as input of well-pretrained FCN-8s model, to learn multi-scale spatial relationship. Though not all polarized characteristics are involved, this feature extracted from deep network is nonlinear, and it is an abstract deep semantic representation.

(d) T3-FCN: in order to compare with the fused algorithm presented in this paper, weighted strategy is also adopted to embed original polarized information into deep multi-scale spatial features.

(e) The proposed method (TSNE-FCN): manifold method-TSNE is firstly employed to reduce dimensionality for original high-dimension features, then we use weighted fusion strategy to integrate manifold representation into learnt deep features of FCN-8s model.

4.2.2. Params Setting

In experiments, specific parameters are quite important for final classification. This section discusses parameters involved thoroughly. The proposed classification algorithm (TSNE-FCN+SVM) mainly includes three stages: TSNE dimensionality reduction for high-dimensional polarized features, spatial features learning by FCN, feature fusion and classification.

In the first phase of manifold dimension reduction, parameters in TSNE affecting mapping efficiency and performance are listed as follows:

(1) Perplexity: a large dataset requires a large perplexity, which is generally set between 5–50, which is set as 50 in this paper.

(2) Clustering degree: to control the compactness of the original clusters and distance between each other in the manifold space. The bigger the degree is, the larger the space between clusters is. It is set to 12 by default in experiment.

(3) Learning rate: usually within the range of [10 1000], it is set to 200 in this paper.

(4) Distance measurement: the metric adopted when calculating the distance between different samples. In this paper, the default method is “European distance”.

(5) Gradient algorithm: when dimension <4, Barnes-Hut approximation method is employed, which can effectively reduce the complexity to $O(dN \log N)$. While the dimension is higher, exact method is used and it requires complexity of $O(dN^2)$. Where d is output dimension and N is sample number.

For different datasets, manifold feature that represents the original PolSAR data may not lie in the same manifold subspace. In experiments, each category(background excluded) is sampled with the same proportion(the chosen training ratio) for supervised training, to determine the optimal manifold subspace dimension. For each PolSAR imagery, the training ratio(for SVM classifier) has to be determined by experiments in advance. when the ratio is determined, the training set is obtained by sampling each category randomly and the rest(no background) are for test. Duing to random

sampling, the OA will be slightly different ($\pm 0.5\%$) even with the same ratio. In fact, for each data set, we repeated experiments with the same parameters 15 times. Considering that the visualization maps can not be averaged, we presented the results corresponding to OA median in the work.

For three PolSAR datasets, the relationship between classification accuracy with training proportion and subspace dimension is shown in Figure 9. With 1% training ratio, the classification accuracy on both three datasets presents obvious fluctuation, showing that SVM classifier is not fully trained with much few samples and the performance is poor. When the proportion increases to 5%, the improvement on accuracy is almost saturated. Considering both classification accuracy and algorithm robustness, the subspace dimensions of Flevoland, Foulum and San Francisco dataset, are set to 33, 30 and 39, and the training proportion is 3%, 5% and 4% respectively.

