Received February 21, 2018, accepted March 25, 2018, date of publication April 5, 2018, date of current version May 24, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2823501

Intelligent Map Reader: A Framework for Topographic Map Understanding With Deep Learning and Gazetteer

HUALI LI[1], JUN LIU[2], AND XIRAN ZHOU[3], (Student Member, IEEE)

[1]College of Electrical and Information Engineering, Hunan University, Hunan 410082, China
[2]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[3]Arizona State University, Tempe, AZ 85281, USA

Corresponding author: Xiran Zhou (xrzhou@asu.edu)

This work was supported in part by the National Natural Science Foundation of China under Grant 61301255, Grant 41271432, Grant 41471340, and Grant 41301403, in part by the China Postdoctoral Science Special Foundation under Grant 2014T70768, in part by the National Key Research and Development Program of China under Grant 2017YFB0504200, in part by the Hunan Provincial Natural Science Foundation of China under Grant S2013J504B, and in part by the Basic Research Program of Shenzhen under Grant JCYJ20150831194441446.

ABSTRACT Text features in topographic maps are important for helping users to locate the area that a map covers and to understand the map's content. Previous works on the optical detection of map text from topographic maps have used geometric features, the Hough transform, and segmentation. However, these approaches still face challenges when detecting map text in complicated contexts, especially when the map text is touching other map features, such as contours or geographical features. Thus, state-of-the-art techniques for map text and feature recognition and manual interpretation and correction are always required to produce accurate results when optically converting topographic maps into a readable format. This paper proposes a methodological framework called the intelligent map reader that enables the automatic and accurate optical understanding of the content of a topographic map using deep learning techniques in combination with a gazetteer. The intelligent map reader framework includes the detection of map text via deep learning, the separation of text units via graph-based segmentation and clustering, optical character recognition (OCR) via an OCR engine, and digital-gazetteer-based map content understanding. Experimental results validate the efficiency and robustness of our proposed methodology for map text recognition and map content understanding. We expect the proposed intelligent map reader to contribute to various applications in the GeoAI field.

INDEX TERMS Optical character recognition, deep convolutional neural network, map feature detection, gazetteer, topographic map understanding.

I. INTRODUCTION

Topographic opographic maps are a commonly used form of data for characterizing the physical origins of natural landforms, the locations where various human activities occur, and how these natural and manmade (or cultural) elements are organized. Due to the rapid progress achieved in Earth observation systems and GIS techniques, massive topographic maps related to various topics and themes are now available for public use. These topographic maps can be accessed from a variety of sources, which include atlases, scanned paper maps, commercial map services (e.g., Google Earth, Google Maps), repositories of volunteered geographical information (e.g., OpenStreetMap), and geo-referenced cyberinfrastructures. The U.S. Geological Survey provides the largest topographic maps, which are generated from a digital GIS database and cover the entire territory of the U.S. In addition to the U.S. Geological Survey, other institutes also make topographic maps available to the public, including the SVG Topographic maps of Europe, the Great Britain AMS Topographic Maps, and the maps available through Tianditu. These digital topographic maps contain text information, such as river names, street names, and place names, that is important for helping users to locate the area that a map covers and to understand its content. However, further investigations are still urgently needed to build an automated and robust approach for recognizing such map text from topographic maps. A crucial challenge in map text recognition is that the text features of many topographic maps are not directly readable by computers [1]–[3]. Currently, most topographic maps are accessed as Portable Document Format (PDF) files, digital images, or paper maps rather than as files that include machine-readable text. The laborious conversion from scanned or digital topographic maps into data in a readable format is very time-consuming, and it can become impossible when large-scale maps require conversion and interpretation [3]. Moreover, the state-of-the-art techniques for map text and feature recognition still require manual interpretation and correction to produce accurate results.

Map text recognition consists of two main tasks [1], [3], [4]: text units separation and text recognition. The main goal of text unit separation is to extract text units from the complicated map background, which may include colour mixing, linear features, overlapping, aliasing, and blurring [5]. Early efforts to separate text units from maps included clustering analysis [6], [7], morphological operations [8], [9], segmentation [10]–[12], connected component labelling [6], [13], and the image pyramid method [14]. However, these approaches do not enable the extraction of map text displayed at multiple angles. Moreover, it is impossible for these approaches to separate text units when they are overlapped or connected to other features in a topographic map [2], [3]. Reference [15] presents an interactive tool for text extraction that allows the user to define rules for colour separation and character clustering. However, such an approach that requires user intervention cannot support text recognition in massive topographic maps that cover large-scale areas. Several attempts to address the problem of separating text and other graphical features in topographic maps have been reported. In reference [16], map text and other map features were separated based on the differences in their constituent strokes. Tembre *et al.* [17] proposed an approach for locating overlapping text by searching for seed strings. Zhong [18] proposed a circle scanning strategy for detecting character-connected regions. However, these two methods fail when all letters in a string intersect with other graphical features. A method called V-lines [19] can be used to remove some non-text features connected to map text. Roy *et al.* [20] claimed that the text units and linear features in topographic maps have distinct widths, which could be useful for separating text from other connected features. Considering the advantages of object-based image analysis (OBIA) in handling remote sensing data, the application of segmentation to topographic maps has been reported. Pouderoux *et al.* [21] extracted text through segmenting topographic maps and performing connected component analysis. Kerle and Leeuw [22] used OBIA techniques to extract map text from sub-regions identified through object-based segmentation. Segmentation on multiple scales might also be useful for resolving the intersections of text with other graphical features. Pezeshk and Tutwiler [3] integrated two hidden Markov models (HMMs) to recognize map text based on its features in the horizontal and vertical dimensions. However, for the case of text oriented in other directions, this method requires further processing and improvements. Overall, the development of an approach that can enable the automatic detection of the positions of text units from the complicated background of topographic maps remains an open problem.

Once text units are available as a result of text unit separation, the task of text recognition consists of text localization and text classification [2], [3], [16], [12]. A direct and simple approach is to exploit optical character recognition (OCR) engines, as reported in many previous works on topographic map text recognition [15], [16], [21]. In particular, the emergence of deep learning approaches has greatly facilitated OCR tasks [24]–[27]. To address the existing challenges facing the recognition of map text from optical topographic maps, this paper proposes a methodological framework that combines deep learning techniques with the use of a gazetteer to enable the automatic and accurate understanding of the content of an optical topographic map. The remainder of this paper is organized as follows. Section 2 discusses previous works related to the topic of this paper. Section 3 introduces the details of the proposed methodological framework, which includes map text detection via deep learning, text unit separation via graph-based segmentation and clustering, OCR via an OCR engine, and digital-gazetteer-based map content understanding. Section 6 presents experimental results obtained for twenty topographic maps. Section 7 summarizes the contributions of this paper and the prospects of automatic map text detection.

II. RELATED WORK

A. FASTER RCNN

As an approach designed for localization and classification based on a cutting-edge deep convolutional neural network (DCNN), Faster R-CNN has become the state of the art in object detection according to the results obtained the PASCAL VOC 2007 and 2012 Test Sets [28]. Faster R-CNN is a deep convolutional neural network developed based on the architecture of fast R-CNN by adding RPN. RPN is a novel fully convolutional network, which refers to a key component of Faster RCNN. Compared with two similar approaches, namely, Region CNN (R-CNN) and Fast R-CNN, Faster R-CNN more efficiently addresses three main challenges of object detection: 1) the prediction of region proposals at multiple scales and ratios, 2) training on regional proposals for object detection, and 3) the unification of the Region Proposal Network (RPN) approach with the convolutional network (Fast R-CNN) approach. Faster R-CNN has been applied for a number of object detection tasks, such as face detection [29], individual recognition [30], target detection from satellite images [31], and urban scene analysis [32].

Figure 1 shows the architecture of Faster R-CNN [28], which consists of two modules: an RPN and a Fast R-CNN detection network. Both networks share a convolutional network, which is labelled as "conv layers" in Figure 1. This convolutional network uses a DCNN approach (such as

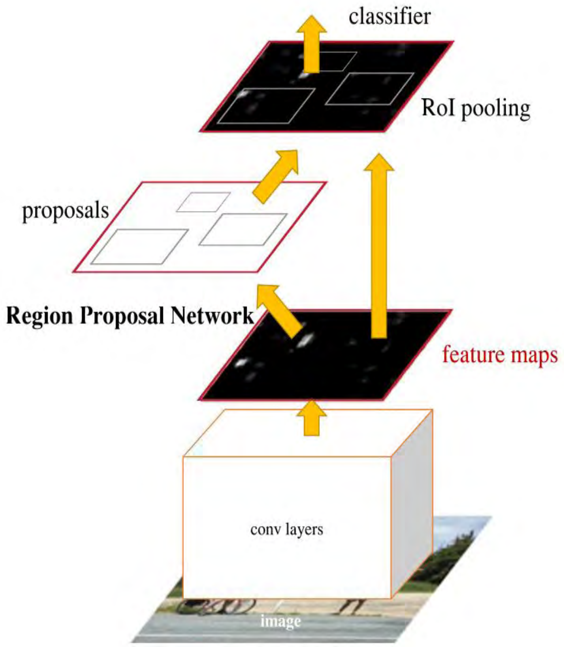


FIGURE 1. Architecture of Faster R-CNN [28].

