A Semi-Supervised Deep Rule-Based Approach for Complex Satellite Sensor Image Analysis

Xiaowei Gu, *Member, IEEE*, Plamen P Angelov, *Fellow, IEEE*, Ce Zhang, and Peter M Atkinson

Abstract—Large-scale (large-area), fine spatial resolution satellite sensor images are valuable data sources for Earth observation while not yet fully exploited by research communities for practical applications. Often, such images exhibit highly complex geometrical structures and spatial patterns, and distinctive characteristics of multiple land-use categories may appear at the same region. Autonomous information extraction from these images is essential in the field of pattern recognition within remote sensing, but this task is extremely challenging due to the spectral and spatial complexity captured in satellite sensor imagery. In this research, a semi-supervised deep rule-based approach for satellite sensor image analysis (SeRBIA) is proposed, where large-scale satellite sensor images are analysed autonomously and classified into detailed land-use categories. Using an ensemble feature descriptor derived from pre-trained AlexNet and VGG-VD-16 models, SeRBIA is capable of learning continuously from both labelled and unlabelled images through self-adaptation without human involvement or intervention. Extensive numerical experiments were conducted on both benchmark datasets and real-world satellite sensor images to comprehensively test the validity and effectiveness of the proposed method. The novel information mining technique developed here can be applied to analyse large-scale satellite sensor images with high accuracy and interpretability, across a wide range of real-world applications.

Index Terms—deep rule-based system, deep learning, satellite sensor image analysis, semi-supervised learning.

1 Introduction

Remotely sensed satellite sensor images provide detailed Earth observations, and play an instrumental role in many real-world applications, such as urban planning, precision agriculture, and environmental management [1], [2]. Large-scale (large-area), fine spatial resolution satellite sensor images present a mosaic of geometrical structures and spatial patterns that can be highly complex and heterogeneous [3]. These images are often composed of sub-regions of different land-use categories, and multiple land cover features can be observed within the same sub-region. Analysis of such large-scale satellite sensor images by human experts is labour-intensive and time-consuming due to the large volume, huge complexity and variability [4]. Moreover, characterizing high-level land-use semantics from satellite sensor images is considered as a challenging task for the machine learning communities, thus, requiring the development of novel techniques to classify land-use categories accurately.

To date, many approaches have been developed for remotely sensed aerial scene classification, and they can be categorized broadly into three major classes: 1) methods based on low-level features, which attempt to distinguish aerial scenes based on low-level visual characteristics extracted from the images [5], [6], [7]; 2) methods based on middle-level information, which encode low-level visual features extracted from local regions of images into middle-level representations (e.g., bag of visual words) [8], [9]; 3) methods based on high-level feature representations, which are often learned through deep neural networks, such as deep convolutional neural networks (DCNNs) [10], [11], [12]. Compared with the first two categories, high-level methods are the most suitable for land-use scene classification, with state-of-the-art results achieved in the computer vision and machine intelligence domains [13], [14], including remote sensing. High-level land-use semantics are learned effectively from DCNNs with highly accurate classification results obtained for aerial scenes. Nonetheless, those DCNN-based approaches lack transparency, and the reasoning process of DCNNs is hidden as a black box that is not interpretable for humans. Besides, DCNN methods are computationally expensive, requiring a huge volume of labelled images to train the DCNNs and for parameter tuning.

The common practice of aerial scene classification is to label each image into a specific land-use category [3], [6], [9]. This is not a critical issue because existing research deals mainly with small-size aerial images [8], [11], [15], [16] with simple geometrical structures, and the patterns are easy to interpret. However, small area aerial scene classification is insufficient for exploiting the extensive details covered by large-scale fine spatial resolution satellite sensor images, thus, resulting in a significant loss of valuable information. In contrast, an alternative and potentially preferable solution is to analyse these large-scale images locally and classify each sub-region into different land-use categories [17], [18].

Most existing aerial scene classification approaches are trained using a wealth of labelled images such as to learn the predictive model in a fully supervised manner [1], [10],

- *X. Gu is with the Department of Computer Science, Aberystwyth University, Aberystwyth, SY23 3DB, UK. E-mail: xig4@aber.ac.uk*
- *P. P. Angelov is with the School of Computing and Communications, Lancaster University, Lancaster, LA1 4WA, UK. E-mail: p.angelov@lancaster.ac.uk.*
- *C. Zhang and P. M. Atkinson are with the Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK. C. Zhang is also with the UK Centre for Ecology & Hydrology, Lancaster, LA1 4AP, UK. E-mails: {c.zhang9, pma}@lancaster.ac.uk.*

Manuscript received September 11, 2019; revised Oct 28, 2020.

[11], [12], [19], [20]. However, labelled images are scarce and expensive to capture, whereas unlabelled images are plentiful. Supervised approaches are unable to utilize unlabelled images for training purposes. In contrast, semi-supervised approaches incorporate both labelled and unlabelled images to build stronger classification models by exploiting the rich information from unlabelled images to a greater extent [21], [22], [23], [24]. As semi-supervised approaches overcome the labelling bottleneck and exhibit greater classification performance with less human labour, they have been increasingly explored for aerial scene classification problems [25], [26], [27], [28].

Semi-supervised deep rule-based (SSDRB) system is a recently proposed generic approach for image classification [29], which pioneers the fusion of traditional fuzzy rule-based (FRB) systems to achieve explainable DCNNs [30]. SSDRB is designed to offer a high level of transparency and a human-interpretable decision-making process, typical characteristics of FRB systems, and it also exhibits high classification accuracy benefiting from DCNNs. By exploiting the idea of "pseudo labelling", the SSDRB system can perform semi-supervised learning from unlabelled images and gain new land-use categories autonomously when new data patterns appear in real time.

To mine the valuable information in large-scale fine spatial resolution satellite sensor images that exists abundantly, but remains to be exploited, a **Se**mi-supervised deep **R**ule-**B**ased approach for satellite sensor **I**mage **A**nalysis (SeRBIA) is proposed in this paper. SeRBIA can analyse autonomously the local semantic content of large-scale satellite sensor images and classify local regions of these images into multiple land-use categories that are most relevant. Furthermore, this approach is capable of self-updating and self-improving the knowledge base "on the fly" without human involvement during the analysis process, and it recognizes unseen data patterns autonomously and updates itself to produce refined results throughout the process. This demonstrates the strong ability of SeRBIA to self-develop and learn continuously from satellite sensor images. Moreover, SeRBIA can represent visually its learned knowledge to users in a human-interpretable form through a set of prototype-based IF...THEN rules. This enables clear understanding of how a decision has been made and how to improve if a specific mistake occurs, coinciding with the current move toward the development of explainable artificial intelligence (xAI) systems.

Main contributions of this paper are: 1) a novel chunk-by-chunk semi-supervised learning technique to learn from unlabelled streaming images; 2) a systematic approach that can analyse large-scale fine spatial resolution satellite sensor images autonomously and classify subregions of these images into one or multiple land-use labels based on high-level semantics observed locally; 3) the utility of the human-interpretable IF...THEN rules is tested through the application of large-scale satellite sensor imagery; 4) the capability to learn life-long (continual learning and self-adaptation to non-stationary environments) from satellite sensor images without human expert involvement, and the continuous extension of the knowledge base to incorporate novel data patterns in real time.

The remainder of this paper is organized as follows. Section 2 describes the methodological details of SeRBIA. The main procedure of large-scale satellite sensor image analysis by SeRBIA is described in Section 3. Extensive numerical experiments are presented in Section 4. Section 5 concludes the paper.

2 Proposed SeRBIA

SeRBIA is a new approach designed for large-scale satellite sensor image analysis. The proposed approach first self-organizes its system structure and meta-parameters, and initiates its knowledge base from benchmark aerial image sets through a supervised learning process. It then learns continuously from large-scale satellite sensor images to self-expand its knowledge base in a semi-supervised manner without human expert involvement and performs analysis on the semantic contents of these images locally.

Different from the original SSDRB proposed in [30], SeRBIA performs semi-supervised learning from unlabelled images (or image segments) on a chunk-by-chunk basis [31]. The chunk-by-chunk semi-supervised learning mechanism allows SeRBIA to interpret the data patterns better while aligning closely to the idea of online learning, giving SeRBIA the ability to handle new data patterns effectively and efficiently in image streams. This ability is of paramount importance for real-world applications, particularly, for satellite sensor image analysis.

Technical details of SeRBIA are presented as follows.

2.1 General Architecture

The proposed SeRBIA method is illustrated in Fig. 1, and the zoomed-in structure inside the grey dash-line box is given in supplementary Fig. 1. The algorithmic procedure of SeRBIA is detailed in the next section.

As shown in Fig. 1, SeRBIA is composed of four components:

1) pre-processing module;
2) ensemble feature descriptor;
3) massively parallel IF...THEN rule base, and;
4) semantic content analyser.

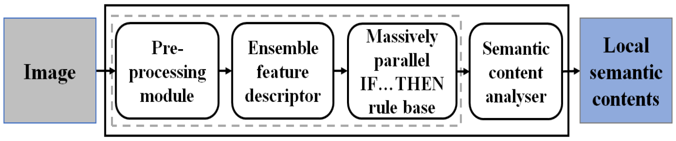


Image
Pre-processing module
Ensemble feature descriptor
Massively parallel IF...THEN rule base
Semantic content analyser
Local semantic contents

Fig. 1: Diagram of SeRBIA.

The pre-processing module consists of the following four sub-layers: *i)* segmentation layer; *ii)* flipping layer; *iii)* mean subtraction layer; and, *iv)* scaling layer. During the supervised priming process with benchmark aerial image sets, the segmentation layer crops each training image into five sub-images/segments, namely, the centre and four corners, to increase generalization ability and reduce over-fitting [10], [32]. During the semi-supervised learning process with large-scale satellite sensor images, the segmentation layer employs a sliding window to partition large-scale images into non-overlapping local regions, which enables SeRBIA to learn the semantic contents of large-scale images locally. The flipping layer, then, flips the obtained segments horizontally to effectively augment the images with increased generalization capability. The mean subtraction layer, which is the third layer of the pre-processing layer, centralizes the R, G, B channels of the image segments around zero mean by subtracting each segment from its mean. This operation accelerates the feature extraction process by the DCNN models since the gradients act uniformly for the three channels. The final layer re-scales the segments of the large-scale image to the sizes required by feature descriptors. In this framework, it rescales the segments to 227×227 pixels size if the AlexNet model [32] is connected or to 224×224 pixels size if the VGG-VD-16 model [33] is concatenated. Details of both DCNN models are given in supplementary section 2.