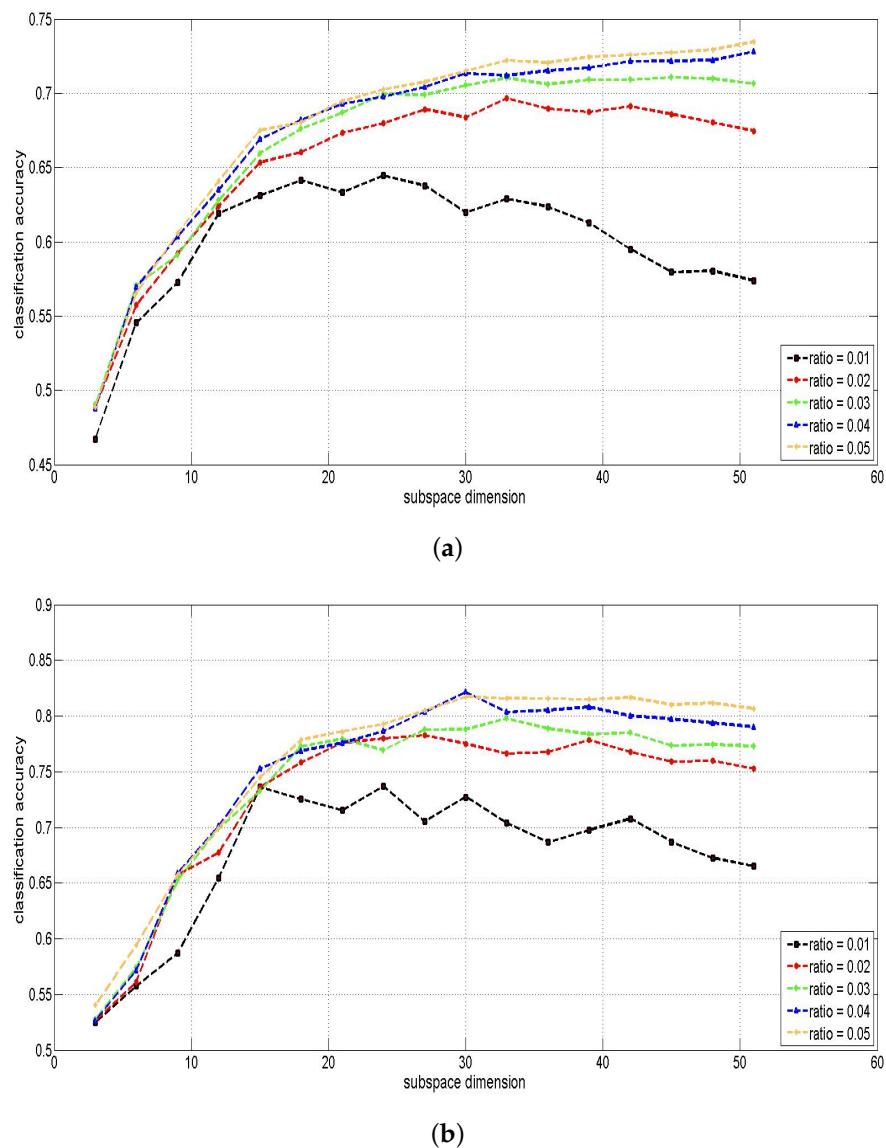
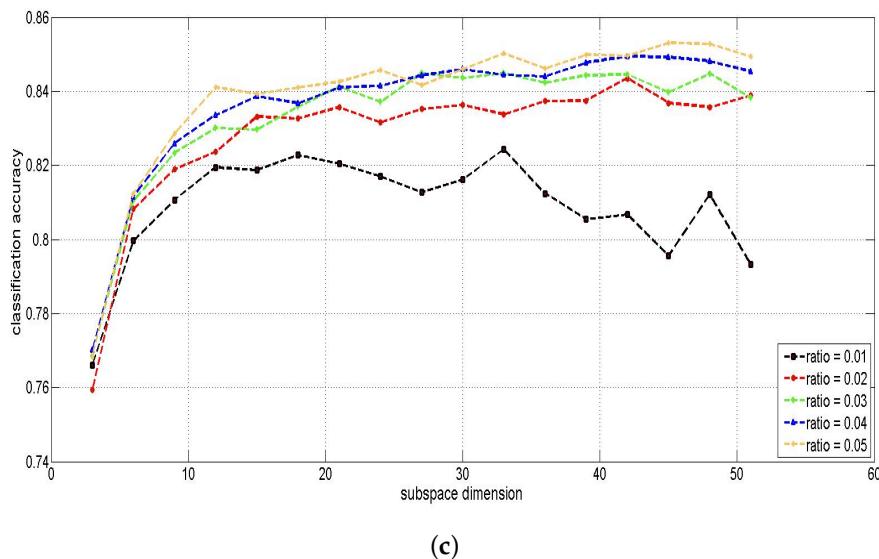


Figure 9. Cont.



(c)

Figure 9. Classification accuracy vs train ratio and subspace dimension. (a) Flevoland dataset. (b) Foulum dataset. (c) San Francisco dataset.

For following feature fusion and classification task, our experiment takes 0.1 as step to adjust the weighting coefficient (between 0–1), and finally sets the weights as 0.75 and 0.25 according to classification accuracy. In order to ensure the robustness of the proposed algorithm, five-fold cross validation is adopted in experiment.

4.3. Experimental Analysis and Results

For the experimental datasets, this paper compares different algorithms from various perspectives, like visual classification results, confusion matrix and different evaluation metrics including classification accuracy of each category, overall accuracy and kappa coefficients.

Table 2 shows the classification performance of Flevoland dataset under different schemes. It can be seen that the combination of T3 feature with SVM classifier has a poor classification effect on each category, and its OA is only 54.10%. The main reason is that T3 feature contains few polarized information and can not describe polarization targets sufficiently. TSNE seeks for compact representation of the original high-dimension polarized feature in the manifold space, whose performance has been greatly improved. In addition to rapeseed and pea, the accuracy of TSNE on other categories has been largely enhanced, especially for bare land and potato, both have increased by more than 30%. Considering spatial relationship, the accuracy of FCN reaches 93.58%, demonstrating the strong ability of nonlinear mapping of activation layers in the network. In contrast, better classification results are achieved by integrating shallow linear characteristics with deep feature with abstract ability. As shown in the results of T3-FCN and TSNE-FCN, both have different degrees of improvement compared with that of single FCN, in which the classification accuracy on almost all categories has risen. However, different linear features and FCN output have shown different positive synergy. It can be noted that T3-FCN (OA 96.34%) has achieved the highest accuracy among all the comparison algorithms, which has exceeded TSNE-FCN. On most categories, the performance of T3-FCN is better than that of TSNE-FCN, indicating that the cooperative effect of manifold representation combined with FCN feature is slightly lower than integration of T3 information and FCN feature, on the Flevoland dataset.

