# *Superpixel-based Radial Basis Function Canonical Correlation Analysis for land cover segmentation*

International Journal of Remote Sensing

ISSN: 0143-1161 (Print) 1366-5901 (Online) Journal homepage: www.tandfonline.com/journals/tres20

# Superpixel-based radial basis function canonical correlation analysis for land cover segmentation

Xia Hong, Hong Wei & James Xiao

**To cite this article:** Xia Hong, Hong Wei & James Xiao (2025) Superpixel-based radial basis function canonical correlation analysis for land cover segmentation, International Journal of Remote Sensing, 46:24, 9411-9433, DOI: 10.1080/01431161.2025.2579806

**To link to this article:** https://doi.org/10.1080/01431161.2025.2579806



Published online: 01 Dec 2025.

INTERNATIONAL JOURNAL OF REMOTE SENSING
2025, VOL. 46, NO. 24, 9411–9433
https://doi.org/10.1080/01431161.2025.2579806

Taylor & Francis
Taylor & Francis Group



# Superpixel-based radial basis function canonical correlation analysis for land cover segmentation

Xia Hong[a], Hong Wei[a] and James Xiao[b]

[a]Department of Computer Science, University of Reading, Reading, UK; [b]Independent Researcher, Vancouver, Canada

**ABSTRACT**

The radial basis functions (RBF) network is popularly used in many machine learning applications. Semantic segmentation of remotely sensed image which can be addressed via a clustering task, is essential in land cover applications. In this study, a novel semi-supervised spectral clustering method, referred to as SLIC-RBF-CCA, is introduced for land cover multi-band image segmentation. The proposed SLIC-RBF-CCA approach is composed of two steps: i) a Simple Linear Iterative Clustering (SLIC) algorithm is initially applied to a set of pseudo-RGB images obtained from the singular vectors of a flattened matrix of a multi-band image; ii) a novel superpixel-based Radial Basis Function Canonical Correlation Analysis (RBF-CCA) generates canonical variables which are used to achieve final image segmentation. Specifically, a superpixel-based radial basis function is defined as the first variable in the framework of Canonical Correlation Analysis, in which the RBF centres are obtained as the local mean from superpixel regions. The second variable of SLIC-RBF-CCA is based on a few labelled pixels. The associated canonical variables, related to the pixels of a full image, are then applied by a *k*-means clustering algorithm. The proposed approach can be interpreted as an example of multi-view machine learning with attention mechanism. Finally, the effectiveness of the proposed algorithm has been validated using experiments on several remotely sensed multi-band images, including two patches from TopoSys GmbH and three patches from City of Potsdam from the ISPRS, showing excellent performance with segmentation accuracy $>85\%$.

**ARTICLE HISTORY**
Received 25 June 2025
Accepted 16 October 2025

**KEYWORDS**
Semi-supervised learning; superpixels; land cover segmentation; canonical correlation analysis; radial basis function

## 1. Introduction

Semantic segmentation of remotely sensed imagery is an essential task in understanding land cover (J. Wang, Ding, and He 2023), such as land use mapping (Taiwo et al. 2023) and precision agriculture (Kganyago et al. 2024). Remotely sensed images involve complex scenes with multiple categories and multiple data types, such as passively collected hyperspectral and multispectral images across a variety of spectral bands for different applications (Vivone 2023). Actively collected data from some remote sensing missions,

**CONTACT** Xia Hong x.hong@reading.ac.uk University of Reading - Department of Computer Science Polly Vacher Building Whiteknights Campus Reading, RG6 6AY UK

*e.g*. LIDAR (Fang et al. 2018), which provide topographical information can also be mapped as 2D images. J. Wang, Ding, and He (2023) employed deep learning models for landcover semantic segmentation, to which boundary enhancement loss functions were incorporated to improve objects boundary precision. However, it may be difficult to obtain a large number of pixel-wise dense image annotations needed for deep learning models for users.

Unsupervised, *e.g*. clustering-based, approaches for image segmentation have attracted wide research interest. Over the past decades, various clustering methods have been proposed, including centroid-based clustering (MacQueen et al. 1967; Zhang and Hepner 2017), density-based clustering (Ester et al. 1996), spectral clustering (Shi and Malik 2000), distribution-based clustering (Wang et al. 2023) and ensemble clustering (Golalipour et al. 2021). Spectral clustering (SC) algorithms are heavily exploited in graph-based image segmentation methods (Hagen and Kahng 1992; Ng, Jordan, and Weiss 2001; Pang et al. 2021; Shi and Malik 2000). By extending the concept of ratio cut Hagen and Kahng (1992) with multiview learning, Zhong and Pun (2023) introduced self-taught multi-view spectral clustering algorithms with excellent performance for various data sets, of which the maximum size is only ten thousands, far below pixel number of a typical image.

Purely unsupervised image segmentation often fails to achieve satisfactory performance due to no ground truth labels being used for training. In order to achieve the performance satisfactory to relevant applications, prior information is incorporated into clustering algorithms to impart prior knowledge into the cluster learning process, which is called semi-supervised clustering methods (Basu, Banerjee, and Mooney 2004; Cai et al. 2023; Ghasemi, Khorshidi, and Aickelin 2021). In some cases, extra information is provided in the clustering output space, *e.g*. through limited labelled samples, to enhance the clustering results. For example, Kim, Lee, and Lee (2013) adopted a semi-supervised learning strategy for developing an efficient algorithm for image segmentation by employing spectral clustering in constructing a sparse multilayer graph with nodes of pixels and over-segmented regions from the image. Alternatively, superpixels (Achanta et al. 2012; W. Wang, Liu, and Mou 2021; Ye et al. 2022) have been widely used as an effective way to reduce the number of image primitives for subsequent processing, such as image segmentation. Given a preset number of superpixels $Q$, the Simple Linear Iterative Clustering (SLIC) algorithm (Achanta et al. 2012) employs $k$-means clustering to group nearby pixels into superpixels based on a 5-dimensional positional and CIELAB color features by relying on $k$-means applied with an appropriate distance mixing spaces of spatial and colour proximity. The SLIC algorithm and its variants are effective (Achanta et al. 2012; Han, Han, and Cheng 2022) in determining an over-segmented image into $Q$ superpixels by labelling $Q$ local regions within which the pixels have similar colours. Consequently, SLIC has very good edge detection ability, and boundaries between distinctive colours often coincide as superpixels edges.

Image segmentation via spectral clustering has become increasingly reliant on superpixels as a pre-processing step. Pont-Tuset et al. (2017) developed a fast normalized cuts algorithm as the initial step to group pixels. They then proposed a multiscale combinatorial grouping algorithm to perform a hierarchical search for similar regions across different scales. In their research, scaled images were used to

generate ultrametric contour maps in the segmentation process. Li, You, and Vidal (2017) proposed structured sparse subspace clustering, in which each data point is expressed as a structured sparse linear combination of all other data points. Then, a joint optimization framework was applied for learning both the affinity learned from the data and the segmentation implemented by spectral clustering. The concept of multi-view spectral clustering was applied by Luo et al. (2019) to deal with a NP-hard optimization problem which occurs in conventional spectral clustering approaches. The developed framework was evaluated for image segmentation along with other applications. Jiao, Chen, and Dong (2020) applied spectral clustering to high-level features, which were outputs from an autoencoder network with its input a set of low-level image features over pre-generated non-overlapping superpixels. In their work, the low-level image features include colour, gradient, local binary pattern, and saliency features. With the unsupervised nature, the evaluation on two data sets showed IoU (Intersection over Union) about 0.6. In addition, spectral clustering was used to generate superpixels as the fundamental unit of images in image segmentation to capture the local data redundancy and to assign each pixel a label in a conservative way by minimizing the risk of merging unrelated pixels (J. Chen, Li, and Huang 2017). Superpixels were used as a weakly supervised semantic image segmentation to guide local and global consistency in dual deep neural networks (Yi et al. 2022). For remote sensing image segmentation, Z. Chen and Wang (2015) attempted using semi-supervised approach to build a spectral-spatial classifier with affinity scoring to segment hyperspectral imagery. Jiang et al. (2022) developed a semi-supervised segmentation method, in which superpixels are generated to construct the graph in a graph convolutional neural network for sea ice classification from remotely sensed synthetic aperture radar (SAR) imagery.

The radial basis functions (RBF) are a cornerstone in approximation theory (Poggio and Girosi 1990) which has been popularly used in many machine learning algorithms such as Support Vector Machines, Artificial Neural Networks (S. Chen, Hong, and Harris 2011) and image processing (Pal et al. 2021). Alternatively, the canonical correlation analysis (CCA) was originally proposed by H. Hotelling in the seminal works (Hotelling 1935, 1936). The subject of this work is on finding the best predictors among the linear functions from each set by maximizing the correlation coefficient between two sets. CCA can simplify the statistical analysis for two sets of variables and properly solve the aforementioned problem, by defining a sequence of pairs of variates as canonical variates and the correlations between them as canonical correlations. The complicated high-dimensional variable sets are projected in the common latent subspace with a desired low dimension. In an attempt to increase the flexibility of CCA for nonlinear relationships between two random variables, kernel CCA (KCCA) has been researched (Akaho 2001). CCA/KCCA was applied for effective feature selection tool as a pre-processing step for classifiers of support vector machine or random forests (Y. Wang, Cang, and Yu 2019). As a powerful tool for multimodal feature fusion, CCA/KCCA has received widespread attention in both theoretical advances and remote sensing applications (Amrani, Bey, and Amamra 2022; Neeti 2014; Yang et al. 2021).

Against this background, we propose a novel semi-supervised segmenting algorithm, referred to as SLIC-RBF-CCA, which employs SLIC algorithm as a pre-

processing stage, in which superpixels are obtained from a multi-band land cover image. A novel RBF is defined in a pixel feature space using superpixels to generate its centres. Then the RBF and a small set of known labels are used in CCA framework to define two sets of random variables, which are in turn used to produce SLIC-RBF-CCA covariates over a full image for final segmentation. The CCA framework was also applied in our previous work (Wei et al. 2024), with significant differences. In (Wei et al. 2024), the first set of random variable in CCA is based on t-SNE dimensional reduction of multiple image processing features, rather than SLIC-RBF. The proposed algorithm is different from Zhong and Pun (2023) since our work is not derived from ratio cut (Hagen and Kahng 1992) cost function, but is based on CCA framework. Our work is also different from (J. Wang, Ding, and He 2023), since (J. Wang, Ding, and He 2023) is not a clustering approach.