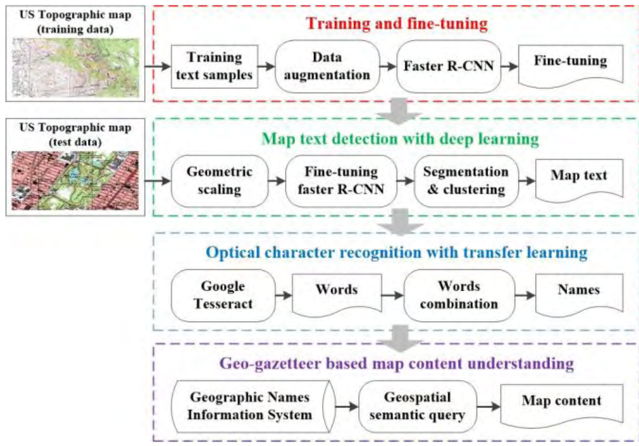


FIGURE 2. Architecture of the proposed methodological framework.

ZF [33] or VGG-16 [34]) to produce feature maps from the original image.

Then, based on the feature maps, the RPN generates region proposals that may contain target objects. It enables end-to-end training to achieve an appropriate trade-off between the accuracy and computational efficiency with which the region proposals are generated. Moreover, in Faster R-CNN, anchor boxes and aspect ratios are used to optimize the positions and shapes of the region proposals. An anchor box will be selected as a positive sample of a region proposal if it has the highest Intersection-over-Union (IoU) overlap with the ground-truth box. The loss function is defined as follows [28]:

$$F_{loss}(P_i, B_i) = \frac{1}{N_{dim}} \sum_i L_{dim}(P_i, WP_i) + \lambda \frac{1}{N_{reg}} \sum_i WP_i \cdot L_{reg}(B_i, WB_i) \quad (1)$$

where i denotes the index of the anchor box and P_i denotes the probability that anchor box i will be predicted to be an object. WP_i is equal to 1 when anchor box i is a positive sample; otherwise, WP_i is equal to 0. N_{dim} and N_{reg} denote the mini-batch size and the total number of anchor boxes, respectively. B_i and WB_i are the vectors of the four corner coordinates of the predicted bounding box and the ground-truth bounding box, respectively. $L_{dim}()$ is the logistic loss between the object class and the non-object class, and $L_{reg}(B_i, WB_i)$ is a robust loss function, the details of which can be found in Reference [35].

Finally, the Fast R-CNN detection network is used to determine whether a region proposal belongs to any one of a set of predefined categories. In Figure 1, region of interest (ROI) pooling is a neural network layer for determining the probability of each region proposal being a predefined category. Since the RPN and the Fast R-CNN each fine-tune the architecture of the convolutional layer, an alternating training strategy is used to enable feature sharing between two convolutional layers. This alternating training strategy serves as the basis for an iterative fine-tuning process that alternates between the RPN and the Fast R-CNN.

B. GOOGLE TESSERACT OCR ENGINE

The success of deep learning techniques in OCR raises the potential for achieving automated text recognition for optical topographic maps. However, several characteristics of map text hinder the application of OCR for map text recognition. First, most OCR models and engines are designed to recognize text that is unaffected by noise, which is never the case in topographic maps. Moreover, unlike the text in typical text documents, the text units in topographic maps are typically rotated at a variety of angles. Most state-of-the-art OCR models and engines are incapable of recognizing such oblique text.

The Google Tesseract OCR engine is an open-source engine that supports the recognition of text information from binary images. It was first developed by Google in 2006 [36], [37]. The Tesseract OCR engine follows a process that includes text line detection, word extraction, and word recognition. The process of text line detection and word extraction consists of three steps: 1) the detection of text lines via blob filtering and line construction, 2) the fitting of baselines with quadratic splines, and 3) the division of each line of text into characters. The word recognition process in Tesseract consists of the classification of characters derived from the detected lines of text using an adaptive classifier. Additional operations are also implemented to handle jointed characters and broken characters.

Version 2 of Tesseract enables the recognition of text in more than 100 languages and supports a variety of formats, including images and plain text. Moreover, Tesseract also offers an interface for conducting further training when

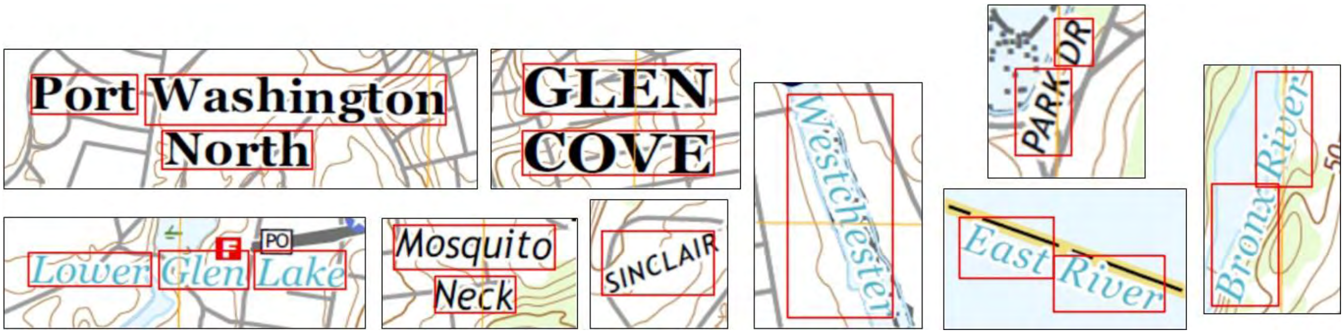


FIGURE 3. Illustrations of training text samples.

additional training text data become available. More details can be found at https://github.com/tesseract-ocr/tesseract. As one of the most accurate OCR engines currently available [37], Tesseract OCR has been used in a number of applications for natural language processing (NLP), text recognition and document analysis [38]–[40].

III. METHODOLOGY FRAMEWORK

A. FRAMEWORK OF THE INTELLIGENT TOPOGRAPHIC MAP READER

Figure 2 illustrates the proposed methodological framework. The first part of the framework is designed to fine-tune the Faster R-CNN architecture based on training samples oriented at different angles and printed by different colors. from different backgrounds, which are manually selected from the optical topographic map. Moreover, data augmentation is used to increase the diversity of the training samples and further boost the performance of the neural network. In the second part of the framework, based on the fine-tuned Faster R-CNN architecture, the regions of the topographic map that contain map text are detected, and the text units are then separated from other graphical features via graph-based segmentation [41] and DBSCAN clustering [42], [43]. In the third part of the framework, we use a pre-trained OCR engine, namely, Google Tesseract, to identify map text as geographical names. In the last part of the framework, the Geographical Names Information System (GNIS), which was developed by the United States Geological Survey (USGS) as a standard repository of domestic geographical name data [44], is accessed via geospatial semantic queries to determine the content and coverage of the topographic map. The details of each part of the framework are introduced in the following subsections.

B. TRAINING DATA SELECTION AND DATA AUGMENTATION

To our knowledge, text datasets for topographic maps are not yet available. Thus, we first selected text samples for training from several optical topographic maps. These topographic maps can be accessed via the portal for the US Topo dataset on the National Map website: https://nationalmap.gov/ustopo/index.html. Additionally, since the dimensions of the original topographic maps are too large for efficient training, we partitioned each entire topographic map into sub-images with dimensions of 600∗600. Then, we used a tool named LabelImg [45] to label the positions of words and terms. Figure 2 illustrates some examples of our selected training samples.

The total number of labelled samples was 10200. Then, we conducted data augmentation on each labelled sample. The data augmentation process included rotations and random cropping. Intensity stretching and modification were not considered since the colours assigned to the map text are fixed. After data augmentation, we had a total of 1673 samples from 12 topographic maps with which to complete the training process.