Table 2. Category accuracy (%), OA (%) and kappa coefficient of different features (classifier: SVM) on Flevoland dataset. (The category with highest accuracy is shown in bold.)

| Categories | T3 | TSNE | FCN | T3-FCN | TSNE-FCN |
|-------------|--------|--------|--------|---------------|--------------|
| 1/rapeseed | 70.16 | 64.20 | 95.61 | 97.40 | 93.89 |
| 2/grassland | 68.84 | 71.09 | 93.49 | 96.71 | 95.40 |
| 3/forest | 80.09 | 86.25 | 96.58 | 99.05 | 98.35 |
| 4/pea | 74.24 | 69.82 | 93.06 | 94.70 | 92.14 |
| 5/alfalfa | 58.30 | 70.48 | 96.36 | 97.76 | 95.65 |
| 6/wheat | 60.21 | 65.83 | 96.26 | 98.84 | 96.90 |
| 7/beet | 28.81 | 55.14 | 89.33 | 93.39 | 88.75 |
| 8/bare land | 33.31 | 67.11 | 95.09 | 97.53 | 95.45 |
| 9/stem bean | 45.44 | 69.60 | 88.91 | 93.66 | 86.42 |
| 10/water | 58.91 | 73.86 | 86.31 | 99.88 | 99.91 |
| 11/potato | 36.76 | 87.51 | 94.67 | 95.44 | 94.19 |
| OA | 54.10 | 70.92 | 93.58 | 96.76 | 94.48 |
| Kappa | 0.4843 | 0.6720 | 0.9275 | 0.9634 | 0.9377 |

For Foulum dataset, we can see that OA of the original coherence matrix and TSNE manifold are 61.71% and 80.05% respectively, and classification accuracy of the latter is almost 20% higher than that of the former. However, when they are fused with deep multi-scale spatial features, they all get better classification performance, the corresponding OA are increased to 89.49% and 91.31% respectively. Compared with FCN, T3-FCN only makes contribution to forest classification, and causes deterioration of different degrees to other categories. Especially for town, the classification accuracy has reduced by nearly 7%, and OA on the whole dataset has decreased to 89.49%. Apart from the fact that accuracy of broad-leaved crop and town is slightly lower than that of FCN, the proposed TSNE-FCN has achieved best classification effect on other categories, also reflected in visual classification of Figure 11. On the basis that FCN alone has excellent classification performance, the fusion of polarized T3 and FCN feature has positive synergistic effect on Flevoland dataset, while significant exclusion occurs on Foulum dataset. Compared with the instability of T3 feature, the combination of manifold representation mined by TSN and FCN spatial features always maintains a robust promoting role.

In the same way, displayed in Table 3, the classification performance of San Francisco dataset under different comparison algorithms shows a quite similar pattern with the first two datasets: FCN alone has already realized uncommon classification effect with OA as high as 92.23%, but the collaborative effect can still improve the classification accuracy. The experiment results in Table 4 has revealed that FCN performances well. Except for fine stem crop, the classification accuracy of other categories is basically higher than 90%, and that of broad-leaved crop is even as high as 97.84%. It has indicated that deep spatial characteristics learned by FCN play a leading role in distinguishing different types of ground objects. In terms of classification accuracy, the performance of two shallow features (T3 polarized information and manifold subspace representation) are weaker than that of FCN network, but the accuracy on low-density area under FCN is the lowest, only 83.53%. Furthermore, after merging two shallow features with FCN's deep spatial features respectively, their classification effects have been improved to different extent. Both T3-FCN and TSNE-FCN's classification accuracy on almost every category have risen, especially for vegetation, the corresponding OA are 9.31% and 10.98% higher than that of FCN, showing that shallow features can boost the discrimination ability of deep multi-scale spatial feature. From the confusion matrix in Table 7, it can be seen that only a few pixels are misclassified by TSNE-FCN. At the same time, it also achieves the highest classification accuracy in all comparison schemes.

Table 3. Category accuracy (%), OA (%) and kappa coefficient of different features (classifier: SVM) on San Francisco dataset. (The category with highest accuracy is shown in bold.)

| Categories | T3 | TSNE | FCN | T3-FCN | TSNE-FCN |
|---------------------------|--------------|--------|--------|--------|---------------|
| 1/high-density urban area | 69.48 | 71.21 | 93.08 | 95.20 | 95.72 |
| 2/vegetation | 77.33 | 79.84 | 82.66 | 91.97 | 93.64 |
| 3/ocean | 86.28 | 91.91 | 96.73 | 99.04 | 99.57 |
| 4/developed urban area | 88.32 | 90.42 | 93.39 | 95.53 | 96.03 |
| 5/low-density urban area | 88.57 | 84.38 | 83.53 | 82.46 | 86.09 |
| OA | 81.03 | 84.53 | 92.23 | 95.96 | 96.77 |
| Kappa | 0.7363 | 0.7838 | 0.8908 | 0.9430 | 0.9546 |

Table 4. Category accuracy (%), OA (%) and kappa coefficient of different features (classifier: SVM) on Foulum dataset. (The category with highest accuracy is shown in bold.)

