

GeoAI-enhanced Techniques to Support Geographical Knowledge Discovery from Big
Geospatial Data

by

Xiran Zhou

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2019 by the
Graduate Supervisory Committee:

Wenwen Li, Chair
Soe W. Myint
Samantha T. Arundel

ARIZONA STATE UNIVERSITY

May 2019

ABSTRACT

Big data that contain geo-referenced attributes have significantly reformed the way that I process and analyze geospatial data. Compared with the expected benefits received in the data-rich environment, more data have not always contributed to more accurate analysis. “Big but valueless” has becoming a critical concern to the community of GIScience and data-driven geography. As a highly-utilized function of GeoAI technique, deep learning models designed for processing geospatial data integrate powerful computing hardware and deep neural networks into various dimensions of geography to effectively discover the representation of data. However, limitations of these deep learning models have also been reported when People may have to spend much time on preparing training data for implementing a deep learning model. The objective of this dissertation research is to promote state-of-the-art deep learning models in discovering the representation, value and hidden knowledge of GIS and remote sensing data, through three research approaches. The first methodological framework aims to unify varied shadow into limited number of patterns, with the convolutional neural network (CNNs)-powered shape classification, multifarious shadow shapes with a limited number of representative shadow patterns for efficient shadow-based building height estimation. The second research focus integrates semantic analysis into a framework of various state-of-the-art CNNs to support human-level understanding of map content. The final research approach of this dissertation focuses on normalizing geospatial domain knowledge to promote the transferability of a CNN’s model to land-use/land-cover classification. This research reports a method designed to discover detailed land-use/land-

cover types that might be challenging for a state-of-the-art CNN's model that previously performed well on land-cover classification only.

DEDICATION

I dedicate my dissertation work to my loving parents, my best friends and my dissertation committee members.

ACKNOWLEDGMENTS

First and foremost, I am indebted to Professor Wenwen Li for a valuable opportunity of pursuing my PhD at Arizona State University. I also convey my sincerest gratitude for all invaluable guidelines, advices and supports given by Professor Wenwen Li in the pursuit of my doctoral research. Thanks are also given to the research opportunity and support from Professor Wenwen Li's project of U.S. Geological Survey. It would not be possible for me to become a well-training researcher without the meticulous guidance given by Professor Wenwen Li. I have benefited much from her meticulous and rigorous scholarships, and dedicated commitments to the pursuit of excellence.

I dearly appreciate of Professor Soe Myint and Dr. Samantha Arundel for agreeing to serve on my dissertation committee members. I am also deeply grateful to the invaluable helps given by Professor Soe Myint and Dr. Samantha Arundel on the occasions that I have greatly changed the foci of my dissertation research. Their sharp perceptions on geography, geomorphology and environment have provided me extensive scientific viewpoints, profound academic sights, and professional career insights.

I also dedicate to all of those faculty members, postdocs, and graduate students with whom I have worked during the five-year doctoral research at Arizona State University. Each of them taught me much regarding both scientific aspects and technical skills.

Last, nobody could be more important to me than my parents. Their loving encouragements are special to the progress of my dissertation research. The completion

of my PhD pursuit dissertation would not be possible without their continued patience,
and endless support.

TABLE OF CONTENTS

	Page
LIST OF TABLES	ix
LIST OF FIGURES	x
CHAPTER	
1 INTRODUCTION	1
1.1 Big Data in GIScience	1
1.2 Challenges of Big Geospatial Data	4
1.3 Deep Learning in Applications Using Big Geospatial Data.....	5
1.4 Motivations	8
1.5 Dissertation Structure	9
2 USING GEOAI TECHNIQUES TO SUPPORT BUILDING HEIGHT ESTIMATION BASED ON OPEN REMOTE SENSING DATA	14
2.1 Introduction.....	14
2.2 Related Works.....	11
2.2.1 Geometrical Relationships Between Building, Building Shadow, Collection Sensor, and Solar Position.....	17
2.2.2 State-of-The-Art Approaches.....	19
2.3 Methods.....	20
2.3.1 Shadow Pattern Determination	20
2.3.2 Height Calculation with Google Earth Pro	25
2.3.3 Height Calculation with Shadow Area Extraction	28

