

A Deep Rule-based Approach for Satellite Scene Image Analysis

Xiaowei Gu¹, *Student Member, IEEE* and Plamen P. Angelov^{1,2*}, *Fellow, IEEE*

1. Data Science Group,
School of Computing and Communications,
Lancaster University, Lancaster, UK

2. Honorary Professor
Technical University,
Sofia, 1000, Bulgaria

*E-mail: p.angelov@lancaster.ac.uk

Abstract— Satellite scene images contain multiple sub-regions of different land use categories; however, traditional approaches usually classify them into a particular category only. In this paper, a new approach is proposed for automatically analyzing the semantic content of sub-regions of satellite images. At the core of the proposed approach is the recently introduced deep rule-based image classification method. The proposed approach includes a self-organizing set of transparent zero order fuzzy IF-THEN rules with human-interpretable prototypes identified from the training images and a pre-trained deep convolutional neural network as the feature descriptor. It requires a very short, nonparametric, highly parallelizable training process and can perform a highly accurate analysis on the semantic features of local areas of the image with the generated IF-THEN rules in a fully automatic way. Examples based on benchmark datasets demonstrate the validity and effectiveness of the proposed approach.

Keywords—deep learning, deep fuzzy rule-based classifier, image analysis.

I. INTRODUCTION

Remote sensing images can provide a detailed land use information about the surface of the Earth and play instrumental role in many applications such as urban development, nature conservation, disaster assessment [1], [2]. Due to the very large volume of high-resolution remote sensing images, automatically recognizing and classifying the semantic content of these images become very important [3]. However, this is a very challenging task because of the complexity and variability of the scene imagery.

Moreover, it is commonly seen that satellite scene images, especially the large size ones, contain multiple sub-regions of different land use classes. For such images, classifying them into a particular single class is not the best solution. On the

other hand, manually analyzing and understanding such a single image can be quite difficult and time consuming. Therefore, it is of great importance to develop an automatic approach for analyzing and understanding satellite scene images.

Deep convolutional neural networks (DCNNs) are the state-of-the-art approaches in the field of machine learning and computer vision [4], [5], and they have demonstrated very high performance on remote sensing scene classification [3], [6], [7]. Compared with the traditional low- and medium-level feature descriptors, i.e. GIST [8], HOG [9], SIFT [10], etc., the DCNNs are capable of capturing more abstract and discriminative high-level semantic features and, therefore, the DCNN-based approaches are able to achieve greater accuracy on the land use scene classification [1]. Nonetheless, the success comes at the price of losing transparency, a lot of computational resources and time. In addition, these approaches require a full re-training if new pattern appears.

Deep Rule-Based (DRB) system was initially introduced in [11], [12] as a deep learning approach for image classification, and it has been successfully applied in different problems including handwritten digits recognition, face recognition, object recognition and remote sensing scene classification, etc. Involving the widely used image transformation and computer vision techniques, the DRB system is able to self-organize a set of transparent, massively parallel zero-order AnYa type IF-THEN rules [13], [14] consisting of visual prototypes identified from the training image set in a nonparametric, non-iterative and transparent way. Furthermore, the DRB system is able to offer highly accurate classification results even starting from the very first image of each class, just in the same way as humans do [11], [12].