| Categories | T3 | TSNE | FCN | T3-FCN | TSNE-FCN |
|---------------------|--------|--------|--------------|--------|---------------|
| 1/broad-leaved crop | 60.20 | 82.53 | 97.84 | 95.45 | 97.18 |
| 2/fine-stem crop | 49.76 | 68.98 | 83.88 | 82.17 | 84.14 |
| 3/bare land | 61.93 | 81.65 | 89.86 | 89.03 | 89.97 |
| 4/town | 55.37 | 71.57 | 90.12 | 83.46 | 87.38 |
| 5/forest | 73.60 | 88.78 | 92.88 | 94.08 | 95.49 |
| OA | 61.71 | 80.05 | 91.06 | 89.49 | 91.31 |
| Kappa | 0.5157 | 0.7468 | 0.8868 | 0.8667 | 0.8899 |

Figures 10–12 display visual classification results of aforementioned three PolSAR datasets under different algorithms. It can be noted that when T3 polarization information is directly adopted for classification, the obtained result is not smooth enough and it seems ‘noise’ occurs in the classification map. This is because of the scattering mechanism of the original PolSAR image, the coherent echoes of distributed targets are superimposed on different pixel points. Meanwhile, some areas in classification maps are magnified for more clear visualization, as shown in Figure 13. On these three PolSAR datasets, the classification result of FCN is relatively smooth, but it is obvious that there are some misclassified pixels on the boundary. Quite obvious in Figure 13, the green part in San Francisco dataset is almost unrecognized. As shown in Flevoland classification map Figures 10 and 13, there are some misclassified samples at the junction of different categories, represented by different colors, and the edges of different color blocks (categories) present apparent hybrid. Since FCN has a large receptive field and can hardly perceive local structure in the image, it cannot deal with details such as edges well. However, by embedding the manifold subspace representation into FCN spatial features, the defect of FCN can be enhanced obviously. With the cooperation of FCN and TSNE together, the classification accuracy of these three PolSAR datasets are increased to 94.48%, 91.31% and 96.77% respectively.

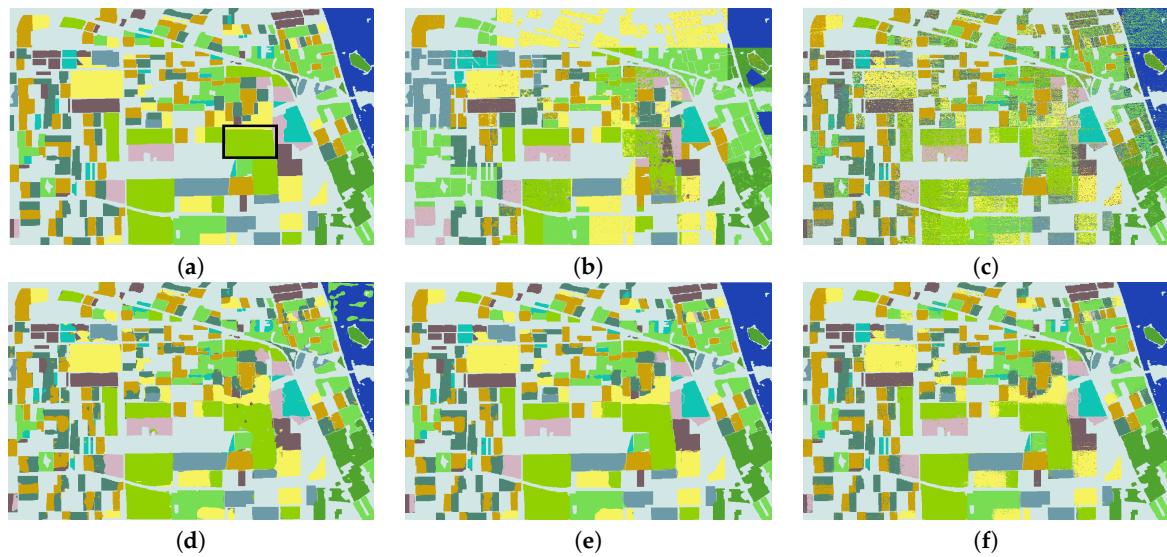


Figure 10. Visual classification of different algorithms on Flevoland dataset. (a) GT. (b) T3: 54.10%. (c) TSNE: 70.92%. (d) FCN: 93.58%. (e) T3-FCN: 96.76%. (f) TSNE-FCN: 94.48%.

