Knowledge-Based Systems 243 (2022) 108469

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Combining deep learning and ontology reasoning for remote sensing image semantic segmentation

Yansheng Li [a,*], Song Ouyang [a,b,**], Yongjun Zhang [a]

[a] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
[b] Institute of Remote Sensing and GIS, Peking University, Beijing, China

ARTICLE INFO

Article history:
Received 4 May 2020
Received in revised form 13 February 2022
Accepted 16 February 2022
Available online 22 February 2022

Keywords:
Collaboratively boosting framework (CBF)
Deep learning
Ontology reasoning
Deep semantic segmentation network (DSSN)
Remote sensing (RS) imagery

ABSTRACT

Because of its wide potential applications, remote sensing (RS) image semantic segmentation has attracted increasing research interest in recent years. Until now, deep semantic segmentation network (DSSN) has achieved a certain degree of success on semantic segmentation of RS imagery and can obviously outperform the traditional methods based on hand-crafted features. As a classic data-driven technique, DSSN can be trained by an end-to-end mechanism and is competent for employing low-level and mid-level cues (i.e., the discriminative image structure) to understand RS images. However, its interpretability and reliability are poor due to the nature weakness of the data-driven deep learning methods. By contrast, human beings have an excellent inference capacity and can reliably interpret RS imagery with the basic RS domain knowledge. Ontological reasoning is an ideal way to imitate and employ the domain knowledge of human beings. However, it is still rarely explored and adopted in the RS domain. As a solution of the aforementioned critical limitation of DSSN, this study proposes a collaboratively boosting framework (CBF) to combine the data-driven deep learning module and knowledge-guided ontology reasoning module in an iterative manner. The deep learning module adopts the DSSN architecture and takes the integration of the original image and inferred channels as the input of the DSSN. In addition, the ontology reasoning module is composed of intra- and extra-taxonomy reasoning. More specifically, the intra-taxonomy reasoning directly corrects misclassifications of the deep learning module based on the domain knowledge, which is the key to improve the classification performance. The extra-taxonomy reasoning aims to generate the inferred channels beyond the current taxonomy to improve the discriminative performance of DSSN in the original RS image space. On the one hand, benefiting from the referred channels from the ontology reasoning module, the deep learning module using the integration of the original image and referred channels can achieve better classification performance than only using the original image. On the other hand, better classification results from the deep learning module further improve the performance of the ontology reasoning module. As a whole, the deep learning and ontology reasoning modules are mutually boosted in the iterations. Extensive experiments on two publicly open RS datasets such as UCM and ISPRS Potsdam show that our proposed CBF can outperform the competitive baselines with a large margin.

1. Introduction

As a fundamental task of remote sensing (RS) image interpretation, RS image semantic segmentation [1,2], which aims to annotate each pixel of the RS imagery with one type of land-use/land-cover (LULC) type, plays an important role in wide applications such as land-cover mapping, natural resource protection, intelligent agriculture, and ecological assessment [3]. Generally, RS image semantic segmentation is similar to natural image semantic segmentation [4]. Compared with natural images, RS images often present more complex image structures [5], which lead to additional challenges in RS image semantic segmentation [6].

Based on hand-crafted features, shallow classifiers such as support vector machine (SVM) [7,8], maximum likelihood estimate (MLE) [9], and decision tree (DT) [10] have been widely applied to RS image semantic segmentation [11]. However, the performance of these handcrafted feature-based semantic segmentation methods is still very limited. Along with the rapid

* Corresponding author.
** Corresponding author at: School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China.
E-mail addresses: yansheng.li@whu.edu.cn (Y. Li), song.ouyang@whu.edu.cn (S. Ouyang).

https://doi.org/10.1016/j.knosys.2022.108469
0950-7051/

development of deep networks [12] such as deep detection network [13,14], deep recognition network [15], deep hashing network [16,17], and deep semantic segmentation network (DSSN) [18], deep networks have led to remarkable improvements for RS image semantic segmentation [19] because of their obvious superiority in learning the network parameters via an end-to-end manner. However, due to the black-box characteristics of deep learning, the interpretability and reliability of DSSN are still extremely weak [20]. Hence, how to further improve the semantic segmentation performance by reinforcing DSSN deserves much more attention.

DSSN is competent for employing low-level and mid-level cues to interpret RS images but lacks the high-level inference ability [21]. By contrast, human beings, who have an excellent inference capacity, can reliably interpret the RS imagery. The reason why RS experts can interpret RS images quickly and accurately is that they have the necessary prior domain knowledge to make the interpretation decisions through knowledge reasoning. Although data-driven learning attracts wide research interest, knowledge-driven reasoning is still regarded as one of the most important research directions by the RS community [22]. As well known, ontology [23] has a strong capability in knowledge representation, commonsense inference, semantic cognition, and knowledge sharing [24]. However, as not all of the domain experts' prior knowledge can be fully modeled, the ontology reasoning performance is still very limited in the RS image interpretation task.

Generally speaking, the data-driven methods have advantages in terms of accuracy and accessibility while their reliability is poor. On the contrary, the knowledge-driven methods have strong interpretability but insufficient performance. Thus, the advantages and disadvantages of data-driven methods and knowledge-driven methods are complementary to a considerable extent. The combination of ontology reasoning and deep learning can make full use of the advantages of knowledge-driven and data-driven methods. Therefore, coupling data-driven deep learning and knowledge-guided ontology reasoning is a promising way to achieve truly intelligent interpretation of RS imagery [25,26]. On the one hand, ontology reasoning helps to directly correct misclassifications and improve the interpretability of the classifications. On the other hand, additional information such as DEM, shadow data, and infrared band data can effectively enhance the anti-interference ability of deep networks against visual similarity [21]. However, additional information is usually difficult to obtain. Generating estimated data of the additional information through ontology reasoning will considerably reduce the pressure on data acquisition. Therefore, ontology reasoning effectively solves the interpretability problem of data-driven methods.

Based on the above analysis, we propose a collaborative boosting framework (CBF) to combine data-driven deep learning and knowledge-guided ontology reasoning for RS image semantic segmentation, which realizes the interaction between the data-driven and knowledge-guided methods. It not only adopts the DSSN to learn the low-level and mid-level cues from RS images, but also applies ontology reasoning to make the classification result interpretable and credible with high-level knowledge. Intra- and extra-taxonomy ontology reasoning are designed for the reasoner in the CBF. The former directly corrects the misclassification of the DSSN. The latter, performed on the corrected classification result from the intra-taxonomy reasoning, aims to provide a better estimation of shadow and elevation as additional channels to enhance the anti-interference capability of the DSSN, which makes the classification result more reliable. In the CBF, the DSSN autonomously learns low-level and mid-level features from RS images. As a whole, ontology reasoner uses high-level domain knowledge to guide interpretation including directly correcting misclassifications of the DSSN and indirectly extracting additional information to assist the DSSN. The whole process forms a closed loop and performs continuous iterations until the CBF converges. The proposed CBF has been evaluated on two publicly open RS datasets such as UCM and ISPRS Potsdam. The experimental results show that it not only improves the interpretability and reliability of the classification result but also further promotes the classification accuracy. As a whole, the main contributions of this paper are summarized as follows:

- A unified framework called CBF is proposed to mutually reinforce data-driven deep learning and knowledge-guided ontology reasoning in an iterative way, which organically realizes coupling of deep learning and knowledge reasoning. Notably, our CBF is a general framework, and more variants can be designed based on the specific task requirements.
- This study presents a new unified ontology reasoning approach which includes intra- and extra-taxonomy ontology reasoning. The former directly corrects misclassifications of the deep learning module, which is the key to improve the classification performance. The latter improves the deep learning module with additional estimated information from the input perspective.

The rest of this paper is organized as follows. Section 2 describes the related work. The proposed methods are detailed in Section 3. Section 4 analyzes and discusses the experimental results. Finally, Section 5 summarizes the work of this paper and points out some potential research directions.

2. Related work

In recent years, deep learning has been widely employed in semantic segmentation of high-resolution RS imagery [27–30]. Basaeed et al. used a convolutional neural network (CNN) to perform multi-scale analysis on each channel [31], which included fusion and morphological operations on the boundary confidence map to obtain a hierarchical segmentation map. Langkvist et al. applied DSSN to achieve fast and accurate pixel-by-pixel classification on multispectral images [32]. Audebert et al. trained variants of the SegNet structure and introduced multi-core convolutional layers to quickly aggregate predictions on multiple scales [33]. Maggiori et al. designed a CNN that combined features at different resolutions to integrate local and global information in an efficient manner [34]. Kampffmeyer et al. mapped the urban land-cover on RS imagery by a novel deep CNN [35], which detected small objects effectively while achieving high overall accuracy. Generally, these existing deep learning-based methods follow a data-driven learning mechanism [36,37]; however, they still cannot make full use of the high-level knowledge of domain experts, which leads to poor interpretability and reliability of the classification.

On another research avenue, ontology-based knowledge models have great advantages in expressing and applying knowledge. By solving the major limitations of deep learning methods in knowledge cognition, ontology-based knowledge models have the potential to promote the long-term development of RS [38]. Sarker et al. proposed a system based on a knowledge model for explaining the classification [24]. Codescu et al. applied the OSM ontology to geographic information system (GIS) [39]; however, the application of ontology was limited. Gui et al. extracted buildings from TerraSAR-X imagery by the RS ontology [40]. Andres et al. used the ontology-based spectral rules to classify the Landsat images [41], without considering other information such as the shape, texture, and spatial relationship of objects. Khitem et al. constructed regional adjacency maps and completed the classification according to the spatial and spectral attributes of

ontological regions [42]. Geographic object-based image analysis (GEOBIA) adopts the object information and expert knowledge to interpret images. Gu et al. proposed an ontology-based semantic segmentation method for high-resolution RS imagery [43], which aims to make full use of the advantages of GEOBIA and ontology. Ontology reasoning in the above methods enhances the interpretability and reliability of classification; however, its performance is very limited compared with that of deep learning-based methods because modeling all the domain knowledge is still an open problem.

Apparently, combining deep learning and ontology reasoning is a promising way to coordinate between data-driven and knowledge-driven methods [44]. As the first attempt toward this direction, Alirezaie et al. proposed a method called Semantic Referee which combined the ontology reasoner and a DSSN-based classifier [21], where the ontology reasoner was responsible for generating an estimation of shadow, elevation, and inconsistent spatial relationship and returned the estimation as additional channels to DSSN to assist the classification. The research showed that, as an additional input to the DSSN, extra information can effectively improve classification accuracy. However, Semantic Referee has two limitations: (1) the ontology reasoner only assists DSSN from the input perspective but does not directly correct misclassifications of the classifier, which limits the performance and interpretability of the classification. (2) Extra information obtained by directly reasoning on the results which contain misclassifications is not sufficiently accurate for DSSN. Therefore, how to leverage knowledge reasoning to improve the interpretability and reliability of the interpretation requires further exploration.

3. Methodology

In view of the aforementioned limitations, this paper proposes a unified framework (i.e., CBF) to comprehensively couple data-driven deep learning and knowledge-guided ontology reasoning, where the ontology reasoning module connects intra- and extra-taxonomy reasoning in series. The intra-taxonomy reasoning module is designed to directly correct misclassifications of the DSSN, which improves the interpretability of classification. In addition, the extra-taxonomy reasoning module is performed on the corrected classification result to provide better estimation information for the DSSN, which indirectly makes the interpretation more reliable. Different from Semantic Referee, our CBF focuses on predicting the information of shadow and elevation as the estimation from the extra-taxonomy reasoning is sufficiently accurate for the DSSN, which effectively reduces the computation.

As visually shown in Fig. 1, our proposed CBF includes two main modules, i.e., the DSSN-based classification module and ontology reasoning module. These two modules interact in an iterative manner to optimize the output of the entire system, where the DSSN-based classification module completes training of DSSN and the initial semantic segmentation of images, and the ontology reasoning module comprises intra- and extra-taxonomy reasoning for estimating additional information. More specifically, our CBF runs via an iterative manner. Remote sensing ontology (RSOntology) is established for the interpretation of RS images. According to RS domain knowledge, we design ontology rules for intra- and extra-taxonomy ontology reasoning. In each iteration of the training phrase, DSSN is firstly trained and outputs the initial semantic segmentation. Then, the intra-taxonomy ontology reasoner directly corrects misclassifications from DSSN according to the ontology rules, and the extra-taxonomy ontology reasoner extracts the estimation of shadow and elevation. The estimation is more accurate because the extra-taxonomy ontology reasoning operates on the corrected classification. Finally, estimation information is employed as an additional channel for training DSSN in the next iteration, which indirectly enhances the classifier's anti-interference ability to the visual similarity. Notably, only the trained DSSN and the intra-taxonomy ontology reasoning are adopted to perform classification in the testing phase. Given an iteration, the output of DSSN is denoted as the result of Stage **I**, and the refined output from the intra-taxonomy ontology reasoning is denoted as the result of Stage **II**.

