Research Article

An Ontology-Based Framework for Complex Urban Object Recognition through Integrating Visual Features and Interpretable Semantics

Xiao Xie,^{1,2} Xiran Zhou ^{1,3} Jingzhong Li,^{2,4} and Weijiang Dai⁵

¹School of Environment Science and Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, China

²Key Lab for Environmental Computation and Sustainability of Liaoning Province, Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang 110016, China

³Artificial Intelligence Research Institute, China University of Mining and Technology, Xuzhou 221116, China

⁴College of Urban and Environmental Sciences, Xuchang University, Xuchang 461000, China

⁵Nuclear Industry Huzhou Engineering Investigation Institute, Huzhou 313000, China

Correspondence should be addressed to Xiran Zhou; xrzhou@cumt.edu.cn

Received 15 May 2020; Revised 2 July 2020; Accepted 14 July 2020; Published 10 September 2020

Guest Editor: Jun Yang

Copyright © 2020 Xiao Xie et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Although previous works have proposed sophisticatedly probabilistic models that has strong capability of extracting features from remote sensing data (e.g., convolutional neural networks, CNN), the efforts that focus on exploring the human's semantics on the object to be recognized are required more explorations. Moreover, interpretability of feature extraction becomes a major disadvantage of the state-of-the-art CNN. Especially for the complex urban objects, which varies in geometrical shapes, functional structures, environmental contexts, etc, due to the heterogeneity between low-level data features and high-level semantics, the features derived from remote sensing data alone are limited to facilitate an accurate recognition. In this paper, we present an ontology-based methodology framework for enabling object recognition through rules extracted from the high-level semantics, rather than unexplainable features extracted from a CNN. Firstly, we semantically organize the descriptions and definitions of the object as semantics (RDF-triple rules) through our developed domain ontology. Secondly, we exploit semantic web rule language to propose an encoder model for decomposing the RDF-triple rules based on a multilayer strategy. Then, we map the low-level data features, which are defined from optical satellite image and LiDAR height, to the decomposed parts of RDF-triple rules. Eventually, we apply a probabilistic belief network (PBN) to probabilistically represent the relationships between low-level data features and high-level semantics, as well as a modified TanH function is used to optimize the recognition result. The experimental results on lacking of the training process based on data samples show that our proposed approach can reach an accurate recognition with high-level semantics. This work is conducive to the development of complex urban object recognition toward the fields including multilayer learning algorithms and knowledge graph-based relational reinforcement learning.

1. Introduction

Urban scenes characterize the representations of land cover and land use on the Earth's surface and the interaction of natural forces and human activities. Recently, because of the emergence of Internet of Things and cyber infrastructure, urban areas are passing through a rapid development forwarding digital cities and even smart cities, making the urban landscape and objects in urban areas become the most

complicated ever [1–4]. The objects locating in the urban scenes, which varies in geometrical shapes, functional structures, environmental contexts, etc., which are named complex urban objects in this paper, might inherit different social roles and urban functions. To effectively and efficiently recognize the complex characteristics of the complex urban object that might be useful for landscape planning, land management, and traffic monitoring in the era of digital economy. A large number of scholars have focused on using

remotely sensed imagery for land scene recognition [2, 5–9]. For example, the emergence of deep learning approaches has significantly shaped the frameworks of object recognition from remote sensing data. A variety of CNN models continuously have been breaking records of the state-of-the-art accuracy and precision regarding object recognition from remote sensing [10].

However, these approaches that heavily rely on features might lead to weak performance in urban object recognition. Rules, knowledge, and explainable features are back to attract researchers' attentions [11, 12]. Recently, the community of pattern recognition in remote sensing proposes some discussions that whether the deep learning models are overtrusted. On one hand, researchers notice the expensive cost of training data preparation for the deep learning-enhanced object recognition [13]. On other hand, the amount of object type that a CNN specially designed for object recognition can recognize seems limited and heavily relies on human's labeling. Thus, how to extend the scope of the deep learning-enhanced object recognition becomes a main concern in the recent works [14, 15].

To deal with the challenge mentioned above, multiple types of probabilistic relational models (PRMs), involving Bayesian belief network, Markov chain model, Monte Carlo, etc., have been successfully exploited by a wide variety of tasks for object recognition, such as road extraction [16], terrestrial modeling [17], urban pattern [18], and change detection [19]. PRMs have some advantages on modeling the relationships included in data, such as addressing the uncertain parts of data, handling with overfitting, and exhibiting a robustness of noises. However, the features extracted by PRMs still belong to low-level features involving color, texture, shape, etc. These low-level features cannot accurately represent the characteristics of objects to be recognized, since the characteristics are described by high-level human's semantics involving concepts and definitions [20, 21].

The substantial difference between high-level semantics and low-level data features has become the critical challenge to generate accurate object recognition [8]. Thus, the approaches for explicitly representing the content or the semantics of remote sensing data have attracted considerable attentions. Initially, several commonly used relational models (e.g., Entity-Relationship model and Graph-based model) have been applied to model the relationships between different objects in remote sensing images [22, 23]. However, the relational model cannot explicitly reveal the uncertain and hidden relationships in remote sensing data. For example, the relational model is limited to represent the information, involving bright color, complicated structure, etc., the inherit meaning of which cannot be directly extracted from low-level features. More important, the relational model cannot support to produce extra knowledge through logical reasoning.

Compared with traditional relational models, ontology has been viewed a commonly used tool for effectively organizing semantic information and knowledge based on high-end semantic web techniques [24, 25]. Ontology-based methodologies for remote sensing data analysis have

attracted wide-ranging discussions in recent years [26–29]. Reviewing the ontology-based approaches that are applicable for object recognition, we classify these approaches into three categories. The first type of approach semantically defines the characteristics and the relationships of each object for interpreting the result of classification or recognition [30–33]. This type of approach only applies knowledge from the ontological definition to postprocess the recognition result, rather than using the knowledge to assist the computational processing. Then, the second type of approach focuses on predefining some rules to extract low-level image features [34–39]. However, the knowledge and rules in these approaches act as the auxiliary information for object recognition alone, which still lacks incorporating rules into the process of computation [27, 40, 41]. Recently, the third type of approach has been created to exploit the rules in ontology to support the computational procedure of recognition, through establishing parameters that includes high-level semantics from ontology and the low-level features from remote sensing image [42, 43].

Above of all, based on ontology and semantic web techniques, the existing methods have capability to semantically model and formally organize the knowledge about the objects to be recognized [44]. Additionally, the low-level features for specific object can be explicitly modeled in ontology to support the computational processing of recognition [44, 45]. However, many further investigations are still needed. First, the previous ontology-based approaches lack to apply probability to quantitatively model the semantics of objects to be recognized in the computation. Second, an unsupervised model, which is created based on ontology to map high-level semantics and low-level features, has not been explored yet for object recognition.

Aiming to combining the external knowledge and the techniques of data-driven probabilistic analysis [46, 47], this paper proposes an unsupervised approach for object recognition from remote sensing image, which integrates the ontology that formally defines the knowledge of remote sensing image and object, and a probabilistic belief network (PBN) for probabilistically representing the features of remote sensing image and the semantics of objects by a more accurate way. The remainder of this paper is organized as follows. Section 2 shows the details of our proposed methodology framework. Section 3 demonstrates the experiment by using different popular PRMs and our proposed approach. Section 4 discusses the contribution of our work and the prospective of related works.

