

Article

Multi-Label Classification Based on Low Rank Representation for Image Annotation

Qiaoyu Tan ¹, Yezi Liu ², Xia Chen ¹ and Guoxian Yu ^{1,*}

¹ College of Computer and Information Science, Southwest University, Chongqing 400715, China; tqy1995119@email.swu.edu.cn (Q.T.); xchen@email.swu.edu.cn (X.C.)

² College of Hanhong, Southwest University, Chongqing 400715, China; iryzbyz45@email.swu.edu.cn

* Correspondence: gxyu@swu.edu.cn; Tel.: +86-23-6825-4396

Academic Editors: Gonzalo Pajares Martinsanz and Prasad S. Thenkabail

Received: 7 November 2016; Accepted: 22 January 2017; Published: 27 January 2017

Abstract: Annotating remote sensing images is a challenging task for its labor demanding annotation process and requirement of expert knowledge, especially when images can be annotated with multiple semantic concepts (or labels). To automatically annotate these multi-label images, we introduce an approach called Multi-Label Classification based on Low Rank Representation (MLC-LRR). MLC-LRR firstly utilizes low rank representation in the feature space of images to compute the low rank constrained coefficient matrix, then it adapts the coefficient matrix to define a feature-based graph and to capture the global relationships between images. Next, it utilizes low rank representation in the label space of labeled images to construct a semantic graph. Finally, these two graphs are exploited to train a graph-based multi-label classifier. To validate the performance of MLC-LRR against other related graph-based multi-label methods in annotating images, we conduct experiments on a public available multi-label remote sensing images (Land Cover). We perform additional experiments on five real-world multi-label image datasets to further investigate the performance of MLC-LRR. Empirical study demonstrates that MLC-LRR achieves better performance on annotating images than these comparing methods across various evaluation criteria; it also can effectively exploit global structure and label correlations of multi-label images.

Keywords: remote sensing images; image annotation; multi-label classification; low-rank representation; graph construction; semantic graph

1. Introduction

Remote sensing image annotation [1] aims to assign one or several predefined labels (semantic concepts) to a remote sensing image. It is the basis of remote sensing image indexing for organizing and locating images of interest from a large database. With the development of satellite sensor technology, remote sensing images can be easily accumulated. On the other hand, manually annotating so many images is time consuming and expensive, even infeasible. Therefore, it is urgent to develop techniques that can effectively and efficiently annotate remote sensing images. Most existing remote sensing image classification methods [2–5] assume that each image is annotated with only one label within a number of candidate labels and target at annotating only one label to an image. However, in real-world applications, a remote sensing image can and should be annotated with multiple labels due to the mixture of multiple signals, the interactions of photons with matter, the atmosphere and phenomenon attributed to the physical properties of light [6]. For example, in Figure 1, the left image is annotated with two labels: “Giza Pyramids” and “Egypt”, and the right image is tagged with labels “Train Station”, “Zurich” and “Switzerland”. Given that, it is necessary to take into account the multi labels’ characteristics of remote sensing images and annotate a set of relevant labels to these images, instead

of a single label. Recently, multi-label classification [7,8] that studies the problem where an instance annotated with a set of labels has been applied to annotate multi-label remote sensing images [1,9–11] and has shown appealing performance in remote sensing applications.

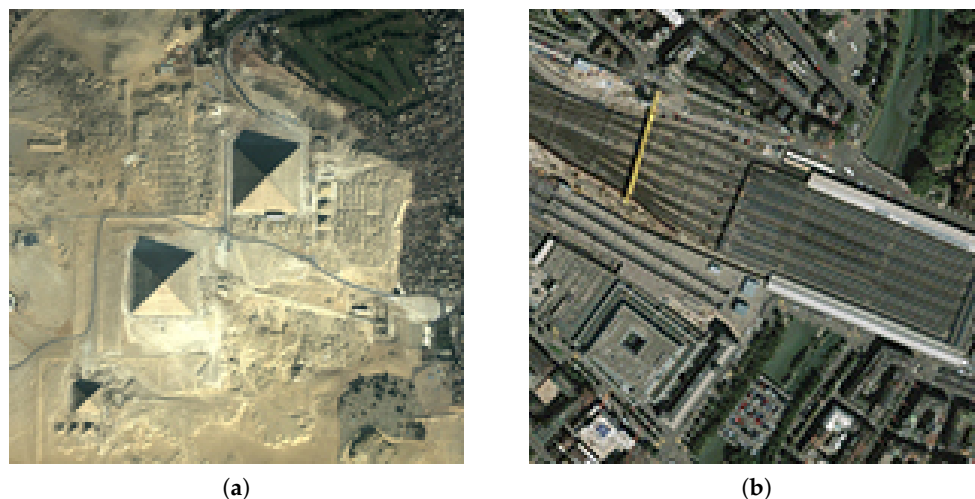


Figure 1. Examples of multi-label remote sensing images collected by QuickBird satellite and their annotations. (a) Annotations: Giza Pyramids and Egypt; (b) annotations: Train Station, Zurich and Switzerland.

The multi-label classification framework has been applied to numerous real-world applications [12]. One key challenging problem [7,8] in multi-label classification is how to make use of label correlations. As shown in Figure 1, since the left image is annotated with “Giza Pyramids”, this image has a higher probability to be annotated with the label “Egypt” than “Switzerland”. Existing strategies on exploiting label correlation could be roughly divided into three categories based on the order of correlations [8]. First-order correlation [13] assumes labels are independent and ignores intertwined effects of other labels. Second-order correlation [14] captures pairwise relationships between labels. However, there are certain real-world applications that label correlations go beyond the second-order assumption. High-order correlation [15] imposes all other labels’ influences on each label and has stronger correlation modeling capabilities than first-order and second-order strategies. However, it is computationally more demanding and less scalable than them. Various methods have been proposed to exploit label correlations for multi-label classification, and corroborated label correlations can improve the accuracy of multi-label classification. A comprehensive coverage of them is beyond the scope of this paper; readers can refer to [7,8] and the references therein.

Recently, graph-based classification methods have been introduced to remote sensing image classification for simplicity and effectiveness [3,16,17]. A common basic assumption for these methods is the good structure of the constructed graph. However, it is difficult to construct a graph that can faithfully describe the relationship between instances, especially in image annotation, since each image can be represented by various types of features (i.e., color histograms and color layout) from different aspects. Moreover, all of these methods assume that each image is annotated with only one label. In this paper, a novel graph-based multi-label classification approach called Multi-Label Classification based on Low Rank Representation for image annotation (MLC-LRR) has been proposed. Unlike traditional graph construction techniques, we take advantage of LRR for graph construction in both the feature space and label space of images. Specifically, MLC-LRR first computes the coefficient matrix of images (including both labeled and unlabeled images) in the feature space by LRR and then makes use of the matrix to define a feature-based graph. In addition to that, MLC-LRR constructs another semantic graph by using LRR again in the label space of labeled images to explore the global relationships among labels. Next, MLC-LRR fuses the feature-based graph and semantic graph into a

graph-based multi-label classifier to annotate unlabeled images. The main contributions of this paper are listed as follows:

1. We apply graph-based multi-label classification to annotate remote sensing images associated with multiple concepts (labels).
2. We exploit LRR for graph construction in the feature space and label space of images, respectively.
3. The semantic graph constructed in the label space can effectively capture global label correlation and improve the accuracy of image annotation.
4. The proposed MLC-LRR can take advantage of limited labeled images and abundant unlabeled images and shows improved performance compared to other related methods on annotating images.

The reminder of this paper is organized as follows. Section 2 briefly reviews related work on graph-based multi-label classification. Section 3 describes the proposed approach, including: (i) feature-based graph construction via LRR in the image feature space; (ii) semantic graph construction using LRR in the image label space; (iii) graph-based multi-label classification. In Section 4, we present the used image datasets, experimental results and discussions. Conclusions and future work are provided in Section 5.

2. Related Work

Graph-based multi-label classification [18,19] models all instances (or images) in a graph $G = (\mathcal{V}, \mathcal{E}, \mathbf{W})$, where \mathcal{V} is a vertex set, and each vertex represents an image; \mathcal{E} is the set of edges; and \mathbf{W} is a nonnegative adjacency matrix storing the weight of edges (or similarity) between images. Wang et al. [20] proposed an l_1 graph-based sparse coding method for image annotation called MSC. MSC constructs a label-space graph based on the overlap of label vectors. Particularly, if two images are annotated with the same labels, then the similarity between them is set to one, otherwise zero. After that, MSC takes available labels of images as features and constructs an l_1 graph based on sparse representation coefficients regularized by an l_1 norm. Next, MSC incorporates the label-space graph and l_1 graph into a linear embedding framework to project these images into low dimensional space. For a query image, MSC firstly projects the image into a low dimensional space and then computes the corresponding sparse representation coefficients with respect to labeled images. In the end, MSC predicts the labels of the query image via linear combination of label vectors of labeled images weighted by the coefficients. However, MSC, like the solutions in [1,9–11], is a supervised approach that asks for sufficient labeled images for sparse representation and dimensionality reduction. Nevertheless, it is very time consuming and expensive to collect sufficient labeled images. We may often have scarce labeled images and a large amount of unlabeled images. Therefore, many researchers have applied semi-supervised learning techniques [21–23] to leverage limited labeled images and abundant unlabeled images for annotating large-scale unlabeled remote sensing images, but these techniques still assume that each remote sensing image is annotated with a single label.