In this paper, a DRB approach is proposed for satellite

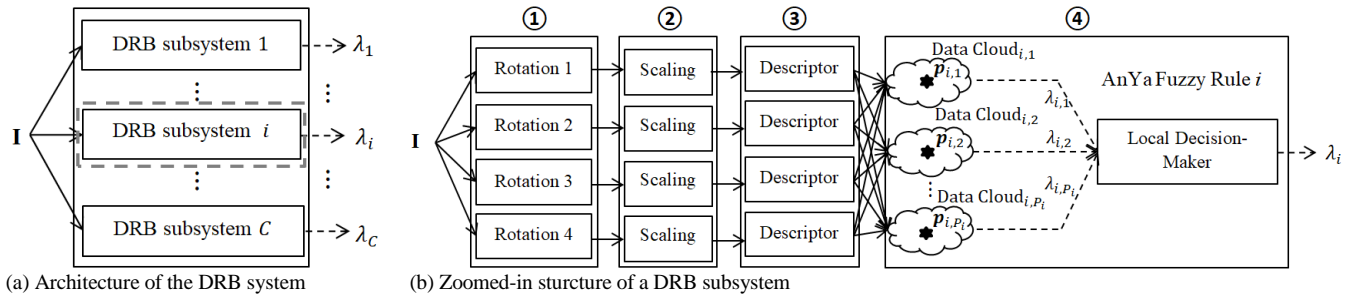


Fig. 1. General architecture of the DRB system

scene image analysis, which takes our previous works one step further by conducting classification on local regions of the images instead of the whole images themselves [11], [12] and results in a deeper understanding of the semantic contents on them. It employs a pre-trained DCNN [15] for feature extraction and thus, is able to grasp more abstract and discriminative semantic features within the satellite scene images, meanwhile, avoids the time- and resource-consuming training. Thanks to its prototype-based nature, the proposed approach can perform the analysis on the semantic contents in the local areas of the image in a fully autonomous way after a highly transparent, efficient and parallelizable training process. Examples based on benchmark datasets demonstrate the strong ability of the proposed DRB approach in understanding satellite scene images.

II. DEEP RULE-BASED CLASSIFIER

In this section, the general architecture and the IF-THEN rule identification process of the DRB classifier are briefly described.

A. General Architecture

The general architecture of the DRB classifier used in this paper is depicted in Fig. 1. From Fig. 1(a) one can see that the DRB classifier consists of C subsystems corresponding to C different categories/classes in the training set. Each DRB subsystem is trained in parallel with others and there is no interdependence between any two of them. Thus, one can update or remove the DRB subsystems freely without influencing other DRB subsystems. The zoomed-in internal structure of a DRB subsystem is presented in Fig. 1(b), which has the following layers:

- ① Rotation layer;
- ② Scaling layer;
- ③ Feature descriptor layer
- ④ Fuzzy rule layer.

The rotation layer rotates each training image at four different angles 1) 0°; 2) 90°; 3) 180° and 4) 270°. The rotation operation is very important for improving the generalization ability of the DRB system. The scaling layer is for resizing the training image into the size that the feature descriptor requires. The pre-trained VGG-VD-16 [15], which is one of the best performed pre-trained DCNN models on remote sensing scene classification, is employed by the DRB system as the feature descriptor due to its simple structure and high performance [1]. In this paper, there is no further tuning involved, and it has to be stressed that alternative feature descriptor can be used as well, i.e. GIST [8], HOG [9], GoogLeNet [16], or VGG-VD-19 [15], etc. As the pre-trained VGG-VD-16 requires the input image to be with size of 227×227 pixels, the scaling layer transforms each training image into this uniform size. Following the common practice, the 1×4096 dimensional activations from the first fully connected layer are extracted as the feature vector of each training image (one 1×4096 dimensional vector per image).

The core of each DRB subsystem is the zero order fuzzy rule of AnYa type [13], [14]. Each subsystem contains one massively parallel IF-THEN rule set formulated around the

prototypes generalized or learned from the images of the corresponding class. These rules have the following form:

$$\begin{aligned} & IF(\mathbf{I} \sim \mathbf{P}_{i,1}) OR (\mathbf{I} \sim \mathbf{P}_{i,2}) OR \dots OR (\mathbf{I} \sim \mathbf{P}_{i,N_i}) \\ & THEN(Class^i) \end{aligned} \quad (1)$$

where \sim denotes similarity, which can also be seen as a fuzzy degree of membership; \mathbf{I} is a particular image, and its corresponding feature vector is denoted by \mathbf{x} ; $\mathbf{P}_{i,j}$ ($j = 1, 2, \dots, N_i, i = 1, 2, \dots, C$) denotes the j^{th} prototype of the i^{th} class with the corresponding feature vector $\mathbf{p}_{i,j}$; N_i corresponds to the number of prototypes of the i^{th} class.

Each IF-THEN rule contains a number of prototypes identified from the training images from the same class, which are connected by a local decision-maker using the “winner-takes-all” principle. Therefore, each AnYa type fuzzy rule can be represented as a massively parallel series of simpler fuzzy rules with a single prototype connected by a logical “OR” operator.