2. Methodology

The proposed methodology mainly focuses on employing the semantics to build an unsupervised approach for object recognition, rather than designing a complex architecture including feature extraction, feature learning, and feature-based classification. The framework of our proposed methodology consists of four sequential steps. (1) Creating region proposal. (2) Developing domain ontology to model the semantics of object to be recognized and generating

knowledge (rules) from the semantics defined in ontology. (3) Creating the low-level features corresponding to these rules. (4) Mapping the low-level features and high-level semantics. (5) Recognizing the object through measuring the similarity between the features from ontological definitions and the low-level features. Additionally, generating knowledge from ontology is a procedure based on top-down architecture, as well as the procedure of creating low-level features corresponding to the knowledge. The details of the methodology framework are elaborated in the following sections.

2.1. Region Proposal Generation. We design a strategy that contains two different procedures for generating region proposals. Firstly, we segment the whole optical satellite image through object-based segmentation named simple linear iteration cluster-SLIC [48]. Then, based on the LiDAR height raster generated from LiDAR cloud point data, we segment the whole height raster through aspect difference. The following expression shows the approach for computing aspect difference based on eight types of connectivity:

$$\text{asptdif}_d \begin{cases} 1, & \max(|p - p_d|) \geq \theta, \\ 0, & \max(|p - p_d|) < \theta, \end{cases} \quad d = [1, 8], \quad (1)$$

where d denotes the index of direction ranging from 1 to 8, which, respectively, represents the directions starting clockwise at the north, p denotes a pixel of the local window, p_d denotes its neighbored pixels over direction d , and θ denotes the threshold. The value of this threshold relies on the shape of the object to be recognized. Generally, the θ applicable for man-made objects and natural objects are 90 and 45 [49, 50], separately.

If the object to be recognized has distinct altitude variation, its segmented result by aspect difference serves as the region proposal. Otherwise, the parameters proposed by Uijlings et al [51] are used to generate region proposal based on the result by object-based segmentation. Specifically, larger scale is applied during the object-based segmentation to reduce the number of segmented fragments.

2.2. Ontology-Based Knowledge Generation. Domain ontology can transform the descriptions of to-be-recognized object into the semantics and to facilitate logic reasoning based on the semantics [4, 9]. Based on Protégé software [52], we developed the domain ontology by semantic standards which include Extensible Markup Language (XML), Resource Description Framework Schema (RDFS), and Web Ontology Language (OWL). In this paper, we develop two domain ontologies: a domain ontology for transforming the descriptions of to-be-recognized object into the semantics (GeoOnto) and another domain ontology for transferring the low-level features into semantics (ImgOnto). Specifically, ImgOnto is developed based on the image region, rather than image pixels. The conceptual model of these two domain ontologies is shown in Figure 1, which includes four components: *Classes*, *Instances*,

Properties, and *Functions*. *Classes* refer to set or type of individuals. *Instances* refer to individuals. *Properties* refer to the relationships that define the connection between classes and individuals. *Functions* refer to the attribute of relationships.

The semantics in GeoOnto and ImgOnto are encoded as semantics by RDF triples, the structure of which is shown as follows:

$$\text{subj} - P - \text{obj}, \quad (2)$$

where subj , P , and obj denote *subject*, *predicate*, and *object*, respectively. Specifically, Protégé software and some literature define *predicate* as *property* or *relation*. *subject* and *object* belong to classes or instances, and *predicate* refers to relationships and its corresponding restrictions. For example, the description “road is one type of impervious surface” is written as RDF triple: $\text{road-typeOf-impervious surface}$, where “road,” “typeOf” and “impervious surface” are *subject*, *predicate*, and *object*, respectively.

Then, based on the RDF triples, the knowledge is represented by rules, which are generated through logic reasoning through semantic web rule languages (SWRL), the structure of which is shown as follows:

$$P_1(?x_1, ?x_2) \wedge \dots \wedge P_k(?x_{k-1}, ?x_k) \Rightarrow P(?x_1, ?y), \quad (3)$$

where $P(?x_1, ?y)$ refers to the RDF triple: $x_1 - P - y$. The same applies to $P_1(?x_1, ?x_2)$, $\dots$, and $P_k(?x_{k-1}, ?x_k)$, as well. k is the total number of RDF triples that lead to $P(?x_1, ?y)$.

2.3. Low-Level Features from Optical Satellite Image and LiDAR Height. Each waveband of the optical satellite image and LiDAR height serves as the sources for generating low-level features. The low-level features are divided into three types: color features (height features), textural features (height textural features), and shape features (height shape features). The methods for computing these low-level features based on different sources are shown in Table 1. The *parameters (rule)* refer to the low-level features defined from descriptions which are modeled by ontology, and the *parameters (test)* refer to the low-level features extracted from each waveband of optical satellite image and LiDAR height raster.

2.3.1. Shape Features. The boundary of region proposal which is obtained by aspect difference from LiDAR height is used as the shape features for the object that has distinct altitude variation (e.g., buildings and trees). After acquiring the boundaries by aspect difference from LiDAR height, we further compute the statistical and texture features in the region proposal.

Otherwise, for the object that has no obvious altitude changes, the region proposal of the object created refers to the minimum bounding rectangle (MBR) of this object, rather than its true boundaries. After acquiring the boundaries of this object from the optical satellite image, we further compute the color and texture features in the region proposal.

FIGURE 1: Conceptual model of our developed ontologies.

TABLE 1: Low-level features and computing methods.

Features	Parameters (rule)	Parameters (test)
Color features (each waveband of optical satellite image and)	Max	Mean
	Min	Standard deviation
	Standard deviation (value range)	Skewness
	Skewness (value range)	Kurtosis
	Kurtosis (value range)	
Height features (LiDAR height raster)	Max	Mean
	Min	Standard deviation
	Standard deviation (value range)	Skewness
	Skewness (value range)	Kurtosis
	Kurtosis (value range)	
Textural features (fused waveband of optical satellite image and)	Normalized symmetrical matrix of grey-level co-occurrence matrix (GLCM)*	Normalized symmetrical matrix of GLCM
	GLCM features**	GLCM features
Height textural features (LiDAR height raster)	Normalized symmetrical matrix of grey-level co-occurrence matrix (GLCM)*	Normalized symmetrical matrix of GLCM
	GLCM features**	GLCM features
Shape features (fused waveband of optical satellite image)	Segmentation fusion [51]	Segmentation fusion [51]
Height shape features (LiDAR height raster)	Region proposal generated by aspect difference	Region proposal generated by aspect difference

*There are four normalized symmetrical matrixes of GLCM in north, east, south, and west directions, respectively. ** GLCM features are computed based on each normalized symmetrical matrix.

2.3.2. Image Color and Height Statistics. For each region proposal, we computed its color based on each waveband of optical satellite image and its height features based on the LiDAR height raster. For the regional proposal generated from optical satellite image, we compute an independent color feature based on each waveband of optical satellite

image. Otherwise, for the region proposal generated from LiDAR height raster, we compute the height feature based on LiDAR height alone. Both color feature and eight feature includes commonly used statistical parameters involving max, min, mean, and standard deviation and skewness and kurtosis, the expressions of which are shown as follows:

$$\left\{ \begin{array}{l} \text{skew} = \frac{1}{N} \times \sum_{n=1}^N \frac{(V_n - \text{mean})^3}{\text{std_dev}^3}, \\ \text{kurt} = \frac{1}{N} \times \sum_{n=1}^N \frac{(V_n - \text{mean})^4}{\text{std_dev}^4}, \end{array} \right. \quad (4)$$

where skew denotes skewness, kurt denotes kurtosis, N denotes the total number of pixel in orthogonal retrieve window, V_n denotes value of each pixel, and mean and std_dev denote the result of mean and standard deviation, respectively.