CHAPTER	Page
2.3.4 Building Height Estimation.....	35
2.4 Experiments	36
2.4.1 Experimental dataset	36
2.4.2 Shadow Extraction Results	37
2.4.3 Accessing Parameters from Google Earth Pro	41
2.4.4 Results of Building Height Estimation	44
2.5 Conclusions	46
3 INTEGRATING DEEP LEARNING AND SEMANTIC ANALYSIS TO SUPPORT HUMAN-LEVEL DIGITAL MAP RECOGNITION	48
3.1 Introduction	48
3.2 Methodology	53
3.2.1 Map-text Recognition.....	53
3.2.2 Map-text Detection.....	54
3.2.3 Map-text Unit Separation and Classification.....	55
3.2.4 Map-type Classification	58
3.3 Human-level Map Understanding	59
3.3.1 Ontology Development and Integration	61
3.3.2 Semantic Query and Reasoning	63
3.4 Experiments	65
3.4.1 deepMap: A Benchmark Data Set for Map-text Recognition.....	65
3.4.2 Experimental Design	67
3.4.3 Demonstrative Results.....	72

CHAPTER	Page
3.4.4 Discussions	74
3.5 Conclusions	77
4 DOMAIN KNOWLEDGE-ENHANCED LAND-COVER/LAND-USE CLASSIFICATION WITH IMAGE-SEMANTIC MODEL	80
4.1 Introduction	80
4.2 Benchmark Dataset for Coastal Scene Recognition	85
4.3 CNNs-enhanced Image-semantic Model	89
4.3.1 Multi-label Land-cover/land-use Classification	89
4.3.2 Image-semantic Model	100
4.4 Experiments	104
4.4.1 Study Area	105
4.4.2 Coastline Land-use Similarity Analysis	107
4.4.3 Coastline land-use scene retrieval	112
4.5 Conclusions	116
5 CONCLUSIONS	119
5.2 Conclusions	119
5.3 Future Works	120
REFERENCES	121

LIST OF TABLES

Table	Page
1. Dissertation Structure	9
2. Three Types of Methods for Building Height Estimation and Shadows	19
3 Building Height Estimation Using Two Different Methods	44
4. Comparison of the Previous Version of Intelligent Map Reader and the Proposed Method	65
5. Statistics Used for the Experiment	68
6. Evaluation of Map Recognition	73
7. Brief Information on Remote-Sensing Datasets' Existing Benchmarks	87
8. Brief Information on Data Augmentation	92
9. Comparison of Coastline Scene Analysis by Various PNASNet-enhanced Classification Strategies	108
10. Target LULC in the Selected Images	113
11. Comparison of Coastline Retrieval by Various PNASNet-enhanced Classification Strategies	115

LIST OF FIGURES

Figure	Page
1. Geometrical Relationship among Building Height, Sun Position, and Sensor Position. (A) First Relation Including the Solar and the Sensor Elevation (Profile View); (B) Second Relation Including the Solar and the Sensor Elevation; (C) Relation Including the Solar and Sensor Azimuth (Plan View); (D) 3d Geometrical Relation	18
2 Illustration of the Pattern Categories in ShadowClass	22
3. Illustration of Inception-resnet-V2	24
4. Illustration of Accessing the Solar and the Sensor Azimuth from Google Earth Pro. (A) Original Building Image. (B) Accessing the Length of BO ₂ and the Sensor Azimuth (θ_2) Using the Ruler Tool. (C) Accessing the Length of AO ₂ and the Solar Azimuth Using the Ruler Tool (θ_1)	27
5. Illustration of the (A) 16 Possible Configurations of the MSA. (B) Sample Workflow of the Ramer–Douglas–Peucker Algorithm	33
6. Illustration of the Selected Test VHR Images	37
7. Illustration of Buildings Visualized by VHR Images and Image Processing Results by the Gabor Filter, Histogram Equalization, Traditional AGC, and AGCWD	39
8. Illustration of Buildings Visualized by VHR images, Enhanced Images by AGCWD, and Results of Raw Shadow Extraction, Shadow Contour Extraction, and Shadow Polygon Simplification	41

Figure	Page
9. Illustration of Buildings Visualized by the VHR image, the Line Segment for Accessing the Length of AO ₂ and the Sensor Azimuth (θ_1) shown in Figure 4(C), and the Length of BO ₂ and the Solar Azimuth (θ_2) Shown in Figure 4(C)..	42
10. Illustration of the Buildings Visualized by VHR Images, the Results of Shadow Extraction, the Shadow Pattern to Which an Extracted Shadow Was Assigned, and the Selected Shadow Length for Building Height Estimation	43
11 Methodological Framework of Map-text Recognition	54
12 (A) Aligned Map-text String Straightening and (B) Curved Map-text String Straightening	57
13 From Extracted Map Features to Map Content With Semantic Analysis	61
14 Enriching Individuals in the Integrated Ontology in the Case of Arizona State University	63
15 Illustration of the Selected Samples from the <i>deepMap</i> Benchmark Data set	66
16 Illustration on the Selected Maps Accessed via Google Image Search	70
17. Illustration on the Selected Semantic Query	74
18. Illustration of the Selected Image Samples in the Benchmark Dataset	89
19. Comparison of Binary Classification, Multi-class Classification, Multi-label Classification, and Multi-label Classification with Spatial Weights	90
20. Illustration of NAS Strategy	95

