VPRS-based regional decision fusion of CNN and MRF classifications for very fine resolution remotely sensed images

Ce Zhang, Isabel Sargent, Xin Pan, Andy Gardiner, Jonathon Hare, and Peter M. Atkinson

***Abstract*—Recent advances in computer vision and pattern recognition have demonstrated the superiority of deep neural networks using spatial feature representation, such as convolutional neural networks (CNN), for image classification. However, any classifier, regardless of its model structure (deep or shallow), involves prediction uncertainty when classifying spatially and spectrally complicated very fine spatial resolution (VFSR) imagery. We propose here to characterise the uncertainty distribution of CNN classification and integrate it into a regional decision fusion to increase classification accuracy. Specifically, a variable precision rough set (VPRS) model is proposed to quantify the uncertainty within CNN classifications of VFSR imagery, and partition this uncertainty into positive regions (correct classifications) and non-positive regions (uncertain or incorrect classifications). Those "more correct" areas were trusted by the CNN, whereas the uncertain areas were rectified by a Multi-Layer Perceptron (MLP)-based Markov random field (MLP-MRF) classifier to provide crisp and accurate boundary delineation. The proposed MRF-CNN fusion decision strategy exploited the complementary characteristics of the two classifiers based on VPRS uncertainty description and classification integration. The effectiveness of the MRF-CNN method was tested in both urban and rural areas of southern England as well as Semantic Labelling datasets. The MRF-CNN consistently outperformed the benchmark MLP, SVM, MLP-MRF and CNN and the baseline methods. This research provides a regional decision fusion framework within which to gain the advantages of model-based CNN, while overcoming the problem of losing effective resolution and uncertain prediction at object boundaries, which is especially pertinent for complex VFSR image classification.**

***Index Terms*—rough set, convolutional neural network, Markov random field, uncertainty, regional fusion decision.**

Manuscript received September 27, 2017; revised November 17, 2017, January 12, 2018 and March 27, 2018; accepted March 31, 2018. This work was supported by National Key Research and Development Program of China (No. 2017YFB0503602) and co-funded by Ordnance Survey and Lancaster University. (*Corresponding author: Peter M. Atkinson*)

C. Zhang and P. M. Atkinson are with the Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, U.K. (e-mail: c.zhang9@lancaster.ac.uk and pma@lancaster.ac.uk)

I. Sargent and A. Gardiner are with the Ordnance Survey, Adanac Drive, Southampton SO16 0AS, U.K. (e-mail: Isabel.Sargent@os.uk and Andy.Gardiner@os.uk)

X. Pan is with the School of Computer Technology and Engineering, Changchun Institute of Technology, Changchun 130021, China and the Northeast Institute of Geography and Agroecology, Chinese Academic of Science, Changchun 130102, China. (e-mail: 101103991@qq.com)

J. Hare is with Electronics and Computer Science (ECS), University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: jsh2@ecs.soton.ac.uk)

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

I. Introduction

REMOTE sensing technologies have evolved greatly since the launch of the first satellite sensors, with a significant change being the wide suite of very fine spatial resolution (VFSR) sensors borne by diverse platforms (satellite, manned aircraft or unmanned aerial vehicles UAV) [1]. These technical advances have resulted in immense growth in the available VFSR remotely sensed imagery typically acquired at sub-metre spatial resolution [2], such as QuickBird, GeoEye-1, Pleiades-1, and WorldView-2, 3, and 4. The fine spatial detail presented in VFSR images offer huge opportunities for extracting a higher quality and larger quantity of information, which may underpin a wide array of geospatial applications, including urban land use change monitoring [3], precision agriculture [4], and tree crown delineation [5], to name but a few. One of the bases of these applications is image classification where information embedded at the pixel level is captured, processed and classified into different land cover classes [6]. Image classification applied to VFSR imagery, however, can be a very complicated task due to the large spectral variation that the same land cover class can produce, which increases the difficulty of discriminating complex and ambiguous image features [7]. The increased spatial resolution, often in conjunction with a limited number of wavebands, can lead to reduced spectral separability amongst different classes. As a consequence, it is of prime concern to develop robust and accurate image classification methods to fully exploit and analyse such data effectively and to keep pace with the technological advances in remote sensors.

Over the last few decades, a vast array of computer-based image classification methods have been developed [8], ranging from unsupervised methods such as K-means clustering, supervised statistical approaches such as maximum likelihood classification, and non-parametric machine learning algorithms, such as the multilayer perceptron (MLP), support vector machine (SVM) and random forest (RF), amongst others. Non-parametric machine learning is currently considered as the most promising and evolving approach [9]. The MLP, as a typical non-parametric neural network classifier, is designed to learn the non-linear spectral feature space at the pixel level irrespective of its statistical properties [10]. The MLP has been used widely in remote sensing applications, including VFSR-based land cover classification (e.g. Del Frate *et al.* (2007), Pacifici *et al.* (2009)). However, a pixel-based MLP classifier does not make use of the spatial patterns implicit in images, especially for VFSR imagery with unprecedented spatial detail. Thus, limited classification performance can be obtained by the pixel-based MLP classifier (and related algorithms, e.g. SVM, RF, etc.) that purely relies on spectral differentiation.

To better exploit the potential in VFSR remotely sensed imagery, many researchers proposed to incorporate spatial information to distinguish spatial features through context. These spatial features may be associated with a regular spatial organization specific to particular types of land cover [12]. For example, the juxtaposition of buildings and roads can create a specific spatial pattern. Similarly, the periodic row structure in cereals can be a useful cue in classifying VFSR image data. These spatial patterns can be captured directly through spatial contextual information in the classification process. A typical example of such is the Markov Random field (MRF) [13], that has been used widely in the field of remote sensing. The MRF models the conditional spatial dependencies within a pixel neighbourhood to support prediction for the central pixel, to increase classification accuracy [14]. However, the contextual MRF often uses small neighbourhood windows to achieve the robustness as well as to balance the computational complexity, which might downgrade the performance for the classification of VFSR imagery that requires wider contexts to handle the rich spatial details.

Recent advances in computer vision and machine learning have suggested that spatial feature representation can be learnt hierarchically at multiple levels through deep learning algorithms [15]. These deep learning approaches learn the spatial contexts at higher levels through the models themselves to achieve enhanced generalization capabilities. The convolutional neural network (CNN), as a well-established deep learning method, has produced state-of-the-art results for multiple domains, such as visual recognition [16], image retrieval [17] and scene annotation [18]. CNNs have been introduced and actively investigated in the field of remote sensing over the past few years, focusing primarily on object detection [19] and scene classification [20]. Recent work has demonstrated the feasibility of CNNs for remote sensing image classification, as here. For example, Zhao and Du (2016) used an image pyramid of hyperspectral imagery to learn deep features through the CNN at multiple scales. Chen *et al.* (2016) introduced a 3D CNN to jointly extract spectral–spatial features, thus, making full use of the continuous hyperspectral and spatial spaces. Längkvist *et al.* (2016) used a CNN model with different contextual sizes to classify and segment VFSR satellite images. Volpi and Tuia (2017) used deep CNNs to perform a patch-based semantic labelling of VFSR aerial imagery together with normalized DSMs. All of these works demonstrated the superiority of CNNs by using contextual patches as their inputs and the convolutional operations for spatial feature representation.

The contextual-based CNN classifiers, however, might introduce uncertainties along object boundaries, leading to over-smoothness to some degree [25]. Besides, objects with little spatial information are likely to be misclassified, even for those with distinctive spectral characteristics [25]. In fact, any classifier, regardless of its model structure, predicts with uncertainty when handling spatially and spectrally complex VFSR imagery. A key problem to be addressed is, thus, for a given classification map, which areas are correctly classified and which are not? This information is important for classification map *producers* who need to further increase classification accuracy. Information on uncertainty is also very useful for classification map *users*, because if it is available, at least in some generalised form, users can better target their attention and effort. Currently, classification model uncertainty is assessed mainly using measures such as the difference between the first and second largest class membership value [26], Shannon's entropy [27], α-quadratic entropy [28], and so on, but there is generally a lack of objective and automatic approaches to partition and label the correct and incorrect classification regions.

The real problem with image classification, using a CNN or any other classifier, is, thus, how to reasonably describe and partition the geometric space given the inherent prediction uncertainties in a classification map. We previously proposed to create rules to threshold the classification results and deal with uncertainties through decision fusion [25]. This method, although having potential to achieve desirable classification results, involves a large amount of trial and error and prior knowledge of feature characteristics, thus was hard to be generalized and applied in an automatic fashion. As a well-established mathematical tool, rough set theory is proposed here as a means of providing an uncertainty description with no need for prior knowledge, and this can be applied to model uncertainties of classification results.

Rough set theory, as proposed by Pawlak (1982), is an extension of conventional set theory that describes and models the vagueness and uncertainty in decision making [30]. It has been applied in diverse domains such as pattern recognition [31], machine learning [32], knowledge acquisition [33], and decision support systems [34]. Unlike other approaches that deal with vague concepts such as fuzzy set theory, rough set theory provides an objective form of analysis without any preliminary assumptions on membership association, thus, demonstrating power in information granulation [35] and uncertainty analysis [36]. In the field of remote sensing and GIS, rough set theory has been applied in rule-based feature reduction and knowledge induction [30], [37], land use spatial relationship extraction [38], spatio-temporal outlier detection [39], and land cover classification and knowledge discovery [40]. However, description of the uncertainty in remote sensing image classification results, as identified as a need and proposed here, has not been addressed through rough set theory, except for the pioneering work of Ge *et al.* (2009) on classification accuracy assessment. In fact, as one of the basic theories of granular computation, the predominant role of rough sets is to transform an original target granularity (i.e., continuous and intricate) into a simpler and more easily analysable variable. Thus, by using rough sets, the uncertainty of remote sensing classification can be simplified and the resulting data is more readily used to support decision-making.

In this paper, a variant of rough set theory, variable precision rough set (VPRS) [30], is introduced for the first time to model and quantify the uncertainties in CNN classification of VFSR imagery with a certain level of error tolerance, which is more suitable for the remote sensing domain than standard rough set theory due to its complexity. Through the VPRS theory, these classification uncertainties are partitioned and labelled automatically into positive regions (correct classifications), negative regions (misclassifications) and boundary regions (uncertain areas), respectively. These labelled regions are then used to guide the regional decision fusion for final classification. Specifically, the positive regions are trusted directly by the CNN, whereas the non-positive regions (negative and boundary regions) with high uncertainty (often occurring along object edges) are rectified by the results of an MLP-based MRF (MLP-MRF). Such a region-based fusion decision strategy performs classification integration at the regional level, as distinct from the commonly used pixel-based strategies. The proposed VPRS-based MRF-CNN regional decision fusion aims to capture the mutual complementarity between the CNN in spatial feature representation and the MLP-MRF in spectral differentiation and boundary segmentation.

