



A top-down approach for semantic segmentation of big remote sensing images

Wadii Boulila^{1,2}

Received: 8 May 2018 / Accepted: 26 December 2018 / Published online: 5 January 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

The increasing amount of remote sensing data has opened the door to new challenging research topics. Nowadays, significant efforts are devoted to pixel and object based classification in case of massive data. This paper addresses the problem of semantic segmentation of big remote sensing images. To do this, we proposed a top-down approach based on two main steps. The first step aims to compute features at the object-level. These features constitute the input of a multi-layer feed-forward network to generate a structure for classifying remote sensing objects. The goal of the second step is to use this structure to label every pixel in new images. Several experiments are conducted based on real datasets and results show good classification accuracy of the proposed approach. In addition, the comparison with existing classification techniques proves the effectiveness of the proposed approach especially for big remote sensing data.

Keywords Semantic segmentation · Remote sensing images · Neural networks · Big data

Introduction

Analysis and interpretation of Remote Sensing (RS) images helped people to understand many phenomena related to the earth. Information extracted from RS images can be used in many fields such as weather forecasting, resources management, regional planning, traffic monitoring, and environmental risk assessment. Many analysis and interpretation tasks rely on understanding the content of an RS image or scene. In literature, several techniques were proposed to help understand the content of RS images. Among these techniques, we list object detection, object recognition, image segmentation, and semantic image segmentation. Often, these techniques may lead to some confusion. Object detection in RS images aims to determine if an image contains one or more objects belonging to the class of interest and to locate their positions (Lei et al. 2012; Cheng and Han 2016). Object

recognition aims to detect all objects in RS images and locate their positions (Durand et al. 2007; Haiyang and Fuping 2009; Diao et al. 2015). In image segmentation, the RS image will be divided into regions; however, these regions will not be labeled (Ming et al. 2015; Zhang et al. 2017). Semantic image segmentation will label each pixel in the RS image according to a class of objects such as urban, forest, water, etc. (Athanasiadis et al. 2007; Shotton and Kohli 2014; Zheng et al. 2017). In this paper, we are interested in the semantic RS image segmentation.

The semantic segmentation of RS images has been a core topic for the last years. Several methods for semantic image segmentation act at the pixel-level by classifying each pixel independently. Other methods try to group pixels into clusters and assign a label to these clusters. For instance, Ma et al. (2014) proposed an object-oriented based approach that combines a pixel-based classification and a segmentation technique. The goal of this approach is to classify polarimetric Synthetic Aperture Radar (SAR) images. Authors in this work developed a soft voting strategy to fuse multiple classifiers. The approach is validated through a set of experiments that are conducted on two quadpolarimetric SAR images. In (Rau et al. 2014), authors proposed an object-oriented analysis scheme for landslide recognition using existing software. The input data comprised only multispectral optical ortho-images and the digital elevation mode. Rau et al. developed

Communicated by: H. Babaie

✉ Wadii Boulila
wadii.boulila@riadi.rnu.tn

¹ IS Department, College of Computer Science and Engineering, Taibah University, Medina, Saudi Arabia

² RIADI Laboratory, National School of Computer Sciences, University of Manouba, Manouba, Tunisia

a semiautomatic method that detects the landslide seeds and then performs a region growth and false-positive elimination for these seeds. Zhang et al. (2016a) proposed to overcome the semantic gap between low-level visual features and high-level semantics of images. In this study, authors developed an object-based mid-level representation method for semantic classification. The proposed algorithm is based on the bag-of-visual-words that generates mid-level features to bridge the two levels. In (Zhang et al. 2016c), authors developed a higher order potential function based on nonlocal shared constraints within the framework of a conditional random field model. The proposed approach combines classification knowledge from labeled data with unsupervised segmentation cues derived from the test data. The conditional random field model integrates low-level and high-level contextual cues from labeled and unlabeled test datasets. In (Andrés et al. 2017), authors presented an approach based on ontology to classify RS images. Andrés et al. developed spectral rules for a pixel-based classification of Landsat images. The proposed prototype is coupled with an open source image processing software at the pre-processing step and it uses a reasoner algorithm to perform image classification. The major limitation highlighted for the proposed system is related to the processing time. Zheng et al. (2017) detailed a semantic segmentation of high spatial resolution RS images. The proposed approach is based on an object MRF (Markov random field) model with auxiliary label fields for semantic segmentation of the RS image. The idea of the proposed approach is to define a label field and two auxiliary label fields on the same region adjacency graph with different class numbers. Then, a net structure is built to describe the interactions between label fields and messages passed between each label field and the two other auxiliary label fields. Marmanis et al. (2018) presented a deep convolutional neural network for semantic segmentation with boundary detection. Authors proposed to combine semantic segmentation with semantically informed edge detection by adding boundary detection to the encoder-decoder architecture. In (Boulila et al. 2018), authors developed a decision support tool for RS big data analytics. The main idea, in this work, is to assist users in making decisions in many RS-related fields. The proposed tool provides descriptive, predictive and prescriptive analytics. Boulila et al. proposed to overcome the complexity of RS data by implementing an iterative and incremental process of data integration. Additionally, they designed a multidimensional model based on a star schema for the image data warehouse, and they proposed techniques such as distribution, indexing and partitioning to enhance the retrieval of RS big data. Experiments were conducted based on three different applications (clustering, decision tree, and association rules).

However, recent years have witnessed an increasing amount of RS images with different spectral and spatial resolutions (Liu et al. 2016). This increasing amount of images has

open the door to new challenging problems facing the RS community such as how to extract valuable information from the various kinds of RS data? How to deal with the increasing data types and volume? (Zhang et al. 2016b). The problem of semantic segmentation of big RS images is becoming a challenging research topic. In the current manuscript, we propose a top-down approach for semantic segmentation. The main goal of the top-level is to compute features for objects extracted from RS images. While the goal of the down-level is to determine the class of each pixel using information computed in the previous level.

The remainder of this paper is organized as follows. In Section 2, we detail the proposed approach for semantic segmentation of RS big images. The presented method is experimented and evaluated through different real datasets in Section 3. Finally, Section 4 concludes the paper and discusses some issues for further research.

Proposed approach

The goal of the proposed approach is to partition RS image into meaningful objects and assign a class to each of them. This goal is ensured through a top-down approach composed by an object processing (top-level) and pixel processing (down-level). The proposed approach is described in the following parts in detail.