C. MAP TEXT DETECTION WITH FASTER R-CNN

In this paper, each map text unit is regarded as an object in a topographic map, and Faster R-CNN is exploited for the localization and recognition of topographic map text units. The DCNN used in this study was a pretrained VGGNet model. In the first module, the key parameters include the dimensionality of the anchor boxes, the aspect ratios and the IoU in the loss function. The details of the anchor boxes, aspect ratios, and IoU can be found in [28]. Several previous papers have also discussed the influences of the box sizes and aspect ratios in localization with Faster R-CNN [29], [46], [47]. In the original Faster R-CNN, the anchor box sizes are 128∗128, 256∗256 and 512∗512, and the aspect ratios are 1:1, 1:2 and 2:1. Considering the typical geometric shapes, scales and positions of map text units, we adopt three box sizes: 64∗64, 128∗128, and 256∗256. This means that the anchor scales are 4, 8 and 16, respectively. The box size of 512∗512 is discarded because the sizes of map text units are always relatively small compared with the entire topographic map.

In addition to the changes in box sizes, we adopt eight aspect ratios: 1:4, 2:5, 1:2, 1:3, 2:1, 5:2, 3:1, and 4:1. The original aspect ratio of 1:1 is discarded because single letters are rarely observed in digital maps and because the minimal bounding box of most map text units has a long and narrow geometric shape. We also tested other aspect ratios of the forms 1:$*$ and $*$:1 ($* \geq 5$) with respect to the anchor boxes; however, we observed that including these aspect ratios yielded little improvement in the accuracy of the generated region proposals while significantly increasing the computational load. Moreover, in this study, a positive label was assigned to an anchor box with an IoU overlap greater than 0.7, which is similar to the value adopted in the original model, as reported in [28].

After obtaining the region proposals with the RPN, we apply the alternating training strategy to complete the sharing of the convolutional layers between the Fast R-CNN and the RPN. The Fast R-CNN is pretrained with ImageNet, the details of which can be found in [35].

Figure 4 illustrates an example of the results generated by this stage of the proposed intelligent map reader. Figure 4(A) shows the original topographic map. Specifically, the map name and map extent are not available here. Figure 4(B) shows the detection results on the entire map. The red boxes indicate correct detection results, whereas the purple and green boxes indicate false negative results and false positive results, respectively. Only one word, ‘‘Ramanda’’, is not completely detected and is instead detected as two separate text units: ‘‘Raman’’ and ‘‘a’’. The false negative results include US Route numbers, which are mistakenly classified as map texts units rather than as components of the US Route symbols in the topographic map.

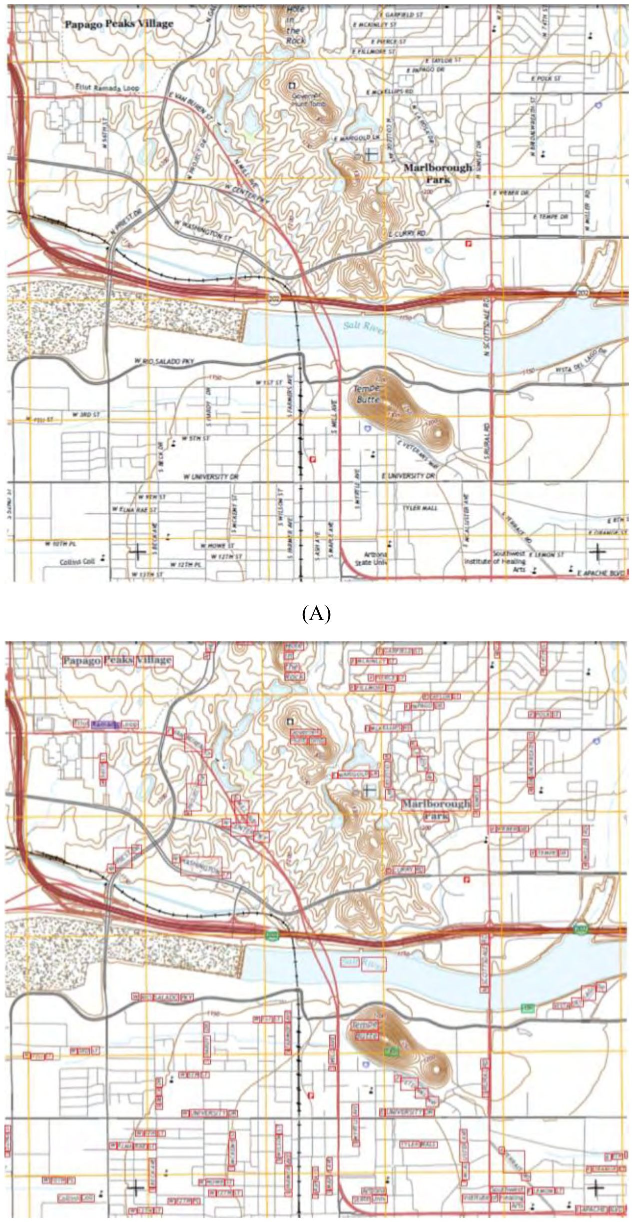
D. TEXT UNIT SEPARATION WITH SEGMENTATION AND CLUSTERING

For the identification of text information, the text units in the detection results should be separated from the map background and map noises. We propose an approach that combines graph-based segmentation and spectral unmixing. Considering that the colours and shapes of text units are generally distinct from the background in topographic maps, it is efficient to perform graph-based segmentation by creating a graph to represent the relationships among the meaningful image regions throughout the entire image. Moreover, most of noises have smaller area and a distance to text units in the result of map text detection. Thus, these noises can be removed based on the area of segmented sub-regions. Such a graph-based segmentation for the detection area can be represented as follows:

$$G = (V, E) \tag{2}$$

where G, V and E denote the graph (or detection area), the set of nodes in the graph, and the set of undirected edges that connect nodes in the graph. We also define $w_{i,j}$ as the weight of the edge between nodes v_i and v_j.

The graph-based segmentation partitions G into regions $g_k = (V, E_k)$, where $E_k \subset E$ and k is the index of the segmented regions.



(A)

(B)

FIGURE 4. Illustration of the detection results. (A) Original topographic map. (B) Detection results.

For each pair of segmented regions, a function called $f()$ is used to predict whether there is a boundary between them:

$$f(E_{k1}, E_{k2}) = \begin{cases} true, & if\ dif\ (E_{k1}, E_{k2}) > int\_dif(E_{k1}, E_{k2}) \\ false, & otherwise \end{cases} \tag{3}$$

where $dif\ (E_{k1}, E_{k2})$ is the difference between E_{k1} and E_{k2} and is calculated as follows:

$$dif\ (E_{k1}, E_{k2}) = \min\left(w_{i,j}\right), v_i \in E_{k1}, v_j \in E_{k2} \tag{4}$$

Another element in Equation (2), $int\_dif(E_{k1}, E_{k2})$, is the internal difference between E_{k1} and E_{k2} and is calculated

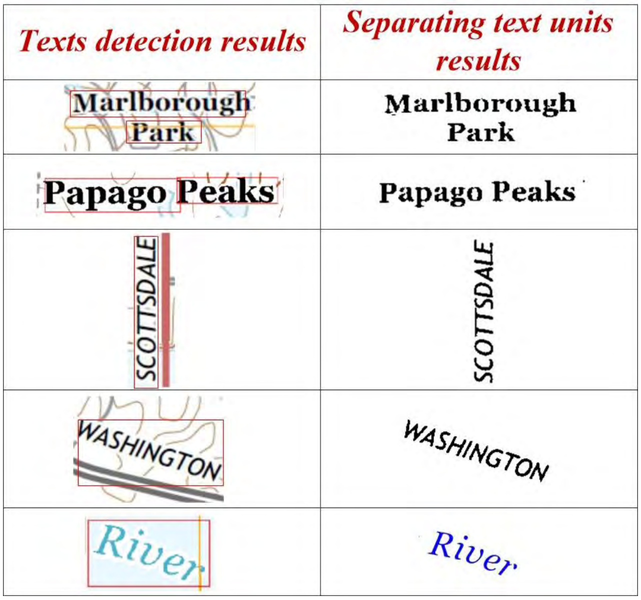


FIGURE 5. Illustration of text units separation.

using the following equation:

$$\begin{cases} true, \ if \ dif\,(E_{k1}, E_{k2}) > int\_dif\,(E_{k1}, E_{k2}) \\ false, \ otherwise \end{cases} \tag{5}$$

We assume that the image regions after graph-based segmentation are R_x, $x \in N$, where N is the total number of segmented image regions.

Then, we cluster all segmented regions into multiple meaningful groups. First, the HSV colour space is normalized as a new three-dimensional space. Second, we use DBSCAN to perform clustering on this new three-dimensional space. The DBSCAN clustering process depends on two parameters, *eps* and *minPts*. *eps* defines the maximal distances in terms of the hue, saturation and value for which two samples can be considered part of the same group. *minPts* defines the minimal number of samples that must be included in the neighbourhood for a sample to be considered a core point. Finally, we select the group that is closest to the map text in the colour space. We illustrate some separation results for detected map texts in Figure 5.