2.3.3. Texture Features. Textural features and height textual features for the regional proposal are, respectively, computed based on the fused waveband of optical satellite image and LiDAR height raster. The fused waveband of the optical satellite image is obtained by

$$\text{img}_{\text{gray}} = R \times 0.299 + G \times 0.587 + B \times 0.114, \quad (5)$$

where img_{gray} refers to the gray image after fusing three wavebands and R , G , and B , respectively, denote red, green, and blue waveband of the optical satellite image.

We exploit a gray-level co-occurrence matrix (GLCM) to, respectively, extract textural features and height textual features from the fused waveband and LiDAR height [53]. The details of computing GLCM can be read in this link: <http://www.fp.ucalgary.ca/mhallbey/tutorial.htm>. In this paper, the texture features need to consider the influence of rotation, which are not available in the traditional GLCM. Thus, we compute two sets of the normalized symmetrical matrix: one set includes four matrices over the cardinal directions involving north, west, south, and east, and

another set contains four matrices over the diagonal directions involving northeast, northwest, southeast, and southwest. The expressions for computing these GLCM features are shown in the following two equations:

$$\left\{ \begin{array}{l} \text{energy} = \sum_{x=1, y=1}^N (E_{x,y})^2, \\ \text{entropy} = \sum_{x=1, y=1}^N -\ln E_{x,y} \times E_{x,y}, \\ \text{contrast} = \sum_{x=1, y=1}^N E_{x,y} \times (x - y)^2, \\ \text{homogeneity} = \sum_{x=1, y=1}^N \frac{E_{x,y}}{1 + (x - y)^2}, \\ \text{correlation} = \sum_{x=1, y=1}^N E_{x,y} \times \frac{(x - \mu) \times (y - \mu)}{v}, \\ \mu = \sum_{x=1, y=1}^N x \times E_{x,y}, \quad v = \sum_{x=1, y=1}^N (x - \mu)^2 \times E_{x,y}, \end{array} \right. \quad (6)$$

where $E_{x,y}$ refers to the element of the normalized symmetrical matrix in location x and y and N refers to the total number of gray levels specified for the image region included in the boundaries. Based on equation (6), another two GLCM features, shade and prominence, are computed by the following expression:

$$\left\{ \begin{array}{l} \text{shade} = f(S) \times S^{1/3}, \quad S = \sum_{x=1, y=1}^N \frac{(i + j - 2\mu)^3 \times E_{x,y}}{v^{3/2} \times (2 \times (1 + \text{correlation}))^{3/2}}, \\ \text{prominence} = f(P) \times P^{1/4}, \quad P = \sum_{x=1, y=1}^N \frac{(i + j - 2\mu)^4 \times E_{x,y}}{4 \times v^4 \times (2 \times (1 + \text{correlation}))^2} \end{array} \right., \quad f(X) = \begin{cases} -1: f(X) < 0 \\ 0: f(X) = 0 \\ 1: f(X) > 0 \end{cases} \quad (7)$$

Moreover, to offset the limit of GLCM, the textural result obtained by GLCM is sensitive to rotation variation, we compute the GLCM features for the original image region, as well as its rotated ones by pivoting 90, 180, and 270°.

2.3.4. Data Structure of Low-Level Features. Above all, the low-level feature (F), which includes color (statistic), texture, and shape features, is a three-dimensional array that has the following data structure:

$$\left\{ \begin{array}{l} F = [F_{\text{img}}, F_{\text{ldr}}], \\ F_d = [f_{\text{stat}, n1}, f_{\text{glcm}, n2}, f_{\text{shap}, n3}], \quad n1 = 1, \dots, 4, \\ n2 = 1, \dots, 16, \quad n3 = 1, 2. \end{array} \right. \quad (8)$$

In equation (8), the first dimension includes F_{img} and F_{ldr} , where F_{img} and F_{ldr} , respectively, refer to the low-level features extracted from optical satellite image and LiDAR height. The second dimension is F_d , where $d \in [\text{img}, \text{ldr}]$, which includes $f_{\text{stat}, n1}$, $f_{\text{glcm}, n2}$, and $f_{\text{shap}, n3}$. The structure

inside of F_d is the third dimension. $f_{\text{stat},n1}$ is a list that records color (statistical) features of each waveband of optical satellite image ($f_{\text{stat},1}$, $f_{\text{stat},2}$, and $f_{\text{stat},3}$) and height feature ($f_{\text{stat},4}$). $f_{\text{glcm},n2}$ is a list including GLCM features, where $n2 \in [1, 8]$ and $n2 \in [9, 16]$, respectively, are the index of GLCM features generated from the fused waveband of the optical satellite image and the index of height GLCM features derived from LiDAR height. $f_{\text{shap},n3}$ contains two matrixes that, respectively, refers to the shape features from the optical satellite image ($f_{\text{shap},1}$) and the height shape features created by aspect difference from LiDAR height ($f_{\text{shap},2}$).

2.4. Ontological Rules for Object Recognition

2.4.1. Mapping between High-Level Semantics to Low-Level Features. In ontology, high-level semantics are explicitly represented by RDF-triples rules. To map the high-level semantics to low-level features, as shown in Table 1, we aim to decompose these RDF-triples rules into indecomposable terms (*subject*, *predicate*, and *object*) and create low-level features corresponding to each term. Since the semantics of a to-be-recognized object generally consists of multiple RDF triples (e.g., the demo in Table 1), the decomposition is gradually completed using a layer-by-layer procedure through the proposed SWRL-based fuzzy encoder shown in Figure 2.

In this encoder model, the encoding outcome of the former layer acts as the input for encoding of the next layer; through this way each layer is mutually connected as multilayers processing architecture. The initial input of the encoder model is the rule set, which includes a set of rules that has the similar structure shown in equation (2). In the 1st layer, the rule set is decomposed into multiple rules: $R_1, \dots, R_n$. Each of these decomposed rules is as one node in the 1st layer. In the 2nd layer, each rule is further decomposed into multiple independent RDF triples that carry the similar structure shown in equation (3). Furthermore, the next layer further decomposes each independent RDF triple into three terms: *subject*, *property*, and *object*.

2.4.2. Probabilistic Prediction on Similarity between High-Level Semantics to Low-Level Features. PBN is a probabilistic graphical model, among which belief and network, respectively, is the probability of observing a state, and a directed acyclic graph of probability inference [54]. In object recognition, belief refers to the probability of recognizing (believing) an image object as the target object, and network refers to the model of recognition. Moreover, the network is comprised of nodes and links, which refer to features and relationships, respectively. PBN has two advantages based on this network architecture. First, one node can be connected with other nodes, which makes it available for multivariable-based probabilistic inference. Second, the graph of PBN and conceptual graph of ontology is easier to be integrated to enable ontological rules using in the procedure of recognition. This integration is shown in the procedure between the final layer and low-level features of Figure 2.

The conceptual model of our proposed approach based on PBN for recognition is completed by the following expression:

$$\Pr(O | w^T \times F^{\text{com}}), \quad (9)$$

where w^T denotes the weighting matrix for controlling the impacts of each feature of F in equation (8). The approach for computing w will be elaborated in Section 2.4.3. F^{com} refers to the low-level features matched between the low-level features defined from ontology and the ones extracted from the image regions in orthogonal retrieve window. Thus, $\Pr(O | w^T \times F^{\text{com}})$ means the probability of recognizing (believing) an object as the target object, given assemblies of multiple matched features ($w^T \times F^{\text{com}}$).