Figure	Page
21. Image Gridding	99
22. Matrix of VSM; (A) Land-cover Category Frequency and Inverse Image Frequency; (B) Matrix Structure; (C) an Example of a VSM Matrix	101
23. Visualization of Study Area; (A) Study Area Map; (B) Samples of Coastal LULC.	106
24. Illustration of the Selected Image Pairs Used for Scenic Similarity Evaluation. (A) Selected Examples on Different Coastline LULC Types. (B) Selected Examples on Coastline LULC Type Measurement. (C) Illustration on Challenges for the Proposed Method.....	111
25. The Selected Images for Image Retrieval	112

CHAPTER 1

INTRODUCTION

1.1 Big Data in GIScience

Earth observation systems (Acker and Leptoukh, 2007; Li et al., 2016), web crawlers (Li, Yang and Yang, 2010), social media (Sui and Goodchild, 2011), cyberinfrastructure (Yang et al, 2010; Wright and Wang, 2011), and cloud-based data catalogs (Gorelick et al., 2017) generate uncountable volumes of data each day. This data-rich environment significantly reshapes the structure of geospatial research, and provides great possibilities for many geospatial data analyses, such as large-scale land change detection (Hansen and Loveland, 2012), global climate analysis (Faghmous and Kumar, 2014). The “3Vs” of big data (Lee and Kang, 2015)—volume, velocity and variety, respectively describe the 1) massive amounts of data created each day, 2) the rapidity of data generation and acquisition, and 3) the large number of data types to be analyzed. As integrated techniques for big data analysis progress, value and veracity become other essential characteristics of big data (Emani, Cullot and Nicolle, 2015). Value reflects the usability and practicality of a dataset. In practice, a majority of datasets might be useless for a data analysis task without their collection specifically planned to address that task (LaValle et al., 2011). Veracity is a significant measurement of data quality and reliability. In the context of big data, veracity not only refers to the accuracy of ground truth data, but also indicates the trustworthiness of data sources and data generation.

Claims that approximately 80% of data are geographically referenced signify the importance of the spatial dimension in big data (Franklin and Hane, 1992; Morais, 2012).

Big data containing geo-referenced attributes, which are viewed as big geospatial data, transform geography and related fields into data-driven disciplines (Graham and Shelton, 2013; Kitchin, 2013; Miller and Goodchild, 2015; Robinson et al., 2017). To better understand the influences of big geospatial data, Miller and Goodchild (2015) categorized big geospatial data sources into five types—location-aware devices, in-situ sensors, satellite and aircraft-based Earth observation systems, radio frequency identification, and social media. Data generated from these sources have not only reformed the way in which people collect and process geospatial data, they have also reshaped the sight of data analysis and data visualization. Big geospatial data have already impressed remarkable changes on different dimensions of geography.

(1) Location. Global navigation satellite systems (GNSS), such as GPS, Galileo, and Beidou, enable users to receive real-time, accurate geographical positioning services with portable devices. For example, timely-updated GPS signals help people understand real-time traffic patterns between a city and its neighborhoods using their smartphone (Luo et al., 2013; Pang et al., 2013). Aside from GNSS, inertial navigation systems that incorporate data derived from rotation sensors and motion sensors can support precise determination of indoor position, user orientation, and speed prediction when GNSS signals are blocked by building materials.

(2) Places and regions are essential attributes for understanding natural and constructed elements of the Earth's surface. Field investigations are often impractical to collect data for large areas, or under dangerous conditions such as flooding and other hazards. Various Earth observation data that are generated at a relatively rapid speed help people monitor and analyze events and phenomena without the workload of field

investigations. Examples of fields requiring huge amounts of geospatial data to explore a place or a region include the monitoring of global climate (Hansen et al., 2013), the terrestrial carbon cycle (Schimel et al., 2015), sea surface temperatures (Donlon et al., 2012), and melting polar ice caps (Hall et al., 2013).

(3) Physical systems and human systems. Much research has reported the significance of big geospatial data in a variety of geospatial applications that involve natural powers or human activities. For example, spatial details and spectral information in Earth observation data can help people receive in-depth knowledge about physical systems or human activities (Hansen and Loveland, 2012; Gómez, White and Wulder, 2016) at different spatial scales. Such efforts associated with Earth observation data also include hurricane path tracking (Wang, Zhao and Shen, 2012), seismic monitoring (Mordret et al., 2016), land cover mapping (Gómez, White and Wulder, 2016), and land change detection (Hansen and Loveland, 2012). Moreover, big social media data and real-time GPS signals are potentially useful for predicting the trajectory and movement of citizens (Spangenberg, 2014), predicting human behavior (Majid et al., 2013), analyzing public health (Houston, et al., 2015), monitoring urbanization (Srivastava et al., 2012) and discovering areas of interest (Jiang et al., 2015).