3.1. Learning deep semantic segmentation network (DSSN)

In the DSSN-based classification module, DSSN is employed as a classifier to perform the initial classification. DSSN usually takes images with RGB channels as an input. Different from the traditional practice, two estimated input channels are added to enhance the DSSN's anti-interference ability. Thus, we train DSSN by RGB channel images with the inferenced additional channels. The specific implementation details are introduced as follows.

Assume that original RS image is $\boldsymbol{I}$ and its corresponding additional channels are $\boldsymbol{E}$ ($\boldsymbol{E}$, which represents the estimated information of shadow and elevation, is set to $\mathbf{0}$ in the first iteration step and can be adaptively generated in the subsequent iteration steps). $\boldsymbol{\theta}$ denotes the parameters of DSSN. After forward propagation, DSSN outputs the classification confidence map $\boldsymbol{F}$ and the category map $\boldsymbol{C}$.

$$\boldsymbol{F} = \varphi\left((\boldsymbol{I}, \boldsymbol{E}), \boldsymbol{\theta}\right) \tag{1}$$

$$\boldsymbol{C} = \mathrm{ArgMax}(\boldsymbol{F}) \tag{2}$$

where φ represents the hierarchical mapping function of DSSN.

Based on the optimization function J, $\boldsymbol{\theta}$ is updated by the backward propagation algorithm. Generally, the cross-entropy loss function in Eq. (3) is adopted as J.

$$J = -\sum_{i}\sum_{j}\left(\sum_{c=1}^{n} y_{i,j}^{c}\log\left(p_{i,j}^{c}\right)\right) \tag{3}$$

where $c = \boldsymbol{C}_{i,j}$ and $p_{i,j}^{c} = \boldsymbol{F}_{i,j}^{c}$. n is the number of classes, and $p_{i,j}^{c}$ denotes the classification confidence of the pixel with the cth class located at the image coordinate (i, j). If the forward prediction is the same as the label, $y_{i,j}^{c}$ is 1, otherwise it is 0.

3.2. Generating inference units

After the RS image semantic segmentation using the DSSN, we need to refine the initial classification result by ontology reasoning. The ground object exists in the form of an individual; thus, reasoning is based on objects. The appropriate size is crucial to the selection of the object units. The most ideal situation is that the segmentation block is segmented along the boundary of the ground object. However, it is difficult to achieve such accurate segmentation using the existing methods [45]. Taking a step back, the superpixel scale is between the pixel scale and real object scale. Each superpixel is composed of spatially neighboring pixels with similar spectra. Therefore, the interior of one superpixel is homogeneous and retains part of the boundary of the ground object. Therefore, the superpixel can represent a part of the ground object. According to the above reasoning, we performed superpixel segmentation on the original image. To restore the ground object as much as possible, we cluster superpixels in the neighborhood according to the initial classification. The specific process is as follows.

Superpixel segmentation is performed on the original image $\boldsymbol{I}$ to obtain the superpixel set $\boldsymbol{G}$ with k superpixels, as shown in Eq. (4).

$$\boldsymbol{G} = \{S_1, S_2, \ldots, S_k | S_i = Segment\left(\boldsymbol{I}\right), 1 \le i \le k\} \tag{4}$$

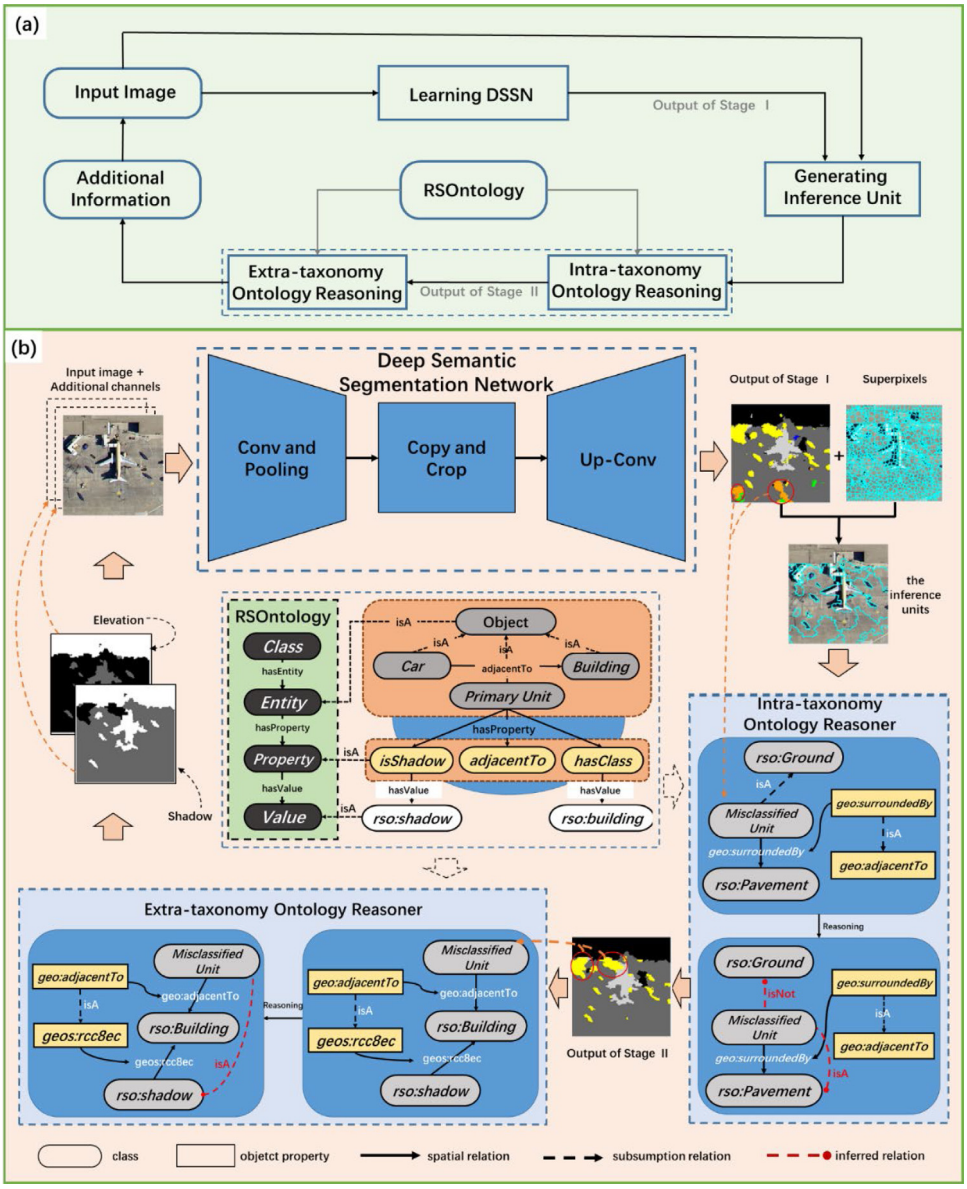

Fig. 1. Workflow of the proposed CBF. (a) General diagram of the workflow, which is shown in detail in (b). The workflow includes two main modules: DSSN-based classification module and ontology reasoning module. RSOntology was constructed for the interpretation of RS imagery. In each iteration, the deep semantic segmentation network is first leaned, inference units are generated, and then ontology reasoning including the intra- and extra-taxonomy ontology reasoning are performed finally.

where S_i denotes the ith superpixel.

Next, the number of pixels of each category in each superpixel area are calculated according to the initial classification, and the category with the largest proportion is used as the characterization category of the superpixel. Furthermore, in the spatial neighborhood, superpixels with the same characterization category are merged into a new superpixel in Fig. 2. After clustering-based aggregation, the new superpixels are taken as the inference units S', as shown in Eq. (5).

$$S' = \{S_i | C_i = C, S_i \text{ Adjacent to } S, 1 \le i \le k\} \tag{5}$$

From $\boldsymbol{F}$, the network outputs values representing the classification confidence for each pixel. The larger the value, the higher the classification reliability. Similarly, we need a value that characterizes the classification confidence of the inference unit to distinguish areas in which the classification is unreliable. Therefore, we denote the average confidence of the pixels in each inference unit as the classification confidence of the inference unit. The classification confidence can be used as a basis for judging whether the classification is correct. An inference unit with low confidence (i.e., $F < F_t$) is taken as the misclassification unit where F_t is an empirical threshold.

3.3. Constructing ontology of remote sensing (RSOntology)

Ontology, as a formal expression of concepts and semantic relationships in the domain of interest, has significant advantages in knowledge representation, semantic reasoning, and knowledge sharing [46]. Ontology is constituted by a set of object types, attributes, and relationships, which describe the research domain with semantics. Ontology can be implemented with standard language such as the Web Ontology Language (OWL) which is a World Wide Web Consortium (W3C) standard related to description logics (DLs). In OWL, semantics consists of three parts: classes (or concepts), properties (or relationships) and entities (or instances). Concepts are the abstraction of the category of

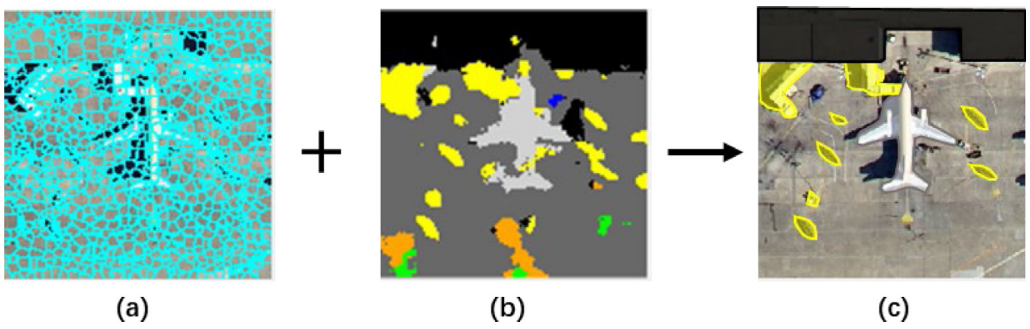

Fig. 2. Schematic for the generation of the inference unit. (a) Superpixels from the raw image. In the clustering process, the reference units (c) are generated according to (a) and initial classification (b).

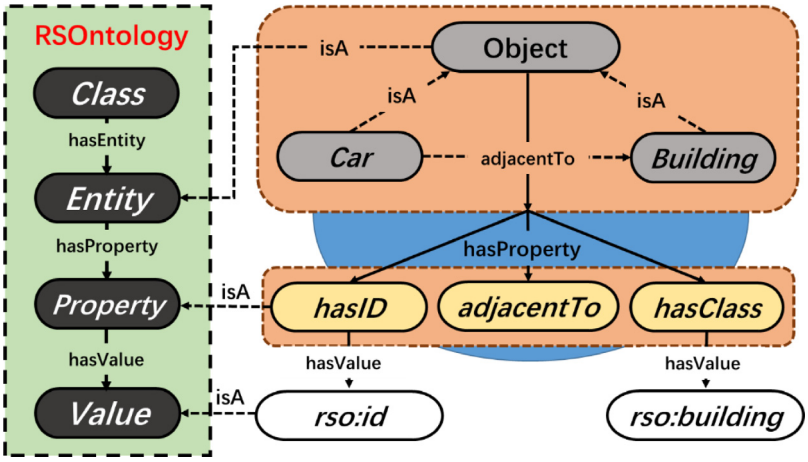

Fig. 3. Example of applying RSOntology to define the geographic object from RS images. Ontology is composed of three parts: class, entity, and property with value. In this framework, relationships between objects (e.g. car and building) are denoted by properties (e.g. hasID, adjacentTo and hasClass) with the corresponding values.

objects (e.g., "Car", "Airplane" and "Ship" belong to the class of "Transportation"). Properties are used to express the inherent characteristics of objects and relationships between objects (e.g., "ID" and "adjacent to"). Entities, which are the instances of concepts, represent objects in the real world (e.g., "Yangtze River" and "Yellow River" are the instances of "Water").