In the next subsection, the identification process of the AnYa type IF-THEN rules used in the proposed DRB classifier will be described.

B. IF-THEN Rules Identification

The identification process of the AnYa type IF-THEN rules is autonomous, nonparametric, transparent, self-evolving and highly efficient. The detailed identification process has been given in [11], [12]. To make this paper self-contained, the main procedure of the identification process of the i^{th} IF-THEN rule ($i = 1, 2, \dots, C$) is summarized in the form of the following pseudo-code.

While new feature vector \mathbf{x}_k of the i^{th} class is available

i. $\mathbf{x}_k \leftarrow \mathbf{x}_k / \|\mathbf{x}_k\|$ ($\|\mathbf{x}_k\|$ is the norm of \mathbf{x}_k);

ii. **If** ($k = 1$) **Then**

1. $\boldsymbol{\mu}_i \leftarrow \mathbf{x}_k$ ($\boldsymbol{\mu}_i$ is the global mean);
2. $N_i \leftarrow 1$ (N_i is the number of prototypes);
3. $\mathbf{p}_{i,1} \leftarrow \mathbf{x}_k$ ($\mathbf{p}_{i,1}$ is the first prototype);
4. $S_{i,1} \leftarrow 1$ ($S_{i,1}$ is the corresponding support);

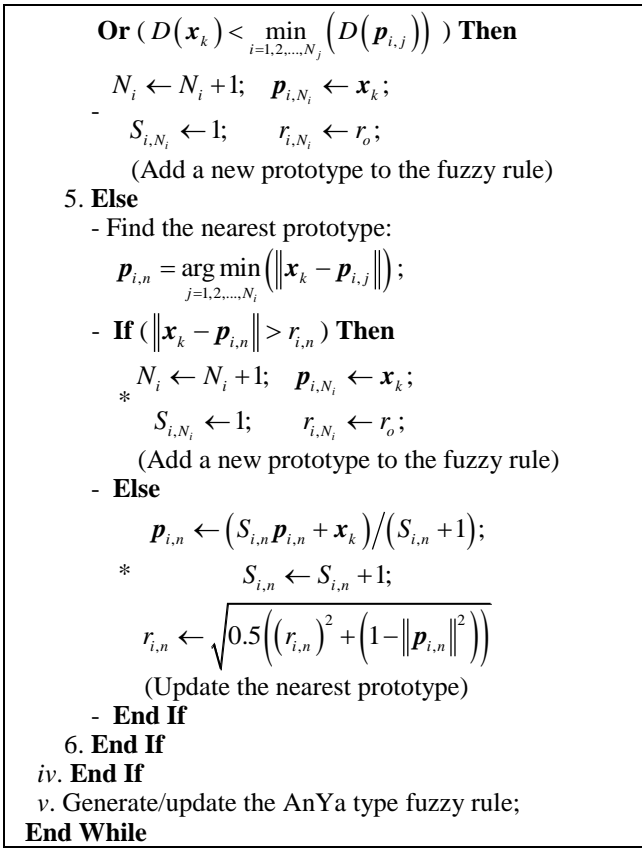
5. $r_{i,1} \leftarrow r_o$ ($r_{i,1}$ is the radius, $r_o = \sqrt{2 - 2\cos\left(\frac{\pi}{6}\right)}$);

iii. **Else**

1. Update global mean: $\boldsymbol{\mu}_i \leftarrow \frac{k-1}{k} \boldsymbol{\mu}_i + \frac{1}{k} \mathbf{x}_k$;
2. Calculate the density of \mathbf{x}_k [17], [18]:

$$D(\mathbf{x}_k) = 1 / \left(1 + \|\mathbf{x}_k - \boldsymbol{\mu}_i\|^2 / (1 - \|\boldsymbol{\mu}_i\|^2) \right);$$
3. Update the densities of prototypes ($j = 1, \dots, N_i$):

$$D(\mathbf{p}_{i,j}) = 1 / \left(1 + \|\mathbf{p}_{i,j} - \boldsymbol{\mu}_i\|^2 / (1 - \|\boldsymbol{\mu}_i\|^2) \right);$$
4. **If** ($D(\mathbf{x}_k) > \max_{j=1,2,\dots,N_i} (D(\mathbf{p}_{i,j}))$)



Once the IF-THEN rules are identified, the image analysis can be performed.