(4) Environment and society. The expansion of human activities renders the isolation of natural phenomena from cultural elements difficult in any geographical application. Big geospatial data offer hyperdimensional geographical aspects to explore the interaction, configuration and organization of a natural/human system (Graham and Shelton, 2013). Traditional field investigations and data analysis approaches lack a data-rich environment, meaning that those methods might be insufficient to describe the

interaction between environments and societies. For example, an accurate urban air quality estimation should consider geographical location and sophisticated air pollutant emissions (Gupta et al., 2006). Understanding other interactions between environment and society, such as urban heat islands (Estoque, Murayama and Myint, 2017), water quality (Gholizadeh, Melesse and Reddi, 2016), and deforestation (Ishtiaque, Myint and Wang, 2016), also relies on big geospatial data

1.2 Challenges of Big Geospatial Data Analysis

While the benefits of big data have been reported, challenges associated with big geospatial data analysis have attracted considerable attention. In the big data era, it is generally believed that more data always lead to more accurate data analysis results. This belief prompts people to collect as much of as many types of data as possible, ignoring the cost of data collection and storage (Goodchild, 2013; Chen and Zhang, 2014). For instance, although big data are now stored and available for public use through large-scale data portals and cyberinfrastructures, only limited amounts of the countless remote sensing images and spatio-temporal GIS datasets have been accessed or utilized. Data deficiency is still a major concern when designing geographical research. Compared to an emphasis on data collection, it may be more valuable to discover new uses of existing datasets. “Big but valueless” has become a decisive obstacle to benefiting from big geospatial data.

Moreover, studies regarding feature engineering have confirmed the significance of sparse data features and data quality in big data analysis (Kasun, et al., 2013; LeCun, Bengio and Hinton, 2015; Najafabadi et al., 2015). The efficiency of traditional machine learning relies heavily on the quality of manually-prepared data, which is limited when

open-source data and volunteer geographical information contain disorganized content and untruthful information. For example, map elements and map metadata are generally missing in the majority of maps on the Internet (ref). These maps also may contain incorrect place names and map features. Lastly, a large portion of valuable content might be hidden in the data itself. An example related to this concern is the difference between land cover mapping and land use interpretation. Researchers generally focus on mapping different land cover types from a remote sensing imagery, while ignoring the functionality and configuration of this land cover that impacts the LULC properties.

1.3 Deep Learning in Applications Using Big Geospatial Data

Artificial Intelligence (AI) is a discipline that creates intelligent machines that approach human intelligence in perception, learning, reasoning, and problem solving. AI techniques were initially developed in the 1950s, and have experienced three boom and bust cycles thanks to advances in computing engineering and automation. Software companies like Microsoft and Esri may have been the first to propose the term “GeoAI”, which attempts to integrate AI techniques and geospatial data into various dimensions of geography. Currently, the integration of GeoAI techniques into various geographical disciplines has been reported in a variety of applications, including automatic map recognition (Li, Liu and Zhou, 2018; Zhou, 2018), environmental health analysis (VoPham et al., 2018), air quality estimation (Li et al., 2017), geo-location discovery (Lin et al., 2015; Tian, Chen and Shah, 2017), ecological activity analysis (Miller-Rushing, Gallinat and Primack, 2019), rural area development prediction (Jean et al., 2016), the detection of interesting targets (Cheng and Han, 2016), and so on.

Recently, deep learning and its derivatives, such as reinforcement learning and graph learning, have become a new orientation in the third boom of AI development. The term “deep” in deep learning describes the deep architecture of a neural network to support a multi-layer data processing. The deep architecture increases the strength of processing data features into multi-level representations—from low-level data features to high-level abstract features (Bengio, Courville and Vincent, 2013). This means that deep learning can effectively handle more complex data features and patterns. Deep learning can be divided into three categories: unsupervised deep learning, supervised deep learning, and reinforcement learning.