Proposed approach for semantic segmentation of big RS images

Figure 1 describes the proposed approach. The process of the proposed semantic segmentation is divided into two levels: 1) top-level and 2) down-level. The first level aims to ensure training of the multi-layer feed-forward neural network (MLFFNN). We start by computing features of object extracted from RS images. These features constitute the input of the MLFFNN module to generate a structure for classifying RS objects. In the second step, down-level, the generated structure is used to perform semantic segmentation at the pixel-level. For an input RS image, an 8-matrix centered in every pixel is considered when computing features related to that pixel. The same features computed at the object-level are computed at the pixel-level (we compute these features based on the 3×3 window surrounding the pixel). The computed features will be entered to the MLFFNN to determine the most similar trained class and assign classes to each pixel.

The proposed process of semantic image segmentation is depicted in the algorithm 1.

Algorithm 1 has as input: 1) imgs, which represent a set of RS images, 2) nbrclass, which represents the number of classes for the input images, 3) connectivity, which means the number of pixels connected horizontally, vertically and

diagonally to every pixel in the considered image, 4) `minNumberPixels`, which corresponds to the minimum number of pixels that every extracted object should contain, 5) `hiddenLayerSize`, which corresponds to hidden layer sizes

for the MLFFNN (see section 2.e for more details), 6) `trainFcn` and `performFcn`, which represent, respectively, the training and the performance functions for the neural network. The output of the algorithm 1 are classified input images.

Algorithm 1 Proposed RS semantic image segmentation

Input: `imgs`, `nbrclass`, `connectivity`, `minNumberPixels`, `hiddenLayerSize`, `trainFcn`, `performFcn`

Output: classified RS images

// Extract features from RS images

Foreach `img` in `imgs` **do**

 Perform a segmentation of `img` into `nbrclass` using k-means

 Assign labels to classes in `img`

 Divide classes in `img` into objects (subclasses) according to `connectivity` and `minNumberPixels`

 Declare a variable `inputNN`

Foreach `class` in `classes` **do**

Foreach `object` in `class` **do**

 Compute features for object

 Add features of object and corresponding label to `inputNN`

End foreach

End foreach

End foreach

//Create a pattern recognition network

`net` = `patternnet(hiddenLayerSize, trainFcn, performFcn)`

Divide data for training, validation and testing

Train the net network

Test the net neural network

//Use neural network to classify pixels in RS images; `inputImg` is an example of image to be classified

Foreach `pixel` in `inputImg` **do**

 Compute features for pixel

 Determine the class of pixel using `net`

End foreach

Top level: object processing

This level aims to generate a neural network structure that will be used for semantic segmentation. The object processing is divided into five steps: a) image segmentation, b) object extraction, c) computation of object features, d) object labeling, and e) generation of neural network structure.

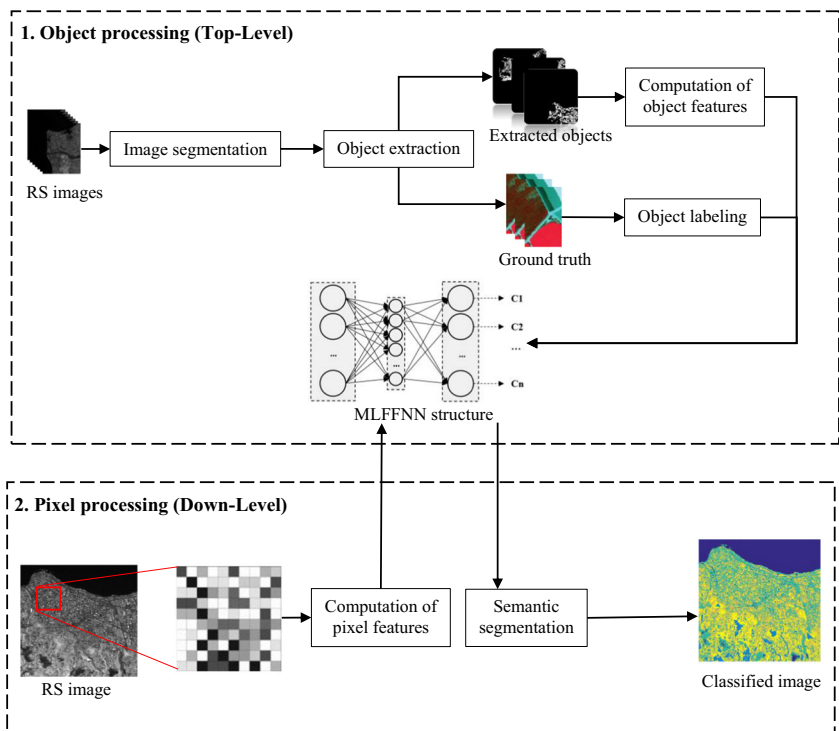
RS image segmentation

It is trivial that results of subsequent steps deeply depends on results provided by the image segmentation step. The

success of image interpretation is strongly related to our reliability on segmentation. Today, the problem of accurate partitioning of RS images is generally a challenging problem. Many works have been achieved on RS image segmentation. Among these works, we can list (Haralick and Shapiro 1985; Ryherd and Woodcock 1996; Trias-Sanz et al. 2008; Akcay and Aksoy 2008; Boulila et al. 2010; Cheng and Han 2016).

In this paper, the k-means method is used to segment images (MacQueen 1967). This method can be replaced by any other segmentation method.

Fig. 1 Proposed approach for semantic segmentation of big RS images



Object extraction

After image segmentation, we obtain a set of objects that collectively cover the entire image. Pixels belonging to the same object have the same label. The goal of this step is to determine meaningful objects from segmented images. Small objects are removed from segmented images. In this paper, we chose two parameters, connectivity and minimum number of pixels (respectively connectivity and minNumberPixels in the algorithm 1), to perform this task. All connected objects that have less than a given pixel value from the segmented image are disregarded. This operation is known as an area opening (Vincent 1993). For the connectivity, we consider the context of 8-connected pixels (3×3 window containing the pixel and its neighbors that are connected to it horizontally, vertically and diagonally). The object extraction task starts by determining connected components. Then, computing the area of each component. Finally, removing small objects (less than minNumberPixels). Another operation is performed in this step is removing isolated pixels from the segmented image.