The second module of SeRBIA is an ensemble of pre-trained DCNN models for feature extraction [34]. In accordance with our previous research [35], both the AlexNet [32] and VGG-VD-16 [33] models are employed to create an ensemble feature descriptor. This is because this particular combination has demonstrated stronger descriptive abilities and results in a higher accuracy in classifying aerial images compared with other benchmark approaches. In this paper, the pre-trained models are used directly without further tuning. The 4096×1 dimensional activations from the first fully-connected layer of the DCNN models are used as the feature vectors. For a particular image segment denoted as $\mathbf{s}$, its feature vectors extracted by the two DCNN models are fused together by addition into a more discriminative representation, $\mathbf{x}$ [36]:

$$\mathbf{x} = \frac{\mathbf{DR}(\mathbf{s})}{||\mathbf{DR}(\mathbf{s})||};\ \ \mathbf{DR}(\mathbf{s}) = \frac{\mathbf{AN}(\mathbf{s})}{||\mathbf{AN}(\mathbf{s})||} + \frac{\mathbf{VN}(\mathbf{s})}{||\mathbf{VN}(\mathbf{s})||}; \qquad (1)$$

where $\mathbf{AN}(\mathbf{s})$ and $\mathbf{VN}(\mathbf{s})$ represent the 4096×1 dimensional feature vectors extracted by the AlexNet and VGG-VD-16 models, respectively; $||\mathbf{x}||$ denotes the L_2-norm of $\mathbf{x}$. However, it should be highlighted that, one can use any types of feature descriptor to create an ensemble, which can be low-level, for example, scale-invariant feature transform (SIFT) [38] and histograms of oriented gradients (HOG) [37], or high-level, for example, ResNet [39] and VGG-VD-19 [33]. Alternatively, one may also train a DCNN from scratch for feature extraction to boost the classification accuracy.

The third and most important module of SeRBIA is the massively parallel IF...THEN rule base [29], [30], which is composed of a set of prototype-based zero-order IF...THEN rules. Each IF...THEN rule consists of a (possibly large) number of human-understandable prototypes identified directly from segments of images of the corresponding land-use category through a "one pass", nonparametric, non-iterative learning process. These prototypes are the most representative samples and they represent local peaks of the multimodal distribution of the data. As a consequence, they represent intuitively the knowledge learned from data that are always meaningful.

Assuming there are N different aerial scene land-use categories, SeRBIA will identify N prototype-based massively parallel IF...THEN rules (one rule per category) from training images in the following form ($n = 1, 2, \ldots, N$):

$$\begin{aligned}\mathbf{R}_n:\ & IF\ (\mathbf{s} \sim \mathbf{P}_{n,1})\ OR\ (\mathbf{s} \sim \mathbf{P}_{n,2})\ OR\ ...\ OR\ (\mathbf{s} \sim \mathbf{P}_{n,L_n}) \\ & THEN\ (Category_n)\end{aligned} \qquad (2)$$

where "$\sim$" represents similarity, which can be interpreted as a fuzzy degree of membership; $\mathbf{s}$ denotes a particular image segment with $\mathbf{x}$ as the discriminative representation; $\mathbf{P}_{n,i}$ represents the i^{th} prototype of the n^{th} category with $\mathbf{p}_{n,i}$ as the discriminative representation; L_n is the number of identified prototypes from the observed images of the n^{th} category.

Prototypes of each massively parallel IF...THEN rule (equation (2)) category are connected by a local decision-maker that follows the "nearest prototype" principle. During operation, for each segment, $\mathbf{s}$ of a particular unlabelled image, $\mathbf{I}$, the local decision-maker will produce a confidence score by identifying the most similar prototype, denoted by $\mathbf{P}_{n,j*}$ to $\mathbf{s}$:

$$\lambda_n(\mathbf{s}) = \max_{l=1,2,\ldots,L_n}(e^{-||\mathbf{x}-\mathbf{p}_{n,l}||^2}) = e^{-||\mathbf{x}-\mathbf{p}_{n,j*}||^2}; \qquad (3)$$

where $\mathbf{x}$ is the discriminative representation of $\mathbf{s}$ extracted by equation (1); $\mathbf{p}_{n,j*}$ is the discriminative representation of $\mathbf{P}_{n,j*}$; $n = 1, 2, \ldots, N$.

The final module of SeRBIA analyses the semantic contents of the satellite sensor image locally. This procedure is based on the respective confidence scores that the IF...THEN rules assign to each segment of the image.

In the following two subsections, the supervised learning and semi-supervised learning processes of SeRBIA are described in detail.

2.2 Supervised Learning

SeRBIA self-organizes an IF...THEN rule base from labelled training images for initialization. Each IF...THEN rule is identified separately in a supervised manner and the identification process of the n^{th} ($n = 1, 2, \ldots, N$) rule is described in this subsection as an example. The supervised learning is performed based on the labelled training images (or, image segments). The same principles apply to all other rules in the rule base [30]. For better illustration, a flow diagram of the main procedure is depicted in supplementary Fig. 2.

Supervised Learning Algorithm

For each labelled training image (assuming the k^{th} one) of the n^{th} category denoted by $\mathbf{s}_{n,k}$, the algorithm firstly extracts its semantic representation by equation (1) as $\mathbf{x}_{n,k}$.

If $\mathbf{s}_{n,k}$ is the very first image of this category (namely, $k = 1$), the n^{th} IF...THEN rule, $\mathbf{R}_n$ is initialized with its global meta-parameters set by equation (4):

$$L_n \leftarrow 1;\ \ \boldsymbol{\mu}_n \leftarrow \mathbf{x}_{n,k}; \qquad (4)$$

where L_n is the current number of available training images that belong to the n^{th} category; $\boldsymbol{\mu}_n$ is the global mean of the corresponding semantic representations. Meta-parameters of the first cluster, $\mathbf{C}_{n,L_n}$ are initialized by $\mathbf{s}_{n,k}$ as follows.

$$\begin{aligned}\mathbf{C}_{n,L_n} \leftarrow \{\mathbf{s}_{n,k}\};\ \ \mathbf{P}_{n,L_n} \leftarrow \{\mathbf{s}_{n,k}\};\ \ \mathbf{p}_{n,L_n} \leftarrow \mathbf{x}_{n,k}; \\ S_{n,L_n} \leftarrow 1;\ \ r_{n,L_n} \leftarrow r_o;\end{aligned} \qquad (5)$$

where $\mathbf{P}_{n,L_n}$ is the visual prototype of $\mathbf{C}_{n,L_n}$; $\mathbf{p}_{n,L_n}$ is the corresponding semantic prototype; S_{n,L_n} is the cardinality of $\mathbf{C}_{n,L_n}$; r_{n,L_n} is radius of the area of influence of $\mathbf{p}_{n,L_n}$; r_0 is a constant for stabilizing new-born clusters, and, in this paper, $r_0 = \sqrt{2(1 - cos(30^o))}$ is used, as in [30]. After this, $\mathbf{R}_n$ is initialized in the form of equation (6).

$$\mathbf{R}_n:\ IF\ (\mathbf{s} \sim \mathbf{P}_{n,L_n})\ THEN\ (Category_n). \qquad (6)$$

Otherwise (namely, $k > 1$), the algorithm firstly calculates data density values of $\mathbf{s}_{n,k}$ and the previously identified prototypes, $\mathbf{P}_{n,i}$ $(i = 1, 2, \ldots, L_n)$ using equation (7):

$$D(\mathbf{Z}) = \frac{1}{1 + \frac{||\mathbf{z}-\boldsymbol{\mu}_n||^2}{1-||\boldsymbol{\mu}_n||^2}}; \tag{7}$$

where $\mathbf{Z} = \mathbf{s}_{n,k}, \mathbf{P}_{n,1}, \mathbf{P}_{n,2}, ..., \mathbf{P}_{n,L_n}$ and $\mathbf{z} = \mathbf{x}_{n,k}, \mathbf{p}_{n,1}, \mathbf{p}_{n,2}, ..., \mathbf{p}_{n,L_n}$.

The prototype, which is the nearest to $\mathbf{s}_{n,k}$, denoted as $\mathbf{P}_{n,j*}$, is then identified based on the similarity between their corresponding semantic representations as:

$$j^* = \underset{i=1,2,\ldots,L_n}{\arg\min} \; (||\mathbf{p}_{n,i} - \mathbf{x}_{n,k}||). \tag{8}$$

Then, **Condition 1** is checked to evaluate the potential of $\mathbf{s}_{n,k}$ to become a new prototype:

$$\begin{aligned} &\textbf{Cond. 1: } \mathit{If}\ (D(\mathbf{s}_{n,k}) > \max_{i=1,2,\ldots,L_n}(D(\mathbf{P}_{n,i}))) \\ &\quad \mathit{Or}\ (D(\mathbf{s}_{n,k}) < \min_{i=1,2,\ldots,L_n}(D(\mathbf{P}_{n,i}))) \\ &\quad \mathit{Or}\ (||\mathbf{p}_{n,j*} - \mathbf{x}_{n,k}|| > r_{n,j*}) \\ &\quad \mathit{Then}\ (\mathit{Add\ a\ new\ prototype}) \end{aligned} \tag{9}$$

If **Condition 1** is satisfied, a new cluster is added to the system $(L_n \leftarrow L_n + 1)$ with $\mathbf{s}_{n,k}$ as its visual prototype. Meta-parameters of this new cluster are initialized by equation (5).