Unsupervised machine learning aims to discover hidden patterns from unlabeled data. Since the process of feature learning is not included, unsupervised machine learning is simple and quick for data analysis. Unsupervised machine learning has developed into a number of geo-referenced algorithms, such as spatially-aware clustering (Wu, Zurita-Milla and Kraak, 2015; Yin et al., 2017), anomaly detection (Xiong and Zuo, 2018), dimensionality reduction (Romero, Gatta and Camps-Valls, 2016; Steiger, Resch and Zipf, 2016), and expectation–maximization optimization (Zhang et al., 2016). However, insufficient accuracy is the major restriction of these traditional unsupervised methods. For example, the accuracy of social media photo recognition might be lower than 60% by spectral unmixing (Hu et al., 2015; Zhou, Zhang and Wu, 2019), which is insufficient to support an accurate tourism AOI analysis. Deep neural networks (DNNs) significantly promote the power of unsupervised machine learning. A variety of DNNs such as autoencoders, deep belief nets, generative adversarial networks (GANs), and self-organizing maps have revealed promise in dimensionality reduction (Han, Zhong and

Zhang, 2017; Su et al., 2019), geotagged image and GPS fusion (Jiang, Kong and Fu, 2017), indoor navigation (Khatab, Hajihoseini and Ghorashi, 2018), spatial interpolation (Li et al., 2017), urban growth modelling (Zhou et al., 2017), ground image generation (Deng, Zhu and Newsam, 2018), and others.

Supervised machine learning algorithms are prominent and popular in GeoAI since they allow users to predefine criteria for the relationship between input data and output results. Compared with “shallow” machine learning techniques such as support machine vector, random forest, and adaptive boosting, deep supervised learning better manages complex representations of data and hyperdimensional data features, to support a great number of geospatial applications like land cover classification, land change detection, and object recognition (Cheng and Han, 2016; Zhang, Zhang and Du, 2016; Zhu et al., 2017).

The focus of reinforcement learning is to maximize the possibility of obtaining the best results through numerous iterative feedbacks in an interactive environment. Deep learning aims to create a model to predict the outputs from new input data, while reinforcement learning uses positive and negative signals to iteratively adjust the actions of an agent to obtain the richest cumulative rewards.

In a complex computing environment, reinforcement learning may become too formidable to effectively deal with problems of infinite probabilities. Thus, deep reinforcement learning—the product of integrating DNNs into reinforcement learning, becomes a promising AI technique (Mnih et al., 2015; Henderson et al., 2018). In the architecture of a deep reinforcement learning model, DNNs act as a powerful agent to recognize the state of complex patterns or features. The strength of cutting-edge deep

reinforcement learning has just been recognized in the community of GeoAI, and has been reported in a few applications (Liu et al., 2017; Peng et al., 2017; Hu et al., 2018).

1.4 Motivations

The motivation of this dissertation research is to promote the state-of-the-art deep learning models in big geospatial data analysis. The dissertation proposes three research questions associated with the current challenge of deep learning-based big geospatial data analysis.

Research question 1: how can useful input data promote deep learning models for a specific geospatial application?

Section 1.3 identifies the significance of deep learning in discovering high-level data feature. Some data features might be hidden from a deep learning model. For example, LULC properties may be out of scope of a CNN that relies on the training data of different land-cover types. Moreover, two characteristics of big data—value and veracity - acknowledge that not all features derived from big geospatial data would be useful for a specific geographical application. In the case of building height estimation with shadows, Chapter 2 proposes a classification system to unify varied shadow shapes into limited numbers of categories, and then organizes each category as a useful input data for building height estimation.

Research question 2: how does the integration of semantics with a deep learning model support geospatial knowledge discovery?

The success of deep learning relies on the amount and quality of training data. Semantics, which refers to the meaning of words, supports the formalization of human-level understanding based on the characteristics of phenomena. Thus, semantic

information can provide substantial clues to the explicit organization of information derived from geospatial data. In case of map content recognition, Chapter 3 exploits the technique of semantic analysis to formalize map text and map features derived from digital maps to help people understand the content of a map.

Research question 3: How can domain knowledge assist deep learning models in discovering geospatial knowledge with a limited amount of training data?

Subsection 1.2 discusses the challenge of preparing large quantities of labeled data. Deep learning requires large amounts of data for representing complicated systems. Creating rules and “geospatial common sense” is crucial to raise the transferability of a deep learning model to other geospatial applications with limited data. Considering the difference between land-cover categories and LULC types, Chapter 5 proposes an image-semantic model to exploit rules and knowledge to semantically organize land-cover categories into more usable LULC types.

1.5 Dissertation Structure

The content of this dissertation includes four independent, mutually-related papers, which are presented in Chapter 2, 3 and 4, respectively. Table 1 lists the relevance of these three chapters.

Table 1

Dissertation structure

	Chapter 2	Chapter 3	Chapter 4
Academic disciplines	Oblique Photogrammetry	Cartography	Remote Sensing & GIS data
Applications	Building height estimation	Digital map recognition	Coastal scene recognition
Techniques	1. Shadow extraction	Optical character recognition	1. None
	2. CNNs-powered shape classification	2. CNNs-powered multi-label classification	2. CNNs-powered multi-task classification
	3. Geometrical relationship	3. Semantic query	3. Vector space models

	between shadow and building		(VSM)
Data	1. Google Earth Pro	1. Internet maps	1. Satellite images 2. Coastline domain taxonomy
Motivations	Useful input data	Integration of semantics and deep learning	Domain knowledge and limited training data for deep learning models

Chapter 2 aims to create useful input data, rather than raw geospatial data, to support deep learning models for geospatial data analysis. This chapter proposes an integrated framework to support building height estimation with open very high-resolution (VHR) images based on representative shadow patterns.