Computation of object features

Computing features aims to find a mapping from pixel-level to a high-level data space. Feature extraction plays an important role in RS image analysis and interpretation. In the present paper, we choose five features computed on objects extracted from RS images.

Let us consider an object *obj* extracted from a satellite image *img*. The features used in the proposed study are:

- The radiometry of the centroid of the object

$$f_1 = \text{img}(\text{centroid}(\text{obj})) \tag{1}$$

Where *centroid* is the function that returns the centroid of the object *obj*.

- The five features coming from GLCM (Gray-Level Co-Occurrence Matrix) of an object. These features are the contrast, the correlation, the energy, the homogeneity, and the entropy (Haralick et al. 1973; Conners et al. 1984; Yang et al. 2012). The GLCM computes the number of different combinations of gray levels occurring in the object *obj*. Features extracted from the GLCM give a measure of the variation in intensity at the pixel of interest.

Let us consider $p(i,j)$ the element that has the coordinate (i,j) in the normalized symmetrical GLCM.

Contrast: measures the contrast intensity between a pixel and its neighbor over the object. The formula of the contrast is given by the following equation:

$$f_2 = \sum_{i,j} (i-j)^2 p(i,j) \tag{2}$$

Correlation: measures the correlation between a pixel and its neighbor over the object. The formula of the contrast is given by the following formula:

$$f_3 \sum_{i,j} p(i,j) \frac{(i-\mu)(j-\mu)}{\sigma^2} \tag{3}$$

Where μ is the mean of the GLCM, calculated as $\mu = \sum_{i,j} p(i,j)i$, and $\sigma^2 = \sum_{i,j} p(i,j)(i-\mu)^2$

Energy (also known as uniformity): calculates the sum of squared elements in the moment. The formula of the energy is given by the following formula:

$$f_4 \sum_{i,j} (p(i,j))^2 \tag{4}$$

Homogeneity: measures how often the distribution of GLCM elements are close to the GLCM diagonal. The formula of the homogeneity is given by the following formula:

$$f_5 \sum_{i,j} \frac{p(i,j)}{1 + (i-j)^2} \tag{5}$$

Entropy: quantifies the randomness of the gray-level intensity distribution. The formula of the entropy is given by the following formula:

$$f_6 \sum_{i,j} -p(i,j) \ln(p(i,j)) \tag{6}$$

Object labeling

The objective of this step is to link features computed for a given object to its label. For the purpose of this work, five land cover classes of interest were identified: water, forest, urban, bare soil, and non-dense vegetation. A set of manually labeled regions given by experts were acquired by careful visual interpretation over the studied area. Polygons of the studied area are digitized to derive the thematic information using a topographic map with the scale of 1/50000. Topographic information is used to determine thematic classes in the studied area (Boulila et al. 2011). The labeled regions were divided into training, validation and test data.

At the end of this step, two outputs are provided. The first contains features corresponding to each object and the second contains label of this object. These two

outputs are provided to the neural network to build the classification structure.

Generation of neural network structure

The goal of the neural network structure is to build a process able to determine the class of an object extracted from the RS image according to its features.

In this study, we choose to work with a MLFFNN (Svozil et al. 1997; Ashwini Reddy et al. 2011; Wang et al. 2015a, b; Ulyanov et al. 2016). Our choice is argued due to the ability of MLFFNN to adapt without a continuous assistance of the user. In addition, MLFFNN reduce considerably the computational effort and memory capacity needed to store weights. Moreover, this type of neural network is very robust in the presence of uncertainty and noise, which is the case of RS image field. In the current work, imperfection modeling is not considered. Readers interested in modeling imperfection related to RS images can refer to our previous works. In (Ferchichi et al. 2017a), we detailed sources of imperfection related to RS images and their main types. We proposed to reduce imperfection by using 1) image fusion (Farah et al. 2008; Boulila et al. 2009), 2) imperfection propagation (Boulila et al. 2014; Ferchichi et al. 2017b), and 3) sensitivity analysis (Boulila et al. 2017; Ferchichi et al. 2018).

Figure 2 depicts the proposed MLFFNN architecture. The input are object features and their corresponding land cover types. We have one hidden layer, one output layer and five outputs (different land cover types).

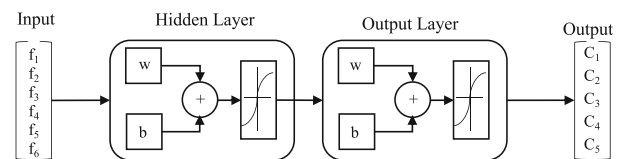
w and b denote, respectively, the parameter (or weight) and the bias associated with the connection between the different units in the two layers.

Down-level: pixel processing

The goal of the down-level is to determine the class of every pixel in an input image according to its features. This is ensured using the already built neural network structure.

Computation of pixel features

The features described at the section (2.c) are object-based features and cannot be computed at the pixel-level. However, our objective in this paper is to determine the class



Object Features
Fig. 2 Proposed MLFFNN architecture

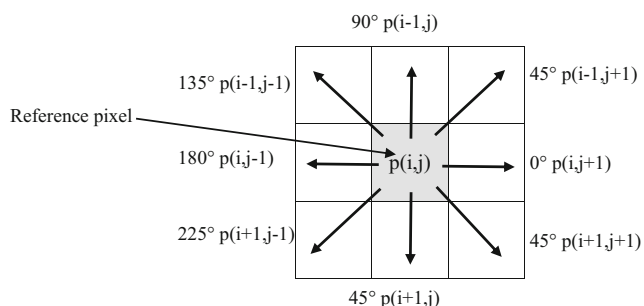


Fig. 3 The 8-connected matrix centered on the reference pixel

of every pixel for a given RS image. To achieve this, we consider an 8-connected matrix centered on that pixel as shown in the Fig. 3. Then, the considered features are computed to this matrix.

The features describing the pixel are the radiometry, the contrast, the correlation, the energy, the homogeneity, and the entropy.

For example, if we consider the following matrix centered at the reference pixel (2,2) as follow:

120	125	127
90	128	127
80	129	129

The value of the radiometry, the contrast, the correlation, the energy, the homogeneity, and the entropy will be, respectively, 128, 1.2597, 0.0493, 0.1136, 0.6444 and -0.0072696 .