Recently, graph-based semi-supervised multi-label classification, an important branch of semi-supervised learning, has been widely applied for annotating images [18,24,25], for its flexibility and ease of application. Chen et al. [18] proposed an l_2 graph-based semi-supervised classifier for multi-label classification; the l_2 graph is constructed by utilizing neighborhood instances of an instance and neighborhood instances of its reciprocal neighbors. Specifically, this method first constructs an instance level graph in the feature space and constructs another category level graph by using cosine similarity in the label space. Then, it combines these two graphs into a regularization framework and makes a prediction for unlabeled instances by solving a Sylvester equation [26]. Yu et al. [27] proposed a Transductive Multi-label Classifier (TMC) on a directed bi-relational graph that takes both instances and labels as nodes. This graph contains three kinds of edges: between instances, between labels and between instances and labels. TMC predicts the labels of unlabeled instances based on Random Walk with Restarts (RWR) [28] on the bi-relational graph. Wang et al. [29] proposed a graph-based multi-label classification method (FCML) for multi-functional protein function prediction. FCML

uses the Green function [30] to incorporate the function-function correlations based on the theory of Reproducing Kernel Hilbert Space (RKHS) to infer protein functions. Kong et al. [19] proposed a k NN graph-based transductive multi-label classifier called Tram. Tram introduces the label composition concept and assumes similar instances should have a similar label composition. It formulates the transductive multi-label classification as an optimization problem of estimating label composition and provides a closed-form solution. Wang et al. [31] proposed a Dynamic Label Propagation (DLP) to infer labels of unlabeled images by using an l_2 -graph, which is constructed by exploiting neighborhood images of an image and neighborhood images of its reciprocal neighbors. However, all of these aforementioned methods for graph construction do not adequately exploit the global structure of data. Low Rank Representation (LRR) was recently introduced for graph construction in remote sensing image classification [32,33]. It is recognized that LRR is an appropriate approach to explore the global structure and global mixture of the subspace structure of images [34]. Jing et al. [25] proposed a graph-based low-rank mapping method for multi-label classification. This method constructs a k NN graph on both labeled and unlabeled instances and defines a mapping matrix as a linear transformation from feature space to label space. Then, this method incorporates the mapping matrix and k NN graph into a multi-label classification framework based on manifold regularization [35], but it does not explicitly utilize label correlations.

In this paper, we introduce a Multi-Label Classifier based on Low Rank Representation (MLC-LRR) for automatically annotating images. MLC-LRR takes advantage of low rank representation for graph construction and constructs a feature-based graph and a semantic graph. The feature-based graph is constructed in the feature space of labeled and unlabeled images, while the semantic graph is defined in the label space of labeled images. Next, MLC-LRR makes use of both the feature-based graph and the semantic graph to predict the labels of unlabeled images.

3. Methodology

In this section, we will briefly introduce LRR, the construction of the feature-based graph and the semantic graph and transductive multi-label classification based on the feature-based graph and the semantic graph derived from labeled images. Before that, we give some notations that will be used throughout the paper. Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbb{R}^{d \times N}$ be a set of images; $\mathbf{x}_i \in \mathbb{R}^d$ is the feature vector for the i -th image. $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N] \in \mathbb{R}^{N \times C}$ is the label indicative matrix for these N images, where C is the number of distinct labels of these images. $\mathbf{y}_i \in \mathbb{R}^C$ is the label vector of the i -th image, if image i is annotated with label c , $y_{ic} = 1$; otherwise, $y_{ic} = 0$. Without loss of generality, suppose among N images in \mathbf{X} , the first l images are labeled, and the left u images are unlabeled, $N = l + u$. Our goal is to use all of the images in \mathbf{X} to train a graph-based multi-label classifier and to predict the labels of these unlabeled images.

3.1. Low Rank Representation for Feature-Based Graph Construction

Graph-based semi-supervised classification depends on a well-structured graph [36]. However, in real applications, it is difficult to construct a graph that effectively and correctly captures the relationship between images. In this section, we take advantage of LRR to construct a feature-based graph and to capture the global relationships among images. Given that $\mathbf{X} \in \mathbb{R}^{d \times N}$ represents the feature space of both labeled and unlabeled images, where each column corresponds to an image, each image can be viewed as a linear combination of bases from a dictionary $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \dots, \mathbf{a}_M] \in \mathbb{R}^{d \times M}$. Similar to the work in [37], we set $\mathbf{A} = \mathbf{X}$ in this paper. LRR encodes each image by a linear combination of the bases in \mathbf{A} as follow:

$$\mathbf{X} = \mathbf{AZ}_1 \quad (1)$$

where $\mathbf{Z}_1 \in \mathbb{R}^{N \times N}$ is the coefficient matrix with each $\mathbf{Z}_1(\cdot, i) \in \mathbb{R}^N$ being the representation coefficient vector for image \mathbf{x}_i with respect to N images. Entry $\mathbf{Z}_1(j, i)$ can be viewed as the contribution of \mathbf{x}_j to the reconstruction of \mathbf{x}_i with \mathbf{A} as the dictionary. Different from sparse representation [38] that may not

capture the global relationship of images in \mathbf{X} , LRR enforces \mathbf{Z}_1 to be low rank and solves the following optimization problem:

$$\begin{cases} \min_{\mathbf{Z}_1} \text{rank}(\mathbf{Z}_1) \\ \text{s.t. } \mathbf{X} = \mathbf{A}\mathbf{Z}_1, \mathbf{Z}_1 \geq 0 \end{cases} \quad (2)$$

Here, \mathbf{X} is reconstructed by the low rank constrained matrix \mathbf{Z}_1 . Equation (2) is called low rank representation [34]. Since non-negativity can guarantee the physical meanings of graph weights and often results in better performance for graph construction [39], similar to Zhuang et al. [40], $\mathbf{Z}_1 \geq 0$ is added. In the iterative solution process of nonnegative LRR, negative values in \mathbf{Z}_1 are substituted with zeros in each iteration.

Due to the discrete nature of the rank function, Equation (2) in general is NP-hard [41]. One popular approach is to replace the rank function by the trace norm (or nuclear norm) [42], and Equation (2) can be relaxed to Equation (3) as follows:

$$\begin{cases} \min_{\mathbf{Z}_1} \|\mathbf{Z}_1\|_* \\ \text{s.t. } \mathbf{X} = \mathbf{A}\mathbf{Z}_1, \mathbf{Z}_1 \geq 0 \end{cases} \quad (3)$$

where $\|\mathbf{Z}_1\|_*$ denotes the nuclear norm of \mathbf{Z}_1 , and it is the sum of singular values of \mathbf{Z}_1 [43]. Recently, Zhang et al. [44] proved that Equations (2) and (3) have a closed-form solution, and Equation (3) can be further relaxed to take into account noisy features as follows:

$$\begin{cases} \min_{\mathbf{Z}_1} \|\mathbf{Z}_1\|_* + \lambda \|\mathbf{E}\|_{2,1} \\ \text{s.t. } \mathbf{X} = \mathbf{A}\mathbf{Z}_1 + \mathbf{E}, \mathbf{Z}_1 \geq 0 \end{cases} \quad (4)$$

where $\|\mathbf{E}\|_{2,1} = \sum_{j=1}^N \sqrt{\sum_{i=1}^d (\mathbf{E}_{ij})^2}$ is called the $l_{2,1}$ -norm and the parameter λ is used to trade off the effect of low-rank part and the noise tolerance part. To solve Equation (4), Liu et al. [34] exploited the well-known alternating direction method [45]. However, this method suffers from $O(n^3)$ computation complexity due to the matrix multiplication and matrix inversion. Moreover, the alternating direction method introduces auxiliary variables and constraints and, thus, downgrades the convergence rate. Such a heavy computational load of the alternating direction method prevents LRR for large-scale datasets. Fortunately, Lin et al. [46] introduced a Linearized Alternating Direction Method with Adaptive Penalty (LADMAP) to accelerate the solution of LRR. LADMAP belongs to the alternating direction method of multipliers family [45]; it solves LRR by linearizing the quadratic penalty term and adding a proximal term when solving the sub-problems. In addition, LADMAP represents \mathbf{Z}_1 as its skinny single-value decomposition and utilizes an advanced functionality of the PROPACK [47] package to accelerate the solving process. The time complexity of LADMAP for LRR is $O(rn^2)$; r is the rank of the optimal \mathbf{Z}_1 , since there is no full rank matrix multiplication. Next, each column of \mathbf{Z}_1 is normalized via $\mathbf{Z}_1(\cdot, i) = \mathbf{Z}_1(\cdot, i) / \|\mathbf{Z}_1(\cdot, i)\|_2$, and then, each negative entry of \mathbf{Z}_1 is set to zero. Since LRR jointly finds the low-ranked coefficient matrix \mathbf{Z}_1 for all images in \mathbf{X} , here, similar to [48], we adapt \mathbf{Z}_1 to define a feature-based graph whose weighted adjacent matrix is \mathbf{W} , where $\mathbf{W} = (\mathbf{Z}_1 + \mathbf{Z}_1^T) / 2$.

3.2. Low Rank Representation for Semantic Graph Construction

Graph construction is important for graph-based semi-supervised learning approaches, and many graph construction methods have been proposed [16,17,38]. All of these methods mainly construct a graph in the feature space and focus on single-label classification, where each image is restricted to be annotated with only one label. However, remote sensing images can often be associated with multiple labels. For example, one can think of two images with the same blue background; the first one is annotated with “blue color” and “ship”, while the latter is tagged with “blue color” and “airplane”.