In terms of knowledge expression, ontology provides a unified interface for multi-source heterogeneous data, which eliminates the gap between data. In terms of semantic reasoning, ontology expresses highly abstract domain knowledge in a regular form, which is helpful for integrating data-type information and knowledge-type information. In terms of knowledge sharing, ontology is a multi-level open semantic model, which is easy to extend, reuse, and transfer for experts in different domains. In view of the importance of expression and application of geographic knowledge, we design a RS ontology (RSOntology) with OWL for RS image interpretation. As shown in Fig. 3, RSOntology is a generalized knowledge model for representing ground objects in terms of conceptual, semantic, and contextual aspects.

The top-down and bottom-up approaches are adopted to implement collaborative domain ontologies [47]. The top-down method summarizes the entities of the whole domain and constructs a hierarchy of ontology with abstract concepts, whereas the bottom-up approach allows domain experts to formally express relevant data and knowledge in a local ontology. In fact, both methods should be considered as complementary and necessary, providing ontologies an essential role in the articulation of knowledge-driven and data-driven approaches in RS. From the top-down perspective, we constructed a hierarchical system of classes and relationships for the instantiation of RS entities. As shown in Fig. 4, the hierarchical system of RSOntology is divided into three layers, namely the domain abstraction layer, the category collection layer, and the object category layer. At top of the class hierarchy, ***rso*:*GeoObject*** denoting RS objects is the root class, which has the broadest representation, in which prefixes ***rso***, ***geo***, and ***oe*** stand for the uniform resource identifiers (URIs) of RSOntology, the geographic relationship, and the ontology entity, respectively. In the category collection layer, ***rso*:*ManmadeArea***, ***rso*:*Ground***, ***rso*:*Transportation***, ***rso*:*Vegetation***, ***rso*:*Water***, and ***rso*:*Segment*** are all subclasses of ***rso*:*GeoObject***. Man-made spatial entity possesses the class of ***rso*:*ManmadeArea*** which includes ***rso*:*Building***, ***rso*:*Pavement***, and ***rso*:*Railway***. ***rso*:*Bareland*** and ***rso*:*Desert*** belong to ***rso*:*Ground***, but there is a difference between the two, that is, the former is the wasteland left over by urban construction, whereas the latter is the ground formed by natural degradation. ***rso*:*Airplane***, ***rso*:*Car***, ***rso*:*Ship***, and ***rso*:*Train*** are subclasses of ***rso*:*Transportation***. The class ***rso*:*Water*** is composed of ***rso*:*River***, ***rso*:*Lake***, and ***rso*:*Sea*** in the RSOntology. Similarly, the class ***rso*:*Vegetation*** includes ***rso*:*Grass*** and ***rso*:*Tree***. In particular, categories of some ground objects are uncertain or inaccurate in the process of classification. Therefore, a category must be defined to represent these objects, and then, the corresponding class must be assigned to them according to the classification. We adopt the class ***rso*:*UnknownObject*** to represent ground objects whose category needs to be determined. ***rso*:*ClassifiedObject***, ***rso*:*MisClassifiedObject***, and ***rso*:*UnClassifiedObject***, which belong to ***rso*:*UnknownObject***, denote the correctly classified object, the misclassified object, and the unclassified object, respectively. To generalize the inherent

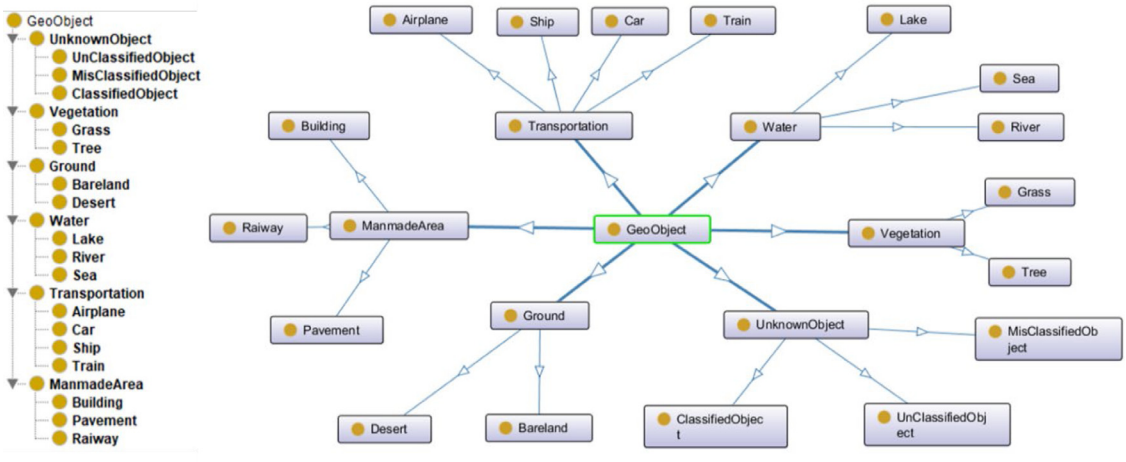

Fig. 4. Visual illustration of hierarchy of RSOntology. The hierarchical system of RSOntology is divided into three levels: the first level is the domain abstraction layer which denotes geographic objects by ***rso*** : ***GeoObject***. The second and third levels are the category collection layer and object category layer, respectively.

characteristics of objects and relationships between objects, abstract properties are defined in RSOntology. The following are the core properties: in terms of the subordinate relationship, ***geo*** : ***isA*** and ***geo*** : ***hasSubclass*** are the subclass attribute and the upper class attribute, respectively, which are used to build the hierarchical relationship of the ontology. In terms of individual characteristics, ***geo*** : ***hasID***, ***geo*** : ***Size***, and ***geo*** : ***hasClass*** denote the ID, area, and category of objects, respectively. The property ***geo*** : ***MaxClass*** represents the most likely category of the object. In terms of relationships between objects, the attributes of spatial relationship are composed of adjacency (***geo*** : ***adjacentTo***), surrounding (***geo*** : ***surroundedBy***), disjointing (***geo*** : ***disjointWith***), intersection (***geo*** : ***intersectWith***), and direction (***geo*** : ***hasDirectionOf***). From the bottom-up perspective, the inference units in Section 3.2, which represent ground objects, are used to instantiate the ontology and construct the relationship network with the defined properties. Moreover, we regularize expert knowledge under the framework of RSOntology to facilitate knowledge reasoning in Section 3.4.

3.4. Ontology reasoning

As a bridge connecting knowledge-driven and data-driven methods, ontology reasoning plays an important role in the CBF, which simulates the way humans work to process data with the domain knowledge. As shown in Fig. 5, in the proposed method, ontology reasoning is divided into two parts: intra- and extra-taxonomy ontology reasoning. Expert knowledge is usually abstract and vague, which is difficult to directly apply in the procedural process. For example, objects adjacent to each other in space have similar characteristics, while it is difficult to quantify machines. In view of the powerful knowledge expression of ontology, we regularized expert knowledge under the framework of RSOntology. According to ontology reasoning rules representing expert knowledge, intra-taxonomy ontology reasoning directly corrects misclassification within the category taxonomy and extra-taxonomy ontology reasoning extracts the estimated information of shadow and elevation as the additional channel of DSSN to indirectly assist in classification. In RSOntology, the inference units in Section 3.2 are used to instantiate the ontology for reasoning. ***oe*** : ***entity*** and ***oe*** : ***entity*1** are instances of the inference unit that is correctly classified, satisfying Eqs. (6) and (7). ***oe*** : ***misEntity*** is the instance of the misclassified unit, satisfying Eq. (8).

$$rso : ClassifiedObject\ (oe : entity) \tag{6}$$

$$rso : ClassifiedObject\ (oe : entity1) \tag{7}$$

$$rso : MisClassifiedObject\ (oe : misEntity) \tag{8}$$

3.4.1. Intra-taxonomy ontology reasoning

Because deep learning is a data-driven method, it is easily affected by noise, leading to classification errors that are usually reflected in the following aspects: (1) hole phenomenon (marked as 1 in Fig. 6): small heterogeneous areas appear in a large homogeneous area, for example, there are buildings in the middle of a large area of water or vegetation. (2) Inconsistent spatial relationships (marked as 2 in Fig. 6): the repulsiveness of spatial relationship between specific ground objects is not satisfied; for example, an airplane is adjacent to water. Notably, these two types of misclassification are common and representative in the image interpretation. Thus, we apply ontology reasoning called intra-taxonomy ontology reasoning to directly correct misclassifications within the category taxonomy.

As shown in Table 1, expert knowledge for the reasoner is symbolized as ontology rules divided into two types. One type is prepared for eliminating the hole phenomenon during classification. According to the principle of spatial correlation, the misclassified holes are more likely to be similar to the surrounding objects. Therefore, rules 1–6 are designed to adjust the category of the misclassified objects to be consistent with the category of the surrounding objects. For example, if an area misclassified as vegetation is surrounded by buildings, the true category of this area should be a building. The other type is used to correct misclassifications of inconsistent spatial relationships. The true category of misclassification is most likely to be the category in the neighborhood, which is expressed by rules 7–9. For example, misclassified water is adjacent to one car, so its class is more likely the category with the largest proportion in its neighborhood. Fig. 5 illustrates the process of intra-taxonomy ontology reasoning, where units in red are misclassified as ground: ***rso*** : ***Ground*** (***oe*** : ***misEntity***). After reasoning, the misclassifications are corrected as the pavement according to the category of surrounding units: ***rso*** : ***Pavement*** (***oe*** : ***Entity***).

3.4.2. Extra-taxonomy ontology reasoning

DSSN has a strong learning ability, whereas its poor anti-interference against visual similarity is obvious, e.g. dividing shadow areas into water and classifying buildings as pavements in Fig. 7. One possible solution to this problem is to include additional sources of information as part of the input data to

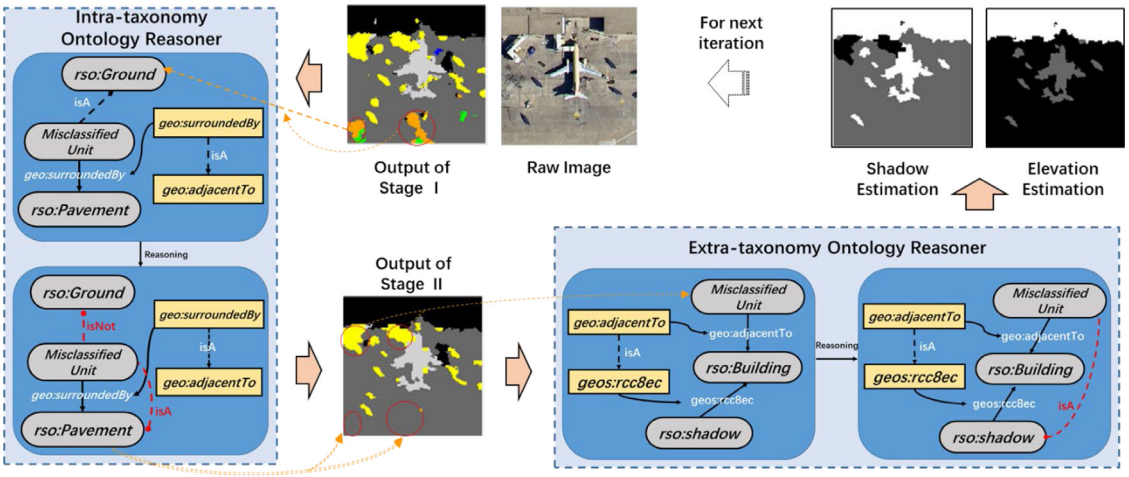

Fig. 5. Example of ontology reasoning. Intra- and extra-taxonomy ontology reasoning are shown, where intra-taxonomy ontology reasoning corrects misclassifications directly and the extra-taxonomy ontology reasoning extracts the estimation of shadow and elevation for the next iteration.

Table 1
Ontology rules of the intra-taxonomy ontology reasoning.