Semantic pixel segmentation

Once the MLFFNN is trained, validated and tested, we can use it to determine the class type of every pixel. The features (radiometry, contrast, correlation, energy, homogeneity, and entropy) of every pixel for an input RS image are computed. Then, these features are provided to the MLFFNN structure. Based on the determination of the pixel class, we obtain a semantic image segmentation.

Experimental results

In this section, we firstly describe the dataset used to validate the proposed approach. Then, we present the MLFFNN architecture used in the experimental results. The third part is devoted to the semantic segmentation of RS images. Finally, we evaluate the performance of the proposed approach with regard to traditional classification methods.

Table 1 Land cover classes with number of samples

Class	Land cover type	No. of samples
C1	Water	45160
C2	Forest	247120
C3	Urban	151640
C4	Bare soil	131920
C5	Non-dense vegetation	270560

Dataset description

The proposed approach is tested and evaluated based on Reunion Island site. This site is located in the South-west of the Indian Ocean (21°06' South and 55°32' East; it is 700 km from Madagascar to the West and 180 km from Mauritius to the Northeast). The Reunion Island is 63 km long and 45 km wide.

Experiments are conducted based on a real dataset belonging to the Kalideos¹ database set up by CNES.²

Due to the limitation of sensors and the influence of the atmospheric condition, RS images need, in general, a preprocessing step to enhance the quality of images before any processing subsequent tasks. In the current study, RS images are preprocessed according to: 1) radiometric preprocessing which is achieved by converting the pixel values into reflectance (Chander and Markham 2003; Chander et al. 2009). Then, the inversion of the reflectance at ground level is performed by comparing the estimated reflectance with simulations made at the top of the atmosphere for the geometric and atmospheric conditions corresponding to the measurement, and 2) geometric preprocessing which aims to provide image series which are perfectly coregistered. The goal is to build a reference image through a validation process including field measurements collected by the scientists. After that, a superposition of RS images compared to the reference image is performed in order to refine the corresponding sensor attitude model.

In this paper, experiments have been carried out on a dataset containing a total number of 293 images. The spatial resolution of the data is 10 m per pixel. The size of each image varies from (3000 × 3000 pixels) to (6000 × 6000 pixels). Several thumbnail images are extracted from these images. Then, segmentation is performed to these thumbnail images using k-means algorithm.

The goal of k-means algorithm is to partition the RS images into k segments (objects). It starts by selecting initial cluster centers (known as centroid) randomly for a given image. Then, it assigns each pixel in this image to the segment that has nearest centroid from the respective pixel (in this paper,

¹ <http://kalideos.cnes.fr>

² Centre National d'Etudes Spatiales – Distribution Airbus DS.

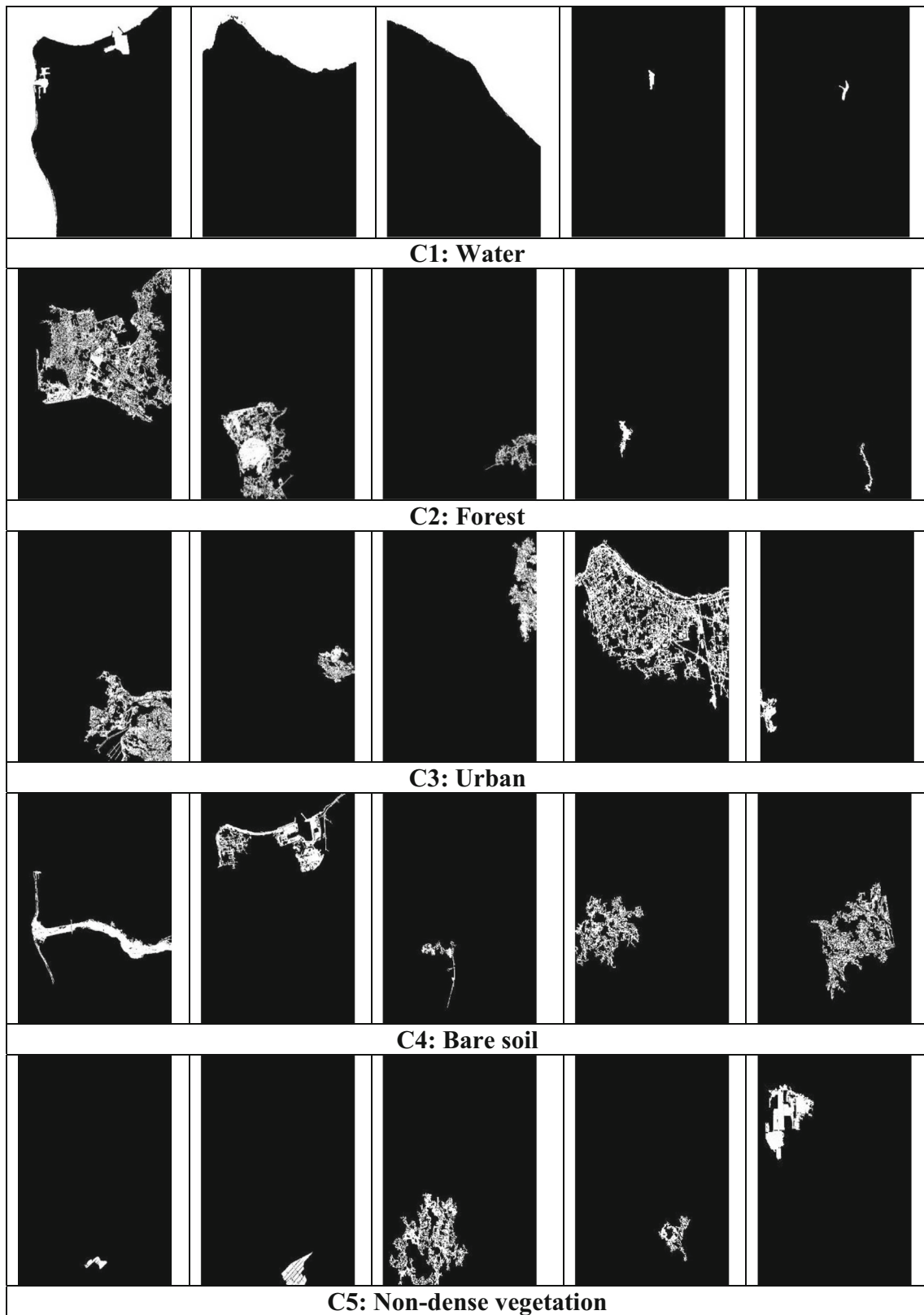
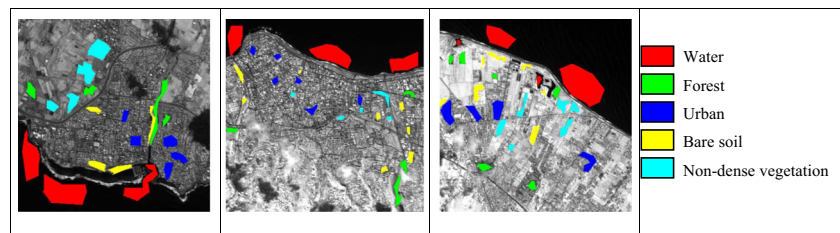


