Article

A Comparative Review of Manifold Learning Techniques for Hyperspectral and Polarimetric SAR Image Fusion

Jingliang Hu [1,2,*] , **Danfeng Hong** [1,2] , **Yuanyuan Wang** [2] **and Xiao Xiang Zhu** [1,2,*]

1   Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 82234 Wessling, Germany; danfeng.hong@dlr.de
2   Signal Processing in Earth Observation (SiPEO), Technical University of Munich (TUM), 80333 Munich, Germany; y.wang@tum.de
*   Correspondence: jingliang.hu@dlr.de (J.H.); xiaoxiang.zhu@tum.de (X.X.Z.)

check for updates

Abstract: In remote sensing, hyperspectral and polarimetric synthetic aperture radar (PolSAR) images are the two most versatile data sources for a wide range of applications such as land use land cover classification. However, the fusion of these two data sources receive less attention than many other, because of their scarce data availability, and relatively challenging fusion task caused by their distinct imaging geometries. Among the existing fusion methods, including manifold learning-based, kernel-based, ensemble-based, and matrix factorization, manifold learning is one of most celebrated techniques for the fusion of heterogeneous data. Therefore, this paper aims to promote the research in hyperspectral and PolSAR data fusion, by providing a comprehensive comparison between existing manifold learning-based fusion algorithms. We conducted experiments on 16 state-of-the-art manifold learning algorithms that embrace two important research questions in manifold learning-based fusion of hyperspectral and PolSAR data: (1) in which domain should the data be aligned—the *data domain* or the *manifold domain*; and (2) how to make use of existing *labeled data* when formulating a graph to represent a manifold—supervised, semi-supervised, or unsupervised. The performance of the algorithms were evaluated via multiple accuracy metrics of land use land cover classification over two data sets. Results show that the algorithms based on manifold alignment generally outperform those based on data alignment (data concatenation). Semi-supervised manifold alignment fusion algorithms performs the best among all. Experiments using multiple classifiers show that they outperform the benchmark data alignment-based algorithms by ca. 3% in terms of the overall classification accuracy.

Keywords: data fusion; generalized graph fusion; hyperspectral image; data alignment; locality preserving projections; manifold alignment; manifold learning; MAPPER-induced manifold alignment; polarimetric SAR; manifold alignment; MIMA

1. Introduction

1.1. Related Work

Multi-modal data fusion [1–7] continuously draws attention in the remote sensing community. The fusion of optical and synthetic aperture radar (SAR) data, two important yet intrinsically different data sources, has also began to appear frequently in the context of multi-modal data fusion [8–14]. With the rapid development of Earth observation missions, such as the Sentinel-1 [15], Sentinel-2 [16], and the upcoming EnMAP [17], the availability of both data sources create a huge potential for Earth-oriented information retrieval. Among all optical data [18–20], hyperspectral data are well

known for their distinguishing power that originates from their rich spectral information [21–24]. Similarly, polarimetric SAR (PolSAR) data are a popular choice for classification task in the field of SAR because it can reflect the geometric and the dielectric property of the scatterers [25–29]. It is of great interest to investigate the fusion of hyperspectral and PolSAR images, especially with the application to land use land cover classification (LULC).

Few studies attempted to address the challenge of fusing hyperspectral and PolSAR data. Jouan and Allard [30] proposed a hierarchical fusion strategy for land cover classification using PolSAR and hyperspectral images. In their work, hyperspectral images are firstly used to distinguish vegetation and non-vegetation area. The PolSAR data are used to classify the non-vegetation area to man-made objects, water, or bare soil. Li et al. [31] applied feature level and decision level fusion using hyperspectral and PolSAR data. They combined the parameters of scattering mechanism of the PolSAR data and the features of the hyperspectral image to create a concatenation of features. The classification results of multiple classifiers are then merged using decision fusion. An application of spilled oil detection was studied by Dabbiru et al. [32] using hyperspectral and PolSAR data. They applied pixel level concatenation on the data, and employed support vector machine (SVM) as the classifier. Hu et al. [33] proposed a framework for fusing hyperspectral and PolSAR data based on the segmented objects that provide spatial information. A two-stream convolutional neural network (CNN) was introduced in [34] that takes advantage of the feature extraction power of CNN.

Among the existing fusion methods, including manifold learning-based [35,36], kernel-based [37], ensemble-based [38,39], tensor-based [40,41], and matrix factorization [42], manifold learning is one of most celebrated techniques. However, although it has been proven as a powerful technique in the field of data fusion, it is barely studied in the scope of fusing hyperspectral and PolSAR data. Generally, manifold-based fusion techniques attempt to find a shared latent space where the original data sets can be fused or aligned. Wang and Mahadevan [35,43–45] proposed several manifold-based techniques to find the correspondence of data sets which describe the same object from different aspects via the latent space. A kernel based manifold alignment [46] searches the latent space from a kernel space of the original data, because the kernel space has a better representation of the data than the original feature space of the data. In remote sensing, it was introduced in [36,47] that the manifold latent space is able to align multiple optical data sets and improve the LULC classification. A manifold-based data alignment technique was introduced in [48] for the fusion of hyperspectral and Lidar data with application to classification. Besides data fusion, various manifold techniques can be found in remote sensing field for detection [49], visualization [50], and dimension reduction [51].

1.2. Scope of This Paper

When fusing data with manifold techniques, one technical question is that: in which domain should the fusion be carried on? We categorized the existing techniques into two types: (1) data alignment-based approach and (2) manifold alignment-based approach. As shown in the left of Figure 1, the data alignment approach carries out the fusion in the data domain. As the simplest example, it fuses the data by concatenation, and carries out a manifold-based dimension reduction. Essentially, this approach assumes that an intrinsic manifold exists in the concatenated data. Representatives of this approach are the locality preservation projection (LPP) [52] and the generalized graph fusion (GGF) [48]. On the contrary, the manifold alignment-based approach carries out the fusion on manifolds which are separately derived from different data sources. This is demonstrated in the right of Figure 1. The assumption of this approach is that different manifold exists in each data source. Those manifolds can be aligned in a latent space. Representative algorithms are the manifold alignment (MA) [36] and the MAPPER-Induced manifold alignment (MIMA) [53].
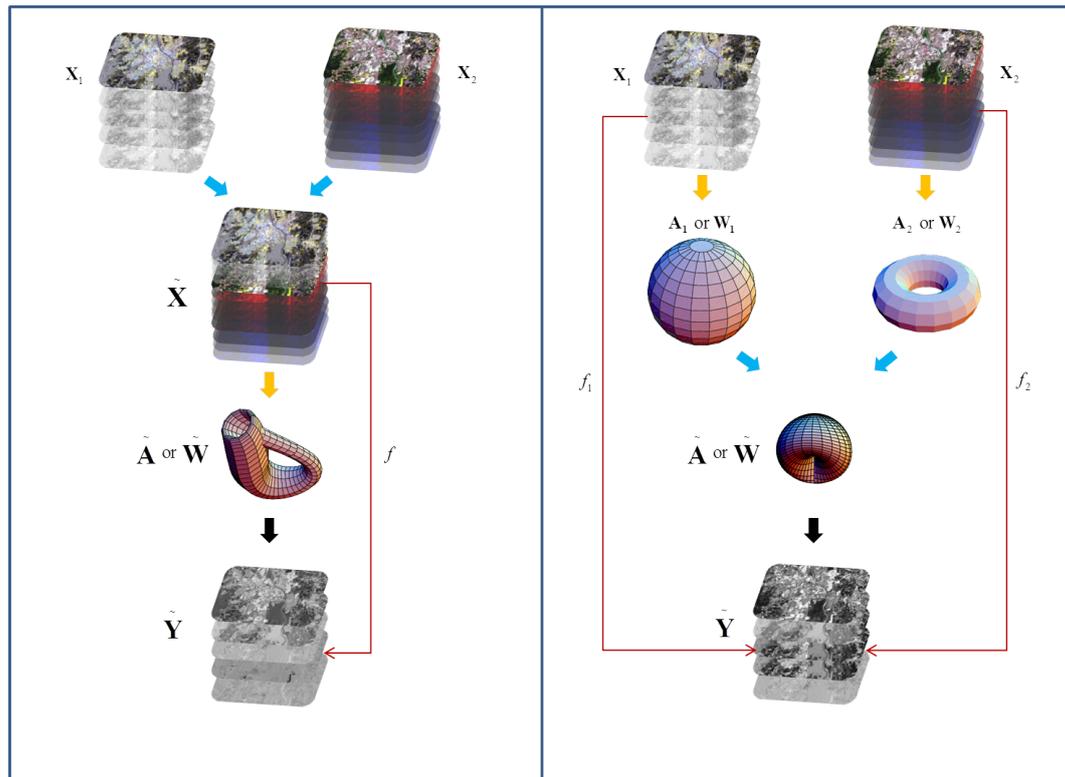
Figure 1. Frameworks of manifold learning fusion techniques. Left: the data alignment fusion; Right: the manifold alignment fusion. The blue arrow indicates the fusion step. The yellow arrow indicates where the modeling of manifold takes place. The black arrow indicates the feature extraction. $\mathbf{X}_i$: ith data source; $\mathbf{A}$ or $\mathbf{W}$: mathematical modeling of manifolds; $\mathbf{Y}$: fused feature; $\mathbf{f}$: the learned projection. $\tilde{}$: a fusion of certain form.

The other essential research question of manifold-based fusion is: how should the manifold be extracted? We categorize the existing techniques into three learning strategies in terms of the usage of labeled data: unsupervised, semi-supervised, and supervised. When modeling a manifold, a general assumption is that, hidden in the data representation, there exists an underlying lower dimensional manifold where the data truly distributes [54]. Early studies [54–57] model manifolds by following a geometric assumption that the Riemannian manifold can be locally approximated by Euclidean measures. The geometric assumption models the manifold in an unsupervised manner by using k-nearest-neighbor (kNN). With the presence of labeled data, the manifold can be jointly modeled by the Riemannian manifold and the labeled data. For example, one can construct the manifold in a semi-supervised fashion by using both kNN and the labeled data [35,36]. The manifold can also be modeled in a supervised manner by using only the labeled data. One of the main goals of this paper is to investigate the impact of these learning strategies on the classification performance on the fused data.

1.3. Contribution of This Paper

This paper investigates the performance of manifold learning technique on the fusion of hyperspectral and PolSAR data, based on four state-of-art algorithms, locality preservation projection (LPP) [52], generalized graph fusion (GGF) [48], manifold alignment (MA) [36,44], and MAPPER-induced manifold alignment (MIMA) [53]. We implemented 16 variants of the four algorithms which involve the above-mentioned two alignment approaches and the three manifold learning strategies. These algorithms were tested on two study areas for a LULC classification task with five classifiers: one nearest neighbor (*1NN*) [58], linear SVM (*LSVM*) [59,60], Gaussian kernel SVM (*KSVM*) [59,60], random forest (*RF*) [61], and canonical correlation forest (*CCF*) [62]. We avoided

any deep network classifiers, because the goal of this article is to solely evaluate the performance of multi-sensory data fusion. In total, 80 classification maps were produced for each study area, based on which a comprehensive discussion was carried out. The main contributions of this paper are as follows:

- An exhaustive investigation of existing manifold learning techniques. A sufficient number of manifold techniques and classifiers were tested on the fusion of hyperspectral and PolSAR data in terms of classification. It provides a reliable demonstration on the performance of the manifold technique regarding hyperspectral and PolSAR data fusion.
- An objective comparison of the performance of different manifold data fusion algorithms. To avoid any fortuity, five classifiers were applied for the classification. A grid search was applied to all tunable hyperparameters of those algorithms. The best classification accuracies are compared.
- A comprehensive analysis of the results. The experiment results were analyzed in regard to two fusion approaches, three manifold learning strategies, four basic algorithms, and five classifiers.

1.4. Structure of This Paper

The second section recalls the theory of manifold technique and the four selected state-of-art algorithms. The third section describes the data sets used in the experiments, introduces the experiment setting, and carries out the discussion. The fourth section concludes the paper. Table 1 also lists the symbols used in this article for a better understanding of the content of the article.

Table 1. The notations used in this article.

Notation	Explanation
$\mathbf{X}_i$	The ith data source
$\mathcal{M}_i$	The manifold of the $\mathbf{X}_i$
$\mathbf{x}_i^p$	The pth instance of the $\mathbf{X}_i$
m_i	The number of dimensions of the $\mathbf{X}_i$
$\mathbf{E}_i$	The labeled subset of the $\mathbf{X}_i$
$\mathbf{y}_i^p$	The pth instance of the $\mathbf{Y}_i$
l_i	The number of dimensions of the $\mathbf{Y}_i$
$\mathbf{F}$	The filter function in MAPPER
$\mathbf{W}$	The weight matrix that models a manifold
$\mathbf{D}$	The degree matrix of a graph
$\tilde{\ }$	The fusion at certain form
$\mathcal{L}$	The loss fuction
dn	The dimension of underlying manifold
b	The number of bins in MAPPER
λ	The eigenvalue of generalized eigenvalue decomposition
K	The total number of data sources
$\mathbf{Y}_i$	The data representation of the $\mathcal{M}_i$
$\mathbf{x}_i^q$	The qth instance of the $\mathbf{X}_i$
n_i	The number of instances of the $\mathbf{X}_i$
n_i^*	The number of instances of the $\mathbf{E}_i$, $n_i^* < n_i$
$\mathbf{y}_i^q$	The qth instance of the $\mathbf{Y}_i$
$\mathbf{f}$	The projection $\mathbf{Y} = \mathbf{f}^T \mathbf{X}$
$\mathbf{A}$	The binary matrix that models a manifold
σ	The filtering parameter of weight matrix
$\mathbf{L}$	The Laplacian matrix of a graph
$\mathcal{D}$	The pairwise distance matrix
k	The number of local neighbors
μ	The weighting of topology structure in MA
c	The overlap rate in MAPPER

2. Materials and Methods

In this section, the general concept of the manifold technique is introduced, with the help of necessary mathematical notations. Meanwhile, the theoretical impact of different learning strategies

to the fusion result is discussed. Then, the following up sub-sections recall the principles of the four selected state-of-art manifold fusion techniques, namely LPP [52], GGF [48], MA [36,44], and MIMA [53]. Pseudo-codes of these four algorithms are listed in the Appendixes A–D, which provides the technical details.Finally, the data sets and the experiment settings are introduced in detail.

2.1. Manifold Technique, Learning Strategy, and Notations

Let $\mathbf{X}_i = [\mathbf{x}_i^1, ..., \mathbf{x}_i^p, ..., \mathbf{x}_i^{n_i}] \in \mathbb{R}^{m_i \times n_i}$ be a matrix representing the *i*th data source, with m_i dimensions by n_i instances. For simplicity, the subscript *i* is omitted in the following content when only one data source is involved. The m_i-dimensional data space is named as the feature space of data $\mathbf{X}_i$ in this paper. The term $\mathbf{x}_i^p$ denotes the *p*th instance of the *i*th data source. Let *K* denote the total number of data sources.