Our contributions to knowledge are listed below:

(1) A novel systematic framework is introduced based on adapting SLIC as a pre-processing tool for remotely sensed land cover imagery to obtain superpixels for pseudo-RGB images based on multi-band grayscaled imagery, which are then segmented using a novel SLIC-RBF-CCA algorithm only with a small amount of labelled data.
(2) To reap the benefits of SLIC algorithm which can efficiently identify superpixel groups with good edge detection performance, a novel RBF affinity function is defined using multi-band features representation centre relating to superpixel locations, which forms the first set of variables of RBF-CCA using the complete input image pixels. A small set of labels are used to form the second set of variables in CCA.
(3) Compared with a supervised deep learning model using a large amount of labelled pixels, the proposed SLIC-RBF-CCA algorithm obtains a projection to canonical variables using CCA, to which *k*-means clustering is applied similar to a spectral clustering algorithm (Luxburg 2007). It is summarized that the proposed algorithm can be interpreted as a multi-view learning approach (Sun 2013) with attention mechanism (Bahdanau, Cho, and Bengio 2015).

The rest of the paper is organized as follows: Section 2 introduces preliminary of both canonical correlation analysis and SLIC algorithm. Section 3 introduces the proposed semi-supervised segmentation approach based on SLIC-RBF-CCA algorithm, followed by discussions. In Section 4, experiments and results are presented and discussed. Finally, Section 5 is devoted to conclusions.

## 2. Preliminary

For ease of exposition, the mathematical background of canonical correlation analysis (Hardle and Simar 2007) is briefly outlined in Section 2.1, and the method of Simple Linear Iterative Clustering (SLIC) (Achanta et al. 2012) is summarized in Section 2.2.

### 2.1. Canonical correlation analysis

Canonical correlation analysis (CCA) (Hotelling 1935, 1936) is a way of measuring the linear relationship between two sets of multidimensional variables. Consider two sets of random variables $\boldsymbol{\Phi} = \{\phi i\}, i = 1, \ldots, d_1$ and $\Psi = \{\mathbf{y}_i\}, i = 1, \ldots, d_2$ with zero mean. It is assumed that $\Phi$, $\Psi$ are full rank, and $d = \min(rank(\Phi), rank(\Psi))$.

Denote the total Covariance matrix as

$$E\left[\begin{pmatrix}\Phi\\ \psi\end{pmatrix}\begin{pmatrix}\Phi\\ \psi\end{pmatrix}^{\mathrm{T}}\right] = \begin{bmatrix}C_{\Phi\Phi} & C_{\Phi\psi}\\ C_{\psi\Phi} & C_{\psi\psi}\end{bmatrix} \tag{1}$$

Define a set of two projection matrices $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d] \in \Re^{d_1 \times d}$ and $\boldsymbol{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d] \in \Re^{d_2 \times d}$, which generate a set of linear combinations named $\boldsymbol{U} = \Phi\boldsymbol{A} = [\boldsymbol{u}_1, \ldots \boldsymbol{u}_d]$ and $\boldsymbol{V} = \Psi\boldsymbol{B} = [\boldsymbol{v}_1, \ldots \boldsymbol{v}_d]$. Each member of $\boldsymbol{U}$ is paired with a member of $\boldsymbol{V}$, as a set of canonical variables pairs $(\boldsymbol{u}_i, \boldsymbol{v}_i)$.

The task in CCA is to find $\boldsymbol{A}, \boldsymbol{B}$ such that the correlations $\rho_i(\boldsymbol{u}_i, \boldsymbol{v}_i)$ are maximized. Represent

$$\rho_i = \frac{\boldsymbol{a}_i^T C_{\Phi\psi} \boldsymbol{b}_i}{\sqrt{\boldsymbol{a}_i^T C_{\Phi\Phi} \boldsymbol{a}_i}\sqrt{\boldsymbol{b}_i^T C_{\psi\psi} \boldsymbol{b}_i}}, \quad i = 1, \ldots, d \tag{2}$$

Equivalently,

$$\max_{\boldsymbol{a}_i, \boldsymbol{b}_i} \rho_i = \boldsymbol{a}_i^T C_{\Phi\psi} \boldsymbol{b}_i, \forall i \tag{3}$$

subject to $\boldsymbol{a}_i^T C_{\Phi\Phi} \boldsymbol{a}_i = 1$, $\boldsymbol{b}_i^T C_{\psi\psi} \boldsymbol{b}_i = 1$.

To obtain the CCA solution, initially defineHardle2007

$$K = C_{\Phi\Phi}^{-1/2} C_{\Phi\psi} C_{\psi\psi}^{-1/2} \tag{4}$$

and perform singular value decomposition (SVD) of $K$ as

$$K = \Gamma \Lambda \Delta^{\mathrm{T}} \tag{5}$$

with $\Gamma = [\gamma_1, \ldots \gamma_{\mathrm{d}}]$, $\Delta = [\delta_1, \ldots, \delta_{\mathrm{d}}]$. $\Lambda = \mathrm{diag}\{\lambda_1^{1/2}, \ldots \lambda_{\mathrm{d}}^{1/2}\}$. and $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_d$ are the nonzero eigenvalues of $K^T K$. $\gamma_i$ and $\delta_i$ are the left and right eigenvectors of $K$.

Now define

$$\boldsymbol{a}_i = C_{\Phi\Phi}^{-1/2} \gamma_i$$

$$\boldsymbol{b}_i = C_{\psi\psi}^{-1/2} \delta_i \tag{6}$$

so that

$$Cov(\boldsymbol{u}_i, \boldsymbol{u}_j) = \boldsymbol{a}_i^T C_{\Phi\Phi} \boldsymbol{a}_j = \gamma_i^T \gamma_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{7}$$

$$Cov(\boldsymbol{v}_i, \boldsymbol{v}_j) = \boldsymbol{b}_i^T C_{\psi\psi} \boldsymbol{b}_j = \delta_i^T \delta_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{8}$$

and the correlation between $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ has maximal of

$$\rho(\boldsymbol{u}_i, \boldsymbol{v}_i) = \gamma_i^T \Gamma \Lambda \Delta^{\mathrm{T}} \delta_{\mathrm{i}} = \lambda_{\mathrm{i}}^{1/2} \tag{9}$$

### 2.2. Simple Linear Iterative Clustering (SLIC) (Achanta et al. 2012)

A superpixel is a group of pixels with similar colours that are spatially close to each other in a coloured image. The SLIC algorithm (Achanta et al. 2012) segments an image into superpixel regions. Computer vision applications have become increasingly reliant on superpixels which groups pixels into local regions as a pre-processing step to reduce computational costs. Given an RGB image represented in the CIELAB space (where RGB is converted to *l, a, b*) by $\mathbf{I} \in \Re^{W \times H \times 3}$, where $W, H$ denote the width, height of the image. Let $\mathbf{I}(w, h)$ be the LAB vector corresponding to the pixel at width and height position of $(w, h)$, with $w \in [1..,W]$, $h \in [1..,H]$. The SLIC algorithm is initialized by putting initial superpixel positions on a regular grid inside the image, according to a desired number of superpixels $Q$ chosen by the user. Each approximately equally-sized superpixel is therefore $N/Q$ pixels, where $N = W \times H$ is the number of pixels of an image. Define the grid interval $S = \sqrt{N/Q}$. At the onset of SLIC algorithm, there are $Q$ superpixel cluster centers $G_q = [l_q, a_q, b_q, w_q, h_q]^T$ with $q = 1, \ldots, Q$ with equal spacing $S$, which are moved to seed locations corresponding to the lowest gradient position in a $3 \times 3$ neighbourhood to avoid edges. Image gradients are computed as:

$$g(w, h) =$$

$$\| \mathbf{I}(w+1, h) - \mathbf{I}(w-1, h)\|^2 + \| \mathbf{I}(w, h+1) - \mathbf{I}(w, h-1)\|^2 \tag{10}$$

$\| \cdot \|$ is Euclidean norm. The SLIC pseudocode is given in Algorithm 1, in which $Q$ superpixel cluster centers $G_q = [l_q, a_q, b_q, w_q, h_q]^T$ are adjusted by a locally adapted algorithm using a distance measure $D_s$ which is defined as follows:

$$d_{lab} = \sqrt{(l_q - l_i)^2 + (a_q - a_i)^2 + (b_q - b_i)^2}$$

$$d_{w,h} = \sqrt{(w_q - w_i)^2 + (h_q - h_i)^2}$$

$$D_s = d_{lab} + \frac{\alpha}{S} d_{w,h} \tag{11}$$

Minimizing $D_s$ offers a good balance between colour similarity and spatial proximity. The greater the value of $\alpha$, the more spatial proximity is emphasized and the more compact the cluster. This value can be in the range [1, 20]. Based on $D_s$, it is safely assumed that pixels that are associated with this cluster center lie within a $2S \times 2S$ area around the superpixel center on the $w - h$ plane, which is used as the search area for the pixels nearest to each cluster centre in Algorithm 1. Note that line 9 "Enforces connectivity'' refers to a post-processing step of relabelling disjoint segments with the labels of the largest neighbouring cluster.