Fig. 4 Excerpt of samples for each land cover type

the Euclidean distance is calculated between each center and each pixel to assign them to the segment having the minimum

distance). Once the segmentation is achieved, the process is taken back to recalculate the new centroid of new segments.

Fig. 5 Ground truth for three thumbnail images extracted from the study region



After that, pixels in the image are reassigned to the new segments. This process is repeated iteratively until a stopping criteria is met (e.g. no pixel changes its cluster, the sum of the distances is minimized, or some maximum number of iterations is reached).

According to the 8-pixels connectivity and to 100 pixels as minimum number of pixels in each extracted object, we obtain a number of labeled samples depicted in the Table 1.

Figure 4 presents an excerpt of samples for each land cover type.

Figure 5 describes ground truth images for three thumbnail images extracted from the study region. To get the ground truth images, information was extracted by experts over the studied areas. Polygons of studied regions of Reunion Island are digitized to derive the thematic information using a topographic map with the scale of 1/50000. Topographic information is used to determine thematic classes in the studied areas. Five thematic classes are identified which are the following: urban, water, forest, bare soil, and non-dense vegetation areas.

MLFFNN description

We used `nprtool` provided by Matlab R2008a (`nprtool` 2018). This tool uses a function named `patternnet` to classify RS images. This function is based on a feed-forward network that is trained to classify pixels according to

target classes. The `patternnet` has three input parameters which are `hiddenLayerSizes`, `trainFcn` and `performFcn`, and returns a pattern recognition neural network. `hiddenLayerSizes` are hidden layer sizes set to 10 in the present paper.

`trainFcn` is the training function set to `trainbr`. `trainbr` is the bayesian regularization backpropagation. It is a network training function that updates the weight and bias values according to Levenberg-Marquardt optimization (Levenberg 1944). It minimizes a combination of squared errors and weights, and then determines the correct combination to produce a network that generalizes well.

`performFcn` is the performance function set to `crossentropy`. It calculates the MLFFNN performance given targets and outputs.

The data used in this paper is divided into 70% for training (205 images), 15% for validation (44 images) and 15% for testing (44 images). The objective of the validation set is to monitor the classification error and stop training before overfitting occurs. The test set is then used independently to evaluate the classification quality (Heisel et al. 2017).

The training process of MLFFNN is performed iteratively 100 times on NVIDIA's GeForce GTX 1080 with 8GB of GPU memory.

Figure 6 describes the performance plot of the MLFFNN which shows the training, validation and testing errors. It can be noted from this figure that the best validation performance

Fig. 6 Performance plot of the MLFFNN

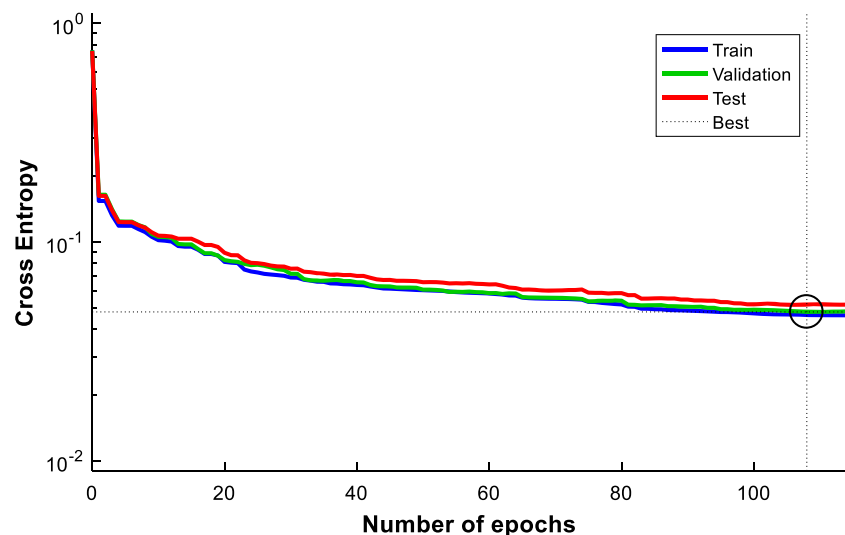
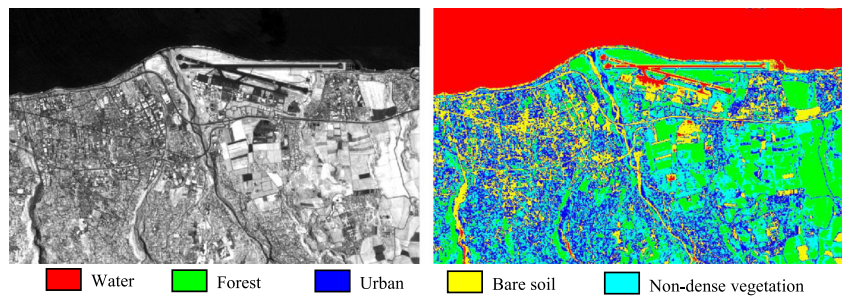


Fig. 7 Satellite image acquired on January 31, 2015 (left) and the semantic image segmentation performed by the proposed approach (right)



was achieved at epoch 108 with an error rate of 0.047969. Moreover, the validation and test curves are very similar which implies there was no significant overfitting occurred (Samuel et al. 2017).

Semantic segmentation of RS images

The goal of this section is to illustrate the applicability of the proposed approach for semantic image segmentation.