If $\mathbf{s}_{n,k}$ fails to satisfy **Condition 1**, meta-parameters of the nearest cluster, $\mathbf{C}_{n,j*}$ are updated with $\mathbf{s}_{n,k}$ as follows:

$$\begin{aligned} &\mathbf{C}_{n,j*} \leftarrow \mathbf{C}_{n,j*} \cup \{\mathbf{s}_{n,k}\}; \;\; \mathbf{p}_{n,j*} \leftarrow \frac{S_{n,j*}\mathbf{p}_{n,j*} + \mathbf{x}_{n,k}}{S_{n,j*} + 1}; \\ &S_{n,j*} \leftarrow S_{n,j*} + 1; \;\; r_{n,j*} \leftarrow \sqrt{\frac{r_{n,j*}^2 + (1 - ||\mathbf{p}_{n,j*}||^2)}{2}}; \end{aligned} \tag{10}$$

Afterwards, $\mathbf{R}^n$ is updated in the same form as equation (2), and the algorithm starts a new learning cycle for the next labelled training image $(k \leftarrow k + 1)$ or is terminated if requested by the user.

2.3 Semi-Supervised Learning

The semi-supervised learning mechanism of SeRBIA is described as follows. In addition, a flow diagram is given in supplementary Fig. 3 to summarize the overall semi-supervised learning process.

Without losing generality, the learning process is performed on the h^{th} $(h = 1, 2, 3, ...)$ chunk of unlabelled training images (or, image segments), denoted as $\{\mathbf{s}\}^h = \{\mathbf{s}_1^h, \mathbf{s}_2^h, ..., \mathbf{s}_W^h\}$; W is the chunk size. There are two user-controlled parameters for the semi-supervised learning, namely, φ and γ. Both parameters carry a clear meaning. φ determines the degree of rigour in pseudo-labelling of SeRBIA. γ controls its sensitivity to unfamiliar data patterns in unlabelled images.

Semi-Supervised Learning Algorithm

Given a new image chunk, $\{\mathbf{s}\}^h$, the algorithm firstly extracts the semantic representations, denoted by $\mathbf{x}_k^h$ $(k = 1, 2, ..., W)$ of all unlabelled images within this chunk by equation (1).

Then, confidence scores for each unlabelled image, $\mathbf{s}_k^h$ $(\mathbf{s}_k^h \in \{\mathbf{s}\}^h)$ are produced by the IF…THEN rules within the rule base using equation (3), denoted as: $\boldsymbol{\lambda}(\mathbf{s}_k^h) = [\lambda_1(\mathbf{s}_k^h), \lambda_2(\mathbf{s}_k^h), \ldots, \lambda_N(\mathbf{s}_k^h)]^T$ $(k = 1, 2, ..., W)$, and **Condition 2** is checked to identify the unlabelled images from $\{\mathbf{s}\}^h$ where SeRBIA is confident about the categories they belonging to $(\mathbf{s} \in \{\mathbf{s}\}^h)$:

$$\begin{aligned} &\textbf{Cond. 2: } \mathit{If}\ (\lambda_{1^{st}max}(\mathbf{s}) > \varphi\lambda_{2^{nd}max}(\mathbf{s})) \\ &\quad \mathit{Then}\ (\mathbf{s}\ \mathit{belongs\ to\ Category}_{1^{st}max}) \end{aligned} \tag{11}$$

where $\lambda_{1^{st}max}(\mathbf{s})$ and $\lambda_{2^{nd}max}(\mathbf{s})$ are the highest and second highest confidence scores assigned to $\mathbf{s}$ by the massively parallel IF…THEN rules. All the unlabelled images satisfying **Condition 2** are denoted as $\{\mathbf{s}\}_0^h$ $(\{\mathbf{s}\}_0^h \subseteq \{\mathbf{s}\}^h)$.

Each image within $\{\mathbf{s}\}_0^h$ is then used for updating the corresponding IF...THEN rule that gives the highest confidence score using the ***Supervised Learning Algorithm*** as described in Section 2.2. Then, $\{\mathbf{s}\}_0^h$ is removed from $\{\mathbf{s}\}^h$: $\{\mathbf{s}\}^h \leftarrow \{\mathbf{s}\}^h / \{\mathbf{s}\}_0^h$, and the confidence scores of the remaining images are updated. After this, the algorithm continues to identify more unlabelled images from $\{\mathbf{s}\}^h$ that satisfy **Condition 2** and use them to self-expand the knowledge base. The same process repeats until no unlabelled image within $\{\mathbf{s}\}^h$ that can satisfy **Condition 2** any more, namely, $\{\mathbf{s}\}_0^h = \emptyset$.

For the remaining unlabelled images within $\{\mathbf{s}\}^i$, the algorithm is much less confident about the categories they belong to, some of which may belong to some unknown categories. Here **Condition 3** is used to identify such images. $\mathbf{s}_{j*}$ is firstly identified from $\{\mathbf{s}\}^h$ by equation (12) as the image that the algorithm is the least confident with:

$$\mathbf{s}_{j*} = \underset{\mathbf{s} \in \{\mathbf{s}\}_0^h}{\arg\min}(\lambda_{1^{st}max}(\mathbf{s})). \tag{12}$$

Condition 3 is then examined to see whether $\mathbf{s}_{j*}$ can represent a new category:

$$\begin{aligned} &\textbf{Cond. 3: } \mathit{If}\ (\lambda_{1^{st}max}(\mathbf{s}_{j*}) < \gamma) \\ &\quad \mathit{Then}\ (\mathbf{s}_{j*}\ \mathit{belongs\ to\ a\ new\ category}) \end{aligned} \tag{13}$$

If $\mathbf{s}_{j*}$ satisfies **Condition 3**, it belongs to a new category that has not been identified beforehand. Thus, a new IF…THEN rule corresponding to the new category $\mathbf{R}_N$ $(N \leftarrow N+1)$ is initialized in a similar form as equation (6) by $\mathbf{s}_{j*}$ and added to the rule base. Global meta-parameters of $\mathbf{R}_N$ are initialized by equation (4). Meta-parameters of the first cluster, $\mathbf{C}_{N,L_N}$ of the new category are then initialized by equation (5), $\mathbf{s}_{j*}$ is removed from $\{\mathbf{s}\}^i$.

After $\mathbf{R}_N$ has been initialized by $\mathbf{s}_{j*}$, the algorithm uses **Condition 4** to identify more images from $\{\mathbf{s}\}^h$ that belong to this new category.

$$\begin{aligned} &\textbf{Cond. 4: } \mathit{If}\ (\lambda_N(\mathbf{s}) > \varphi \max_{n=1,2,\ldots,N-1}(\lambda_n(\mathbf{s}))) \\ &\quad \mathit{Then}\ (\mathbf{s}\ \mathit{belongs\ to\ Category}_N) \end{aligned} \tag{14}$$

where $\mathbf{s} \in \{\mathbf{s}\}^h$; $\lambda_N(\mathbf{s})$ is the confidence score assigned by $\mathbf{R}_N$. The collection of images satisfying **Condition 4** is denoted as $\{\mathbf{s}\}_1^h$ $(\{\mathbf{s}\}_1^h \subseteq \{\mathbf{s}\}^h)$. $\{\mathbf{s}\}_1^h$ is used for updating $\mathbf{R}_N$ using the ***Supervised Learning Algorithm***. Subsequently, $\{\mathbf{s}\}_1^h$ is removed from $\{\mathbf{s}\}^h$: $\{\mathbf{s}\}^h \leftarrow \{\mathbf{s}\}^h / \{\mathbf{s}\}_1^h$ and the confidence scores that $\mathbf{R}_N$ give to $\{\mathbf{s}\}^h$ are recalculated. Then, the algorithm repeats the same process by identifying more images from $\{\mathbf{s}\}^h$ to update $\mathbf{R}_N$ until no image can satisfy **Condition 4**.

At the end, the algorithm continues to find the next unlabelled image that may initialize a new IF...THEN rule by equation (12). If **Condition 3** is satisfied by this image, the IF...THEN rule base is further expanded with another new rule, and the same process as for $\mathbf{R}_N$ is performed for updating the meta-parameters of this new rule. However, if $\mathbf{s}_{j*}$ fails to satisfied **Condition 3**, it indicates that no more new rules will be added to the rule base and thus, the current learning cycle enters the final phase.

Finally, after the IF...THEN rule base has been fully expanded with the unlabelled images of the current chunk, the algorithm will compare each new rule that represents a new category, denoted as $\mathbf{R}_j$ with the original rules that represent the known categories using **Condition 5**.

$$\textbf{Cond. 5: } \textit{If } (\Lambda_{k^*}(\mathbf{R}_j) > \varphi \max_{\substack{n=1,2,...,N^*;\\ n\neq k^*}} (\Lambda_n(\mathbf{R}_j)) \qquad (15)$$
$$\textit{Then } (\mathbf{R}_j \textit{ is merged into } \mathbf{R}_{k^*})$$

where $j = N^* + 1, N^* + 2, ..., N$; N^* is the number of categories in the labelled training set; $0 \leq k^* \leq N^*$; and $\Lambda_n(\mathbf{R}_j)$ is the average confidence score that $\mathbf{R}_n$ assigns to the L_j prototypes of $\mathbf{R}_j$:

$$\Lambda_n(\mathbf{R}_j) = \frac{1}{L_j}\sum_{l=1}^{L_j}\lambda_n(\mathbf{P}_{j,l}); \qquad (16)$$

and there is: $\lambda_n(\mathbf{P}_{j,l}) = \max_{t=1,2,...,L_n}(e^{-||\mathbf{p}_{j,l}-\mathbf{p}_{n,t}||^2})$. If **Condition 5** is satisfied, the prototypes of $\mathbf{R}_j$ share very high similarity with the prototypes of $\mathbf{R}_{k^*}$ and, meanwhile, are distinctive from prototypes of other categories. Thus, it is expected that the j^{th} category is the same as the category represented by $\mathbf{R}_{k^*}$. As a result, $\mathbf{R}_j$ is merged into $\mathbf{R}_{k^*}$ by updating $\mathbf{R}_{k^*}$ with $\mathbf{P}_{j,1}, \mathbf{P}_{j,2}, ..., \mathbf{P}_{j,L_j}$ using the ***Supervised Learning Algorithm***.

On the other hand, if **Condition 5** is not met, $\mathbf{R}_j$ is kept in the rule base for the next learning cycle, and the algorithm continues to check the next IF...THEN rule.

After all the IF...THEN rules representing new categories have been examined by **Condition 5**, the current learning cycle is completed, and the algorithm begins a new learning cycle to process the next available image chunk ($h \leftarrow h+1$).