**Algorithm 1.** SLIC pseudo code.

**Require:** An RGB image represented in CIELAB space (lab) as $\mathbf{I} \in \Re^{W \times H \times 3}$. A desired superpixels number $Q$.
1: Initialize cluster centers $G_q = [l_q, a_q, b_q, w_q, h_q]^T$ by sampling pixels at regular grid steps $S$, $q = 1, ..., Q$.
2: Perturb cluster centers in a $3 \times 3$ neighbourhood, to the lowest gradient position
3: **repeat**
4: **for** each cluster center $G_q$ **do**
5: Assign the best matching pixels from a $2S \times 2S$ square neighbourhood around the cluster center according to the distance measure (Eq. 11).
6: **end for**
7: Compute new cluster centers and residual error $E$ (Distance between previous centers and recomputed centers)
8: **until** $E \leq$ threshold
9: Enforce connectivity.
10: **Return**: Labelling matrix $L \in \Re^{W \times H}$ assigning each pixel in $\mathbf{I}$ to $q \in [1, ..Q]$.

Due to the clustering criterion of SLIC, it is known that neighbouring SLIC superpixels may coincide well with edges/boundary within an image. Clearly, pixels with similar colours at different spatial regions are allocated to different superpixels, thus SLIC produces superpixels which over segment an image without any semantic inference.

## 3. Method

Figure 1 outlines the system framework. Given a multi-band land cover image, the proposed method starts with using singular value decomposition (SVD) to create pseudo-RGB images onto which SLIC is applied, as detailed in Section 3.1. Our proposed semi-supervised clustering method of SLIC-RBF-CCA algorithm is based on a novel radial basis function using SLIC superpixels' regions. The idea of constructing radial basis function on superpixels is explained in Section 3.2. We propose our semi-supervised segmentation approach of SLIC-RBF-CCA in Section 3.3.

### 3.1. Using SLIC for multi-band remote sensing image

Given a grey scaled multi-band remote sensing image $I^{full} \in \Re^{W \times H \times M}$, where $M$ is the number of bands in $I^{full}$. Each gray scaled image is denoted as $\boldsymbol{I}_m$, $m = 1, \ldots, M$ and each pixel in band $m$ is denoted as $\boldsymbol{I}_m(w, h)$. For remote sensing applications, additionally derived bands may be created based on physical properties. For example, a *NDVI* (Normalized Difference Vegetation Index) is computed as the difference between near-infrared (IR) and red (RED) reflectance divided element-wise by their sum:


**Figure 1.** Overview of the proposed SLIC-RBF-CCA approach for semi-supervised land cover image segmentation.

$$NDVI = \frac{IR - R}{IR + R} \tag{12}$$

where $IR$ and $R$ denote near-infrared and red bands, respectively. In this study, we used the NDVI as an additional band specifically contributing to distinct vegetation from other in land cover classification. It is an essential feature for measuring greenness exposure (de la Iglesia Martinez and Labib 2023) in remote sensing imagery. Alternatively, we generate a grayscale image via

$$GRAY = 0.299R + 0.587G + 0.114B \tag{13}$$

from which an additional band is created by scaling it into eight levels, referred to as scaled grey image (SGI). The SGI is often associated with Gray-Level Co-occurrence Matrix (GLCM) calculation (Haralick, Shanmugam, and Dinstein 1973), so it may help further segmentation. LIDAR data are also employed in the feature space where available, with its first echo (FE), last echo (LE), and intensity, in which LIDAR FE and LE correspond to the first and last points from where the laser beam is reflected, and LIDAR intensity represents the laser pulse amplitude of the LIDAR FE. When the surface height information of a Digital Surface Model (DSM) is given, it is also used as a band in the feature space as they add information which may help to segment similar objects in different image areas such as buildings as the same cluster.

The goal of SLIC is to partition these co-registered pixels into disjoint sets as superpixels. Initially, a set of pseudo-RGB images is created from the multi-band remote sensing image so that the SLIC algorithm is applicable to each pseudo-RGB image, as follows. Each $\boldsymbol{I}_m$ is vectorized to $\tilde{I}_m \in \Re^N$. For the full multi-band remote sensing image $I^{full}$, we have

$$X = \left[\tilde{I}_1, \tilde{I}_2, \ldots, \tilde{I}_M\right] \in \Re^{N \times M} \tag{14}$$

The mean removed $\boldsymbol{X}$ undergoes singular vector decomposition (SVD), and the first left singular vectors are retrieved. Each singular vector is then multiplied by the sign of the entry with maximum absolute value. Then these are normalized to the range of [0,1], followed by being reshaped into a set of pseudo-RGB images in the same dimension of $\boldsymbol{I}_m$. The rationale of the above operations is to ensure that each of the singular vectors are ready to be converted to a grayscale image for use by SLIC, and to ensure the largest absolute value singular vector becomes the leading value in respective band of the pseudo-RGB image. Each pseudo-RGB image is from $N \times 3$ matrix of a set of three singular vectors. The resulting $p$ th pseudo-RGB are denoted as $\mathbf{I}^p \in \Re^{W \times H \times 3}$, to which SLIC is applied to obtain the labelling matrices $L^{(p)}$, with $L^{(p)}(w,h) \in [1,\ldots,Q]$. In this study, we used two pseudo-RGB images from 1st-3rd and then 4-6th singular vectors, respectively with $p = 1, 2$. To provide a visual illustration, two pseudo-RGB images are shown in Figure 2 by taking a data set from Section 5 as an example. It is visible that two SLIC images provide complementary segmentation viewpoints to the original multiband image. Note that from SLIC algorithm, two sets of superpixel labeling matrices relating to a co-registered image are produced to store the SLIC

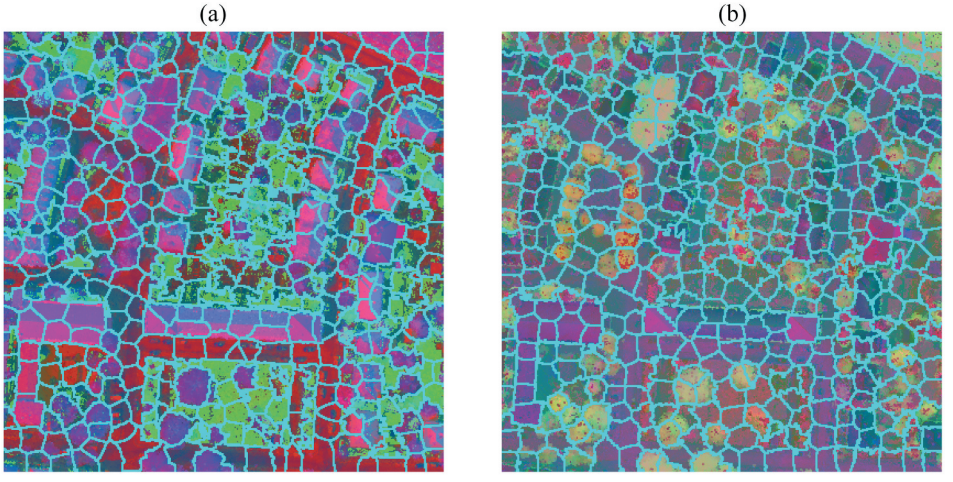


**Figure 2.** Two pseudo-RGB images and SLIC superpixel boundary are generated using the first data set TopoSys GmbH RS (see Section 5), with $Q = 400$. (a) $\mathbf{I}^1$ and (b) $\mathbf{I}^2$.

boundaries, which impart different viewpoints into our superpixle-based RBF, as proposed next.

### 3.2. Radial basis functions on superpixels

For ease of exposition, each row of $\boldsymbol{X}$ is defined as $\boldsymbol{x}_i = [x_{i,1}, \ldots, x_{i,M}]^{\mathsf{T}} \in \Re^M$ which forms a feature vector of a pixel in co-registered bands, so that $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]^{\mathsf{T}}$. The vectorized co-registered multi-band images, together with SLIC algorithm output labelling matrices are denoted as data matrix $D = \{\boldsymbol{x}_i, L_i^{(1)}, L_i^{(2)}\}_{i=1}^N$, where $L_i^{(p)} \in [1, \ldots, Q]$ is vectorized from labelling matrix $L^{(p)}$. In the sequel, we preset $p = 1, 2$ for simplicity (two pseudo-RGB images).

The radial basis functions (RBF) are a cornerstone in approximation theory Girosi 1990). The aim of using RBF in our proposed approach is to introduce a pairwise similarity function between a pixel and a superpixel in multiple pseudo-RGB images to achieve semantic segmentation of the original image. This helps to semantically link distant superpixels in a land cover image by feature representation association. It is noted that provided $Q$ is large enough, the pixels within each superpixel group will be close in terms of original multi-band features space due to continuity. To represent the features of each superpixel in a pseudo-RGB image, we define

$$\boldsymbol{c}_{j+Q(p-1)} = \frac{1}{N_j^{(p)}} \sum_{L_i^{(p)}=j} \boldsymbol{x}_i, \quad j = 1, \ldots, Q \tag{15}$$

for $p = 1, 2$, where $N_j^{(p)}$ is the number of pixels within the $j$ th superpixel group in the $p$ th pseudo-RGB image. The radial basis function (RBF) is given as:

$$RBF(\boldsymbol{x}, \boldsymbol{c}_j) = \exp\left(-(\boldsymbol{x} - \boldsymbol{c}_j)^2/(2\sigma^2)\right) \tag{16}$$

where $\boldsymbol{c}_j$, $j$, are the centers of RBFs. $\sigma > 0$ is an RBF width parameter which an adjustable hyperparameter adjustable by user. In this work, we simply set as average pairwise distance between pixel $\boldsymbol{x}_i$ and $\boldsymbol{c}_j$, defined as

$$\sigma = \frac{1}{2NQ} \sum_i \sum_j \| \boldsymbol{x}_i - \boldsymbol{c}_j \| \tag{17}$$

This will make $\sigma$ adaptive to data and the algorithm automatic. Note that the approximation capability of RBF is robust to a wide range of $\sigma$.

The output of a radial basis function reduces as the distance between the input and its fixed centre $\boldsymbol{c}_j$ increases, giving its ability to discriminate superpixels in the multi-band input feature space, since pixels in other superpixels may have similar/dissimilar features to their own superpixel group. The derivation of $\boldsymbol{c}_j$ in (15) can be explained in relation to superpixel regions in Figure 2, each $\boldsymbol{c}_j$ is obtained by averaging multi-band features subject to its association to a specified local region, which is obtained by applying the SLIC algorithm to multiple pseudo-RGB images. By doing so, a large number of feature representations is created using local statistics generated by SLIC pixels which may be closely related to semantic objects and useful for enhancing final segmentation performance.