A manifold $\mathcal{M}$ is a smooth hyper-surface embedded in a higher dimensional space [56], e.g., the surface of a sphere is a 2D manifold in a 3D space. The underlying assumption of the manifold technique is that, for a data $\mathbf{X} \in \mathbb{R}^{m \times n}$ of redundant *m* dimensions, there exists a low dimensional intrinsic manifold $\mathcal{M}$ where the data distributes [54,57,63,64]. The goal of a manifold technique is to pursue a representation, realized by a projection $\mathbf{Y} = [\mathbf{y}^1, ..., \mathbf{y}^p, ..., \mathbf{y}^n] \in \mathbb{R}^{l \times n}, l < m$ of the original data, of the manifold $\mathcal{M}$. In order to approximate $\mathbf{Y}$, the bridging property is that the data point $\mathbf{y}^p$ on the manifold is locally homeomorphic to its counterpart $\mathbf{x}^p$ in the feature space [56]. It means that a data point has identical local structures in its intrinsic manifold and in its feature space. With this property, a variety of methods [54,55,57,65] extract the local structure of a data [52,66–69] in its feature space as an estimation to the local structure in its intrinsic manifold, with different locality criterion. All those methods pursue an optimized projection $\mathbf{f}$ which maps data from the feature space to a representation ($\mathbf{Y} = \mathbf{f}^T \mathbf{X}$) of the intrinsic manifold $\mathcal{M}$. In terms of the manifold technique for data fusion [36,44,48], the aim is to find the projection which maps multiple data sources $\{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_K\}$ into a fused manifold $\tilde{\mathcal{M}}$ where the fused data locates.

The centerpiece of the abovementioned algorithms is the modeling of the manifold. Usually, an intrinsic manifold of the data is modeled by an $n \times n$ symmetric binary matrix $\mathbf{A}$ that describes the connection among the data points. $\mathbf{A}(p,q) = 1$ for a confirmed connection between $\mathbf{x}_i^p$ and $\mathbf{x}_i^q$ while $\mathbf{A}(p,q) = 0$ otherwise. $\mathbf{A}$ can be generalized to an $n \times n$ symmetric weight matrix $\mathbf{W}$. Different from $\mathbf{A}$, $\mathbf{W}(p,q)$ takes a real value in $[0,1]$, which describes the strength of the connection between $\mathbf{x}_i^p$ and $\mathbf{x}_i^q$. Essentially, $\mathbf{A}$ and $\mathbf{W}$ are the adjacency and weight matrix of a graph that captures the topology of the manifold. As introduced in [52], the manifold structure ($\mathbf{A}$ or $\mathbf{W}$) can be defined from different perspectives. In this paper, we would like to categorize these perspectives based on how the labeled data is utilized for modeling the manifold, namely the unsupervised learning, supervised learning, and semi-supervised learning.

- The unsupervised learning takes the original geometric assumption that the manifold and the original data space share the same local property. Besides the geometric measure, model-based similarity measurement can also be used to build up the structure of the manifold. The key point is that the definition of the similarity measurement is capable of revealing the underlying distribution of the data or the physical information in the data.

- The supervised learning assumes that a given set of labeled data includes sufficient amount of inter- and intra-class connections among the data points, so that they can well capture the topology of the manifold. As a result, the underlying manifold is directly defined by the label information. Thus, the quality of the label has a great impact.

- The semi-supervised learning pursues a manifold where the data distribution partially correlates to the label information and partially associates to the distribution predefined by a similarity measurement. This manifold implicitly propagates the label information to the unlabeled data.

2.2. Locality Preservation Projection (LPP)

LPP aims to find a lower dimensional representation $\mathbf{Y}$ of the original data $\mathbf{X}$ which reflects the intrinsic manifold $\mathcal{M}$. According to the geometric assumption that the intrinsic manifold and the original data share the same local properties, the lower dimensional representation $\mathbf{Y}$ achieved by LPP preserves the local structure of the original data $\mathbf{X}$. The locality is defined by either the k-nearest-neighbor or the ϵ-neighborhoods [52] and is mathematically described in a weight matrix $\mathbf{W}$ as Equation (1).

$$\mathbf{W}(p,q) = \begin{cases} e^{-\frac{||x^p - x^q||^2}{\sigma}} & x^p \text{ and } x^q \text{ are local neighbors} \\ 0 & x^p \text{ and } x^q \text{ are not local neighbors} \end{cases} \tag{1}$$

where σ is a filtering parameter.

LPP pursues an optimized projection $\mathbf{f}$ which maps the data $\mathbf{X}$ to a lower dimensional representation $\mathbf{Y} = \mathbf{f}^T\mathbf{X}$. As the local structure of the intrinsic manifold is modeled by Equation (1), minimizing the objective function expressed by Equation (2) encourages the preservation of the derived local structure in the intrinsic manifold.

$$\mathcal{L} = \sum_{pq}(y^p - y^q)\mathbf{W}(p,q) = \sum_{pq}(\mathbf{f}^T x^p - \mathbf{f}^T x^q)\mathbf{W}(p,q). \tag{2}$$

Thus, the optimization is formulated as follow:

$$\min_{\mathbf{f}} \sum_{pq}(\mathbf{f}^T x^p - \mathbf{f}^T x^q)\mathbf{W}(p,q). \tag{3}$$

Proven in [52], the solution that minimizes the objective function $\mathcal{L}(\mathbf{f})$ is given by the minimum eigenvalue solution to the generalized eigenvalue problem expressed in Equation (4).

$$\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{f} = \lambda\mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{f}, \tag{4}$$

where $\mathbf{D}$ is the degree matrix; if $p = q$, $\mathbf{D}(p,q) = \sum_{p=1}^{p=n}\mathbf{W}(p,q)$, otherwise, $\mathbf{D}(p,q) = 0$; and the $\mathbf{L}$ is the Laplacian matrix, $\mathbf{L} = \mathbf{D} - \mathbf{W}$.

As brief described above, LPP is originally designed to as a dimension reduction algorithm, instead of data fusion. However, it is essential to include this algorithm in the scope of this paper, because (1). When conducting manifold fusion, the dimension reduction is also accomplished as a side effect. Due to the well-known curse-of-dimensionality [70], classification on selective subset of dimensions could result in better performance than using the data with all dimensions [71]. LPP can serve as a baseline algorithm to reduce the dimension of the data; and (2). LPP is essentially a manifold learning technique. Some data fusion algorithms [48,72] are developed on the idea of data alignment using the LPP.

2.3. Generalized Graph-Based Fusion (GGF)

GGF is originally proposed to fuse hyperspectral data and LiDAR data for land cover classification [48]. Its fusion strategy comprises a joint LPP dimension reduction and an additional constraint that captures the common local structure that exists in all data sources.

Technically, GGF concatenates K data sources ($\mathbf{X}_i = [\mathbf{x}_i^1, ..., \mathbf{x}_i^p, ..., \mathbf{x}_i^n] \in \mathbb{R}^{m_i \times n}, i \in \{1, 2, ..., K\}$) into a stack ($\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}^1, ..., \tilde{\mathbf{x}}^p, ..., \tilde{\mathbf{x}}^n] \in \mathbb{R}^{(m_1 + m_2 + ... + m_K) \times n}$) which are treated as one data sources in its high dimensional feature space. Therefore, GGF is essentially a LPP carried out on the data stack $\tilde{\mathbf{X}}$, with an additional constraint. The constraint assumes that the connectivity $\tilde{\mathbf{A}}$ of the fused intrinsic manifold $\tilde{\mathcal{M}}$ should be a complete subset of the connectivity matrices of the manifolds $\mathcal{M}_i$ of the individual data sources $\mathbf{X}_i, i \in \{1, 2, ..., K\}$. Thus, the assumption is formulated as Equation (5).

$$\tilde{\mathbf{A}} = \mathbf{A}_1 \odot \mathbf{A}_2 \odot, ..., \odot \mathbf{A}_K \tag{5}$$

where $\odot$ indicates element-wise multiplication.

The manifold constraint $\tilde{\mathbf{A}}$ is embedded into a n by n pairwise distance matrix $\tilde{\mathcal{D}}$ ($\tilde{\mathcal{D}}(p,q) = ||\tilde{\mathbf{x}}^p - \tilde{\mathbf{x}}^q||$), which is expressed by Equation (6) where $\neg$ means logical operator negative, and $max(\cdot)$ means the maximum value of all elements in '$\cdot$'. The distance between any two data points that are not connected according to $\tilde{\mathbf{A}}$ is penalized with the maximum distance value of $\tilde{\mathcal{D}}$. The final distance matrix is named as $\tilde{\mathcal{D}}_{\mathbf{GGF}}$.

$$\tilde{\mathcal{D}}_{\mathbf{GGF}} = \tilde{\mathcal{D}} + (\neg\tilde{\mathbf{A}})max(\tilde{\mathcal{D}}) \tag{6}$$

The weight matrix $\tilde{\mathbf{W}}$ of the intrinsic manifold is then as follows.

$$\tilde{\mathbf{W}}(p,q) = \begin{cases} e^{-\tilde{\mathcal{D}}_{GGF}(p,q)} & x^p \text{ and } x^q \text{ are local neighbors} \\ 0 & x^p \text{ and } x^q \text{ are not local neighbors} \end{cases} \tag{7}$$

After achieving the weight matrix $\tilde{\mathbf{W}}$, similar to the LPP, the optimized projection $\mathbf{f}$ is given by the minimum eigenvalue solution to the generalized eigenvalue problem in Equation (8).

$$\tilde{\mathbf{X}}\tilde{\mathbf{L}}\tilde{\mathbf{X}}^T\mathbf{f} = \lambda\tilde{\mathbf{X}}\tilde{\mathbf{D}}\tilde{\mathbf{X}}^T\mathbf{f} \tag{8}$$

where $\tilde{\mathbf{D}}$ is the degree matrix. If $p = q$, $\tilde{\mathbf{D}}(p,q) = \sum_{p=1}^{p=n}\tilde{\mathbf{W}}(p,q)$, otherwise, $\tilde{\mathbf{D}}(p,q) = 0$. $\tilde{\mathbf{L}}$ is the Laplacian matrix, $\tilde{\mathbf{L}} = \tilde{\mathbf{D}} - \tilde{\mathbf{W}}$.

2.4. Manifold Alignment (MA)

Manifold alignment [35,36,44] aims to learn a set of projections $\{\mathbf{f}_1, ..., \mathbf{f}_K\}$ that (1) apply to individual data sources $\mathbf{X}_i$ in order to obtain their individual manifolds $\mathcal{M}_i$, and (2) align those obtained manifolds $\{\mathcal{M}_1, ..., \mathcal{M}_K\}$ to each other.

Designed in [36,44], three properties hold in the fused manifold: (*a*) data of the same class should locate close to each other; (*b*) data of different classes should locate far from one another; and (*c*) the intrinsic manifolds of individual data are preserved. These three properties are respectively formulated by the following three connection matrices $\tilde{\mathbf{A}}_s$ (9), $\tilde{\mathbf{A}}_d$ (10), and $\tilde{\mathbf{A}}_g$ (11).

$$\tilde{\mathbf{A}}_s = \begin{pmatrix} \mathbf{A}_s^{1,1} & ... & \mathbf{A}_s^{1,K} \\ ... & ... & ... \\ \mathbf{A}_s^{K,1} & ... & \mathbf{A}_s^{K,K} \end{pmatrix} \tag{9}$$

The connection matrix of similarity (9) is computed by the labeled information to pursue property (*a*).

$$\tilde{\mathbf{A}}_d = \begin{pmatrix} \mathbf{A}_d^{1,1} & ... & \mathbf{A}_d^{1,K} \\ ... & ... & ... \\ \mathbf{A}_d^{K,1} & ... & \mathbf{A}_d^{K,K} \end{pmatrix} \tag{10}$$

The connection matrix of dissimilarity is modeled as (10) to accomplish property (*b*), which is also computed from the label.

$$\tilde{\mathbf{A}}_g = \begin{pmatrix} \mathbf{A}_g^{1,1} & 0 & 0 \\ 0 & ... & 0 \\ 0 & 0 & \mathbf{A}_g^{K,K} \end{pmatrix} \tag{11}$$

The connection matrix (11) describes the manifolds of individual data sources by using kNN, which aims at the property (*c*). All of the matrices (9)–(11) have the size of $(n_1 + n_2 + ... + n_k) \times (n_1 + n_2 + ... + n_k)$. In each matrix, the superscript i, j, e.g., $\mathbf{A}^{i,j}$, represents the relationship between the ith and jth data sources.

With connection matrices (9)–(11), three terms are formulated as Equations (12)–(14) to preserve the three properties, respectively.

$$\mathcal{A} = \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{p=1}^{n_i} \sum_{q=1}^{n_j} \|\mathbf{f}_i^T x_i^p - \mathbf{f}_j^T x_j^q\|^2 \tilde{\mathbf{A}}_s^{i,j}(p,q). \tag{12}$$

Minimizing Equation (12) pulls data of the same class together, which meets property (*a*).

$$\mathcal{B} = \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{p=1}^{n_i} \sum_{q=1}^{n_j} \|\mathbf{f}_i^T x_i^p - \mathbf{f}_j^T x_j^q\|^2 \tilde{\mathbf{A}}_d^{i,j}(p,q). \tag{13}$$

Maximizing Equation (13) pushes data of different classes away, which meets property (*b*).

$$\mathcal{C} = \sum_{i=1}^{K} \sum_{p=1}^{n_i} \sum_{q=1}^{n_i} \|\mathbf{f}_i^T x_i^p - \mathbf{f}_i^T x_i^q\|^2 \tilde{\mathbf{A}}_g^{i,i}(p,q). \tag{14}$$

Minimizing Equation (14) preserves the geometric structure of individual data sources, which corresponds to property (*c*). The terms (12)–(14) jointly construct the objective unction (15):

$$\mathcal{L}(\mathbf{f}_1, ..., \mathbf{f}_K) = (\mathcal{A} + \mathcal{C})/\mathcal{B}, \tag{15}$$

and hence an optimization problem (16) can be written as

$$\min_{\mathbf{f}_1, ..., \mathbf{f}_K} \mathcal{L}(\mathbf{f}_1, ..., \mathbf{f}_K). \tag{16}$$

Proven in [35], the solution $\{\mathbf{f}_1, ..., \mathbf{f}_K\}$ that minimizing the cost function $\mathcal{L}(\mathbf{f}_1, ..., \mathbf{f}_K)$ is given by the smallest non-zero eigenvectors of the generalized eigenvalue decomposition of (17).

$$\tilde{\mathbf{X}}(\mu\tilde{\mathbf{L}}_g + \tilde{\mathbf{L}}_s)\tilde{\mathbf{X}}^T\mathbf{f} = \lambda\tilde{\mathbf{X}}\tilde{\mathbf{L}}_d\tilde{\mathbf{X}}^T\mathbf{f}, \tag{17}$$

where

$$\tilde{\mathbf{X}} = \begin{pmatrix} \mathbf{X}_1 & 0 & ... & 0 \\ ... & ... & ... & ... \\ 0 & ... & 0 & \mathbf{X}_K \end{pmatrix},$$

$$\tilde{\mathbf{L}}_{\{s,d,g\}} = \tilde{\mathbf{A}}_{\{s,d,g\}} - \tilde{\mathbf{D}}_{\{s,d,g\}},$$

$$\tilde{\mathbf{D}}_{\{s,d,g\}}(p,q) = \begin{cases} \sum_{q=1}^{m_1+...+m_k} \tilde{\mathbf{A}}_{\{s,d,g\}}(p,q) & p = q \\ 0 & p \neq q \end{cases}.$$

The matrices $\tilde{\mathbf{D}}$ and $\tilde{\mathbf{L}}$ with subscript s, d, and g are the degree matrices and the Laplacian matrices, respectively.

2.5. MAPPER-Induced Manifold Alignment (MIMA)

MIMA is designed to fuse optical and PolSAR data for the purpose of LULC classification [53]. It follows the framework of MA [36,44] yet introduces a novel constraint term which originates from a recent field of topological data analysis (TDA). TDA has emerged as a new mathematical sub-field of big data analysis that aims to derive relevant information from the topological property of a data [73–77]. One TDA tool, named MAPPER [78], has been proven capable of revealing unknown insights in medical studies, by interpreting topological structures of data sets [79–82]. As a brief introduction, the MAPPER requires a filter function as an input which projects the data into a parameter space.

The original data is sorted into overlapping bins guided by the projected parameter. MAPPER carries out clustering of data points in each of the data bins, respectively. Afterwards, MAPPER models a graph where a node represents a cluster and an edge links two clusters that share common data points. Finally, a simplified graph is built up to represent the shape of the data. Such graph is an approximation of the Reeb graph [83] .