Note that although SeRBIA is able to recognize images of land-use categories with semantic features that are distinctive from the labelled training set, it is not able to assign semantic labels to these new categories. Therefore, SeRBIA will name them automatically as "$New\ Category_1$", "$New\ Category_2$", "$New\ Category_3$", etc. Optionally, human experts can be involved to examine the newly learnt IF. . . THEN rules and assign meaningful labels to them accordingly.

3 SeRBIA for Complex Satellite Sensor Image Analysis

In this section, the operation mechanism of SeRBIA for complex satellite sensor image analysis is presented.

To start with, SeRBIA is primed with a benchmark image set using the ***Supervised Learning Algorithm*** (see Section 2.2). After this, for a given satellite sensor image, the pre-processing module of SeRBIA crops the image into K segments, denoted as $\{\mathbf{s}\}_K = \{\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_K\}$: each segment represents a sub-region of the whole image. Each $\mathbf{s}_k \in \{\mathbf{s}\}_K$ is further flipped horizontally to create a mirror image, $\mathbf{s}'_k$ for augmentation, and the augmented set is re-denoted as $\{\mathbf{s}\}'_K = \{\mathbf{s}_1, \mathbf{s}'_1, \mathbf{s}_2, \mathbf{s}'_2, ..., \mathbf{s}_K, \mathbf{s}'_K\}$. SeRBIA then organizes $\{\mathbf{s}\}'_K$ into chunks and performs the ***Semi-Supervised Learning Algorithm*** (see Section 2.3) on a chunk-by-chunk basis to update its massively parallel IF. . . THEN rule base from these image segments.

After completing the semi-supervised learning process, SeRBIA is able to analyse the sub-regions of the satellite sensor image. Given a particular image segment, $\mathbf{s}_k \in \{\mathbf{s}\}_K$, the IF. . . THEN rule base of SeRBIA produces two sets of confidence scores using (3) on both the segment itself and the corresponding mirror image, namely, $\boldsymbol{\lambda}(\mathbf{s}_k) = [\lambda_1(\mathbf{s}_k), \lambda_2(\mathbf{s}_k), \ldots, \lambda_N(\mathbf{s}_k)]^T$ and $\boldsymbol{\lambda}(\mathbf{s}'_k) = [\lambda_1(\mathbf{s}'_k), \lambda_2(\mathbf{s}'_k), \ldots, \lambda_N(\mathbf{s}'_k)]^T$, respectively.

Both $\boldsymbol{\lambda}(\mathbf{s}_k)$ and $\boldsymbol{\lambda}(\mathbf{s}'_k)$ are, then, passed to the semantic content analyser to generate the overall confidence score:

$$\hat{\boldsymbol{\lambda}}(\mathbf{s}_k) = \boldsymbol{\lambda}(\mathbf{s}_k) + \boldsymbol{\lambda}(\mathbf{s}'_k). \qquad (17)$$

Based on $\hat{\boldsymbol{\lambda}}(\mathbf{s}_k)$, the analyser identifies one or multiple land-use categories sharing distinctive high-level semantic features with $\mathbf{s}_k$ using **Condition 6** ($n = 1, 2, ..., N$):

$$\textbf{Cond. 6: } \textit{If } (\varphi\hat{\lambda}_n(\mathbf{s}_k) \geq \hat{\lambda}_{1^{st}max}(\mathbf{s}_k)) \qquad (18)$$
$$\textit{Then } \left(\begin{matrix}\mathbf{s}_k \ possesses\ distinctive\\ semantic\ features\ of\ Category_n\end{matrix}\right)$$

where φ is the same parameter used in **Condition 2** (see Section 2.3). The rationales of **Condition 6** and **Condition 2** are the same. The land-use category corresponding to $\hat{\lambda}_{1^{st}max}(\mathbf{s}_k)$, denoted as $Category_{1^{st}max}$ is the dominant land-use category that $\mathbf{s}_k$ belongs to. If there are other land-use categories satisfying **Condition 6**, one can expect that $\mathbf{s}_k$ is highly likely to have the most distinctive semantic features of these land-use categories as well.

For M_k land-use categories satisfying **Condition 6**, denoted as $Category^*_l$ ($l = 1, 2, ..., M_k$), the likelihoods (the ratios of the importance of different semantic contents) of the M_k most relevant categories that appear in $\mathbf{s}_k$ are given by:

$$\ell^*_l = \frac{\tilde{\lambda}^*_l(\mathbf{s}_k)}{\sum_{j=1}^{M_k}\tilde{\lambda}^*_j(\mathbf{s}_k)}; \qquad (19)$$

where ℓ^*_l is the likelihood of $Category^*_l$; $\tilde{\lambda}^*_l(\mathbf{s}_k)$ is the corresponding confidence score standardized by the mean υ_k and standard deviation δ_k of $\hat{\boldsymbol{\lambda}}(\mathbf{s}_k)$: $\tilde{\lambda}^*_l(\mathbf{s}_k) = \frac{\hat{\lambda}^*_j(\mathbf{s}_k)-\upsilon_k}{\delta_k}$.

Thereafter, SeRBIA analyses the next segment ($k \leftarrow k+1$) by repeating the same procedure from equations (17) to (19). After all the sub-regions of the image have been analysed, SeRBIA moves on to process the next satellite sensor image.

The main procedure of SeRBIA is summarized by the following six steps and a flow diagram is given as supplementary Fig. 4 for better illustration.

Step 1. Perform the ***Supervised Learning Algorithm*** to prime the massively parallel IF. . . THEN rule base with a benchmark dataset;

Step 2. Crop a large-scale satellite sensor image into segments, $\{\mathbf{s}\}_K$ using the sliding window and rescale them to the required sizes;

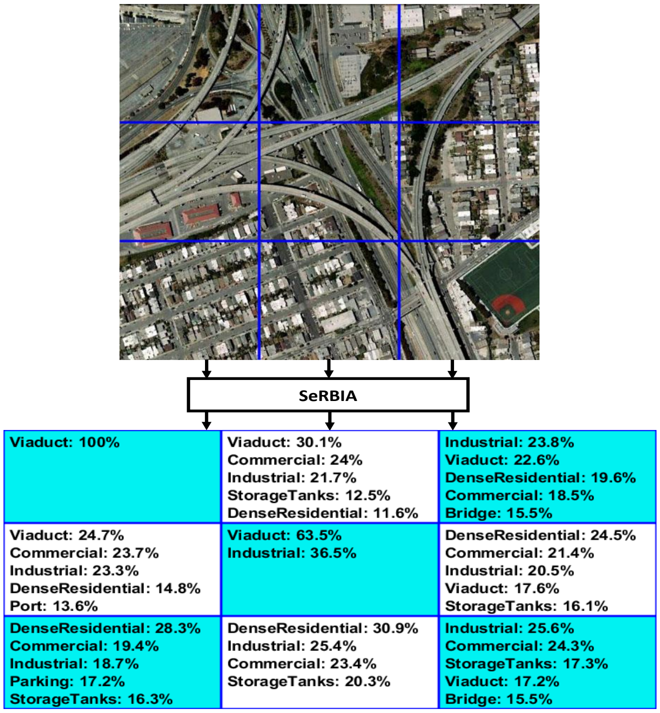
Step 3. Update the system structure and meta-parameters using the ***Semi-Supervised Learning Algorithm*** with $\{\mathbf{s}\}'_K$ chunk-by-chunk;

Step 4. Calculate the overall confidence scores $\hat{\boldsymbol{\lambda}}(\mathbf{s}_k)$ for each segment, $\mathbf{s}_k \in \{\mathbf{s}\}_K$ using equation (17) and identify the M_k most relevant land-use categories to $\mathbf{s}_k$ using equation (18) ;

Step 5. Estimate the likelihoods of the M_k land-use categories associated with each segment, $\mathbf{s}_k$ ($\mathbf{s}_k \in \{\mathbf{s}\}_K$) using equation (19);

Step 6. Go back to **Step 2** and start the analysis process for the next large-scale satellite sensor image.

An illustrative example of the end-product of SeRBIA is given in Fig. 2, where the window size of the sliding window is 200 × 200 pixels and the step size is 200 pixels. In this example, the AID dataset [3] is used to train SeRBIA. An image from WHU-RS dataset [16] is used for validating the proposed approach. Details of the two datasets involved for this example are provided in supplementary section 3. For visual clarity, the maximum value of M_k is set to be 5 throughout this paper.



Viaduct: 100%	Viaduct: 30.1% Commercial: 24% Industrial: 21.7% StorageTanks: 12.5% DenseResidential: 11.6%	Industrial: 23.8% Viaduct: 22.6% DenseResidential: 19.6% Commercial: 18.5% Bridge: 15.5%
Viaduct: 24.7% Commercial: 23.7% Industrial: 23.3% DenseResidential: 14.8% Port: 13.6%	Viaduct: 63.5% Industrial: 36.5%	DenseResidential: 24.5% Commercial: 21.4% Industrial: 20.5% Viaduct: 17.6% StorageTanks: 16.1%
DenseResidential: 28.3% Commercial: 19.4% Industrial: 18.7% Parking: 17.2% StorageTanks: 16.3%	DenseResidential: 30.9% Industrial: 25.4% Commercial: 23.4% StorageTanks: 20.3%	Industrial: 25.6% Commercial: 24.3% StorageTanks: 17.3% Viaduct: 17.2% Bridge: 15.5%

Fig. 2: Illustrative example of SeRBIA.

4 Numerical Examples and Discussion

In this section, numerical examples are presented to demonstrate the ability of SeRBIA for understanding the semantic content of large-scale satellite sensor images. Initially, quantitative analysis is performed on benchmark datasets to justify the validity and effectiveness of the proposed approach for classifying land-use categories from aerial sensor images. Numerical experiments are, then, performed on large-scale satellite sensor images obtained from Google Earth (Google Inc.) to demonstrate the proposed concept and principles.

In this section, the following six popular benchmark datasets for quantitative analysis are used to evaluate the performance of SeRBIA.

1) Singapore dataset [15];
2) UCMerced dataset [8];
3) WHU-RS dataset [16];
4) RSSCN7 dataset [11];
5) AID dataset [3];
6) NWPU45 dataset [40].