III. SATELLITE IMAGE ANALYSIS

As it has been stated in section I, a large size satellite image usually contains multiple sub-regions of different land use classes. Instead of classifying it into one of the certain classes, in this section, we will describe the new approach for automatically analyzing and understanding better these satellite images. The diagram of the image analysis process is given in Fig. 2.

A. Local Information Extraction

For a certain large-size satellite image, one of the most commonly used ways for extracting sub-regions is to partition the image with a grid net. In this paper, we use the $V \times H$ grid net to separate each image into $V \times H$ segments evenly.

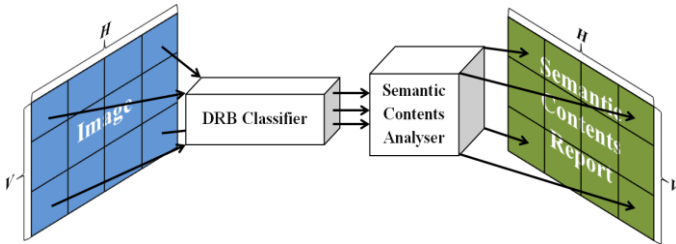


Fig. 2. Diagram of the proposed approach for image analysis

However, it has to be stressed that other segmentation methods can also be considered.

After the large size image is partitioned into $V \times H$ local areas, the trained DRB classifier scans segments one by one. In this stage, the rotation layer (layer ① in Fig. 1(b)) of the DRB classifier is not needed anymore.

For each segment, a $1 \times C$ dimensional vector of scores of confidence, denoted by $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_C]$, is produced as follows.

$$\lambda_i(\mathbf{sg}) = \max_{j=1,2,\dots,N_i} \left(\exp \left(-\|\mathbf{y} / \|\mathbf{y}\| - \mathbf{p}_{i,j}\|^2 \right) \right) \quad (2)$$

where \mathbf{sg} denotes the segment; \mathbf{y} is the feature vector of the segment extracted by the feature descriptor.

As the image has been divided into $V \times H$ local areas, the DRB system will pass $V \times H$ confidence score vectors to the semantic contents analyzer for further evaluation. As the locations of the sub-regions are also very important, their coordinates on the grid net will also be sent.

B. Local Information Analysis

After the semantic contents analyzer received the information from the DRB system, for each segment, \mathbf{sg} , the semantic contents analyzer firstly identifies the top three scores indicating the most likely three land use categories this particular local area belongs to, denoted by $\lambda_{\max}^{(1)}(\mathbf{sg})$, $\lambda_{\max}^{(2)}(\mathbf{sg})$ and $\lambda_{\max}^{(3)}(\mathbf{sg})$ ($\lambda_{\max}^{(1)}(\mathbf{sg}) \geq \lambda_{\max}^{(2)}(\mathbf{sg}) \geq \lambda_{\max}^{(3)}(\mathbf{sg})$).

Then, the Condition 1 is examined to see whether the analyzer is able to declare the land use category of this segment with a high confidence.

$$\text{Cond.1. } IF \left(\lambda_{\max}^{(1)}(\mathbf{sg}) > \varepsilon_1 \right) AND \left(\lambda_{\max}^{(1)}(\mathbf{sg}) > \varepsilon_2 \cdot \lambda_{\max}^{(2)}(\mathbf{sg}) \right) \\ THEN \left(\mathbf{sg} \text{ belongs to } Class_{\max}^{(1)} \right) \quad (3)$$

If Condition 1 is met, it indicates that this local area is mainly used for this particular purpose. Otherwise, Condition 2 is checked:

$$\text{Cond.2. } IF \left(\lambda_{\max}^{(2)}(\mathbf{sg}) > \varepsilon_2 \cdot \lambda_{\max}^{(3)}(\mathbf{sg}) \right) \\ THEN \left(\mathbf{sg} \text{ belongs to } Class_{\max}^{(1)} \text{ and } Class_{\max}^{(2)} \right) \quad (4)$$

In this paper, we use $\varepsilon_1 = 0.4$ and $\varepsilon_2 = 1.2$. However, we have to stress that the free parameters ε_1 and ε_2 can be adjusted to meet the specific requirements of different applications.