The key innovations of this research can be summarized as: 1) a novel variable precision rough set model is proposed to quantify the uncertainties in CNN classification of VFSR imagery, and 2) a spatially explicit regional decision fusion strategy is introduced for the first time to improve the classification in uncertain regions using the distribution characteristics of the CNN classification map.

The effectiveness of the proposed method was tested on images of both an urban scene and a rural area as well as semantic labelling datasets. A benchmark comparison was provided by pixel-based MLP and SVM, spectral-contextual based MLP-MRF as well as contextual-based CNN classifiers, together with mainstream baseline methods.

II. METHODS

A novel VPRS-based method for regional decision fusion of CNN and MRF (MRF-CNN) is proposed for the classification of VFSR remotely sensed imagery. The methodology consists of the following steps:

1. perform CNN and MLP classification using a training sample set (*T*1) and validate them using a testing sample set (*T*3),
2. estimate the uncertainty of the CNN classification result to achieve a CNN classification confidence map (CCM), and perform MLP-based MRF (MLP-MRF) classification,
3. construct a VPRS fusion decision model to partition the CCM into positive regions and non-positive (i.e. boundary and negative) regions using a test sample set (denoted as *T*2), and
4. obtain the final classification result by taking the classification results of the CNN for the positive regions and those of MLP-MRF for the non-positive regions.

Principles and major workflows are detailed hereafter.

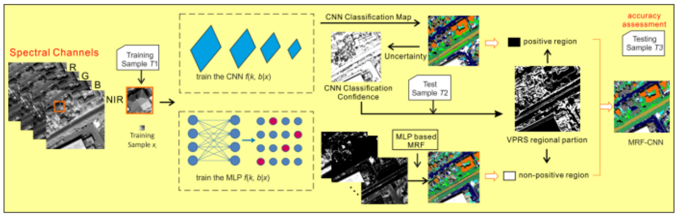


Fig. 1. A workflow illustrating the proposed MRF-CNN methodology.

A. *Convolutional Neural Network (CNN)*

A Convolutional Neural Network (CNN) is a multi-layer feed-forward neural network that is designed specifically to process large scale images or sensory data in the form of multiple arrays by considering local and global stationary properties [42]. The main building block of a CNN is typically composed of multiple layers interconnected to each other through a set of learnable weights and biases [43]. Each of the layers is fed by small patches of the image that scan across the entire image to capture different perspectives of features at local and global scales. Those image patches are generalized through a convolutional layer and a pooling/subsampling layer alternatively within the CNN framework, until the high-level features are obtained on which a fully connected classification is performed [42]. Additionally, several feature maps may exist in each convolutional layer and the weights of the convolutional nodes in the same map are shared. This setting enables the network to learn different features while keeping the number of parameters tractable. Moreover, a nonlinear activation (e.g. sigmoid, hyperbolic tangent, rectified linear units) function is taken outside the convolutional layer to strengthen the non-linearity [44]. Specifically, the major operations performed in the CNN can be summarized as:

$$O^l = pool_p(\sigma(O^{l-1} * W^l + b^l)) \tag{1}$$

where the O^{l-1} denotes the input feature map to the *l*th layer, the W^l and the b^l represent the weights and biases of the layer, respectively, that convolve the input feature map through linear convolution*, and the $\sigma(\cdot)$ indicates the non-linearity function outside the convolutional layer. These are often followed by a max-pooling operation with $p \times p$ window size ($pool_p$) to aggregate the statistics of the features within specific regions, which forms the output feature map O^l at the *l*th layer [43].

B. *Multilayer Perceptron based Markov Random Field (MLP-MRF) Classification*

A multilayer perceptron (MLP) is a classical neural network model that maps sets of input data onto a set of outputs in a feed-forward manner [45]. The typical structure of a MLP is cascaded by interconnected nodes at multiple layers (input, hidden and output layers), with each layer fully connected to the preceding layer as well as the succeeding layer [11]. The outputs of each node are weighted units and biases followed by a non-linear activation function to distinguish the data that are not linearly separable [9]. The weights and biases at each layer are learned by supervised training using a back-propagation algorithm to approximate an unknown input-output relation between the input feature vectors and the desired outputs [11].

The predictive output of the MLP is the membership probability/likelihood to each class at the pixel level, which forms the conditional probability distribution function according to the Bayesian theorem [46]. The objective of Bayesian prediction is to achieve the maximum posterior probability by combining the prior and conditional probability distribution functions, so as to solve the classification problem effectively. The MRF classifier provides a convenient way to model the local properties of an image into positivity, Marknovianity and Homogeneity as its prior probability, together with the learnt likelihood from the MLP, which constitutes the MLP-MRF [47], [48]. Such local neighbourhood information can further be converted into its global equivalence of the Gibbs random field as an energy function based on the Hamersley-Clifford theorem [14]. The MLP-MRF is hence iteratively solved by minimizing the energy function to search for the global minima. See [48] and [49] for more theoretical concepts on MLP-based MRF and its application to image classification.

C. Variable precision rough set based decision fusion between CNN and MRF

1) Introduction to variable precision rough set theory: In rough set theory [29], a dataset is represented as a table, which is called an information system, denoted as $S = (U, A)$, where U is a non-empty finite set of objects known as the universe of discourse, and A is a non-empty finite set of attributes, such that $U \rightarrow Va$ exists for each $a \in A$. The set Va denotes the set of attribute values that a may take. A decision table is an information system in the form of $S = (U, A \cup \{d\})$, where $d \notin A$ is the decision attribute. For any attribute set $P \subseteq A$, there is an indiscernible relation R between two objects x and y:

$$R = \{(x, y) \in U^2 \mid \forall a \in P, a(x) = a(y)\} \tag{2}$$

where R explains that the x and y are indiscernible by the attributes from P (i.e. both x and y share the same attribute values).

The equivalence classes of the indiscernible relation based on R can be defined as:

$$[x]_R = \{y \in U \mid (x, y) \in R\} \tag{3}$$

Given a target set $X \subseteq U$, X can then be approximated by using the equivalence classes of the indiscernible relation R, including a R-lower approximation: $\underline{R}X = \{x \mid [x]_R \subseteq X\}$ and a R-upper approximation: $\overline{R}X = \{x \mid [x]_R \cap X \neq \phi\}$. If $\underline{R}X \neq \overline{R}X$, then the tuple $(\underline{R}X, \overline{R}X)$ forms a rough set. The positive ($POS_R(X)$), negative ($NEG_R(X)$) and boundary ($BND_R(X)$) regions can be defined as:

$$POS_R(X) = \underline{R}X \tag{4}$$

$$NEG_R(X) = U - \overline{R}X \tag{5}$$

$$BND_R(X) = \overline{R}X - POS_R(X) \tag{6}$$

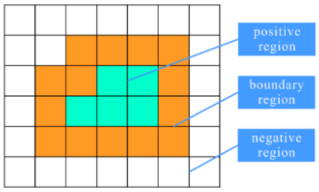


Fig. 2. An illustration of the standard rough set with positive, boundary and negative regions

However, the above standard definition of the set inclusion relation is too rigorous to represent any "almost" complete set inclusion [50] (i.e., equation (4) is difficult to be satisfied strictly). Thus, a variable precision rough set (VPRS) model was proposed to allow a certain number of inclusion errors. Let X and Y be two non-empty subsets of a finite universe U, the degree (or level) of inclusion error of Y within X can be defined as [36]:

$$e(Y, X) = 1 - \frac{Card(Y \cap X)}{Card(Y)}, \ (Y \neq \phi) \tag{7}$$

where the $Card(*)$ denotes the cardinality of a set. The $e(Y, X) = 0$ if and only if $Y \subseteq X$, that is, the case of standard rough set theory (Fig. 2). Suppose $e(Y, X) \neq 0$, then a level of inclusion error β is introduced to tolerate a certain level of inclusion. Given a level of inclusion error β, Y being included by X can be defined as:

$$Y \subseteq_\beta X \quad iff \quad e(Y, X) \leq \beta, \quad 0 \leq \beta \leq 1 \tag{8}$$

Having defined the relative inclusion error β, the β-lower approximation and the β-upper approximation can be characterized as:

$$\underline{R}_\beta X = \{x \in U \mid e([x]_R, X) \leq \beta\} \tag{9}$$

$$\overline{R}_\beta X = \{x \in U \mid e([x]_R, X) \leq 1 - \beta\} \tag{10}$$

Given equations (9) and (10), the positive ($POS_{R,\beta}(X)$), negative ($NEG_{R,\beta}(X)$) and boundary ($BND_{R,\beta}(X)$) regions with a level of inclusion error β can be inferred as:

$$POS_{R,\beta}(X) = \underline{R}_\beta X \tag{11}$$

$$NEG_{R,\beta}(X) = U - \overline{R}_\beta X \tag{12}$$

$$BND_{R,\beta}(X) = \overline{R}_\beta X - POS_{R,\beta}(X) \tag{13}$$

2) VPRS-based MRF-CNN fusion decision: Suppose the membership prediction of the CNN at each pixel is an n-dimensional vector $C = (c_1, c_2, ..., c_n)$, where n represents the number of classes, while each dimension $c_i (i \in [1, n])$ corresponds to the pixel's probability of a specific (i-th) class with certain membership association. Ideally, the probability value of the classification prediction is 1 for the target class but 0 for the other classes, which is usually unobtainable due to the extensive uncertainty in the process of remotely sensed image classification. The probability value C is, therefore, denoted as:

$$f(z) = \{c_z \mid z \in (1, 2, ..., n)\} \ c_z \in [0,1], \sum_1^n c_z = 1 \tag{14}$$

By default, the classification model simply takes the maximum membership association as the predicted output label (denoted as class(C)):

$$class(C) = \arg\max_z(\{f(z) = c_z \mid z \in (1, 2, ..., n)\}) \tag{15}$$

The confidence of being determined as *class*(C) is derived from one minus the normalized Shannon Entropy [41]:

$$conf = 1 - \frac{E_i}{E_{max} \text{-} E_{min}} = 1 - \frac{-\sum_{z=1}^{n} f_i(z) \log_2(f_i(z))}{E_{max} \text{-} E_{min}} \tag{16}$$

where, $E_i = -\sum_{z=1}^{n} f_i(z) \log_2(f_i(z))$ denotes the entropy value of the ith pixel, whereas the E_{max} and the E_{min} refer to the maximum and minimum entropy values, respectively, of the entire classification map. When the entropy of a pixel is maximized (i.e., E_{max} in (16)), $f(z)$ approximates a uniform probability distribution, representing that there is a strong possibility that the pixel is wrongly classified, and therefore the confidence value *conf* tends to be small (i.e., the level of the corresponding uncertainty tends to be higher) and *vice versa*. Therefore, the *conf* ($\in[0,1]$) is inversely correlated with the normalized entropy.