E. MAP TEXT IDENTIFICATION WITH GOOGLE TESSERACT OCR ENGINE

After text detection and text unit separation, we have an image that includes map text alone. However, in practice, the OCR results of Google Tesseract are sensitive to rotation and irrelevant noise. In Subsection 3.4, we described the removal of irrelevant noise with our proposed method for text unit separation. Additionally, to address oblique text, we propose an approach based on the creation of a minimal bounding box (MBB), as shown in Figure 6.

Figure 6(A) illustrates an oblique text unit derived from a topographic map. The blue and green rectangles represent this oblique text unit's MBB and its rotated MBB, respectively.

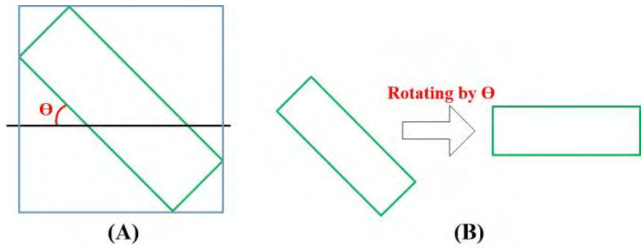


FIGURE 6. The proposed approach for addressing oblique text. (A) Illustration of the proposed approach. (B) Examples of rotation.

Θ denotes the angle between the horizontal dimension and the long edge of the rotated MBB. Using the following equation, we rotate the oblique text unit into an upright orientation.

$$[x_{new}, y_{new}, z]^T = \begin{bmatrix} cos\theta, -sina\theta, 0 \\ sin\theta, cos\theta, 0 \\ 0, 0, 1 \end{bmatrix} \bigotimes [x, y, z]^T \tag{6}$$

where x, y and z denote the horizontal coordinate, the vertical coordinate and the intensity of the corresponding pixel in the original bounding box, respectively, whereas x_{new} and y_{new} denote the horizontal and vertical coordinates, respectively, of a pixel in the bounding box after rotation. Generally, in a topographic map, the angle of each character in the same place name may be slightly different. For example, each character of the word "River" in Figure 3 is placed at a different angle. We ignore this issue since in practice, the text will be sufficiently close to upright after rotation for correct identification by Tesseract OCR.

Then, we exploit the Google Tesseract OCR engine to identify the detected map text units. The identification results are machine-readable text units that can be subjected to natural language processing (NLP) and imported into a database. Based on the example text units shown in Figure 6, the results of map text identification are given in the column titled "OCR results" in Figure 7.

F. DIGITAL GAZETTEER-BASED MAP CONTENT UNDERSTANDING

Digital gazetteer is a tool for organizing knowledge and information about named places for the standardization of place names and spellings, or toponyms [48]. Due to the growing number of data acquisition and cartography techniques, conflations of place names are commonly observed in digital maps generated at different times, with different focuses, or with other differences [49]. Thus, gazetteers that enable the searching and retrieval of place names are receiving increasing attention from government agencies, educational institutions, and commercial enterprises. Generally, the information and knowledge included in a gazetteer consist of [48] place names, geographic locations, place descriptions, and geographical features and/or place types.

A topographic map is regarded as an important type of map that characterizes the details of the terrestrial surface over a large area. To facilitate the use of topographic maps, the U.S.

FIGURE 7. Illustration of map text identification with the Google Tesseract OCR engineer.

Board on Geographic Names has established an administrative gazetteer service called the GNIS to standardize the places named in topographic maps. In GNIS, all place names and geographic features printed in topographic maps have been semantically organized and defined. Moreover, names in the GNIS are organized as a hierarchy, which might be useful to support semantic queries. The impacts of the GNIS on information sharing and management have been reported in a previous study [48]–[50]. In the framework proposed in this paper, the contents of the map text units recognized from topographic maps are used as keywords to access relevant geo-information and knowledge from the GNIS (https://geonames.usgs.gov/apex/f?p = 138:1:11458161770145). The useful information for automatic map understanding that is available by querying the GNIS is listed as follows:

- Basic information, including IDs, place names, place types, citations, entry data, and elevations.
- Variant names, including all names by which a place is or has been known, regardless of whether this name is used currently.
- County information, including the area in which a named place is, such as the county name, county code, state name, state code, and country name.
- Geographical coordinates, including the latitude and longitude of a place.

Figure 8 illustrates how the information accessed from the GNIS supports map understanding. Figure 8(A) lists the information obtained by accessing the GNIS with respect to the detected map text. Figure 8(B) shows the contents of this topographic map, which are organized in a hierarchy based on the contents of the map text recognized from the topographic map. The text phrases enclosed in purple rectangles are those detected from the topographic map, the text phrases enclosed in blue rectangles are those obtained through reasoning based on the detected text, and the text phrases enclosed in grey rectangles are the detected text for which no descriptions or attributes are available in the GNIS. Based on the names Tempe Butte and Arizona State University, we can deduce that this map covers part of Tempe. Additionally, based on the names Marlborough Park, Papago Peak Village, and others, we can deduce that this map also covers part of Scottsdale. Finally, by combining the names Tempe, Scottsdale and Salt River, we can conclude that the coverage of this map includes the southern areas of Scottsdale and the northern areas of Tempe where it touches the Salt River. Additionally, we can further estimate the coverage of this map based on the geographical coordinates of each detected feature (or map text unit).

However, the information accessed from the GNIS still seems inadequate to support semantic understanding based on timely updates of geographical information. Previous works [51]–[53] have reported efforts to integrate ontologies and gazetteers to generate geospatial knowledge. Information accessed from the GNIS is useful to facilitate knowledge discovery through the integration of ontologies developed based on geographical context, such as the Global Change Master Directory [54], USTopographic [55], SWEET [56], and other geographical domain ontologies [57]. Based on the logical reasoning offered by ontologies, the GNIS enabled information queries can be extended to semantic queries by means of geoSPARQL [58]. In addition to geo-semantic reasoning and ontologies, we can also use linked geodatabases for knowledge discovery. Linked geodatabases provide massive amounts of spatial data in a variety of contexts, which is crucial for enabling geo-semantic queries. Finally, the emergence of volunteered geographic information [59] offers an increased potential to integrate detection results for map text (place names) from topographic maps with other geographical information, such as population data and disaster data. The information contained in volunteered data from social media (e.g., Twitter, Facebook) and digital map tagging (e.g., Google Maps) carries great potential to facilitate the understanding of topographic map content.

IV. EXPERIMENTS

A. RESULT OF MAP TEXT DETECTION

As mentioned in Subsection 3.2, 1673 samples selected from 12 topographic maps were used to train the Faster R-CNN architecture to support map text recognition. In this experiment, an Nvidia TITAN X GPU was used for training and testing. Then, we use the fine-tuned Faster R-CNN to detect the map text from twenty topographic maps, which were selected from different states in the U.S. and included urban and rural scenes, mountains, coasts, lakes, rivers, and other features. The differing scenes meant that the contents and backgrounds of these twenty topographic maps varied significantly. Table 1 lists the names of the topographic maps used

Machine-readable text after conversion	Information from GNIS						
	Class	County	State	Country	Map name	Elevation	Geographical coordinates
Tempe butte	Summit	Maricopa	Arizona	U.S.	Tempe	...	...
Arizona State University	School	Maricopa	Arizona	U.S.	Tempe	...	...
Marlborough Park	Populated Place	Maricopa	Arizona	U.S.	Tempe	...	...
Salt River	Stream	Maricopa	Arizona	U.S.	Mesa Tolleson	...	...
Papago Peak Village	Populated Place	Maricopa	Arizona	U.S.	Tempe	...	...
Governor Hunt Tomb	Cemetery	Maricopa	Arizona	U.S.	Tempe	...	...
Hole in the Rock	Arch	Maricopa Coconino	Arizona	U.S.	Tempe	...	...
...	...	...	...	...	...	...	...

(A)

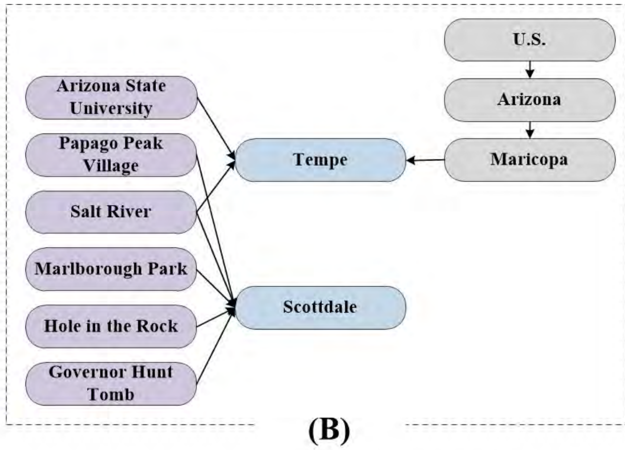


FIGURE 8. Illustration of accessing the GNIS for map understanding. (A) Information regarding detected map text accessed from the GNIS. (B) Relationships created based on GNIS information.

in the experiment and their dimensionalities, which ranged from 8900×11400 to 9100×11600.