### 3.3. Proposed semi-supervised segmentation based on SLIC-RBF-CCA

The SLIC algorithm itself acts as over-segmentation with clustering local region only, the aim of proposed semi-supervised algorithm based on SLIC-RBF-CCA is to help link pixels in distant SLIC superpixel groups to the same semantic labels with feature similarity via semi-supervised clustering. The two sets of random variables $\Phi$ and $\Psi$ are defined in our proposed canonical correlation analysis (SLIC-RBF-CCA) as follows.

Initially, over $\boldsymbol{X}$, let

$$\phi_{i,j} = \frac{RBF(\boldsymbol{x}_i, \boldsymbol{c}_j)}{\sum_{j=1}^{2Q} RBF(\boldsymbol{x}_i, \boldsymbol{c}_j)} \tag{18}$$

and construct a matrix

$$\Phi_{full} = \begin{bmatrix} \phi_{1,1} & \cdots & \phi_{1,2Q} \\ \vdots & \ddots & \vdots \\ \phi_{N,1} & \cdots & \phi_{N,2Q} \end{bmatrix} \tag{19}$$

followed by mean removal as

$$\Phi_{full} \leftarrow \Phi_{full} - \text{mean}(\Phi_{full}). \tag{20}$$

By examining (18), we can see that this creates a pairwise relationship between all possible multi-band pixel feature vectors to all possible superpixels (whose location were detected by the pre-processing SLIC algorithm). Dissimilar pixels in

the feature space will yield a near zero response. Similar pixels in the feature space but in another superpixel group will generate a higher response. The pixels' RBF will be excited within their own superpixel group, as well as other superpixel groups if the respective mean superpixel features are close to each other.

Given a subset of $N_s$ labelled data points, whose indices vector is denoted as *Index*, each with known cluster $C(Index(i)) \in [1..,K], i = 1,\ldots,N_s$, we define

$$\Phi = \Phi_{\text{full}}[\text{Index},:] \in \Re^{N_s \times 2Q} \tag{21}$$

Over the set of labelled data points *Index*, the second variable of RBF-CCA $\Psi \in \Re^{N_s \times K}$ is defined as

$$\Psi = \begin{bmatrix} \psi_{1,1} & \cdots & \psi_{1,K} \\ \vdots & \ddots & \vdots \\ \psi_{N_s,1} & \cdots & \psi_{N_s,K} \end{bmatrix} \tag{22}$$

where

$$\psi_{i,j} = \begin{cases} 1 & \text{If } C(Index(i)) = j \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

Then $\Psi$ is followed by mean removal, as

$$\Psi \leftarrow \Psi - \text{mean}(\Psi). \tag{24}$$

which simply set pixels within the same superpixel to excite the same column of $\Psi$, hence to focus on the pixel's spatial closeness information of the SLIC algorithm.

The proposed semi-supervised segmentation of SLIC-RBF-CCA is summarized in Algorithm 2. Algorithm 3 summarizes the steps of obtaining SLIC-RBF-CCA canonical variables (see Section 2.1) based on two sets of random vectors as defined above. Note that Line 4 of Algorithm 3 returns the canonical variable corresponding to the full input features. Line 6–9 of Algorithm 3 is the k-means clustering step applied to normalized canonical variables, which in CCA are obtained using SVD algorithm related to mathematic spectral theory. Similarly the k-means clustering clustering is applied to eigenvectors of a Laplacian matrix in spectral clustering approach (Luxburg 2007). Hence the proposed approach can be viewed as a generalized spectral clustering approach.

It is worth noting that the proposed algorithm is simple to use, as there are only two preset parameters: numbers of superpixels $Q$ and number of final clusters $K$. Note that the number of super-pixels $Q$ is user predefined hyperparameter as in SLIC (Achanta et al. 2012) with the aims to achieve improve computational efficiency with a smaller $Q$, as well as the ability to identify object boundary. Note that if $Q$ is too small, SLIC will fail to identify object boundary for subsequent image processing tasks, hence for practitioners, it is suggested that intermediate semantical analysis to the outcome of STEP 1 in Algorithm 2, based on image size and image scene, could be carried to in order to decide an appropriate sufficiently small $Q$. Note that there should be no supervised optimization of $Q$ since the approach is a clustering algorithm.

**Algorithm 2.** Proposed semi-supervised segmentation of SLIC-RBF-CC.

**Require:** A multi-band remote sensing image $I^{full}$. Number of superpixels $Q$. Number of final clusters $K$ to construct.

**STEP 1** : Construct multiple pseudo-RGB images. $\mathbf{I}^{(p)} \in \Re^{W\times H\times 3}$ as described in Section 3.1.

Apply Algorithm 1 to the $p$th pseudo-RGB image $\mathbf{I}^{(p)}$ and return $L^{(p)}(w,h)$, $\forall p$.

From $I^{full}$ and $L^{(p)}(w,h)$ to form $D = \{\boldsymbol{x}_i, L_i^{(p)}\}_{i=1}^{N}$.

**STEP 2** : Call Algorithm 3 to obtain $\boldsymbol{U}$.

Form the matrix $\boldsymbol{Z} = \{z_{i,j}\} \in \Re^{N\times K}$ by normalising the rows of top $K$ columns of $\boldsymbol{U}$, *i.e.* to set

$$z_{i,j} = u_{i,j} \Big/ \sqrt{\sum_{j=1}^{K} u_{i,j}^2} \tag{25}$$

**for** $i = 1, ..., N$ **do**

  Let $\hat{\boldsymbol{z}}_i \in \Re^K$ be the vector corresponding to the $i$th row of $\boldsymbol{Z}$.

**end for**

Cluster the points $\hat{\boldsymbol{z}}_i$, $i = 1, ..., N$ with the $k$-means algorithm (Haykin 2009) into clusters $C_1, ..., C_K$.

**Return:** Find clusters $k \in \{1, ..., K\}$ with $\{k, \hat{\boldsymbol{z}}_i \in C_k\}$ and assign original data points $\boldsymbol{x}_i$ according to clusters' index set of $k = 1, ..., K$.

**Algorithm 3.** Constructing SLIC-RBF-CCA canonical variables.

**Require:** Vectorized multi-band image pixels and superpixel label data $D = \{\boldsymbol{x}_i, L_i^{(p)}\}_{i=1}^{N}$.

Construct zero mean $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ using (15) -(24).

Perform CCA to obtain $\boldsymbol{A} \in \Re^{2Q\times K}$.

Calculate $\boldsymbol{U} = \boldsymbol{\Phi}_{full}\boldsymbol{A} \in \Re^{N\times K}$.

Return $\boldsymbol{U}$.

The computing costs of Step 1 consist of $O(N^2)$ of generating pseudo-RGB images plus $O(N)$ of SLIC algorithm. The computing costs of Step 2 consist of $O(NQ)$ for generating RBFs, plus $O(N_s^2 Q)$ of CCA algorithm, and plus $O(N)$ of final $K$-means clustering step. Putting together, it can be concluded that the algorithm's costs are in the order of $O(N^2)$, since $Q \ll N$, which makes the proposed approach a highly efficient spectral clustering algorithm.

### 3.4. Discussion

Two sets of random variables, CCA is regarded as one of the earliest multi-view learning approaches, a general concept in solving machine learning problems (Sun 2013), of which the aims are to learn from data represented by multiple distinct feature sets. The proposed algorithm's relationship with the multi-view learning is twofold: Firstly, it learns from one view which are the co-registered pixel based RBF features and also from another view which are their respective semantic labels via a small set of labelled points. Secondly, the use of multiple SLIC pseudo-RGB images can also be interpreted as multi-view learning to create many RBFs from the set of singular vectors for revealing multi facets information to the image. Clearly, variants of multiple SLIC pseudo-RGB images may be employed.

Alternatively, it is worth noting that the proposed SLIC-RBF-CCA algorithm is related to the attention mechanism originating in machine translation (Bahdanau, Cho, and Bengio 2015). Fundamental to the attention mechanism is the concept of the triplet {Query, Key, Value} introduced in the transformer (Vaswani et al. 2017), which is regarded as a main contributor to the modern artificial intelligence since the transformer approach has become the main architecture of large language models. It can be seen that (18) has used attention mechanism in the mathematical form of the softmax function, in that $\boldsymbol{x}_i$ can map onto a query and $\boldsymbol{c}_j$ onto a key. $\phi_{i,j}$ of (18) relates to attention Score of $\boldsymbol{x}_i$ to each RBF center $\boldsymbol{c}_j$. All attention scores sum up to one. Moreover, the final covariate $\boldsymbol{u}_i$ can be interpreted as a Value in the self-attention mechanism, since CCA operates in a linear projection space, so it is a linear combiner of attention score. Therefore, the final spectral segmentation step can be interpreted as clustering of a set of attention Values, which have been mapped from a set of attention Query using attention mechanism.

## 4. Experiments

Two sets of experiments are conducted using multi-band remotely sensed land cover images of urban areas, consisting of buildings, vegetation, trees, and other surface areas. The RGB view of all data sets can be visualized in Figures 4(a), 6(a) and 7(a)–9(a)More details on additional bands are explained in each example. The respective hand-labelled ground truth is used for evaluation purposes. In semi-supervised comparative methods, randomly selected five percent of data along with their ground truth ($N_s = 5\%N$) are used for training. Several unsupervised and semi-supervised approaches are compared. To obtain the final performance metrics, the predicted cluster labels are mapped into the given ground truth via the well known Kuhn-Munkres algorithm (Lovász and Plummer 2009). The first baseline is the k-means algorithm, applied to the original multi-bands data. Based on a single pseudo-RGB as described in Section 3.1, the *k*-means and spectral clustering (SC) (Ng, Jordan, and Weiss 2001) are also compared, referred to as *k*-means-SLIC and SC-SLIC respectively, in which the final pixel segmentation results are obtained from that of the associated SLIC superpixels. Additionally, two semi-supervised approaches are devised to compare with proposed SLIC-RBF-CCA. The main differences are that $\Phi^{full}$ is formed differently. Specifically:

(1) Linear CCA algorithm: The original multi-bands are used without any nonlinear transformation to form $\Phi^{full}$.
(2) Polynomial CCA algorithm: The original multi-bands together with their quadratic terms are applied to form $\Phi^{full}$.