Details of the six benchmark datasets are given in supplementary section 3. Note that, NWPU45 is currently the largest benchmark dataset for aerial scene classification. All the reported numerical results were obtained after 15 Monte-Carlo experiments to allow a certain degree of randomness.

4.1 Quantitative Results on Benchmark Datasets

In the numerical experiments presented in this subsection, the segmentation and flipping layers of the pre-processing module create M_0 new segments from each aerial image, $\mathbf{s}_k$ for both training and validation purposes ($M_0 = 10$, namely, five cropped from the original image plus another five created by horizontal flipping). A 4096×1 dimensional representation of the image is obtained as the average of the discriminative representations of the 10 segments [10]:

$$\mathbf{x}_k = \frac{1}{M_0}\sum_{j=1}^{M_0}\mathbf{x}_{k,j} = \frac{1}{M_0}\sum_{j=1}^{M_0}\frac{\mathbf{DR}(\mathbf{s}_{k,j})}{||\mathbf{DR}(\mathbf{s}_{k,j})||}; \qquad (20)$$

where $\mathbf{s}_{k,j}$ is the j^{th} segment of $\mathbf{s}_k$.

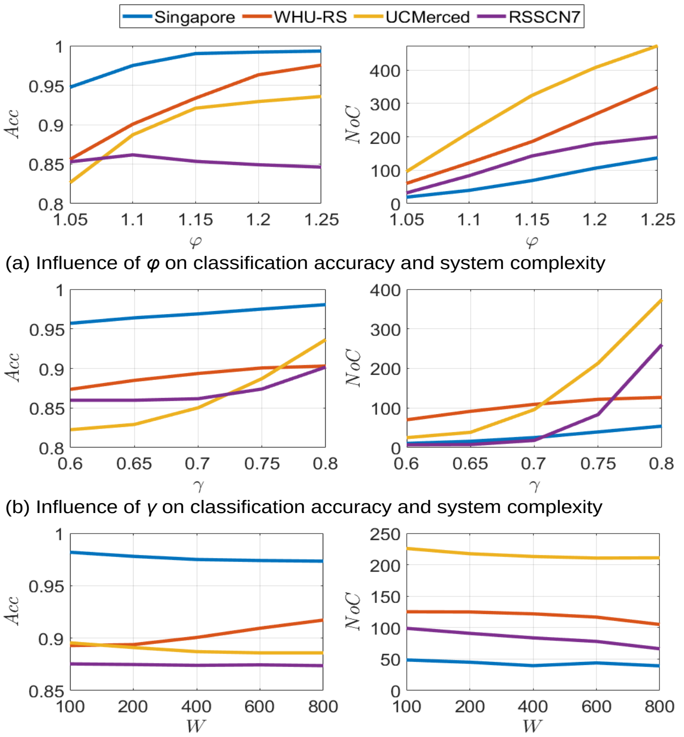
For each validation image $\mathbf{s}_k$, the IF...THEN rule base produces N confidence scores, $\lambda_n(\mathbf{s}_k)$ $(n = 1, 2, ..., N)$. Based on this, the semantic content analyser determines the dominant land-use category as the unique land-use label of s based on the "winner-takes-all" principle:

$$Category(\mathbf{s}_k) \leftarrow Category_{i*};\ \ i^* = \underset{n=1,2,...,N}{\arg\max}\,(\lambda_n(\mathbf{s}_k)). \qquad (21)$$

During the semi-supervised learning process, SeRBIA is able to learn new categories and gain new IF. . . THEN rules. To calculate the accuracies of the classification results obtained by SeRBIA, during the validation process, the dominant land-use category of each image segment associated with the newly gained IF. . . THEN rules are used as the true semantic labels of these new rules.

Firstly, the influence of φ, γ and W on classification accuracy and system complexity of SeRBIA were investigated using the Singapore, UCMerced, WHU-RS and RSSCN7 datasets. For each dataset, 10% of the images per class were randomly selected as the labelled set and the remaining images were used as the unlabelled set. In the first example, the influence of φ on the system performance was investigated, where the value of φ varied from 1.05 to 1.25, and γ and W were set to be 0.75 and 400. The classification accuracy rates (Acc) on the unlabelled set and the average numbers of IF...THEN rules (NoR) after the semi-supervised learning process with different values of φ were depicted in Fig. 3(a). In the second example, the influence of γ on the performance of SeRBIA was investigated, where the value of γ varied from 0.6 to 0.8, φ and W were set to be 1.1 and 400. In the third example, the influence of W was investigated, where its value varied from 100 to 800, and φ and γ were set to be 1.1 and 0.75, respectively. The results of the second and third examples were presented in Fig. 3(b) and (c), respectively. The obtained numerical results were also tabulated in supplementary Tables 2 and 3 for clarity.

It can be observed from Fig. 3, supplementary Tables 2 and 3 that, given a fixed γ, a smaller value of φ allows SeRBIA to identify more prototypes from unlabelled images by **Condition 2** and less new land-use categories, but may introduce more pseudo-labelling errors and decrease the overall classification accuracy. A greater value of φ leads to less prototypes being identified, but may also lower the overall classification accuracy because SeRBIA cannot make full use of the unlabelled images leading to the loss of valuable information. In contrast, a greater value of γ increases the sensitivity of SeRBIA to capture unfamiliar data patterns during the semi-supervised learning process, resulting greater performance and higher system complexity (more new categories being identified by **Condition 3**). Considering the trade-off between the classification accuracy and system complexity, the recommanded value ranges of φ and γ are $[1.1, 1.2]$ and $[0.60, 0.75]$, respectively. Comparing with φ and γ, the value of W has a minor influence on both the classification accuracy and system complexity of SeRBIA. In general, a smaller value of W allows SeRBIA to react quickly to unfamiliar data patterns with the price of higher system complexity. The recommanded value range of W is $[400, 600]$.



(a) Influence of φ on classification accuracy and system complexity

(b) Influence of γ on classification accuracy and system complexity

(c) Influence of W on classification accuracy and system complexity

Fig. 3: Investigation of influence of φ, γ and W on the performance of SeRBIA.

An illustration of the IF…THEN rule base of SeRBIA during one particular experiment ($\varphi = 1.1$, $\gamma = 0.6$ and $W = 400$) was given in supplementary Fig. 6, where one can see that SeRBIA self-organizes nine massively parallel IF…THEN rules from labelled training images during the supervised learning stage. After the semi-supervised learning process, two new IF…THEN rules were identified, and the knowledge base of the previous nine IF…THEN rules were expanded largely by using unlabelled training images.

In the following example, the performance of SeRBIA was compared with the eight benchmark approaches under the same experimental protocol based on the Singapore, UCMerced, WHU-RS, RSSCN7 and AID datasets:

1) Deep rule-based classifier (DRB) [30];
2) k-nearest neighbor classifier (kNN) [41];
3) Support vector machine (SVM) [42];
4) Anchor graph regularization with kernel weights (AnchorK) [24];
5) Anchor graph regularization with local anchor embedding weights (AnchorL) [24];
6) Local and global consistency (LGC) [43];
7) Greedy gradient Max-Cut (GGMC) [22] , and;
8) Laplacian SVM (LapSVM) [23].

DRB was used as the baseline for comparing with SeRBIA. kNN and SVM are both used widely in pre-trained DCNN-based approaches, and they are able to perform state-of-the-art classification results. In this paper, the value of k for kNN was set to be 5. SVM used the linear kernel function. AnchorK, AnchorL, LGC, GGMC and LapSVM are popular semi-supervised classification approaches. In this paper, the user-controlled parameter of AnchorK and AnchorL, s (number of the closest anchors) was set to be $s = 3$, and the iteration number of the local anchor embedding for AnchorL was set to 10 [24]. The user-controlled parameter α of LGC was set to 0.99 as suggested by [43]. The parameter μ of GGMC was set to $\mu = 0.99$ as suggested by [22]. Both LGC and GGMC used the kNN graph with $k = 5$. LapSVM employed the "one versus all" strategy, and it used a radial basis function kernel with $\sigma = 10$. The other two user-controlled parameters μ_I and μ_A were set to 1 and 10^{-6}, respectively; the number of neighbours, k for computing the graph Laplacian was set to 15, as suggested by [23]. Since the performance of LapSVM is highly subject to its parameter settings, the following two alternative experimental settings were considered as well: 1) $\sigma = 10$, $\mu_I = 0.5$, $\mu_A = 10^{-6}$, $k = 15$; and 2) $\sigma = 1$, $\mu_I = 1$, $\mu_A = 10^{-5}$, $k = 10$. Thus, the three LapSVMs with the respective settings were re-denoted as LapSVM$_1$, LapSVM$_2$ and LapSVM$_3$. In this example, SeRBIA used $\varphi = 1.1$, $\gamma = 0.75$ and $W = 400$ for achieving higher classification accuracy. For fair comparison, all the comparative methods used the same 4096×1 dimensional semantic representations extracted by SeRBIA for training and validation.

During the experiments, for each dataset, 10% and 20% of the images per class were randomly selected as the labelled training images and the remaining images were used as the unlabelled training images. Average classification accuracy rates on the unlabelled sets of the five benchmark datasets by the 11 approaches were shown in Table 1. A comparison between the numbers of IF...THEN rules of SeRBIA and DRB was given in supplementary Table 4.