Technically, MIMA pursues the solution $\{\mathbf{f}_1, ..., \mathbf{f}_K\}$ by solving the same generalized eigenvalue decomposition as in Equation (17), except the the connection matrix of geometry $\tilde{\mathbf{A}}_g$ (Equation (11)) is replace by the MAPPER-derived connection matrix $\tilde{\mathbf{A}}_{MIMA}$ where $\mathbf{A}_{MIMA}^{i,i}(p,q) = 1$ if x_i^p and x_i^q belongs to the same cluster or belongs to two separated but linked clusters; $\mathbf{A}_{MIMA}^{i,i}(p,q) = 0$ elsewhere. Comparing to $\tilde{\mathbf{A}}_g$, $\tilde{\mathbf{A}}_{MIMA}$ introduces some unique properties that are listed as follows:

- Field knowledge. An expertise knowledge is introduced by the selection of the filter function. It defines a perspective of viewing the data while deriving the structure.
- A regional-to-global structure. Clustering in each data bin provides a regional structure. The design of overlapping bins combines the regional structures into a global one. It makes the derived structure more robust to outliers than the one derived by *kNN*.
- A data-driven regional structure. A spectral clustering is applied in the step, which is capable of detecting the number of clusters by the concept of eigen-gap [84]. It allows the derived structure constraining to the data distribution.

2.6. Data Description

Two sets of real data were used to investigate the manifold learning techniques for the fusion of hyperspectral and PolSAR data. The two data sets are in city of Berlin, Germany, and Augsburg, Germany.

2.6.1. The Berlin Data Set

In the Berlin data set, the hyperspectral image is a synthetic spaceborne EnMAP scene synthesized from airborne HyMap data. It has a size of 817 by 220 pixels, a 30-m ground sampling distance (GSD), and 244 spectral bands ranging from 400 nm to 2500 nm [85]. The dual-channel PolSAR data is a VH-VV polarized Sentinel-1 data acquired in interferometric wide swath mode. The Sentinel-1 SLC data is preprocessed using ESA SNAP toolbox and filtered by a non-local mean filter [86]. The PolSAR data has a GSD of 13 m and a size of 1723 by 476 pixels. The ground truth is a land use land cover map derived from Open Street Map (OSM) data [87]. The ground truth labels are spatially separated into a training data set and a testing data set shown in Figure 2. The details of the training and testing data sets are summarized in Table 2.

Table 2. Summary of the training data and the testing data for the scene of city Berlin.

Class	# of Training Sample	# of Testing Sample
Forest	298	52,455
Residential area	756	262,903
Industrial area	296	17,462
Low plants	344	56,683
Soil	428	14,505
Allotment	281	11,322
Commercial area	560	20,909
Water	153	5539
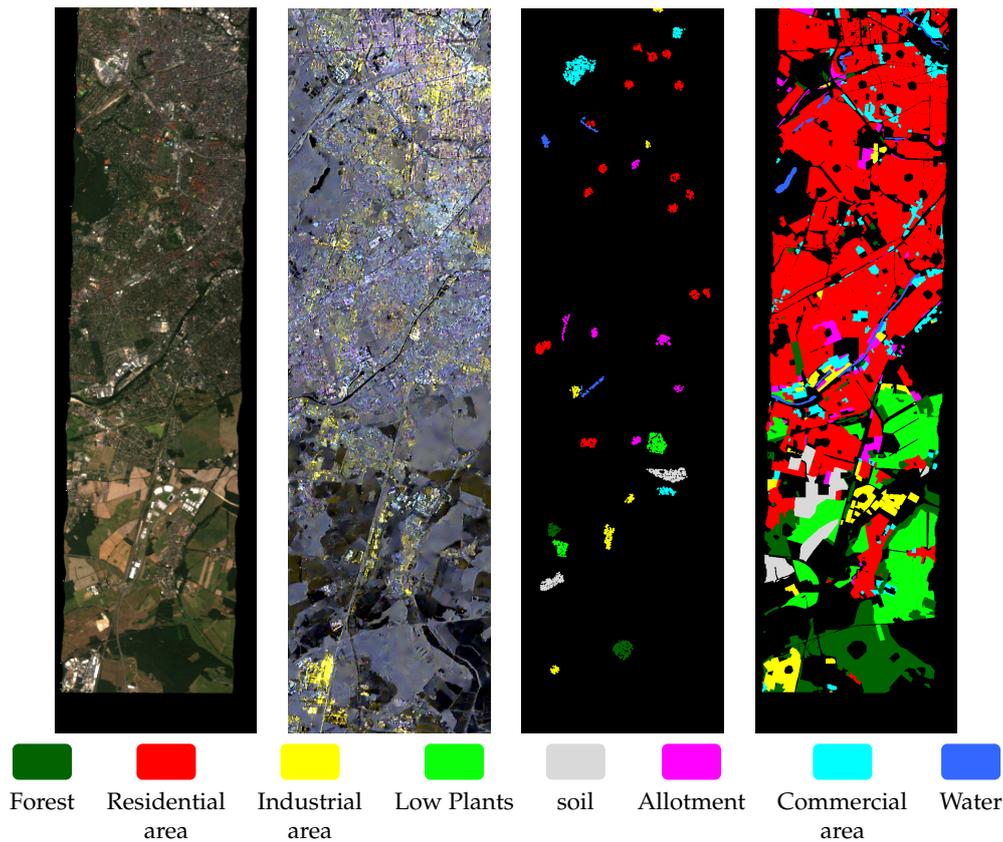
Figure 2. The Berlin data set. From left to right: RGB components of the simulated EnMAP data; Sentinel-1 dual-Pol data; The training data; The testing data.

2.6.2. The Augsburg Data Set

Similar to the Berlin data set, the hyperspectral image in the Augsburg data set is a synthetic spaceborne imagery simulated based on an airborne HySpex data. It has a GSD of 30 m, a size of 332 by 485 pixels, and 180 bands ranging from 400 nm to 2500 nm. Same as the Berlin data set, the PolSAR data is a VH-VV polarized Sentinel-1 image with a GSD of 10 m and a size of 997 by 1456 pixels. The training data and the testing data are shown in Figure 3 which are spatially separated. The details of the training and testing data sets are summarized in Table 3.

Table 3. Summary of the training data and the testing data for the scene of city Augsburg.

Class	# of Training Sample	# of Testing Sample
Forest	200	4100
Residential area	200	4100
Industrial area	200	4100
Low plants	200	4100
Soil	-	-
Allotment	200	4100
Commercial area	200	4100
Water	200	4100
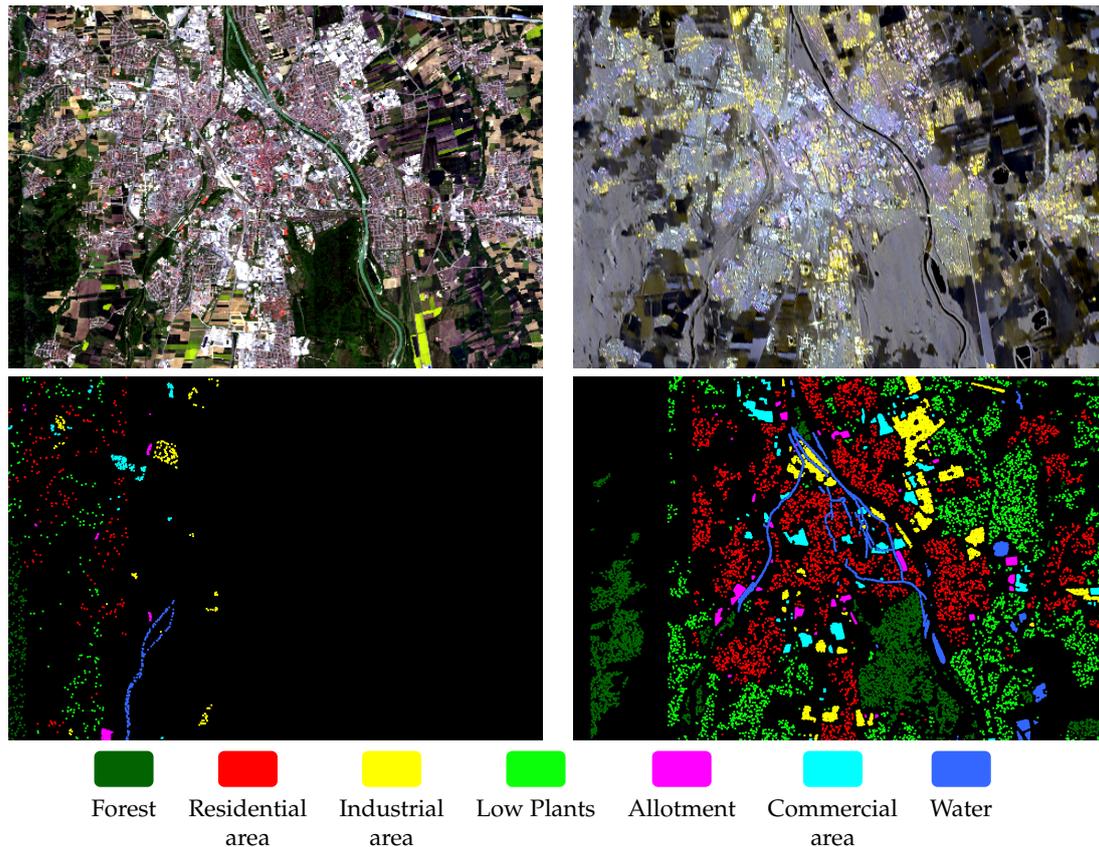
Figure 3. The Augsburg data set. From left to right, top to bottom: RGB components of the hyperspectral image; Sentinel-1 dual-Pol data; The training data; The testing data.

2.7. Experiment Setting

We start with a reasonable feature selection and extraction strategy from the original data, since it is well known that feature selection and extraction promote the classification performance of remote sensing data. The spectral-spatial feature extraction was employed for the hyperspectral image because of its excellent performance on classification tasks [88–91]. Specifically, the first four and six principal components (PCs) which occupy 99% of the variance of the data were extracted from the hyperspectral images of Berlin and Augsburg, respectively. The morphological profiles with radius of one, two, and three were employed to extract the spatial information on each PC. Thus, in total, 28 features and 42 features were extracted from the hyperspectral images of Berlin and Augsburg, respectively. For the feature extraction of Sentinel-1 dual-Pol data, four polarimetric features were extracted. They are the intensity of the VH channel, the intensity of the VV channel, the coherence of VV and VH, and the intensity ratio of VV and VH. Since the morphological profile was proven to promote classification of PolSAR [92,93], it is also used to extract spatial information from the four polarimetric features with radius equal to one, two, and three. In addition, the local statistics including the mean and standard deviation were extracted using a sliding window of 11 by 11 pixel on those four polarimetric features. In total, 36 features were extracted from the dual-Pol SAR data for both data sets of Berlin and Augsburg, respectively.

To carry out a comprehensive comparison of the fusion algorithms, in total 16 algorithms were implemented. Listed in Table 4, they are (1) PolSAR data only (POL), (2) hyperspectral image only (HSI), (3) feature stacking of hyperspectral and PolSAR data (HSI+POL), (4) data alignment using the original locality preserving projections (LPP) [52], (5) supervised version of LPP (LPP_SU), (6) semi-supervised version of LPP (LPP_SE), (7) the generalized graph-based fusion (GGF) [48], (8) supervised version of GGF (GGF_SU), (9) semi-supervised version of GGF (GGF_SE), (10) manifold

alignment (MA) [36,44], (11) unsupervised version of MA (MA_UN), (12) supervised version of MA (MA_SU), (13) MAPPER-Induced manifold alignment with first two principal components as filter functions (MIMA) [53], (14) unsupervised MIMA (MIMA_UN), (15) MIMA with local density as filter function (MIMA-D), and (16) unsupervised MIMA with local density as filter function (MIMA-D_UN).

Table 4. Technical summary of the selected algorithms. 'SU', 'UN', and 'SE' represent the learning strategy of supervised, unsupervised, or semi-supervised, respectively. **W** and **A** represent the weight matrix and the connection matrix, respectively. The hyperparameter set $\{k, dn, \mu, b\}$ indicates the number of neighbors, the number of dimensions, the topology weighting parameter, and the number of bins.

	Algorithm	Data		Learning Strategy			Fusion	Manifold	Hyper-
		HSI	POL	SU	UN	SE	Concept		Parameter
1	POL	-	✓	-	-	-	-	-	-
2	HSI	✓	-	-	-	-	-	-	-
3	HSI+POL	✓	✓	-	-	-	Concatenation	-	-
4	LPP	✓	✓	-	✓	-	data alignment	**W**	$\{k, dn\}$
5	LPP_SU	✓	✓	✓	-	-	data alignment	**W**	$\{k, dn\}$
6	LPP_SE	✓	✓	-	-	✓	data alignment	**W**	$\{k, dn\}$
7	GGF	✓	✓	-	✓	-	data alignment	**W**	$\{k, dn\}$
8	GGF_SU	✓	✓	✓	-	-	data alignment	**W**	$\{k, dn\}$
9	GGF_SE	✓	✓	-	-	✓	data alignment	**W**	$\{k, dn\}$
10	MA	✓	✓	-	-	✓	manifold alignment	**A**	$\{\mu, k, dn\}$
11	MA_UN	✓	✓	-	✓	-	Constrained dimension reduction	**A**	$\{k, dn\}$
12	MA_SU	✓	✓	✓	-	-	manifold alignment	**A**	$\{dn\}$
13	MIMA	✓	✓	-	-	✓	manifold alignment	**A**	$\{\mu, b, dn\}$
14	MIMA_UN	✓	✓	-	✓	-	Constrained dimension reduction	**A**	$\{b, dn\}$
15	MIMA-D	✓	✓	-	-	✓	manifold alignment	**A**	$\{\mu, b, dn\}$
16	MIMA-D_UN	✓	✓	-	✓	-	Constrained dimension reduction	**A**	$\{b, dn\}$

These manifold algorithms listed in Table 4 are categorized into the two approaches (data alignment or manifold alignment) mentioned in Section 1.2. LPP and GGF belong to the category of data alignment which concatenates data as a stack, and applies manifold learning on the stacked data. MA and MIMA belong to the category of manifold alignment which independently project K data sources to a latent space where the data are aligned.

The hyperparameters of each algorithm were tuned via a grid search, so that each algorithm reaches its best performance. The k was set in a range of 10 to 120 with an interval of 10. The number of dimension dn is set in a range of 5 to 50 with an interval of 5. The topology weighting parameter μ is set in a range of 0.5 to 3 with an interval of 0.5. The number of bins b is set in a range of 5 to 55 with an interval of 5.

After the data being fused, five different shallow classifiers were applied to the fused data set in the classification step. They are: one nearest neighbor (*1NN*) [58], linear SVM (*LSVM*) [59,60], Gaussian kernel SVM (*KSVM*) [59,60], random forest (*RF*) [61], and canonical correlation forest (*CCF*) [62]. The parameter tuning of *LSVM* is done in a heuristic procedure [60]. LIBSVM [94] is employed for the implementation of the *KSVM*. The number of trees was set as 40 for both *RF* and *CCF*.

3. Experiment Results

The discussion of experiment result mainly focus on the following three aspects:

- Manifold learning strategy. The experiment result supports the discussion of the impact that causes by different learning strategies, the unsupervised learning, the supervised learning, and the semi-supervised learning.
- Data fusion approach. The result supports the discussion of the two fusion approaches, the data alignment-based and the manifold alignment-based, for the fusion of the hyperspectral image and PolSAR data.

- Performance on classification. The experiment result reveals how manifold techniques perform on fusing hyperspectral images and PolSAR data and how different these manifold techniques perform.