Given a CNN classification map, the confidence value of an object is spatially heterogeneous: the central region is often accurately classified, but the boundary region is likely to be misclassified [25]. The two regions (i.e., patch centre and patch boundary) can then be described theoretically by using rough set theory [30]. That is, the correctness, incorrectness and uncertainty of image classification can be modelled via the positive (4), negative (5) and boundary (6) regions, respectively.

The decision attribute $\{d\}$ of the rough set model, commonly referred to as the attribute for the identification of a specific land cover class, is used here to describe whether a test sample is correctly classified (i.e., a strength and weakness analysis on the classification results of the region corresponding to the sample). The confidence value (*conf*) of any two samples within this region should belong to the same indiscernible relation, of which they should be treated simultaneously. For the confidence map of CNN classification (i.e., the image with a *conf* value at each pixel), it can, therefore, be partitioned into a series of intervals, each of which represents a particular indiscernible relation:

$$[0, step),[step, step\times 2),...,[step\times floor(conf\ /\ step),1] \qquad (17)$$

Where, *step* is the atomic granule representing the least unit of indiscernible relation. Each interval forms an indiscernible region (denoted as IND_{Area}) on the CNN classification map. By checking the consistency of the classification results with respect to the test samples (*T*2), the partitions can then be characterized as: the positive region (the negative region, respectively) where the entirety of *T*2 lying in the region are correctly (incorrectly, respectively) classified, and the boundary region in which the *T*2 are partially correctly classified.

There exists extensive uncertainty and inconsistency in remotely sensed image classification, especially for VFSR imagery. A small amount of error (even with only one misclassified sample) could inevitably turn a positive region into a boundary region. Thus, equation (4) is too restrictive and might not be sufficiently satisfied. Therefore, the introduction of the VPRS model with a relative classification error β is necessary to allow for some degree of misclassification in the largely correct classification. Based on the VPRS model, the CNN classification confidence map can be partitioned into indiscernible regions (i.e. IND_{Area}). The accuracy of each region is evaluated further using the test sample sets (*T*2) to quantify the ratio of the labelled samples that are consistent or inconsistent to the categories of the classification results. Those indiscernible regions that meet the accuracy requirements of (11) are labelled as positive regions, whereas those fitting equations (12) and (13) are characterised as non-positive regions.

As illustrated by equation (7), the real level of inclusion error (denoted as *error*) in a specific IND_{Area} is essentially the classification error of the test sample (*T*2), that is, the ratio between the number of misclassified samples and the total number of the samples within the region. The IND_{Area} can then be identified either as a *positive region* or a *non-positive region* based on the relative inclusion error β:

$$IND_{Area} = \begin{cases} positive\ region & error \le \beta \\ non\text{-}positive\ region & error > \beta \end{cases} \qquad (18)$$

The final classification results of all pixels within the region can then be determined by using either the results ($class_{cnn}$) of CNN (in the case of *positive region*), or the results ($class_{mlp\text{-}mrf}$) of MLP-MRF (in the case of *non-positive region*). The positive region and the non-positive region are, therefore, allocating priority to the CNN and the MLP-MRF accordingly.

Following the strategy mentioned above, the VPRS-based decision fusion algorithm for remotely sensed image classification is illustrated using pseudo-code in Table I:

TABLE I
DETAILED DESCRIPTION OF THE VPRS-BASED REGIONAL DECISION FUSION ALGORITHM FOR REMOTELY SENSED IMAGE CLASSIFICATION

```
VPRS-BASED REGIONAL DECISION FUSION ALGORITHM
Input: remotely sensed (RS) image, level of inclusion error $\beta$, training sample set $T1$, rough set test sample set $T2$, atomic granule step
Output: classification result resultImg
Model$_{cnn}$ = The CNN model trained by sample set $T1$
Model$_{mlp-mrf}$ = The MLP-MRF model trained by sample set $T1$
fuzzyMatrix = The RS image classified by using Model$_{cnn}$ to obtain decision vector
conf = The uncertain level within fuzzyMatrix (1 – Normalized Entropy)
For each region $IND_{Area}$ partitioned from conf using an atomic granule step
    using error (derived from $T2$) and $\beta$ to determine $IND_{Area}$ (18)
    If   $error \le \beta$ then $IND_{Area}$ belongs to positive region
        resultPixels = each pixel within $IND_{Area}$ is classified by CNN ($class_{cnn}$)
    Else $IND_{Area}$ belongs to non-positive region
        resultPixels = each pixel within $IND_{Area}$ is classified by MLP-MRF ($class_{mlp-mrf}$)
    End if
    resultImg = resultImg ∪ resultPixels
End for
Return resultImg
End
```

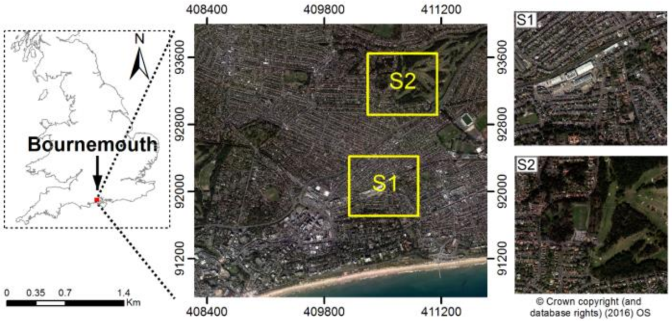
III. EXPERIMENTAL RESULTS AND ANALYSIS

A. Data description and experimental design

Experiment 1: The city of Bournemouth, UK and its surrounding environment, located on the southern coast of England, was selected as a case study area (Fig. 3). The urban area of Bournemouth city is very developed with a high density of anthropogenic structures such as residential houses, commercial buildings, roads and railways. In the contrast, the suburban and rural areas near Bournemouth are less densely populated, predominantly covered by natural and semi-natural environments.

An aerial image was captured on 20 April 2015 using a Vexcel UltraCam Xp digital aerial camera with 25 cm spatial resolution and four multispectral bands (Red, Green, Blue and

Near Infrared), referenced to the British National Grid coordinate system (Fig. 3). Two subsets of the imagery with different environmental settings, including S1 (2772×2515 pixels) within Bournemouth city centre and S2 (2639×2407 pixels) in the rural and suburban area were chosen to test the classification algorithms. S1 consists mainly of nine dominant land cover classes, including Clay roof, Concrete roof, Metal roof, Asphalt, Railway, Grassland, Trees, Bare soil and Shadow, listed in Table II. S2 includes Queen's Park Golf Course and is comprised of large patches of woodland, grassland and bare soil speckled with small buildings and roads. There are seven land cover categories in this study site, namely, Clay roof, Concrete roof, Road-or-track, Grassland, Trees, Bare soil and Shadow (Table II).



© Crown copyright (and database rights) (2016) OS

Fig. 3. Location of study area at Bournemouth within the UK, and aerial imagery showing zooms of the two study sites S1 and S2.

TABLE II

LAND COVER CLASSES AT TWO STUDY SITES WITH TRAINING AND TESTING SAMPLE SIZE PER CLASS. TRAINING SAMPLE *T*1 AND TESTING SAMPLE *T*3 WERE USED FOR MODEL CONSTRUCTION AND ACCURACY VALIDATION, WHILE TEST SAMPLE *T*2 WAS USED FOR BUILDING THE VARIABLE PRECISION ROUGH SET.

Study Sites	Class	Training Sample *T*1	Test Sample *T*2	Testing Sample *T*3
S1	Clay roof	110	156	110
	Concrete roof	107	148	107
	Metal roof	103	139	103
	Asphalt	107	148	107
	Grassland	114	162	114
	Trees	104	141	104
	Bare soil	103	139	103
	Shadow	103	139	103
	Railway	102	137	102
S2	Clay roof	82	104	82
	Concrete roof	90	115	90
	Road-or-track	85	108	85
	Grassland	86	110	86
	Trees	98	124	98
	Bare soil	84	106	84
	Shadow	86	110	86

Sample points were collected using a stratified random scheme from ground data provided by local surveyors in Bournemouth, and split into 50% training samples (Training Sample *T*1 at Table II) and 50% testing samples (Testing Sample *T*3 at Table II) for each class. In addition, a set of test samples (Test Sample *T*2, see Table II) with which to construct the variable precision rough set (VPRS) model were stratified randomly collected throughout the imagery and manually labelled into different land cover classes. The sample labelling was based on expert knowledge and historical references provided by local surveyors and photogrammetrists. Field survey was conducted on April 2015 to further check the validity and precision of the selected samples. Moreover, a highly detailed vector map from the Ordnance Survey, namely the MasterMap Topography Layer [51], was fully consulted and cross-referenced to gain a comprehensive appreciation of the land cover and land use within the study area.

Experiment 2: Two well-known semantic labelling datasets, the Vaihingen dataset and the Potsdam dataset, were used to further evaluate the effectiveness of the proposed method.

The Vaihingen dataset contains 33 true orthophoto tiles with a spatial resolution of 9 cm. For each tile, four channels are provided, namely near-infrared (NIR), red (R) and green (G), together with digital surface models (DSMs). Six semantic categories were manually classified by ISPRS, including impervious surfaces, building, low vegetation, tree, car, and clutter/background. As previously with other authors (e.g. [24], [52]), the clutter/background class (mainly involving water bodies, background and others) was not considered in the experiments since it accounts only for 0.88% of the total number of pixels.

Following the same training and testing procedures set by FCN [52] and SegNet [24], we used the sixteen annotated tiles in our experiments. Eleven tiles (areas: 1, 3, 5, 7, 13, 17, 21, 23, 26, 32, 37) were selected for training, while the other five tiles (areas: 11, 15, 28, 30, 34) were reserved for testing.