Moreover, based on equation (8), we have $F^{\text{com}} \subset [F_{\text{img}}^{\text{com}}, F_{\text{ldr}}^{\text{com}}]$. Equation (9) is extended as the following expression:

$$\Pr(O | w^T \times F^{\text{com}}) = \prod \Pr(O | w^T \times F^{\text{com}}), \quad d = [\text{img}, \text{ldr}]. \quad (10)$$

Then, based on $F_d = [f_{\text{stat},n1}, f_{\text{glcm},n2}, f_{\text{shap},n3}]$, from equation (8), $\Pr(O | w^T \times F^{\text{com}})$ in equation (10) is extended as the following expression:

$$\begin{aligned} \Pr(O | w_d^T \times F_d^{\text{com}}) &= \Pr(O | w_{\text{stat},n1}^T \times f_{\text{stat},n1}^{\text{com}}, w_{\text{glcm},n2}^T \times f_{\text{glcm},n2}^{\text{com}}, w_{\text{shap},n3}^T \times f_{\text{shap},n3}^{\text{com}}) \\ &= \Pr(O | w_{\text{stat},n1}^T \times f_{\text{stat},n1}^{\text{com}}) \times \Pr(O | w_{\text{glcm},n2}^T \times f_{\text{glcm},n2}^{\text{com}}) \times \Pr(O | w_{\text{shap},n3}^T \times f_{\text{shap},n3}^{\text{com}}), \end{aligned} \quad (11)$$

where the implicate content of matched low-level features $f_{\text{stat},n1}^{\text{com}}$, $f_{\text{glcm},n2}^{\text{com}}$, and $f_{\text{shap},n3}^{\text{com}}$, respectively, corresponds to the features $f_{\text{stat},n1}$, $f_{\text{glcm},n2}$, and $f_{\text{shap},n3}$ in equation (8).

We apply two separate approaches to measure the similarity for these matched low-level features (F^{com}). The first approach is used for statistical (color) features and GLCM features, or $f_{\text{stat},n1}^{\text{com}}$ and $f_{\text{glcm},n2}^{\text{com}}$ in equation (11). Similarity is measured by the following expression:

$$f^{\text{com}} = \text{sim}(f_{\text{onto}}, f_{\text{test}}) = \begin{cases} 1: f_{\text{test}} \in f_{\text{onto}}, \\ 0: f_{\text{test}} \notin f_{\text{onto}}, \end{cases} \quad (12)$$

where $f^{\text{com}} \in f_{\text{stat},n1}^{\text{com}}$ or $f_{\text{glcm},n2}^{\text{com}}$ and f_{onto} and f_{test} , respectively, denote one statistical (color) or GLCM feature derived from ontology and the image regions in orthogonal retrieve window. Additionally, f_{test} is a single value, whereas f_{onto} is a range of values. $f_{\text{test}} \in f_{\text{onto}}$ denotes that f_{test} is inside the value range of f_{onto} .

FIGURE 2: Architecture of SWRL-based encoder.

The second approach is designed for measuring shape features, or $f_{\text{shap},n3}^{\text{com}}$, in equation (11). Since shape features are represented as a matrix, rather than a single value, the similarity of shape features is measured by the modified mean absolute error (MAE). In a two-dimensional space, traditional approach calculates the MAE based on horizontal and vertical dimension, respectively. In our modified MAE, we, respectively, divide the horizontal and vertical dimensional MAE into eastern and western dimensions and northern and southern dimensions. Then, the similarity of shape features is measured based on four MAEs being oriented to the north, south, east, and west.

Assuming the boundaries defined from ontology and derived from the image regions in orthogonal retrieve window as m_{onto} and m_{test} . The modified MAE between m_{onto} and m_{test} includes $\text{mae}(n|m_{\text{onto}}, m_{\text{test}})$, $\text{mae}(s|m_{\text{onto}}, m_{\text{test}})$, $\text{mae}(e|m_{\text{onto}}, m_{\text{test}})$, and $\text{mae}(w|m_{\text{onto}}, m_{\text{test}})$, where n , s , e , and w , respectively, refers to north, south, east, and west. Then, the similarity is measured based on an idea; if m_{onto} and m_{test} were more similar at multiple scales, they should be mutually closer to be paralleled. Being paralleled is represented by the following modified MAE:

$$f^{\text{com}} = \text{sim}(m_{\text{onto}}, m_{\text{test}}) = \begin{cases} 1: \sum \text{stad_dev}(\text{mae}(\text{dir}|M_{\text{onto}}, M_{\text{test}})) \leq \vartheta, \\ 0: \sum \text{stad_dev}(\text{mae}(\text{dir}|M_{\text{onto}}, M_{\text{test}})) > \vartheta, \end{cases} \quad (13)$$

where dir is the direction that includes north, south, west, and east.

2.4.3. Weighted Probability Optimization. In previous probabilistic relational models without supports of semantics, the weighted vector w^T in equation (9) is acquired from data, rather than the semantics. As discussed previously, however, the low-level features derived from an image are different from the semantics. Thus, in this paper, the w^T is given based on high-level semantics. In equation (11), previous probabilistic models semantically define the weight of image-level features for the result of $\text{Pr}(O|w^T \times F^{\text{com}})$; these models fail to give adequate attention to a critical issue: the impact of each low-level features for to-be-recognized object are varied. Generally speaking, the impact of each low-level feature cannot be represented by the linear function. This critical issue drives us to focus on optimizing the probabilistic result from PBN through computing each

weight based on two sequential steps: qualitative computing (semantics) and quantitative computing (nonlinear regression). Qualitative computing is conducted through the following strategy that involves three conditions:

- (i) Condition 1: if a low-level feature was not semantically defined in ontology, then the weight for this feature equals to 0
- (ii) Condition 2: if the importance of m low-level features are equally defined in ontology, then the weight for this feature equals to $1/m$
- (iii) Condition 3: if the importance of a low-level feature is semantically defined in a biased way, then the weight for this feature equals ω ($\omega \in [0, 1]$), which is derived from the restrictions for the property of these triples

Condition 3 is more common to be observed among these three conditions. According to the research achievements on selective attention recognition [55, 56], due to the

restriction of conscious experience, humans can only recognize one thing at one time. This means that the features or object described prior are more important than the ones described later. Thus, in Condition 3, the weight can be determined by not only the degree defined in ontology but also the significant degree of features. For example, in the case of the definition in Table 2, three definitions are ranked based on two indexes:

- (i) Degree: generally $> a$ majority of $>$ restriction “in a city or town”
- (ii) The sequence as these definitions appear

Then, the weights ranked by qualitative computing are further quantitatively processed through implementing a modified activation function. Many multilayer processing models have successfully applied a great number of activation functions, such as logistic, Gaussian, Tan, softsign [57], and Re [58, 59], to probabilistically predict the uncertain conditions. In this paper, TanH is exploited due to two reasons. First, previous literature had claimed that TanH was the appropriate form as an activation function in a nonlinear classifier [60]. Moreover, TanH presses close to the nonlinear variations in modeling the probability of recognition based on multiple features.

To make TanH function more fitting to the proposed algorithm, we make some modifications. Based on the weight of each matched feature are ranked by qualitative computing as $w_1, w_2, \dots, w_r$, where r is the total number of matched low-level features. From the number of $f_{stat,n1}^{com}$, $f_{glcm,n2}^{com}$, and $f_{shap,n3}^{com}$ in equation (13), we have $r \leq 20$, since $n1 + n2 + n3 = 20$. The following expression is proposed to quantitatively measure the impact of each weight:

$$\left\{ \begin{array}{l} w_{\sigma} = \text{our_tanh}(\sigma) = \frac{2}{1 + e^{-(\sigma) \times a}} + b, \\ \sum_{n=1}^{\sigma} \text{our_tanh}(n) = 1, \end{array} \right. \quad (14)$$

where σ refers to the ranking index of each feature by qualitative computing. The outline of equation (14) is shown in Figure 3. In equation (14), a and b are two constants, which are used to create a nonlinear TanH function to accurately fit the priority of each low-level feature.