Due to page limitations, we selected one these maps to present the results of map text detection, text unit separation, and text recognition. This map is the USGS USTopo 7.5-minute map for Onego, WV 2016. Figure 9(A) shows the detection results for the entire map. In Figure 9(A), the red boxes indicate the correctly detected text units, or the true positive results. False positive and false negative results are indicated by purple and green boxes, respectively. Three text units are detected incorrectly; these three text units are components of the U.S. Route symbol and indicate the numbers of U.S. Routes. The only false positive result is for a place named "ENTERPRISE", for which the text is curved. The bounding box produced by the Faster R-CNN only partially covers the entire word because of its length. Other than this place name, all names in this topographic map are successfully detected. Figure 9(B) shows the details of example detection results selected from all twenty topographic maps. For each detected text unit, the red box represents the MBB, which contains the majority of the characters of the map text in each case. Detection with the Faster R-CNN architecture is robust to variations in colour and texture, as seen from the fact that the detection results include map text units of different colours and textures. Moreover, the detected map text units shown in Figure 9(B) are oriented at different angles, demonstrating robustness to rotational variations. Finally, although these map text units touch other map features, such as contours, map lines, and water bodies, the Faster R-CNN architecture has the ability to distinguish true text features from a variety of background features. Thus, the results shown in Figure 9(A) and Figure 9(B) validate the capability of the Faster R-CNN approach for detecting map text from complicated backgrounds.

Moreover, the evaluations of the detection results for the other topographic maps are given in Table 1. We provide the average precision (AP) of the detection of map text from each of the twenty topographic maps, the details of which are listed in Table 1. Topographic maps of urban scenes contain significantly more map text. Consequently, the AP values for map text recognition that are obtained on topographic maps of urban scenes are lower than those achieved for topographic maps of natural or rural scenes. Generally, the AP values for map text recognition range between 87% and 94%, thereby proving the stability of the Faster R-CNN method in producing accurate detection results for map text in topographic maps.

B. RESULT OF TEXT UNIT SEPARATION

Now, we present the results of using our proposed approach to separate the text in the detected text units from the background and other geographic features depicted in the topographic maps. Figure 10 shows selected separation results for text units from all twenty topographic maps. In Figure 10, the original data are the detection results, and the separated results are the results obtained after text unit separation. The text to be separated had a variety of characteristics and appeared against complicated backgrounds. First, these lines of text were printed in different colours, including black and blue, and some even included attached shadows (e.g., CHELAN). Additionally, they were placed at different angles. For example, the place name "ENTERPRISE" was arranged on a curve. Finally, the greatest challenge was to separate map text from complicated background contexts, especially for text touching other map features, such as "Columbia", "Smith", and "SKYLINE". It can be observed that all detected map texts were successfully separated from a variety of backgrounds and other map features.

Before using the Tesseract OCR engine to identify the detected map text units, we applied the proposed approach for straightening oblique map text, the results of which are shown in Figure 11. In Figure 11, the original data are the map text units at different angles obtained after text unit separation. The map text samples in Figure 11 have diverse angles, as is typical of topographic maps. The straightening results confirm that the proposed approach can effectively

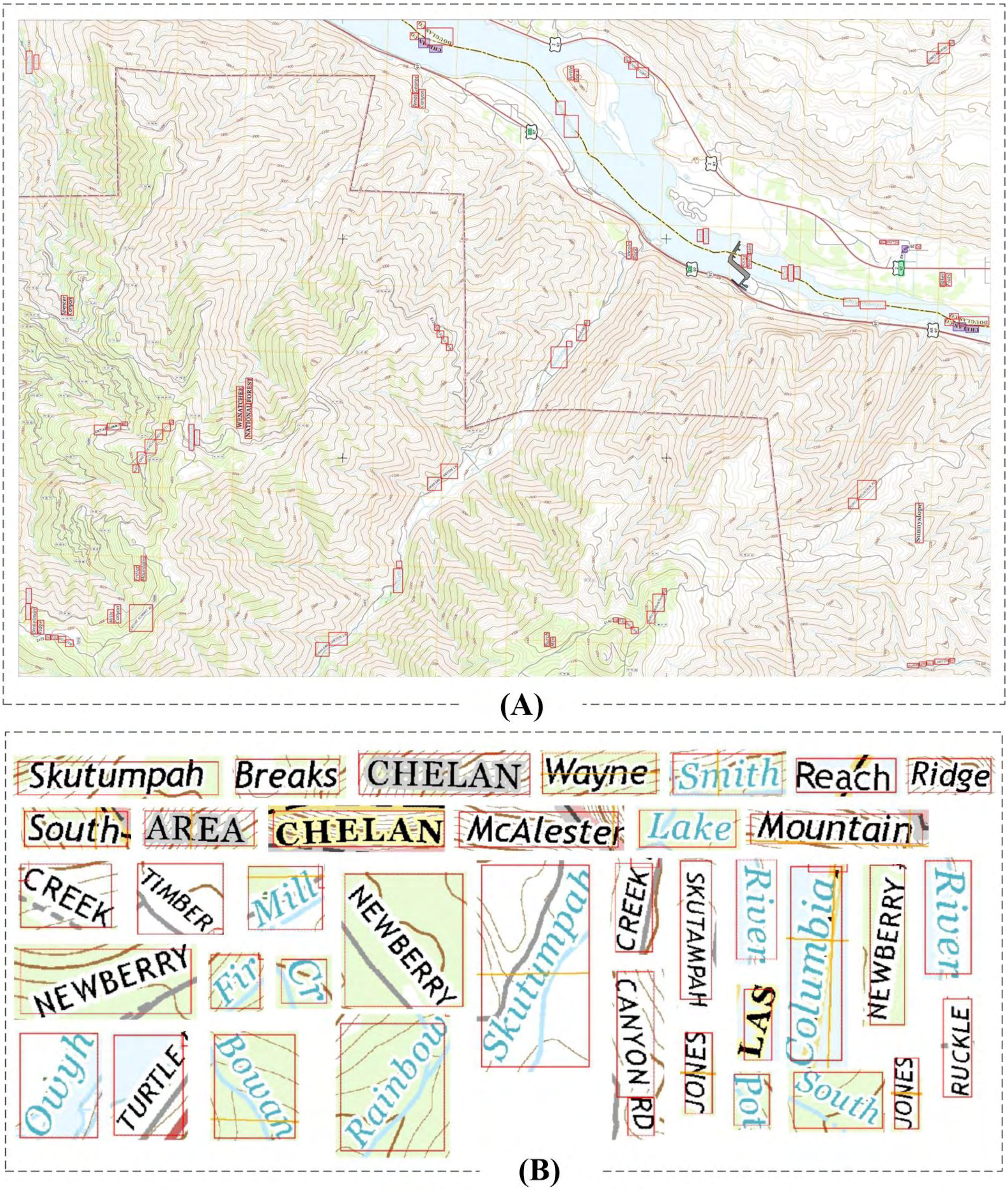


FIGURE 9. Detection results. (A). The entire map. (B) Detected map text.

straighten map text placed at different angles. Such straightening might be particularly challenging when a place name is designed with a curved shape. However, place names that are printed with non-linear shapes are rare in topographic maps.