It is easy to find that these two images are similar to each other in the feature space, since they have the same background, which may account for the majority feature similarity. However, these two images are different from each other in the semantic perspective; the first image describes a ship on the sea, while the second image describes flying in the sky.

Therefore, directly calculating the similarity between two images based on their feature vectors could not distinguish polysemous labels with different concepts in different images. To improve the performance of graph-based multi-label classification, it is necessary to develop techniques that can additionally utilize semantic information for graph construction. Most multi-label learning methods [7,8] additionally make use of pairwise (or high-order) label correlations to improve the performance; for example, pairwise label correlation computed by cosine similarity [29] and empirical conditional probability [14], high-order label correlation by classifier chain [49] and a random label set [50]. Different from these popular techniques, some multi-label classifiers take each label as a node and construct a hybrid graph with label nodes and images to represent the intra-relationship between images, between labels and inter-relationships between images [27]. The semantic information of an image can be additionally encoded by the labels annotated to that image. These label correlation-based methods do not take into account all of the labels annotated to an image, and thus, they do not make proper use of the semantic relationship between images. The semantic similarity between two multi-label images can be derived from the labels associated with these two images. Wang et al. [20] proposed a reconstruction-based semantic graph by utilizing sparse representation in the label space of labeled images. Nevertheless, the graph is constructed by separately treating each labeled image with respect to other labeled images; thus, it may be ineffective in capturing the global semantic relationship between images. Here, we construct a semantic graph to capture the global semantic similarity between images by reusing LRR on labeled images in \mathbf{Y} . Label vectors of labeled images are reconstructed with respect to \mathbf{Y} by LRR (similar to Equation (4)) as a whole, and then, the low rank coefficients are used to define a semantic graph $\mathbf{S} \in \mathbb{R}^{N \times N}$, where $\mathbf{S} = (\mathbf{Z}_2 + \mathbf{Z}_2^T)/2$ and \mathbf{Z}_2 is the low rank representation coefficient matrix with respect to labeled images in \mathbf{Y} . In practice, we just utilize labeled images to construct the semantic graph between l labeled images. To be consistent with the feature-based graph \mathbf{W} , we extend \mathbf{S} to be an $N \times N$ matrix, with the upper left sub-matrix with size $l \times l$ initialized by low rank coefficients in \mathbf{Z}_2 , and the other entries of \mathbf{S} are set to zeros.

3.3. Graph-Based Multi-Label Classification

To this end, we introduce a graph-based multi-label classifier to use the feature-based graph and the semantic graph for multi-label classification. At first, we define a matrix $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_N]^T \in \mathbb{R}^{N \times C}$, where $\mathbf{f}_i = [\mathbf{f}_{i1}, \mathbf{f}_{i2}, \dots, \mathbf{f}_{iC}]^T \in \mathbb{R}^C$ is the predicted likelihood of the i -th image with respect to C labels. Here, we consider a graph-based multi-label classifier as follows:

$$\Psi(\mathbf{F}) = \arg \min_{\mathbf{F}} \sum_{i=1}^l \Omega(\mathbf{f}_i, \mathbf{y}_i) + \alpha \|\mathbf{f}\|_F^2 + \beta \|\mathbf{f}\|_L^2 \quad (5)$$

where the first term is the empirical loss function measuring the approximation error between the annotated multi-label images and the predicted likelihoods, the second term is the regularization on the global structure of labeled and unlabeled images and the last term is to take advantage of the semantic relationship between labeled images. α and β are two parameters to balance these three terms. The first term is defined as follows:

$$\Omega(\mathbf{f}, \mathbf{y}) = \sum_{i=1}^l \|\mathbf{f}_i - \mathbf{y}_i\|^2 = \sum_{i=1}^N ((\mathbf{f}_i - \mathbf{y}_i)^T \mathbf{H}_{ii} (\mathbf{f}_i - \mathbf{y}_i)) = \text{tr}((\mathbf{F} - \mathbf{Y})^T \mathbf{H} (\mathbf{F} - \mathbf{Y})) \quad (6)$$

where $tr()$ is the matrix trace operator and $\mathbf{H} \in \mathbb{R}^{N \times N}$ is a diagonal matrix with $\mathbf{H}_{ii} = 1$ if \mathbf{x}_i is labeled, $\mathbf{H}_{ii} = 0$, otherwise. The second term of Equation (5) can be computed as:

$$\|\mathbf{f}\|_F^2 = \frac{1}{2} \sum_{i,j=1}^N \|\mathbf{f}_i - \mathbf{f}_j\|^2 \mathbf{W}_{ij} = tr(\sum_{i=1}^N (\mathbf{f}_i \mathbf{W}_{ii} \mathbf{f}_i^T) - \sum_{i,j=1}^N (\mathbf{f}_i \mathbf{W}_{ij} \mathbf{f}_j^T)) = tr(\mathbf{F}^T (\mathbf{D} - \mathbf{W}) \mathbf{F}) = tr(\mathbf{F}^T \mathbf{L} \mathbf{F}) \quad (7)$$

where \mathbf{W}_{ij} is the weight of the edge between images \mathbf{x}_i and \mathbf{x}_j . \mathbf{D} is a diagonal matrix with $\mathbf{D}_{ii} = \sum_{j=1}^N \mathbf{W}_{ij}$; $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the graph Laplacian matrix [51].

Similar to the assumption in [52], in this paper, we assume that the label of an image can be reconstructed by other related images, while the reconstructed coefficients are derived from low rank representation in the label space. Based on this assumption, the last term of Equation (5) is defined as:

$$\|\mathbf{f}\|_L^2 = \frac{1}{2} \sum_{i,j=1}^N \|\mathbf{f}_i - \sum_{j \neq i} \mathbf{S}_{ij} \mathbf{f}_j\|^2 = tr(\mathbf{F}^T (\mathbf{I} - \mathbf{S}) (\mathbf{I} - \mathbf{S})^T \mathbf{F}) = tr(\mathbf{F}^T \mathbf{M}_L \mathbf{F}) \quad (8)$$

where $\mathbf{S} \in \mathbb{R}^{N \times N}$ is the adjacent matrix of the semantic-based graph defined in Section 3.2. \mathbf{S}_{ij} represents the similarity of label vector \mathbf{y}_i and \mathbf{y}_j . $\mathbf{I} \in \mathbb{R}^{N \times N}$ is an identity matrix, and $\mathbf{M}_L = (\mathbf{I} - \mathbf{S})(\mathbf{I} - \mathbf{S})^T$. The motivation to use the last term is to replenish possible missing labels of labeled images by taking advantage of the semantic graph. If a labeled image has some missing labels and its semantic neighbors are annotated with these labels, this term can replenish missing labels of that image to some extent. In addition, by joint work with the second term on the right of Equation (5), the available labels and replenished labels of labeled images can further propagate to other unlabeled images, and thus, MLC-LRR can more completely predict the labels of unlabeled images.

Based on Equations (6)–(8), we can rewrite Equation (5) as:

$$\Psi(\mathbf{F}) = tr((\mathbf{F} - \mathbf{Y})^T \mathbf{H} (\mathbf{F} - \mathbf{Y})) + \alpha tr(\mathbf{F}^T \mathbf{L} \mathbf{F}) + \beta tr(\mathbf{F}^T \mathbf{M}_L \mathbf{F}) \quad (9)$$

The reason to minimize the first term in Equation (9) is to force the predicted outputs \mathbf{f}_i to be similar to the original labels. The motivation to minimize the second term is to ensure images with similar feature vectors having similar outputs. The last term is motivated by the label reconstruction assumption [52], where the labels of an image can be reconstructed by those of other images. In other words, if the label vectors of two images \mathbf{y}_i and \mathbf{y}_j are similar to each other, then \mathbf{S}_{ij} has a large value. In this way, the third term can capture and employ semantic information between images.

Equation (9) can be solved by taking the partial derivative of $\Psi(\mathbf{F})$ with respect to \mathbf{F} as follows:

$$\frac{\partial \Psi(\mathbf{F})}{\partial \mathbf{F}} = 2\mathbf{H}(\mathbf{F} - \mathbf{Y}) + 2\alpha \mathbf{L} \mathbf{F} + 2\beta \mathbf{M}_L \mathbf{F} \quad (10)$$

Let $\frac{\partial \Psi(\mathbf{F})}{\partial \mathbf{F}} = 0$; we can obtain the analytic solution of \mathbf{F} as:

$$\mathbf{F} = (\mathbf{H} + \alpha \mathbf{S} + \beta \mathbf{M}_L)^{-1} \mathbf{H} \mathbf{Y} \quad (11)$$

Given \mathbf{X} and \mathbf{Y} are already known, $\mathbf{S} = \mathbf{D} - \mathbf{W}$, and $\mathbf{M}_L = (\mathbf{I} - \mathbf{S})(\mathbf{I} - \mathbf{S})^T$, \mathbf{F} is mainly determined by \mathbf{W} and \mathbf{S} . Here, \mathbf{W} and \mathbf{S} are the weighted adjacent matrices of graphs constructed by LRR in the feature space and label space of images, respectively. Figure 2 briefly lists the procedure of the MLC-LRR algorithm.