TABLE 1: CLASSIFICATION PERFORMANCE COMPARISON ON BENCHMARK DATASETS

Algorithm	Singapore		WHU-RS		UCMerced		RSSCN7		AID	
	10%	20%	10%	20%	10%	20%	10%	20%	10%	20%
SeRBIA	**0.9750**	**0.9883**	**0.9006**	**0.9563**	**0.8871**	**0.9227**	0.8740	**0.8946**	0.8053	0.8351
DRB	0.9468	0.9673	0.8454	0.8972	0.8025	0.8578	0.8310	0.8683	0.7688	0.8159
SVM	0.9084	0.9582	0.8632	0.9206	0.8211	0.8798	0.8251	0.8563	0.7832	0.8316
kNN	0.9183	0.9550	0.7872	0.8714	0.7740	0.8363	0.8307	0.8689	0.7654	0.8159
AnchorK	0.9589	0.9643	0.8578	0.8956	0.8210	0.8607	0.8302	0.8573	0.7768	0.8169
AnchorL	0.9501	0.9606	0.8600	0.9039	0.8280	0.8690	0.8342	0.8610	0.7804	0.8188
LGC	0.9663	0.9656	0.8931	0.9258	0.8594	0.8844	**0.8848**	0.8925	**0.8290**	**0.8422**
GGMC	0.9669	0.9749	0.8618	0.8904	0.7810	0.8387	0.7538	0.8037	0.7981	0.8203
LapSVM$_1$	0.8322	0.8924	0.8600	0.9284	0.8476	0.8788	0.8351	0.8568	0.6270	0.6731
LapSVM$_2$	0.8802	0.9379	0.8750	0.9303	0.8485	0.8627	0.8429	0.8403	0.6683	0.7027
LapSVM$_3$	0.8633	0.9264	0.8518	0.9225	0.8618	0.8896	0.8480	0.8706	0.6735	0.7395

It is noticeable from Table 1 that there is a significant increase in terms of classification accuracy of SeRBIA by self-learning from unlabelled images. Most importantly, SeRBIA can achieve high classification accuracy (up to 98.83%) surpassing the alternatives in the majority of cases. Nonetheless, it is important to investigate whether the higher performance of SeRBIA over the alternative approaches is of statistical significance. Therefore, statistical pairwise Wilcoxon tests between SeRBIA and the 10 alternative approaches were conducted. The Fisher's method was employed to combine the p-values returned from the hypothesis tests on 15 Monte Carlo experiments:

$$X^2 = -2\sum_{j=1}^{15}\ln(p_j); \tag{22}$$

where p_j is the p-value returned from the j^{th} hypothesis test. The X^2 values returned from the pairwise Wilcoxon tests between SeRBIA and the 10 comparative approaches were given in supplementary Table 5. Note that the value of X^2 is greater when all the p-values are small, suggesting that the null hypotheses are not true for all the tests. If the obtained 15 p-values are all greater than 0.05, X^2 is smaller than $-2 \times 15 \times \ln(0.05) \approx 89.87$. From the X^2 values returned from the 15 statistical tests one can conclude that SeRBIA is significantly more accurate than alternatives across different problems.

In addition, the performance of SeRBIA was further compared with SSDRB, which performs online semi-supervised learning on a sample-by-sample basis, under the same experimental protocol based on the five datasets used by the previous example. Performance comparison in terms of average classification accuracy rates, numbers of IF...THEN rules and overall training time consumptions (in seconds, s) was presented in supplementary Table 6. It can be observed from this table that in general, the chunk-by-chunk semi-supervised learning mechanism allows SeRBIA to achieve higher prediction precision with lower system complexity on complex problems. Most importantly, the time consumptions of SeRBIA is much less than SSDRB, and the difference becomes even larger with the increase of problem size (it only takes 5 milliseconds to process each image in AID). This demonstrates the very strong capability of SeRBIA on handling unlabelled image streams.

4.2 Comparison with the State-of-the-Art Approaches

In this subsection, the performance of SeRBIA is compared with the state-of-the-art methods in the literature on WHU-RS, UCMerced, RSSCN7, AID and NWPU45 datasets under the commonly-used experimental protocols.

In the experiments presented in this subsection, SeRBIA uses the same architecture as Section 4.1. However, to maximize the strength of SeRBIA, the training and validation processes are performed based on segments of the aerial images instead. During the validation process, the semantic content analyser will receive, in total, M_0N confidence scores for each unlabelled image $\mathbf{s}_k$ (N scores per segment). Based on these scores, the dominant land-use category of $\mathbf{s}_k$ is determined by equation (23) as its label:

$$Category(\mathbf{s}_k) \leftarrow Category_{i*};\ i^* = \underset{n=1,2,...,N}{\arg\max}\,(\sum_{j=1}^{M_0}\lambda_n(\mathbf{s}_{k,j})). \tag{23}$$

In this example, the following parameter setting was used $\varphi = 1.1$, $\gamma = 0.7$ and $W = 400$.

Regarding the splitting of labelled and unlabelled training sets, common practice is followed [3], [40]. For WHU-RS, the ratio of labelled training images per category was set to be 40% and 60%, respectively, and the rest were controlled as unlabelled. For UCMerced, the ratios were set to be 50% and 80%. For RSSCN7, the ratios were set to be 20% and 50% per category. For AID, the ratios of labelled training images per category were set to be 20% and 50%, and the ratios were set to be 10% and 20% for NWP45.

Numerical results obtained by the state-of-the-art approaches on the five datasets were listed in Table 2 for benchmark comparison. The results obtained by DRB were also reported in the same table as the baseline. It is clear from this table that SeRBIA is able to perform highly accurate classification (94.29% accuracy rate on AID with 20% labelled images and 87.32% on NWPU45 with only10% labelled images) on unlabelled aerial images surpassing, or on par with, the state-of-the-art approaches.

4.3 Application to Large-Scale Satellite Sensor Images

In this subsection, numerical examples are given to demonstrate the general concept and principles of SeRBIA for application to large-scale satellite sensor images.

10 large-scale satellite sensor images of urban and rural areas of UK were downloaded from Google Earth (Google

TABLE 2: CLASSIFICATION PERFORMANCE COMPARISON WITH THE STATE-OF-THE-ART METHODS ON BENCHMARK DATASETS UNDER COMMONLY USED EXPERIMENTAL PROTOCOLS

Algorithm	WHU-RS		UCMerced		RSSCN7		AID		NWPU45	
	40%	60%	50%	80%	20%	50%	20%	50%	10%	20%
SeRBIA	0.9757 *(0.0060)*	0.9802 *(0.0054)*	0.9636 *(0.0041)*	0.9786 *(0.0099)*	**0.9485** ***(0.0037)***	**0.9619** ***(0.0057)***	**0.9429** ***(0.0020)***	**0.9503** ***(0.0025)***	**0.8732** ***(0.0017)***	0.8811 *(0.0014)*
DRB	0.9442 *(0.0115)*	0.9536 *(0.0087)*	0.9340 *(0.0049)*	0.9684 *(0.0123)*	0.8780 *(0.0055)*	0.9158 *(0.0058)*	0.8358 *(0.0038)*	0.8769 *(0.0044)*	0.7411 *(0.0001)*	0.7812 *(0.0001)*
CaffeNet [3]	0.9511 *(0.0120)*	0.9624 *(0.0056)*	0.9398 *(0.0067)*	0.9502 *(0.0081)*	0.8557 *(0.0095)*	0.8885 *(0.0062)*	0.8686 *(0.0047)*	0.8953 *(0.0031)*	-	-
GoogLeNet [3], [40]	0.9312 *(0.0082)*	0.9471 *(0.0133)*	0.9270 *(0.0060)*	0.9431 *(0.0089)*	0.8255 *(0.0111)*	0.8584 *(0.0092)*	0.8344 *(0.0040)*	0.8639 *(0.0055)*	0.7619 *(0.0038)*	0.7848 *(0.0026)*
VGG-VD-16 [3], [40]	0.9544 *(0.0060)*	0.9605 *(0.0091)*	0.9414 *(0.0069)*	0.9521 *(0.0120)*	0.8398 *(0.0087)*	0.8718 *(0.0094)*	0.8659 *(0.0029)*	0.8964 *(0.0036)*	0.7647 *(0.0018)*	0.7979 *(0.0015)*
BoVW(SIFT) [3], [40]	0.7526 *(0.0139)*	0.8013 *(0.0201)*	0.7190 *(0.0079)*	0.7412 *(0.0330)*	0.7633 *(0.0088)*	0.8134 *(0.0055)*	0.6140 *(0.0041)*	0.6765 *(0.0049)*	0.4172 *(0.0021)*	0.4497 *(0.0028)*
LLC(SIFT) [3], [40]	0.7332 *(0.0213)*	0.7742 *(0.0185)*	0.6941 *(0.0114)*	0.7117 *(0.0209)*	0.7329 *(0.0097)*	0.7657 *(0.0077)*	0.5636 *(0.0068)*	0.5992 *(0.0063)*	0.3881 *(0.0023)*	0.4003 *(0.0034)*
SalM^{3}LBP-CLM [20]	0.9535 *(0.0076)*	0.9638 *(0.0082)*	0.9421 *(0.0075)*	0.9575 *(0.0080)*	-	-	0.8692 *(0.0035)*	0.8976 *(0.0045)*	-	-
salM^{3}LBP [20]	0.8974 *(0.0184)*	0.9258 *(0.0089)*	0.8997 *(0.0085)*	0.9314 *(0.0100)*	-	-	0.8231 *(0.0019)*	0.8759 *(0.0038)*	-	-
salCLM(eSIF) [20]	0.9381 *(0.0091)*	0.9592 *(0.0095)*	0.9293 *(0.0092)*	0.9452 *(0.0079)*	-	-	0.8558 *(0.0083)*	0.8841 *(0.0063)*	-	-
TEX-Net-LF [44]	0.9761 *(0.0036)*	0.9800 *(0.0046)*	0.9589 *(0.0037)*	0.9662 *(0.0049)*	0.8861 *(0.0046)*	0.9125 *(0.0058)*	0.9081 *(0.0011)*	0.9296 *(0.0018)*	-	-
TSDA-ELM [45]	**0.9823** ***(0.0056)***	**0.9892** ***(0.0052)***	**0.9697** ***(0.0075)***	0.9802 *(0.0103)*	-	-	0.9232 *(0.0041)*	0.9458 *(0.0025)*	0.8022 *(0.0022)*	0.8316 *(0.0018)*
VGG-16-CapsNet [46]	-	-	0.9533 *(0.0018)*	**0.9881** ***(0.0022)***	-	-	0.9163 *(0.0019)*	0.9478 *(0.0017)*	0.8508 *(0.0013)*	**0.8918** ***(0.0014)***