Building height is valuable for a variety of foci in urban studies, such as flight safety control, urban air pollution, local temperature prediction, and residential energy consumption. Although traditional field investigations can obtain accurate building height information, this laborious and time-consuming work is not practical to support the update of numerous building heights in large urban areas. Moreover, elevation-related digital products (e.g. LiDAR data, DEM, DSM, and DTM) created for public use only cover selected areas, as updating them is quite costly for a large-scale region. Given the relationship between building structures and their shadow sizes, building shadows have been affirmed to provide an alternative data source to support building height estimation. The process of successful building height estimation with shadow size consists of two major steps: detecting building shadows and calculating the geometrical relationship between building shadow and building height.

Compared with the numerous research activities associated with shadow detection, only a few investigations have been performed to predict the height of a building using its shadow size. First, this chapter creates a shadow pattern classification system

(*ShadowClass*) not only to determine the pattern category to which a particular shadow belongs but also to draw the corresponding shadow-derived line useful for building height estimation. Then, the proposed framework provides two separate strategies for calculating building height, including building height estimation with Google Earth Pro, and building height estimation with shadow area extraction. Experimental results of building height estimation were close to the ground truth values. Without the support of elevation products (e.g., LiDAR, DTM, etc.), the shadow in both oblique and orthorectified VHR images could support the height prediction for low-, mid-, and high-floor buildings.

Chapter 3 aims to integrate a deep learning model with semantics to support geospatial knowledge discovery. This chapter proposes a framework to support the discovery of on-demand maps from Internet resources through convolutional neural networks, optical character recognition, and semantic analysis.

A map is an essential medium that provides symbolic representation and geographical information about 1) the characteristics of a place in terms of georeferenced location, 2) the distribution and pattern of phenomena over space, 3) the configuration of cultural and natural elements, and 4) the relationships between a variety of objects, areas, and phenomena. Over the last two decades, the progress of surveying, mapping, and web-service techniques have facilitated much of the efficiency of map generation and map sharing, and the benefits of maps have been identified in many geospatial interpretations, analyses, visualizations, and communications.

The considerable supply of maps currently available has encouraged researchers to focus on the efficiency of map retrieval and discovery, since the capability of

inefficient traditional interactive tools for map interpretation cannot meet the qualifications to process them. Map content recognition not only requires the conversion of map features into machine-readable map text, but also the semantic organization of this text into human understanding of map content.

Chapter 4 aims to use domain knowledge to reinforce deep learning models for geospatial knowledge discovery with a limited amount of training data. This chapter proposes an integrated framework called an *image-semantic model*, which identifies land cover in a remote sensing image using a convolutional neural network (CNN)-powered multi-label classification with spatial weights, and then employs a vector space model to convert the resulting information into more comprehensible LULC types.

Frequently updated remote sensing images provide the potential to support large-scale land-cover classification and reinforce competence in understanding, estimating, and predicting the influences of natural forces and artificial activities on land surface. The rapid progress of CNNs provides significant possibilities in the extraction of high-level abstract features from a remote sensing image to characterize complicated land-cover scenarios. Thus, land-cover attributes derived from remote sensing images are insufficient for directly predicting the functionality and organization of different land parcels.

The proposed Chapter 4 framework comprises three sections: (1) building a benchmark remote sensing dataset within limited land-cover categories, (2) performing multi-label land-cover classification with a pretrained CNN based on the training images available in the benchmark remote sensing dataset, and (3) organizing the land-cover categories to measure similarity with the vector space model, and then recognizing the

target land-cover/land-use scenarios. recognizing land-cover/land-use scenarios through organizing the land-cover categories to measure the similarity among different images with the vector space model.

The remainder of this dissertation is organized as follows. In Chapter 2, I report my research on building height estimation with shadows. In Chapter 3, I present my research on integrating semantics into a framework of deep learning to support map content recognition. The research focus of Chapter 4 is related to predict unknown land cover/land-use types with the proposed method that combines domain knowledge and the state-of-the-art CNNs for classification. In Chapter 5, I summarize the highlights of my dissertation research, and provide some future works worthy of attentions.