Num	Description	Expression based on DL
Rule 1	If an inference unit, misclassified as vegetation, is surrounded by ground, pavements, buildings, or water, its class should be corrected to the category of objects around it.	$rso:Vegetation\ (oe:misEntity)$, $rso:geoClass \sqsubseteq rso:Ground \sqcup rso:geoClass \sqsubseteq rso:Pavement$ $\sqcup\, rso:geoClass \sqsubseteq rso:Building \sqcup rso:geoClass \sqsubseteq rso:Water$, $\forall oe:entity \in (\forall geo:adjacentTo.oe:misEntity)$ $rso:geoClass\ (oe:entity)$, $\Longrightarrow rso:geoClass(oe:misEntity)$
Rule 2	If an inference unit, misclassified as ground, is surrounded by pavements, buildings, or water, its class should be corrected to the category of objects around it.	$rso:Ground\ (oe:misEntity)$, $rso:geoClass \sqsubseteq rso:Pavement \sqcup rso:geoClass \sqsubseteq rso:Building$ $\sqcup\, rso:geoClass \sqsubseteq rso:Water$, $\forall oe:entity \in (\forall geo:adjacentTo.oe:misEntity)$ $rso:geoClass\ (oe:entity)$, $\Longrightarrow rso:geoClass(oe:misEntity)$
Rule 3	If an inference unit, misclassified as a building, is surrounded by ground or water, its class should be corrected to the category of objects around it.	$rso:Building\ (oe:misEntity)$, $rso:geoClass \sqsubseteq rso:Ground \sqcup rso:geoClass \sqsubseteq rso:Water$, $\forall oe:entity \in (\forall geo:adjacentTo.oe:misEntity)$ $rso:geoClass\ (oe:entity)$, $\Longrightarrow rso:geoClass(oe:misEntity)$
Rule 4	If an inference unit, misclassified as water, is surrounded by vegetation, buildings, or pavements, its class should be corrected to the category of objects around it.	$rso:Water\ (oe:misEntity)$, $rso:geoClass \sqsubseteq rso:Vegetation \sqcup rso:geoClass \sqsubseteq rso:Building$ $\sqcup\, rso:geoClass \sqsubseteq rso:Pavement$ $\forall oe:entity \in (\forall geo:adjacentTo.oe:misEntity)$ $rso:geoClass\ (oe:entity)$, $\Longrightarrow rso:geoClass(oe:misEntity)$
Rule 5	If an inference unit, misclassified as airplane, is surrounded by vegetation, buildings, or water, its class should be corrected to the category of objects around it.	$rso:Airplane\ (oe:misEntity)$, $rso:geoClass \sqsubseteq rso:Vegetation \sqcup rso:geoClass \sqsubseteq rso:Building$ $\sqcup\, rso:geoClass \sqsubseteq rso:Water$, $\forall oe:entity \in (\forall geo:adjacentTo.oe:misEntity)$ $rso:geoClass\ (oe:entity)$, $\Longrightarrow rso:geoClass(oe:misEntity)$
Rule 6	If an inference unit, misclassified as car, is surrounded by vegetation or water, its class should be corrected to the category of objects around it.	$rso:Car\ (oe:misEntity)$, $rso:geoClass \sqsubseteq rso:Vegetation \sqcup rso:geoClass \sqsubseteq rso:Water$ $\forall oe:entity \in (\forall geo:adjacentTo.oe:misEntity)$ $rso:geoClass\ (oe:entity)$, $\Longrightarrow rso:geoClass(oe:misEntity)$
Rule 7	If an inference unit is misclassified as an airplane, with none of the correctly classified objects which have adjacent pavements, its class should be the category with the most correctly classified objects in its neighborhood.	$rso:Airplane\ (oe:misEntity)$, $\forall oe:entity \in (\forall geo:adjacentTo.oe:misEntity)$ $oe:entity \in \neg rso:Pavement$, $\Longrightarrow oe:misEntity \in geo:MaxClass(oe:entity)$
Rule 8	If an inference unit is misclassified as a car, with none of the correctly classified objects which have adjacent pavements, its class should be the category with the most correctly classified objects in its neighborhood.	$rso:Car\ (oe:misEntity)$, $\forall oe:entity \in (\forall geo:adjacentTo.oe:misEntity)$ $oe:entity \in \neg rso:Pavement$, $\Longrightarrow oe:misEntity \in geo:MaxClass(oe:entity)$
Rule 9	If an inference unit is misclassified as a ship, with none of the correctly classified objects which have adjacent water, its class should be the category with the most correctly classified objects in its neighborhood.	$rso:Ship\ (oe:misEntity)$, $\forall oe:entity \in (\forall geo:adjacentTo.oe:misEntity)$ $oe:entity \in \neg rso:Water$, $\Longrightarrow oe:misEntity \in geo:MaxClass(oe:entity)$

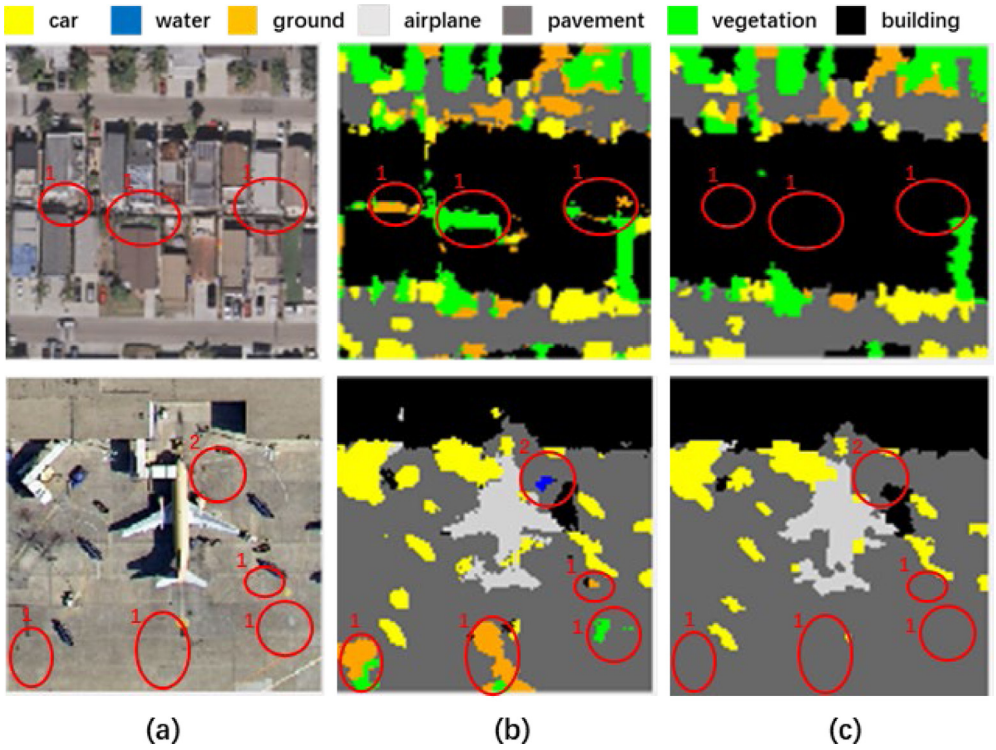

Fig. 6. Major misclassifications of the DSSN: hole phenomenon (marked as 1) and inconsistent spatial relationships (marked as 2). (a) Raw image and (b) classification of the DSSN. (c) Result of ontology reasoning on (b) from (b) to (c); both holes and inconsistent spatial relationships are corrected by the ontology reasoning.

DSSN. Additional information such as DEM, shadow data, and infrared band data can effectively enhance the deep network's anti-interference ability against visual similarity. Because of light occlusion, shadows are commonly present in RS images. When shadows are projected on the ground objects, it will change the reflection spectrum of ground objects. Because DSSN mainly adopts spectral information as the basis for discrimination, the appearance of shadows will seriously affect the performance of the classifier. In addition, although some objects have similar spectral characteristics in images, such as buildings, roads, grass, and trees, their elevations are distinguishable. Therefore, elevation information is beneficial to improve the discriminative ability of the classifier. However, additional information is usually difficult to obtain. Thus, generating estimation of the additional information through ontology reasoning considerably reduces the pressure on data acquisition. Based on the above considerations, we select the shadow estimation and the elevation estimation for the classifier, which indicates that the input of DSSN will have five color channels (three RGB channels + 1 shadow estimation channel + 1 elevation estimation channel) instead of the original three RGB channels.

In the CBF, the extra-taxonomy ontology reasoning is applied to estimate shadow and elevation. The ontology rules for the extra-taxonomy ontology reasoning in Table 2 are divided into two types. One type is used to estimate shadows, including rules 1–4. For example, if an object that is misclassified as pavements, ground, water, or cars is adjacent to correctly classified buildings, there is most likely a shadow in the corresponding area. In the extra channel, pixels are assigned with values of 1 (shadow), 0 (uncertain), and −1 (not shadow). The other type is used to extract the estimation of elevation, including rules 5–7. For example, if an object is correctly classified as a building, it has a high elevation. In the extra channel, pixels are assigned values of 2 (high elevation), 1 (medium elevation), and 0 (low elevation). Fig. 5 illustrates the process of the extra-taxonomy ontology reasoning, where units in red are misclassified as a car and are adjacent to buildings. According to Rule 1, the corresponding areas exist as shadows (marked with black).

3.5. Overview of the proposed CBF

In general, the process of the CBF is composed of three steps above including initial semantic segmentation based on DSSN, generating inference units, and ontology reasoning, which is briefly summarized in Algorithm 1.

Algorithm 1. Collaboratively boosting data-driven deep learning and knowledge-guided ontology reasoning for semantic segmentation of RS imagery

Input: RSOntology, the RS image dataset $\boldsymbol{D}_T$; the number of superpixels k; the number of iteration N.
1. Regions segmented by the unsupervised segmentation algorithm are used as superpixels $\boldsymbol{S} = \{s_1, s_2, \cdots, s_k\}$ for each image in $\boldsymbol{D}_T$.
2. **while** $i < N$ **do**
3. Train DSSN with samples from $\boldsymbol{D}_T$.
4. Generate initial classification $\boldsymbol{O}_1$ (the output of **Stage I**) by DSSN.
5. Cluster superpixels according to $\boldsymbol{O}_1$ to generate the inference units.
6. Implement the intra-taxonomy ontology reasoning on $\boldsymbol{O}_1$ to correct misclassifications and get the corrected results $\boldsymbol{O}_2$ (the output of **Stage II**).
7. Implement the extra-taxonomy ontology reasoning to produce estimation of shadow and elevation.
8. Return the estimation information to DSSN as the additional input channel.
9. Iteration: $i = i + 1$
10. **end while**

Output: $\boldsymbol{O}_1$, $\boldsymbol{O}_2$ and the estimation of shadow and elevation.

Table 2
Ontology rules of the extra-taxonomy ontology reasoning.

Num	Description	Expression based on DL
Rule 1	The misclassification is pavement, ground, water, or car. If there is a correctly classified building in its neighborhood, there is a shadow in the area.	$rso:geoClass \sqsubseteq rso:Pavement \sqcup rso:Ground \sqcup rso:Water \sqcup rso:Car$, $rso:geoClass\ (oe:misEntity)$, $\exists oe:entitiy \in (\forall geo:adjacentTo.oe:misEntity)$ $rso:Building\ (oe:entitiy)$, $\Longrightarrow rso:Shadow(oe:misEntity)$
Rule 2	The misclassification is car, ship, vegetation, or airplane. If there are no correctly classified buildings in its neighborhood, there is no shadow in the area.	$rso:geoClass \sqsubseteq rso:Vegetation \sqcup rso:Car \sqcup rso:Ship \sqcup rso:Airplane$, $rso:geoClass\ (oe:entity)$, $\exists oe:entitiy \in (\forall geo:adjacentTo.oe:misEntity)$ $oe:entitiy \in \neg rso:Building$, $\Longrightarrow rso:NonShadow(oe:entity)$
Rule 3	The correct classification is ground. If there are no correctly classified buildings and vegetation in its neighborhood, there is no shadow in the area.	$rso:Ground\ (oe:entity)$, $\exists oe:entitiy1 \in (\forall geo:adjacentTo.oe:entity)$ $oe:entitiy1 \in (\neg rso:Building \sqcap \neg rso:Vegetation)$, $\Longrightarrow rso:NonShadow(oe:entity)$
Rule 4	If the correct classification is a building, there is no shadow in the area.	$rso:Building\ (oe:entity)$, $\Longrightarrow rso:NonShadow(oe:entity)$
Rule 5	If an object is correctly classified as vegetation, ground, pavements, or water, it has low elevation.	$rso:geoClass \sqsubseteq rso:Vegetation \sqcup rso:Ground \sqcup rso:Pavement \sqcup rso:Water$, $rso:geoClass\ (oe:entity)$, $\Longrightarrow geo:hasLowElevation(oe:entity)$
Rule 6	If an object is correctly classified as an airplane, car, or ship, it has medium elevation.	$rso:geoClass \sqsubseteq rso:Airplane \sqcup rso:Car \sqcup rso:Ship$, $rso:geoClass\ (oe:entity)$, $\Longrightarrow geo:hasMediumElevation(oe:entity)$
Rule 7	If an object is correctly classified as a building, it has high elevation.	$rso:Buliding\ (oe:entity)$, $\Longrightarrow geo:hasHighElevation(oe:entity)$