The Potsdam 2D segmentation dataset includes 38 tiles of fine spatial resolution remote sensing images. All of them feature a spatial resolution of 5 cm and have a uniform resolution of 6000×6000 pixels. Twenty-four tiles are provided with Ground Reference pixel labels, using the same five classes as in the Vaihingen dataset without the clutter/background class. In the experiments, Following the practice in [52], six tiles (02_12, 03_12, 04_12, 05_12, 06_12, 07_12) were selected as the testing set, while the other eighteen among the annotated tiles were used for training.

Sample points for both datasets were acquired using a stratified random scheme from the Ground Reference with a stride of 300 pixels to ensure the adequacy of GPU memory, and these were partitioned into 30%, 40% and 30% sets for Training Sample *T*1, Test Sample *T*2 and the Testing Sample *T*3. SVM and other mainstream methods, such as FCN [52], SegNet [24] and Deeplab-v2 [53], were applied as benchmarks.

B. Model Architectures and Parameter Settings

Since the MRF used in this research was based on the probabilistic output from a pixel-based MLP, good choices for the model architectures and parameter settings of the MLP and CNN are essential for the proposed MRF-CNN approach. To make a fair comparison, both CNN and MLP models were assigned the same parameters for the learning rate as 0.1, the momentum factor as 0.7, the logistic non-linearity function, and the maximum iteration number of 1000 to allow the networks to fully converge to a stable state through back-propagation. In the MLP, the numbers of nodes and hidden layers were tuned with 1-, 2-, and 3-hidden layers through cross-validation, and the best predicting MLP was found using two hidden layers with 20 nodes in each layer. For the CNN, a range of parameters including the number of hidden layers, the input image patch size, the number and size of convolutional filter, need to be tuned [43]. Following the discussion by Längkvist *et al.* (2016), the input patch size was chosen from {12×12, 14×14, 16×16, 18×18, 20×20, 22×22, 24×24} to evaluate the influence of context area on classification performance. In general, a small-sized contextual area results in overfitting of the model, whereas a large one often leads to under-segmentation. In consideration of the image object size and contextual relationship coupled with a small amount of trial and error, the optimal input image patch size was set to 16×16 in this research. Besides, as discussed by Chen *et al.* (2014) and Längkvist *et al.* (2016), the depth plays a key role in classification accuracy because the quality of learnt feature is highly influenced by the level of abstraction and representation. As suggested by Längkvist et al. (2016), the number of CNN hidden layers was chosen as four to balance the network complexity and robustness. Other parameters were tuned empirically based on cross-validation accuracy, for example, the kernel size of the convolutional filters within the CNN was set as 3×3 and the number of filters was tuned as 24 at each convolutional layer.

The MLP-MRF requires to predefine a fixed size of neighbourhood and a parameter γ that controls the smoothness level. The window size of the neighbourhood in the MLP-MRF model was chosen optimally as 7×7 in consideration of the spatial context and the fidelity maintained in the classification output. Due to the fine spatial detail contained in the VFSR imagery, the parameter γ controlling the level of smoothness was set as 0.7 to achieve an increasing level of smoothness in terms of the MRF. The simulated annealing optimization using a Gibbs sampler [55] was employed in MLP-MRF to maximize the posterior probability through iteration.

An SVM classifier was further used as a benchmark comparator to test the classification performance. The SVM model involves a penalty value C and a kernel width σ that needs to be parameterised. Following the recommendation by Zhang *et al.* (2015), a grid search with 5-fold cross-validation was implemented to exhaustively search within a wide parameter space (C and σ within [2^{-10}, 2^{10}]). Such parameter settings would lead to high validation accuracy using support vectors to formulate an optimal classification hyperplane.

C. Decision Fusion Parameter Setting and Analysis

The decision fusion between the MLP-MRF and the CNN, namely, the MRF-CNN, based on the VPRS model, involves parameters β (the level of inclusion error) and *step* (the atomic granule). The two parameters were optimized through grid search with cross-validation using Training Sample 2 (Listed in Table II). Specifically, β was varied from 0 to 1 with incremental steps of 0.01, while the *step* was tuned between 0 to 0.5 through a small step of 0.025 (i.e. with a wider parameter searching space) to obtain a higher validation accuracy. By doing so, β and *step* were chosen optimally as 0.1 and 0.075, respectively.

Both of the fusion decision parameters (β and *step*) jointly determined the partition of the positive and non-positive regions. As shown in Fig. 4, these parameter settings, reflected by variation between the ratios of VPRS non-positive and positive regions (horizontal axis coordinates ranging from 0 to 1), have an impact on the CNN classification confidence values (blue dots) and the overall accuracies (boxplots). From the figure, it can be seen that along with the increase of the non-positive ratio, the CNN classification confidence decreases constantly, except for the non-positive ratio from 0.3 to 0.55; whereas the overall accuracy initially increases from around 0.86 to around 0.9 and then decreases constantly until around 0.81. Another observation is that the boxplot tends to be wider as the ratio of non-positive to positive region becomes larger, with more credits being given from the CNN to the MLP-MRF. The optimal non-positive ratio (determined by decision fusion parameter setting) was found to be 0.3 (marked by the red dotted line in Fig. 4).

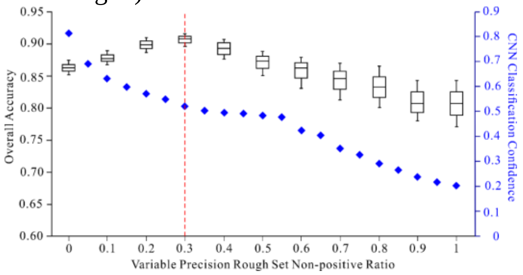


Fig. 4. The CNN classification confidence value and the overall accuracy influenced by the fusion decision parameter setting (in the form of the non-positive to positive ratio)

D. Classification Results and Analysis

Experiment 1: The classification performance of the MRF-CNN and the other benchmark methods, including the MLP, SVM, MLP-MRF and the CNN, were compared using the Testing samples of Bournemouth dataset. Table III lists the detailed accuracy assessment of both S1 for Bournemouth city centre and S2 for the rural and suburban areas with overall accuracy (OA), Kappa coefficient (κ) as well as per-class mapping accuracy. Clearly, the MRF-CNN achieved the best overall accuracy of 90.96% for S1 and 89.76% for S2 with Kappa coefficients of 0.89 and 0.88 respectively, consistently higher than the CNN (85.37% and 86.39% OA with κ of 0.84 and 0.83, respectively), the MLP-MRF (83.76% and 84.52% with corresponding κ of 0.79 and 0.80), the SVM (81.65% and 81.24% with corresponding κ of 0.77 and 0.78), and the MLP (81.52% and 80.32% with the same κ of 0.77) (Table III). In addition, a McNemar z-test that accounts for the pair-wise classification comparison further demonstrates that a statistically significant increase has been achieved by the MRF-CNN over the MLP, SVM, MLP-MRF and the CNN, with z-value = 3.27, 3.02, 2.74 and 2.02 in S1 and z-value =

3.89, 3.51, 3.06 and 2.05 in S2 respectively, greater than 1.96 at 95% confidence level (Table IV). Moreover, the class-wise classification accuracy of MRF-CNN constantly reports the most accurate results highlighted by the bold font in Table III, except for the trees in S2 (89.32%) for which accuracy is slightly lower than for the CNN (90.42%). In Particular, the mapping accuracies of most land covers classified by the MRF-CNN were higher than 90%, with the greatest accuracy achieved in grassland at both study sites S1 and S2, up to 93.57% and 92.94%, respectively.

With respect to the four benchmark classifiers themselves (i.e., MLP, SVM, MLP-MRF and CNN), it can be seen from Table III that their classification accuracies are ordered as: MLP < SVM < MLP-MRF < CNN. For the urban area at S1, the accuracy of the MLP-MRF and the SVM is closer to the MLP (<2%), but with larger difference (>3%) from the CNN. This is further demonstrated by the McNemar *z*-test in Table IV where the CNN is significantly different from the MLP, the SVM and the MLP-MRF (z = 3.12, 2.85 and 2.14, respectively), but the increase of the MLP-MRF is not significant compared with the MLP (z = 1.57) and the SVM (z = 1.68). In the rural area at S2, on the contrary, the accuracy of the MLP-MRF is remarkably higher (>4%) than that of the MLP and SVM with statistical significance (z = 2.12 and 2.04), and only slightly lower than that of the CNN (<2%) without significant difference (z = 1.59).

Figs. 5 and 6 demonstrate visual comparisons of the five classification results using three subset images at each study site (S1 and S2). For the Concrete roof class, from the upper right of Fig. 5(*a*), it is clear that the MLP and SVM classification results maintain the rectangular geometry of the building, but at the same time present very noisy information with salt-and-pepper effects in white throughout the Concrete roof (see the red circles at the figure). Such noise has been largely reduced by the MLP-MRF but still not yet completely eliminated (shown by red circle). The noise has been erased thoroughly by the CNN. However, some serious mistakes have been introduced by misclassifying the asphalt on top of the Concrete roof (highlighted by red circle). Fortunately, the MRF-CNN removed all of the noise while keeping the correctness of the semantic segmentation (yellow circle). A similar pattern was found in the middle of Fig. 5(*b*), where the MLP-MRF is less noisy than the MLP and the SVM (red circles), and the CNN obtains the smoothest classification result, but tends to be under-segmented along the object boundaries (highlighted by red circle). The MRF-CNN, in contrast, keeps the central regions smooth while preserving the precise boundary information (e.g. the rectangularity of the concrete roofs and the shadow next to them; shown in yellow circle). Similar situations are found in the Clay roof, as shown in Fig. 6(*a*) and 6(*c*), where the MLP, SVM and MLP-MRF introduced some noise in the central region, whereas the CNN eradicated them but with obvious geometric distortions. The MRF-CNN, surprisingly, removes all the noise while keeping the crisp boundaries with accuracy. In terms of the railway class illustrated in the middle of Fig. 5(*a*), it was noisily classified by the MLP, the SVM and the MLP-MRF (red circles). This noise was eliminated by the CNN as well as the MRF-CNN (yellow circles). Moreover, some small Road-or-tracks exemplified by Fig. 6(*a*) and 6(*b*) were successfully maintained by the MLP, SVM, MLP-MRF as well as MRF-CNN, yet omitted by CNN due to the convolutional operations.