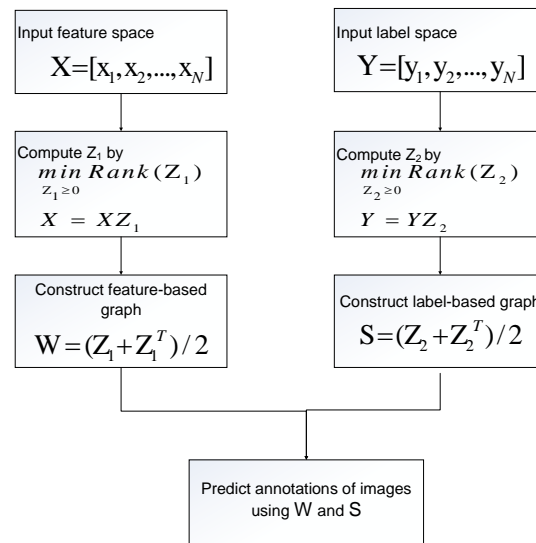


Figure 2. Flowchart of Multi-Label Classification based on Low Rank Representation (MLC-LRR).

4. Results and Discussion

We conduct experiments on Land Cover and five other publicly-available multi-label image datasets, Flags [53], Scene [13], Corel5k [54], MIRFlickr [55] and ESPGame[56], to validate the performance of our proposed MLC-LRR and compare it with five representative and related graph-based multi-label classifiers: MSC [20], TMC [27], FCML [29], Tram [19] and DLP [31]. MSC is a supervised multi-label classifier, and the other four are semi-supervised multi-label classifiers. These classifiers have been introduced in Section 2.

Land Cover is a multi-label remote sensing image dataset collected by Karalas et al. [11]. This dataset combines real satellite data from the Moderate Resolution Imaging Spectroradiometer instrument and high spatial resolution ground data from the CORINE Land Cover (CLC) project supported by European Environment Agency. We use the same features and labels as suggested in [11] for image annotation; each image is represented with 57 features with respect to 20 distinct labels, as depicted in Table 1.

Flags is a multi-label toy image dataset with 194 images in seven object classes. Each image is represented by a vector with 19 features. Scene includes 2407 images in six object classes; each image has 294 features. Corel5k, MIR FLICKRand ESPGame are three representative and popular image datasets. Similar to MLR-GL [57], we remove the images that are annotated with fewer than three labels from Corel5k and images that are tagged with fewer than five labels from MIR Flickr and ESPGame. Each image in the last three datasets is represented by dense SIFT descriptors [58]. The statistical information of these datasets used for experiments is revealed in Table 2.

In the experiments, Unless otherwise specified, we set the parameters of the comparing methods according to what the author supposed in the original papers or codes. As for MLC-LRR, we perform a cross-validation on the Flags and Scene datasets in the preliminary experiments by varying α and β from 0.01 to one with a stepsize of 0.01. Results show that MLC-LRR yields stable performance around $\alpha = 0.5$ and $\beta = 0.01$; these two values are used for MLC-LRR in the following experiments.

Table 1. Twenty ground-truth CORINE Land Cover (CLC) labels for remote sensing images in the Land Cover dataset.

No.	CLC Code	Description
1	111	Continuous urban fabric
2	121	Industrial or commercial units
3	122	Road and rail networks and assoc.land
4	124	Airports
5	131	Mineral extraction sites
6	132	Dump sites
7	133	Construction sites
8	141	Green urban areas
9	142	Sport and leisure facilities
10	212	Permanently irrigated land
11	213	Rice fields
12	223	Olive groves
13	241	Annual crops assoc. with perm.crops
14	322	Moors and heathland
15	331	Beaches, dunes, sands
16	332	Bare rocks
17	411	Inland marshes
18	412	Peat bogs
19	421	Salt marshes
20	521	Coastal lagoons

Table 2. Datasets used in the experiments. N is the number of images; D is the dimensionality of images; C is the number of distinct labels of images; Avg is the average number of labels per image.

Datasets	N	D	C	Avg
Land Cover	12,291	57	20	2.037
Flags	194	19	7	3.392
Scene	2407	294	6	2.158
Core15k	4395	1000	260	3.61
MIRFlickr	4951	1000	457	7.31
ESPGame	10,457	1000	268	6.41

4.1. Evaluation Metrics

Performance evaluation for multi-label image annotation is somewhat complicated, as each image is annotated with a set of labels, instead of a single one. Various metrics have been used to evaluate the performance of multi-label classification. Given C different labels, all of these comparing methods result in a predicted likelihood vector with respect to C labels. In this paper, we use five widely-used evaluation metrics (*RankLoss*, *AvgPrec*, *Coverage*, *MicroAvg* and *AUC*) [7,8] to quantitatively compare the performance of these multi-label classification methods in annotating images.

RankLoss evaluates the average fraction of misordered label pairs, i.e., an irrelevant label of an image is ranked ahead of a relevant label. The smaller the value of *RankLoss*, the better the performance. Its formal definition is:

$$RankLoss = \frac{1}{u} \sum_{i=1}^N \frac{|\{(Y_{ik}, Y_{ij}) | f(x_i, k) \leq f(x_i, j), (k, j) \in \mathcal{Y}_i \times \bar{\mathcal{Y}}_i\}|}{|\mathcal{Y}_i| |\bar{\mathcal{Y}}_i|} \quad (12)$$

$i \leq k, j \leq C$, $\bar{\mathcal{Y}}_i$ is the complementary set of \mathcal{Y}_i and \mathcal{Y}_i is the known label set of the i -th image.

AvgPrec evaluates the average fraction of relevant labels ranked ahead of a particular label $\mathcal{Y}_{ik} \in \mathcal{Y}_i$. The larger the value of *AvgPrec*, the better the performance. Its formal definition is:

$$AvgPrec = \frac{1}{u} \sum_{i=l+1}^N \frac{1}{|\mathcal{Y}_i|} \sum_{\mathcal{Y}_{ik} \in \mathcal{Y}_i} \frac{|\Phi(x_i, \mathcal{Y}_{ik})|}{rank(x_i, \mathcal{Y}_{ik})}, \text{ where} \quad (13)$$

$$\Phi(x_i, \mathcal{Y}_{ik}) = \{\mathcal{Y}_{ij} \mid rank(x_i, \mathcal{Y}_{ij}) \leq rank(x_i, \mathcal{Y}_{ik}), \mathcal{Y}_{ij} \in \mathcal{Y}_i\}$$

$rank(x_i, \mathcal{Y}_{ik})$ returns the rank (from largest to lowest) of \mathcal{Y}_{ik} in $f(\mathbf{x}_i)$.

MicroAvg evaluates both the micro average of precision and the micro average of recall with equal importance. The bigger the value, the better the performance. Its formal definition is:

$$MicroAvg = \frac{2 \times \sum_{i=l+1}^N |f(\mathbf{x}_i) \cap \mathcal{Y}_i|}{\sum_{i=l+1}^N |f(\mathbf{x}_i)| + \sum_{i=l+1}^N |\mathcal{Y}_i|} \quad (14)$$

MicroAvg requires the vector to be a binary indicator vector. Here, we consider the labels corresponding to the r largest entries of $f(\mathbf{x}_i)$ as the predicted labels of image \mathbf{x}_i , where r is determined as the average number of labels (rounded to the next integer) of annotated images. From Table 1, r for Land Cover, Flags, Scene, Core15k, MIR Flickr and ESPGame is 3, 4, 3, 4, 8 and 7, respectively.

The adapted Area Under the Curve (AUC) was proposed and utilized in [57]. AUC firstly sorts the predicted likelihood scores vector for each image in descending order; it then varies the number of predicted labels from one to C and computes the receiver operating characteristic curve by calculating the true positive rate and the false positive rate for each number of predicted labels. It finally computes the area under the curve to evaluate the performance of multi-label classification.

Coverage evaluates how many steps are needed, on average, to move down the ranked label list of an image to cover all of its relevant labels. Its formal definition is:

$$Coverage = \frac{1}{u} \sum_{i=l+1}^N \max_{\mathcal{Y}_{ik} \in \mathcal{Y}_i} rank(x_i, \mathcal{Y}_{ik}) - 1 \quad (15)$$

Obviously, different from the above four metrics, *Coverage* can be greater than one, and the smaller the value of *Coverage*, the better the performance is.

To maintain consistency with *AUC*, *MicroAvg* and *AvgPrec*, we use $1 - RankLoss$ instead of *RankLoss*. Thus, the larger the value of these metrics (except *Coverage*), the better the performance is. We want to remark that these metrics evaluate the performance of multi-label image annotation from different aspects; it is difficult for an approach to consistently perform better than other methods across all of these metrics.

4.2. Experimental Results on Annotating Remote Sensing Images

In this section, we conduct experiments on the Land Cover dataset to investigate the performance of MLC-LRR on annotating remote sensing images by using both labeled and unlabeled images and compare the performance of MLC-LRR with other comparing methods. We randomly partition the images of the Land Cover dataset into two sets; one is used as labeled images, and the other is used as unlabeled images for validation. Here, we consider two different label ratios: 10% and 15%. 10% (or 15%) means we randomly select 10% (or 15%) of images in a dataset as the labeled training set and take the remaining images as unlabeled images to be annotated by these comparing methods. Table 3 reports the results of these comparing methods on the Land Cover dataset. To reduce any random effect, for each fixed label ratio, we repeat independent experiments for each method 10 times and report the average results. In the table, \bullet/\circ indicates whether MLC-LRR is statistically (pairwise t -test at 95% significance level) superior/inferior to the comparing methods under a particular evaluation metric. \downarrow along with *Coverage* means the lower the value, the better the performance.

Table 3. Experimental results (mean \pm std) on the Land Cover dataset with 10% and 15% labeled images. DLP, Dynamic Label Propagation (DLP); TMC, Transductive Multi-label Classifier. \bullet/\circ indicates whether MLC-LRR is statistically (pairwise t -test at 95% significance level) superior/inferior to the comparing methods under a particular evaluation metric.