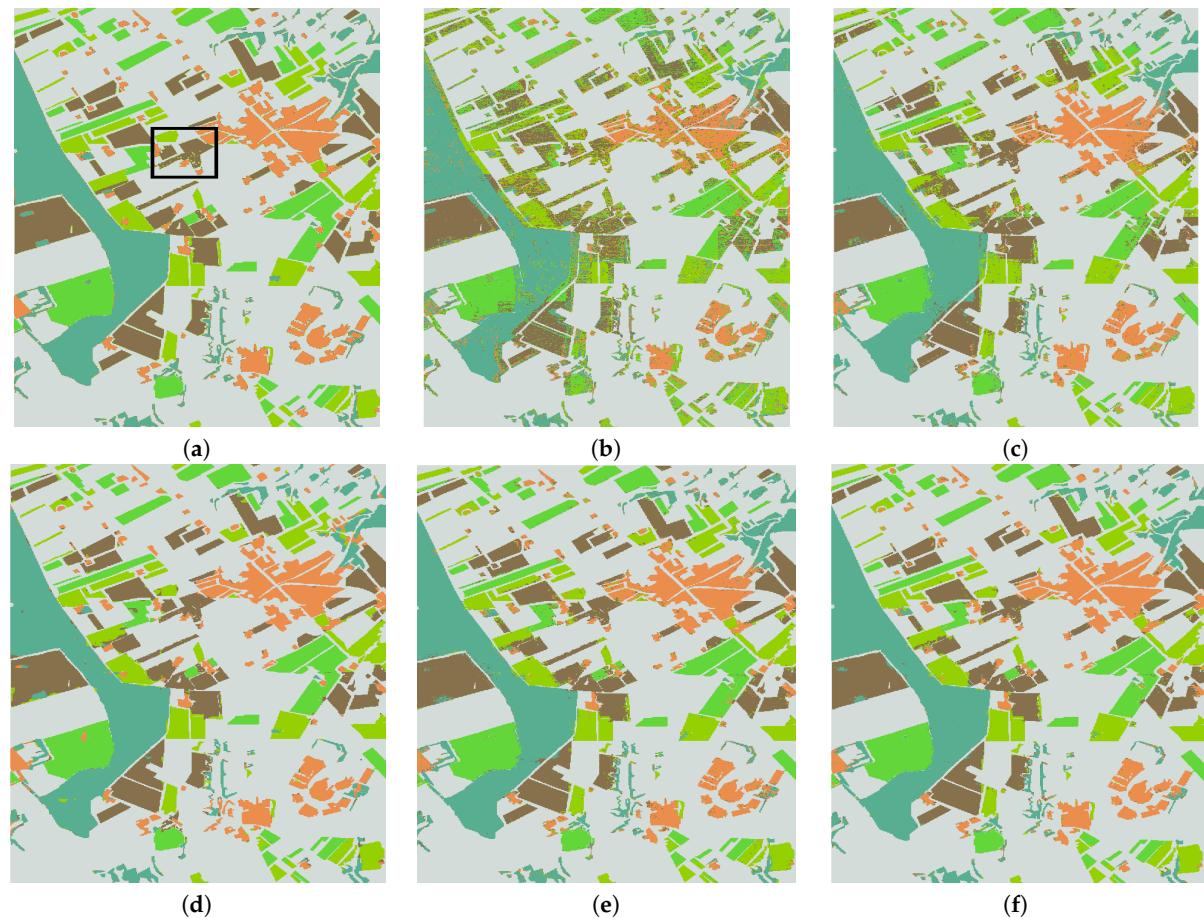


Figure 11. Visual classification of different algorithms on Foulum dataset. (a) GT. (b) T3: 61.71%. (c) TSNE: 80.05%. (d) FCN: 91.06%. (e) T3-FCN: 89.49%. (f) TSNE-FCN: 91.31%.

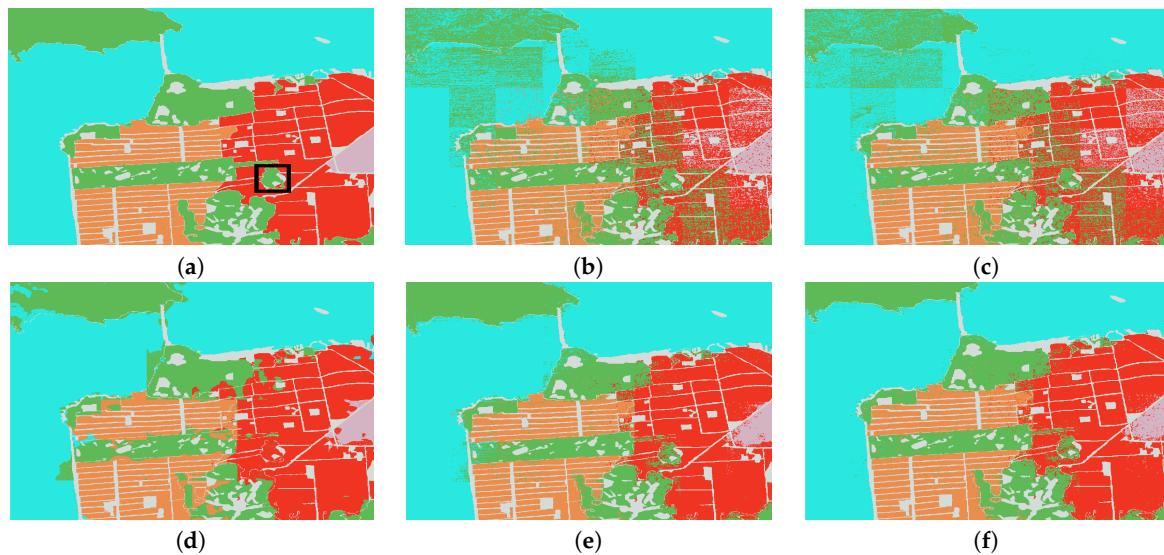


Figure 12. Visual classification of different algorithms on San Francisco dataset. (a) GT. (b) T3: 81.03%. (c) TSNE: 84.53%. (d) FCN: 92.23%. (e) T3-FCN: 95.96%. (f) TSNE-FCN: 96.77%.