If the segment satisfies Condition 2, it means that this local area is a mixture of two types of land uses. And the percentage of each type is given as:

$$pct_j = \frac{\lambda_{\max}^{(j)}(\mathbf{sg})}{\lambda_{\max}^{(1)}(\mathbf{sg}) + \lambda_{\max}^{(2)}(\mathbf{sg})}; \quad j = 1, 2 \quad (5)$$



Fig. 3. Examples of the UCMerced Image Set

If Condition 2 is not satisfied, it indicates that this particular segment has the distinctive features of three different land use categories ($Class_{\max}^{(1)}$, $Class_{\max}^{(2)}$ and $Class_{\max}^{(3)}$), and the semantic contents analyzer will report the three categories for this segment and give the corresponding percentages:

$$pct_j = \frac{\lambda_{\max}^{(j)}(\mathbf{sg})}{\lambda_{\max}^{(1)}(\mathbf{sg}) + \lambda_{\max}^{(2)}(\mathbf{sg}) + \lambda_{\max}^{(3)}(\mathbf{sg})}; \quad j = 1, 2, 3 \quad (6)$$

After the analyzer has analyzed all the segments, a detailed report of the results will be given in the form of a $V \times H$ table with the land use categories and the respective percentages of each segment tabulated in the corresponding cell (see Figs. 5-7).

IV. APPLICATION EXAMPLES

In this section, examples on satellite image analysis are given. The experiments were conducted using MATLAB R2017a on a PC with dual core processor 3.60 GHz \times 2 and 16 GB RAM.

Firstly, we use the UCMerced image set [19] to train the DRB system. UCMerced image set is one of the most widely used benchmark dataset in the field of remote sensing. It consists of fine spatial resolution remote sensing images of 21 commonly seen land use categories (see Fig. 3), which is sufficient for demonstrating the core idea of this paper. Each category of the UCMerced dataset contains 100 images with the uniform image size of 256 \times 256 pixels. Due to the relatively smaller image size, the images of this dataset have simpler structures and show less variety in terms of the semantic contents within a single image.

Nonetheless, one may find that some categories share quite similar semantic features, i.e. the images of category “tennis court” have similar local features as the images from the categories “dense residential”, “parking lot”, “freeway”, etc. As the similar features between different classes will create confusion for the analyzer, three categories (“baseball diamond”, “mobile home park” and “tennis court”) are removed to avoid the significant overlaps. We also rename the category “golf course” to “meadow” for a better interpretability about the results.

After the DRB system is trained on the images of the 18 categories, one can get 18 AnYa type zero order IF-THEN rules consisting of a large number of prototypes identified from the images. The illustrative examples of the 18 IF-THEN rules are visualized in Table I. The respective time

consumptions for training each rule is tabulated in Table II. One can see that, the training process is highly efficient, it takes maximum 3.5 seconds to training one fuzzy rule with

TABLE I. ILLUSTRATIVE EXAMPLE OF THE IF-THEN RULES

#	Fuzzy Rules			
1		OR		THEN (Agricultural)
2		OR		THEN (Airplane)
3		OR		THEN (Beach)
4		OR		THEN (Buildings)
5		OR		THEN (Chaparral)
6		OR		THEN (Dense Residential)
7		OR		THEN (Forest)
8		OR		THEN (Freeway)
9		OR		THEN (Meadow)
10		OR		THEN (Harbour)
11		OR		THEN (Intersection)
12		OR		THEN (Medium Residential)
13		OR		THEN (Overpass)
14		OR		THEN (Parking Lot)
15		OR		THEN (River)
16		OR		THEN (Runway)
17		OR		THEN (Sparse Residential)
18		OR		THEN (Storage Tanks)

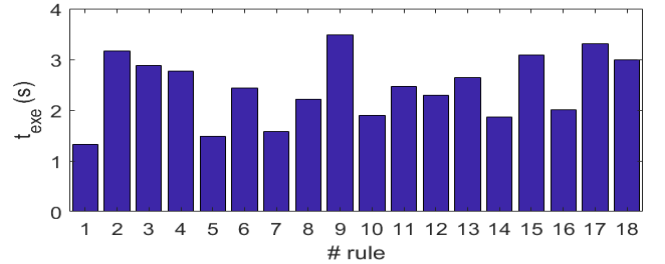


Fig. 4. Time required for training each rule

100 remote sensing images.