For comparison of three semi-supervised clustering approaches, we used randomly selected five percent data to pair with of the ground truth ($N_s = 5\%N$) for training. Twenty realizations are repeatedly experimented, with the results reported as mean and standard deviation.

Example 1:Two data sets obtained from Trimble Holdings GmbH (formerly TopoSys GmbH, Germany) are experimented. Each data set contains six bands, which are LIDAR first echo FE, last echo LE, colour image red (R), green (G), and blue (B) and LIDAR

intensity(IN). Two additional bands of NDVI and SGI were generated, so eight bands are used for each data set, $M = 8$.

• *The first data set* has a size of 400 × 400 pixels per band, with six measured data plus two derived bands NDVI, SGI, as shown in Figure 3, so $I^{full} \in \Re^{400\times400\times8}$, resulting in $N = 160000$. The RGB view of the image is shown in Figure 4(a). The ground truth of this image contains four classes (1 – Building; 2 – Tree; 3 – Low vegetation; and 4 – Impervious surface or Car) that were manually labelled pixel-wise are used for validation. The proposed SLIC-RBF-CCA algorithm is compared against several unsupervised and semi-supervised clustering approaches with their segmentation accuracy shown in Table 1. Weexperimented $Q = 100, 200, 400, 800$ for SLIC algorithm which are applicable to SLIC algorithms used in the experiments. It can be seen that the proposed SLIC-RBF-CCA algorithm outperforms other methods. The second best is Polynomial CCA, which does not use the SLIC algorithm. Over all values of $Q$, the accuracy performance is insensitive to $Q$ within the 1% differences, but $Q = 200, 400$ are the best. The ground

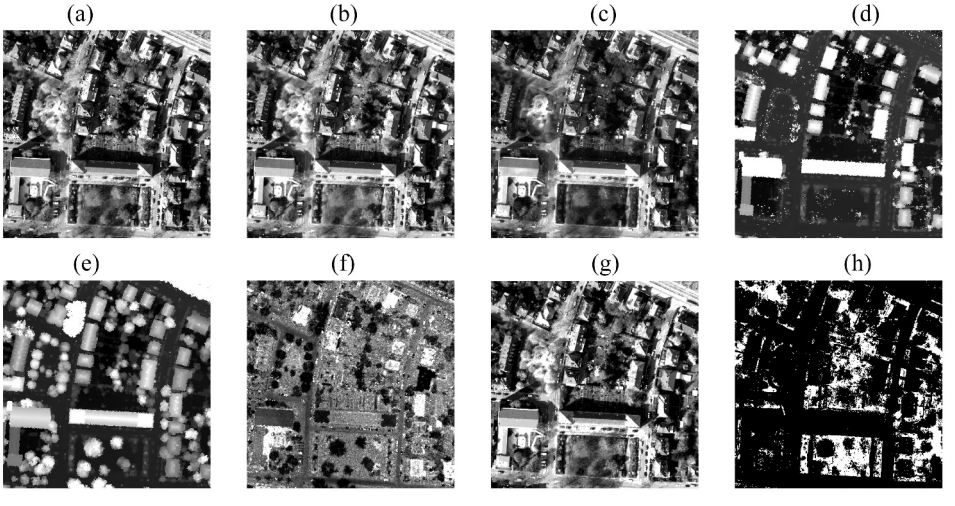


**Figure 3.** Eight multispectral images of TopoSys GmbH (the first data set, example 1) (a) red; (b) green; (c) Blue; (d) last echo; (e) first echo; (f) LIDAR intensity; (g) SGI; (h) NDVI.

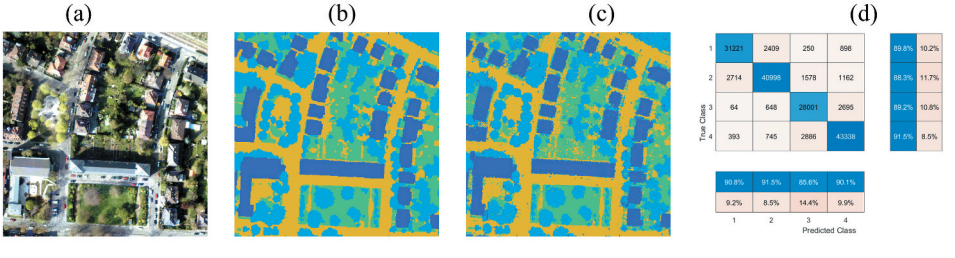


**Figure 4.** Segmentation results of TopoSys GmbH using $Q = 400$ superpixels (the first data set, example 1): (a) RGB view (b) ground truth; (c) segmentation; and (d) confusion matrix (1- Building; 2 -Tree; 3 - low vegetation; and 4- Impervious surface OR Car).

**Table 1.** Example 1: comparison of segmentation accuracy (percentage of correctly identified pixel labels). For semi-supervised approaches, the results are given as *mean* $\pm$ *std* over 20 repetitive random sampled labelled data sets.

| Data set | $Q$ | $k$-means | $k$-means-SLIC | SC- SLIC | Linear CCA | Polynomial CCA | Proposed SLIC-RBF-CCA |
|---|---|---|---|---|---|---|---|
| | – | 48.60 | – | – | $76.83 \pm 1.40$ | $87.82 \pm 0.08$ | – |
| First | 100 | – | 47.77 | 47.89 | – | – | $89.51 \pm 0.14$ |
| data | 200 | – | 49.81 | 52.77 | – | – | $89.77 \pm 0.08$ |
| set | 400 | – | 52.06 | 47.10 | – | – | $89.75 \pm 0.12$ |
| | 800 | – | 46.65 | 47.47 | – | – | $88.49 \pm 0.15$ |
| | – | 50.53 | – | – | $75.22 \pm 0.50$ | $81.29 \pm 0.51$ | – |
| Second | 50 | – | 52.54 | 50.61 | – | – | $85.78 \pm 0.17$ |
| data | 100 | – | 59.27 | 56.13 | – | – | $86.24 \pm 0.18$ |
| set | 200 | – | 58.24 | 49.93 | – | – | $86.29 \pm 0.26$ |
| | 300 | – | 47.15 | 59.97 | – | – | $84.95 \pm 1.13$ |

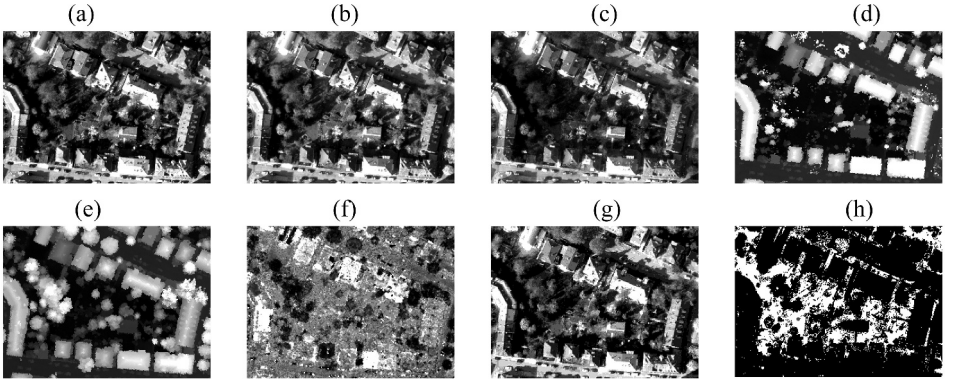


**Figure 5.** Eight multispectral images of TopoSys GmbH (the second data set, example 1) (a) red; (b) green; (c) Blue; (d) last echo; (e) first echo; (f) LIDAR intensity; (g) SGI; and (h) NDVI.

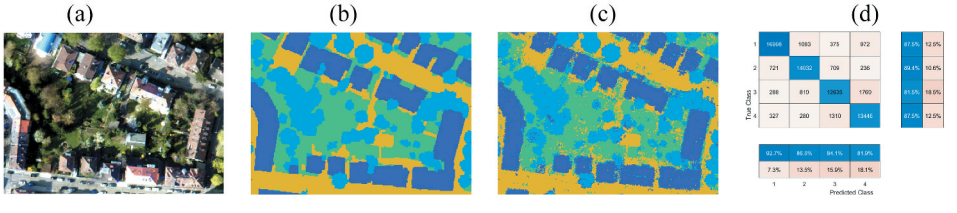


**Figure 6.** Segmentation results of the TopoSys GmbH using $Q = 200$ superpixels (the second data set, example 1); (a) RGB view (b) ground truth; (c) segmentation; and (d) confusion matrix ((1 - Building; 2 - Tree; 3 - low vegetation; and 4 - Impervious surface OR Car).

truth, segmentation representation and classification confusion matrix for $Q = 400$ are shown in Figure 4(b–d), respectively.