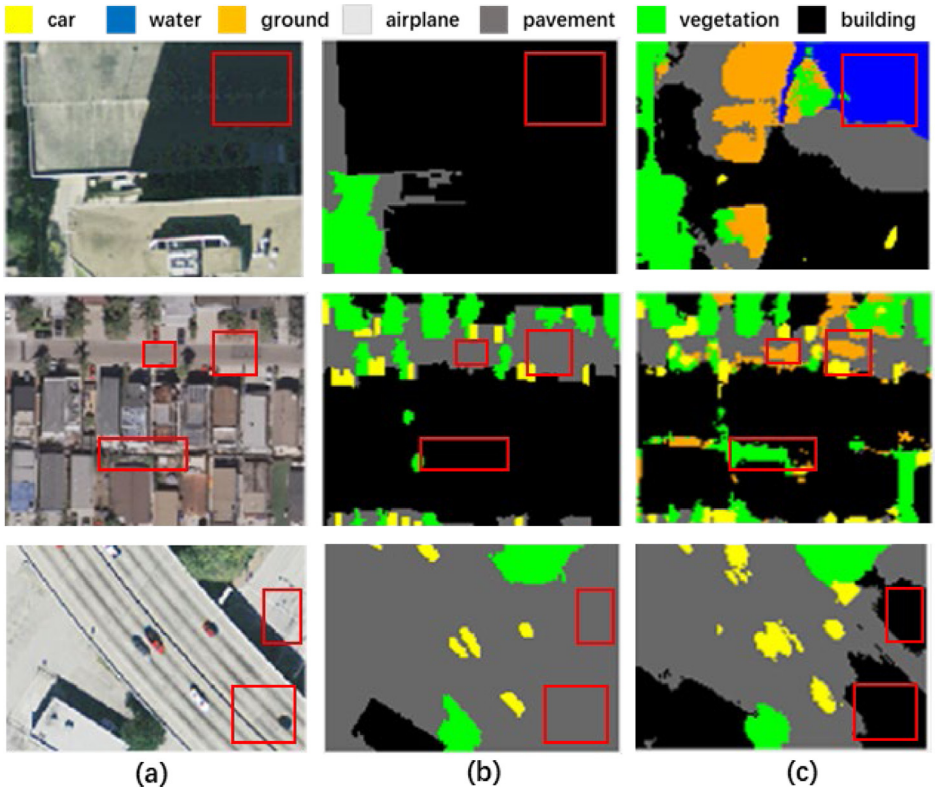

Fig. 7. DSSN's poor anti-interference against visual similarity. (a) Raw image. (b) Ground truth. (c) DSSN classification. Visual similarity leads the DSSN to make misclassification (red box), e.g. dividing shadow areas into water, labeling pavements as ground, and classifying pavements as buildings.

4. Experiments and discussion

In this section, data description and details of the experimental settings are introduced first. The experimental results and analysis are given after that.

4.1. Datasets and evaluation metrics

To fairly show the superiority of our method, experiments were performed on the UCM dataset and the ISPRS Potsdam dataset.

As visually illustrated in Fig. 8, the UCM dataset includes 2100 RS images with 0.3 m spatial resolution, and each image

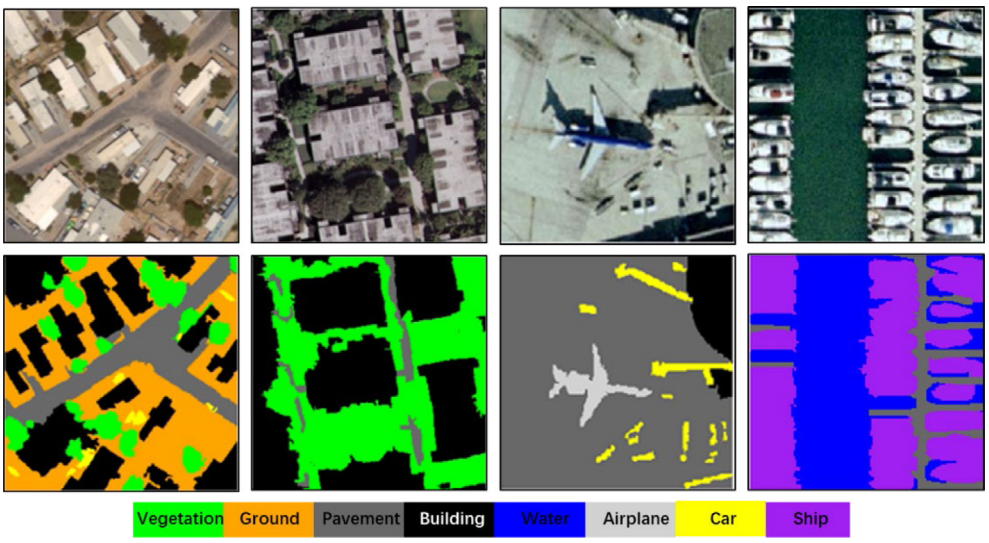

Fig. 8. Raw images and ground truth masks of the UCM dataset.

in UCM is labeled with 17 categories for semantic segmentation [48]. Following [21], categories with similar semantics are merged to fit the category system of reasoning. The new categories include Vegetation (trees, grass), Ground (bare soil, sand, chaparral), Pavement (pavement, dock), Building (building, mobile home, tank), Water (water, sea), Airplane (airplane), Car (car), and Ship (ship). Images containing field or tennis court are removed due to the category system of reasoning. Each category is a combination of the original categories in parentheses. The filtered images are randomly divided into the training set, validation set, and test set, each with 1513, 189, and 190 images in the proportions of 80%, 10%, and 10%, respectively.

As shown in Fig. 9, the Potsdam dataset was divided into six common land-cover classes: Impervious Surfaces, Building, Low Vegetation, Tree, Car, and Clutter (background). Clutter includes water, containers, tennis courts, swimming pools, etc. The dataset contains 38 aerial orthographic images of urban areas with 0.05 m spatial resolution and a pixel size of 6000 × 6000. Because of the limitation of GPU memory, multiple 512 × 512 images are cropped from each image. These cropped images are randomly divided into the training set, validation set, and test set, each with 2758, 919, and 921 images in the proportions of 60%, 20%, and 20%, respectively.

4.2. Experimental setup and evaluation metrics

As a classic deep network for semantic segmentation, U-Net [49] was adopted as the DSSN backbone in this implementation. The cross-entropy loss function and the Adam backward propagation optimization algorithm [50] with a learning rate of $10e^{-4}$ are applied for training DSSN. To generate inference units, we used simple linear iterative clustering (SLIC) [51] to obtain superpixels. The number K of superpixel segmentation and the confidence threshold F_t are set to 1000 and 0.7, respectively. For the evaluation of classification, the overall accuracy (OA), the intersection over union (IoU), the mean intersection over union (MIoU), and the frequency weighted intersection over union (FWIoU) were adopted as the evaluation metrics [52].

$$OA = \frac{TP + TN}{TP + FP + TN + FN} \tag{9}$$

$$IoU_i = \frac{TP_i}{TP_i + FP_i + FN_i}, i = 1, 2, \ldots, n \tag{10}$$

$$MIoU = \frac{1}{n}\sum_{1}^{n} IoU_i \tag{11}$$

$$FWIoU = \sum_{1}^{n}(IoU_i \cdot \frac{TP_i + FN_i}{TP_i + FP_i + TN_i + FN_i}) \tag{12}$$

where TP, TN, FP, and FN are the number of true positive points, true negative points, false positive points, and false negative points, respectively. n is the number of classes.

4.3. Sensitivity analysis of critical parameters

In the proposed CBF, the DSSN-based classification module and the ontology reasoning module interact in an iterative manner to optimize the output of the complete system. These two modules continue to promote each other in the iterations until the accuracy converges. The former provides the initial semantic segmentation for the latter, while the latter corrects misclassifications and generates the estimation of shadow and elevation from the corrected segmentation. The estimation information is used as additional input for the former to enter the next iteration. The iteration plays a vital role in the loop. As shown in Fig. 10, along with an increase in the iterations, the classification accuracy of Stage I and Stage II will continue to improve until the best is achieved, which shows the effectiveness of the iterative strategy and demonstrates that the classification module and the ontology reasoning module promote each other. Two main reasons exist for the decrease of accuracy at the end iteration: (1) In subsequent iterations, the estimation information becomes accurate. Compared to complex RGB images, DSSN tends to learn from the estimation information, which weakens the discriminative power of the network. (2) Because of the influence of closed loop iteration, some misclassifications that cannot be corrected by the system will continue to accumulate. When the negative effects of these errors exceed the positive effects of reasoning, the segmentation performance of the network will reduce. To circumvent this problem, we can obtain the best performance by determining the optimal iteration. The optimal iteration on the UCM dataset and the Potsdam dataset is 3 and 4, respectively. Moreover, the performances of both Stage I and Stage II are better than that of the backbone, which demonstrate the good performance of the CBF. The success of the intra-taxonomy ontology reasoning is reflected in the fact that the accuracy of Stage II is much higher than that of Stage I both in the UCM dataset and the Potsdam

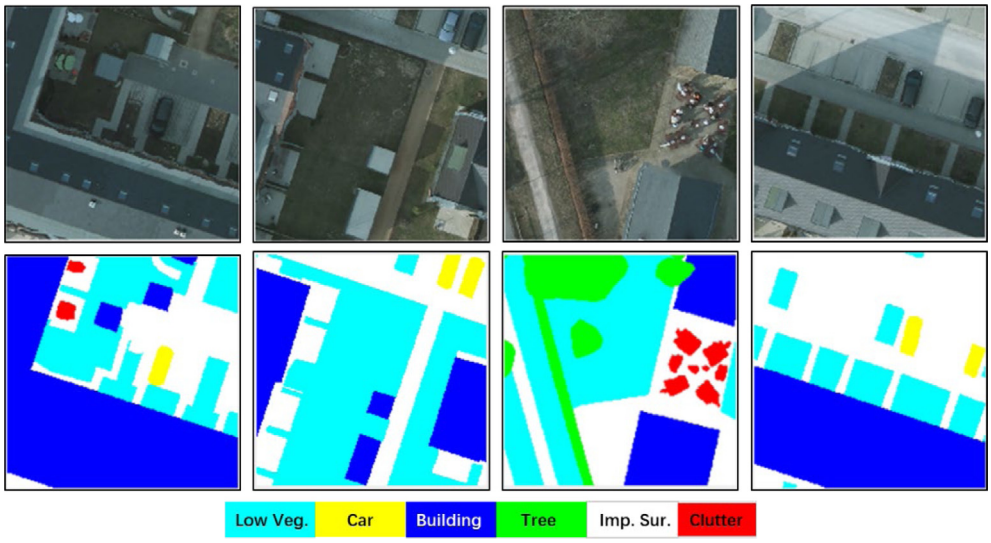

Fig. 9. Raw images and ground truth masks of the Potsdam dataset.