The classification result is quantitatively evaluated by the class-specific accuracy, the average accuracy, the overall accuracy, and the kappa coefficient. The class-specific accuracy provides the percentage of correct predictions for a specific class. The average accuracy is the mean value of class-specific accuracy. The overall accuracy indicates the percentage of correctness for all predictions. And kappa coefficient also evaluates the overall correctness, yet is more robust than the overall accuracy [95].

3.1. Experiment on the Berlin Data Set

As shown in Figure 4 and Table 5, for the data alignment-based fusion algorithms (LPP and GGF), the unsupervised versions outperform the supervised and the semi-supervised versions. However, for the manifold alignment-based fusion algorithms (MA, MIMA, and MIMA-D), the semi-supervised versions have the best performance comparing to the supervised and the unsupervised ones. Surprisingly, in both type of fusion algorithms, the fully supervised strategy performs the worst.
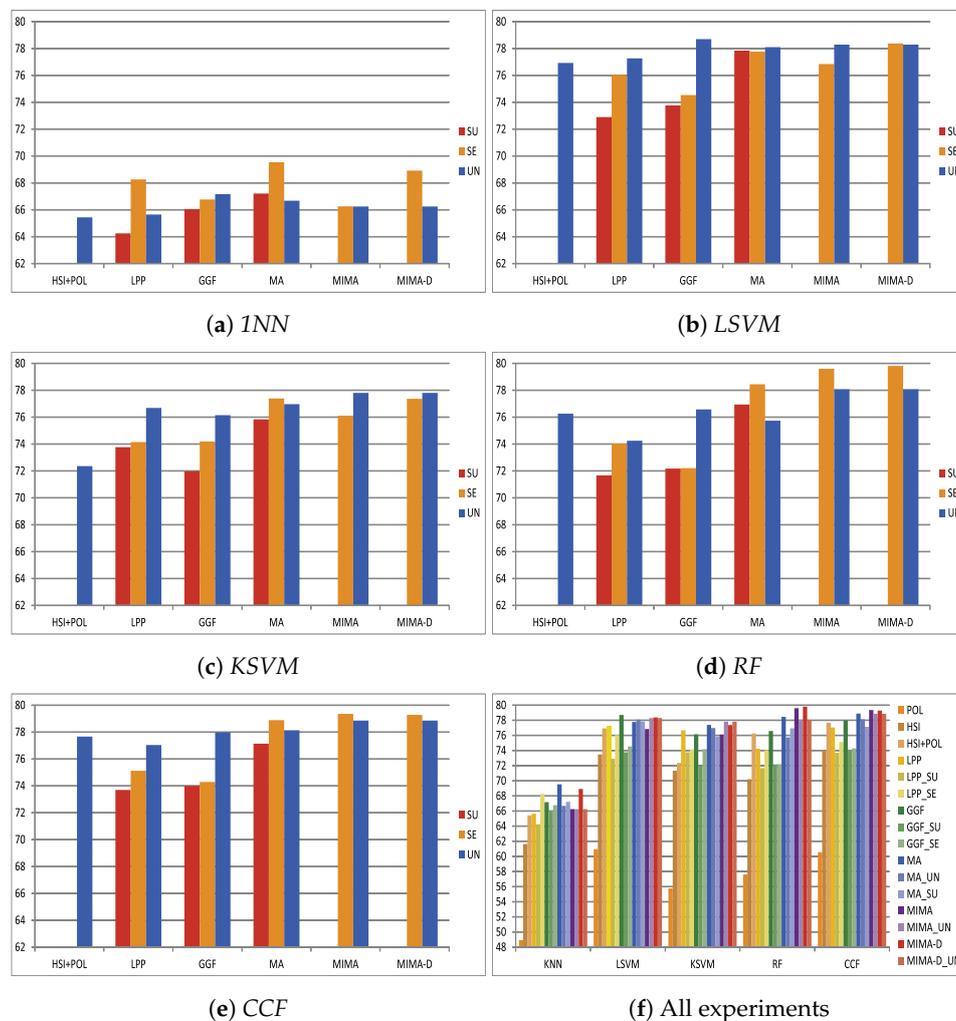
Figure 4. Comparison of the classification accuracies of different classifiers applied on the Berlin data set. Each chart is resulted from a corresponding classifier. The right bottom chart demonstrates all the overall accuracy resulted by applying five classifiers on every fused data achieved from the selected 16 algorithms. The *y*-axis report the overall accuracy in percentage (%). The 'SU', 'SE', and 'UN', represent supervised, semi-supervised, and unsupervised, respectively.

Taking the result of the simple concatenation (HSI+POL in Table 5) as reference, the data alignment-based fusion algorithms (LPP and GGF) marginally improve the classification accuracy. Sometimes the performance even drops below the reference accuracy. On the contrary, the manifold alignment-based fusion algorithms (MA, MIMA, and MIMA-D) have a more consistent improvement of the classification accuracy by ca. 3%. In fact, MIMA and MIMA-D have a considerable improvement comparing to LPP, GGF, and MA, especially when *RF* and *CCF* are employed as the classifier. This can be seen in Figure 4. Among all the algorithms, MIMA and MIMA-D have the best overall performance. Shown in Table 5, their best performance reach over 0.66, 65%, and 79%, for the kappa coefficient, the average accuracy, and the overall accuracy, respectively. For a visual comparison, Figure 5 plots the ground truth and the classification maps predicted by the 16 algorithms with *CCF*.
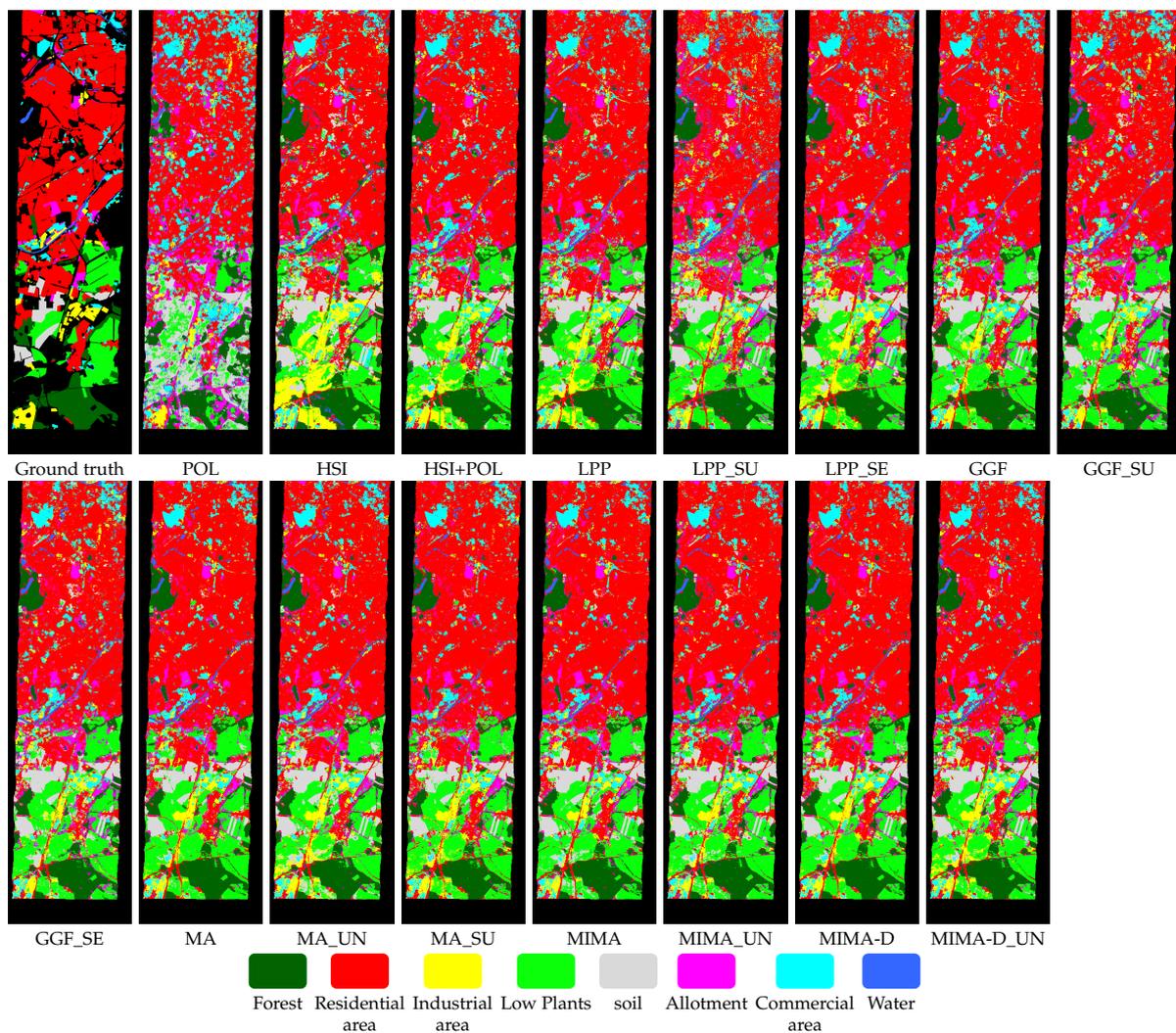
Figure 5. Visualization of the classification maps and the ground truth. The 16 classification maps are provided by applying *CCF* on fused data of the 16 algorithms, for the Berlin data set. Classification maps achieved by manifold alignment fusion methods are more accurate than the maps achieved by data alignment fusion methods.

Table 5. Quantitative performance comparison on the Berlin data, in terms of class-specific accuracy, kappa coefficient, average accuracy, overall accuracy, and mean overall accuracy. The mean overall accuracy is calculated based on the overall accuracies achieved by the five classifiers. The listed indications are achieved after hyperparameter tunning. The hyperparameters of each algorithm are listed under the name of the algorithm and their values are listed in the table. The kappa coefficient, average accuracy, and the overall accuracy that larger than 0.66, 65%, and 79% are marked in bold. And the three highest mean overall accuracies are also marked in bold.

Algorithm	Parameter	Classifiers	Forest	Residential Area	Industrial Area	Low Plants	Soil	Allotment	Commercial Area	Water	KAPPA	AA	OA	Mean OA
POL	-	1NN	40.64	57.67	25.14	32.94	56.88	32.19	30.37	33.85	0.2927	38.71	48.92	56.76
		LSVM	33.02	77.92	13.85	36.46	72.6	40.64	32.23	37.68	0.4012	43.05	60.94	
		KSVM	34.36	69.94	20.38	30.61	68.27	38.62	32.79	42.82	0.3566	42.23	55.76	
		RF	35.61	72.3	25.63	28.66	66.38	43.9	37.87	45.39	0.3789	44.47	57.61	
		CCF	37.96	76.87	24.87	30.69	64.72	38.82	36.88	41.34	0.4035	44.02	60.56	
HSI	-	1NN	68.78	63.87	30.01	57.58	90.73	55.76	32.86	73.89	0.4599	59.18	61.64	70.14
		LSVM	69.2	82.5	18.55	65.7	79.06	53.59	44.77	72.81	0.585	60.77	73.48	
		KSVM	72.58	78.68	35.43	63.74	74.18	56.87	31.58	74.29	0.5625	60.92	71.34	
		RF	66.65	79.64	30.25	57.44	75.33	47.77	35.17	78.1	0.5437	58.79	70.21	
		CCF	71	81.86	31.54	68.95	81.36	53.47	38.35	74.81	0.597	62.67	74.03	
HSI+POL	-	1NN	64.83	69.7	32.89	65.27	83.81	54.77	34.59	63.51	0.4975	58.67	65.44	73.73
		LSVM	66.57	86.24	30.48	75.3	79.61	53.52	40.12	76.11	0.6329	63.49	76.93	
		KSVM	67.27	80.93	41.78	64.02	72.37	57.58	33	74.6	0.5764	61.44	72.36	
		RF	63.46	84.99	37.79	74.38	82.72	56.26	40.61	82.09	0.6266	**65.29**	76.26	
		CCF	71.51	86.27	34.05	72.03	83.24	56.3	44.33	77.7	0.6445	**65.68**	77.67	
LPP $\{k, dn\}$	{60, 15}	1NN	69.53	69.07	34.56	66.09	80.27	57.51	32.18	64.56	0.5009	59.22	65.65	74.18
	{20, 30}	LSVM	70.1	87.05	32.52	70.97	79.26	58.88	36.48	72.61	0.6354	63.48	77.27	
	{30, 25}	KSVM	71.19	85.77	41.43	70.95	82.36	53.97	30.77	72.68	0.6297	63.64	76.69	
	{10, 20}	RF	56.2	85.87	28.9	69.28	76	49.9	38.64	67.07	0.5874	58.98	74.25	
	{10, 15}	CCF	68.41	86.68	34.35	71.96	80.07	54.07	37.54	75.93	0.6325	63.63	77.04	
LPP_SU $\{dn\}$	{10}	1NN	63.86	67.04	34.79	71.42	79.06	54.39	28.17	72.32	0.4817	58.88	64.25	71.26
	{30}	LSVM	64.41	81.51	34.12	70.1	81.56	56.74	29.1	71.38	0.578	61.11	72.9	
	{50}	KSVM	67.06	81.6	43.96	72.17	82.34	57.81	25.04	69.69	0.5908	62.46	73.77	
	{25}	RF	64.71	80.89	30.98	65.55	72.26	55.27	32.9	69.36	0.5596	58.99	71.67	
	{25}	CCF	64.25	81.99	33.72	74.47	75.59	55.89	33.77	69.76	0.5883	61.18	73.7	
LPP_SE $\{k, dn\}$	{80, 10}	1NN	68.22	72.17	38.92	73.21	73.43	58.09	30.65	74.02	0.5327	61.09	68.26	73.52
	{120, 40}	LSVM	64.68	85.37	38.15	74.36	79.63	59.18	29.75	77.41	0.6194	63.57	76.04	
	{120, 40}	KSVM	69.02	81.93	41.67	70.74	77	59.76	30.77	76.17	0.6001	63.38	74.15	
	{120, 30}	RF	66.96	83.15	29.66	72.12	66.45	56.39	34.17	74	0.5919	60.36	74.03	
	{120, 25}	CCF	64.86	85.09	34.63	71.85	66.83	56.05	34.33	75.05	0.6044	61.09	75.12	
GGF $\{k, dn\}$	{20, 30}	1NN	69.28	71.37	36.65	66.54	83.51	56.94	31.34	63.82	0.5186	59.93	67.17	75.31
	{90, 30}	LSVM	68.11	88.76	34.14	76.11	79.29	54.93	36.54	75.14	0.655	64.13	78.7	
	{20, 30}	KSVM	72.18	84.64	37.08	70.29	81.88	57.25	34.49	74.44	0.6254	64.03	76.15	
	{10, 20}	RF	68.97	86.55	29.13	70.39	81.23	49.45	41.85	62.88	0.6242	61.31	76.58	
	{10, 25}	CCF	70.53	87.51	31.29	76.34	70.86	51.95	42.06	67.95	0.6448	62.31	77.98	