	MicroAvg	1-RankLoss	AvgPrec	AUC	Coverage \downarrow
10%					
MSC	0.269 \pm 0.004 \bullet	0.699 \pm 0.001 \bullet	0.396 \pm 0.002 \bullet	0.721 \pm 0.001 \bullet	8.882 \pm 0.030 \bullet
DLP	0.226 \pm 0.021 \bullet	0.613 \pm 0.057 \bullet	0.327 \pm 0.030 \bullet	0.635 \pm 0.059 \bullet	8.389 \pm 0.770 \bullet
Tram	0.263 \pm 0.033 \bullet	0.109 \pm 0.082 \bullet	0.259 \pm 0.048 \bullet	0.607 \pm 0.031 \bullet	11.530 \pm 0.784 \bullet
TMC	0.253 \pm 0.003 \bullet	0.652 \pm 0.001 \bullet	0.379 \pm 0.001 \bullet	0.672 \pm 0.001 \bullet	9.700 \pm 0.032 \bullet
FCML	0.020 \pm 0.000 \bullet	0.286 \pm 0.001 \bullet	0.123 \pm 0.000 \bullet	0.304 \pm 0.001 \bullet	16.354 \pm 0.023 \bullet
MLC-LRR	0.341 \pm 0.017	0.740 \pm 0.009	0.467 \pm 0.016	0.761 \pm 0.009	7.867 \pm 0.236
15%					
MSC	0.276 \pm 0.003 \bullet	0.702 \pm 0.001 \bullet	0.397 \pm 0.001 \bullet	0.724 \pm 0.001 \bullet	8.810 \pm 0.032 \bullet
DLP	0.231 \pm 0.009 \bullet	0.666 \pm 0.006 \bullet	0.341 \pm 0.017 \bullet	0.692 \pm 0.005 \bullet	9.486 \pm 0.143 \bullet
Tram	0.406 \pm 0.057 \circ	0.438 \pm 0.132 \bullet	0.462 \pm 0.081 \bullet	0.737 \pm 0.052 \bullet	8.229 \pm 1.314 \bullet
TMC	0.253 \pm 0.003 \bullet	0.653 \pm 0.001 \bullet	0.381 \pm 0.001 \bullet	0.673 \pm 0.001 \bullet	9.657 \pm 0.028 \bullet
FCML	0.020 \pm 0.000 \bullet	0.286 \pm 0.001 \bullet	0.123 \pm 0.000 \bullet	0.304 \pm 0.001 \bullet	16.367 \pm 0.024 \bullet
MLC-LRR	0.369 \pm 0.010	0.759 \pm 0.009	0.488 \pm 0.012	0.780 \pm 0.009	7.382 \pm 0.213

We additionally investigate the performance of MLC-LRR with a normalized graph Laplacian and introduce $\text{MLC-LRR}_{\tilde{L}}$, which is adopted from MLC-LRR by replacing \mathbf{L} in Equation (7) with a normalized graph Laplacian matrix $\tilde{\mathbf{L}} = \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$. Similar to previous experimental protocols, we randomly select some images of the Land Cover dataset as labeled images and take the remaining images as unlabeled images for annotation. Figure 3 reports the results of MLC-LRR and $\text{MLC-LRR}_{\tilde{L}}$ with respect to MicroAvg, 1-RankLoss, AvgPrec and Coverage, respectively. These reported results are the average of ten independent experiments for each fixed label ratio (from 10% to 35% with the stepsize as 5%).

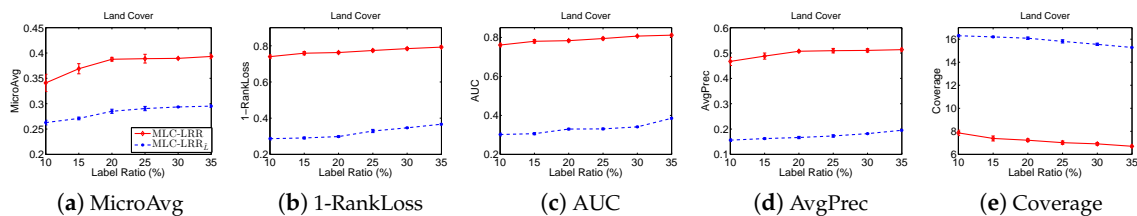


Figure 3. MLC-LRR vs. $\text{MLC-LRR}_{\tilde{L}}$ on the Land Cover dataset under different label ratios across five evaluation metrics (a–e).

4.3. The Benefit of the Semantic Graph

In this subsection, to further study the effectiveness of MLC-LRR in employing semantic graph \mathbf{S} , we conduct another set of experiments on a variant of MLC-LRR: MLC-LRR_F is adopted from MLC-LRR by only using feature-based graph \mathbf{W} . Similar to previous experimental protocols, we randomly draw 10% to 35% of the images of the Land Cover dataset as labeled images and use the remaining images as unlabeled ones for annotation. Ten independent experiments are conducted under each label ratio (10% to 35% with a stepsize of 5%), and meanwhile, the mean value of each evaluation metric is recorded under each label ratio. Figure 4 shows the performance of MLC-LRR and MLC-LRR_F with respect to MicroAvg, 1-RankLoss, AUC, AvgPrec and Coverage, respectively.

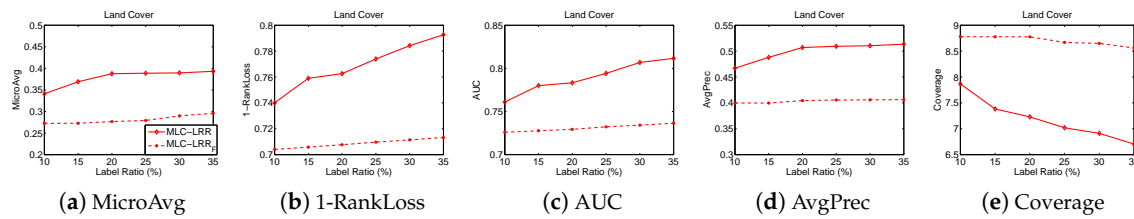


Figure 4. MLC-LRR vs. MLC-LRR_F on the Land Cover dataset under different label ratios across five evaluation metrics (a–e).

4.4. Experimental Results on Other Multi-Label Image Datasets

In this section, we conduct experiments on five public image datasets (Flags, Scene, Core15k, MIR Flickr and ESPGame) to validate the performance of MLC-LRR in annotating multi-label images and compare it with other related methods. Here, we randomly partition 30% images of each dataset as the labeled training set and take the remaining images as the unlabeled images. Table 4 reports the results of these comparing methods across various evaluation metrics and: MicroAve, 1-RankLoss, AvgPrec, AUC and Coverage. In the table, ●/○ indicates whether MLC-LRR is statistically (pairwise *t*-test at 95% significance level) superior/inferior to the comparing methods under a particular evaluation metric. ↓ along with *Coverage* means the lower the value, the better the performance.

Table 4. Experimental results (mean ± std) on Flags, Scene, Core15k, MIR Flickr and ESPGame with 30% labeled images. ●/○ indicates whether MLC-LRR is statistically (pairwise *t*-test at 95% significance level) superior/inferior to the comparing methods under a particular evaluation metric.

	MicroAvg	1-RankLoss	AvgPrec	AUC	Coverage ↓
Flags					
MSC	0.689 ± 0.004	0.712 ± 0.006 ●	0.799 ± 0.004	0.734 ± 0.005 ●	3.567 ± 0.035
DLP	0.685 ± 0.004	0.715 ± 0.004 ●	0.801 ± 0.002	0.735 ± 0.002 ●	3.590 ± 0.033
Tram	0.693 ± 0.005	0.704 ± 0.009 ●	0.789 ± 0.007	0.725 ± 0.006 ●	3.603 ± 0.037
TMC	0.680 ± 0.002	0.710 ± 0.003 ●	0.800 ± 0.003	0.735 ± 0.002 ●	3.604 ± 0.040
FCML	0.473 ± 0.003 ●	0.285 ± 0.004 ●	0.541 ± 0.003 ●	0.414 ± 0.003 ●	4.704 ± 0.009 ●
MLC-LRR	0.689 ± 0.004	0.726 ± 0.005	0.804 ± 0.003	0.747 ± 0.004	3.510 ± 0.035
Scene					
MSC	0.268 ± 0.012 ●	0.519 ± 0.018 ●	0.447 ± 0.008 ●	0.603 ± 0.021 ●	2.511 ± 0.093 ●
DLP	0.229 ± 0.006 ●	0.497 ± 0.004 ●	0.409 ± 0.002 ●	0.583 ± 0.005 ●	2.586 ± 0.017 ●
Tram	0.615 ± 0.001 ○	0.919 ± 0.001 ○	0.856 ± 0.002 ○	0.942 ± 0.001 ○	0.489 ± 0.005 ○
TMC	0.114 ± 0.013 ●	0.300 ± 0.009 ●	0.288 ± 0.010 ●	0.371 ± 0.011 ●	3.589 ± 0.047 ●
FCML	0.128 ± 0.010 ●	0.383 ± 0.010 ●	0.337 ± 0.008 ●	0.452 ± 0.012 ●	3.128 ± 0.046 ●
MLC-LRR	0.547 ± 0.016	0.837 ± 0.017	0.720 ± 0.027	0.896 ± 0.011	0.909 ± 0.089
Core15k					
MSC	0.204 ± 0.014 ●	0.697 ± 0.032 ●	0.223 ± 0.010 ●	0.724 ± 0.010 ●	59.579 ± 1.906 ●
DLP	0.215 ± 0.004 ●	0.748 ± 0.001 ●	0.238 ± 0.001 ●	0.751 ± 0.001 ●	54.750 ± 0.170 ●
Tram	0.326 ± 0.001 ○	0.853 ± 0.001 ○	0.391 ± 0.001 ○	0.856 ± 0.001 ○	34.309 ± 0.146 ○
TMC	0.224 ± 0.001 ●	0.733 ± 0.001 ●	0.235 ± 0.002 ●	0.734 ± 0.001 ●	57.141 ± 0.078 ●
FCML	0.009 ± 0.000 ●	0.242 ± 0.001 ●	0.033 ± 0.001 ●	0.247 ± 0.001 ●	103.46 ± 0.078 ●
MLC-LRR	0.279 ± 0.000	0.833 ± 0.001	0.333 ± 0.002	0.837 ± 0.001	38.655 ± 0.324
MIR Flickr					
MSC	0.088 ± 0.002 ●	0.649 ± 0.006 ●	0.093 ± 0.002 ●	0.667 ± 0.003 ●	273.514 ± 1.830 ●
DLP	0.091 ± 0.003 ●	0.695 ± 0.001	0.101 ± 0.001	0.699 ± 0.001	257.810 ± 0.409 ○
Tram	0.060 ± 0.001 ●	0.656 ± 0.001 ●	0.074 ± 0.000 ●	0.658 ± 0.001 ●	267.765 ± 0.338 ●
TMC	0.104 ± 0.000 ●	0.698 ± 0.021	0.109 ± 0.000	0.702 ± 0.000	257.980 ± 0.352 ○
FCML	0.008 ± 0.000 ●	0.301 ± 0.001 ●	0.021 ± 0.000 ●	0.299 ± 0.012 ●	349.629 ± 0.092 ●
MLC-LRR	0.104 ± 0.000	0.698 ± 0.001	0.109 ± 0.000	0.703 ± 0.001	258.955 ± 0.606