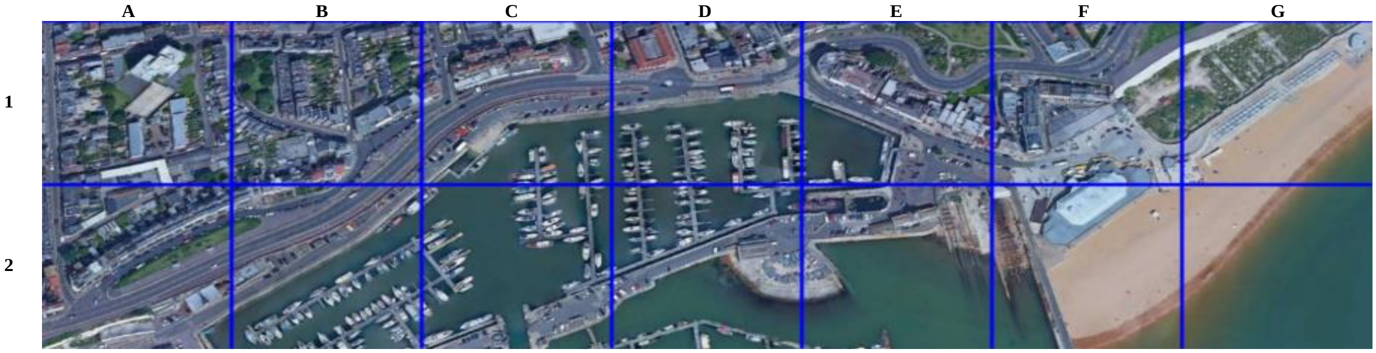
Inc.). Spatial resolutions of these images varies from 30 cm (Worldview-3) to 30 m (Landsat 8). Due to the limited space of this paper, a subregion of satellite sensor image 1 is presented in Fig. 4 as an example. All 10 images are given in supplementary Figs. 7(a)-7(j). These satellite sensor images vary greatly in terms of semantic content; for example, including harbour, parking lots, residential areas, freeways, commercial and forest areas. The local regions of each image also demonstrate strong variability. The size of each satellite sensor image is 800 × 1400 pixels. The 10 satellite sensor images are challenging with high complexity in geometrical structures and spatial patterns, and they are, thus, particularly suitable for testing the performance of SeRBIA. The scale and step size of the sliding window used by SeRBIA was set the same as for the example in Fig. 2, and each satellite sensor image was segmented into 4×7 sub-regions. Note that the segmentation scale used in this paper was determined empirically. Alternative segmentation scales can be considered as well, but one needs to ensure that the scale and spatial resolution of the segments are similar to those of the training images.

AID dataset [3] was used to train SeRBIA in a supervised manner. This dataset was chosen for priming SeRBIA because this dataset contains aerial images from a wide variety of land-use categories, which cover most of the commonly seen categories. The supervised training process followed the same principles as described in Section 3. The size of the images was rescaled to 400×400 pixels. The aerial images of different categories in the AID dataset show a great variety in scale and, the local features of images of some categories show very strong similarities with other categories. Thus, 16 of the 30 land-use categories with similar spatial resolutions and lower inter-class similarity were intentionally selected to train SeRBIA. The chosen land-use categories were listed in supplementary Table 7. After the supervised training process, 16 IF. . . THEN rules were acquired as illustrated in supplementary Fig. 8(a).

After the above, SeRBIA was used to analyse the 10 large-scale satellite sensor images following the same algorithmic procedure presented in section 3. SeRBIA analysed the sub-regions of the satellite sensor images one-by-one, and calculated the likelihoods of the most-likely land-use categories that each sub-region belongs to. The results for the 10 satellite sensor images were given in supplementary Figs. 7(a)-7(j) in the form of a 4 × 7 table containing the background in white and blue colours. The corresponding result of the subregion of satellite sensor image 1 was also given in Fig. 4. During the experiment, SeRBIA used $\varphi = 1.1$ and $\gamma = 0.6$. Each large-scale satellite sensor image was analysed independently, and all the segments from the same image were fed to SeRBIA as a single chunk.

The experimental process was repeated by using the NWPU45 dataset [40] to initialize SeRBIA. Similarly, 18 land-use categories of the original dataset were selected for training (listed in supplementary Table 7) and 18 IF. . . THEN rules of the corresponding classes were obtained after the supervised training process described in Section 2.2. The IF. . . THEN rules were visualized in supplementary Fig. 8(b). The pre-trained SeRBIA was applied to the 10 large-scale satellite sensor images following the same experimental protocol. The results were also shown in supplementary Figs. 7(a)-7(j), which were presented in a similar form but with the background in white and yellow colours.

From Fig. 4 and supplementary Figs. 7(a)-7(j), it is clear that SeRBIA is able to identify multiple most-likely land-use categories of the sub-regions of the large-scale satellite



	A	B	C	D	E	F	G
1	DenseResidential: 31.9% Commercial: 25.3% Industrial: 24.8% MediumResidential: 18.1%	DenseResidential: 27.1% Industrial: 22.3% Commercial: 20.5% Viaduct: 19.5% Port: 10.6%	Viaduct: 24.3% Port: 20.6% Industrial: 19.3% DenseResidential: 18.5% Commercial: 17.4%	Port: 22.1% Industrial: 21.4% StorageTanks: 20.4% Parking: 18.9% DenseResidential: 17.1%	Viaduct: 25.5% Commercial: 23% Industrial: 17.6% StorageTanks: 17% Port: 16.9%	Viaduct: 24.1% Commercial: 23.8% Industrial: 19.6% StorageTanks: 17.6% DenseResidential: 15%	Bridge: 32.1% Viaduct: 25.9% Pond: 15.5% Industrial: 15% MediumResidential: 11.5%
2	Viaduct: 22.3% Commercial: 20.4% Industrial: 20% DenseResidential: 19% StorageTanks: 18.3%	Port: 28.1% DenseResidential: 21.5% Viaduct: 20.6% Parking: 17.3% Industrial: 12.5%	Port: 57.9% Parking: 42.1%	Port: 46.8% Bridge: 28.6% StorageTanks: 24.6%	Port: 33.5% Pond: 24.5% Commercial: 15.3% Industrial: 13.4% Bridge: 13.2%	Port: 24% Viaduct: 19.8% Bridge: 19.6% Pond: 19.2% Industrial: 17.5%	Beach: 100%

Fig. 4: Classification result on a subregion of satellite sensor image 1 using SeRBIA.

sensor images accurately in most cases. For a sub-region, the proposed approach is able to provide a ratio for the importance of different high-level semantic features of the most-likely land-use categories within the sub-region. Furthermore, new land-use categories with the data patterns that did not appear during the supervised training process are identified and the newly gained knowledge is used for analysis.

4.4 Discussion and Future Research

The quantitative analysis in Sections 4.1 and 4.2 undertaken based on the well-known benchmark datasets demonstrates that SeRBIA is able to perform highly accurate classification, surpassing, or on par with, state-of-the-art benchmark methods. Unlike other approaches, SeRBIA not only learns from unlabelled training images to update the system structure and meta-parameters, but also identifies new land-use categories that show less similarity with the land-use categories learned previously from labelled training images. This autonomous semi-supervised learning procedure significantly increases the classification accuracy. Numerical examples on large-scale satellite sensor images presented in Section 4.3 justify the validity and effectiveness of the proposed approach. SeRBIA trained either on the AID or NWPU45 dataset can perform highly accurate analysis on satellite sensor images with highly complex high-level semantic features. Meanwhile, SeRBIA has the capability of continuously self-developing through the analysis process without human intervention, which allows SeRBIA to learn life-long from new images.

However, the results presented in Section 4.3 also show that SeRBIA made some incorrect categorizations occasionally, due to similarity amongst high-level semantic features shared by different land-use categories. For example, SeRBIA sometimes confused land-use categories such as “forest” and “sparse residential”, “port/habor” and “parking/parking lot”, and “pond” and “meadow”. In some cases, SeRBIA may not be able to produce highly accurate estimation on the ratios of importance between multiple land-use categories identified within the same sub-regions due to similar high-level semantic content across the images of different land-use categories in the large-scale AID and NWPU45 remotely sensed scene datasets. For example, similar semantic content can be observed in images of land-use categories “port/habor” and “industrial/industrial area”. Thus, SeRBIA may potentially miss some land-use categories in these particular sub-regions where such high-level semantic content plays a dominant role. This issue could be addressed in the future by making an automatic pre-selection on benchmark datasets and removing the less representative images before training SeRBIA.

One attractive functionality of SeRBIA is the semantic interpretability offered by the massively parallel IF. . . THEN rule base. SeRBIA self-organizes a transparent prototype-based structure by identifying the most representative samples from data, which resembles the learning process of human beings. Although there is no off-the-shelf ground reference to validate the extracted rules and benchmark comparisons, their scientific credibility and interpretability are visually consistent with expert knowledge and human visual interpretation. Through the insights generated by SeRBIA, users can check the learned knowledge by examining the prototypes visually and improve the performance and correctness of the proposed approach by simply adding, deleting and/or merging prototypes, which is much more straightforward than parameter fine-tuning for common “black box”-type DCNNs. This is a very attractive mode of learning in contrast with DCNNs and shines a light on a feasible direction for developing the next generation of explainable artificial intelligence.

This paper, therefore, offers a very promising approach for analysing autonomously large-scale satellite sensor images, providing a useful tool for geospatial data scientists and practitioners. It needs to be stressed that the main purpose of this paper was to deliver the general concept and principles of this new method. Therefore, only standard image pre-processing techniques and pre-trained DCNN models are employed. In fact, SeRBIA is a generic framework, where more advanced techniques can be utilized for pre-processing and feature extraction to further enhance its performance, but this is beyond the scope of this paper.

Future research should further increase the classification accuracy of SeRBIA and its utility for a broad range of real-world applications. To summarize, the future program of research involves:

1) In this paper, DCNN models pre-trained on natural images are used for feature extraction. Although these models have demonstrated great potential in the remote sensing domain, it can be expected that a DCNN trained specifically on aerial images can perform feature extraction better.

2) SeRBIA is pre-trained based on benchmark datasets; however, these datasets might not be the most suitable ones for priming the system. A more suitable training set composed of images with different levels of scale, illumination and resolution needs to be considered.

3) The relationship between the neighbouring segments (in terms of their locations on the original satellite sensor image) is not taken into consideration in this paper. By incorporating this in the decision-making process, the overall accuracy and utility of SeRBIA can be increased further.