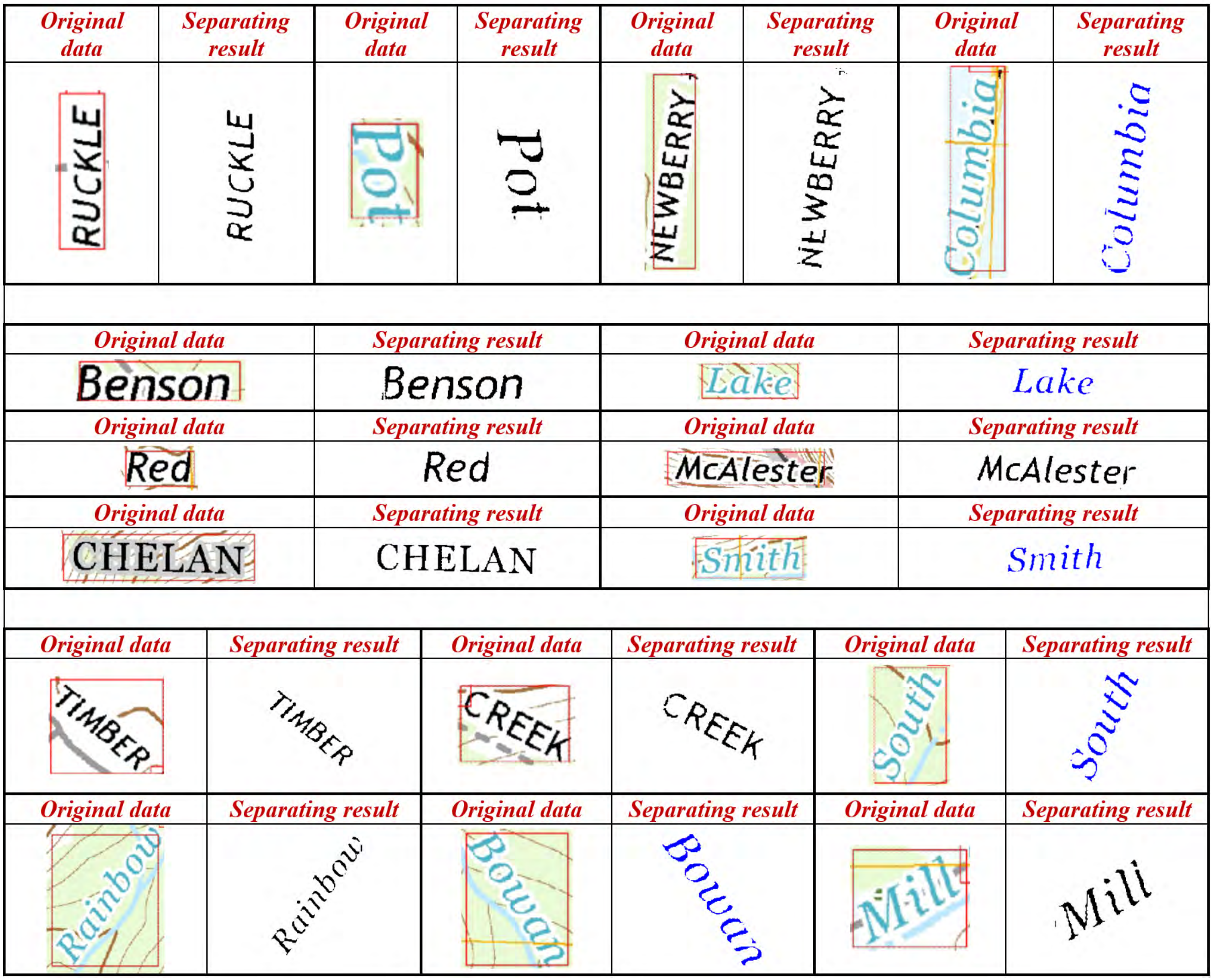


FIGURE 10. Illustration of the clustering results.

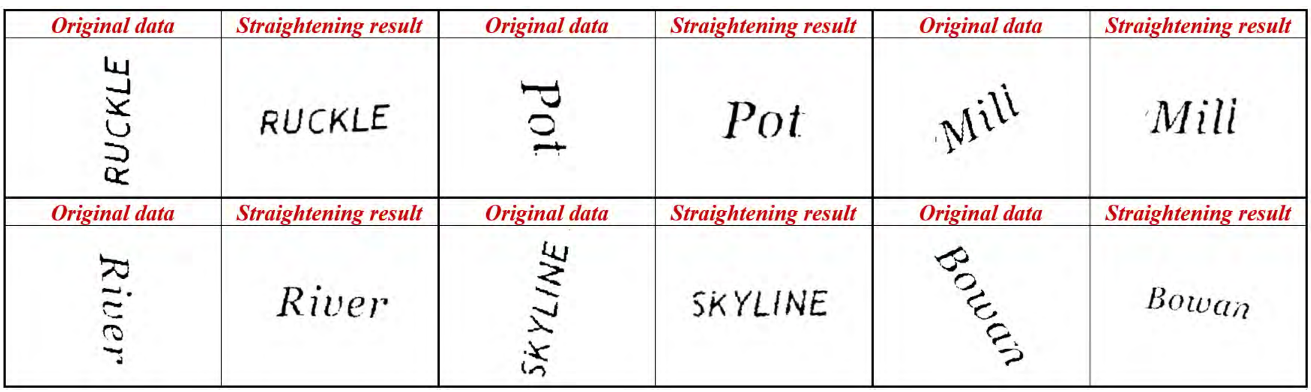


FIGURE 11. Illustration of the straightening results.

Although some straightened text units might not be exactly horizontal, our experimental results show that this will not affect the performance of Google Tesseract OCR engine in identifying map text.

The reason for this success is that graph-based segmentation is invariant to rotation. Although graph-based segmentation is prone to over-segmentation, the over-segmented subregions can subsequently be grouped into meaningful clusters

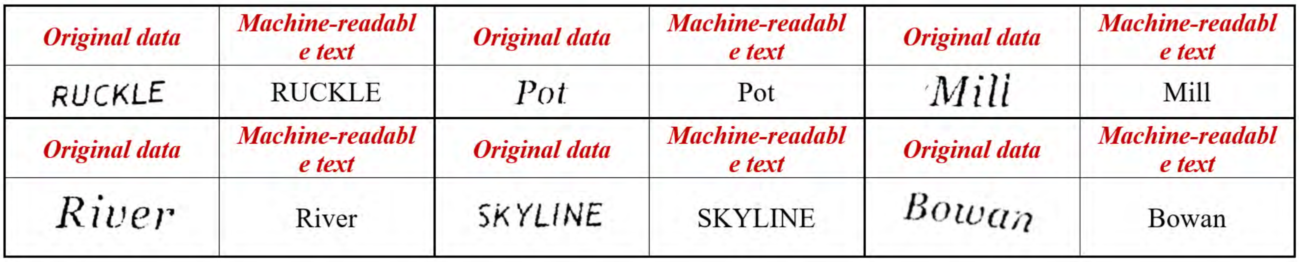


FIGURE 12. Illustration of the OCR results.

TABLE 1. Evaluation of the map text detection and identification results from 20 topographic maps, with anchor scales of 4, 8 and 16 and aspect ratios of 1:4, 2:5, 1:2, 1:3, 2:1, 5:2, 3:1 and 4:1.

Topographic map	AP of map text recognition (%)
USGS US Topo 7.5-minute map for Onego, WV 2016	95.57
USGS US Topo 7.5-minute map for Eagle Nest, ID 2017	95.13
USGS US Topo 7.5-minute map for Terrell Creek, MT 2017	95.35
USGS US Topo 7.5-minute map for Rocky Reach Dam, WA 2017	94.84
USGS US Topo 7.5-minute map for Skutumpah Creek, UT 2017	94.70
USGS US Topo 7.5-minute map for Billerica, MA 2015	92.97
USGS US Topo 7.5-minute map for Boston North, MA 2015	91.75
USGS US Topo 7.5-minute map for Brooklin, NY 2016	92.52
USGS US Topo 7.5-minute map for Jamaica, NY 2016	92.55
USGS US Topo 7.5-minute map for Flagstaff West, AZ 2014	94.27
USGS US Topo 7.5-minute map for Sunland, CA 2015	93.40
USGS US Topo 7.5-minute map for South Lake Tahoe, CA-NV 2015	94.73
USGS US Topo 7.5-minute map for Sled Springs, OR 2017	93.67
USGS US Topo 7.5-minute map for Evergreen, CO 2016	94.89
USGS US Topo 7.5-minute map for Marsh-Miller Lake, WI 2015	96.51
USGS US Topo 7.5-minute map for Hamilton, NY 2016	92.31
USGS US Topo 7.5-minute map for San Jose East, CA 2015	94.65
USGS US Topo 7.5-minute map for El Monte, CA 2015	94.86
USGS US Topo 7.5-minute map for Wappapello, MO 2017	94.71
USGS US Topo 7.5-minute map for Hilo, HI 2017	94.93
All 20 maps	94.78

by means of DBSCAN based on their intensity similarity. Thus, the results shown in Figure 11 validate the performance and robustness of our proposed approach for separating text units from other map features.

After straightening, we applied the Google Tesseract OCR engine for map text recognition. The accuracy results for text identification are listed in Table 1. Tesseract achieved an accuracy of almost 100% in recognizing the detected, separated and straightened map text in this experiment. Figure 11 shows the results of converting the detected map text into a machine-readable format.

Overall, the robustness and effectiveness of our proposed approach for map text recognition are validated by the results shown in Figures 9 to 12 and Table 1. The experimental results demonstrate that the proposed methodology, including map text detection, map text separation and map text recognition, enables the successful recognition of almost all map text in a topographic map.