- *The second data set* is measured in $220 \times 300$ per band, as shown in Figure 5, so $I^{full} \in \Re^{220\times300\times8}$, resulting in $N = 66000$. A coloured RGB is shown in Figure 4(a). Similar to the first data set, four classes for segmentation are labelled as (1 – Building; 2 -Tree; 3 – Low vegetation; and 4 – Impervious surface OR Car). The segmentation accuracy of the

**Table 2.** Intersection over union (IoU) of example 1 (TopoSys GmbH).

| Class | First data set mean IoU | Second data set mean IoU |
|---|---|---|
| 1- Building | 0.8227 | 0.8186 |
| 2- Tree | 0.8158 | 0.7848 |
| 3- Low vegetation | 0.7752 | 0.7057 |
| 4- Impervious surface OR Car | 0.8316 | 0.7332 |

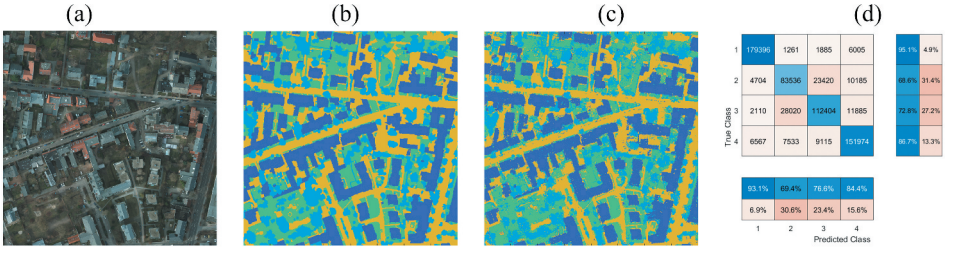


**Figure 7.** Segmentation results of Potsdam 4–11 data set (example 2); (a) RGB view; (b) ground truth; (c) segmentation; and (d) confusion matrix (1 - Building; 2 - Tree; 3 - low vegetation; 4 - Impervious surface OR Car).

proposed SLIC-RBF-CCA algorithm in also listed in Table 1 compared to other clustering approaches. Here the choice of $Q$ must be smaller due to a smaller pixel $N$ than the first data set. We experimented $Q = 50, 100, 200, 300$ for SLIC algorithm which are applicable to SLIC used in the experiments. It can be seen that the proposed SLIC-RBF-CCA significantly outperforms other methods. The second best is Polynomial CCA, which is not SLIC based. Over all values of $Q$, the accuracy performance is insensitive to $Q$ within the 1.5% differences, where $Q = 100, 200$ return the best performance. The ground truth, segmentation representation and classification confusion matrix are illustrated in Figure 8(b–d), respectively for $Q = 200$.

The quality of segmentation of Example 1 over both data sets, as measured by intersection over union (IoU) from set theory, is shown in Table 2, and the mean IoU is calculated using one-against-all for each class. It can be seen that all categories have good performances.

Example 2:Three data sets showing City of Potsdam from the ISPRS 2D Semantic Labelling Contest are used in this experiment Potsdam. The website contains 38 patches of the same size. The patch 4–11, patch 6–14 and patch 7–8 were used to demonstrate the effectiveness of the proposed approaches. Each of the original patch has a high resolution 6000 by 6000 pixels, with multi-band images consisting of R, G, B, IR and normalized DSM. The high-resolution multi band images were resized to 800 by 800 pixels, which are then divided into four 400 by 400 pixels sub-images. The original ground truth was impervious surfaces (RGB: 255, 255, 255), Building (RGB: 0, 0, 255), Low vegetation (RGB: 0, 255, 255), Tree (RGB: 0, 255, 0), Car (RGB: 255, 255, 0), Clutter/background (RGB: 255, 0, 0). This is also resized to 800 by 800 pixels, followed by rounding to the closest (255 or 0) in each R, G, B band. Two additional bands of NDVI and SGI were generated, so there are seven bands

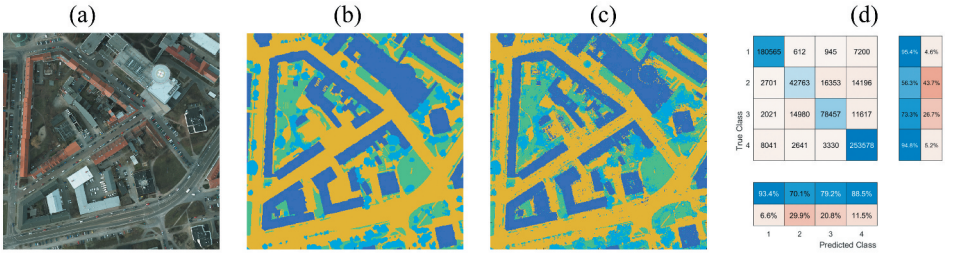


**Figure 8.** Segmentation results of Potsdam 6–14 data set (example 2); (a) RGB view; (b) ground truth; (c) segmentation; and (d) confusion matrix (1- Building; 2 - Tree; 3- low vegetation; 4 - Impervious surface OR Car).

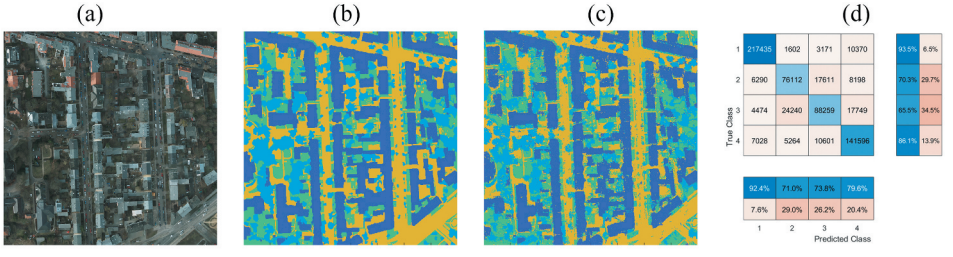


**Figure 9.** Segmentation results of Potsdam 7–8 data set (example 2); (a) RGB view; (b) ground truth; (c) segmentation; and (d) confusion matrix (1 - Building; 2 - Tree; 3 - low vegetation; 4 - Impervious surface OR Car).

are used for each data set, $M = 7$. The RGB view of the image is shown in Figures 7(a)–9(a). The ground truth of this image contains four classes (1 - Building; 2 - Tree; 3 - Low vegetation; and 4 - Impervious surface or Car) that were manually labelled pixel-wise are used for validation. For three patches, the proposed SLIC-RBF-CCA algorithm is compared against several unsupervised and semi-supervised clustering approaches with their segmentation accuracy shown in Tables 3–5, for each quarter of the patches. We experimented $Q = 100, 200, 400, 800$ for the SLIC algorithm used in the experiments. It can be seen that the proposed SLIC-RBF-CCA outperforms other methods. Without using SLIC algorithm, the Polynomial CCA is worse than SLIC-RBF-CCA, but other approaches are significantly worse. Similar to Example 1, we noted that segmentation accuracy is insensitive to a large range of Q used for testing, since for most patches the variation of segmentation accuracy $< 1\%$ for different $Q$. It is also noted that variance is larger for $Q = 800$, hence $Q$ should be set up a smaller value if possible. For illustration, the segmentation results are combined from four sub-images, and the ground truth, combined segmentation representation and combined classification confusion matrix of the proposed SLIC-RBF-CCA from one realization using $Q = 400$ are shown in Figures 7(b–d)–9(b–d), respectively. The quality of segmentation of Example 2 as combined from four sub-images, as measured by intersection over union (IoU) from set theory, is shown in Table 6 using the same realization. The mean IoU is calculated using one-against-all for each class. It is seen that the segmentation performance of buildings and others are much better than that of tree and low vegetation categories. However, the resultant IOU around

**Table 3.** Example 2 (Patch 4–11) – comparison of segmentation accuracy (percentage of correctly identified pixel labels). For semi-supervised approaches, the results are given as *mean* ± *std* over 20 repetitive random sampled labelled data sets.

| Patch 4–11 | *Q* | *k*-means | *k*-means-SLIC | SC- SLIC | Linear CCA | Polynomial CCA | Proposed SLIC-RBF-CCA |
|---|---|---|---|---|---|---|---|
| | – | 54.86 | – | – | 67.65 ± 3.74 | 80.47 ± 0.12 | – |
| Top | 100 | – | 53.47 | 51.30 | – | – | 82.81 ± 0.08 |
| left | 200 | – | 61.68 | 58.28 | – | – | 82.96 ± 0.11 |
| Quarter | 400 | – | 64.33 | 59.24 | – | – | 82.53 ± 0.14 |
| | 800 | – | 64.95 | 61.93 | – | – | 81.21 ± 0.62 |
| | – | 52.76 | – | – | 65.43 ± 3.81 | 77.10 ± 0.13 | – |
| Bottom | 100 | – | 52.26 | 53.68 | – | – | 80.94 ± 0.16 |
| left | 200 | – | 54.17 | 47.89 | – | – | 81.65 ± 0.13 |
| Quarter | 400 | – | 53.52 | 58.18 | – | – | 81.51 ± 0.09 |
| | 800 | – | 54.67 | 53.35 | – | – | 79.83 ± 1.44 |
| | – | 57.11 | – | – | 68.90 ± 2.69 | 81.12 ± 0.16 | – |
| Top | 100 | – | 54.12 | 52.23 | – | – | 82.98 ± 0.11 |
| right | 200 | – | 55.73 | 55.26 | – | – | 82.73 ± 0.09 |
| Quarter | 400 | – | 55.25 | 60, 77 | – | – | 82.69 ± 0.13 |
| | 800 | – | 59.68 | 64.57 | – | – | 81.42 ± 0.43 |
| | – | 56.78 | – | – | 72.66 ± 1.85 | 81.55 ± 0.14 | – |
| Bottom | 100 | – | 53.55 | 57.48 | – | – | 83.94 ± 0.13 |
| right | 200 | – | 56.47 | 57.87 | – | – | 84.41 ± 0.11 |
| Quarter | 400 | – | 58.32 | 57.72 | – | – | 83.90 ± 0.24 |
| | 800 | – | 56.08 | 56.33 | – | – | 82.51 ± 0.43 |

**Table 4.** Example 2 (Patch 6–14) – comparison of segmentation accuracy (percentage of correctly identified pixel labels). For semi-supervised approaches, the results are given as *mean* ± *std* over 20 repetitive random sampled labelled data sets.