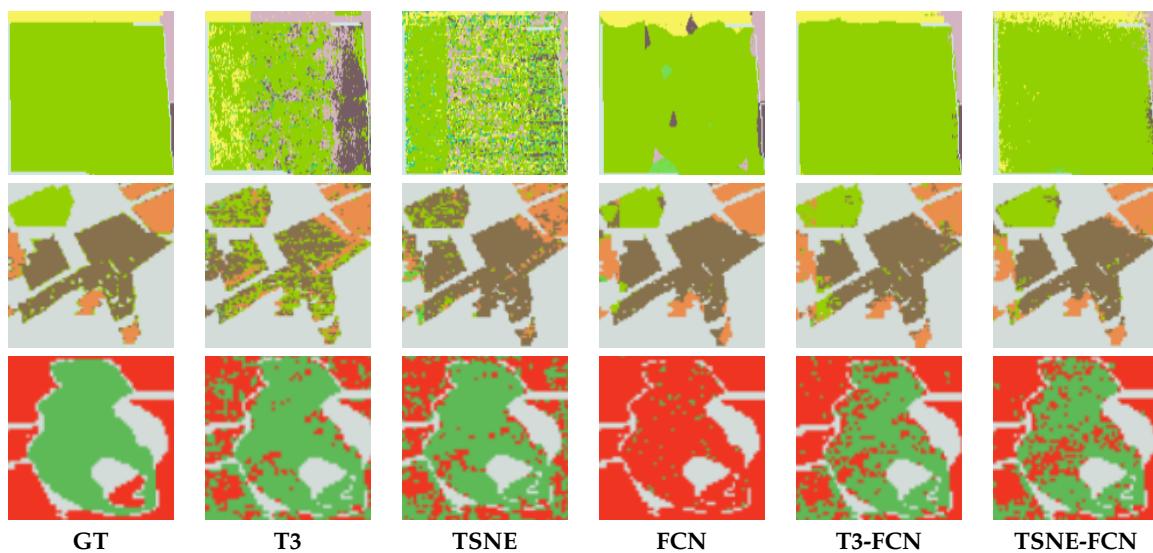


Figure 13. Three rows are magnified classification areas of Flevoland, Foulum and San Francisco datasets in order, the magnified areas are marked ‘black’ in their corresponding GT (see Figures 10–12). Best viewed on screen.

Tables 5–7 present confusion matrix of the proposed TSNE-FCN algorithm on three PolSAR datasets. The classification accuracy of different categories are basically over 85%, OA of some categories are even up to 97%, proving that the algorithm proposed in this paper has outstanding classification performance.

Table 5. Confusion matrix (%), category accuracy is shown in bold) of TSNE-FCN+SVM algorithm on Flevoland dataset.

| Classes | Rapeseed | Grassland | Forest | Pea | Alfalfa | Wheat | Beet | Bare Land | Stem Bean | Water | Potato |
|-----------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| rapeseed | 93.89 | 1.25 | 0 | 0.24 | 0.63 | 1.44 | 1.26 | 0.61 | 0.16 | 0 | 0.52 |
| grassland | 0.71 | 95.40 | 0.19 | 0.03 | 0.21 | 0.61 | 0.96 | 0.31 | 0.61 | 0.04 | 0.92 |
| forest | 0.05 | 0.69 | 98.35 | 0 | 0.21 | 0.04 | 0.24 | 0.01 | 0.01 | 0.15 | 0.26 |
| pea | 0.96 | 0.20 | 0.01 | 92.14 | 0.02 | 2.12 | 2.60 | 0.61 | 0.82 | 0 | 0.53 |
| alfalfa | 0.96 | 0.64 | 0.15 | 0 | 95.65 | 1.37 | 0.30 | 0.35 | 0.15 | 0 | 0.44 |
| wheat | 0.93 | 0.50 | 0 | 0.52 | 0.48 | 96.90 | 0.13 | 0.22 | 0.02 | 0 | 0.28 |
| beet | 1.36 | 1.70 | 0.11 | 0.52 | 0.12 | 0.51 | 88.75 | 0.54 | 0.88 | 0.02 | 5.49 |
| bare land | 1.12 | 0.83 | 0.01 | 0.35 | 0.27 | 0.62 | 0.65 | 95.45 | 0.11 | 0 | 0.59 |
| stem bean | 0.58 | 3.41 | 0 | 0.60 | 0.07 | 0.33 | 4.73 | 0.69 | 86.42 | 0 | 3.17 |
| water | 0.02 | 0.04 | 0.01 | 0 | 0 | 0 | 0.02 | 0 | 0 | 99.91 | 0 |
| potato | 0.29 | 0.73 | 0.06 | 0.16 | 0.12 | 0.14 | 3.42 | 0.22 | 0.65 | 0.01 | 94.19 |

Table 6. Confusion matrix (%), category accuracy is shown in bold) of TSNE-FCN+SVM algorithm on Foulum dataset.