For the natural land cover classes, the grassland patch shown in Fig. 5(*b*) is shaped approximately square (see the original image in Fig. 5(*b*)). The MLP and SVM produced noisy results confused with the surrounding tree species (shown in red circles). A similar pattern was found in the result of the MLP-MRF but with less noise (marked by red circle). The CNN and the MRF-CNN did not show any noise in the classification map. However, the CNN did not maintain the squared shape of the grassland (shown in red circle), whereas the MRF-CNN successfully kept the geometric fidelity as a square shaped object (highlighted by yellow circle). With regard to the Trees indicated in Fig. 6(*a*) and 6(*b*), the MLP, SVM and MLP-MRF produce different noise: the MLP tends to misclassify the trees as grassland (shown in red circle), whereas the SVM and MLP-MRF sometimes falsely considers the leaf-off trees or the shade of trees as the shaded Clay roof (marked by red circle). All these misclassifications are rectified by the CNN and the MRF-CNN (in yellow circle).

As for the other land cover classes (e.g., bare soil and shadow) the four classification methods do not show significant differences, although some increases in classification accuracy were still obtained by the MRF-CNN. For example, the bare soil shown in Fig. 6(*c*) is highly influenced by the cars and other small objects, which results in over-segmented noise by the MLP and the SVM (shown in red circles) or false identification into Clay roof by the CNN (marked in red circle). The MLP-MRF and the proposed MRF-CNN, fortunately, addressed those challenges with smooth yet semantically accurate geometric results (in yellow circle).

TABLE III

CLASSIFICATION ACCURACY COMPARISON AMONGST MLP, SVM, MLP-MRF, CNN AND THE PROPOSED MRF-CNN APPROACH FOR BOURNEMOUTH CITY CENTRE (S1) AND THE SUBURBAN AREA (S2) USING THE PER-CLASS MAPPING ACCURACY, OVERALL ACCURACY (OA) AND KAPPA COEFFICIENT (κ). THE BOLD FONT HIGHLIGHTS THE GREATEST CLASSIFICATION ACCURACY PER ROW.

Study Site	Class	MLP	SVM	MLP-MRF	CNN	MRF-CNN
S1	Clay roof	91.37%	91.45%	90.58%	88.56%	**92.68%**
	Concrete roof	68.52%	68.74%	72.23%	74.37%	**78.25%**
	Metal roof	89.75%	89.52%	90.12%	91.42%	**92.23%**
	Asphalt	88.59%	88.55%	88.67%	85.98%	**91.26%**
	Grassland	73.51%	74.28%	76.42%	88.63%	**93.57%**
	Trees	65.68%	65.79%	72.28%	82.28%	**88.53%**
	Bare soil	80.46%	80.51%	80.82%	85.23%	**90.24%**
	Shadow	91.56%	91.23%	91.23%	90.14%	**92.16%**
	Railway	82.14%	82.35%	83.57%	90.23%	**91.56%**
	OA	81.52%	81.65%	83.26%	86.37%	**90.96%**
	κ	0.77	0.77	0.79	0.84	**0.89**

S2	Clay roof	88.56%	88.27%	86.75%	82.37%	**90.16%**
	Concrete roof	79.84%	79.62%	81.26%	84.17%	**88.27%**
	Road-or-track	83.02%	83.36%	83.17%	86.54%	**92.38%**
	Grassland	72.11%	73.64%	80.57%	88.58%	**92.94%**
	Trees	79.31%	79.24%	85.26%	**90.42%**	89.32%
	Bare soil	76.18%	76.42%	78.25%	81.36%	**88.75%**
	Shadow	89.42%	89.56%	89.42%	88.25%	**89.58%**
	OA	80.32%	81.24%	84.52%	86.39%	**89.76%**
	κ	0.77	0.78	0.80	0.83	**0.88**

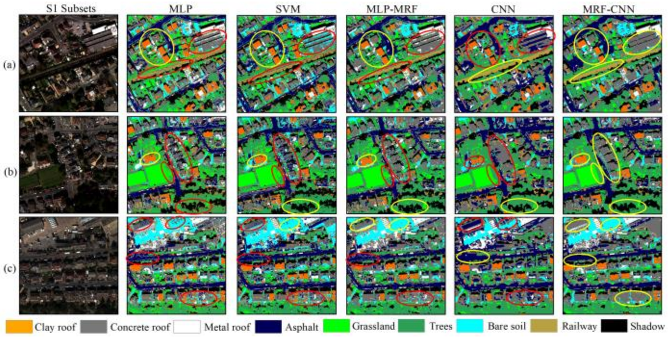


Fig. 5. Three typical image subsets (a, b and c) in study site S1 with their classification results. Columns from left to right represent the original images (R G B bands), the MLP, the SVM, the MLP-MRF, the CNN, and the MRF-CNN classification results. The red and yellow circles denote incorrect and correct classification, respectively.

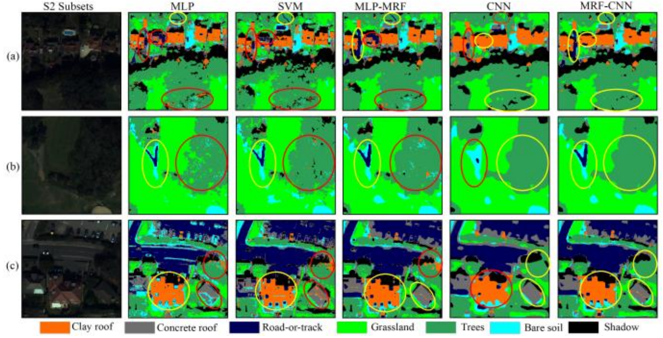


Fig. 6. Three typical image subsets (a, b and c) in study site S2 with their classification results. Columns from left to right represent the original images (R G B bands), the MLP, the SVM, the MLP-MRF, the CNN, and the MRF-CNN classification results. The red and yellow circles denote incorrect and correct classification, respectively.

TABLE IV

MCNEMAR *Z*-TEST COMPARING THE PERFORMANCE OF THE FOUR CLASSIFIERS FOR TWO STUDY SITES S1 AND S2. SIGNIFICANTLY DIFFERENT ACCURACIES WITH CONFIDENCE OF 95% (*Z*-VALUE > 1.96) ARE INDICATED BY *.

Study sites	Classifiers	McNemar Z-test				
		MLP	SVM	MLP-MRF	CNN	MRF-CNN
S1	MLP	—				
	SVM	1.32	—			
	MLP-MRF	1.57	1.68	—		
	CNN	3.12*	2.85*	2.14*	—	
	MRF-CNN	3.27*	3.02*	2.74*	2.02*	—
S2	MLP	—				
	SVM	1.66	—			
	MLP-MRF	2.12*	2.04*	—		
	CNN	2.42*	2.15*	1.59	—	
	MRF-CNN	3.89*	3.51*	3.06*	2.05*	—

Experiment 2: The proposed MRF-CNN and its sub-modules (MLP, MLP-MRF and CNN) as well as other benchmark methods were validated on the Vaihingen and Potsdam semantic segmentation datasets. Table V and VI present the classification accuracies of all four methods together with the four benchmark methods (SVM, FCN, SegNet and Deeplab-v2). The MRF-CNN achieved the largest OA of 88.4% and 89.4% for the two datasets, larger than its sub-modules (86.2% and 86.5%, 82.1% and 83.7%, and 81.4% and 82.1% OA of CNN, MLP-MRF and the MLP, respectively). The MRF-CNN also demonstrates greater accuracy than the benchmarks, including the Deeplab-v2 with an OA of 86.7% and 88.2%, the FCN with an OA of 85.9% and 86.2% [52], the SegNet with an OA of 82.8% and 83.6% [24], and the SVM with an OA of 81.7% and 82.4%.

The per-class mapping accuracy (Table V and VI) shows the effectiveness of the proposed MRF-CNN for the majority of classes. Significant increases in accuracy are realized for the classes of Impervious surfaces, Low vegetation, Building and Car relative to the individual classifier CNN and MLP-MRF, with an average large margin of 3.9%, 4%, 5.55% and 8.75%, respectively. The Tree class accuracy, however, was less significantly increased compared to the CNN, with small margins of 0.8% and 0.6%. In terms of benchmark methods, the MRF-CNN demonstrates higher accuracy for the majority of classes, except for the Car class (79.6% and 80.3%), for which the accuracy is less than for the state-of-the-art Deeplab-v2 (84.7% and 83.9%).

TABLE V

PER-CLASS ACCURACY AND OVERALL ACCURACY (OA) FOR THE MLP, SVM, MLP-MRF, CNN AND THE PROPOSED MRF-CNN APPROACH, AS WELL AS BASELINE METHODS, FOR THE VAIHINGEN DATASET. THE BOLD FONT HIGHLIGHTS THE LARGEST CLASSIFICATION ACCURACY PER ROW.

Method	Imp Surf	Build ing	Low Veg	Tree	Car	OA
MLP	83.5%	82.1%	68.3%	86.1%	64.2%	81.4%
SVM	82.7%	82.4%	69.2%	84.3%	66.5%	81.7%
MLP-MRF	84.3%	83.6%	72.7%	83.9%	71.7%	82.1%
CNN	86.2%	89.2%	76.9%	86.9%	69.7%	86.2%
FCN [52]	87.1%	91.8%	75.2%	86.1%	63.8%	85.9%
SegNet [24]	82.7%	89.1%	66.3%	83.9%	55.7%	82.8%
Deeplab-v2 [53]	88.5%	93.3%	73.9%	86.9%	**84.7%**	86.7%
MRF-CNN	**89.7%**	**93.8%**	**80.1%**	**87.7%**	79.6%	**88.4%**

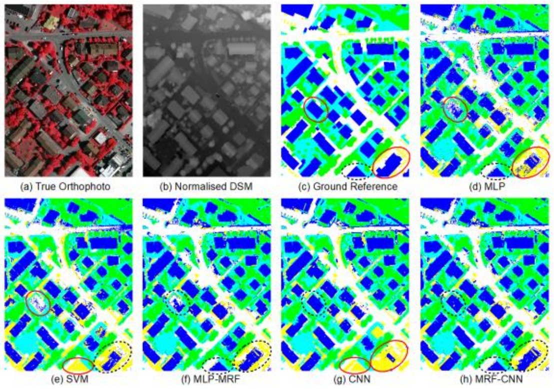


Fig. 7. Full tile prediction for No. 30. Legend on the Vaihingen dataset: white=impervious surface; blue=buildings; cyan=low vegetation; green=trees;

yellow=cars. (a) True Orthophoto; (b) Normalised DSM; (c) Ground Reference, ground reference labelling; (d, e, f, g) the inference result from MLP, SVM, MLP-MRF, CNN, respectively; (f) the proposed MRF-CNN classification result. The red and dashed circles denote incorrect and correct classification, respectively.