Eventually, by substituting each weight obtained from equation (14) into equation (13), we obtain the $\text{Pr}(O|w^T \times F^{com})$ between an image object and the object to be recognized. According to $\text{Pr}(O|w^T \times F^{com})$, we determine if this image region in orthogonal retrieve window belonged to the object to be recognized.

The materials and methods section should contain sufficient detail so that all procedures can be repeated. It may be divided into headed sections if several methods are described.

3. Experimental Analysis

The first experimental data is from CAP LTER project [61], which has been managed by Julie Ann Wrigley Global Institute of Sustainability, Arizona State University. Figure 4(a)

shows two types of remote sensing data: multispectral Quickbird and LiDAR height, which cover a partial area of Tempe, Arizona. The spatial resolution and image size of multispectral Quickbird image are 0.63 meters and 1295×1843 pixels, and this image was produced on March 30th, 2008. LiDAR height has a spatial resolution of 1 meter, an image size of 1295×1843 pixels, and was produced on May 5th, 2008. Specifically, there is a missing area over both optical satellite image and LiDAR height data. Moreover, resolution difference exists between the optical satellite image and LiDAR height. Thus, we conducted a registration between LiDAR height and the optical satellite imaged by ERDAS software.

The second experimental datasets are shown in Figure 4(b), which consist of an optical satellite image obtained from Google Earth and LiDAR height accessed from a cyber infrastructure named OpenTopography [62]. This experimental data covers main body of Embry-Riddle Aeronautical University-Prescott, Arizona. The spatial resolution and image size of optical satellite image are 1 meter and 1100×1100 pixels, and this image was produced on September 2009. LiDAR height has a spatial resolution of 1 meter, an image size of 1100×1100 pixels, and was produced on January 2014. In the previous work, we have testified land cover classification based on LiDAR data alone [49]. We recognized the object based on the semantics converted from the optical satellite image and LiDAR height.

Based on the classification system and remote sensing data involving optical satellite image and LiDAR height, the objects to be recognized in the first experimental data consists of residential buildings and commercial buildings, trees have height greater than 10 meters, trees have height over 5 meters to 10 meters, trees have height over 1.5 meters to 5 meters, circular roads, and swimming pools. Figure 5 illustrates the samples of these objects to be recognized.

We followed the workflow mentioned in Section 2 to create low-level features from the semantics of each category of the object. Table 2 shows an example of transforming the descriptions on object “street” to semantics and rules in domain ontology.

Then, 2e uses Rules 2 and 5 in Table 2 to illustrate how to map the high-level semantics and low-level features through the proposed multilayer architecture. The mapping procedure is shown in Figure 4. The initial input is the rule set involving Rule 2 and Rule 5. In the 1st layer, this rule set is decomposed into two independent rules: Rule 2 and Rule 5. In the 2nd layer, these two rules are decomposed into three RDF triples. Especially, Rule 2 and Rule 5 has the same triple: covers (*?image_object*, *?asphalt*). Then, these three RDF triples are decomposed into multiple *subjects*, *predicates*, and *objects* in the next layer. In the 4rd layer, the rules involving low-level features are created for each of these terms. Specifically, we use the node named *asphalt* as the illustration. In the 4th layer, the low-level features for *asphalt* are organized as a rule that contains four RDF triples. In the 5th layer, this rule is decomposed as four rules. Next, each of these rules is decomposed into *subjects*, *predicates*, and *objects*. These *subjects*, *predicates*, and *object* in the 6th layer are the ultimate results of the proposed multilayer decomposition.

TABLE 2: Demo of generating semantics and rules in ontology.

Definition	Description 1: generally, the covered materials are asphalt or bitumen.
	Description 2: a majority of street contains yellow or white center lines
Semantics (RDF triples)	Description 3: a road in a city or town that has houses or other buildings on one or both sides (from Merriam-Webster)
	Triple 1 from description 1: <i>street</i> covers <i>asphalt</i> Triple 2 from description 1: <i>street</i> covers <i>bitumen</i> Triple 3 from description 2: <i>street</i> contains <i>center_line</i> Triple 4 from description 3: <i>street</i> isA <i>road</i> Triple 5 from description 3: <i>house</i> isOnSidesOf <i>street</i> Triple 6 from description 3: <i>building</i> isOnSidesOf <i>street</i>
Rule set by SWRL	Rule 1 from triple 4: $\text{isA}(\text{?image_object}, \text{?road}) \Rightarrow \text{isA}(\text{?image_object}, \text{?street})$ Rule 2 from triple 1 and 5: $\text{isOnSidesOf}(\text{?house}, \text{?image_object}) \wedge \text{covers}(\text{?image_object}, \text{?asphalt}) \Rightarrow \text{isA}(\text{?image_object}, \text{?street})$ Rule 3 from triple 1 and 6: $\text{isOnSidesOf}(\text{?building}, \text{?image_object}) \wedge \text{covers}(\text{?image_object}, \text{?asphalt}) \Rightarrow \text{isA}(\text{?image_object}, \text{?street})$ Rule 4 from triple 2 and 5: $\text{isOnSidesOf}(\text{?house}, \text{?image_object}) \wedge \text{covers}(\text{?image_object}, \text{?bitumen}) \Rightarrow \text{isA}(\text{?image_object}, \text{?street})$ Rule 5 from triple 1 and 3 and 5: $\text{isOnSidesOf}(\text{?house}, \text{?image_object}) \wedge \text{covers}(\text{?image_object}, \text{?asphalt}) \wedge \text{contains}(\text{?image_object}, \text{?center_line}) \Rightarrow \text{isA}(\text{?image_object}, \text{?street})$...

FIGURE 3: Illustration on mapping the terms of RDF triple and low-level features.

FIGURE 4: Visual illustration on (a) experimental dataset 1 and (b) experimental dataset 2.

FIGURE 5: Visual illustration on (a) experimental dataset 1, and (b) experimental dataset 2.

The next step is to create the corresponding low-level features for each *subject*, *predicate*, and *object* in the 6th layer. Dissimilar to the low-level features extracted from remote sensing data, the parameters of low-level features here are given according to the semantic information obtained through fuzzy logic reasoning from descriptions on these terms, results of visual interpretation and experts' knowledge.

Uijlings et al [51] proposed a strategy that discovers the representative parameter for appropriate segmentation generation. In SLIC segmentation, the scale of segmentation studied in [51] refers to the number of segmentation and compactness. Based on the complexity of image content, we employed the strategy defined by Uijlings et al [51] to determine the parameter regarding the number of segmentation and compactness. Table 3 lists the priority of each low-level feature, which was semantically defined by transforming the descriptions on the object to be recognized to semantics in ontology. The descriptions on each category of the object for transforming semantics are mainly collected from visual interpretations on the optical satellite image. We also selected some text information from external resources including Wikipedia, Merriam Webster dictionary, and land classification systems involving National Oceanic and Atmospheric Administration (NOAA) and National Land Cover Database (NCLD).

Based on the information selected from multiple sources, we assign the priority of low-level features for each category. The color of roof and height refer to the key indicators for distinguishing buildings and other categories. Additionally, the differences in geometrical shape and size observed from the optical satellite image between residential and commercial buildings are used to distinguish these two types of building. From LiDAR height, the roof structures of many residential buildings are distinctly different from the ones of commercial buildings. For circular roads, its color and unique shape observed from the optical satellite image are mainly used. Not only the optical satellite image but also the text information about the materials of road that has specific color and unique geometric shape are given in some external resources, such as asphalt, bitumen, and circle. For swimming pools, water is mentioned in the external sources at high frequency. In the optical satellite image, the color of water is much more significant than geometric shape, since swimming pool carries a variety of shape. The color of trees is distinct compared with other categories in the optical satellite image. Moreover, the area

being covered by the boundaries derived from LiDAR height refers to another significant indicator for distinguishing trees and other categories that have obvious altitude variation (e.g., building). Due to the classification system that classifies trees over different height, the height derived from LiDAR height is also used for classifying these different types of trees.