Figure 7 (left) represents an image taken from the previously described Kalideos database. This image is not among the training dataset. The image is acquired on January 31, 2015 and coming from SPOT 5 (Satellite Pour L’observation de la Terre) satellite. The considered image has a spatial resolution of 10 m and a size of 800 × 500 pixels. Figure 7 (right) depicts the semantic image segmentation performed by the proposed approach.

The goal of this section is to perform a semantic segmentation of RS images using the proposed approach. Then, results of the semantic segmentation are compared to the ground truth image representing the same region at the same date. The comparison is carried out using two criteria: overall accuracy (OA) and kappa coefficient (K). OA is the sum of the correctly classified pixels divided by the total number of image pixels. K is an accuracy measure that compares proposed results of classification to the real ones. It takes values from zero to one (higher values of kappa coefficient means a good classification). K is defined as follow (Congalton and Green 2008):

$$K = \frac{n \sum_{i=1}^k n_{ii} - \sum_{i=1}^k n_{i+} n_{+i}}{n^2 - \sum_{i=1}^k n_{i+} n_{+i}} \tag{7}$$

Where.

- k denotes the number of classes.
- n is the total number of pixels in images.
- n_{ii} is the sum of correctly classified pixels for the class i (the number of pixels belonging to class i in the ground truth that have also been classified as class i in the classified image).
- n_{i+} is the sum of pixels classified into class i in the proposed image classification.
- n_{+i} is the number of pixels classified into class i in the ground truth image.

Table 2 depicts the confusion matrix of the proposed semantic segmentation for the image presented in the Fig. 7. Rows denote classes for the ground truth image, whereas columns represent classes for the proposed semantic image segmentation. As indicated in Table 2, the proposed approach performs a good semantic segmentation of the image with an OA = 91.85% and a K = 0.8982.

Evaluation of the proposed approach

To further evaluate the performance of the top-down approach for semantic image segmentation, we compare results of the proposed approach with well-used traditional classification methods. The comparison includes SVM (Support Vector

Table 2 Confusion matrix of the proposed semantic image segmentation

		Ground truth image					User Accuracy
		Water	Forest	Urban	Bare soil	Non-dense vegetation	
Semantic image segmentation	Water	94.31	0.45	2.05	2.29	0.9	94.31
	Forest	0.09	89.05	2.75	3.01	5.1	89.05
	Urban	1.3	0.26	90.85	4.56	3.03	90.85
	Bare soil	2.09	0.2	4.5	91.18	2.03	91.18
	Non-dense vegetation	0.04	2.71	1.62	1.75	93.88	93.88
Producer Accuracy		96.40	96.09	89.26	88.70	89.46	
Overall Classification Accuracy = 91.85, Kappa = 0.8982							

Table 3 Comparison of the classification accuracy between the proposed method, SVM, and MLC

Method	OA	K
SVM	88.34	0.8590
MLC	86.11	0.8277
MLFFNN	91.85	0.8982

Machines) (Huang et al. 2002; Mitra et al. 2004) and Maximum Likelihood Classification (MLC) (Bruzzone and Prieto 2001; Murthy et al. 2003).

Table 3 depicts a comparison of image classification between SVM, MLC, and the MLFFNN according to the overall classification and kappa coefficient. As we note, the proposed approach outperforms the two others methods for the image presented in Fig. 7.

Figure 8 illustrates the overall accuracies of image classification according to the training set size for the three methods: SVM, MLC and the proposed method. The size of training set varies between 100 to 800,000 samples. The important observation from this figure is that all the three methods was positively influenced by the size of the training set. Both SVM and the proposed approach provide the best results in all cases. The OA moves from 79.8% (case of 100 samples) to 87.4% (case of 800,000 samples) for the SVM method. Whereas, OA moves from 71.5% (case of 100 samples) to 84.1% (case of 800,000 samples) for the MLC method and from 74.2% (case of 100 samples) to 91.6% (case of 800,000 samples) for the proposed method. Further, the proposed approach provides the best results especially when the training size set become more important (when the size of the training set is greater than 200,000).

Figure 9 describes the error in image classification between the SVM, MLC, and the proposed approach. We can note that

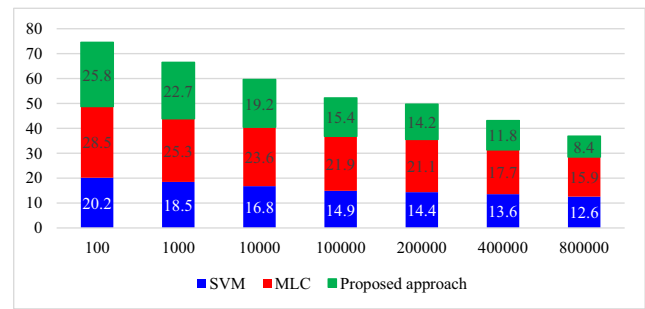


Fig. 9 Performance plot of the MLFFNN

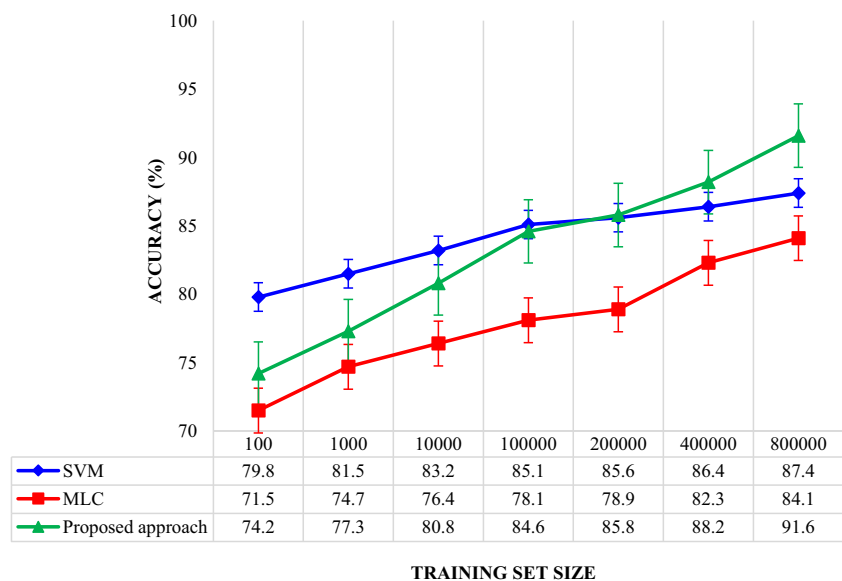
the SVM is less sensitive to the size of the training set with a difference of 7.6% between the size of 100 and 800,000 samples. The MLC comes in second place with a difference of 12.6% and the proposed approach in the third place with a difference of 17.4%.