| Classes | Broad-Leaved Crop | Fine-Stem Crop | Bare Land | Town | Forest |
|-------------------|-------------------|----------------|--------------|--------------|--------------|
| broad-leaved crop | 97.18 | 1.18 | 0.57 | 0.50 | 0.57 |
| fine-stem crop | 1.40 | 84.14 | 7.21 | 4.21 | 3.04 |
| bare land | 0.45 | 2.49 | 89.97 | 4.18 | 2.91 |
| town | 0.92 | 3.14 | 4.00 | 87.38 | 4.56 |
| forest | 0.43 | 1.39 | 1.14 | 1.55 | 95.49 |

Table 7. Confusion matrix (%), category accuracy is shown in bold) of TSNE-FCN+SVM algorithm on San Francisco dataset.

| Classes | High-Density Urban Area | Vegetation | Ocean | Developed Urban Area | Low-Density Urban Area |
|-------------------------|-------------------------|--------------|--------------|----------------------|------------------------|
| high-density urban area | 95.72 | 2.81 | 0.06 | 0.95 | 0.47 |
| vegetation | 3.16 | 93.64 | 1.21 | 1.99 | 0 |
| ocean | 0.03 | 0.37 | 99.57 | 0.03 | 0 |
| developed urban area | 0.67 | 3.21 | 0.10 | 96.03 | 0 |
| low-density urban area | 13.91 | 0 | 0 | 0 | 86.09 |

4.4. Discussion

Aimed at essential representation of high dimensional polarization characteristics, an effective PolSAR imagery classification method is proposed in this paper. The nonlinear manifold space representation and deep multiscale spatial features learned by FCN are fused for classification. In the experiment, some interesting points are discussed.

- The feature effectiveness. The experimental results of FCN alone have shown that spatial features learnt by migrated FCN are of great significance in classification. Meanwhile, considering the coherence principle of PolSAR image, nonlinear manifold learning is adopted to model ground objects more effectively. Two types of features are fused for classification and good results are achieved. However, the original features in our work do not take phase information into account. Phase information might play an important role in distinguishing some object types, in future research, how to encode phase information into pseudo color images should be explored.
- The synergism of FCN and nonlinear manifold learning. From the experimental results of Flevoland and Foulum datasets, it can be found that a method that shows synergistic effect on one dataset may presents a totally different pattern on the other dataset. On the one hand, it is due to the property of data itself. But on the other hand, this also reflects the instability of the ‘collaborative’ method, such as T3-FCN. However, our method has been showing robust enhancement, which demonstrates the superiority of the proposed scheme.

- In fact, the success of our approach can provide a common framework to integrate any other existing approach that is consistent with our theory. We can substitute a better deep neural network for FCN, TSNE can also be replaced if the new manifold is good enough. However, most manifold methods are not explicitly expressed. Embedding manifold into neural networks through approximation for joint optimization can be a good direction. Current valuable work will lay a solid foundation for the research of automatic classification network.

5. Conclusions

Traditional PolSAR interpretation mainly relies on statistical and scattering features based on electromagnetic wave interaction, to establish the imaging physical model. With improvement of PolSAR image's resolution, this underlying feature shows limitations of poor adaptability and low utilization. Therefore, researchers have considered the introduction of feature learning method for PolSAR image classification. However, when applying deep learning models pre-trained on optical images to PolSAR data, over fitting and weak generalization are prone to occur. Therefore, from the perspective of manifold hypothesis and mapping classification, this paper studies to construct effective classification features that adapt to PolSAR data characteristics. In the proposed method, low-dimensional representation learned by nonlinear manifold method is embedded into FCN's deep multi-scale spatial features, then the fused features are fed into the discriminant model SVM for classification. Compared with convolutional transform in common network, nonlinear manifold learning can map high-dimensional features to its deep core subspace. In this way, the manifold feature can describe the original PolSAR data more efficiently, so as to complete effective classification. The comparative experiments on a series of PolSAR datasets have verified the following points: (1) In contrast to underlying polarized features, feature engineering can realize abstraction for them and produce more complete representations; (2) Drawing on transfer learning, FCN can adaptively learn deep multi-scale spatial features of PolSAR image, which plays an import role for final classification; (3) Nonlinear manifold learning reveals the most essential structure of high-dimensional polarimetric data, and forms complementary advantages with FCN. Their synergistic effect can improve the representation and discrimination ability of the fused features. In the future, the proposed method can be integrated into the remote sensing interpretation system, or serve as the cornerstone of following research for PolSAR imagery classification.

Author Contributions: C.H. and M.T. conceived and designed the experiments; D.X. performed the experiments and analyzed the results; C.H. and M.T. wrote the paper; M.L. revised the paper. All authors have read and agree to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (No. 2016YFC0803000) and the National Natural Science Foundation of China (No. 41371342 and No. 61331016).