Table 5. *Cont.*

Algorithm	Parameter	Classifiers	Forest	Residential Area	Industrial Area	Low Plants	Soil	Allotment	Commercial Area	Water	KAPPA	AA	OA	Mean OA
GGF_SU {dn}	{10}	1NN	65.57	69.99	37.73	68.89	80.13	51.96	28.71	76.62	0.5013	59.95	66.05	
	{50}	LSVM	63.6	82.87	36.49	69.8	82.34	56.58	29.62	76.22	0.5906	62.19	73.77	
	{50}	KSVM	69.99	80.63	46.43	60.43	77.21	53.92	25.15	78.77	0.5695	61.57	71.98	71.59
	{50}	RF	62.01	81.42	32.09	67.3	74.08	53.3	38.83	65.17	0.5678	59.28	72.17	
	{40}	CCF	65.54	83.4	31.38	70.58	72.26	51.15	37.02	68.24	0.5906	59.95	74	
GGF_SE {k,dn}	{10, 15}	1NN	66.96	70.63	36.07	69.65	80.62	55.65	29.49	76.35	0.5119	60.68	66.77	
	{120, 45}	LSVM	63.06	83.52	37.69	73.01	81.94	55.48	29.11	79.87	0.6007	62.96	74.54	
	{40, 40}	KSVM	70.19	82.26	41.52	67.92	80.35	54.38	31.22	82.51	0.5988	63.79	74.19	72.40
	{20, 40}	RF	65.27	80.56	34.49	67.01	75.85	54.57	38.72	66.98	0.5716	60.43	72.21	
	{70, 30}	CCF	60.15	83.94	35.07	74.18	74.3	51.22	35.51	68.64	0.5942	60.37	74.29	
MA {μ,k,dn}	{2, 90, 10}	1NN	69.83	73.8	38	75.68	69.64	60.09	29.41	72.27	0.5474	61.09	69.54	
	{2.5, 20, 25}	LSVM	65.49	86.97	37.63	79.08	80.06	55.63	34.46	73.37	0.6445	64.09	77.77	
	{2.5, 90, 35}	KSVM	69.38	85.81	37.49	78.3	80.54	55.42	33.29	73.21	0.6405	64.18	77.39	**76.40**
	{2, 10, 50}	RF	64.5	90.08	30.25	77.68	65.58	49.41	36.85	67.95	0.644	60.29	78.45	
	{2, 10, 20}	CCF	66.66	89.12	33.05	79.51	68.95	54.91	39.47	71.01	0.6557	62.84	78.89	
MA_UN {k,dn}	{120, 15}	1NN	68.46	69.61	32.87	72.87	78.51	54.88	34.76	67.95	0.5159	59.99	66.68	
	{90, 30}	LSVM	66.86	87.58	35.97	77.55	78.59	55.44	36.3	76.15	0.649	64.3	78.1	
	{40, 50}	KSVM	70.55	85.61	36.23	74.18	79.83	57.57	35.55	73.14	0.6346	64.08	76.97	75.13
	{100, 30}	RF	58.91	87.37	26.35	69.77	80.7	53.14	41.94	60.71	0.6079	59.86	75.74	
	{30, 30}	CCF	67	88.05	33.05	74.11	81.85	55	41.91	70.52	0.6467	63.94	78.14	
MA_SU {dn}	{5}	1NN	69.88	71.34	34.87	68.69	71.01	57.88	32.38	73.52	0.5199	59.94	67.21	
	{50}	LSVM	67.56	86.73	38.76	79.67	77.21	56.87	32.27	75.45	0.6457	64.31	77.85	
	{50}	KSVM	71.6	83.96	35.72	75.92	61.57	59.59	37.1	72.65	0.6204	62.26	75.84	75
	{50}	RF	60.53	87.82	33.22	77.13	70.16	52.42	38.82	63.66	0.6242	60.47	76.94	
	{50}	CCF	64.09	88.37	30.57	76.73	62.56	51.86	36.99	59.9	0.6257	58.89	77.14	
MIMA {μ,b,dn}	{1, 15, 5}	1NN	69.91	70.2	33.39	69.63	61.94	53.49	35.07	68.62	0.5055	57.78	66.26	
	{1, 15, 15}	LSVM	67.76	84.97	36.22	78.36	79.08	57.74	38	70.25	0.6328	64.05	76.85	
	{1, 15, 15}	KSVM	71.06	84.24	41.01	76.11	69.87	55.82	32.97	68.97	0.6233	62.51	76.11	**76.22**
	{1.5, 25, 40}	RF	65.1	90.31	32.54	80	82.77	50.79	35.08	71.01	**0.6642**	63.45	**79.6**	
	{2, 25, 20}	CCF	70.86	88.06	36.54	80.42	76.88	57.21	39.61	73.21	**0.667**	**65.35**	**79.36**	
MIMA_UN {b,dn}	{10, 20}	1NN	72.57	68.39	35.96	70.18	79.27	62.58	30.73	67.41	0.513	60.89	66.25	
	{10, 35}	LSVM	68.21	88.59	36.62	74.6	80.79	55.87	29.86	76.08	0.6495	63.83	78.29	
	{10, 35}	KSVM	71.78	87.1	36.85	73.13	82.31	58.05	31.79	73.14	0.6449	64.27	77.81	75.85
	{55, 30}	RF	67.92	88.44	27.36	77.22	81.32	50.9	35	61.04	0.6417	61.15	78.08	
	{30, 20}	CCF	71.06	88.19	29.72	77.55	79.81	55.71	39.99	69.67	0.658	63.96	78.86	
MIMA-D {μ,b,dn}	{1.5, 30 ,15}	1NN	71.31	72.3	35.31	74.51	76.66	57.37	33.48	71.84	0.5423	61.6	68.92	
	{1.5, 45, 20}	LSVM	67.59	86.85	36.8	81.07	78.3	56.4	38.97	75.88	0.6549	**65.23**	78.38	
	{2.5, 55, 30}	KSVM	70.01	85.33	36.79	78.84	78.52	56.83	36.44	76.08	0.6425	64.86	77.37	**76.75**
	{1, 30, 30}	RF	67.02	89.85	33.09	80.46	83.21	50.61	37.95	74.27	**0.6698**	64.56	**79.81**	
	{1, 45, 30}	CCF	68.91	89.18	34.79	78.63	75.48	51.74	39.85	69.45	**0.6628**	63.5	**79.28**	
MIMA-D_UN {b,dn}	{55, 15}	1NN	72.57	68.39	35.96	70.18	79.27	62.58	30.73	67.41	0.513	60.89	66.25	
	{55, 25}	LSVM	68.21	88.59	36.62	74.6	80.79	55.87	29.86	76.08	0.6495	63.83	78.29	
	{40, 20}	KSVM	71.78	87.1	36.85	73.13	82.31	58.05	31.79	73.14	0.6449	64.27	77.81	75.52
	{45, 30}	RF	67.92	88.44	27.36	77.22	81.32	50.9	35	61.04	0.6417	61.15	78.08	
	{45, 25}	CCF	71.06	88.19	29.72	77.55	79.81	55.71	39.99	69.67	0.658	63.96	78.86	

3.2. Experiment on Augsburg Data Set

The findings of the Augsburg data set are consistent with that of the Berlin data set. For the data alignment-based fusion algorithms, the unsupervised learning strategy works the best among the three learning strategies. For the manifold alignment-based fusion algorithms, the semi-supervised learning strategy performs the best. Comparing the results to that of simple concatenation (HSI+POL), the data alignment-based fusion (LPP and GGF) barely has any improvement, while the manifold alignment fusion has a 2% improvement comparing to the LPP and GGF. These findings can be seen in Figure 6, and Table 6. Among all the algorithms, combining MIMA or MIMA-D with *RF* or *CCF* provide the best classification performance. Their kappa coefficient, the average accuracy, and the overall accuracy, reach 0.56, 62.5%, and 62.5% respectively. A visual comparison of the results is shown in Figure 7. Similar to Figure 5, the classification maps were predicted by *CCF*.

(**a**) *1NN*

(**b**) *LSVM*

(**c**) *KSVM*

(**d**) *RF*

(**e**) *CCF*

(**f**) *All experiments*
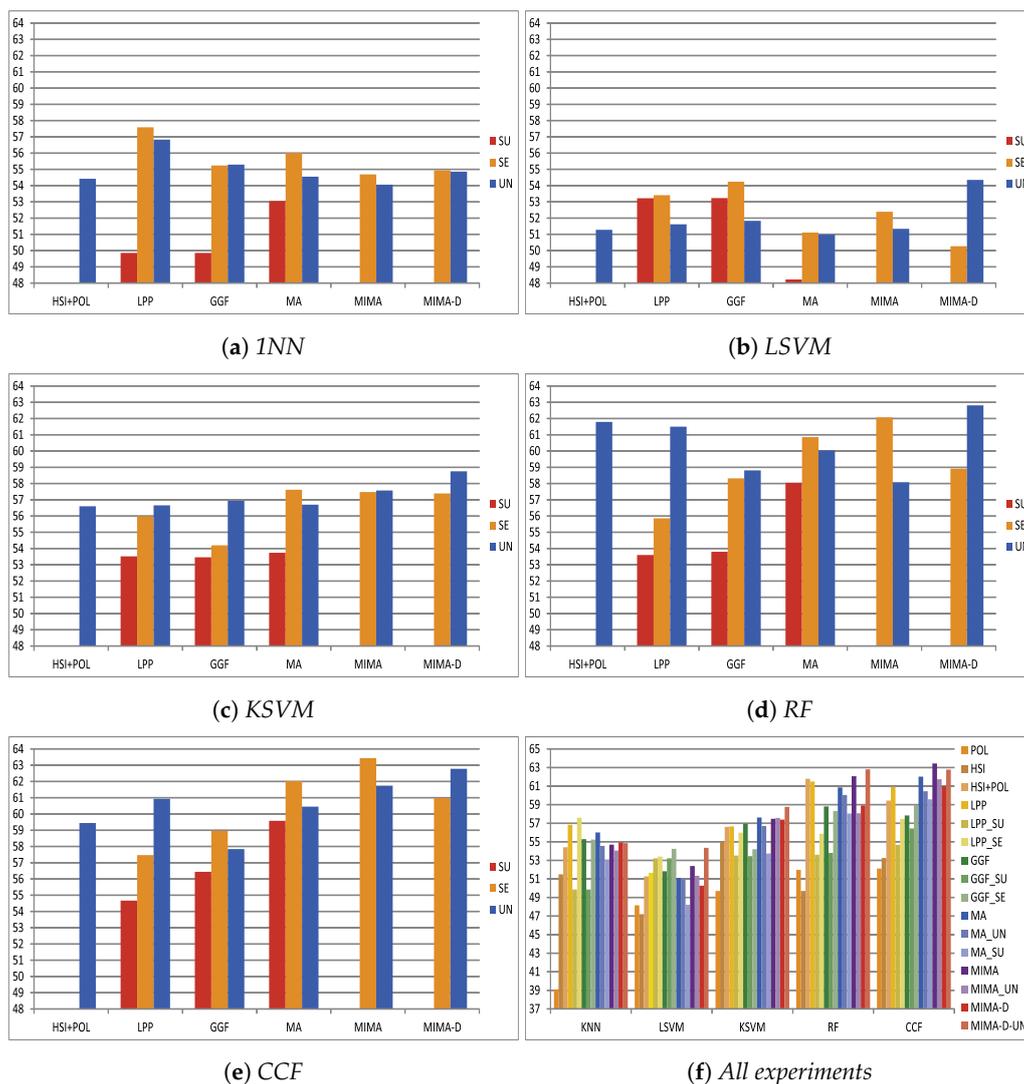
Figure 6. Comparison of the classification accuracies of different classifiers applying on the Augsburg data set. Each chart is resulted from a corresponding classifier. The right bottom chart demonstrates all the overall accuracy resulted by applying five classifiers on the fused data achieved from each of the 16 algorithms. The *y*-axis is the overall accuracy in percentage (%). The 'SU', 'SE', and 'UN', represent supervision, semi-supervision, and un-supervision, respectively.

Table 6. Quantitative performance comparison on the Augsburg data, in terms of class-specific accuracy, kappa coefficient, average accuracy, overall accuracy, and mean overall accuracy. The mean overall accuracy is calculated based on the overall accuracies achieved by the five classifiers. The listed indications are achieved after hyperparameter tunning. The hyperparameters of each algorithm are listed under the name of the algorithm and their values are listed in the table. The kappa coefficient, average accuracy, and the overall accuracy that larger than 0.56, 62.5%, and 62.5% are marked in bold. And the three highest mean overall accuracies are also marked in bold.

Algorithm	Parameter	Classifiers	Forest	Residential Area	Industrial Area	Low Plants	Allotment	Commercial Area	Water	KAPPA	AA	OA	Mean OA
POL	-	1NN	64	35.88	38.8	55.02	22.54	38.9	18.66	0.2897	39.11	39.11	
		LSVM	86.93	46.44	39.15	73.17	25.37	44.29	21.8	0.3952	48.16	48.16	
		KSVM	86.51	64.49	31.41	81.98	22.39	41.98	19.12	0.4131	49.7	49.7	48.21
		RF	81.88	63.44	47.76	88.46	28.88	38.71	14.63	0.4396	51.97	51.97	
		CCF	82.29	61.85	47.8	88.37	30.34	38.07	16.1	0.4414	52.12	52.12	
HSI	-	1NN	27.9	52.49	61.1	78.2	60.66	24.9	55.24	0.4341	51.5	51.5	
		LSVM	25.44	50.22	75.93	67.46	38.32	15.15	57.93	0.3841	47.21	47.21	
		KSVM	31.2	65.2	70.71	86.37	55.98	20.8	54.63	0.4748	54.98	54.98	51.33
		RF	25.59	58.29	70.29	84.34	40.41	15.98	52.98	0.4131	49.7	49.7	
		CCF	27.29	64.56	75.71	84.68	48.29	16.54	55.66	0.4546	53.25	53.25	
HSI+POL	-	1NN	34.76	58.17	55.93	84.56	57.73	34.9	54.88	0.4682	54.42	54.42	
		LSVM	31	65.95	73.29	83.85	36.9	25.07	42.85	0.4315	51.28	51.28	
		KSVM	40.59	67.83	67.07	92.59	45.24	27.1	55.78	0.4937	56.6	56.6	56.71
		RF	61.27	73.88	70.1	94.98	47.51	25.63	59.17	0.5542	61.79	61.79	
		CCF	46.07	75.63	78.05	95.51	58.07	18.49	44.22	0.5267	59.44	59.44	
LPP $\{k, dn\}$	{10, 40}	1NN	44.9	60.61	53.29	86.56	61.37	34.76	56.32	0.4963	56.83	56.83	
	{20, 20}	LSVM	28.17	64.93	76.63	81.54	38.27	17.88	53.93	0.4356	51.62	51.62	
	{40, 50}	KSVM	40.98	67.98	73.49	92.32	45.49	22.68	53.66	0.4943	56.66	56.66	57.42
	{10, 30}	RF	73.66	66.15	65.8	89.54	51.24	25.78	55.17	0.5456	61.05	61.05	
	{10, 35}	CCF	59.63	70.71	72.8	92.2	51.9	22.78	56.51	0.5442	60.93	60.93	
LPP_SU $\{dn\}$	{5}	1NN	31.93	55.83	56.95	78.51	49.07	33.98	42.76	0.415	49.86	49.86	
	{10}	LSVM	40.85	63.1	63.29	87.46	49.17	32.61	36.05	0.4542	53.22	53.22	
	{40}	KSVM	54.24	63.93	66.32	87.2	45.05	28.49	29.41	0.4577	53.52	53.52	52.97
	{35}	RF	44.46	60.93	62.78	90.07	44.15	30.95	41.88	0.4587	53.6	53.6	
	{35}	CCF	52.07	62.15	64.66	90.17	44.24	28.88	40.51	0.4711	54.67	54.67	
LPP_SE $\{k, dn\}$	{20, 45}	1NN	49.76	59.15	53	85.05	60.98	40.05	55.15	0.5052	57.59	57.59	
	{10, 10}	LSVM	43.49	65.51	77.22	85.07	40.76	20.8	41.05	0.4565	53.41	53.41	
	{120, 35}	KSVM	37.66	71.27	75.22	93.22	48.44	20.54	45.49	0.4864	55.98	55.98	56.06
	{30, 15}	RF	27.17	63.22	72.2	91.78	54.46	26.54	55.66	0.485	55.86	55.86	
	{80, 40}	CCF	47.2	66.46	73.22	90.93	56.07	23.27	45.17	0.5039	57.47	57.47	
GGF $\{k, dn\}$	{20, 50}	1NN	41.37	57.22	49.68	82.63	61.61	38.2	56.32	0.4784	55.29	55.29	
	{30, 15}	LSVM	29.17	63.76	74.83	82.12	36.54	19.71	56.71	0.438	51.83	51.83	
	{20, 15}	KSVM	34.51	69.22	73.71	92.34	45.32	23.9	59.61	0.4977	56.94	56.94	55.81
	{40, 45}	RF	60.22	65.61	61.29	89.73	46.46	31.78	56.56	0.5194	58.81	58.81	
	{40, 35}	CCF	47.9	70.9	72.22	92.44	43.34	23.05	55	0.5081	57.84	57.84	