CHAPTER 2

USING GEOAI TECHNIQUES TO SUPPORT BUILDING HEIGHT

ESTIMATION BASED ON OPEN REMOTE SENSING DATA

2.1 Introduction

Building height is valuable for a variety of foci in urban studies, such as flight safety control, urban air pollution studies (Hang et al., 2012), local temperature prediction (Perini and Magliocco, 2014), and residence energy consumption (Abohela, Hamza and Dudek, 2013). Although a traditional field investigation can obtain accurate building height information, the laborious and time-consuming work is not practical to support massive building height updates in large-scale urban areas. Moreover, the elevation-related digital products (e.g. LiDAR data, DEM, DSM, and DTM) created for public use only cover selected places, as updating these data are costly. Given the relationship between building structures and their shadow sizes, building shadow has been affirmed to be an alternative data source to support building height estimation (Liasis and Stavrou, 2016; Qi, Zhai and Dang, 2016). This method becomes even more practical when geometrical shadows are visible in newly emerging very high resolution (VHR) images, such as GeoEye and Worldview. Such geometrical properties reinforce the significance of building shadows in places where elevation-related digital products are not updated in a timely manner (Comber et al., 2012; Raju, Chaudhary and Jha, 2014).

The process of successful building height estimation using shadow shape consists of two major steps: 1) detecting building shadows and 2) calculating the geometrical relationship between building shadow and building height. Compared with the numerous research activities associated with shadow detection (Liu, Fang and Li, 2011; Zhang, Sun

and Li, 2014), only a few investigations have been performed to predict the height of a building using its shadow shape. The early-stage attempts were reported by Irvin and McKeown (1989), Cheng and Thiel (1995), and Shettigara and Summerling (1998). The accuracy of building height estimation may be limited in these works because of the restriction of image spatial resolution. However, these attempts proved the practicability of building height estimation using shadow shape. Massalabi et al. (2004) summarized the key parameters derived from shadow that could support building height prediction, including the elevation and azimuth of the sun, the elevation and azimuth of a sensor/camera, and the relative position of a building. Wang and Wang (2009) used these key parameters to create the geometrical relationship between the solar azimuth, the satellite azimuth, and the building shadows. Shao, Taff, and Walsh (2011) created a linear function involving building position and shadow properties to easily calculate building height. On the basis of the key parameters mentioned by Massalabi et al. (2004), Kim, Javzandulam, and Lee (2007) and Lee and Kim (2013) promoted building height estimation by matching the projected shadow to the actual shadow. However, their proposed algorithms could be faced with the challenge of dealing with shadow delineation in a complicated three-dimensional (3D) space.

Thus, several investigations considered creating a 3D geometrical relationship involving the sun position, sensor position, and building position to measure building height (Wang, Yu and Ling, 2014; Qi, Zhai and Dang, 2016; Wang et al., 2017). To further simplify the calculation process, Izadi and Saeedi (2012) estimated building height by calculating the relative geometrical position of the sun, sensor, and building. They also reported an approach to promote the accuracy of building height using the

properties associated with building wall areas. Qi, Zhai, and Dang (2016) presented a framework to calculate the slope angle, the solar elevation and azimuth, and the satellite elevation and azimuth using information accessed from Google Earth. Hodul, Knudby, and Ho (2016) incorporated the effect of an undulating ground surface into the 3D geometrical relationship. They proposed a sky view factor model to predict building height using the amount of obstructed sky derived from satellite images and the slope of the ground surface. However, the resolution of Landsat data seemed inefficient to support accurate building height estimation.

Spatial resolution in a VHR image presents a new issue to be considered in building height estimation. Building shadows in a VHR image may have similar textures, patterns, and illuminations to other dark land cover (e.g., asphalt surface, water, etc.). Such similar features increase the difficulty in visually distinguishing building shadows and their surrounding land covers. Thus, a number of studies reported their efforts in designing additional features and applying advanced machine learning models. Comber et al. (2012) proposed a rule-based classification to determine the stories of residential buildings based on shadow width. This work did not take the geometrical positions of the sun and the sensor into account in the input features used in its machine learning model. Qi and Wang (2014) proposed a method called corner–shadow–length ratio to calculate building height considering both the geometrical relationships between the building roof structure and solar, sensor, and building positions. Thus, this method supported the measurement of building height differently between flat roofs and pitched (sloping) roofs. However, slope roofs might not be visually recognizable in the geometrically corrected VHR image.

Currently, the approaches for building height estimation vary according to the availability of image metadata, spatial resolution, and other factors. A comprehensive framework should consider these variations to support accurate shadow delineation and building height estimation. This study proposes an integrated framework to support building height estimation based on various conditions. The remainder of the chapter is organized as follows. Section 2.2 discusses the works that can help in building height estimation from VHR images. Section 2.3 reports the proposed methodological framework. Section 2.4 presents the results of height estimation of multiple-story buildings. Section 2.5 summarizes the contributions of the work, along with prospective future efforts.