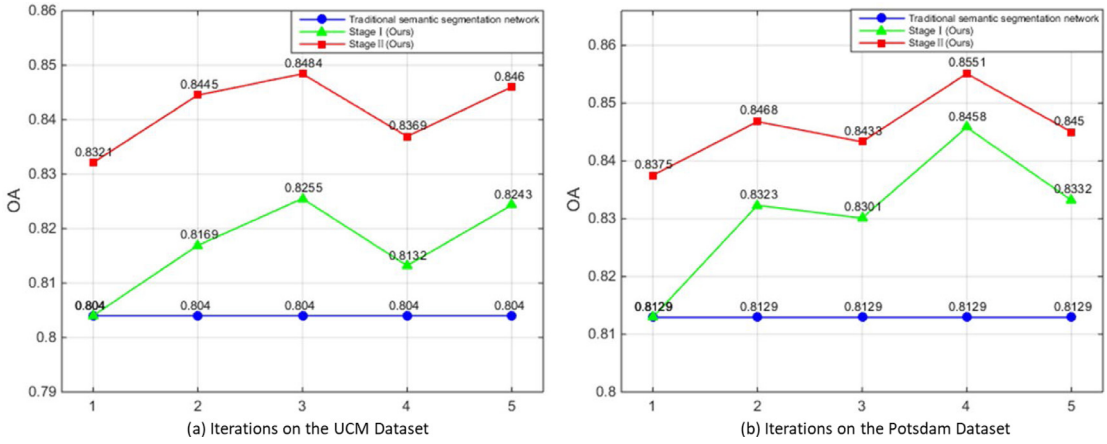

Fig. 10. Performance variation of the proposed method under different iterations. (a) and (b) Classification accuracy on the UCM dataset and the Potsdam dataset, respectively. The blue line represents the classification accuracy of the backbone network (baseline). The green line depicts the trend of output of Stage I in the CBF. After ontology reasoning on Stage I, Stage II outputs the refined results (red line).

dataset. This shows the feasibility of using knowledge reasoning to directly correct the misclassification.

The semantic segmentation result and estimation information of the CBF on the UCM dataset are visually shown on Fig. 11. In the first and third iterations, the segmentation result of Stage II is significantly better than that of Stage I mainly because of the help of intra-taxonomy ontology reasoning. The red cycles in Fig. 11 represent places to be corrected. Misclassifications of Stage I are corrected by reasoning ((c) to (d), (e) to (f)), such as ground changed to pavement, water divided into buildings, and buildings corrected as pavement. Simultaneously, the intra-taxonomy ontology reasoning eliminates the noises of the initial classification, making classification more accurate ((d) and (f)). After three iterations, the network performance is optimal where the classification errors are significantly reduced. At the same stage, with an increase in the iteration, the segmentation becomes better. This is because of the direct correction of ontology reasoning and the indirect assistance of estimation information. This clearly shows that the estimation of shadow and elevation generated by the extra-taxonomy ontology reasoning is more accurate in the iterative process ((g) to (h), (i) to (j)), which is attributed to the corrected output of Stage II and the improvement of DSSN's anti-interference. Conversely, accurate estimation information also improves the network performance. Therefore, in the CBF, the ontology reasoning and the classification modules complement each other in iterations.

Fig. 12 presents the semantic segmentation and estimation information on the Potsdam dataset of the CBF. In Stage I, misclassifications, which are presented by the red cycles, are corrected by ontology reasoning ((d) to (c), (f) to (e) in the figure), such as buildings divided into impervious surface, clutters classified as buildings, and low vegetation corrected as buildings. With the help of intra-taxonomy ontology reasoning, the segmentation of Stage II is significantly better than that of Stage I in the first and third iterations. Meanwhile, noise holes in the classification are eliminated by intra-taxonomy ontology reasoning ((d) and (f)), which improves the accuracy of classification. With an increase in the iteration, the segmentation improves. After three iterations, the network performance is optimal where the classification errors are significantly reduced. It can be clearly seen that the segmentation results ((c) to (e), (d) to (f)) and the estimation of shadow and elevation ((g) to (h), (i) to (j)) are more accurate in the iterative process, which demonstrates the effectiveness of the CBF framework.

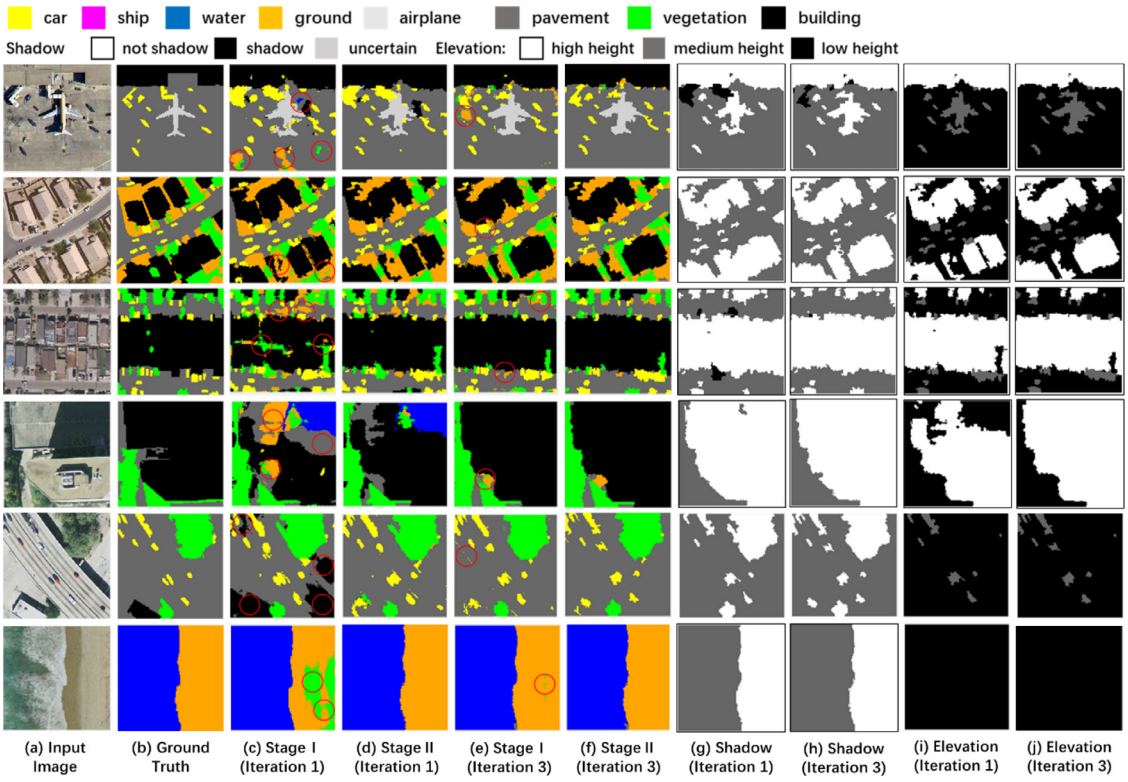

Fig. 11. Semantic segmentation and estimation information of the CBF on the UCM dataset. (a) and (b) Input images and ground truth, respectively. In the first iteration, (c) and (d) show the classifications of Stage I and Stage II, respectively. (g) and (i) Estimation of shadow and elevation, respectively. In the third iteration, (e) and (f) present the classification of Stage I and Stage II, respectively. (h) and (j) Estimation of shadow and elevation, respectively. The red cycles of Stage I represent the places to be corrected in Stage II.

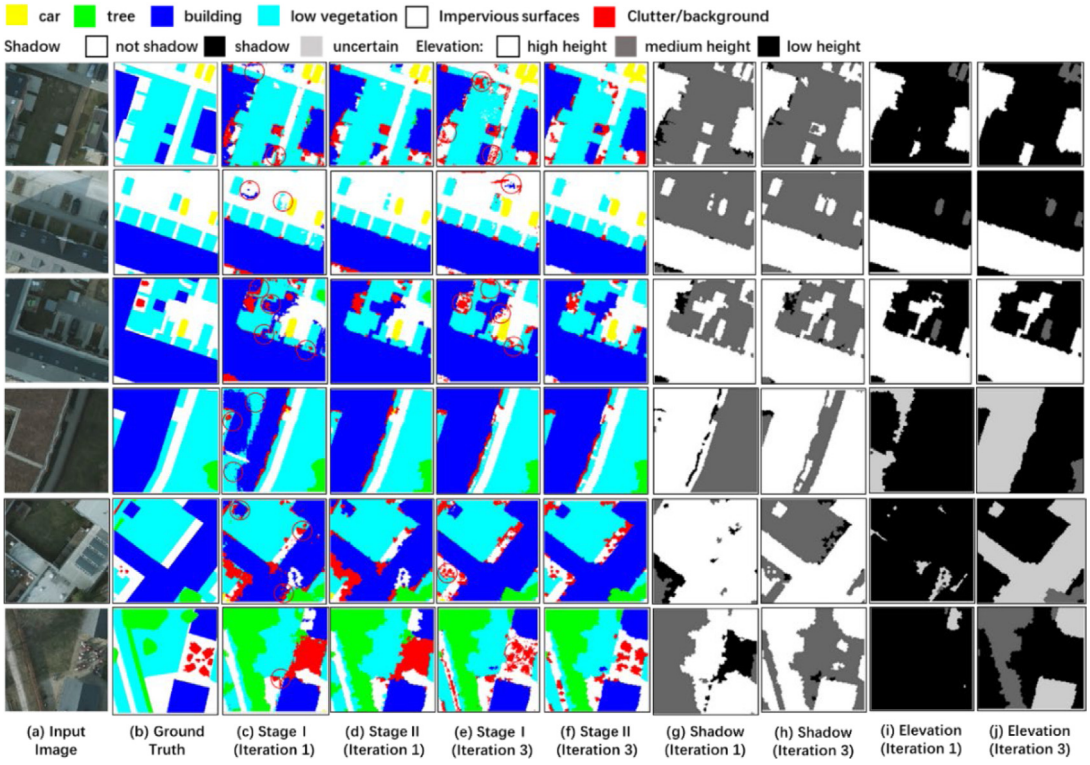

Fig. 12. Semantic segmentation result and estimation information of the CBF on the Potsdam dataset. (a) and (b) Input images and ground truth, respectively. In the first iteration, (c) and (d) show classification of Stage I and Stage II, respectively. (g) and (i) Estimation of shadow and elevation, respectively. In the third iteration, (e) and (f) present the classification of Stage I and Stage II, respectively. (h) and (j) Estimation of shadow and elevation, respectively. The red cycles of Stage I represent the misclassifications to be corrected in Stage II.

Table 3
Overall accuracy (OA) (%) of semantic segmentation on the UCM dataset.

Method	Vegetation	Ground	Pavement	Building	Water	Car	Ship	Airplane	Overall (OA)
Baseline	75.76	81.79	76.08	86.27	85.36	92.38	97.66	**86.26**	80.26
Semantic Referee	77.60	78.57	80.34	**92.38**	88.64	**92.61**	97.66	73.21	82.28
Stage I (Ours)	81.71	80.21	81.44	91.08	88.27	91.78	**98.19**	63.58	83.74
Stage II (Ours)	**86.41**	**81.94**	**83.95**	90.93	**90.05**	87.87	95.93	61.63	**85.92**

4.4. Comparison with the state-of-the-art methods

To fairly show the effectiveness of the proposed CBF, experiments were conducted on two datasets. As aforementioned, few studies have been conducted on coupling deep learning and knowledge reasoning in the field of RS image interpretation. Semantic Referee [21], which couples DSSN and knowledge reasoning for RS image semantic segmentation, is adopted as a comparison method. For fair comparison, U-Net [49] is adopted as the network backbone of the Semantic Referee and our CBF. In addition, U-Net is taken as the baseline in the following experiments.

4.4.1. Results on the UCM dataset

The OA of different semantic segmentation methods on the UCM dataset is shown in the last column of Tables 3 and 4. Stage I corresponds to the initial semantic segmentation result of the CBF, and Stage II denotes the result corrected by ontology reasoning on the classification of Stage I. Compared with the baseline, the OA/MIoU/FWIoU increases by 3.48%/2.79%/4.87% for Stage I and Stage II raises the OA/MIoU/FWIoU by 5.66%/4.92%/8.12%. This significant improvement proves the effectiveness of integrating DSSN and ontology reasoning in the CBF. Compared with Semantic Referee, the OA/MIoU/FWIoU of Stage I increases by 1.46%/1.05%/2.15% and that of Stage II increases by 3.64%/3.18%/5.4%, which demonstrates the superiority of the CBF. In addition, Stage II raises the OA/MIoU/FWIoU by 2.18%/2.13%/3.25% compared to Stage I, which indicates that misclassifications are directly corrected by the intra-taxonomy ontology reasoning and shows that the intra-taxonomy ontology reasoning is the key to improve the classification performance. Tables 3 and 4 report the OA and the IoU of semantic segmentation of each category on the UCM dataset, respectively. Among all methods, the CBF (including Stage I and Stage II) almost achieves the best classification accuracy for each category. After intra-taxonomy ontology reasoning, Stage II completes better semantic segmentation on the basis of Stage I, especially for Vegetation, Ground, Pavement, and Water, which account for a large proportion in the UCM dataset. However, it can be easily seen that the proposed method does not work well on the Car, Ship, and Airplane categories. Especially, a considerable reduction in the Airplane category exists when compared to the baseline. This is because these three types of objects are so small that they are easily misclassified by the reasoner. In addition, the airplane category occupies the smallest proportion in the dataset, so its segmentation accuracy is more susceptible to reasoning correction. If it is necessary in practical applications, particular ontology reasoning rules can be designed for such categories to avoid performance degradation.