Table 6. *Cont.*

Algorithm	Parameter	Classifiers	Forest	Residential Area	Industrial Area	Low Plants	Allotment	Commercial Area	Water	KAPPA	AA	OA	Mean OA
GGF_SU {dn}	{5}	1NN	31.93	55.83	56.95	78.51	49.07	33.98	42.76	0.415	49.86	49.86	
	{10}	LSVM	40.85	63.15	63.29	87.46	49.2	32.61	36.05	0.4543	53.23	53.23	
	{35}	KSVM	51.17	64.07	65.46	86.78	44.37	30.9	31.49	0.4571	53.46	53.46	53.36
	{45}	RF	44.93	61.05	60.15	89.93	42.88	32.68	45.02	0.4611	53.8	53.8	
	{45}	CCF	51.2	62.61	66.93	90.24	46.78	28.76	48.59	0.4918	56.44	56.44	
GGF_SE {k, dn}	{120, 10}	1NN	44.8	58.17	63.32	84.54	56.05	32.95	46.83	0.4778	55.24	55.24	
	{20, 30}	LSVM	53.02	66.54	66.95	84.61	47.27	29.41	31.88	0.4661	54.24	54.24	
	{10, 50}	KSVM	67.54	68.24	66.8	87.12	41.32	23.98	24.41	0.4657	54.2	54.2	56.19
	{90, 15}	RF	42.88	64.9	68.07	92.56	56.68	26.63	56.54	0.5138	58.32	58.32	
	{120, 40}	CCF	47	65.83	67.88	92.51	57.29	27.07	55.15	0.5212	58.96	58.96	
MA {μ, k, dn}	{2, 70, 35}	1NN	30.88	58.68	61.39	82.05	77.27	27.78	54.02	0.4868	56.01	56.01	
	{2.5, 60, 35}	LSVM	26.22	66.63	78.2	72.44	42.9	16.1	55.27	0.4296	51.11	51.11	
	{2, 70, 25}	KSVM	31.44	69.54	78.8	93	59.05	17.73	53.76	0.5055	57.62	57.62	**57.52**
	{1, 110, 45}	RF	75.34	72.15	64.66	91.61	48.88	30.12	43.24	0.5433	60.86	60.86	
	{1, 110, 45}	CCF	65.85	73.24	72.61	93.61	55	23.8	50.05	0.557	62.02	62.02	
MA_UN {k, dn}	{100, 25}	1NN	31.61	56.85	57.29	80.71	73.98	26.61	54.83	0.4698	54.55	54.55	
	{100, 30}	LSVM	26.51	67.12	76.8	73.78	41.07	15.71	55.83	0.428	50.98	50.98	
	{100, 20}	KSVM	32.56	68.2	74	89.29	58.63	18.88	55.32	0.4948	56.7	56.7	56.54
	{20, 25}	RF	75.15	67.93	63.17	87.93	44.29	31	50.83	0.5338	60.04	60.04	
	{20, 40}	CCF	75.27	69.07	60.95	89.83	50.07	32.41	45.56	0.5386	60.45	60.45	
MA_SU {dn}	{50}	1NN	26.71	52.78	61.15	80.22	69.93	26.07	54.46	0.4522	53.05	53.05	
	{50}	LSVM	25.2	57.2	77.56	70.29	36.85	16.68	53.76	0.3959	48.22	48.22	
	{50}	KSVM	28.68	60.68	74.83	87.9	56.2	17.46	50.39	0.4602	53.74	53.74	54.53
	{50}	RF	49.76	67.1	67.12	91.9	47.27	28.85	54.32	0.5105	58.05	58.05	
	{45}	CCF	64.07	69.12	66.78	92.41	52.39	27.63	44.63	0.5284	59.58	59.58	
MIMA {μ, b, dn}	{2.5, 35, 35}	1NN	27.68	57.07	62.56	81.39	72.17	26.46	55.51	0.4714	54.69	54.69	
	{3, 25, 5}	LSVM	23.61	71.93	78.63	79.98	44.29	13.76	54.51	0.4445	52.39	52.39	
	{1.5, 35, 15}	KSVM	34.15	68.12	72.9	92.27	53.51	22.07	59.34	0.5039	57.48	57.48	**58.01**
	{0.5, 40, 35}	RF	66.22	76.88	65.51	92.8	47.77	26.27	59.02	0.5575	62.07	62.07	
	{0.5, 55, 40}	CCF	76.78	77.49	65.12	92.73	50	28.78	53.15	**0.5734**	**63.44**	**63.44**	
MIMA_UN {b, dn}	{5, 35}	1NN	34.34	55.24	54.85	80.76	71.41	28.44	53.39	0.4641	54.06	54.06	
	{10, 50}	LSVM	28.39	67.78	76.05	74.17	40.73	19.22	53.02	0.4323	51.34	51.34	
	{5, 30}	KSVM	31.9	68.12	74.78	90.78	59.44	20.1	57.9	0.505	57.57	57.57	56.56
	{20, 30}	RF	58.95	66.54	71.76	89.68	51.68	25.24	42.68	0.5109	58.08	58.08	
	{15, 45}	CCF	83.71	67.61	68.07	89.73	57.98	27.68	37.37	0.5536	61.74	61.74	
MIMA-D {μ, b, dn}	{3, 50, 35}	1NN	28.76	57.63	62.68	80.22	74.83	24.98	55.49	0.4743	54.94	54.94	
	{3, 40, 15}	LSVM	25.27	67.44	78.46	73.29	39.85	15.41	52.17	0.4198	50.27	50.27	
	{3, 40, 15}	KSVM	33.12	68.95	70.41	92.8	55.8	20.05	60.59	0.5029	57.39	57.39	56.5
	{3, 35, 30}	RF	52.54	72.27	73.56	92.24	49.83	24.29	47.71	0.5207	58.92	58.92	
	{2.5, 30, 40}	CCF	55.51	73.66	72.98	92.85	53.61	24.05	54.22	0.5448	60.98	60.98	
MIMA-D_UN {b, dn}	{20, 30}	1NN	34.93	56.02	56.54	80.46	75.29	26.39	54.41	0.4734	54.86	54.86	
	{55, 5}	LSVM	87.22	55.24	48	57.41	36.2	41.95	54.41	0.4674	54.35	54.35	
	{15, 30}	KSVM	35.27	67.93	77.54	91.83	66.56	17.37	54.73	0.5187	58.75	58.75	**60.29**
	{20, 50}	RF	82.95	65.17	58.12	88.41	54.32	34.05	56.63	**0.5661**	**62.81**	**62.81**	
	{20, 30}	CCF	78.54	72.29	65.85	92.63	49.88	26.68	53.56	**0.5657**	**62.78**	**62.78**	
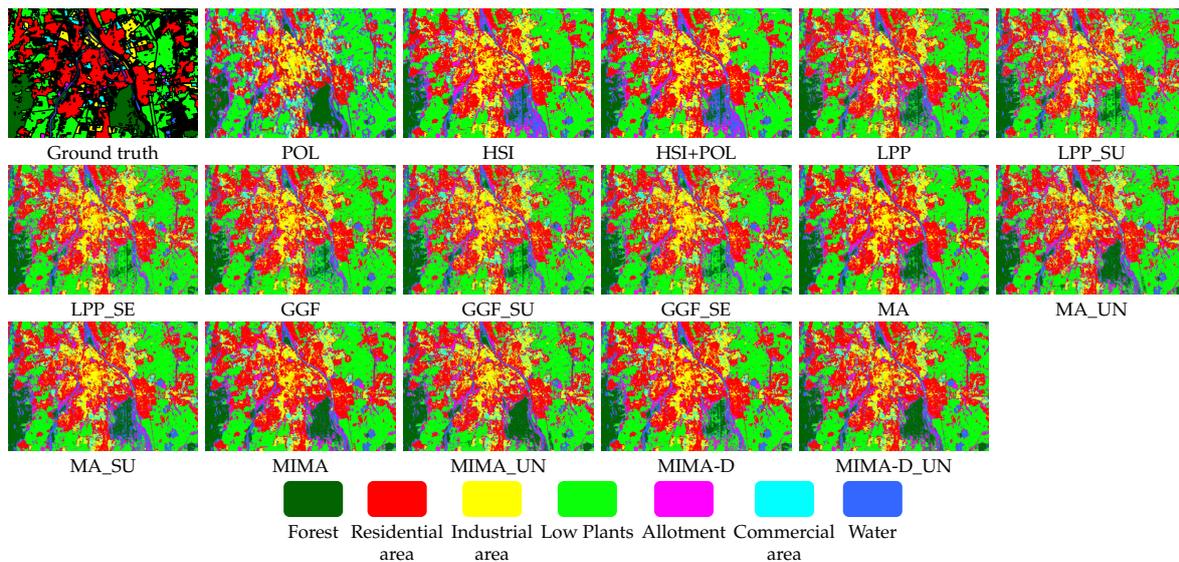
Figure 7. Visualization of the achieved classification maps and the ground truth. The 16 classification maps are provided by applying *CCF* on fused data of the 16 algorithms, for the Augsburg data set. Classification maps achieved by manifold alignment fusion methods are more accurate than the maps achieved by data alignment fusion methods, especially MIMA based methods.

4. Discussion

4.1. The Setting of the Training and Testing Samples

As shown in Figures 2 and 3, the training and testing samples are spatially separated as a standard machine learning practice. However, the distribution of training samples of the Berlin and the Augsburg data set are slightly different. For each class of the Berlin data set, the training data are block-wisely scattered over the whole area. For the Augsburg data set, the training data only covers on the western part of the area. There is no sample from the eastern half of the site where the testing data distribute. Both scenarios are common in remote sensing applications. The latter one is naturally more challenging. This is why the overall accuracy of the Augsburg data set fluctuates around 56%, while it is around 76% for the Berlin data set.

4.2. The Data Alignment Fusion

An unsupervised data alignment-based fusion in this article pursues an intrinsic manifold of a concatenation of the hyperspectral and PolSAR data. Intuitively, making use of the additional label information in the manifold learning (semi-supervision) should be improve the classification accuracy. However, we observed the exact opposite in our experiments. We believe it is due to the misalignment of image pixels of optical and SAR images caused by their distinct imaging geometry. This pixel misalignment leads to extra difficulty in learning a joint manifold. Adding one more manifold defined by the misaligned label will only lead to destructive effects. Therefore, the data alignment-based fusion algorithm is not competent for fusing hyperspectral and PolSAR data with the resolution similar to our dataset. This finding should also be able to generalized to high resolution optical and SAR data, although we have not conducted any experiment.

4.3. The Manifold Alignment Fusion

Different to the data alignment-based fusion, the semi-supervised manifold alignment-based fusion outperforms the unsupervised manifold alignment fusion. This fusion concept is able to introduce the advantage of label information while pursuing the intrinsic manifold. The reason is that

this fusion concept models the manifold of individual data sources independently which suits the fact that hyperspectral and PolSAR data are severely dissimilar in geometry and content. The label information is merged into the two manifolds in a way that the two manifolds are separately link to the label and are then aligned to each other by the label. In such manner, the advantage of label data appears on the classification results. Comparing to the data alignment-based fusion, the manifold alignment-based fusion introduces considerable improvements to the classification accuracy, which shows its competence for the fusion of hyperspectral and PolSAR data.

4.4. The Filter Function of MIMA

As introduced in Section 2.5, the filter function of MIMA introduces an expertise knowledge while deriving the manifold structure of data. MIMA and MIMA-D in this paper employed PCA and a density estimation as the filter function, respectively. The principal components are frequently used in classification. It is proven to be effective [82,96]. The density function is an important property for classification or clustering tasks [97,98]. However, from the experiment in this paper, it is inconclusive which choice is more suitable to serve as the filter function.

5. Conclusions and Outlook

This paper compares 16 variants of four state-of-the-art multi-sensory data fusion algorithms based on manifold learning. The comparison was done via a rigorous evaluation of the performance of the 16 algorithms on land use land cover classification on two sets of spaceborne hyperspectral images and PolSAR data. To carry out an objective comparison, the hyperparameters of the 16 algorithms were optimized via a grid search. Five different shallow classifiers were applied on the data sets fused by the 16 algorithms. We avoided any deep network classifiers, because the goal of this article is to solely evaluate the performance of multi-sensory data fusion algorithms. The experiments conclude that (1) data alignment-based (data concatenation) manifold techniques are less competent for the fusion of hyperspectral images and PolSAR data, or in general optical and SAR images fusion, because a concatenation of the two data sets with distinct imaging geometries causes difficulty even destructive effects when optimizing the target manifold; On the contrary, manifold alignment-based techniques are more competent for the task of optical and SAR images fusion, because the manifolds of the two data are separately modeled and aligned; (2) Among the manifold alignment-based manifold techniques, semi-supervised methods are able to effectively make use of both the structure of data and the existing label information; (3) the MIMA algorithm cooperating with the *CCF* classifier provides the best classification accuracy among all the algorithms.

Based on our current research, our future research directions can include:

- In the current algorithms, the learned manifold is specific to the very input data sets. We would like to study the generalization of such manifold on data sets of the same sensors. Eventually, we aim at big data processing where one common manifold can be applied to all the data sets of the same type.
- Graph CNN has been an emerging filed in deep lerning. It is also of great interest to combine it with the traditional manifold learning techniques described in this article.
- Because of the data availability of spaceborne hyperspectral and PolSAR data, they have not been extensively applied to real world problems. We would like to address more real world applications especially those for social good using those two types of data, for example, contributing to the monitoring of Unite Nation's sustainable development goals.

Author Contributions: Conceptualization, J.H., D.H., and X.X.Z.; methodology, J.H., and D.H.; software, J.H.; validation, Y.W.; formal analysis, J.H.; investigation, J.H.; resources, J.H.; data curation, J.H.; writing—original draft preparation, J.H.; writing—review and editing, D.H. and Y.W.; visualization, J.H.; supervision, X.X.Z.; project administration, J.H.; funding acquisition, X.X.Z.

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

Appendix A. Pseudo-Code of LPP

Algorithm A1: LPP($\mathbf{X}$,k,σ)

Input:

    $\mathbf{X}$: the data source $\mathbf{X} = [\mathbf{x}^1, ..., \mathbf{x}^p, ..., \mathbf{x}^n] \in \mathbb{R}^{m \times n}$ with n instances and m dimensions

    k: the number of local neighbors

    σ: the filtering parameter

Output:

    $\mathbf{Y}$: the representation of data $\mathbf{X}$ on the intrinsic manifold $\mathcal{M}$.