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

References

1. Lee, J.S.; Pottier, E. *Polarimetric Radar Imaging: From Basics to Applications*; CRC Press: Boca Raton, FL, USA, 2009.
2. Lee, J.S.; Grunes, M.R.; Pottier, E.; Ferro-Famil, L. Unsupervised terrain classification preserving polarimetric scattering characteristics. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 722–731.
3. Kong, J.; Swartz, A.; Yueh, H.; Novak, L.; Shin, R. Identification of terrain cover using the optimum polarimetric classifier. *J. Electromagn. Waves Appl.* **1988**, *2*, 171–194.
4. Novak, L.; Burl, M. Optimal speckle reduction in polarimetric SAR imagery. *IEEE Trans. Aerosp. Electron. Syst.* **1990**, *26*, 293–305. [[CrossRef](#)]
5. Cloude, C.S. Group Theory and Polarisation Algebra. *Optik* **1986**, *75*, 26–36.
6. Freeman, A.; Durden, S.L. A Three-Component Scattering Model for Polarimetric SAR Data. *IEEE Trans. Geosci. Remote Sens.* **2002**, *36*, 963–973. [[CrossRef](#)]

7. Zhang, L.; Zhang, L.; Du, B. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 22–40. [[CrossRef](#)]
8. Hansch, R.; Hellwich, O. Classification of Polarimetric Sar Data by Complex Valued Neural Networks. 2014. Available online: https://www.isprs.org/proceedings/xxxviii/1_4_7-W5/paper/Haensch-147.pdf (accessed on 28 December 2019).
9. He, C.; Liu, X.; Feng, D.; Shi, B.; Luo, B.; Liao, M. Hierarchical terrain classification based on multilayer bayesian network and conditional random field. *Remote Sens.* **2017**, *9*, 96. [[CrossRef](#)]
10. Sellami, A.; Farah, M.; Farah, I.R.; Solaiman, B. Hyperspectral imagery classification based on semi-supervised 3-D deep neural network and adaptive band selection. *Expert Syst. Appl.* **2019**, *129*, 246–259. [[CrossRef](#)]
11. Li, Y.; Zhang, H.; Shen, Q. Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network. *Remote Sens.* **2017**, *9*, 67. [[CrossRef](#)]
12. Goodman, J. Statistical Properties of Laser Speckle Patterns. In *Laser Speckle and Related Phenomena*; Springer: Berlin/Heidelberg, Germany, 1975.
13. Lee, J.S. Speckle suppression and analysis for synthetic aperture radar images. In Proceedings of the International Conference on Speckle, San Diego, CA, USA, 20–23 August 1985.
14. Song, H.; Yang, W.; Zhong, N.; Xu, X. Unsupervised Classification of PolSAR Imagery via Kernel Sparse Subspace Clustering. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1487–1491. [[CrossRef](#)]
15. Tu, S.T.; Chen, J.Y.; Yang, W.; Sun, H. Laplacian Eigenmaps-Based Polarimetric Dimensionality Reduction for SAR Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 170–179. [[CrossRef](#)]
16. Roweis, S.T.; Saul, L.K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **2000**, *290*, 2323–2326. [[CrossRef](#)] [[PubMed](#)]
17. Balasubramanian, M.; Schwartz, E.L. The Isomap Algorithm and Topological Stability. *Science* **2002**, *295*, 7. [[CrossRef](#)] [[PubMed](#)]
18. Ainsworth, T.L.; Lee, J.S. Optimal Polarimetric Decomposition Variables-non-linear Dimensionality Reduction. In Proceedings of the IGARSS 2001—IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217), Sydney, Australia, 9–13 July 2001; pp. 928–930.
19. Ainsworth, T.L.; Lee, J.S. Polarimetric SAR Image Classification Exploiting Optimal Variables Derived from Multiple Image Data Sets. In Proceedings of the IGARSS 2004, 2004 IEEE International Geoscience and Remote Sensing Symposium, Anchorage, AK, USA, 20–24 September 2004; pp. 188–191.
20. Belkin, M.; Niyogi, P. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Comput.* **2003**, *15*, 1373–1396. [[CrossRef](#)]
21. Shi, L.; Zhang, L.; Yang, J.; Zhang, L.; Li, P. Supervised Graph Embedding for Polarimetric SAR Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 216–220. [[CrossRef](#)]
22. Mhaskar, H. Approximation properties of a multilayered feedforward artificial neural network. *Adv. Comput. Math.* **1993**, *1*, 61–80. [[CrossRef](#)]
23. Olah, C. Neural Networks, Manifolds, and Topology. Available online: <https://colah.github.io/posts/2014-03-NN-Manifolds-Topology/> (accessed on 10 January 2020).
24. Wang, Y.; He, C.; Liu, X.; Liao, M. A hierarchical fully convolutional network integrated with sparse and low-rank subspace representations for polsar imagery classification. *Remote Sens.* **2018**, *10*, 342. [[CrossRef](#)]
25. Jonathan, L.; Evan, S.; Trevor, D. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
26. Yamaguchi, Y.; Moriyama, T.; Ishido, M.; Yamada, H. Four-component scattering model for polarimetric SAR image decomposition. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 1699–1706. [[CrossRef](#)]
27. Hinton, G. Visualizing high-dimensional data using t-sne. *Vigiliae Christ.* **2008**, *9*, 2579–2605.
28. Bengio, Y. Deep Learning of Representations for Unsupervised and Transfer Learning. *JMLR: Workshop Conf. Proc.* **2012**, *27*, 17–36.