Table 4. Cont.

	MicroAvg	1-RankLoss	AvgPrec	AUC	Coverage ↓
ESPGame					
MSC	0.125 ± 0.046 •	0.766 ± 0.001 •	0.211 ± 0.001 •	0.770 ± 0.001 •	158.048 ± 0.461 •
DLP	0.214 ± 0.002 •	0.778 ± 0.001 •	0.213 ± 0.001 •	0.781 ± 0.001 •	151.113 ± 0.425 •
Tram	0.255 ± 0.001 •	0.818 ± 0.012 •	0.266 ± 0.001 •	0.821 ± 0.021 •	128.114 ± 0.303
TMC	0.210 ± 0.005 •	0.773 ± 0.001 •	0.210 ± 0.001 •	0.776 ± 0.001 •	154.484 ± 0.538 •
FCML	0.005 ± 0.000 •	0.218 ± 0.001 •	0.021 ± 0.000 •	0.217 ± 0.001 •	250.081 ± 0.061 •
MLC-LRR	0.261 ± 0.007	0.823 ± 0.007	0.266 ± 0.012	0.826 ± 0.007	126.290 ± 0.421

4.5. Discussion

4.5.1. Results Analysis on Land Cover Images

As we can see from Table 3, the performance of all methods increases with the increase of the number of labeled images. By comparing with the five related methods, MLC-LRR performs better than or comparable to them in most cases. In summary, out of 10 configurations (1 dataset × 5 evaluation metrics × 2 settings of label ratios), MLC-LRR always outperforms MSC, DLP, TMC and FCML, outperforms TRAM in nine configurations and ties with it only in one configuration. These results demonstrate the advantage of MLC-LRR in annotating multi-label remote sensing images. MSC is an inductive multi-label classifier. Both MSC and MLC-LRR explicitly utilize label correlation and semantic information for graph construction, where MSC constructs an l_1 graph in the label space and MLC-LRR constructs a semantic graph by low rank representation in the label space. However, MLC-LRR always outperforms MSC. This is principally because MSC assumes a large amount of labeled images is available, and it only exploits labeled images, discarding abundant unlabeled images in the training process. Moreover, constructing a graph by low rank representation can better capture the global relationship among images than sparse representation. These results hint that unlabeled images can be used to boost the performance of multi-label classification. DLP is a dynamic label propagation method. In fact, MLC-LRR also makes a prediction of unlabeled images by label propagation. However, DLP is outperformed by MLC-LRR in many cases. One possible reason is that DLP just makes use of the relationship between images in feature space without using label correlations. In contrast, MLC-LRR additionally utilizes semantic information among images by low rank representation. Both Tram and MLC-LRR employ unlabeled images, but Tram is outperformed by MLC-LRR. There are two passable reasons: (i) Tram does not explicitly utilize label correlation; (ii) Tram constructs a k NN graph in the feature space to capture the relationship between images while MLC-LRR constructs a feature-based graph by low-rank representation. It may not be reliable to construct a predefined k NN graph in the feature space for automatically annotating remote sensing images. This observations suggest that the feature-based graph adopted by MLC-LRR can effectively capture the global relationship of remote sensing images.

TMC constructs a directed bi-relational graph to capture three types of relationships: relationship between labels, relationship between images and association between images and labels. Both TMC and MLC-LRR utilize unlabeled images and explicitly employ label correlation, but TMC is outperformed by MLC-LRR in many cases. The passable reason is that TMC just considers pairwise label correlation, whereas MLC-LRR exploits LRR to explore and employ high order label correlation. FCML is another transductive multi-label classifier that takes advantage of pairwise label correlation. Both FCML and MLC-LRR explicitly employ label correlations; however, MLC-LRR significantly outperforms FCML. The cause is that FCML utilizes cosine similarity to capture the relationship between different labels; MLC-LRR constructs a semantic graph to capture the global relationship among labels by LRR, which can effectively capture the global semantic relationship among images. These results not only corroborate the effectiveness of LRR in capturing the global structure of remote sensing images, but

also indicate the significance of the semantic graph in capturing the correlation among labels and the semantic relationship between images.

Experimental results in Figure 3 also support the effectiveness of the semantic graph in utilizing the semantic relationship between images. From this figure, we can see that the performance of MLC-LRR and MLC-LRR_F increases with the increase of label ratio (from 10% to 35%). Both MLC-LRR and MLC-LRR_F utilize unlabeled remote sensing images for training; however, MLC-LRR outperforms MLC-LRR_F in all evaluation metrics due to additionally employing the semantic graph. In other words, the semantic graph can further improve the performance of MLC-LRR_F. From these results, we can observe that MLC-LRR performs significantly better than MLC-LRR_F in all evaluation metrics. Taking 1-RankLoss in Figure 3b for example, with labeled images increasing from 10% to 35%, MLC-LRR_F increases by 1.00%, while MLC-LRR increases by 5.19%. The additional improvement of MLC-LRR with respect to MLC-LRR_F is attributed to the semantic graph. These results suggest that the semantic graph derived from labels of annotated remote sensing images can be used to boost the performance of image annotation. In addition, these results also justify our motivation to combine both the feature-based graph and the semantic graph for image annotation.

As shown in Figure 4, we can observe that the performance of MLC-LRR and MLC-LRR_L increases with the increase of label ratio. However, MLC-LRR always performs better than MLC-LRR_L across five evaluation metrics. This is possible because that un-normalized graph Laplacian is more suitable for the Land Cover dataset. Another interesting observation is that MLC-LRR_L generally has smaller variance than MLC-LRR. This observation shows that the normalized graph Laplacian provides additional stability.

4.5.2. Results Analysis on Other Multi-Label Images

From Table 4, we can also observe that MLC-LRR performs better (or is comparable to) than the five other related graph-based multi-label classifiers in annotating multi-label images in most cases. The overall observation is similar as that in Table 3. Particularly, out of 25 configurations (5 datasets \times 5 evaluation metrics), MLC-LRR outperforms MSC, DLP, Tram and TMC in 88.00%, 72.00%, 44.00% and 72.00% of cases, ties with them in 12.00%, 24.00%, 20.00% and 24.00% of cases and loses to DLP, Tram and TMC in 4.00%, 40.00% and 4.00% of cases, respectively. MLC-LRR outperforms FCML in all configurations as on the Land Cover dataset. We also observe that MLC-LRR performs significantly better than MSC, DLP, Tram and TMC in 100.00%, 73.33%, 60.00% and 73.33% of cases on three high-dimensional image datasets: Core15k, MIR Flickr and ESPGame. MLC-LRR on average improves them by 32.00%, 16.00%, 28.00% and 16.00% on the last three high-dimensional image datasets, respectively. Taking experimental results with respect to MIR Flickr in Table 4 for example, we can observe that MLC-LRR achieves the best (or comparable) performance among all comparing methods across various evaluation metrics. These results demonstrate the advantage of MLC-LRR in annotating high-dimensional multi-label images.

We also observe that the performance of MLC-LRR increases as the average number of labels per image (*Avg*) increases, compared with the five other related graph-based methods. This is principally because label correlation is more prominent with the increase of the average number of labels of an image; thus, the semantic graph of MLC-LRR can more effectively capture the semantic relationship between images. This fact implies that the semantic graph is useful for capturing label correlation and for image annotation, which coincides with the results on Land Cover images revealed in Figure 3.