Visible semantic segmentation results of different methods on the UCM dataset are visually shown in Fig. 13 to qualitatively verify the aforementioned statement. Fig. 13 shows that Stage I of the proposed CBF achieves more accurate and consistent classification results compared to the baseline and Semantic Referee. Moreover, semantic segmentation of Stage II is significantly better than that of Stage I, which highlights the effectiveness of ontology reasoning to improve the results of semantic segmentation.

4.4.2. Results on the Potsdam dataset

The last column of Tables 5 and 6 shows the OA of different semantic segmentation methods on the Potsdam dataset. The OA/MIoU/FWIoU of Stage I is higher than the baseline by 3.69%/4.64%/5.52%, which shows that the estimation information generated by the extra-taxonomy ontology reasoning improves the discrimination performance of DSSN. It verifies the effectiveness of the proposed CBF that Stage II raises the OA/MIoU/FWIoU by 4.63%/5.56%/4.76% compared with the baseline. Compared to Semantic Referee, the OA/MIoU/FWIoU of Stage I increases by 1.98%/2.12%/3.49%, and the OA/MIoU/FWIoU of Stage II increases by 2.92%/3.04%/2.73%, which demonstrates the advancement of the CBF. In addition, Stage II raises the OA/MIoU by 0.94%/0.92% compared to Stage I, which indicates that the intra-taxonomy ontology reasoning directly corrects misclassifications. The FWIoU of Stage II is slightly lower than that of Stage I is because of the bad performance on Clutter caused by the intra-taxonomy ontology reasoning's focus on main categories rather than Clutter. The significant improvement of classification in Stage II shows that the intra-taxonomy ontology reasoning is the key to the CBF. The OA and the IoU of semantic segmentation of each category on the Potsdam dataset are shown in Tables 5 and 6, respectively. Among all the methods, the CBF (including Stage I and Stage II) almost achieves the best classification accuracy for each category. After intra-taxonomy ontology reasoning, Stage II completes better semantic segmentation on the basis of Stage I, especially in Impervious Surface, Building, and Low Vegetation, which account for a large proportion in the Potsdam dataset.

The visible semantic segmentation results of different methods on the Potsdam dataset are visually shown in Fig. 14. Stage I of the proposed CBF achieves better classification compared to the baseline and the Semantic Referee. In addition, semantic segmentation of Stage II is significantly better than that of Stage I, which shows the effectiveness of ontology reasoning to correct misclassifications in the initial semantic segmentation.

5. Conclusion

To improve the interpretability and reliability of deep learning-based RS image semantic segmentation methods, this study presents a novel CBF to couple data-driven deep learning and knowledge-guided ontology reasoning for RS image semantic segmentation, which realizes an interaction between the data-driven and knowledge-guided methods in an iterative manner. In the CBF, the deep learning module autonomously learns low-level and mid-level features from RS images, and ontology reasoning comprises intra- and extra-taxonomy ontology reasoning. The intra-taxonomy reasoning directly corrects misclassifications of the deep learning module, which helps to improve the classification performance. The extra-taxonomy reasoning calculates the estimates of shadow and elevation from the corrected results as additional information to enhance the anti-interference capability of the deep learning module, thus improving the reliability of the interpretation. Ontology reasoner applies high-level domain knowledge to guide the interpretation including directly correcting misclassifications and indirectly extracting additional information to assist the deep learning module. The whole process forms a closed loop and iterates continuously until the

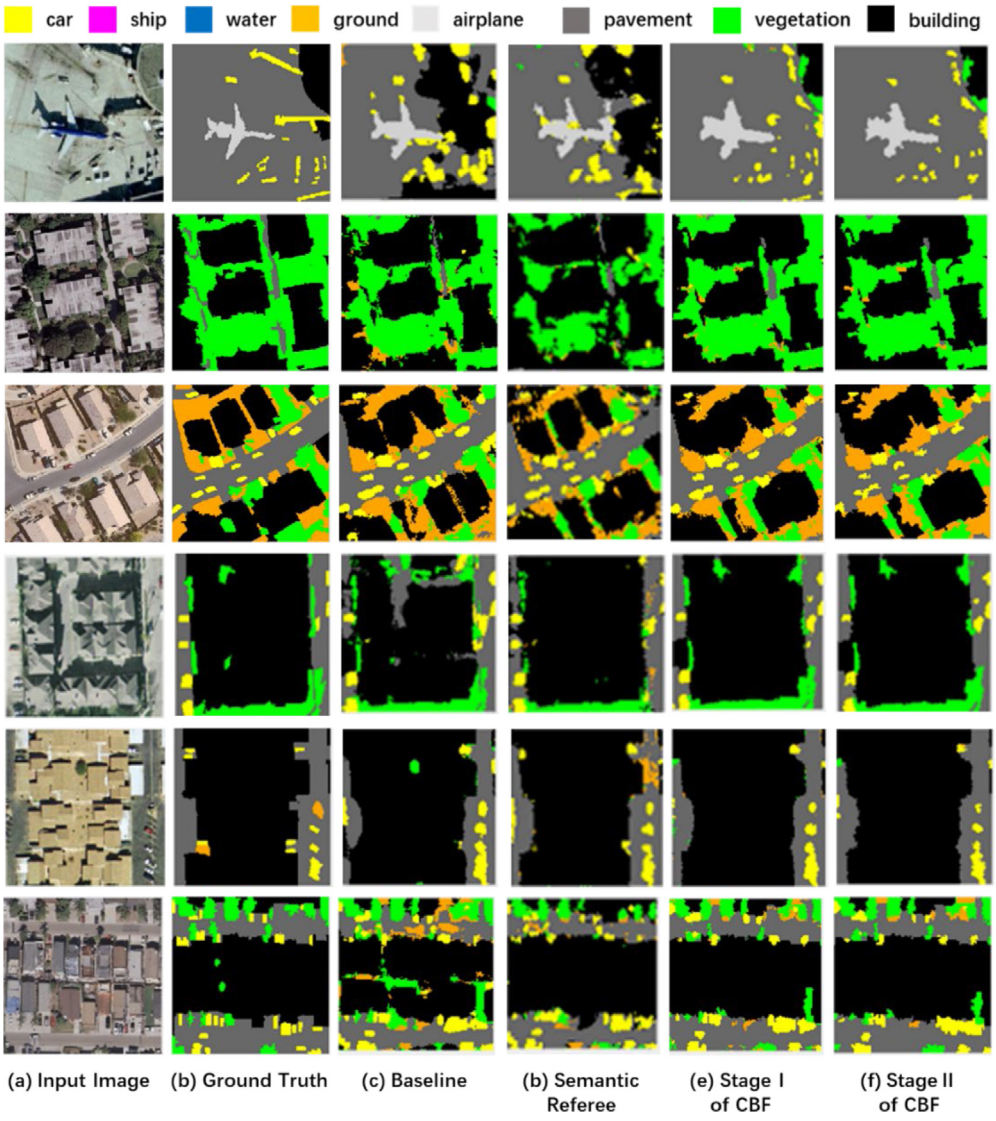

Fig. 13. Visible semantic segmentation on the UCM dataset. (a) and (b) Input images and ground truth data, respectively. (c) and (d) Output of baseline and output of Semantic Referee, respectively. In the proposed CBF, semantic segmentation results of Stage I and Stage II are shown in (e) and (f), respectively.

Table 4
Intersection over union (IoU) (%) of semantic segmentation on the UCM dataset.

Method	Vegetation	Ground	Pavement	Building	Water	Car	Ship	Airplane	Overall (MIoU)	Overall (FWIoU)
Baseline	67.00	64.10	68.66	64.16	81.97	62.22	68.34	**52.04**	66.06	67.34
Semantic referee	68.53	65.44	71.64	72.33	82.88	63.35	66.98	51.23	67.80	70.06
Stage I (Ours)	70.95	66.18	74.22	**75.26**	83.60	**69.75**	67.37	43.42	68.85	72.21
Stage II (Ours)	**75.83**	**72.20**	**76.06**	75.23	**87.02**	69.46	**67.60**	44.45	**70.98**	**75.46**

Table 5
Overall accuracy (OA) (%) of semantic segmentation on the Potsdam dataset.

Model	Imp. Sur.	Building	Low Veg.	Tree	Car	Clutter	Overall (OA)
Baseline	78.83	88.28	**82.81**	73.20	**95.67**	67.50	81.05
Semantic Referee	76.90	93.75	76.93	**88.52**	91.90	65.20	82.76
Stage I (Ours)	83.23	93.82	80.96	81.40	94.25	70.17	84.74
Stage II (Ours)	**85.42**	**93.97**	82.38	81.38	88.50	**70.78**	**85.68**

Table 6
Intersection over union (IoU) (%) of semantic segmentation on the Potsdam dataset.

Model	Imp. Sur.	Building	Low Veg.	Tree	Car	Clutter	Overall (MIoU)	Overall (FWIoU)
Baseline	71.90	81.83	65.06	63.33	69.27	67.50	64.17	71.37
Semantic Referee	72.68	88.75	65.61	63.01	**72.13**	65.20	66.69	73.40
Stage I (Ours)	76.88	90.26	68.73	**68.95**	71.61	**70.17**	68.81	**76.89**
Stage II (Ours)	**78.26**	**90.31**	**69.86**	68.53	68.58	42.84	**69.73**	76.13

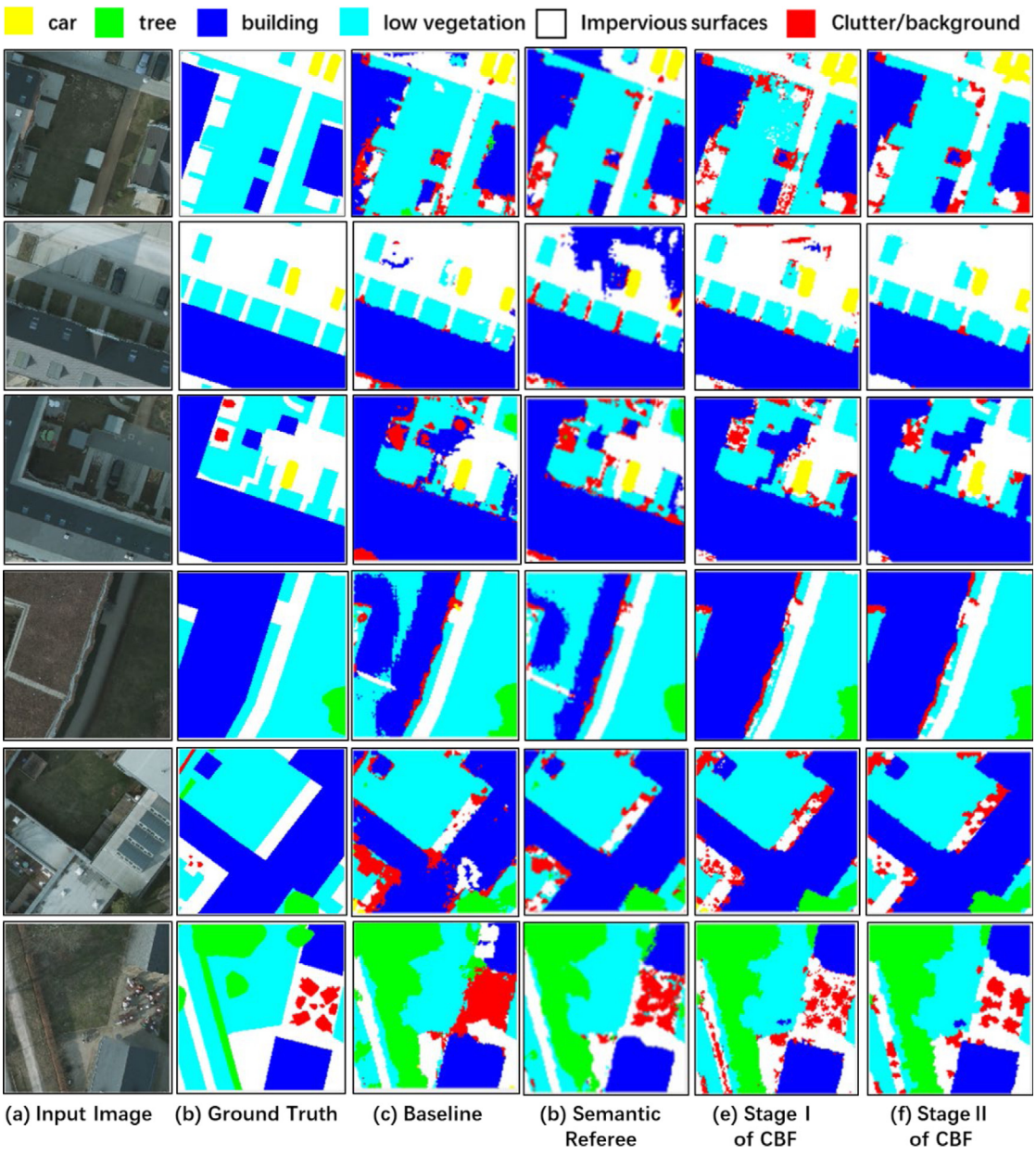

Fig. 14. Visible semantic segmentation on the Potsdam dataset. (a) and (b) Input images and ground truth, respectively. (c) and (d) Output of the baseline and Semantic Referee, respectively. In the proposed CBF, semantic segmentation results of Stage I and Stage II are shown in (e) and (f), respectively.

classification framework converges, which effectively improves the classification performance. Extensive experiments on two publicly open datasets highlight that the proposed CBF not only improves the classification interpretability and reliability but also further promotes the classification accuracy.