C. RESULT OF MAP UNDERSTANDING

Based on the map text recognized from the USGS US Topo 7.5-minute map for Onego, WV 2016, Table 2 lists the information obtained from the GNIS with respect to each recognized phrase. Here, we present the class, state, county and map name information for each phrase of map text derived from the topographic map. Based on the information listed in Table 2, we conclude that the location covered by this map is West Virginia, U.S. The regions contained in this map mainly cover Pendleton County, which includes dozens of streams, arroyos, or valleys; four populated places called Onego, Teterton, Seneca Rocks and Macksville; and Monongahela National Forest. Since no street names or building names were detected in this topographic map, we can deduce that this map represents a large-scale landscape in which most areas are natural scenes. Moreover, because of the inclusion of Monongahela National Forest and the dozens of streams, the land cover in the area represented by this topographic map is expected to have rich forest coverage and adequate water distribution.

In addition to producing series of topographic maps, the Center of Excellence for Geospatial Information Science of the USGS has established an ontology called USTopographic to provide access to formal definitions of classes used in topographic maps. Moreover, extended attributes of the

TABLE 2. Information obtained from the GNIS with respect to the detected map text.

Machine-readable text derived from topographic map	Information accessed from GNIS					
	Class	State	County	Map name	Geographical coordinates	…
Horsecamp Run	Stream Valley Arroyo	West Virginia	Pendleton	Onego Whitmer Laneville Harman	…	…
Onego	Populated Place	West Virginia	Pendleton	Onego	…	…
Teterton	Populated Place	West Virginia	Pendleton	Onego	…	…
Seneca Rocks	Populated Place	West Virginia	Pendleton	Onego Upper Tract	…	…
Seneca Rocks	Cliff	West Virginia	Pendleton	Upper Tract	…	…
Macksville	Populated Place	West Virginia	Pendleton	Macksville	…	…
Monongahela National Forest	Forest	West Virginia Virginia	Randolph Barbour Alleghany Pendleton …	Onego Upper Tract Droop Marlinton …	…	…

TABLE 3. Extended formal definitions and attributes associated with classes and place names listed in Table 2.

Place name or class name	Sources	Content	Relationship
Stream	USTopographic	Linear body of water flowing on the Earth's surface.	*subClassOf* Surface Water *hasPart* Head waters *hasPart* Mouth *hasPart* Streambank
Valley	USTopographic	An elongated depression in the earth's surface which generally slopes from one end to the other	*subClassOf* Terrain
Arroyo	USTopographic	Watercourse or channel through which water may occasionally flow (coulee, draw, gully, wash)	*subClassOf* Surface Water
West Virginia	Wikipedia	West Virginia is a state located in the Appalachian region of the Southern United States. It is bordered by Virginia to the southeast, Kentucky to the southwest, Ohio to the northwest, Pennsylvania to the north (and, slightly, east), and Maryland to the northeast.	
		…	

places named in Table 2 can be accessed from volunteered information and linked databases. Table 3 lists the extended formal definitions and attributes associated with some of the classes and place names listed in Table 2. In the future, we will establish a query system to support semantic queries based on the information accessed from the GNIS.

V. CONCLUSION

This paper reports our efforts to develop a methodological framework to support the automatic understanding of the contents of topographic maps via deep learning and gazetteer-supported information queries. The map text recognition process consists of detecting map text from a map context based on Faster R-CNN, separating the map text from the background in the map area associated with each detected text unit, and recognizing the contents of the text units by using the Google Tesseract OCR engine. The results of map text recognition are then converted into a machine-readable format to facilitate the retrieval of relevant information from a gazetteer called the GNIS. In the future, the map text recognition results obtained with our proposed methodology may be combined with ontologies, open linked databases or volunteered geographic information to efficiently support cartographical applications. Our experimental results validate the efficiency and robustness of our proposed methodology for map text recognition and map content understanding.

Previous works on the detection of map text from optical topographic maps have used geometric features, the Hough transform, and segmentation. However, these approaches still face challenges when detecting map text from complicated contexts, especially when the map text is touching other map features, such as contours or geographical features. The DCNN approach has markedly outperformed other approaches in object recognition and OCR [60], [61]. The investigations conducted in this study prove that the DCNN approach has great potential to facilitate automatic and efficient map text recognition and map content understanding. We expect the proposed intelligent map reader to contribute to a variety of applications in the GeoAI field.

In the future, a number of possible extensions of this research might be worthy of investigation. In the proposed intelligent map reader, map text detection and map text recognition are performed separately. We plan to develop a new DCNN framework for joint map text detection and recognition. Moreover, to enable further knowledge discovery from topographic maps, we will explore a methodological framework in which the proposed intelligent map reader is integrated with ontologies. We hope that the proposed intelligent map reader will contribute to the implementation of AI techniques in cartographical applications and geographical information science.

REFERENCES

[1] A. Pezeshk and R. L. Tutwiler, "Extended character defect model for recognition of text from maps," in *Proc. SSIAI*, Austin, TX, USA, May 2010, pp. 85–88.

[2] A. Pezeshk and R. L. Tutwiler, "Improved multi angled parallelism for separation of text from intersecting linear features in scanned topographic maps," in *Proc. ICASSP*, Dallas, TX, USA, Mar. 2010, pp. 1078–1081.

[3] A. Pezeshk and R. L. Tutwiler, "Automatic feature extraction and text recognition from scanned topographic maps," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 5047–5063, Dec. 2011.

[4] M. P. Deseilligny *et al.*, "Topographic maps automatic interpretation : Some proposed strategies," in *Proc. GREC*, Nancy, France, 1997, pp. 175–193.

[5] A. Khotanzad and E. Zink, "Contour line and geographic feature extraction from USGS color topographical paper maps," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 1, pp. 18–31, Jan. 2003.

[6] H. Wang and H. Yan, "Text extraction from color map images," *J. Electron. Imag.*, vol. 3, no. 4, pp. 390–396, Oct. 1994.

[7] M. Caprioli and P. Gamba, "Detecting and grouping words in topographic maps by means of perceptual concepts," in *Proc. EUSIPCO*, Tampere, Finland, Sep. 2000, pp. 1–4.

[8] H. Yamada, K. Yamamoto, and K. Hosokawa, "Directional mathematical morphology and reformalized Hough transformation for the analysis of topographic maps," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 4, pp. 380–387, Apr. 1993.

[9] Z. Lu, "Detection of text regions from digital engineering drawings," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 4, pp. 431–439, Apr. 1998.

[10] D. Dori and Y. Velkovitch, "Segmentation and recognition of dimensioning text from engineering drawings," *Comput. Vis. Image Understand.*, vol. 69, no. 2, pp. 196–201, 1998.

[11] G. K. Myers, P. G. Mulgaonkar, C.-H. Chen, J. L. DeCurtins, and E. Chen, "Verification-based approach for automated text and feature extraction from raster-scanned maps," in *Proc. GREC*, University Park, PA, USA, 1995, pp. 190–203.

[12] A. Pezeshk and R. L. Tutwiler, "Text segmentation and reorientation from scanned color topographic maps," in *Proc. ICSIP*, Kailua-Kona, HI, USA, 2008, pp. 94–97.

[13] R. Cao and C. L. Tan, "Separation of overlapping text from graphics," in *Proc. ICDAR*, Seattle, WA, USA, Sep. 2001, pp. 44–48.

[14] C. L. Tan, and P. O. Ng, "Text extraction using pyramid," *Pattern Recognit.*, vol. 31, no. 1, pp. 63–72, 1998.

[15] L. Li, G. Nagy, A. Samal, S. Seth, and Y. Xu, "Integrated text and line-art extraction from a topographic map," *Int. J. Doc. Anal. Recognit.*, vol. 2, no. 4, pp. 177–185, 2000.

[16] R. Cao and C. L. Tan, "Text/graphics separation in maps," in *Proc. GREC*, Kingston, ON, Canada, 2001, pp. 167–177.

[17] K. Tombre, S. Tabbone, L. Pélissier, B. Lamiroy, and P. Dosch, "Text/graphics separation revisited," in *Proc. DAS*, Princeton, NJ, USA, 2002, pp. 200–211.

[18] D. X. Zhong, "Extraction of embedded and/or line-touching character-like objects," *Pattern Recognit.*, vol. 35, no. 11, pp. 2453–2466, 2002.

[19] A. Velázquez and S. Levachkine, "Text/graphics separation and recognition in raster-scanned color cartographic maps," in *Proc. GREC*, Barcelona, Spain, 2003, pp. 63–74.

[20] P. P. Roy, E. Vazquez, J. Lladós, R. Baldrich, and U. Pal, "A system to segment text and symbols from color maps," in *Proc. GREC*, Curitiba, Brazil, 2007, pp. 245–256.