The MVC classifier outperforms the MLC method in all situations regardless the size of the training set. Although SVM provides good results for image classification but a large training set may not be very useful for it to work. This observation is compatible with results reported in the literature (Huang et al. 2002; Foody and Mathur 2004). However, in big RS data, the volume of data plays a very important role in semantic image segmentation. Thus, the use of the proposed method in the case of RS big data is more appropriate. Further, our method consistently provides good results even with the training set is very small.

Conclusion

In this paper, we proposed a new approach for semantic segmentation of RS big images. The main idea is to use features

Fig. 8 Performance plot of the MLFFNN



calculated at the object-level to determine the class of every pixel in a new image. To do this an 8-connected matrix centered on that pixel is considered. Determining the pixel class is achieved using a MLFFNN. Input for the neural network are object features and different class labels. The output is the structure that is used for semantic segmentation.

Experimental results were carried out based on a real dataset belonging to the Kalideos database. The classification results obtained by the proposed approach show significant improvements in both the overall and categorical classification accuracies. Besides, comparison with state-of-the-art classification methods proves that the proposed approach provides good performances especially when the volume of data become important.

However, despite of the promising results obtained by the proposed approach, several issues can be addressed in future such as the determination of number of hidden layer nodes and their respective weights. Another challenging topic to be explored is the effect of 8-connected matrix centered on the pixel on the determination of the class label.

Acknowledgements The author would like to thank the anonymous reviewers for their valuable comments which were very helpful to improve the manuscript.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- [nprtool, 18] (2018) <https://www.mathworks.com/help/nnet/ref/nprtool.html>. [Accessed: 08- Jan- 2018]
- Akçay HG, Aksoy S (2008) Automatic detection of geospatial objects using multiple hierarchical segmentations. *IEEE Trans Geosci Remote Sens* 46(7):2097–2111
- Andrés S, Arvor D, Mougenot I, Libourel T, Durieux L (2017) Ontology-based classification of remote sensing images using spectral rules. *Comput Geosci* 102:158–166
- Ashwini Reddy T, Renuka Devi K., Gangashetty SV (2011) Multilayer Feedforward Neural Network Models for Pattern Recognition Tasks in Earthquake Engineering, International Conference on Advanced Computing, Networking and Security, pp. 154–162
- Athanasiadis T, Mylonas P, Avrithis Y, Kollias S (2007) Semantic image segmentation and object labeling. *IEEE Trans Circuits Syst Video Technol* 17(3):298–312
- Boulila W, Farah IR, Saheb Etabaa K, Solaiman B, Ben Ghézala H (2009) Improving spatiotemporal change detection: a high level fusion approach for discovering uncertain knowledge from satellite image databases. *ICDM* 58:222–227
- Boulila W, Farah IR, Saheb Etabaa K, Solaiman B, Ben Ghézala H (2010) Spatio-temporal modeling for knowledge discovery in satellite image databases, CORIA CONFérence en Recherche d'Information et Applications, pp. 35–49
- Boulila W, Farah IR, Saheb Etabaa K, Solaiman B, Ben Ghézala H (2011) A data mining based approach to predict spatiotemporal changes in satellite images. *Int J Appl Earth Obs Geoinf* 13(3): 386–395
- Boulila W, Bouatay A, Farah IR (2014) A probabilistic collocation method for the imperfection propagation: application to land cover change prediction. *J Multimedia Process Technol* 5(1):12–32
- Boulila W, Ayadi Z, Farah IR (2017) Sensitivity analysis approach to model epistemic and aleatory imperfection: application to land cover change prediction model. *J Comput Sci* 23:58–70
- Boulila W, Farah IR, Hussain A (2018) A novel decision support system for the interpretation of remote sensing big data. *Earth Sci Inf* 11(1): 31–45
- Bruzzone L, Prieto DF (2001) Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 2, pp. 456–460
- Chander G, Markham B (2003) Revised Landsat-5 TM radiometric calibration procedures and postcalibration dynamic ranges. *IEEE Trans Geosci Remote Sens* 41(11):2674–2677
- Chander G, Markham BL, Helder DL (2009) Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sens Environ* 113(5):893–903
- Cheng G, Han J (2016) A survey on object detection in optical remote sensing images. *J Photogramm Remote Sens* 117:11–28
- Congalton R, Green K (2008) Assessing the accuracy of remotely sensed data: principles and practices, 2nd edn. CRC Press, Taylor & Francis, Boca Raton
- Connors RW, Trivedi MM, Harlow CA (1984) Segmentation of a high-resolution urban scene using texture operators. *Comput Vis Graph Image Process* 25:273–310
- Diao W, Sun X, Dou F, Yan M, Wang H, Fu K (2015) Object recognition in remote sensing images using sparse deep belief networks. *IEEE Geosci Remote Sens Lett* 6(10):745–754
- Durand N, Derivaux S, Forestier G (2007) Ontology-Based Object Recognition for Remote Sensing Image Interpretation, *IEEE International Conference on Tools with Artificial Intelligence*, pp. 472–479
- Farah IR, Boulila W, Saheb Etabaa K, Solaiman B, Ben Ahmed M (2008) Interpretation of multi-sensor remote sensing images: Multi-approach fusion of uncertain information. *IEEE Trans Geosci Remote Sens* 46(12)
- Ferchichi A, Boulila W, Farah IR (2017a) Propagating aleatory and epistemic uncertainty in land cover change prediction process. *Eco Inform* 37:24–37
- Ferchichi A, Boulila W, Farah IR (2017b) Towards an uncertainty reduction framework for land-cover change prediction using possibility theory. *Vietnam J Comput Sci* 4(3):195–209
- Ferchichi A, Boulila W, Farah IR (2018) Reducing uncertainties in land cover change models using sensitivity analysis. *Knowl Inf Syst* 55(3):719–740
- Foody GM, Mathur A (2004) A relative evaluation of multiclass image classification by support vector machines. *IEEE Trans Geosci Remote Sens* 42(6):1335–1343
- Haiyang Y, Fuping G (2009) Object recognition of high resolution remote sensing image based on PSWT, International Conference on Image Analysis and Signal Processing, pp. 52–56
- Haralick RM, Shapiro LG (1985) Image segmentation techniques. *Comput Vis Graph Image Process* 29(1):100–132
- Haralick RM, Shanmugam K, Dinstein I (1973) Textural features for image classification. *IEEE Trans Syst Man Cybern SMC-3*:610–621
- Heisel S, Kovačević T, Briesen H, Schembecker G, Wohlgemuth K (2017) Variable selection and training set design for particle classification using a linear and a non-linear classifier. *Chem Eng Sci* 173: 131–144
- Huang C, Davis LS, Townshend JRG (2002) An assessment of support vector machines for land cover classification. *Int J Remote Sens* 23(4):725–749