After the training process is finished, we test the DRB system on the satellite images from WHU-RS19 image set [20], which is coincident to the core concept of transfer learning. WHU-RS19 dataset is collected from Google Earth imagery, and the images are in the fixed size of 600×600 pixels with various resolutions up to half a meter. There are 19 scene categories within this dataset; however, due to the wide variety of the semantic contents within the local areas of these satellite images, they are, actually, composed of sub-regions of different land use categories. Therefore, this dataset is very suitable for validating the proposed DRB approach.

As the images from categories “commercial”, “forest”, “industrial”, “park”, “parking”, “port”, “residential”, “river”, and “viaduct” of the WHU-RS19 dataset possess the similar semantic features that also appear in UCMerced dataset as well, we use twelve images from these categories and conducted the sub-region analysis by using the DRB system trained based on the UCMerced dataset. We selected four most interesting images and presented the analysis results in Fig. 5. For the analysis results of other images, one can download them from [21]. In these examples, the grid net of 4×4 is used for partitioning the images. However, alternative segmentation approaches can be considered as well.

As one can see from Fig. 5, the proposed DRB approach can successfully identify the semantic contents of the local regions of the satellite scene images. Moreover, for a sub-region that consists of contents of multiple land use categories, the proposed approach is able to further give a ratio on the importance of different semantic contents playing in this sub-region. The very detailed analysis provided by the proposed approach can be very important and useful for studying the land use through the satellite images.

Meanwhile, one may also notice that during the experiment, the proposed approach occasionally ignored some landmarks or confused some categories with others, i.e. the system has problem in distinguishing between “parking lots” and “harbor” in Fig. 5(b), and some part of the ocean is miscategorized as “meadow” as well; some part in Fig. 5(c) is miscategorized as “medium residential”, where there are no obvious building there. We believe that these omissions and miscategorizations of the proposed approach are due to the fact that we did not feed the DRB approach with sufficient training images and the pre-trained DCNN is also not sensitive enough to the difference between the semantic features of different land use classes.

Therefore, by replacing the pre-trained DCNN with a DCNN specifically trained on remote sensing scene images, and involving more training samples to train the DRB system, the ability of the proposed approach on distinguishing the semantic contents on the scene images can be further strengthen, which leads to better performance.

V. CONCLUSION AND FUTURE WORK

In this paper, a novel deep rule-based approach is introduced for recognizing the semantic contents of the sub-regions of satellite scene images. Based on the recently introduced DRB system and further involving the pre-trained

DCNN for feature extraction only, the proposed approach is able to perform high quality detailed analysis on the images after a very short, transparent, highly parallelizable training process. Illustrative examples conducted on benchmark dataset demonstrate the validity and effectiveness of the proposed approach. This proposed approach can be a very useful tool for assisting geoscientists in analyzing the satellite scene images.

However, at the same time, we have to admit that this paper does not have any quantitative analysis due to the lack of benchmark. There is no convenient way to get the ground truth of the land use labels of the subareas of the dataset. On the other hand, the work presented in this paper is illuminating; it offers an effective approach for assisting geoscientists to analyze large-scale satellite images. As future work, we will conduct more systematic experiments on the satellite scene image analysis using the proposed DRB approach. The semi-supervised learning and active learning mechanisms will also be involved to further enhance the learning ability of the proposed approach.