    $\mathbf{f}$: the projection maps data $\mathbf{X}$ to $\mathbf{Y}$

1 construct the n by n weight matrix with Equation (1)

2 construct the degree matrix $\mathbf{D}$

3 construct the Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{W}$

4 solve the generalized eigenvalue decomposition $\mathbf{XLX}^T\mathbf{f} = \lambda\mathbf{XDX}^T\mathbf{f}$

5 construct $\mathbf{Y}$: $\mathbf{Y} = \mathbf{f}^T\mathbf{X}$

6 **Return $\mathbf{Y}$ and $\mathbf{f}$**

Appendix B. Pseudo-Code of GGF

Algorithm A2: GGF($\mathbf{X}_1$,$\mathbf{X}_2$,k,σ)

Input:

    $\mathbf{X}_1$: the data source $\mathbf{X}_1 \in \mathbb{R}^{m_1 \times n}$ with n instances and m_1 dimensions

    $\mathbf{X}_2$: the data source $\mathbf{X}_2 \in \mathbb{R}^{m_2 \times n}$ with n instances and m_2 dimensions

    k: the number of local neighbors

    σ: the filtering parameter

Output:

    $\tilde{\mathbf{Y}}$: the fused data.

    $\mathbf{f}$: the projection maps data $\tilde{\mathbf{X}}$ to $\tilde{\mathbf{Y}}$

1 stacking data sources on the feature dimension: $\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} = [\tilde{\mathbf{x}}^1, ..., \tilde{\mathbf{x}}^p, ..., \tilde{\mathbf{x}}^n] \in \mathbb{R}^{(m_1+m_2) \times n}$

2 construct binary matrices $\mathbf{A}_i (i \in \{1, 2\})$ to model manifolds of $\mathbf{X}_i$:

    $\mathbf{A}_i(p, q) = \begin{cases} 1 & x_i^p \text{ is one of the } k \text{ nearest neighbor to } x_i^q \\ 0 & \text{otherwise} \end{cases}$

3 construct a fused binary matrix $\tilde{\mathbf{A}}(p, q) = \mathbf{A}_1(p, q) * \mathbf{A}_2(p, q)$

4 calculate a n by n pairwise distance matrix $\tilde{\mathcal{D}}$

5 construct a GGF pairwise distance matrix $\tilde{\mathcal{D}}_{GGF}$ as Equation (6)

6 calculate the n by n weight matrix: $\tilde{\mathbf{W}}$ as Equation (7)

7 calculate the degree matrix $\tilde{\mathbf{D}}$

8 calculate the Laplacian matrix $\tilde{\mathbf{L}} = \tilde{\mathbf{D}} - \tilde{\mathbf{W}}$

9 solve the generalized eigenvalue decomposition $\tilde{\mathbf{X}}\tilde{\mathbf{L}}\tilde{\mathbf{X}}^T\mathbf{f} = \lambda\tilde{\mathbf{X}}\tilde{\mathbf{D}}\tilde{\mathbf{X}}^T\mathbf{f}$

10 calculate $\tilde{\mathbf{Y}} = \mathbf{f}^T\tilde{\mathbf{X}}$

11 **Return $\tilde{\mathbf{Y}}$ and $\mathbf{f}$**

Appendix C. Pseudo-Code of MA

Algorithm A3: MA($\mathbf{X}_1$,$\mathbf{X}_2$,$\mathbf{E}_1$,$\mathbf{E}_2$,k)

Input:

$\mathbf{X}_1$: the data source $\mathbf{X}_1 \in \mathbb{R}^{m_1 \times n_1}$ with n_1 instances and m_1 dimensions

$\mathbf{X}_2$: the data source $\mathbf{X}_2 \in \mathbb{R}^{m_2 \times n_2}$ with n_2 instances and m_2 dimensions

$\mathbf{E}_1$: $\mathbf{E}_1 \in \mathbb{R}^{1 \times n_1^*}$ with $n_1^* < n_1$, labels for the first n_1^* instances of $\mathbf{X}_1$

$\mathbf{E}_2$: $\mathbf{E}_2 \in \mathbb{R}^{1 \times n_2^*}$ with $n_2^* < n_2$, labels for the first n_2^* instances of $\mathbf{X}_2$

k: the number of local neighbors

Output:

$\tilde{\mathbf{Y}}_1$: the projected data of $\mathbf{X}_1$.

$\tilde{\mathbf{Y}}_2$: the projected data of $\mathbf{X}_2$.

$\mathbf{f}_1$: the projection maps data $\mathbf{X}_1$ to $\tilde{\mathbf{Y}}_1$

$\mathbf{f}_2$: the projection maps data $\mathbf{X}_2$ to $\tilde{\mathbf{Y}}_2$

1 construct $(n_1 + n_2)$ by $(n_1 + n_2)$ binary matrices $\tilde{\mathbf{A}}_s$ (Equation (9)) and $\tilde{\mathbf{A}}_d$ (Equation (10)) using $\mathbf{E}_1$ and $\mathbf{E}_2$

2 construct $(n_1 + n_2)$ by $(n_1 + n_2)$ binary matrix $\tilde{\mathbf{A}}_g$ (Equation (11)) using k-nearest-neighbor with the given k

3 construct degree matrices $\tilde{\mathbf{D}}_s$, $\tilde{\mathbf{D}}_d$, and $\tilde{\mathbf{D}}_g$ with $\tilde{\mathbf{A}}_s$, $\tilde{\mathbf{A}}_d$, and $\tilde{\mathbf{A}}_g$, respectively

4 construct Laplacian matrices $\tilde{\mathbf{L}}_s$, $\tilde{\mathbf{L}}_d$, and $\tilde{\mathbf{L}}_g$ as instructed in Equation (17)

5 organize the data matrix $\tilde{\mathbf{X}}$ as instructed in Equation (17)

6 solve the generalized eigenvalue decomposition $\tilde{\mathbf{X}}(\mu\tilde{\mathbf{L}}_g + \tilde{\mathbf{L}}_s)\tilde{\mathbf{X}}^T\mathbf{f} = \lambda\tilde{\mathbf{X}}\tilde{\mathbf{L}}_d\tilde{\mathbf{X}}^T\mathbf{f}$ so that $\mathbf{f}_1$ and $\mathbf{f}_2$ are achieved, $\mathbf{f} = \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \end{bmatrix}$.

7 calculate $\tilde{\mathbf{Y}}_1 = \mathbf{f}_1^T\mathbf{X}_1$ and $\tilde{\mathbf{Y}}_2 = \mathbf{f}_2^T\mathbf{X}_2$

8 **Return** $\tilde{\mathbf{Y}}_1$, $\tilde{\mathbf{Y}}_2$, $\mathbf{f}_1$, $\mathbf{f}_2$

Appendix D. Pseudo-Code of MIMA

Algorithm A4: MIMA-MAPPER($\mathbf{X}$,b,c,$\mathbf{F}$)

Input:

$\mathbf{X}$: the data source $\mathbf{X} \in \mathbb{R}^{m \times n}$ with n instances and m dimensions

b: the number of data bins

c: the overlapping rate

$\mathbf{F}$: the filtering function

Output:

$\mathbf{A}_{MIMA}$: the connection matrix

1 calculate the parameter space $\mathbf{XF}$

2 divide $\mathbf{XF}$ into $\mathbf{b}$ intervals with $\mathbf{c}$% overlap of adjacent intervals

3 divide data $\mathbf{X}$ into $\mathbf{b}$ bins corresponding to intervals achieved in 2

4 **for** (each data bin):

5   Spectral clustering

6 **end for**

7 Construct topological matrix $\mathbf{A}_{MIMA}(p,q) = \begin{cases} 1, & \text{if } p \text{ and } q \text{ in the same cluster;} \\ 1, & \text{if } p \text{ and } q \text{ in the linked clusters;} \\ 0, & \text{otherwise.} \end{cases}$

8 **Return** $\mathbf{A}_{MIMA}$

Algorithm A5: MIMA($\mathbf{X}_1$,$\mathbf{X}_2$,$\mathbf{E}_1$,$\mathbf{E}_2$,k)

Input:

  $\mathbf{X}_1$: the data source $\mathbf{X}_1 \in \mathbb{R}^{m_1 \times n_1}$ with n_1 instances and m_1 dimensions

  $\mathbf{X}_2$: the data source $\mathbf{X}_2 \in \mathbb{R}^{m_2 \times n_2}$ with n_2 instances and m_2 dimensions

  $\mathbf{E}_1$: $\mathbf{E}_1 \in \mathbb{R}^{1 \times n_1^*}$ with $n_1^* < n_1$, labels for the first n_1^* instances of $\mathbf{X}_1$

  $\mathbf{E}_2$: $\mathbf{E}_2 \in \mathbb{R}^{1 \times n_2^*}$ with $n_2^* < n_2$, labels for the first n_2^* instances of $\mathbf{X}_2$

  k: the number of local neighbors

Output:

  $\tilde{\mathbf{Y}}_1$: the projected data of $\mathbf{X}_1$.

  $\tilde{\mathbf{Y}}_2$: the projected data of $\mathbf{X}_2$.

  $\mathbf{f}_1$: the projection maps data $\mathbf{X}_1$ to $\tilde{\mathbf{Y}}_1$

  $\mathbf{f}_2$: the projection maps data $\mathbf{X}_2$ to $\tilde{\mathbf{Y}}_2$

1 construct $(n_1 + n_2)$ by $(n_1 + n_2)$ binary matrices $\tilde{\mathbf{A}}_s$ (Equation (9)) and $\tilde{\mathbf{A}}_d$ (Equation (10)) using $\mathbf{E}_1$ and $\mathbf{E}_2$

2 **for**(i=1:2)

3   $\mathbf{A}_{MIMA}^{i,i} = \textbf{MIMA-MAPPER}(\mathbf{X}_i,b,c)$

4 **end**

5 construct matrix $\tilde{\mathbf{A}}_{MIMA} = \begin{pmatrix} \mathbf{A}_{MIMA}^{1,1} & 0 & 0 \\ 0 & ... & 0 \\ 0 & 0 & \mathbf{A}_{MIMA}^{K,K} \end{pmatrix}$

6 construct degree matrices $\tilde{\mathbf{D}}_s$, $\tilde{\mathbf{D}}_d$, and $\tilde{\mathbf{D}}_{MIMA}$ with $\tilde{\mathbf{A}}_s$, $\tilde{\mathbf{A}}_d$, and $\tilde{\mathbf{A}}_{MIMA}$, respectively

7 construct Laplacian matrices $\tilde{\mathbf{L}}_s$, $\tilde{\mathbf{L}}_d$, and $\tilde{\mathbf{L}}_{MIMA}$ as instructed in Equation (17)

8 organize the data matrix $\tilde{\mathbf{X}}$ as instructed in Equation (17)

9 solve the generalized eigenvalue decomposition $\tilde{\mathbf{X}}(\mu\tilde{\mathbf{L}}_g + \tilde{\mathbf{L}}_s)\tilde{\mathbf{X}}^T\mathbf{f} = \lambda\tilde{\mathbf{X}}\tilde{\mathbf{L}}_{MIMA}\tilde{\mathbf{X}}^T\mathbf{f}$ so that $\mathbf{f}_1$ and $\mathbf{f}_2$ are achieved, $\mathbf{f} = \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \end{bmatrix}$

10 calculate $\tilde{\mathbf{Y}}_1 = \mathbf{f}_1^T\mathbf{X}_1$ and $\tilde{\mathbf{Y}}_2 = \mathbf{f}_2^T\mathbf{X}_2$

11 **Return** $\tilde{\mathbf{Y}}_1$, $\tilde{\mathbf{Y}}_2$, $\mathbf{f}_1$, $\mathbf{f}_2$

References

1. Zhang, J. Multi-source remote sensing data fusion: status and trends. *Int. J. Image Data Fusion* **2010**, *1*, 5–24. [CrossRef]

2. Dalla Mura, M.; Prasad, S.; Pacifici, F.; Gamba, P.; Chanussot, J.; Benediktsson, J.A. Challenges and opportunities of multimodality and data fusion in remote sensing. *Proc. IEEE* **2015**, *103*, 1585–1601. [CrossRef]

3. Yokoya, N.; Grohnfeldt, C.; Chanussot, J. Hyperspectral and multispectral data fusion: A comparative review of the recent literature. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 29–56. [CrossRef]

4. Hong, D.; Yokoya, N.; Chanussot, J.; Zhu, X.X. CoSpace: Common Subspace Learning from Hyperspectral-Multispectral Correspondences. *arXiv* **2018**, arXiv:1812.11501.

5. Dalponte, M.; Bruzzone, L.; Gianelle, D. Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1416–1427. [CrossRef]

6. Swatantran, A.; Dubayah, R.; Roberts, D.; Hofton, M.; Blair, J.B. Mapping biomass and stress in the Sierra Nevada using lidar and hyperspectral data fusion. *Remote Sens. Environ.* **2011**, *115*, 2917–2930. [CrossRef]

7. Khodadadzadeh, M.; Li, J.; Prasad, S.; Plaza, A. Fusion of hyperspectral and LiDAR remote sensing data using multiple feature learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2971–2983. [CrossRef]

8. Merkle, N.; Auer, S.; Müller, R.; Reinartz, P. Exploring the Potential of Conditional Adversarial Networks for Optical and SAR Image Matching. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1811–1820. [CrossRef]

9. Joshi, N.; Baumann, M.; Ehammer, A.; Fensholt, R.; Grogan, K.; Hostert, P.; Jepsen, M.R.; Kuemmerle, T.; Meyfroidt, P.; Mitchard, E.T.; et al. A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. *Remote Sens.* **2016**, *8*, 70. [CrossRef]

10. Wang, Y.; Zhu, X.X.; Zeisl, B.; Pollefeys, M. Fusing Meter-Resolution 4-D InSAR Point Clouds and Optical Images for Semantic Urban Infrastructure Monitoring. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 14–26. [CrossRef]

11. Schmitt, M.; Hughes, L.H.; Zhu, X.X. THE SEN1-2 Dataset for DEEP LEARNING IN SAR-OPTICAL DATA FUSION. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *IV-1*, 141–146. [CrossRef]

12. Hong, D.; Yokoya, N.; Ge, N.; Chanussot, J.; Zhu, X.X. Learnable manifold alignment (LeMA): A semi-supervised cross-modality learning framework for land cover and land use classification. *ISPRS J. Photogramm. Remote Sens.* **2019**, *147*, 193–205. [CrossRef]

13. Chang, Y.L.; Han, C.C.; Ren, H.; Chen, C.T.; Chen, K.S.; Fan, K.C. Data fusion of hyperspectral and SAR images. *Opt. Eng.* **2004**, *43*, 1787–1798.

14. Koch, B. Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 581–590. [CrossRef]

15. Torres, R.; Snoeij, P.; Geudtner, D.; Bibby, D.; Davidson, M.; Attema, E.; Potin, P.; Rommen, B.; Floury, N.; Brown, M.; et al. GMES Sentinel-1 mission. *Remote Sens. Environ.* **2012**, *120*, 9–24. [CrossRef]

16. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [CrossRef]

17. Stuffler, T.; Kaufmann, C.; Hofer, S.; Förster, K.; Schreier, G.; Mueller, A.; Eckardt, A.; Bach, H.; Penne, B.; Benz, U.; et al. The EnMAP hyperspectral imager—An advanced optical payload for future applications in Earth observation programmes. *Acta Astronaut.* **2007**, *61*, 115–120. [CrossRef]

18. Wu, X.; Hong, D.; Ghamisi, P.; Li, W.; Tao, R. MsRi-CCF: Multi-scale and rotation-insensitive convolutional channel features for geospatial object detection. *Remote Sens.* **2018**, *10*, 1990. [CrossRef]

19. Wu, X.; Hong, D.; Tian, J.; Chanussot, J.; Li, W.; Tao, R. ORSIm Detector: A Novel Object Detection Framework in Optical Remote Sensing Imagery Using Spatial-Frequency Channel Features. *arXiv* **2019**, arXiv:1901.07925.

20. Hong, D.; Zhu, X.X. SULoRA: Subspace unmixing with low-rank attribute embedding for hyperspectral data analysis. *IEEE J. Sel. Top. Signal Process.* **2018**, *12*, 1351–1363. [CrossRef]

21. Drumetz, L.; Veganzones, M.A.; Henrot, S.; Phlypo, R.; Chanussot, J.; Jutten, C. Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability. *IEEE Trans. Image Process.* **2016**, *25*, 3890–3905. [CrossRef] [PubMed]

22. Hong, D.; Yokoya, N.; Chanussot, J.; Zhu, X.X. Learning a low-coherence dictionary to address spectral variability for hyperspectral unmixing. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 235–239.