Another interesting observation is that MLC-LRR performs similarly to Tram on the Flags and Scene datasets, but it outperforms Tram on the Core15k, MIR Flickr and ESPGame datasets. In practice, both MLC-LRR and Tram take advantage of the label composition concept for graph-based multi-label classification. Tram constructs a k NN graph by the Euclidean distance between images in the feature space, while MLC-LRR constructs a feature-based graph by low-rank representation. These contrast results can be attributed to the dimensionality of images in the first two datasets being much lower than that of the last three datasets; there may be no significant difference between the k NN graph and




the feature-based graph in these two datasets. This fact indicates that MLC-LRR is more suitable for annotating high-dimensional multi-label images.

4.6. Toy Examples

In this section, we conduct another experiment to visually compare MLC-LRR and the five other related methods in annotating images in the IAPRTC-12 [59] image dataset. The number of distinct labels is 291 in IAPRTC-12, and each image is described by dense SIFT features and is represented by a 1000-dimensional vector. We filter rare labels by keeping the top 30% frequent labels and remove the images that are assigned to fewer than six labels. In the end, we get 3207 images with 88 labels, then we partition these images into two parts. The first part accounts for 30% images used as labeled instances; the second part accounts for 70% images used as unlabeled ones, whose labels are annotated by these comparing methods. In order to transform the predicted likelihood vector into a binary indicator vector, here, we consider the labels corresponding to the r largest entries of $f(\mathbf{x})$ as the predicted labels of image \mathbf{x} , where r is determined as the maximum number of labels of these annotated images. Here, r is fixed as 10.

From Table 5, we can observe that MLC-LRR has better (or comparable) performance than other comparing methods in annotating images. In summary, out of 23 ($9 + 7 + 7$) labels in these three images, MLC-LRR, MSC, Tram, TMC, FCML and DLP correctly predict 21, 15, 14, 16, 2 and 19 of them, respectively.

Table 5. Predicted labels for exemplar images in IAPRTC-12 [59]. The ground-truth labels are highlighted in bold font, and the potentially correct labels are in italic font.

			
True Labels	bike, front, house, mountain people, sky, street, tree, wall	building, car, group people, sky, street, tree	front, house, lamp roof, sky, street, tree
MSC	<i>building, front, house</i> <i>lamp, man, mountain</i> people, sky, square, tree	building, cloud, front house, mountain, people sky, square, tower, tree	<i>building, front, house</i> lamp, man, mountain people, sky, tree, wall
Tram	<i>building, front, house</i> <i>man, mountain, people</i> roof, sky, street, woman	desert, front, house mountain, people, rock sky, square, street, tree	bush, flower, front house, lamp, lawn people, square, tree, street
TMC	<i>building, front, house</i> <i>lamp, man, mountain</i> people, sky, tree, wall	building, front, house <i>lamp, man, mountain</i> people, sky, tree, wall	<i>building, front, house</i> lamp, man, mountain people, sky, tree, wall
FCML	boy, bridge, cliff cycling, cyclist, <i>hair</i> jersey, people , short, <i>sweater</i>	boy, bridge, cliff cycling, cyclist, <i>hair</i> jersey, people , short, <i>sweater</i>	boy, bridge, cliff cycling, cyclist, <i>hair</i> jersey, people, short, <i>sweater</i>
DLP	<i>building, front, house</i> <i>lamp, mountain, people</i> sky, street, tree, wall	building, front, house <i>lamp, mountain, people</i> sky, street, tree, wall	<i>building, front, house</i> lamp, mountain, people sky, street, tree, wall
MLC-LRR	<i>building, front, house</i> <i>lamp, mountain, people</i> sky, street, tree, wall	building, group, house <i>lamp, mountain, people</i> sky, street, tree, wall	<i>building, front, house</i> lamp, mountain, people roof, sky, street, tree

Another interesting observation is that these comparing methods also annotate additionally correct labels to these images. For example, “lamp”, “building”, “man” and “sweater” are not in the

ground-truth label set of the first image; “lamp”, “man” and “hair” are not in the ground-truth label set of the second image; “building” is not in the ground-truth label set of the last image; but we can see that “lamp”, “building”, “man” and “sweater” are annotated to the first image, “lamp”, “man” and “hair” to the second image, and “building” to the last image by these comparing methods. By taking into account these additional labels, MLC-LRR also produces more correct labels than these comparing methods.

These examples show that the proposed method has the potential to provide more complete annotations of images than other comparing methods. These toy examples also indicate the potential application of MLC-LRR in collaborative tagging systems (also known as folksonomy) [60]; for example, folksonomy-based personalized searches [61], folksonomy-based recommender systems [62] and folksonomy-based social media services [63,64].

5. Conclusions

In this paper, we take advantage of low-rank representation for graph construction and introduce a graph-based Multi-Label Classifier based on Low Rank Representation (MLC-LRR) for annotating remote sensing images associated with multiple labels. Unlike existing methods that focus on the single-label problem and just construct graphs in the feature space, we construct a semantic graph based on LRR in the label space. Our empirical study with five related methods on the Land Cover remote sensing images shows that MLC-LRR performs significantly better than these methods. In addition, we find that the synergy of the two graphs constructed by LRR in the feature and in the label space achieves better performance than using the graph in the feature space alone. Extra experimental results on five multi-label image datasets and a case study again show that MLC-LRR performs better than these comparing methods in annotating multi-label images. In the future, we want to investigate the effectiveness of LRR with other graph-based multi-label classifiers and design more effective algorithms for remote sensing image annotation.

Acknowledgments: The authors are grateful to the anonymous reviewers and editor’s suggestions on improving this paper. This work is supported by the Natural Science Foundation of China (61402378), the Natural Science Foundation of Chongqing Science and Technology Commission(cstc2014jcyjA40031 and cstc2016jcyjA0351), the Fundamental Research Funds for the Central Universities of China (2362015XK07, XDJK2016B009 and XDJK2016E076), the National Undergraduate Training Programs for Innovation and Entrepreneurship (201610635047) and the Southwest University Undergraduate Science and Technology Innovation Fund Project (20153601001).

Author Contributions: Guoxian Yu proposed the idea and conceived of the whole program. Qiaoyu Tan implemented the experiments and drafted the manuscript. Guoxian Yu, Yezi Liu and Xia Chen participated in analyzing the experimental data and revising the manuscript. All of the authors read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; nor in the decision to publish the results.

References

1. Chen, K.; Jian, P.; Zhou, Z.; Zhang, D. Semantic annotation of high-resolution remote sensing images via Gaussian process multi-instance multilabel learning. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1285–1289.
2. Lu, D.; Weng, Q. A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.* **2007**, *28*, 823–870.
3. Xia, G.S.; Wang, Z.; Xiong, C.; Zhang, L. Accurate annotation of remote sensing images via active spectral clustering with little expert knowledge. *Remote Sens.* **2015**, *7*, 15014–15045.
4. Huang, L.; Chen, C.; Li, W.; Du, Q. Remote sensing image scene classification using multi-scale completed local binary patterns and Fisher vectors. *Remote Sens.* **2016**, *8*, 483.
5. Ko, C.; Sohn, G.; Rimmel, T.K.; Miller, J.R. Maximizing the diversity of ensemble random forests for tree genera classification using high density LiDAR Data. *Remote Sens.* **2016**, *8*, 646.
6. Keshava, N. A survey of spectral unmixing algorithms. *Linc. Lab. J.* **2003**, *14*, 55–78.