Figure 7 and 8 illustrate full tile predictions of Vaihingen dataset (No. 30) and Potsdam dataset (No. 05_12), with red and dashed circles highlighting broadly incorrect and correct classifications, respectively. Both MLP and SVM classifications result in salt-and-pepper effects due to pixel-level differentiation with subtle differences between them (e.g. red circles shown in Fig. 7(*d*) and 7(*e*)). The MLP-MRF (Fig. 7(*f*) and 8(*f*)) improves on the MLP (Fig. 7(*d*) and 8(*d*)) with homogeneous blocks and crisp boundary differentiation. This can be seen at the lower right side of the Building that has reduced salt-and-pepper effect (dashed circle in Fig. 7(*d*) and 8(*d*)). The CNN acquires the greatest smoothness (Fig. 7(*g*) and 8(*g*)) thanks to higher-level spatial feature representation. However, it makes some blunders by misclassifying Building as Car (red circles in Fig. 7(*g*) or falsely producing some building edge artefacts as Impervious Surface (the red circle in Fig. 8(*g*)). The MRF-CNN (Fig. 7(*h*) and 8(*h*)), solved the aforementioned problems (all dashed circles) by taking advantage of the rough set uncertainty partition as well as the subsequent decision fusion.

TABLE VI
PER-CLASS ACCURACY AND OVERALL ACCURACY (OA) FOR THE MLP, SVM, MLP-MRF, CNN AND THE PROPOSED MRF-CNN APPROACH, AS WELL AS BASELINE METHODS, FOR THE POTSDAM DATASET. THE BOLD FONT HIGHLIGHTS THE LARGEST CLASSIFICATION ACCURACY PER ROW.

Method	Imp Surf	Build ing	Low Veg	Tree	Car	OA
MLP	84.3%	81.3%	71.5%	85.6%	70.4%	82.1%
SVM	83.6%	81.8%	72.2%	84.3%	70.9%	82.4%
MLP-MRF	85.8%	83.6%	73.4%	84.8%	72.3%	83.7%
CNN	86.5%	88.7%	76.7%	87.6%	72.7%	86.5%
FCN [52]	85.5%	90.6%	75.8%	86.1%	69.8%	86.2%
SegNet [24]	82.9%	89.5%	73.1%	84.3%	70.5%	83.6%
Deeplab-v2 [53]	88.7%	93.6%	77.2%	86.5%	**83.9%**	88.2%
MRF-CNN	**90.8%**	**95.2%**	**81.5%**	**88.2%**	80.3%	**89.4%**

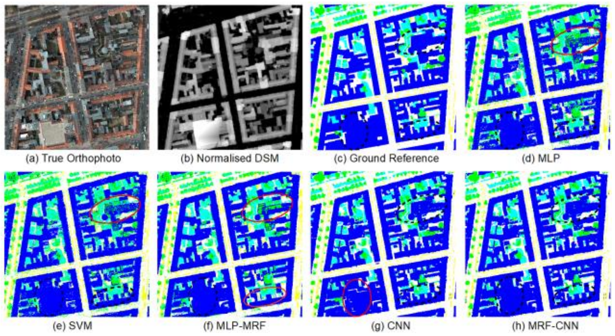


Fig. 8. Full tile prediction for No. 05_12. Legend on the Potsdam dataset: white=impervious surface; blue=buildings; cyan=low vegetation; green=trees; yellow=cars. (a) True Orthophoto; (b) Normalised DSM; (c) Ground Reference, ground reference labelling; (d, e, f, g) the inference result from MLP, SVM, MLP-MRF, CNN, respectively; (f) the proposed MRF-CNN classification result. The red and dashed circles denote incorrect and correct classification, respectively.

E. *Function of the VPRS fusion decision parameter β and step*

The VPRS fusion decision parameters (β and *step*) were analysed separately to investigate each of their contributions in describing and integrating the classification results. As illustrated by Fig. 9(*a*) and 9(*b*), relations between the fused classification accuracy and each of the parameters (while fixing the other) can be plotted. Generally, there are similar trends in terms of the influence of two parameters on classification accuracy: the accuracy increases initially until reaching the maximum accuracy at β = 0.1 and *step* around 0.075-0.1, and then decreases constantly, along with further increases of the inclusion error β (Fig. 9(*a*)) and the atomic granule *step* (Fig. 9(*b*)) respectively. This means that both β and *step* can impact the accuracy. However, compared with the step, the change in accuracy caused by β is greater accompanied by greater accuracy variation, indicating that β is the crucial factor for VPRS parameter setting. It can be imagined that a large value of β can wrongly take the CNN's problematic boundary information as positive regions, whereas the "should-be" positive regions can be eliminated by too small a value of β. In terms of *step*, the smaller its value (i.e. a finer information granularity), the larger the test samples for the VPRS will be required, to provide enough samples within each information granularity level. An atomic granularity should, therefore, ideally match with the sampling density level; otherwise, it will reduce the classification accuracy (Fig. 9(*b*)).

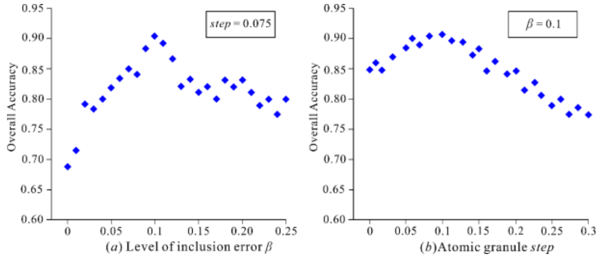


Fig. 9. Accuracies of VPRS (*a*) influenced by β when fixing the *step* as 0.075, (*b*) influenced by *step* when fixing the β as 0.1

IV. DISCUSSION

Due to the spatial and spectral complexity within VFSR imagery, any classification model prediction is inherently uncertain, including the advanced CNN classifier. Thus, for the integration of classifiers, it would be of paramount importance to discriminate the less uncertain and more uncertain results of each individual classification. A VPRS based regional fusion decision strategy was, thus, proposed to integrate the spectral-contextual-based MLP-MRF classifier with precise boundary partitions and the CNN classifier with spatial feature representations for high accuracy classification of VFSR remotely sensed imagery. The proposed MRF-CNN regional decision fusion method takes advantage of the merits of the two individual classifiers and overcomes their respective shortcomings as discussed below.

A. *Characteristics of MLP-MRF classification*

The MLP-MRF classifier is constructed based on the pixel-based MLP as its conditional probability and models the prior probability using its contextual neighbourhood information to achieve a certain amount of smoothness [14]. That is, the MLP-MRF depends primarily on the spectral feature differentiation from the MLP with consideration of its spatial connectivity/smoothness [56]. Such characteristics result in similar classification performance to the result of MLP but with less salt and pepper effect. One positive attribute of the MLP-MRF, inherited from the non-parametric learning classifier MLP, is the ability to maintain precise boundaries of some objects with high accuracy and fidelity. In particular, the classification accuracy of a pixel in the MLP model is not affected by the relative position (e.g. lying on or close to boundaries) of the object it belongs to, as long as the corresponding spectral space is separable. Some land cover classes (e.g. Clay roof, Metal roof and Shadow), with salient spectral properties that are spectrally exclusive to other classes, are therefore not only accurately classified with high classification accuracies (>90% overall accuracy), but also with less noise in comparison with the standard MLP and SVM classification results. At the same time, the MLP-MRF can elaborately identify some components of an object, for example, the Velux^{TM} windows of a building (shown by yellow circle in Fig. 6(*c*)), indicating that the object and its sub-objects might be possibly mapped accurately in future. However, the classification accuracy increase of the MLP-MRF over the MLP is not substantial or less remarkable, with just a 2-3% accuracy increase (see Table III in experiment 1 and Table V in experiment 2). In comparison with the CNN, the MLP-MRF usually demonstrates a much larger intra-class variation, which can be demonstrated by the fact that the boxplots of confidence values are larger when gradually trusting the MLP-MRF (Fig. 4). This is mainly because the MLP-MRF utilizes the spectral information in the classification process without fully exploiting the abundant spatial information appearing in the VFSR imagery (e.g. texture, geometry or spatial arrangement) [57]. Such deficiencies often lead to unsatisfactory classification performance in classes with spectrally mixed but spatially distinctive characteristics (e.g., the confusion and misclassification between Trees and Grassland or Low Vegetation that are spectrally similar, the severe salt and pepper effects on railway with linear textures, etc.).

B. Characteristics of CNN classification

Spatial features in remotely sensed data like VFSR imagery are intrinsically local and stationary that represent a coherent spatial pattern [58]. The presence of such spatial features are detected by the convolutional filters within the CNN, and well generalized into increasingly abstract and robust features through hierarchical feature representations. Therefore, the CNN shows an impressive stability and effectiveness in VFSR image classification [21]. Especially, classes like Concrete roof and Road-or-track that are difficult to distinguish from their backgrounds with only spectral features at pixel level, are identified with relatively high accuracies. In addition, classes with heavy spectral confusion in both study sites (e.g. Trees and Grassland), are accurately differentiated due to their obvious spatial pattern differences; for example, the texture of tree canopies is generally rougher than that of grassland, which is captured by the CNN through spatial feature representations. Moreover, the convolutional filters applied at each layer within the CNN framework remove all of the noise that is smaller than the size of the image patch, which leads to the smoothest classification results compared with the MLP, the SVM and the MLP-MRF (see Figure 5-8). This is also demonstrated by Figure 4, where the boxplots of the CNN are much narrower than those of the MLP-MRF.