Since the proposed approach is an unsupervised model without the support of data training, we show the recognition results using an unsupervised model available in some commercial software (e.g., ENVI and ERDAS): k nearest neighbor (KNN), as well as a commonly-used supervised model SVM in remote sensing data classification [63]. To get the recognition results by KNN, we used ERDAS software to apply the KNN for classifying five classes based on the optical satellite image. Then, we extracted the classified image regions being similar to trees and further classified these regions as three classes based on LiDAR height. For SVM, we applied a popular open library named LIBSVM to create a binary classifier for classifying the image object as either a target object or a nontarget object. Based on the strategy of binary classification, the optical satellite image was used as data source for recognizing swimming pools and circular roads, as well as classifying other five categories by incorporating with LiDAR height. Based on the segmentation results generated from the optical satellite image and LiDAR height, we randomly selected the image objects from the optical satellite image as training samples. The amount of the image objects used for training is less than 15% of the total number of all objects. Table 4 shows the quantitative comparison of recognition accuracy.

In another experimental dataset, based on our previous works in LiDAR data-based classification [49] and the content of the optical satellite image and LiDAR height, the objects to be recognized consists of buildings, parking structure, parking lot, shrub, grass, bare soil, and road. Figure 6 illustrates the samples of these objects to be recognized.

Similarly, we followed the workflow mentioned in Section 2 to create low-level features from the semantics of each category of the object. Table 5 lists the priority of each low-level feature, which was semantically defined by transforming the descriptions on the object to be recognized to semantics in ontology. The descriptions on each category of the object for transforming semantics are mainly collected from visual interpretations on the optical satellite image and the experiments from our efforts for LiDAR data-based land cover classification.

TABLE 3: Priority of low-level features defined from the first experimental datasets.

	Color/statistical features		GLCM features		Shape features	
	Optical satellite image	LiDAR height	Optical satellite image	LiDAR height	Optical satellite image	LiDAR height
Residential buildings	Primary	Primary	Secondary	Secondary		
Commercial buildings		Primary	Secondary	Secondary		
Circular roads	Primary		Secondary		Primary	
Swimming pools	Primary					
Trees (1.5-5m)	Secondary	Primary	Tertiary		Tertiary	Secondary
Trees (5-10 m)	Secondary	Primary	Tertiary		Tertiary	Secondary
Trees (>10 m)	Secondary	Primary	Tertiary		Tertiary	Secondary

TABLE 4: Quantitative evaluation of recognition results based on the first experimental datasets.

	KNN		SVM		Our approach	
	Precision	Recall	Precision	Recall	Precision	Recall
Residential buildings	0.5852	0.6132	0.6877	0.6523	0.9688	0.9798
Commercial buildings	0.6721	0.5547	0.7621	0.6788	0.9167	0.9136
Circular roads	0.3894	0.6998	0.5586	0.5919	0.9705	0.9814
Swimming pool	0.702	0.3485	0.8775	0.9262	0.9333	1
Trees (1.5-5m)	0.5264	0.4416	0.7295	0.8562	0.8824	0.9525
Trees (5-10 m)	0.4799	0.5230	0.7462	0.8025	0.8493	0.9729
Trees (>10 m)	0.5012	0.4611	0.7119	0.837	0.8677	0.9702

FIGURE 6: Illustration on the object to be recognized from the second experimental datasets. (a) Buildings. (b) Parking lot. (c) Shrub. (d) Grass. (e) Bare soil. (f) Road.

Based on the information selected from multiple sources, we assign the priority of low-level features for each category. The height refers the key indicators for distinguishing buildings and other categories. Parking lot and road carry similar color and LiDAR height. However, vehicles and the

texture feature of road plays the key impact to distinguish these two types of object. Moreover, LiDAR height refers to the key parameter to distinguish between the parking structure and parking lot and road. Similar to the first experimental dataset, the shape of the road is significantly

TABLE 5: Priority of low-level features defined from the second experimental datasets.

Color/statistical features		GLCM features		Shape features	
Optical satellite image	LiDAR height	Optical satellite image	LiDAR height	Optical satellite image	LiDAR height
Buildings	Primary		Secondary		Primary
Parking lot	Primary	Secondary			
Shrub	Primary	Secondary	Secondary	Tertiary	
Grass	Primary				
Bare soil	Primary	Secondary			
Road	Primary	Secondary		Tertiary	

TABLE 6: Quantitative evaluation of recognition results based on the first experimental datasets.

	KNN		SVM		Our approach	
	Precision	Recall	Precision	Recall	Precision	Recall
Buildings	0.7556	1*	0.8395	1*	0.9527	1*
Parking lot	0.5208	0.5855	0.7588	0.7295	0.8405	0.82077
Shrub	0.573	0.7801	0.8041	0.7705	0.9133	0.8582
Grass	0.5264	0.7266	0.8315	0.8853	0.9515	0.9474
Bare soil	0.6975	0.8316	0.791	0.7655	0.9097	0.8711
Road	0.701	0.7961	0.7229	0.7383	0.8836	0.8642

*buildings are extracted separately from other types of object.

different from the parking structure and parking lot. Although bare soil and shrub have similar land cover from the optical satellite image and LiDAR intensity, shrub carries specific pattern which can be represented from texture features. For grass, its color is distinct compared with other categories in the optical satellite image.

According to Tables 4 and 6, it can be observed that the recognition result by our proposed approach overperforms an unsupervised KNN approach and supervised SVM approach. Without data training, KNN is not able to effectively support object recognition based on the optical satellite image and LiDAR height. First, without manual setting, the distance metric defined by KNN classifier is difficult to be self-adjusted for the specific application. Second, KNN such as an unsupervised lazy classifier might be challenging to effectively handle complicated variables. The performance of KNN in recognition proves that the features extracted from data alone without people’s interpretation (e.g., distance metric and distance assigned with direction) cannot characterize the object from the optical satellite image and LiDAR height. Although SVM performs well on applying the low-level features for distinguishing a variety of objects through the training process, it still has challenges to conduct perfect recognition without data features. Our proposed approach first defines the representations of to-be-recognized objects based on high-level semantics. Then, it creates low-level features that corresponding to semantics, which facilitate discovering the low-level features more fitting for the representations of to-be-recognized objects. Lacking the training process based on data samples, our proposed approach can also reach the recognition results within better accuracy.

Based on the results obtained from above two experiments, several conclusions can be derived from the comparison of these three methods. First, the existing techniques

that focus on data features alone cannot support efficient object recognition without data training and knowledge assistance. Additionally, even though the low-level features can be extracted through the training process in commonly used models and algorithms, the gap between low-level features and high-level semantics can still significantly impact the recognition accuracy. Last but not least, high-level semantics has been proved as the essential factor to improve the performance of machine learning models in object recognition. Thus, the investigations on high-level semantics have a great potential for facilitating object recognition from remote sensing data.

Meanwhile, the main disadvantage of the proposed method is also observed from the experiment. As mentioned in some papers [27, 44], semantics-driven object recognition might not have a good transferability, which can only be limited in the predefined object type. Moreover, more in-depth investigations are still needed for the mapping between semantics and data features.