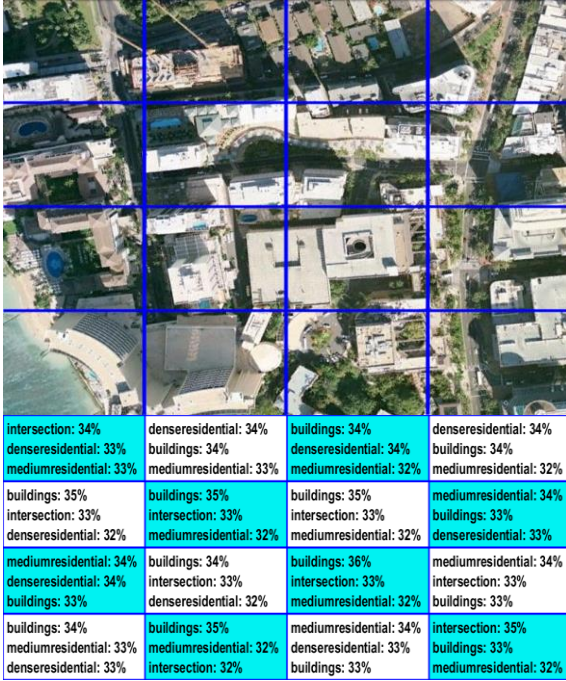
REFERENCES

- [1] G. S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, and L. Zhang, “AID: a benchmark dataset for performance evaluation of aerial scene classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, 2017.
- [2] Q. Weng, Z. Mao, J. Lin, and W. Guo, “Land-use classification via extreme learning classifier based on deep convolutional features,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 704–708, 2017.
- [3] F. Zhang, B. Du, and L. Zhang, “Scene classification via a gradient boosting random convolutional network framework,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1793–1802, 2016.
- [4] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nat. Methods*, vol. 13, no. 1, pp. 35–35, 2015.
- [5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Advances In Neural Information Processing Systems*, 2012, pp. 1097–1105.
- [6] G. J. Scott, M. R. England, W. A. Starns, R. A. Marcum, and C. H. Davis, “Training deep convolutional neural networks for land-cover classification of high-resolution imagery,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 4, pp. 549–553, 2017.
- [7] L. Zhang, L. Zhang, and V. Kumar, “Deep learning for remote sensing data,” *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, 2016.
- [8] A. Oliva and A. Torralba, “Modeling the shape of the scene: A holistic representation of the spatial envelope,” *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.
- [9] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, pp. 886–893.
- [10] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [11] X. Gu, P. Angelov, C. Zhang, and P. Atkinson, “A massively parallel deep rule-based ensemble classifier for remote sensing scenes,” *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 345–349, 2018.
- [12] P. P. Angelov and X. Gu, “Deep rule-based classifier with human-level performance and characteristics,” *Inf. Sci. (Nij.)*, DOI: 10.1016/j.ins.2018.06.048, 2018.
- [13] P. Angelov and R. Yager, “A new type of simplified fuzzy rule-based system,” *Int. J. Gen. Syst.*, vol. 41, no. 2, pp. 163–185, 2011.
- [14] P. Angelov and R. Yager, “Simplified fuzzy rule-based systems using non-parametric antecedents and relative data density,” in *IEEE Symposium Series on Computational Intelligence, IEEE Workshop on Evolving and Adaptive Intelligent Systems*, 2011, pp. 62–69.

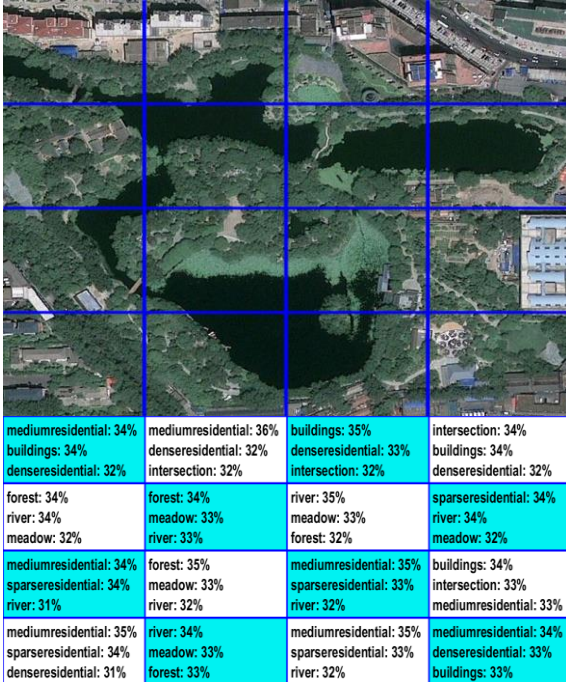
- [15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*, 2015, pp. 1–14.
- [16] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, C. Hill, and A. Arbor, "Going deeper with convolutions," in *IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
- [17] P. Angelov and R. Yager, "Density-based averaging - a new operator for data fusion," *Inf. Sci. (Ny)*, vol. 222, pp. 163–174, 2013.
- [18] P. P. Angelov, "Anomaly detection based on eccentricity analysis," in