The presented CBF mainly considers the spatial distribution knowledge, which has effectively improved the classification performance. Naturally, more types of domain knowledge will help to further improve the classification performance. In future work, we will introduce more kinds of expert knowledge such as Gestalt rules into the ontology reasoning module to further improve the classification performance. Although the intra-taxonomy ontology reasoning effectively improves the classification performance, the performance may be inevitably affected by the aforementioned rigid confidence parameter. How to adaptively determine the confidence parameter will be explored in future work. Moreover, this study only couples deep learning and ontology reasoning from the input and output perspectives. In future work, we will explore how to integrate deep learning and knowledge reasoning at a deeper level.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under grant 41971284; the State Key Program of the National Natural Science Foundation of China under Grant 42030102; and the Foundation for Innovative Research Groups of the Natural Science Foundation of Hubei Province, China under Grant 2020CFA003.

References

[1] E. Basaeed, H. Bhaskar, M. Al-Mualla, Supervised remote sensing image segmentation using boosted convolutional neural networks, Knowl.-Based Syst. 99 (2016) 19–27.

[2] Yansheng Li, Yongjun Zhang, A New Paradigm of Remote Sensing Image Interpretation By Coupling Knowledge Graph and Deep Learning, Geomatics and Information Science of Wuhan University, 2022, http://dx.doi.org/10.13203/j.whugis20210652.

[3] J. Ball, D. Anderson, C.S. Chan, A comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community, J. Appl. Remote Sens. 11 (4) (2017).

[4] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 3431–3440.

[5] Yansheng Li, D. Kong, Y. Zhang, Y. Tan, L. Chen, Robust deep alignment network with remote sensing knowledge graph for zero-shot and generalized zero-shot remote sensing image scene classification, ISPRS J. Photogramm. Remote Sens. 179 (2021) 145–158.

[6] Y. Zhang, M. Ye, Y. Gan, W. Zhang, Knowledge based domain adaptation for semantic segmentation, Knowl.-Based Syst. 193 (4) (2020).

[7] S. Li, H. Wu, D. Wan, J. Zhu, An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine, Knowl.-Based Syst. 24 (1) (2011) 40–48.

[8] G. Mountrakis, I. Jungho, C. Ogole, Support vector machines in remote sensing: A review, ISPRS J. Photogramm. Remote Sens. 66 (3) (2011) 247–259.

[9] S. Wang, S. Zhang, T. Wu, Y. Duan, L. Zhou, H. Lei, Fmdbn: A first-order Markov dynamic Bayesian network classifier with continuous attributes, Knowl.-Based Syst. 195 (11) (2020).

[10] H. Hong, J. Liu, A. Zhu, Modeling landslide susceptibility using LogitBoost alternating decision trees and forest by penalizing attributes with the bagging ensemble, Sci. Total Environ. 718 (2020).

[11] G. Camps-Valls, D. Tuia, L. Bruzzone, J.A. Benediktsson, Advances in hyperspectral image classification, IEEE Signal Process. Mag. 31 (2014) 45–54.

[12] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444.

[13] Y. Li, W. Chen, Y. Zhang, C. Tao, R. Xiao, Y. Tan, Accurate cloud detection in high-resolution remote sensing imagery by weakly supervised deep learning, Remote Sens. Environ. 250 (2020) 112045.

[14] J. Qi, C. Tao, H. Wang, Y. Tang, Z. Cui, Spatial information inference net: road extraction using road-specific contextual information, ISPRS J. Photogramm. Remote Sens. 158 (2019) 155–166.

[15] Y. Li, Y. Zhang, Z. Zhu, Error-tolerant deep learning for remote sensing image scene classification, IEEE Trans. Cybern. 51 (2021) 1756–1768.

[16] Y. Li, J. Ma, Y. Zhang, Image retrieval from remote sensing big data: A survey, Inf. Fusion 67 (2021) 94–115.

[17] Y. Li, Y. Zhang, X. Huang, J. Ma, Learning source-invariant deep hashing convolutional neural networks for cross-source remote sensing image retrieval, IEEE Trans. Geosci. Remote Sens. 56 (11) (2018) 6521–6536.

[18] X. Zhu, T. Devis, L. Mou, Deep learning in remote sensing: A comprehensive review and list of resources, IEEE Geosci. Remote Sens. Mag. 5 (4) (2017) 8–36.

[19] Y. Li, T. Shi, Y. Zhang, W. Chen, Z. Wang, H. Li, Learning deep semantic segmentation network under multiple weakly-supervised constraints for cross-domain remote sensing image semantic segmentation, ISPRS J. Photogramm. Remote Sens. 175 (2021) 20–33.

[20] D. Doran, S. Schulz, T.R. Besold, What does explainable AI really mean? A new conceptualization of perspectives, 2017, eprint arXiv:1710.00794.

[21] M. Alirezaie, M. Längkvist, M. Sioutis, Semantic referee: A neural-symbolic framework for enhancing geospatial semantic segmentation, Semant. Web 10 (2019) 863–880.

[22] D. Arvor, M. Belgiu, Z. Falomir, I. Mougenot, L. Durieux, Ontologies to interpret remote sensing images: why do we need them? GIsci, Remote Sens. (2019) 1–29.

[23] H. Couclelis, Ontologies of geographic information, Int. J. Geogr. Inf. Sci. 24 (12) (2010) 1785–1809.

[24] M.K. Sarker, N. Xie, D. Doran, Explaining trained neural networks with semantic web technologies: First steps, 2017, eprint CoRR abs/1710.04324.

[25] J. Chen, I. Dowman, S. Li, Information from imagery: ISPRS scientific vision and research agenda, ISPRS J. Photogramm. Remote Sens. 115 (2015) 3–21.

[26] N. Moran, S. Nieland, G.T. Suntrup, B. Kleinschmit, Combining machine learning and ontological data handling for multi-source classification of Nature Conservation Areas, Int. J. Appl. Earth Obs. Geoinf. 54 (2017) 124–133.

[27] M. Wurm, T. Stark, X.X. Zhu, M. Weigand, H. Taubenbock, Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks, ISPRS J. Photogramm. Remote Sens. 150 (2019) 59–69.

[28] J. Sherrah, Fully convolutional networks for dense semantic labelling of high-resolution aerial imagery, 2016, eprint arXiv:1606.02585.

[29] E. Maggiori, Y. Tarabalka, G. Charpiat, P. Alliez, Fully convolutional neural networks for remote sensing image classification, in: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Beijing, China, 2016, pp. 5071–5074.

[30] G. Hongmin, Y. Yao, L. Sheng, Multi-branch fusion network for hyperspectral image classification, Knowl.-Based Syst. 167 (2019) 11–25.

[31] E. Basaeed, H. Bhaskar, P. Hill, A supervised hierarchical segmentation of remote sensing images using a committee of multi-scale convolutional neural networks, Int. J. Remote Sens. 37 (7) (2016) 1671–1691.

[32] M. Langkvist, A. Kiselev, M. Alirezaie, Classification and segmentation of satellite ortho imagery using convolutional neural networks, Remote Sens. 8 (4) (2016) 1–21.

[33] N. Audebert, B.L. Saux, S. Lefèvre, Semantic segmentation of earth observation data using multimodal and multi-scale deep networks, 2016, eprint arXiv:1609.06846.

[34] E. Maggiori, Y. Tarabalka, G. Charpiat, High-resolution semantic labeling with convolutional neural networks, 2016, eprint arXiv:1611.01962.

[35] M. Kampffmeyer, A.B. Salberg, 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 (CVPR) Workshops, 2016.

[36] Z. Deng, H. Sun, S. Zhou, J. Zhao, L. Lei, H. Zou, Multi-scale object detection in remote sensing imagery with convolutional neural networks, ISPRS J. Photogramm. Remote Sens. 145 (2018) 3–22.

[37] P. Ding, Y. Zhang, W. Deng, P. Jia, A. Kuijper, A light and faster regional convolutional neural network for object detection in optical remote sensing images, ISPRS J. Photogramm. Remote Sens. 141 (2018) 208–218.

[38] D. Arvor, L. Durieux, S. Andrés, Advances in geographic object-based image analysis with ontologies: A review of main contributions and limitations from a remote sensing perspective, ISPRS J. Photogramm. Remote Sens. 82 (2013) 125–137.

[39] M. Codescu, H. Gregor, OSMonto-an ontology of open street map tags, in: Proceedings of the State of the Map Europe (SOTM-EU) Conference, 2011.

[40] R. Gui, X. Xin, H. Dong, Individual building extraction from terrasar-x images based on ontological semantic analysis, Remote Sens. 8 (9) (2016) 708.

[41] S. Andrés, D. Arvor, I. Mougenot, Ontology-based classification of remote sensing images using spectral rules, Comput. Geosci. 102 (2017) 158–166.

[42] A. Khitem, F. Mohamed, Graph of concepts for semantic annotation of remotely sensed images based on direct neighbors in RAG, Can. J. Remote Sens. 44 (6) (2018) 551–574.

[43] H. Gu, H. Li, L. Yan, Z. Liu, An object-based semantic classification method for high resolution remote sensing imagery using ontology, Remote Sens. 9 (4) (2017).

[44] T.R. Besold, A.S. Garcez, S. Bader, Neural symbolic learning and reasoning: A survey and interpretation, 2017, eprint arXiv:1711.03902.

[45] A. Kolesnikow, C. Lampert, Seed, expand and constrain: three principles for weakly-supervised image segmentation, in: Proceedings of the 14th European Conference on Computer Vision (ECCV), Springer, Amsterdam, The Netherlands, 2016, pp. 695–711.

[46] G.A. Nys, J.P. Kasprzyk, P. Hallot, R. Billen, Towards an ontology for the structuring of remote sensing operations shared by different processing chains, in: Proceedings of Ontologies, Semantics, and Knowledge Representation for Geospatial Information, 2018.

[47] B. Nasri, H. Nefzi, M. Farah, Towards a hybrid approach for remote sensing ontology construction, in: Proceedings of the 4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Sousse, Tunisia, 2018.

[48] Z. Shao, K. Yang, W. Zhou, B. Hu, Performance evaluation of single-label and multi-label remote sensing image retrieval using a dense labeling dataset, Remote Sens. 10 (2018) 964.

[49] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in: Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention, Munich, Germany, 2015, pp. 234–241.

[50] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, Comput. Sci. (2014).

[51] R. Achanta, A. Shaji, K. Smith, Slic: Superpixels compared to state-of-the-art superpixel methods, IEEE Trans. Pattern Anal. Mach. Intell. 34 (2012) 2274–2282.

[52] A. Garcia-Garcia, S. Orts-Escolano, S.O. Oprea, V. Villena-Martinez, J. Garcia-Rodriguez, A review on deep learning techniques applied to semantic segmentation, 2017, eprint arXiv:1704.06857v1.