4. Conclusion

This paper presents an ontology-embedded probabilistic relational model to support object recognition based on geographical knowledge being derived from ontological rules. The first step aims to semantically organize the descriptions of the object to be recognized as rules in ontology. Then, an SWRL-based encoder is used to decompose the rules into three elements of triple (subject, property, and object) and map these elements to low-level features created from an optical satellite image and LiDAR height. Next, a PBN is developed for representing the low-level features corresponding to rules in ontology. Eventually, a modified TanH function is created to compute the distribution of

features-based probability and to determine the object recognized based on the graph of PBN.

A majority of sophisticatedly probabilistic models for object recognition mainly focus on extracting features remote sensing data, rather than rich human semantics. The heterogeneity between low-level data features and high-level human semantics prevents these sophisticated algorithms from producing accurate result on object recognition. Although some previous research has attempted to utilize knowledge to support remote sensing data analysis, few of them have discussed on extracting the low-level features that are applicable for applying the semantics of the object to be recognized. Moreover, integrating knowledge to establish an unsupervised approach for object recognition without data training is still a nonproven research field. In the future, several extended endeavors are worthy studying. We hope our work can further help in the development of object recognition as the field continues moving forward towards multilayer learning algorithms, knowledge graph-based relational reinforcement learning, etc.

Data Availability

The multispectral remote sensing image and LiDAR height data used to support the first experimental findings of this study have been deposited in the Quickbird and the CAP LTER project [61], respectively. The high-resolution remote sensing image and LiDAR height data used to support the second experimental findings of this study have been obtained from Google Earth and LiDAR height accessed from a cyber infrastructure named OpenTopography (<https://opentopography.org/>), respectively. Zhao, Q., Myint, S. W., Wentz, E. A., and Fan, C. [61]. Rooftop surface temperature analysis in an urban residential environment. *Remote Sensing*, 7(9), 12135–12159.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported by “the Fundamental Research Funds for the Central Universities (Grant no. 2020QN28).”

References

- [1] T. Huang, Z. Liu, Y. Pan, and Y. Zhang, “Land-cover survey in northeast China using remote sensing and GIS,” *Chinese Geographical Science*, vol. 8, no. 3, pp. 264–270, 1998.
- [2] W. Kuang, “Spatio-temporal patterns of intra-urban land use change in Beijing, China between 1984 and 2008,” *Chinese Geographical Science*, vol. 22, no. 2, pp. 210–220, 2012.
- [3] B. Xue, L. Zhang, Y. Geng, B. Mitchell, and W. Ren, “Extended land-use coding system and its application in urban brownfield redevelopment: case study of Tiexi district in Shenyang, China,” *Journal of Urban Planning and Development*, vol. 142, no. 3, Article ID 05015014, 2016.
- [4] A. Guo, J. Yang, X. Xiao, J. Xia, C. Jin, and X. Li, “Influences of urban spatial form on urban heat island effects at the community level in China,” *Sustainable Cities and Society*, vol. 53, Article ID 101972, 2020.
- [5] L. I. Yi, S. R. Zhang, L. I. Ting, L. I. U. Yu, and L. Zhang, “Exploration on method of auto-classification for the main types of ground objects in alpine and gorge region based on ASTER,” *Remote Sensing Technology and Application*, vol. 22, no. 3, pp. 389–395, 2011.
- [6] G. E. Hinton, “Learning multiple layers of representation,” *Trends in Cognitive Sciences*, vol. 11, no. 10, pp. 428–434, 2007.
- [7] K. An, J. Zhang, and Y. Xiao, “Object-oriented urban dynamic monitoring—a case study of haidian district of beijing,” *Chinese Geographical Science*, vol. 17, no. 3, pp. 236–242, 2007.
- [8] Y. Liu, M. Li, L. Mao, F. Xu, and S. Huang, “Review of remotely sensed imagery classification patterns based on object-oriented image analysis,” *Chinese Geographical Science*, vol. 16, no. 3, pp. 282–288, 2006.
- [9] M. Li, X. Zhu, W. Guo, P. Yue, and Y. Fan, “tdcA case-based reasoning approach for task-driven remote sensing image discovery under spatial-temporal constraints,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 1, pp. 454–466, 2016.
- [10] G. Cheng and J. Han, “A survey on object detection in optical remote sensing images,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 117, pp. 11–28, 2016.
- [11] X. Gu, P. P. Angelov, C. Zhang, and P. M. Atkinson, “A massively parallel deep rule-based ensemble classifier for remote sensing scenes,” *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 3, pp. 345–349, 2018.
- [12] M. A. Islam, D. T. Anderson, A. Pinar, T. C. Havens, G. Scott, and J. M. Keller, “Enabling explainable fusion in deep learning with fuzzy integral neural networks,” *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 7, pp. 1291–1300, 2019.
- [13] J. Zhang, W. Deng, and S. Liu, “Geographical space development zone classification: an essential guide for transformation of mountain resource cities,” *Chinese Geographical Science*, vol. 25, no. 3, pp. 361–374, 2015.
- [14] J. Yang, S. Jin, X. Xiao et al., “Local climate zone ventilation and urban land surface temperatures: towards a performance-based and wind-sensitive planning proposal in megacities,” *Sustainable Cities and Society*, vol. 47, Article ID 101487, 2019.
- [15] L. Jiang, B. Xue, Z. Ma, L. Yu, B. Huang, and X. Chen, “A life-cycle based co-benefits analysis of biomass pellet production in China,” *Renewable Energy*, vol. 154, 2020.
- [16] J. Hu, A. Razdan, J. C. Femiani, M. Cui, and P. Wonka, “Road network extraction and intersection detection from aerial images by tracking road footprints,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 4144–4157, 2007.
- [17] R. Neher and A. Srivastava, “A Bayesian MRF framework for labeling terrain using hyperspectral imaging,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 6, pp. 1363–1374, 2005.
- [18] M. Wieland and M. Pittore, “Performance evaluation of machine learning algorithms for urban pattern recognition from multi-spectral satellite images,” *Remote Sensing*, vol. 6, no. 4, pp. 2912–2939, 2014.
- [19] C. Benedek and T. Szirányi, “Change detection in optical aerial images by a multilayer conditional mixed Markov model,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 10, pp. 3416–3430, 2009.
- [20] N. Durand, S. Derivaux, G. Forestier et al., “Ontology-based object recognition for remote sensing image interpretation,” in *Proceedings of the 19th IEEE International Conference on Tools*