2014 *IEEE Symposium Series in Computational Intelligence, IEEE Symposium on Evolving and Autonomous Learning Systems, EALS, SSCI 2014*, 2014, pp. 1–8.

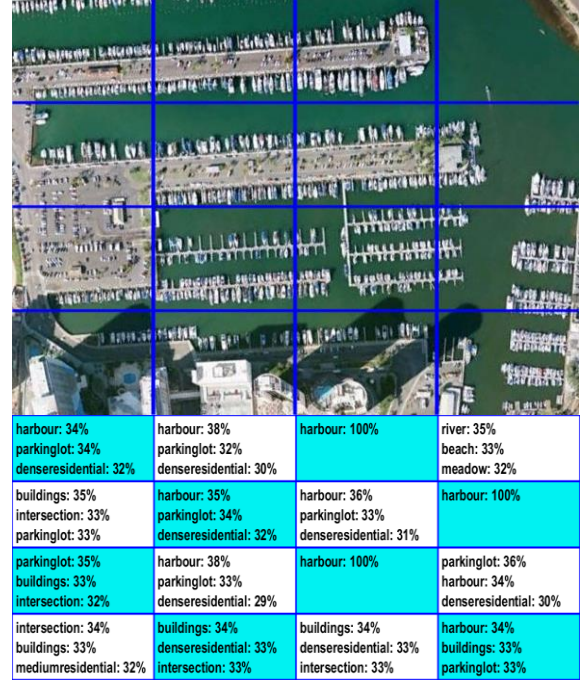
- [19] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *International Conference on Advances in Geographic Information Systems*, 2010, pp. 270–279.
- [20] G. Sheng, W. Yang, T. Xu, and H. Sun, "High-resolution satellite scene classification using a sparse coding based multiple feature combination," *Int. J. Remote Sens.*, vol. 33, no. 8, pp. 2395–2412, 2012.
- [21] <http://empiricaldataanalytics.org/downloads.html>.



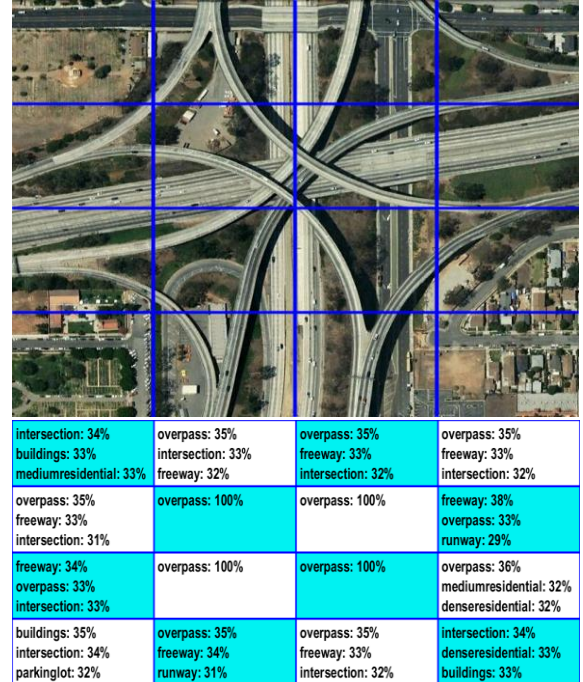
(a) Commercial



(c) Park



(b) Port



(d) Park

Fig.5. Sub-region analysis of the satellite images (part 1)