- Lei Z, Fang T, Huo H, Li D (2012) Rotation-invariant object detection of remotely sensed images based on texton forest and Hough voting. *IEEE Trans Geosci Remote Sens* 50(4):1206–1217
- Levenberg K (1944) A method for the solution of certain non-linear problems in least squares. *Q Appl Math* 2:164–168
- Liu J, Li J, Li W, Wu J (2016) Rethinking big data: a review on the data quality and usage issues. *ISPRS J Photogramm Remote Sens* 115:134–142
- Ma X, Shen H, Yang J, Zhang L, Li P (2014) Polarimetric-spatial classification of SAR images based on the fusion of multiple classifiers. *IEEE J. Sel Topics Appl Earth Observ* 7(3):961–971
- MacQueen JB (1967) Some methods for classification and analysis of multivariate observations, Berkeley symposium on mathematical statistics and probability, University of California Press, pp. 281–297
- Marmanis D, Schindler K, Wegner JD, Galliani S, Datcu M, Stilla U (2018) Classification with an edge: improving semantic image segmentation with boundary detection. *J Photogramm Remote Sens* 135:158–172
- Ming D, Li J, Wang J, Zhang M (2015) Scale parameter selection by spatial statistics for GeOBIA: using mean-shift based multi-scale segmentation as an example. *ISPRS J Photogramm Remote Sens* 106:28–41
- Mitra P, Uma Shankar B, Pal SK (2004) Segmentation of multispectral remote sensing images using active support vector machines. *Pattern Recogn Lett* 25:1067–1074
- Murthy CS, Raju PV, Badrinath KVS (2003) Classification of wheat crop with multi-temporal images: performance of maximum likelihood and artificial neural networks. *Int J Remote Sens* 24(23):4871–4890
- Rau JY, Jhan JP, Rau RJ (2014) Semiautomatic object-oriented landslide, Recognition Scheme From Multisensor Optical Imagery and DEM. *IEEE Trans Geosci Remote Sens* 52(2):1336–1349
- Ryherd S, Woodcock C (1996) Combining spectral and texture data in the segmentation of remotely sensed images. *Photogramm Eng Remote Sens* 62(2):181–194
- Samuel OW, Asogbon GM, Sangaiah AK, Fang P, Li G (2017) An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction. *Expert Syst Appl* 68:163–172
- Shotton J, Kohli P (2014) Semantic Image Segmentation, *Computer Vision: A Reference Guide*, Springer, pp. 713–716
- Svozil D, Kvasnicka V, Pospichal J (1997) Introduction to multi-layer feed-forward neural networks. *Chemom Intell Lab Syst* 39(1):43–62
- Trias-Sanz R, Stamon G, Louchet J (2008) Using colour texture and hierarchical segmentation for high-resolution remote sensing. *J Photogramm Remote Sens* 63(2):156–168
- Ulyanov D, Lebedev V, Vedaldi A, Lempitsky V (2016) Texture networks: feed-forward synthesis of textures and stylized images. *ICML* 48:1349–1357
- Vincent L (1993) Grayscale area openings and closings, their efficient implementation and applications, Workshop on Mathematical Morphology and Its Applications to Signal Processing, J. Serra and P. Salembier, Eds., Barcelona, Spain, pp. 22–27
- Wang S, Zhang Y, Dong Z, Du S, Ji G, Yan J, Yang J, Wang Q, Feng C, Phillips P (2015a) Feed-forward neural network optimized by hybridization of PSO and ABC for abnormal brain detection. *Int J Imaging Syst Technol* 25(2):153–164
- Wang S, Zhang Y, Ji G, Yang J, Wu J, Wei L (2015b) Fruit classification by wavelet-entropy and feedforward neural network trained by fitness-scaled chaotic ABC and biogeography-based optimization. *Entropy* 17(8):5711–5728
- Yang X, Tridandapani S, Beitler JJ, Yu DS, Yoshida EJ, Curran WJ, Liu T (2012) Ultrasound GLCM texture analysis of radiation-induced parotid-gland injury in head-and-neck cancer radiotherapy: an in vivo study of late toxicity. *Med Phys* 39(9):5732–5739
- Zhang J, Li T, Lu X, Cheng Z (2016a) Semantic classification of high-resolution, Remote-Sensing Images Based on Mid-level Features. *IEEE J. Sel Topics Appl Earth Observ* 9(6):2343–2353
- Zhang L, Zhang L, Du B (2016b) Deep learning for remote sensing data: a technical tutorial on the state of the art. *IEEE Geosci Remote Sens Mag* 4(2):22–40
- Zhang T, Yan W, Li J, Chen J (2016c) Multiclass labeling of very high-resolution remote sensing imagery by enforcing nonlocal shared constraints in multilevel conditional random fields model. *IEEE J. Sel Topics Appl Earth Observ* 9(7):2854–2867
- Zhang AZ, Sun GY, Liu SH, Wang ZJ, Wang P, Ma JS (2017) Multi-scale segmentation of very high resolution remote sensing image based on gravitational field and optimized region merging. *Multimed Tools Appl* 76(13):15105–15122
- Zheng C, Zhang Y, Wang L (2017) Semantic segmentation of remote sensing imagery using an object-based Markov random field model with auxiliary label fields. *IEEE Trans Geosci Remote Sens* 55(5):3015–3028