23. Ceamanos, X.; Waske, B.; Benediktsson, J.A.; Chanussot, J.; Fauvel, M.; Sveinsson, J.R. A classifier ensemble based on fusion of support vector machines for classifying hyperspectral data. *Int. J. Image Data Fusion* **2010**, *1*, 293–307. [CrossRef]

24. Hong, D.; Yokoya, N.; Chanussot, J.; Zhu, X.X. An augmented linear mixing model to address spectral variability for hyperspectral unmixing. *IEEE Trans. Image Process.* **2019**, *28*, 1923–1938. [CrossRef] [PubMed]

25. Lee, J.S.; Pottier, E. *Polarimetric Radar Imaging: From Basics to Applications*; CRC Press: Boca Raton, FL, USA, 2009.

26. Cloude, S.R.; Pottier, E. An entropy based classification scheme for land applications of polarimetric SAR. *IEEE Trans. Geosci. Remote Sens.* **1997**, *35*, 68–78. [CrossRef]

27. Moreira, A.; Prats-Iraola, P.; Younis, M.; Krieger, G.; Hajnsek, I.; Papathanassiou, K.P. A tutorial on synthetic aperture radar. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–43. [CrossRef]

28. Schmitt, A.; Wendleder, A.; Hinz, S. The Kennaugh element framework for multi-scale, multi-polarized, multi-temporal and multi-frequency SAR image preparation. *ISPRS J. Photogramm. Remote Sens.* **2015**, *102*, 122–139. [CrossRef]

29. Hu, J.; Ghamisi, P.; Zhu, X. Feature Extraction and Selection of Sentinel-1 Dual-Pol Data for Global-Scale Local Climate Zone Classification. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 379. [CrossRef]

30. Jouan, A.; Allard, Y. Land use mapping with evidential fusion of features extracted from polarimetric synthetic aperture radar and hyperspectral imagery. *Inf. Fusion* **2004**, *5*, 251–267. [CrossRef]

31. Li, T.; Zhang, J.; Zhao, H.; Shi, C. Classification-oriented hyperspectral and PolSAR images synergic processing. In Proceedings of the 2013 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Melbourne, Australia, 21–26 July 2013; pp. 1035–1038.

32. Dabbiru, L.; Samiappan, S.; Nobrega, R.A.A.; Aanstoos, J.V.; Younan, N.H.; Moorhead, R.J. Fusion of synthetic aperture radar and hyperspectral imagery to detect impacts of oil spill in Gulf of Mexico. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015.

33. Hu, J.; Ghamisi, P.; Schmitt, A.; Zhu, X. Object Based Fusion of Polarimetric SAR and Hyperspectral Imaging for Land Use Classification. In Proceedings of the 2016 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Los Angeles, CA, USA, 21–24 August 2016.

34. Hu, J.; Mou, L.; Schmitt, A.; Zhu, X.X. FusioNet: A two-stream convolutional neural network for urban scene classification using PolSAR and hyperspectral data. In Proceedings of the 2017 Joint Urban Remote Sensing Event (JURSE), Dubai, UAE, 6–8 March 2017; pp. 1–4.

35. Wang, C.; Mahadevan, S. A General Framework for Manifold Alignment. In Proceedings of the AAAI Fall Symposium: Manifold Learning and Its Applications, Washington, DC, USA, 14–18 July 2009; pp. 53–58.

36. Tuia, D.; Volpi, M.; Trolliet, M.; Camps-Valls, G. Semisupervised manifold alignment of multimodal remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 7708–7720. [CrossRef]

37. Ghamisi, P.; Benediktsson, J.A.; Phinn, S. Land-cover classification using both hyperspectral and LiDAR data. *Int. J. Image Data Fusion* **2015**, *6*, 189–215. [CrossRef]

38. Xia, J.; Yokoya, N.; Iwasaki, A. Hyperspectral image classification with canonical correlation forests. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 421–431. [CrossRef]

39. Yokoya, N.; Ghamisi, P.; Xia, J.; Sukhanov, S.; Heremans, R.; Tankoyeu, I.; Bechtel, B.; Le Saux, B.; Moser, G.; Tuia, D. Open data for global multimodal land use classification: Outcome of the 2017 IEEE GRSS Data Fusion Contest. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1363–1377. [CrossRef]

40. Makantasis, K.; Doulamis, A.; Doulamis, N.; Nikitakis, A.; Voulodimos, A. Tensor-Based Nonlinear Classifier for High-Order Data Analysis. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; Volume 2, pp. 2221–2225.

41. Makantasis, K.; Doulamis, A.D.; Doulamis, N.D.; Nikitakis, A. Tensor-based classification models for hyperspectral data analysis. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6884–6898. [CrossRef]

42. Yokoya, N.; Chanussot, J.; Iwasaki, A. Nonlinear unmixing of hyperspectral data using semi-nonnegative matrix factorization. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 1430–1437. [CrossRef]

43. Wang, C.; Mahadevan, S. Manifold Alignment without Correspondence. In Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI), Pasadena, CA, USA, 11–17 July 2009; Volume 2, p. 3.

44. Wang, C.; Mahadevan, S. Heterogeneous domain adaptation using manifold alignment. *IJCAI Proc. Int. Jt. Conf. Artif. Intell.* **2011**, *22*, 1541.

45. Wang, C.; Mahadevan, S. Manifold Alignment Preserving Global Geometry. In Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI), Beijing, China, 3–9 August 2013; pp. 1743–1749.

46. Tuia, D.; Camps-Valls, G. Kernel manifold alignment for domain adaptation. *PLoS ONE* **2016**, *11*, e0148655. [CrossRef]

47. Tuia, D.; Munoz-Mari, J.; Gómez-Chova, L.; Malo, J. Graph matching for adaptation in remote sensing. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 329–341. [CrossRef]

48. Liao, W.; Pižurica, A.; Bellens, R.; Gautama, S.; Philips, W. Generalized graph-based fusion of hyperspectral and LiDAR data using morphological features. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 552–556. [CrossRef]

49. Volpi, M.; Camps-Valls, G.; Tuia, D. Spectral alignment of multi-temporal cross-sensor images with automated kernel canonical correlation analysis. *ISPRS J. Photogramm. Remote Sens.* **2015**, *107*, 50–63. [CrossRef]

50. Liao, D.; Qian, Y.; Zhou, J.; Tang, Y.Y. A manifold alignment approach for hyperspectral image visualization with natural color. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3151–3162. [CrossRef]

51. Hong, D.; Yokoya, N.; Zhu, X.X. Learning a Robust Local Manifold Representation for Hyperspectral Dimensionality Reduction. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 2960–2975. [CrossRef]

52. He, X.; Niyogi, P. Locality preserving projections. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2004; pp. 153–160.

53. Hu, J.; Hong, D.; Zhu, X.X. MIMA: MAPPER-Induced Manifold Alignment for Semi-Supervised Fusion of Optical Image and Polarimetric SAR Data. *IEEE Trans. Geosci. Remote Sens.* **2019**, under review.

54. Roweis, S.T.; Saul, L.K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **2000**, *290*, 2323–2326. [CrossRef]

55. Tenenbaum, J.B.; De Silva, V.; Langford, J.C. A global geometric framework for nonlinear dimensionality reduction. *Science* **2000**, *290*, 2319–2323. [CrossRef]

56. Hatcher, A. *Algebraic Topology*; Tsinghua University Press: Beijing, China, 2005.

57. Lin, T.; Zha, H. Riemannian manifold learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 796–809.

58. Friedman, J.H.; Bentley, J.L.; Finkel, R.A. An algorithm for finding best matches in logarithmic time. *ACM Trans. Math. Softw.* **1976**, *3*, 209–226. [CrossRef]

59. Cristianini, N.; Shawe-Taylor, J. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*; Cambridge University Press: Cambridge, UK, 2000.

60. Schölkopf, B.; Smola, A.J.; Bach, F. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*; MIT Press: Cambridge, MA, USA, 2002.

61. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

62. Rainforth, T.; Wood, F. Canonical correlation forests. *arXiv* **2015**, arXiv:1507.05444.

63. Hong, D.; Yokoya, N.; Zhu, X.X. The K-LLE algorithm for nonlinear dimensionality ruduction of large-scale hyperspectral data. In Proceedings of the 2016 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Los Angeles, CA, USA, 21–24 August 2016; pp. 1–5.

64. Hong, D.; Yokoya, N.; Zhu, X.X. Local manifold learning with robust neighbors selection for hyperspectral dimensionality reduction. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 40–43.

65. Belkin, M.; Niyogi, P. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2002; pp. 585–591.

66. Hong, D.; Pan, Z.; Wu, X. Improved differential box counting with multi-scale and multi-direction: A new palmprint recognition method. *Opt.-Int. J. Light Electron Opt.* **2014**, *125*, 4154–4160. [CrossRef]

67. He, X.; Cai, D.; Yan, S.; Zhang, H.J. Neighborhood preserving embedding. In Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05), Beijing, China, 17–21 October 2005; Volume 2, pp. 1208–1213.

68. Hong, D.; Liu, W.; Su, J.; Pan, Z.; Wang, G. A novel hierarchical approach for multispectral palmprint recognition. *Neurocomputing* **2015**, *151*, 511–521. [CrossRef]

69. Hong, D.; Liu, W.; Wu, X.; Pan, Z.; Su, J. Robust palmprint recognition based on the fast variation Vese–Osher model. *Neurocomputing* **2016**, *174*, 999–1012. [CrossRef]

70. Donoho, D.L. High-dimensional data analysis: The curses and blessings of dimensionality. *AMS Math Chall. Lect.* **2000**, *1*, 32.

71. Farrell, M.D.; Mersereau, R.M. On the impact of PCA dimension reduction for hyperspectral detection of difficult targets. *IEEE Geosci. Remote Sens. Lett.* **2005**, *2*, 192–195. [CrossRef]

72. Debes, C.; Merentitis, A.; Heremans, R.; Hahn, J.; Frangiadakis, N.; van Kasteren, T.; Liao, W.; Bellens, R.; Pižurica, A.; Gautama, S.; et al. Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS data fusion contest. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2405–2418. [CrossRef]

73. Chintakunta, H.; Robinson, M.; Krim, H. Introduction to the special session on Topological Data Analysis, ICASSP 2016. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 6410–6414. [CrossRef]

74. Chazal, F.; Michel, B. An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists. *arXiv* **2017**, arXiv:1710.04019.

75. Zomorodian, A.; Carlsson, G. Computing persistent homology. *Discret. Comput. Geom.* **2005**, *33*, 249–274. [CrossRef]

76. Edelsbrunner, H.; Letscher, D.; Zomorodian, A. Topological persistence and simplification. In Proceedings of the 41st Annual Symposium on Foundations of Computer Science, Redondo Beach, CA, USA, 12–14 November 2000; pp. 454–463.

77. Edelsbrunner, H. *A Short Course in Computational Geometry and Topology*; Springer: Cham, Switzerland, 2014.

78. Singh, G.; Mémoli, F.; Carlsson, G.E. Topological methods for the analysis of high dimensional data sets and 3D object recognition. In Proceedings of the Eurographics Symposium on Point-Based Graphics (SPBG), San Diego, CA, USA, 25 May 2007; pp. 91–100.

79. Nicolau, M.; Levine, A.J.; Carlsson, G. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 7265–7270. [CrossRef]

80. Nielson, J.L.; Paquette, J.; Liu, A.W.; Guandique, C.F.; Tovar, C.A.; Inoue, T.; Irvine, K.A.; Gensel, J.C.; Kloke, J.; Petrossian, T.C.; et al. Topological data analysis for discovery in preclinical spinal cord injury and traumatic brain injury. *Nat. Commun.* **2015**, *6*, 8581. [CrossRef]

81. Li, L.; Cheng, W.Y.; Glicksberg, B.S.; Gottesman, O.; Tamler, R.; Chen, R.; Bottinger, E.P.; Dudley, J.T. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Sci. Transl. Med.* **2015**, *7*, 311ra174. [CrossRef]

82. Lum, P.Y.; Singh, G.; Lehman, A.; Ishkanov, T.; Vejdemo-Johansson, M.; Alagappan, M.; Carlsson, J.; Carlsson, G. Extracting insights from the shape of complex data using topology. *Sci. Rep.* **2013**, *3*, srep01236. [CrossRef]

83. Carriere, M.; Michel, B.; Oudot, S. Statistical analysis and parameter selection for Mapper. *J. Mach. Learn. Res.* **2018**, *19*, 478–516.

84. Ng, A.Y.; Jordan, M.I.; Weiss, Y. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2002; pp. 849–856.

85. Okujeni, A.; Van Der Linden, S.; Hostert, P. *Berlin-Urban-Gradient dataset 2009—An EnMAP Preparatory Flight Campaign (Datasets)*; GFZ Data Services: Potsdam, Germany, 2016.

86. Hu, J.; Guo, R.; Zhu, X.; Baier, G.; Wang, Y. Non-local means filter for polarimetric SAR speckle reduction-experiments using TerraSAR-x data. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *2*, 71. [CrossRef]

87. Haklay, M.; Weber, P. Openstreetmap: User-generated street maps. *IEEE Pervasive Comput.* **2008**, *7*, 12–18. [CrossRef]

88. Benediktsson, J.A.; Palmason, J.A.; Sveinsson, J.R. Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 480–491. [CrossRef]

89. Ghamisi, P.; Benediktsson, J.A.; Ulfarsson, M.O. Spectral—Spatial classification of hyperspectral images based on hidden Markov random fields. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2565–2574. [CrossRef]

90. Liao, W.; Chanussot, J.; Dalla Mura, M.; Huang, X.; Bellens, R.; Gautama, S.; Philips, W. Taking Optimal Advantage of Fine Spatial Resolution: Promoting partial image reconstruction for the morphological analysis of very-high-resolution images. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–28. [CrossRef]

91. Rasti, B.; Ghamisi, P.; Gloaguen, R. Hyperspectral and lidar fusion using extinction profiles and total variation component analysis. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3997–4007. [CrossRef]

92. Zhu, Z.; Woodcock, C.E.; Rogan, J.; Kellndorfer, J. Assessment of spectral, polarimetric, temporal, and spatial dimensions for urban and peri-urban land cover classification using Landsat and SAR data. *Remote Sens. Environ.* **2012**, *117*, 72–82. [CrossRef]

93. Wurm, M.; Taubenböck, H.; Weigand, M.; Schmitt, A. Slum mapping in polarimetric SAR data using spatial features. *Remote Sens. Environ.* **2017**, *194*, 190–204. [CrossRef]

94. Chang, C.C.; Lin, C.J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol. (TIST)* **2011**, *2*, 27. [CrossRef]

95. Banerjee, M.; Capozzoli, M.; McSweeney, L.; Sinha, D. Beyond kappa: A review of interrater agreement measures. *Can. J. Stat.* **1999**, *27*, 3–23. [CrossRef]

96. Hong, D.; Yokoya, N.; Xu, J.; Zhu, X. Joint and progressive learning from high-dimensional data for multi-label classification. In *European Conference on Computer Vision (ECCV)*; Springer: Cham, Switzerland, 2018; pp. 478–493.

97. Rodriguez, A.; Laio, A. Clustering by fast search and find of density peaks. *Science* **2014**, *344*, 1492–1496. [CrossRef]

98. Chazal, F.; Guibas, L.J.; Oudot, S.Y.; Skraba, P. Persistence-based clustering in riemannian manifolds. *J. ACM (JACM)* **2013**, *60*, 41. [CrossRef]