[21] J. Pouderoux, J.-C. Gonzato, A. Pereira, and P. Guitton, "Toponym recognition in scanned color topographic maps," in *Proc. ICDAR*, Curitiba, Brazil, Sep. 2007, pp. 531–535.

[22] N. Kerle and J. de Leeuw, "Reviving legacy population maps with object-oriented image processing techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2392–2402, Jul. 2009.

[23] D. B. Dhar and B. Chanda, "Extraction and recognition of geographical features from paper maps," *Int. J. Doc. Anal. Recognit.*, vol. 8, no. 4, pp. 232–245, 2006.

[24] T. Wang, D. J. Wu, A. Coates, and A. Y. Ng, "End-to-end text recognition with convolutional neural networks," in *Proc. ICPR*, Tsukuba, Japan, Nov. 2012, pp. 3304–3308.

[25] M. Jaderberg, A. Vedaldi, and A. Zisserman, "Deep features for text spotting," in *Proc. ECCV*, Zurich, Switzerland, 2014, pp. 512–528.

[26] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman, "Reading text in the wild with convolutional neural networks," *Int. J. Comput. Vis.*, vol. 116, no. 1, pp. 1–20, 2016.

[27] A. Gupta, A. Vedaldi, and A. Zisserman, "Synthetic data for text localisation in natural images," in *Proc. CVPR*, Las Vegas, NV, USA, Jun. 2016, pp. 2315–2324.

[28] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[29] H. Jiang and E. Learned-Miller, "Face detection with the faster R-CNN," in *Proc. FG*, Washington, DC, USA, May/Jun. 2017, pp. 650–657.

[30] G. S. Cheema and S. Anand, "Automatic detection and recognition of individuals in patterned species," in *Proc. ECML PKDD*, Skopje, Macedonia, 2017, pp. 27–38.

[31] P. Zhou, G. Cheng, Z. Liu, S. Bu, and X. Hu, "Weakly supervised target detection in remote sensing images based on transferred deep features and negative bootstrapping," *Multidimensional Syst. Signal Process.*, vol. 27, no. 4, pp. 925–944, 2016.

[32] M. Kampffmeyer, A. B. Salberg and R. Jenssen, "Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks," in *Proc. CVPR*, New York, NY, USA, Jun./Jul. 2006, pp. 680–688.

[33] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. ECCV*, Zurich, Switzerland, 2014, pp. 818–833.

[34] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online.] Available: https://arxiv.org/abs/1409.1556

[35] R. Girshick. (2015). "Fast R-CNN." [Online.] Available: https://arxiv.org/abs/1504.08083

[36] A. Kay, "Tesseract: An open-source optical character recognition engine," *Linux J.*, vol. 2007, p. 2, Jul. 2007.

[37] R. Smith, "An overview of the Tesseract OCR engine," in *Proc. ICDAR*, Parana, Brazil, Sep. 2007, pp. 629–633.

[38] L. Vincent, "Google book search: Document understanding on a massive scale," in *Proc. ICDAR*, Parana, Brazil, Sep. 2007, pp. 819–823.

[39] C. Patel, A. Patel, and D. Patel, "Optical character recognition by open source OCR tool tesseract: A case study," *Int. J. Comput. Appl.*, vol. 55, no. 10, pp. 50–56, 2012.

[40] R. Mithe, S. Indalkar, and N. Divekar, "Optical character recognition," *Int. J. Recent Technol. Eng.*, vol. 2, no. 1, pp. 72–75, 2013.

[41] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, Sep. 2004.

[42] Q. Ye, W. Gao, and W. Zeng, "Color image segmentation using density-based clustering," in *Proc. ICME*, Baltimore, MD, USA, Jul. 2003, pp. II-401–II-403.

[43] J. Jiao, Q. Ye, and Q. Huang, "A configurable method for multi-style license plate recognition," *Pattern Recognit.*, vol. 42, no. 3, pp. 358–369, 2009.

[44] J. A. Kelmelis, M. L. DeMulder, C. E. Ogrosky, N. J. Van Driel, and B. J. Ryan, "The *NationalMap* from geography to mapping and back again," *Photogramm. Eng. Remote Sens.*, vol. 69, no. 10, pp. 1109–1118, 2003.

[45] Tzutalin. (2015). *LabelImg. Git Code*. [Online]. Available: https://github.com/tzutalin/labelImg

[46] X. Chen and A. Gupta. (2017). "An implementation of faster RCNN with study for region sampling." [Online]. Available: https://arxiv.org/abs/1702.02138

[47] X. Sun, P. Wu, and S. C. H. Hoi. (2017). "Face detection using deep learning: An improved faster RCNN approach." [Online]. Available: https://arxiv.org/abs/1701.08289

[48] M. F. Goodchild and L. L. Hill, "Introduction to digital gazetteer research," *Int. J. Geograph. Inf. Sci.*, vol. 22, no. 10, pp. 1039–1044, 2008.

[49] J. T. Hastings, "Automated conflation of digital gazetteer data," *Int. J. Geograph. Inf. Sci.*, vol. 22, no. 10, pp. 1109–1127, 2008.

[50] M. F. Goodchild, P. Fu, and P. Rich, "Sharing geographic information: An assessment of the geospatial one-stop," *Ann. Assoc. Amer. Geograph.*, vol. 97, no. 2, pp. 250–266, 2007.

[51] K. Janowicz and C. Keßler, "The role of ontology in improving gazetteer interaction," *Int. J. Geograph. Inf. Sci.*, vol. 22, no. 10, pp. 1129–1157, 2008.

[52] C. Keßler, P. Maué, J. T. Heuer, and T. Bartoschek, "Bottom-up gazetteers: Learning from the implicit semantics of geotags," in *Proc. ICGeoS*, Mexico City, Mexico, 2009, pp. 83–102.

[53] S. D. Cardoso *et al.*, "SWI: A semantic web interactive gazetteer to support linked open data," *Future Generat. Comput. Syst.*, vol. 54, pp. 389–398, Jan. 2016.

[54] L. M. Olsen *et al.* (2007). *NASA/Global Change Master Directory (GCMD) Earth Science Keywords. Version 6.0.0.0.0*. [Online]. Available: http://gcmd.nasa.gov/Resources/valids/archives/keyword_list.html

[55] E. L. Usery and D. Varanka, "Design and development of linked data from the national map," *Semantic Web*, vol. 3, no. 4, pp. 371–384, 2012.

[56] R. G. Raskin and J. P. Michael, "Knowledge representation in the semantic Web for Earth and environmental terminology (SWEET)," *Comput. Geosci.*, vol. 31, no. 9, pp. 1119–1125, 2005.

[57] M. K. Kavouras, M. Kokla, and E. Tomai, "Comparing categories among geographic ontologies," *Comput. Geosci.*, vol. 31, no. 2, pp. 145–154, 2005.

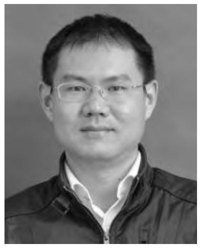
[58] M. Perry and J. Herring, *OGC GeoSPARQL—A Geographic Query Language for RDF Data*, OGC Implementation Standard OGC 11-052r4, 2012.

[59] D. Sui, M. F. Goodchild, and S. Elwood, "Volunteered geographic information, the exaflood, and the growing digital divide," in *Crowdsourcing Geographic Knowledge*. Dordrecht, The Netherlands: Springer, 2013, pp. 1–12.

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

[61] Q. Ye and D. Doermann, "Text detection and recognition in imagery: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 7, pp. 1480–1500, Jul. 2015.

HUALI LI received the B.S. degree from the School of Remote Sensing and Information Engineering, Wuhan University, China, in 2007, and the M.S. and Ph.D. degrees from the State Key Laboratory of Information Engineering, Surveying, Mapping and Remote Sensing, Wuhan University, in 2009 and 2012, respectively. In 2012, she was an Assistant Professor with Hunan University. Her research interests include hyperspectral image processing mainly spectral unmixing and some related applications.



JUN LIU received the B.S. and Ph.D. degrees from the School of Remote Sensing and Information Engineering, Wuhan University, China, in 2007 and 2012, respectively. From 2012 to 2015, he was an Assistant Researcher Fellow with the Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences. Since 2015, he has been an Associate Researcher Fellow with the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences. His research interests include image processing (such as image enhancement and image fusion), data mining and high-performance computing in remote sensing and geography fields.

XIRAN ZHOU (S'14) received the B.S. degree from Ningbo University in 2010 and the M.S. degree from Wuhan University in 2013. He is currently pursuing the Ph.D. degree with Arizona State University. His research interests include computer vision, machine learning, remote sensing, and GIScience.