7. Tsoumakas, G.; Katakis, I.; Vlahavas, I. Mining multi-label data. In *Data Mining and Knowledge Discovery Handbook*; Springer: New York, NY, USA, 2009; pp. 667–685.
8. Zhang, M.L.; Zhou, Z.H. A review on multi-label learning algorithms. *IEEE Trans. Knowl. Data Eng.* **2014**, *26*, 1819–1837.
9. Zhou, T.H.; Wang, L.; Ryu, K.H. Supporting keyword search for image retrieval with integration of probabilistic annotation. *Sustainability* **2015**, *7*, 6303–6320.
10. Karalas, K.; Tsagkatakis, G.; Zervakis, M.; Tsakalides, P. Deep learning for multi-label land cover classification. *Proc. SPIE* **2015**, *96430*, doi:10.1117/12.2195082.
11. Karalas, K.; Tsagkatakis, G.; Zervakis, M.; Tsakalides, P. Land classification using remotely sensed data: Going multilabel. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3548–3563.
12. Santos, A.; Canuto, A.; Neto, A.F. A comparative analysis of classification methods to multi-label tasks in different application domains. *Int. J. Comput. Inf. Syst. Ind. Manag. Appl.* **2011**, *3*, 218–227.
13. Boutell, M.R.; Luo, J.; Shen, X.; Brown, C.M. Learning multi-label scene classification. *Pattern Recognit.* **2004**, *37*, 1757–1771.
14. Ghamrawi, N.; McCallum, A. Collective multi-label classification. In Proceedings of the 14th ACM International Conference on Information and Knowledge Management, ACM, Bremen, Germany, 31 October–5 November 2005; pp. 195–200.
15. Read, J.; Pfahringer, B.; Holmes, G.; Frank, E. Classifier chains for multi-label classification. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Bled, Slovenia, 7–11 September 2009; pp. 254–269.
16. Camps-Valls, G.; Muñoz-Marí, J.; Gómez-Chova, L.; Richter, K.; Calpe-Maravilla, J. Biophysical parameter estimation with a semisupervised support vector machine. *IEEE Geosci. Remote Sens. Lett.* **2009**, *6*, 248–252.
17. Tuia, D.; Verrelst, J.; Alonso, L.; Pérez-Cruz, F.; Camps-Valls, G. Multioutput support vector regression for remote sensing biophysical parameter estimation. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 804–808.
18. Chen, G.; Song, Y.; Wang, F.; Zhang, C. Semi-supervised Multi-label Learning by Solving a Sylvester Equation. In Proceedings of the 8th SIAM International Conference on Data Mining, Atlanta, GA, USA, 24–26 April 2008; pp. 410–419.
19. Kong, X.; Ng, M.K.; Zhou, Z.H. Transductive multilabel learning via label set propagation. *IEEE Trans. Knowl. Data Eng.* **2013**, *99*, 704–719.
20. Wang, C.; Yan, S.; Zhang, L.; Zhang, H.J. Multi-label sparse coding for automatic image annotation. In Proceedings of the 20th IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 1643–1650.
21. Erkan, A.N.; Camps-Valls, G.; Altun, Y. Semi-supervised remote sensing image classification via maximum entropy. In Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing, Kittila, Finland, 29 August–1 September 2010; pp. 313–318.
22. Li, J.; Bioucas-Dias, J.M.; Plaza, A. Semisupervised hyperspectral image classification using soft sparse multinomial logistic regression. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 318–322.
23. Uhlmann, S.; Kiranyaz, S.; Gabbouj, M. Semi-supervised learning for ill-posed polarimetric SAR classification. *Remote Sens.* **2014**, *6*, 4801–4830.
24. Tang, J.; Hong, R.; Yan, S.; Chua, T.S.; Qi, G.J.; Jain, R. Image annotation by k nn-sparse graph-based label propagation over noisily tagged web images. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 135–136.
25. Jing, L.; Yang, L.; Yu, J.; Ng, M.K. Semi-supervised low-rank mapping learning for multi-label classification. In Proceedings of the 28th IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 8–10 June 2015; pp. 1483–1491.
26. Hu, D.Y.; Reichel, L. Krylov-subspace methods for the Sylvester equation. *J. Linear Algebra Its Appl.* **1992**, *172*, 283–313.
27. Yu, G.; Rangwala, H.; Domeniconi, C.; Zhang, G.; Yu, Z. Protein function prediction using multilabel ensemble classification. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2013**, *10*, 1045–1057.
28. Tong, H.; Faloutsos, C.; Pan, J.Y. Fast random walk with restart and its applications. In Proceedings of the 6th IEEE International Conference on Data Mining, Hong Kong, China, 18–22 December 2006; pp. 613–622.
29. Wang, H.; Huang, H.; Ding, C. Function-function correlated multi-label protein function prediction over interaction networks. *J. Comput. Biol.* **2013**, *20*, 322–343.

30. Ding, C.; Simon, H.D.; Jin, R.; Li, T. A learning framework using Green's function and kernel regularization with application to recommender system. In Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, CA, USA, 12–15 August 2007; pp. 260–269.
31. Wang, B.; Tu, Z.; Tsotsos, J.K. Dynamic label propagation for semi-supervised multi-class multi-label classification. In Proceedings of the 15th IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 425–432.
32. Lu, X.; Wang, Y.; Yuan, Y. Graph-regularized low-rank representation for destriping of hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 4009–4018.
33. Shi, Q.; Du, B.; Zhang, L. Domain adaptation for remote sensing image classification: A low-rank reconstruction and instance weighting label propagation inspired algorithm. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1–13.
34. Liu, G.; Lin, Z.; Yan, S.; Sun, J.; Yu, Y.; Ma, Y. Robust recovery of subspace structures by low-rank representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 171–84.
35. Belkin, M.; Niyogi, P.; Sindhwani, V. Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples. *J. Mach. Learn. Res.* **2006**, *7*, 2399–2434.
36. Zhu, X. *Semi-Supervised Learning Literature Survey*; Technical Report 1530, Computer Sciences; University of Wisconsin-Madison: Madison, WI, USA, 2008.
37. Yang, S.; Wang, X.; Wang, M.; Han, Y.; Jiao, L. Semi-supervised low-rank representation graph for pattern recognition. *IET Image Process.* **2013**, *7*, 131–136.
38. Liang, H.; Li, Q. Hyperspectral imagery classification using sparse representations of convolutional neural network features. *Remote Sens.* **2016**, *8*, 99.
39. He, R.; Zheng, W.S.; Hu, B.G.; Kong, X.W. Nonnegative Sparse Coding for Discriminative Semi-supervised Learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA, 20–25 June 2011; pp. 2849–2856.
40. Zhuang, L.; Gao, H.; Lin, Z.; Ma, Y.; Zhang, X.; Yu, N. Non-negative low rank and sparse graph for semi-supervised learning. In Proceedings of the 23th IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 2328–2335.
41. Fazel, M. Matrix Rank Minimization with Applications. Ph.D. Thesis, Stanford University, Stanford, CA, USA, 2002.
42. Vandenberghe, L.; Boyd, S. Semidefinite programming. *SIAM Rev.* **1996**, *38*, 49–95.
43. Cai, J.F.; Candès, E.J.; Shen, Z. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* **2008**, *20*, 1956–1982.
44. Zhang, H.; Lin, Z.; Zhang, C. A counterexample for the validity of using nuclear norm as a convex surrogate of rank. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Prague, Czech Republic, 23–27 September 2013; pp. 226–241.
45. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **2011**, *3*, 1–122.
46. Lin, Z.; Liu, R.; Su, Z. Linearized Alternating Direction Method with Adaptive Penalty for Low-Rank Representation. In Proceedings of Advances in Neural Information Processing Systems Conference, Granada, Spain, 12–14 December 2011; pp. 612–620.
47. Larsen, R.M. *Lanczos Bidiagonalization With Partial Reorthogonalization*; Technical report, DAIMI PB-357; Department of Computer Science, Aarhus University: Aarhus, Denmark, 1998.
48. Zheng, Y.; Zhang, X.; Yang, S.; Jiao, L. Low-rank representation with local constraint for graph construction. *Neurocomputing* **2013**, *122*, 398–405.
49. Read, J.; Pfahringer, B.; Holmes, G.; Frank, E. Classifier chains for multi-label classification. *Mach. Learn.* **2011**, *85*, 333–359.
50. Tsoumakas, G.; Vlahavas, I. Random k-labelsets: An ensemble method for multilabel classification. In *Proceedings of the European Conference on Machine Learning, Warsaw, Poland, 17–21 September 2007*; Springer: Berlin, Germany, 2007; pp. 406–417.
51. Chung, F.R. *Spectral Graph Theory*; Published for the Conference Board of the Mathematical Sciences by the American Mathematical Society; AMS: Providence, RI, USA, 1997.
52. Wang, F.; Zhang, C. Label propagation through linear neighborhoods. *IEEE Trans. Knowl. Data Eng.* **2007**, *20*, 55–67.

53. Goncalves, E.C.; Plastino, A.; Freitas, A.A. A genetic algorithm for optimizing the label ordering in multi-label classifier chains. In Proceedings of the IEEE 25th International Conference on Tools with Artificial Intelligence, Washington, DC, USA, 4–6 November 2013; pp. 469–476.
54. Duygulu, P.; Barnard, K.; de Freitas, J.F.; Forsyth, D.A. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark, 28–31 May 2002; pp. 97–112.
55. Huiskes, M.J.; Lew, M.S. The MIR flickr retrieval evaluation. In Proceedings of the 1th ACM International Conference on Multimedia Information Retrieval, Vancouver, BC, Canada, 30–31 October 2008; pp. 39–43.
56. Von Ahn, L.; Dabbish, L. Labeling images with a computer game. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vienna, Austria, 24–29 April 2004; pp. 319–326.
57. Bucak, S.S.; Jin, R.; Jain, A.K. Multi-label learning with incomplete class assignments. In Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA, 20–25 June 2011; pp. 2801–2808.
58. Wang, J.G.; Li, J.; Yau, W.Y.; Sung, E. Boosting dense SIFT descriptors and shape contexts of face images for gender recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, San Francisco, CA, USA, 13–18 June 2010; pp. 96–102.
59. Grubinger, M.; Clough, P.; Müller, H.; Deselaers, T. The IAPR tc-12 Benchmark: A New Evaluation Resource for Visual Information Systems. In Proceedings of the International Workshop OntoImage'2006 Language Resources for Content-Based Image Retrieval, Genoa, Italy, 22 May 2006; pp. 10–20.
60. De Meo, P.; Quattrone, G.; Ursino, D. Exploitation of semantic relationships and hierarchical data structures to support a user in his annotation and browsing activities in folksonomies. *Inf. Syst.* **2009**, *34*, 511–535.
61. Cai, Y.; Li, Q.; Xie, H.; Min, H. Exploring personalized searches using tag-based user profiles and resource profiles in folksonomy. *Neural Netw.* **2014**, *58*, 98–110.
62. De Meo, P.; Quattrone, G.; Ursino, D. A query expansion and user profile enrichment approach to improve the performance of recommender systems operating on a folksonomy. *User Model. User Adapt. Interact.* **2010**, *20*, 41–86.
63. Kim, H.N.; Rawashdeh, M.; Alghamdi, A.; El Saddik, A. Folksonomy-based personalized search and ranking in social media services. *Inf. Syst.* **2012**, *37*, 61–76.
64. Kim, H.N.; Rocznia, A.; Lévy, P.; El Saddik, A. Social media filtering based on collaborative tagging in semantic space. *Multimed. Tools Appl.* **2012**, *56*, 63–89.



© 2017 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).