- with Artificial Intelligence (ICTAI 2007), vol. 1, pp. 472–479, IEEE, Patras, Greece, 2007 October.
- [21] H. Ma, J. Zhu, M. R. T. Lyu, and I. King, “Bridging the semantic gap between image contents and tags,” *IEEE Transactions on Multimedia*, vol. 12, no. 5, pp. 462–473, 2010.
 - [22] H. Tamura and N. Yokoya, “Image database systems: a survey,” *Pattern Recognition*, vol. 17, no. 1, pp. 29–43, 1984.
 - [23] V. E. Ogle and M. Stonebraker, “Chabot: retrieval from a relational database of images,” *Computer*, vol. 28, no. 9, pp. 40–48, 1995.
 - [24] T. Berners-Lee, J. Hendler, and O. Lassila, “The semantic web,” *Scientific American*, vol. 284, no. 5, pp. 28–37, 2001.
 - [25] R. Bahmanyar, A. Murillo Montes de Oca, and M. Datcu, “The semantic gap: an exploration of user and computer perspectives in earth observation images,” *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 10, pp. 2046–2050, 2015.
 - [26] S. S. Durbha and R. L. King, “Knowledge mining in earth observation data archives: a domain ontology perspective,” in *Proceedings of the Geoscience and Remote Sensing Symposium 2004*, vol. 1, IEEE, Anchorage, AK, USA, 2004 September.
 - [27] D. Arvor, L. Durieux, S. Andrés, and M. A. Laporte, “Advances in geographic object-based image analysis with ontologies: a review of main contributions and limitations from a remote sensing perspective,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 82, pp. 125–137, 2013.
 - [28] M. Belgiu, I. Tomljenovic, T. Lampoltshammer, T. Blaschke, and B. Höfle, “Ontology-based classification of building types detected from airborne laser scanning data,” *Remote Sensing*, vol. 6, no. 2, pp. 1347–1366, 2014.
 - [29] T. Lampoltshammer and S. Wiegand, “Improving the computational performance of ontology-based classification using graph databases,” *Remote Sensing*, vol. 7, no. 7, pp. 9473–9491, 2015.
 - [30] M. Datcu and K. Seidel, “Human-centered concepts for exploration and understanding of earth observation images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 601–609, 2005.
 - [31] G. Forestier, A. Puissant, C. Wemmert, and P. Gançarski, “Knowledge-based region labeling for remote sensing image interpretation,” *Computers, Environment and Urban Systems*, vol. 36, no. 5, pp. 470–480, 2012.
 - [32] G. Forestier, C. Wemmert, and A. Puissant, “Coastal image interpretation using background knowledge and semantics,” *Computers & Geosciences*, vol. 54, pp. 88–96, 2013.
 - [33] R. Oliva-Santos, F. Maciá-Pérez, and E. Garea-Llano, “Ontology-based topological representation of remote-sensing images,” *International Journal of Remote Sensing*, vol. 35, no. 1, pp. 16–28, 2014.
 - [34] D. Kohli, R. Sliuzas, N. Kerle, and A. Stein, “An ontology of slums for image-based classification,” *Computers, Environment and Urban Systems*, vol. 36, no. 2, pp. 154–163, 2012.
 - [35] S. Réjichi, F. Chaabane, and F. Tupin, “Expert knowledge-based method for satellite image time series analysis and interpretation,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 5, pp. 2138–2150, 2015.
 - [36] L. Liu, Y. Xue, J. Guang, and J. Liu, “Description of an ontology-based remote sensing model service with an integrated framework environment for remote sensing applications,” *Remote Sensing Letters*, vol. 6, no. 10, pp. 804–813, 2015.
 - [37] A. J. Pérez-Luque, R. Pérez-Pérez, F. J. Bonet-García, and P. J. Magaña, “An ontological system based on MODIS images to assess ecosystem functioning of natura 2000 habitats: a case study for *Quercus pyrenaica* forests,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 37, pp. 142–151, 2015.
 - [38] S. Nieland, B. Kleinschmit, and M. Förster, “Using ontological inference and hierarchical matchmaking to overcome semantic heterogeneity in remote sensing-based biodiversity monitoring,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 37, pp. 133–141, 2015.
 - [39] H. Luo, L. Li, H. Zhu, X. Kuai, Z. Zhang, and Y. Liu, “Land cover extraction from high resolution ZY-3 satellite imagery using ontology-based method,” *ISPRS International Journal of Geo-Information*, vol. 5, no. 3, p. 31, 2016.
 - [40] S. Andres, D. Arvor, and C. Pierkot, “Towards an ontological approach for classifying remote sensing images,” in *Proceedings of the Signal Image Technology and Internet Based Systems (SITIS), 2012*, IEEE, Naples, Italy, pp. 825–832, 2012 November.
 - [41] S. Hashimoto, T. Tadono, M. Onosato, M. Hori, and T. Moriyama, “A framework of ontology-based knowledge information processing for change detection in remote sensing data,” in *Proceedings of the Geoscience and Remote Sensing Symposium (IGARSS), 2011*, IEEE, Vancouver, Canada, pp. 3927–3930, 2011 July.
 - [42] J. M. Almendros-Jiménez, L. Domene, and J. A. Piedra-Fernández, “A framework for ocean satellite image classification based on ontologies,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 2, pp. 1048–1063, 2013.
 - [43] É. Ranisavljević, F. Devin, D. Laffly, and Y. Le Nir, “Semantic orchestration of image processing services for environmental analysis,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 83, pp. 184–192, 2013.
 - [44] D. Arvor, M. Belgiu, Z. Falomir, I. Mougenot, and L. Durieux, “Ontologies to interpret remote sensing images: why do we need them?” *GIScience & Remote Sensing*, vol. 56, no. 6, pp. 911–939, 2019.
 - [45] M. Shoshany and Y. Cohen, “Inference versus evidence in reasoning remote sensing recognition, with an information foraging perspective,” *International Journal of Remote Sensing*, vol. 28, no. 11, pp. 2613–2634, 2007.
 - [46] A. Baraldi, V. Puzzolo, P. Blonda, L. Bruzzone, and C. Tarantino, “Automatic spectral rule-based preliminary mapping of calibrated Landsat TM and ETM+ images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 9, pp. 2563–2586, 2006.
 - [47] A. Baraldi, L. Durieux, D. Simonetti, G. Conchedda, F. Holec, and P. Blonda, “Automatic spectral-rule-based preliminary classification of radiometrically calibrated SPOT-4/-5/IRS, AVHRR/MSG, AATSR, IKONOS/QuickBird/OrbView/GeoEye, and DMC/SPOT-1/-2 imagery—Part I: system design and implementation,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 3, pp. 1299–1325, 2009.
 - [48] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, “SLIC superpixels compared to state-of-the-art superpixel methods,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.
 - [49] X. Zhou and W. Li, “A geographic object-based approach for land classification using LiDAR elevation and intensity,” *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 669–673, 2017.
 - [50] W. Zhou, “An object-based approach for urban land cover classification: integrating lidar height and intensity data,” *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 4, pp. 928–931, 2013.

- [51] J. R. R. Uijlings, K. E. A. Van De Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *International Journal of Computer Vision*, vol. 104, no. 2, pp. 154–171, 2013.
- [52] J. H. Gennari, M. A. Musen, R. W. Ferguson et al., "The evolution of Protégé: an environment for knowledge-based systems development," *International Journal of Human-Computer Studies*, vol. 58, no. 1, pp. 89–123, 2003.
- [53] V. Sebastian, A. Unnikrishnan, and K. Balakrishnan, "Gray level co-occurrence matrices: generalisation and some new features," 2012, <http://arxiv.org/abs/1205.4831>.
- [54] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [55] M. Saenz, G. T. Buracas, and G. M. Boynton, "Global effects of feature-based attention in human visual cortex," *Nature Neuroscience*, vol. 5, no. 7, pp. 631–632, 2002.
- [56] T. Moore and K. M. Armstrong, "Selective gating of visual signals by microstimulation of frontal cortex," *Nature*, vol. 421, no. 6921, pp. 370–373, 2003.
- [57] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *Aistats*, vol. 9, pp. 249–256, 2010.
- [58] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning*, pp. 807–814, Rio de Janeiro, Brazil, July 2010.
- [59] Y. Bengio, "Learning deep architectures for AI," *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.
- [60] B. Karlik and A. V. Olgac, "Performance analysis of various activation functions in generalized MLP architectures of neural networks," *International Journal of Artificial Intelligence and Expert Systems*, vol. 1, no. 4, pp. 111–122, 2011.
- [61] Q. Zhao, S. Myint, E. Wentz, and C. Fan, "Roof top surface temperature analysis in an urban residential environment," *Remote Sensing*, vol. 7, no. 9, pp. 12135–12159, 2015.
- [62] S. Krishnan, C. Crosby, V. Nandigam et al., "Open-Topography: a services oriented architecture for community access to LIDAR topography," in *Proceedings of the 2nd International Conference on Computing for Geospatial Research & Applications*, p. 7, ACM, Washington DC, USA, May 2011.
- [63] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: a review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 3, pp. 247–259, 2011.