| Patch 6–14 | *Q* | *k*-means | *k*-means-SLIC | SC- SLIC | Linear CCA | Polynomial CCA | Proposed SLIC-RBF-CCA |
|---|---|---|---|---|---|---|---|
| | – | 59.31 | – | – | 60.60 ± 1.87 | 83.08 ± 0.15 | – |
| Top | 100 | – | 57.67 | 52.70 | – | – | 84.45 ± 0.18 |
| left | 200 | – | 57.02 | 58.25 | – | – | 84.54 ± 0.22 |
| Quarter | 400 | – | 56.83 | 50.42 | – | – | 83.95 ± 0.24 |
| | 800 | – | 58.57 | 62.47 | – | – | 83.24 ± 0.79 |
| | – | 56.42 | – | – | 55.05 ± 5.20 | 88.34 ± 0.11 | – |
| Bottom | 100 | – | 69.46 | 52.27 | – | – | 89.98 ± 0.20 |
| left | 200 | – | 67.60 | 59.36 | – | – | 90.40 ± 0.18 |
| Quarter | 400 | – | 68.25 | 56.39 | – | – | 90.14 ± 0.23 |
| | 800 | – | 67.27 | 63.67 | – | – | 89.61 ± 0.67 |
| | – | 59.40 | – | – | 57.86 ± 2.21 | 83.27 ± 0.63 | |
| Top | 100 | – | 58.46 | 51.24 | – | – | 87.98 ± 0.12 |
| right | 200 | – | 55.70 | 60.51 | – | – | 88.30 ± 0.13 |
| Quarter | 400 | – | 63.04 | 55.12 | – | – | 88.02 ± 0.17 |
| | 800 | – | 62.72 | 47.40 | – | – | 87.57 ± 0.20 |
| | – | 56.98 | – | – | 58.72 ± 7.86 | 84.04 ± 0.11 | – |
| Bottom | 100 | – | 65.65 | 55.81 | – | – | 85.63 ± 0.15 |
| right | 200 | – | 57.31 | 49.14 | – | – | 85.86 ± 0.23 |
| Quarter | 400 | – | 58.71 | 50.27 | – | – | 85.23 ± 0.38 |
| | 800 | – | 54.84 | 52.00 | – | – | 84.07 ± 0.72 |

**Table 5.** Example 2 (Patch 7–8) – comparison of segmentation accuracy (percentage of correctly identified pixel labels). For semi-supervised approaches, the results are given as *mean* ± *std* over 20 repetitive random sampled labelled data sets.

| Patch 7–8 | $Q$ | $k$-means | $k$-means-SLIC | SC- SLIC | Linear CCA | Polynomial CCA | Proposed SLIC-RBF-CCA |
|---|---|---|---|---|---|---|---|
| | – | 57.50 | – | – | $72.21 \pm 2.37$ | $80.76 \pm 0.10$ | – |
| Top | 100 | – | 52.85 | 49.88 | – | – | $81.90 \pm 0.10$ |
| left | 200 | – | 56.47 | 57.74 | – | – | $82.19 \pm 0.15$ |
| Quarter | 400 | – | 53.73 | 53.99 | – | – | $81.94 \pm 0.17$ |
| | 800 | – | 55.61 | 55.38 | – | – | $80.98 \pm 0.13$ |
| | – | 47.22 | – | – | $56.70 \pm 1.06$ | $75.72 \pm 0.17$ | – |
| Bottom | 100 | – | 41.68 | 46.99 | – | – | $81.63 \pm 0.12$ |
| left | 200 | – | 44.17 | 51.24 | – | – | $81.85 \pm 0.11$ |
| Quarter | 400 | – | 46.21 | 51.79 | – | – | $81.36 \pm 0.14$ |
| | 800 | – | 46.36 | 49.57 | – | – | $79.60 \pm 0.85$ |
| | – | 55.04 | – | – | $63.89 \pm 1.19$ | $78.84 \pm 0.10$ | – |
| Top | 100 | – | 51.71 | 54.26 | – | – | $83.55 \pm 0.07$ |
| right | 200 | – | 51.74 | 47.74 | – | – | $83.77 \pm 0.08$ |
| Quarter | 400 | – | 53.38 | 55.54 | – | – | $83.20 \pm 0.14$ |
| | 800 | – | 54.47 | 53.29 | – | – | $81.90 \pm 0.20$ |
| | – | 51.42 | – | – | $65.54 \pm 5.96$ | $77.97 \pm 0.12$ | – |
| Bottom | 100 | – | 50.06 | 52.41 | – | – | $80.49 \pm 0.21$ |
| right | 200 | – | 50.23 | 42.65 | – | – | $80.88 \pm 0.21$ |
| Quarter | 400 | – | 49.53 | 48.05 | – | – | $80.45 \pm 0.65$ |
| | 800 | – | 49.73 | 46.78 | – | – | $78.05 \pm 2.39$ |

**Table 6.** Intersection over union (IoU) of example 2 (Potsdam).

| Class | Patch 4–11 mean IoU | Patch 6–14 mean IoU | Patch 7–8 mean IoU |
|---|---|---|---|
| 1- Building | 0.8884 | 0.8935 | 0.8685 |
| 2- Tree | 0.5265 | 0.4537 | 0.5463 |
| 3- Low vegetation | 0.5952 | 0.6144 | 0.5313 |
| 4- Impervious surface OR Car | 0.7477 | 0.8436 | 0.7051 |

0.5 is still reasonable since it can be difficult to distinguish small/sparse trees from low vegetation due to their shared characteristics of NDVI. On the other hand, it may be a common phenomenon for any segmentation method to obtain unbalanced performance for different categories, as these are dependent on many factors from original image source, feature construction methods, class balancenes in data, etc. For example, the multi-band images contains normalized DSM, which would contribute the excellent performance of detection of buildings. While it is hoped that the combination of NDVI and height information can help to detection trees from low vegetation, the trees in the experimenting image may not be green and small in size, results in higher IOU error with other categories.

## 5. Conclusions

Due to its simplicity and the low computational costs, image segmentation by clustering may be a viable alternative to expensive supervised deep learning if satisfactory performance is achievable. Applied to image segmentation of remote sensing image

data, this study has introduced a novel semi-supervised spectral clustering method by assigning each co-registered pixel a cluster label with resort to only a small number of labelled pixels for training. The proposed method consists of two steps. Firstly, a Simple Linear Iterative Clustering (SLIC) step is initially carried out to a set of pseudo-RGB images obtained from singular vectors of a flattened matrix of a multi-band image. Then the proposed SLIC-RBF-CCA is applied for final segmentation. The proposed SLIC-RBF-CCA employs a novel radial basis function based on SLIC superpixels, which acts as the first variable in the CCA framework. Specifically, RBF centres are obtained from local data mean related to SLIC generated superpixel regions. We define the second variable of SLIC-RBF-CCA based on a small number of labelled pixels. The proposed approach is advantageous in that only a minimum of input from users is required: a preset superpixel number and a desired cluster number. Some discussions and insights are provided, in particular on multi-view learning interpretation. It is shown that the proposed algorithm is capable of excellent segmentation for several benchmark land cover data sets compared with other clustering based segmentation approaches. This work contributes a novel approach to remote image land cover segmentation.

It is noted that clustering approaches for image segmentation are fundamentally different from deep learning classification models which are useful since a well trained deep learning model may generalize well to similar images so as to benefit large-scale data applications or facilitate transfer learning. In common to all clustering approaches the proposed method is most useful for data exploratory analysis for given images, since it is not a parameterized classifier. However we believe that the proposed approach may be adaptable to high-resolution complex scenes with little user interactions using their prior knowledge for pre-processing or post-precessing images such as specific characteristics (e.g. types of land cover, specific features in the scenes). In future work, with regard to semi-supervised multispectral image segmentation for remote sensing imagery, different land cover scenarios will be considered. The algorithm will be further developed and optimized in terms of robustness in challenging situations with a focus on imbalanced clusters, scattered important objects, and segmentation of similar objects such as trees and low vegetation.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## References

Achanta, R., A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. 2012. "SLIC Superpixels Compared to State-of-the-Art Superpixel Methods." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 34 (11): 2274–2282. https://doi.org/10.1109/TPAMI.2012.120.

Akaho, S. 2001. "A Kernel Method for Canonical Correlation Analysis." In *International Meeting of Psychometric Society*, 1.

Amrani, M., A. Bey, and A. Amamra. 2022. "New SAR Target Recognition Based on YOLO and Very Deep Multi-Canonical Correlation Analysis." *International Journal of Remote Sensing* 43 (15–16): 5800–5819. https://doi.org/10.1080/01431161.2021.1953719.

Bahdanau, D., K. Cho, and Y. Bengio. 2015. "Neural Machine Translation by Jointly Learning to Align and Translate." In *3rd International Conference on Learning Representations, ICLR 2015, Conference Track Proceedings,* edited by Y. Bengio and Y. LeCun. San Diego, CA, USA. May 7–9.

Basu, S., A. Banerjee, and R. J. Mooney. 2004. "Active Semi-supervision for Pairwise Constrained Clustering." In *Proceedings of the 2004 SIAM International Conference on Data Mining (SDM)*, 333–344.

Cai, J., J. Hao, H. Yang, X. Zhao, and Y. Yang. 2023. "A Review on Semi-Supervised Clustering." *Information Sciences* 632:164–200. https://doi.org/10.1016/j.ins.2023.02.088.

Chen, J., Z. Li, and B. Huang. 2017. "Linear Spectral Clustering Superpixel." *IEEE Transactions on Image Processing* 26 (7): 3317–3330. https://doi.org/10.1109/TIP.2017.2651389.

Chen, S., X. Hong, and C. J. Harris. 2011. "Grey-Box Radial Basis Function Modelling." *Neurocomputing* 74 (10): 1564–1571. https://doi.org/10.1016/j.neucom.2011.01.023.

Chen, Z., and B. Wang. 2015. "Semisupervised Spectral–Spatial Classification of Hyperspectral Imagery with Affinity Scoring." *IEEE Geoscience & Remote Sensing Letters* 12 (8): 1710–1714. https://doi.org/10.1109/LGRS.2015.2421347.

de la Iglesia Martinez, A., and S. Labib. 2023. "Demystifying Normalized Difference Vegetation Index (Ndvi) for Greenness Exposure Assessments and Policy Interventions in Urban Greening." *Environmental Research* 220:115155. https://doi.org/10.1016/j.envres.2022.115155.

Ester, M., H.-P. Kriegel, J. Sander, and X. Xu. 1996. "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise." In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 96, 226–231. AAAI Press.