As discussed above, the CNN classifier demonstrates obvious superiority over the spectral-contextual based MLP-MRF (and the pixel-based MLP and SVM classifiers) for the classification of the spatially and spectrally complex VFSR remotely sensed imagery. However, according to the “no free lunch” theorem [59], any elevated performance in one aspect of a problem will be paid for through others, and the CNN is no exception. the CNN also demonstrates some deficiencies for boundary partition and small feature identification, which is essential for VFSR image classification with unprecedented spatial detail. Such a weakness occurs mainly because of over-smoothness that leads to boundary uncertainties with small useful features being falsely erased, somehow similar to morphological or Gabor filter methods [60], [61]. For example, the human-made objects in urban scenes like buildings and asphalt are often geometrically enlarged with distortion to some degree (See Fig. 5(*b*) and 6(*c*)), and the impervious surfaces and the building are confused with cars being enlarged or misclassified (Fig. 7(*e*)). As for natural objects in rural areas (S2), edges or porosities of a landscape patch are simplified or ignored, and even worse, linear features like river channels or dams that are of ecological importance, are erroneously erased (e.g. Fig. 5(*b*)). Besides, certain spectrally distinctive features without obvious spatial patterns are poorly differentiated. For example, some Concrete roofs are wrongly identified as Asphalt as illustrated in Fig. 5(*c*). Previous work also found that the CNN was inferior to some global low level feature descriptors like Border/ Interior Pixel Classification when dealing with a remote sensing image that has abundant spectral but lacks spatial information [62]. However, the uncertainties in the CNN classification demonstrate regional distribution characteristics, either along the object boundaries (e.g. Fig. 5(*b*)) or entire objects (e.g. Fig. 5(*c*)). These provide the justification of regional decision fusion to further improve the CNN for VFSR image classification.

C. The VPRS based MRF-CNN fusion decision

This paper proposed to explore rough set theory for region-based uncertainty description and classification decision fusion using VFSR remotely sensed imagery. The classification uncertainties in the CNN results were quantified at a regional level, with each region determined as positive or non-positive (boundary and negative) regions by matching the correctness of a group of samples in the Test Sample *T*2. Nevertheless, in the standard rough set, most of the actual positive regions are occupied by boundary (i.e. non-positive) regions due to the huge uncertainty and inconsistency in VFSR image classification results. Such issues limit the practical application of the standard rough set because of its ignorance of the desired positive regions. A variable precision rough set (VPRS) is proposed for uncertainty description and classification integration by incorporating a small level of inclusion error (i.e. parameter β). The VPRS theory is used here as a spatially explicit framework for regional decision fusion, where the non-positive regions in this research represent the spatial uncertainties in the CNN classification result. For those positive regions of CNN classifications, including the very close to 100% correct classifications, are identified and utilized; whereas the rest (i.e. the non-positive) regions are replaced by the MLP-MRF results with crisp and accurate boundary delineation.

To integrate the CNN and the MLP-MRF classifier, the CNN was served as the base classifier to derive the classification confidence, considering its superiority in terms of classification accuracy and the regional homogeneity of classification results. Therefore, the regional decision fusion process is based on the CNN classification results, and the MLP-MRF is only trusted at the regions where the CNN is less believable (i.e. the non-positive regions). Such a fusion decision strategy achieves an accurate and stable result with the least variation in accuracy, as illustrated by the narrow box in Figure 4. The complete correctness of the MLP-MRF results at the non-positive regions are not guaranteed, but one thing is certain: the corresponding MLP-MRF results are much more accurate than those of the CNN. In fact, while the CNN accurately classifies the interiors of objects with spatial feature representations, the MLP-MRF could provide a smooth, but also crisp boundary segmentation with high fidelity [56]. These supplementary characteristics inherent in the MLP-MRF and CNN, are captured well by the proposed VPRS-based MRF-CNN regional decision fusion approach. As shown by Figure 4, although the values of the CNN confidence map decrease gradually from the centre to its boundary (i.e. the edge between the positive and non-positive regions, at 0.3 marked by the red vertical line), the classification accuracies rise constantly until reaching the maximum accuracy. For these MLP-MRF results in the non-positive regions, the corresponding non-positive regions (i.e. the problematic areas of the final fusion decision results) can be further clarified. Moreover, additional improvement might be obtained by means of imposing extra expert knowledge and/or combining other advanced classifiers (e.g. SVM, Random Forest, etc.).

In summary, the proposed method for classification data description and integration is, in fact, a general framework extensively applicable to any classification algorithms (not just for the mentioned individual classifiers), and to any remote sensing images (not just for the VFSR remotely sensed imagery). The general approach, thus, addresses the complex problem of remote sensing image classification in a flexible, automatic and active manner.

The proposed MRF-CNN relies on an efficient and relatively limited CNN network with just four layers (c.f. state-of-the-art networks, such as Deeplab-v2, built on extremely deep ResNet-101). Nevertheless, it still achieves comparable and promising classification performance with the largest accuracy overall. This demonstrates that the proposed method has practical utility, especially when facing the problems of limited computational power with insufficient training data, which are commonly encountered in the remote sensing domain when building a deep CNN network.

V. CONCLUSION

Spatial uncertainty is always a key concern in remote sensing image classification, which is essential when facing the spatially and spectrally complex VFSR remotely sensed imagery. Characterising the spatial distribution of uncertainties has great potential for practical application of the data. In this paper, a novel variable precision rough set (VPRS) based regional fusion decision between CNN and MRF was presented for the classification of VFSR remotely sensed imagery. The VPRS model quantified the uncertainties in CNN classification of VFSR imagery by partitioning the result into spatially explicit granularities that represent positive regions (correct classifications) and non-positive regions (uncertain or incorrect classifications). Such a region-based fusion decision approach reflects the regional homogeneity of the CNN classification map. The positive regions were directly trusted by the CNN, whereas non-positive regions were rectified by the MLP-MRF in consideration of their complementary behaviour in spatial representation. The proposed regional fusion of MRF-CNN classifiers consistently outperformed the standard pixel-based MLP and SVM, spectral-contextual based MLP-MRF as well as contextual-based CNN classifiers, and increased classification accuracy above state-of-the-art methods when applied to the ISPRS Semantic Labelling datasets. Therefore, this VPRS-based regional classification integration of CNN and MRF classification results provides a framework to achieve fully automatic and effective VFSR image classification.

ACKNOWLEDGMENT

This research was supported by the National Key Research and Development Program of China (NO. 2017YFB0503602) and co-funded by PhD Studentship “Deep Learning in massive area, multi-scale resolution remotely sensed imagery” (NO. EAA7369), sponsored by Ordnance Survey and Lancaster University. The authors are grateful to the anonymous referees for their constructive comments on this paper.

REFERENCES

[1] J. A. Benediktsson, J. Chanussot, and W. M. Moon, “Very High-resolution remote sensing: Challenges and opportunities [point of view],” in *Proceedings of the IEEE*, 2012, vol. 100, no. 6, pp. 1907–1910.

[2] X. Yao, J. Han, G. Cheng, X. Qian, and L. Guo, “Semantic Annotation of High-Resolution Satellite Images via Weakly Supervised Learning,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3660–3671, 2016.

[3] H. Shi, L. Chen, F. Bi, H. Chen, and Y. Yu, “Accurate urban area detection in remote sensing images,” *IEEE Geosci. Remote Sens.*

Lett., vol. 12, no. 9, pp. 1948–1952, 2015.

[4] A. Ozdarici-Ok, A. Ok, and K. Schindler, "Mapping of agricultural crops from single high-resolution multispectral images—Data-driven smoothing vs. Parcel-based smoothing," *Remote Sens.*, vol. 7, no. 5, pp. 5611–5638, 2015.

[5] J. P. Ardila, V. A. Tolpekin, W. Bijker, and A. Stein, "Markov-random-field-based super-resolution mapping for identification of urban trees in VHR images," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 6, pp. 762–775, 2011.

[6] T. Zhang, W. Yan, J. Li, and J. Chen, "Multiclass Labeling of Very High-Resolution Remote Sensing Imagery by Enforcing Nonlocal Shared Constraints in Multilevel Conditional Random Fields Model," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 9, no. 7, pp. 2854–2867, 2016.

[7] Z. Lei, T. Fang, and D. Li, "Land cover classification for remote sensing imagery using conditional texton forest with historical land cover map," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 720–724, 2011.

[8] C. Zhang, T. Wang, P. M. Atkinson, X. Pan, and H. Li, "A novel multi-parameter support vector machine for image classification," *Int. J. Remote Sens.*, vol. 36, no. 7, pp. 1890–1906, 2015.

[9] F. Pacifici, M. Chini, and W. J. Emery, "A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification," *Remote Sens. Environ.*, vol. 113, no. 6, pp. 1276–1292, 2009.

[10] P. M. Atkinson and A. R. Tatnall, "Introduction: neural networks in remote sensing," *Int. J. Remote Sens.*, vol. 18, no. 4, pp. 699–709, 1997.

[11] F. Del Frate, F. Pacifici, G. Schiavon, and C. Solimini, "Use of neural networks for automatic classification from high-resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 800–809, 2007.

[12] O. Regniers, L. Bombrun, V. Lafon, and C. Germain, "Supervised Classification of Very High Resolution Optical Images Using Wavelet-Based Textural Features," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3722–3735, 2016.

[13] R. Nishii, "A Markov random field-based approach to decision-level fusion for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 10, pp. 2316–2319, 2003.

[14] L. Wang and J. Liu, "Texture classification using multiresolution Markov random field models," *Pattern Recognit. Lett.*, vol. 20, no. 2, pp. 171–182, 1999.

[15] I. Arel, D. C. Rose, and T. P. Karnowski, "Deep machine learning - A new frontier in artificial intelligence research," *IEEE Comput. Intell. Mag.*, vol. 5, no. 4, pp. 13–18, 2010.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep Convolutional Neural Networks," in *NIPS2012: Neural Information Processing Systems*, 2012, pp. 1–9.

[17] X. Yang, X. Qian, and T. Mei, "Learning salient visual word for scalable mobile image retrieval," *Pattern Recognit.*, vol. 48, no. 10, pp. 3093–3101, 2015.

[18] E. Othman, Y. Bazi, N. Alajlan, H. Alhichri, and F. Melgani, "Using convolutional features and a sparse autoencoder for land-use scene classification," *Int. J. Remote Sens.*, vol. 37, no. 10, pp. 2149–2167, 2016.

[19] Z. Dong, M. Pei, Y. He, T. Liu, Y. Dong, and Y. Jia, "Vehicle type classification using unsupervised Convolutional Neural Network," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 2247–2256, 2015.

[20] 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.

[21] W. Zhao and S. Du, "Learning multiscale and deep representations for classifying remotely sensed imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 113, pp. 155–165, 2016.

[22] Y. Chen, H. Jiang, C. Li, X. Jia, and S. Member, "Deep feature extraction and classification of hyperspectral images based on Convolutional Neural Networks," *IEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, 2016.

[23] M. Längkvist, A. Kiselev, M. Alirezaie, and A. Loutfi, "Classification and segmentation of satellite orthoimagery using Convolutional Neural Networks," *Remote Sens.*, vol. 8, no. 329, pp. 1–21, 2016.