4) The quality of analysis of SeRBIA was judged empirically due to the lack of a benchmark. There is no convenient way to capture ground reference land-use labels for real-world satellite sensor images. In the future, experts in the geospatial science domain should be involved to collect high quality ground datasets for rigorous benchmark comparison.

5 Conclusions

In this paper, a semi-supervised deep rule-based approach for satellite sensor image analysis (SeRBIA) was proposed for remotely sensed satellite sensor image analysis. Through the high-level ensemble feature descriptor, SeRBIA was able to perform high-quality analysis on large-scale satellite sensor images and provide a detailed analysis with the most-likely land-use category/categories at each local region of these images. Moreover, SeRBIA was able to perform continuous self-learning without human intervention and is capable of learning life-long. Therefore, SeRBIA represents a promising technique for assisting human experts to analyse large-scale fine spatial resolution satellite sensor images. More widely, this is a generic method that could be applied to single acquisition, fine resolution RGB imagery captured by unmanned aerial vehicles and airplanes. It also has a great potential to be implemented for other real-world applications concerning image stream/video analysis, such as autonomous driving and surveillance camera. Quantitative analysis on benchmark datasets demonstrated that SeRBIA achieved highly accurate results on unlabelled aerial images surpassing, or on par with, state-of-the-art benchmarks. Numerical examples on large-scale satellite sensor images justified the proposed approach with both high classification accuracy and high interpretability by users. The results further demonstrated the utility of the identified IF. . . THEN rules and the robustness of the learned interpretable knowledge. As such, the proposed SeRBIA approach can be applied to a broad range of image classification problems.

References

[1] 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.

[2] C. Zhang et al., "Joint Deep Learning for land cover and land use classification," *Remote Sens. Environ.*, vol. 221, pp. 173–187, 2019.

[3] G. Xia et al., "AID: a benchmark dataset for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, 2017.

[4] C. Zhang et al., "An object-based convolutional neural network (OCNN) for urban land use classification," *Remote Sens. Environ.*, vol. 216, pp. 57–70, 2018.

[5] J. Yin, H. Li, and X. Jia, "Crater detection based on Gist features," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 8, no. 1, pp. 23–29, 2015.

[6] G. Cheng, J. Han, P. Zhou, and L. Guo, "Scalable multi-class geospatial object detection in high-spatial-resolution remote sensing images," in *International Geoscience and Remote Sensing Symposium*, 2014, pp. 2479–2482.

[7] J. dos Santos, O. Penatti, and R. da Silva Torres, "Evaluating the potential of texture and color descriptors for remote sensing image retrieval and classification," in *International Conference on Computer Vision Theory and Applications*, 2010, pp. 203–208.

[8] 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.

[9] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006, pp. 2169–2178.

[10] F. Hu, G. Xia, J. Hu, and L. Zhang, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," *Remote Sens.*, vol. 7, no. 11, pp. 14680–14707, 2015.

[11] Q. Zou, L. Ni, T. Zhang, and Q. Wang, "Deep learning based feature selection for remote sensing scene classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 11, pp. 2321–2325, 2015.

[12] G. J. Scott, M. R. England, W. Starms, R. Marcum, and C. 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.

[13] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 142–58, 2016.

[14] Y. Guo, J. Zhang, J. Cai, B. Jiang, and J. Zheng, "CNN-based real-time dense face reconstruction with inverse-rendered photo-realistic face images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 6, pp. 1294–1307, 2019.

[15] J. Gan, Q. Li, Z. Zhang, and J. Wang, "Two-level feature representation for aerial scene classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 11, pp. 1626–1630, 2016.

[16] G. Xia, W. Yang, J. Delon, and Y. Gousseau, "Structural high-resolution satellite image indexing," in *ISPRS TC VII Symposium - 100 Years ISPRS*, 2010, pp. 298–303.

[17] X. Gu and P. Angelov, "A deep rule-based approach for satellite scene image analysis," in *IEEE International Conference on Systems, Man and Cybernetics*, 2018, pp. 2778–2783.

[18] Y. Wei et al., "HCP: a flexible CNN framework for multi-label image classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 9, pp. 1901–1907, 2016.

[19] A. Cheriyadat, "Unsupervised feature learning for aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 439–451, 2014.

[20] X. Bian, C. Chen, L. Tian, and Q. Du, "Fusing local and global features for high-resolution scene classification," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 10, no. 6, pp. 2889–2901, 2017.

[21] S. Xiang, F. Nie, and C. Zhang, "Semi-supervised classification via local spline regression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 2039–2053, 2010.

[22] J. Wang, T. Jebara, and S. F. Chang, "Semi-supervised learning using greedy Max-Cut," *J. Mach. Learn. Res.*, vol. 14, pp. 771–800, 2013.

[23] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: a geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, pp. 2399–2434, 2006.

[24] W. Liu, J. He, and S. Chang, "Large graph construction for scalable semi-supervised learning," in *International Conference on Machine Learning*, 2010, pp. 679–689.

[25] L. Bruzzone, M. Chi, and M. Marconcini, "A novel transductive SVM for semisupervised classification of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3363–3373, 2006.

[26] I. Dopido, J. Li, P. R. Marpu, A. Plaza, J. M. Bioucas Dias, and J. A. Benediktsson, "Semisupervised self-learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4032–4044, 2013.

[27] Z. Wang, B. Du, L. Zhang, L. Zhang, and X. Jia, "A novel semisupervised active-learning algorithm for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3071–3083, 2017.

[28] N. Kothari, S. Meher, and G. Panda, "Improved spatial information based semisupervised classification of remote sensing images," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 329–340, 2020.

[29] X. Gu and P. Angelov, "Semi-supervised deep rule-based approach for image classification," *Appl. Soft Comput.*, vol. 68, pp. 53–68, 2018.

[30] 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.

[31] M. Pratama, W. Pedrycz, and E. Lughofer, "Evolving ensemble fuzzy classifier," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 5, pp. 2552–2567, 2018.

[32] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Adv. Neural. Inform. Process Syst*, 2012, pp. 1097–1105.

[33] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*, 2015, pp. 1–14.

[34] L. Zheng, Y. Yang, and Q. Tian, "SIFT meets CNN: a decade survey of instance retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 5, pp. 1224–1244, 2018.

[35] X. Gu and P. Angelov, "Deep rule-based aerial scene classifier using high-level ensemble feature descriptor," in *International Joint Conference on Neural Networks*, 2019, pp. 1-7.

[36] S. Chaib, H. Liu, Y. Gu, and H. Yao, "Deep feature fusion for VHR remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4775–4784, 2017.

[37] 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.

[38] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[40] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: benchmark and state of the art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, 2017.

[41] P. Cunningham and S. J. Delany, "K-nearest neighbour classifiers," *Mult. Classif. Syst.*, vol. 34, pp. 1–17, 2007.

[42] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge: Cambridge University Press, 2000.

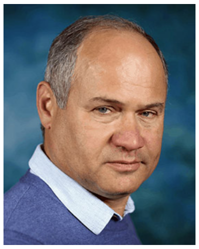
[43] D. Zhou et al., "Learning with local and global consistency," in *Adv. Neural. Inform. Process Syst*, 2004, pp. 321–328.

[44] R. M. Anwer et al., "Binary patterns encoded convolutional neural networks for texture recognition and remote sensing scene classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 138, pp. 74–85, 2018.

[45] Y. Yu and F. Liu, "A two-stream deep fusion framework for high-resolution aerial scene classification," *Comput. Intell. Neurosci.*, vol. 2018, no. 8639367, pp. 1–13, 2018.

[46] W. Zhang, P. Tang, and L. Zhao, "Remote sensing image scene classification using CNN-CapsNet," *Remote Sens.*, vol. 11, no. 5, pp. 494, 2019.

Xiaowei Gu (S16; M19) received the PhD degree in computer science from Lancaster University, U.K., and the MEng degree in communication and information systems and the BEng degree in communication engineering from Hangzhou Dianzi University, China. Dr. Gu is currently a Lecturer in Computer Science at the Department of Computer Science, Aberystwyth University, U.K. His major research interests include machine learning, data analytics and signal processing.



Plamen P Angelov (M99; SM04; F16; MEng 89; PhD 93; DSc 15) holds a Personal Chair (full Professorship) in Intelligent Systems with the School of Computing and Communications, Lancaster University, U.K. where he is also the Director of LIRA (Lancaster Intelligent, Robotic and Autonomous systems) Research Centre (www.lancs.ac.uk/lira). Dr. Angelov is the Vice President of the International Neural Networks Society and a Distinguished Lecturer of IEEE. He was the recipient various awards and is internationally recognized pioneering results into evolving intelligent systems; autonomous, empirical and anthropomorphic machine learning which is focused on human-intelligible results and self-learning.

Ce Zhang received PhD Degree in Geography from Lancaster Environment Centre, Lancaster University, U.K. in 2018. He was the recipient of a prestigious European Union (EU) Erasmus Mundus Scholarship for a European Joint MSc programme between the University of Twente (The Netherlands) and the University of Southampton (U.K.). Dr. Zhang is currently a Lecturer in Geospatial Data Science at Centre of Excellence in Environmental Data Science (CEEDS) joint venture between Lancaster University and UK Centre for Ecology & Hydrology (UKCEH). His major research interests include geospatial artificial intelligence, machine learning, deep learning and remotely sensed image analysis.

Peter M Atkinson received the B.Sc. degree in geography from the University of Nottingham, Nottingham, U.K. in 1986, the Ph.D. degree from The University of Sheffield (NERC CASE award with Rothamsted Experimental Station), Sheffield, U.K. in 1990, and the MBA degree from the University of Southampton, Southampton, U.K. in 2012. Professor Atkinson is currently Distinguished Professor of Spatial Data Science at Lancaster University, Lancaster, U.K. and currently Dean of the Faculty of Science and Technology there. He is a Fellow of the Learned Society of Wales and also a Visiting Professor with the University of Southampton, U.K. and the Chinese Academy of Sciences, Beijing, China. He was previously a Professor of Geography at the University of Southampton. His research interests include remote sensing, geographical information science, and spatial (and space-time) statistics applied to a range of environmental science and socio-economic problems.