Fang, F., B. E. McNeil, T. A. Warner, and A. E. Maxwell. 2018. "Combining High Spatial Resolution Multi-Temporal Satellite Data with Leaf-On LIDAR to Enhance Tree Species Discrimination at the Crown Level." *International Journal of Remote Sensing* 39 (23): 9054–9072. https://doi.org/10.1080/01431161.2018.1504343.

Ghasemi, Z., H. A. Khorshidi, and U. Aickelin. 2021. "A Survey on Optimisation-Based Semi-Supervised Clustering Methods." In *2021 IEEE International Conference on Big Knowledge (ICBK)*, 477–482.

Golalipour, K., E. Akbari, S. S. Hamidi, M. Lee, and R. Enayatifar. 2021. "From Clustering to Clustering Ensemble Selection: A Review." *Engineering Applications of Artificial Intelligence* 104:104388. https://doi.org/10.1016/j.engappai.2021.104388.

Hagen, L., and A. Kahng. 1992. "New Spectral Methods for Ratio Cut Partitioning and Clustering." *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 11 (9): 1074–1085. https://doi.org/10.1109/43.159993.

Han, B., P. Han, and Z. Cheng. 2022. "Superpixel Segmentation of PolSAR Images Based on Improved Local Iterative Clustering." *International Journal of Remote Sensing* 43 (8): 2735–2754. https://doi.org/10.1080/01431161.2021.1993466.

Haralick, R. M., K. Shanmugam, and I. Dinstein. 1973. "Textural Features for Image Classification." *IEEE Transactions on Systems, Man, and Cybernetics SMC* 3 (6): 610–621. https://doi.org/10.1109/TSMC.1973.4309314.

Hardle, W., and L. Simar. 2007. "Canonical Correlation Analysis." In *Applied Multivariate Statistical Analysis*, 321–330. Berlin Heidelberg: Springer.

Hotelling, H. 1935. "The Most Predictable Criterion." *Journal of Educational Psychology* 26 (2): 139–142. https://doi.org/10.1037/h0058165.

Hotelling, H. 1936. "Relations Between Two Sets of Variables." *Biometrika* 28 (3–4): 321–377. 12. https://doi.org/10.1093/biomet/28.3-4.321.

ISPRS. *2D Semantic Labeling Contest - Potsdam*.

Jiang, M., X. Chen, L. Xu, and D. A. Clausi. 2022. "Semi-Supervised Sea Ice Classification of SAR Imagery Based on Graph Convolutional Network." In *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 1031–1034.

Jiao, X., Y. Chen, and R. Dong. 2020. "An Unsupervised Image Segmentation Method Combining Graph Clustering and High-Level Feature Representation." *Neurocomputing* 409:83–92. https://doi.org/10.1016/j.neucom.2020.05.073.

Kganyago, M., C. Adjorlolo, P. Mhangara, and L. Tsoeleng. 2024. "Optical Remote Sensing of Crop Biophysical and Biochemical Parameters: An Overview of Advances in Sensor Technologies and Machine Learning Algorithms for Precision Agriculture." *Computers and Electronics in Agriculture* 218:108730. https://doi.org/10.1016/j.compag.2024.108730.

Kim, T. H., K. M. Lee, and S. U. Lee. 2013. "Learning Full Pairwise Affinities for Spectral Segmentation." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 35 (7): 1690–1703. https://doi.org/10.1109/TPAMI.2012.237.

Li, C.-G., C. You, and R. Vidal. 2017. "Structured Sparse Subspace Clustering: A Joint Affinity Learning and Subspace Clustering Framework." *IEEE Transactions on Image Processing* 26 (6): 2988–3001. https://doi.org/10.1109/TIP.2017.2691557.

Lovász, L., and M. Plummer. 2009. *Matching Theory*. AMS Chelsea Publishing Series. Akadémiai Kiadó.

Luo, M., C. Yan, Q. Zheng, X. Chang, L. Chen, and F. Nie. 2019. "Discrete Multi-Graph Clustering." *IEEE Transactions on Image Processing* 28 (9): 4701–4712. https://doi.org/10.1109/TIP.2019.2913081.

Luxburg, U. 2007. "A Tutorial on Spectral Clustering." *Statistics and Computing* 17 (4): 395–416. dec. https://doi.org/10.1007/s11222-007-9033-z.

MacQueen, J., 1967. "Some Methods for Classification and Analysis of Multivariate Observations." In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Volume 1, 281–297. Oakland, CA, USA.

Neeti, N. 2014. "Extending T-Mode Canonical Correlation Analysis to T-Mode Pre-Filtered Canonical Correlation Analysis: A Novel Approach to Discover Shared Patterns Between Two Image Time Series." *International Journal of Remote Sensing* 35 (5): 1926–1935. https://doi.org/10.1080/01431161.2013.879352.

Ng, A., M. Jordan, and Y. Weiss. 2001. "On Spectral Clustering: Analysis and an Algorithm." *Advances in Neural Information Processing Systems* 2 (11).

Pal, R., H. Begum, S. Mukhopadhyay, S. Sarkar, D. Chakraborty, S. Majumdar, and D. Sengupta. 2021. "Edge Directed Radial Basis Function Based Interpolation Towards PCA Based Pan-Sharpening." *International Journal of Remote Sensing* 42 (23): 9047–9067. https://doi.org/10.1080/01431161.2021.1981560.

Pang, Y., W. Wang, L. Du, Z. Zhang, X. Liang, Y. Li, and Z. Wang. 2021. "Nyström-Based Spectral Clustering Using Airborne LiDAR Point Cloud Data for Individual Tree Segmentation." *International Journal of Digital Earth* 14 (10): 1452–1476. https://doi.org/10.1080/17538947.2021.1943018.

Poggio, T., and F. Girosi. 1990. "Networks for Approximation and Learning." *Proceedings of the IEEE* 78(9): 1481–1497.

Pont-Tuset, J., P. Arbeláez, J. T. Barron, F. Marques, and J. Malik. 2017. "Multiscale Combinatorial Grouping for Image Segmentation and Object Proposal Generation." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 39 (1): 128–140. https://doi.org/10.1109/TPAMI.2016.2537320.

Shi, J., and J. Malik. 2000. "Normalized Cuts and Image Segmentation." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 22 (8): 888–905. https://doi.org/10.1109/34.868688.

Sun, S. 2013. "A Survey of Multi-View Machine Learning." *Neural Computing & Applications* 23 (7–8): 2031–2038. https://doi.org/10.1007/s00521-013-1362-6.

Taiwo, B. E., A. A. Kafy, A. A. Samuel, Z. A. Rahaman, O. E. Ayowole, M. Shahrier, B. M. Duti, M. T. Rahman, O. T. Peter, and O. O. Abosede. 2023. "Monitoring and Predicting the Influences of Land Use/Land Cover Change on Cropland Characteristics and Drought Severity Using Remote Sensing Techniques." *Environmental and Sustainability Indicators* 18:100248. https://doi.org/10.1016/j.indic.2023.100248.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. 2017. "Attention Is All You Need." In *Advances in Neural Information Processing Systems*, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Vol. 30. Curran Associates, Inc.

Vivone, G. 2023. "Multispectral and Hyperspectral Image Fusion in Remote Sensing: A Survey." *Information Fusion* 89:405–417. https://doi.org/10.1016/j.inffus.2022.08.032.

Wang, J., N. Ding, and G. He. 2023. "A Boundary Enhancement Loss Function for Semantic Segmentation of Land Cover." *International Journal of Remote Sensing* 44 (12): 3637–3659. https://doi.org/10.1080/01431161.2023.2224101.

Wang, R., S. Han, J. Zhou, Y. Chen, L. Wang, T. Du, K. Ji, Y.-O. Zhao, and K. Zhang. 2023. "Transfer-Learning-Based Gaussian Mixture Model for Distributed Clustering." *IEEE Transactions on Cybernetics* 53 (11): 7058–7070. https://doi.org/10.1109/TCYB.2022.3177242.

Wang, W., X. Liu, and X. Mou. 2021. "Optimal Superpixel Selection for Hyperspectral Image Classification of Limited Training Samples." *International Journal of Remote Sensing* 42 (23): 9068–9084. https://doi.org/10.1080/01431161.2021.1988184.

Wang, Y., S. Cang, and H. Yu. 2019. "Mutual Information Inspired Feature Selection Using Kernel Canonical Correlation Analysis." *Expert Systems with Applications* 4:100014. https://doi.org/10.1016/j.eswax.2019.100014.

Wei, H., J. Xiao, Y. Zhang, and X. Hong. 2024. "Semi-Supervised Segmentation of Land Cover Images Using Nonlinear Canonical Correlation Analysis with Multiple Features and T-SNE." *ArXiv*: 2401.12164.

Yang, X., W. F. Liu, W. Liu, and D. Tao. 2021. "A Survey on Canonical Correlation analysis." *IEEE Transactions on Knowledge and Data Engineering* 33 (6): 2349–2368. https://doi.org/10.1109/TKDE.2019.2958342.

Ye, C., S. Liu, M. Xu, and Z. Yang. 2022. "Combining Low-Rank Constraint for Similar Superpixels and Total Variation Sparse Unmixing for Hyperspectral Image." *International Journal of Remote Sensing* 43 (12): 4331–4351. https://doi.org/10.1080/01431161.2022.2111532.

Yi, S., H. Ma, X. Wang, T. Hu, X. Li, and Y. Wang. 2022. "Weakly-Supervised Semantic Segmentation with Superpixel Guided Local and Global Consistency." *Pattern Recognition* 124:108504. https://doi.org/10.1016/j.patcog.2021.108504.

Zhang, Y., and G. F. Hepner. 2017. "The Dynamic-Time-Warping-Based K-Means++ Clustering and Its Application in Phenoregion Delineation." *International Journal of Remote Sensing* 38 (6): 1720–1736. https://doi.org/10.1080/01431161.2017.1286055.

Zhong, G., and C.-M. Pun. 2023. "Self-Taught Multi-View Spectral Clustering." *Pattern Recognition* 138:109349. https://doi.org/10.1016/j.patcog.2023.109349.