[24] M. Volpi and D. Tuia, "Dense Semantic Labeling of Subdecimeter Resolution Images With Convolutional Neural Networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 881–893, 2017.

[25] C. Zhang *et al.*, "A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification," *ISPRS J. Photogramm. Remote Sens.*, 2017.

[26] P. Olofsson, G. M. Foody, M. Herold, S. V. Stehman, C. E. Woodcock, and M. A. Wulder, "Good practices for estimating area and assessing accuracy of land change," *Remote Sens. Environ.*, vol. 148, pp. 42–57, 2014.

[27] Q. Wang and W. Shi, "Unsupervised classification based on fuzzy c-means with uncertainty analysis," *Remote Sens. Lett.*, vol. 4, no. 11, pp. 1087–1096, 2013.

[28] F. Giacco, C. Thiel, L. Pugliese, S. Scarpetta, and M. Marinaro, "Uncertainty analysis for the classification of multispectral satellite images using SVMs and SOMs," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3769–3779, 2010.

[29] Z. Pawlak, "Rough sets," *Int. J. Comput. Inf. Sci.*, vol. 11, no. 5, pp. 341–356, 1982.

[30] X. Pan, S. Zhang, H. Zhang, X. Na, and X. Li, "A variable precision rough set approach to the remote sensing land use/cover classification," *Comput. Geosci.*, vol. 36, no. 12, pp. 1466–1473, 2010.

[31] R. W. Swiniarski and A. Skowron, "Rough set methods in feature selection and recognition," *Pattern Recognit. Lett.*, vol. 24, no. 6, pp. 833–849, 2003.

[32] D. G. Chen, Q. He, and X. Z. Wang, "FRSVMs: Fuzzy rough set based support vector machines," *Fuzzy Sets Syst.*, vol. 161, no. 4, pp.

596–607, 2010.

[33] H. Yu, Z. Liu, and G. Wang, "An automatic method to determine the number of clusters using decision-theoretic rough set," *Int. J. Approx. Reason.*, vol. 55, no. 1 PART 2, pp. 101–115, 2014.

[34] J. Zhan and K. Zhu, "A novel soft rough fuzzy set: Z-soft rough fuzzy ideals of hemirings and corresponding decision making," *Soft Computing*, pp. 1–14, 2016.

[35] Y. Qian, X. Liang, G. Lin, Q. Guo, and J. Liang, "Local multigranulation decision-theoretic rough sets," *Int. J. Approx. Reason.*, vol. 82, pp. 119–137, 2017.

[36] Y. Chen, Y. Xue, Y. Ma, and F. Xu, "Measures of uncertainty for neighborhood rough sets," *Knowledge-Based Syst.*, vol. 120, pp. 226–235, 2017.

[37] Y. Leung, M. M. Fischer, W. Z. Wu, and J. S. Mi, "A rough set approach for the discovery of classification rules in interval-valued information systems," *Int. J. Approx. Reason.*, vol. 47, no. 2, pp. 233–246, 2008.

[38] Y. Ge, F. Cao, Y. Du, V. C. Lakhan, Y. Wang, and D. Li, "Application of rough set-based analysis to extract spatial relationship indicator rules: An example of land use in Pearl River Delta," *J. Geogr. Sci.*, vol. 21, no. 1, pp. 101–117, 2011.

[39] A. Albanese, S. K. Pal, and A. Petrosino, "Rough sets, kernel set, and spatiotemporal outlier detection," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 1, pp. 194–207, 2014.

[40] I. U. Sikder, "A variable precision rough set approach to knowledge discovery in land cover classification," *Int. J. Digit. Earth*, vol. 9, no. 12, pp. 1206–1223, 2016.

[41] Y. Ge, H. Bai, F. Cao, S. Li, X. Feng, and D. Li, "Rough set-derived measures in image classification accuracy assessment," *Int. J. Remote Sens.*, vol. 30, no. 20, pp. 5323–5344, 2009.

[42] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[43] A. Romero, C. Gatta, G. Camps-valls, and S. Member, "Unsupervised deep feature extraction for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1349–1362, 2016.

[44] D. Strigl, K. Kofler, and S. Podlipnig, "Performance and scalability of GPU-based Convolutional Neural Networks," in *2010 18th Euromicro Conference on Parallel, Distributed and Network-based Processing*, 2010, pp. 317–324.

[45] P. M. Atkinson and A. R. L. Tatnall, "Introduction Neural networks in remote sensing," *Int. J. Remote Sens.*, vol. 18, no. 4, pp. 699–709, 1997.

[46] G. M. Foody, "Mapping land cover from remotely sensed data with a softened feedforward neural network classification," *J. Intell. Robot. Syst.*, vol. 29, no. 4, pp. 433–449, 2000.

[47] R.A. Dunne and N.A. Campbell, "Neighbour-Based MLPs," in *IEEE International Conference on Neural Networks*, 1995, vol. 4, no. i, pp. 270–274.

[48] B. Tso and P. M. Mather, *Classification methods for remotely sensed data*, 2nd ed. Boca Raton, FL: CRC Press, 2009.

[49] S. Z. Li, *Markov Random Field Modeling in Image Analysis*, Third Edit. London: Springer, 2009.

[50] W. Ziarko, "Variable precision rough set model," *J. Comput. Syst. Sci.*, vol. 46, no. 1, pp. 39–59, 1993.

[51] N. Regnauld and W. a. Mackaness, "Creating a hydrographic network from its cartographic representation: a case study using Ordnance Survey MasterMap data," *Int. J. Geogr. Inf. Sci.*, vol. 20, no. 6, pp. 611–631, 2006.

[52] M. Kampffmeyer, A.-B. Salberg, and R. Jenssen, "Semantic Segmentation of Small Objects and Modeling of Uncertainty in Urban Remote Sensing Images Using Deep Convolutional Neural Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2016, pp. 1–9.

[53] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," *arXiv*, 2016. [Online]. Available: http://arxiv.org/abs/1606.00915.

[54] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, 2014.

[55] M. Berthod, Z. Kato, S. Yu, and J. Zerubia, "Bayesian image classification using Markov random fields," *Image Vis. Comput.*, vol. 14, no. 4, pp. 285–295, 1996.

[56] C. Wang, N. Komodakis, and N. Paragios, "Markov Random Field modeling, inference & learning in computer vision & image understanding: A survey," *Comput. Vis. Image Underst.*, vol. 117, no. 11, pp. 1610–1627, 2013.

[57] L. Wang, C. Shi, C. Diao, W. Ji, and D. Yin, "A survey of methods incorporating spatial information in image classification and spectral unmixing," *Int. J. Remote Sens.*, vol. 37, no. 16, pp. 3870–3910, 2016.

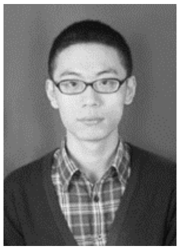
[58] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, 2016.

[59] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 67–82, 1997.

[60] S. Reis and K. Tasdemir, "Identification of hazelnut fields using spectral and gabor textural features," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 5, pp. 652–661, 2011.

[61] T. J. Pingel, K. C. Clarke, and W. A. McBride, "An improved simple morphological filter for the terrain classification of airborne LIDAR data," *ISPRS J. Photogramm. Remote Sens.*, vol. 77, pp. 21–30, 2013.

[62] K. Nogueira, O. A. B. Penatti, and J. A. dos Santos, "Towards better exploiting convolutional neural networks for remote sensing scene classification," *Pattern Recognit.*, vol. 61, pp. 539–556, 2017.



Ce Zhang received European Union Erasmus Mundus M.Sc. Degrees from both the Faculty of ITC, University of Twente, The Netherlands and the University of Southampton, U.K. in 2015. He is currently pursing Ph.D. degree in geography at Lancaster Environment Centre, Lancaster University, U.K.

His PhD project (sponsored by Ordnance Survey and Lancaster University) is about developing novel deep learning methods for the classification of very fine spatial resolution remotely sensed imagery, where he jointly works with the ImageLearn team from Ordnance Survey (Britain's Mapping Agency) and the University of Southampton, U.K.

His major research interests include geospatial data mining, machine learning, deep learning and remotely sensed image analysis.

Isabel Sargent received the Ph.D. degree from University of Southampton, UK. She is currently the Senior Research Scientist at Ordnance Survey and Visiting Academic at University of Southampton, with a specific interest in machine learning for automatic data capture and quality measurement, particularly using imagery and creating 3D spatial products.

Her major role entails generating novel research ideas, undertaking independent investigations and leading collaborative projects across academic and commercial organizations.

Xin Pan received the Ph.D. Degree from Northeast Institute of Geography and Agroecology, Chinese Academic of Science, Changchun 130102, China. He is currently an associate professor at the School of Computer Technology and Engineering, Changchun Institute of Technology, Changchun 130021, China. His main research interests include machine learning, data mining, deep learning and remotely sensed image processing.

Andy Gardiner received the B.Sc. degree in Geography from the University of Edinburgh, U.K. He is the Senior Research Scientist at Ordnance Survey, U.K. His research interests include urban and rural land cover mapping and monitoring, change detection using object-based image analysis (OBIA) and machine learning.

Jonathon Hare is a Lecturer in the Vision, Learning and Control research group in Electronics and Computer Science at the University of Southampton, U.K.

Jonathon's research interests lie in the area of multimedia data mining, analysis and retrieval, with a particular focus on large-scale multimodal approaches. This research area is at the convergence of machine learning and computer vision, but also encompasses other modalities of data, such as remotely sensed imagery. The long-term goal of his research is to innovate techniques that can allow machines to understand the information conveyed by multimedia data and use that information for fulfilling the information needs of humans. He has published over 70 articles in peer-reviewed conferences and journals.

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 Dean of the Faculty of Science and Technology, Lancaster University, Lancaster, U.K. and a Professor of Spatial Data Science at Lancaster Environment Centre. He is also a Visiting Professor with Queen’s University Belfast, Belfast, U.K. and the Chinese Academy of Sciences, Beijing, China. He was previously a Professor of Geography at the University of Southampton (for 21 years; 13 years as a full Professor), where he is currently a Visiting Professor. He has authored over 250 peer-reviewed articles in international scientific journals and around 50 refereed book chapters. He has also edited nine journal special issues and eight books. 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.

Professor Atkinson is an Associate Editor of Computers and Geosciences and serves on the editorial boards of several journals including *Geographical Analysis*, *Spatial Statistics*, *International Journal of Applied Earth Observation and Geoinformation*, and *Environmental Informatics*. He also serves on various international scientific committees.