2.2 Related Works

2.2.1 Geometrical Relationships Between Building, Building Shadow, Collection Sensor, and Solar Position.

Shadows are always observed from elevated artificial architectures, elevated hills and mountains, and clouds in a remote sensing image when these elevated objects block the visible light emitted by the sun. Currently, the availability of VHR images increased the accuracy of shadow height estimation using remote sensing data. Building shadows have become an inevitable artifact to be explored, as they may overlay other objects, such as buildings, roads, and vehicles, among others. This superimposition renders image analysis more difficult for a variety of applications involving land cover classification, land parcel segmentation, and object detection (Dare, 2005; Zhang, Sun and Li, 2014). According to how a shadow generates, Arevalo, González, and Ambrosio (2005, 2008) divided shadows into self-shadow and cast shadow. Self-shadow is the dark area of an

object itself where light is not available. Conversely, cast shadow is the dark area that is approximately similar to the projection of an object shape, where illumination is blocked by this object. The shadow visible in a VHR image is considered a cast shadow.

Therefore, accessing the height of an elevated building based on the projection of its shadow shape is potentially possible.

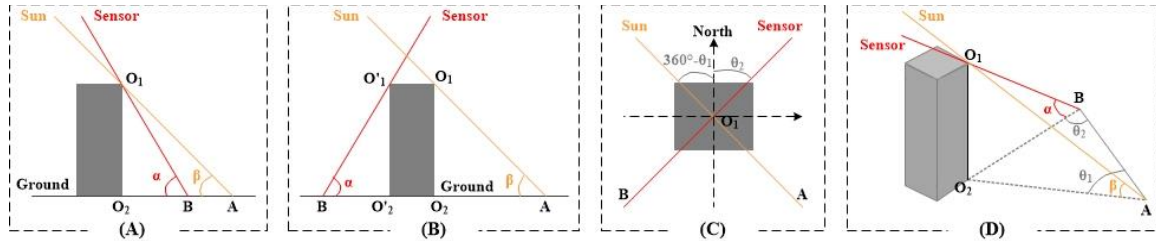


Figure 1. Geometrical relationship among building height, sun position, and sensor position. (A) First geometrical relationship between solar and sensor elevation (profile view); (B) Second relationship between solar and sensor elevation; (C) Relationship between solar and sensor azimuth (plan view); (D) 3D geometrical relationship.

Figure 1 presents the plan view and the profile view to illustrate the geometrical relationship among building height, solar and sensor elevation, and solar and sensor azimuth. α and β denote the solar elevation and the sensor elevation, and θ_1 and θ_2 are the solar azimuth and the sensor azimuth, respectively. Figure 1(A) and (B) illustrate the profile view of the two types of geometrical relationships between building height, sensor elevation, and solar elevation. In Figure 1(A), the sun and the sensor are located on the same side. Conversely, the sun and the sensor are located on opposite sides in Figure 1(B). In these two subgraphs, the shadow edges visible from a VHR image are the line segments L_{AB} and L_{AO_2} , respectively. Figure 1(C) illustrates the plan view of the geometrical relationships among building position, sensor azimuth, and solar azimuth. As

shown in Figure 1(A), (B), and (C), in practice, the solar and sensor elevation and the solar and sensor azimuth may affect the shadow shape of a building. Figure 1(D) illustrates the geometrical relationship in a 3D space of the solar and sensor azimuth, the solar and sensor elevation, and the building position. In this subgraph, the visible shadow edge is the line segment AO₂. Based on the geometrical relationships shown in Figure 1(D), the following subsection summarizes the previous approaches for building height estimation with shape detection.

2.2.2 State-of-The-Art Approaches.

Depending on the availability of data sources and the content of VHR images, the state-of-the-art approaches consist of three types of implementation of shadow-based building height estimation. Table 2 compares these three classes of methods.

Table 2

Three types of methods for building height estimation and shadows.

	Data			Cost	Computing complex	Automatic degree	Result precision
	Google Earth Pro (Open VHR image)	Commercial VHR image	Image metadata				
Method 1		√	√	High	Difficult	Auto	High
Method 2	√			Free	Easy	Manual	Moderate
Method 3	√	√		High	Difficult	Semi-auto	High

Method 1: Only commercial VHR images with metadata are available. Spatial resolution and solar elevation are available under this condition. The workflow of building height estimation is composed of the following steps: 1) extract building shadows from the VHR image, 2) calculate the shadow length, 3) convert the pixel–unit

shadow length into the meter–unit one, and 4) predict the building height based on the shadow length and solar elevation (Shao, Taff and Walsh, 2011; Comber et al., 2012; Liasis and Stavrou, 2016). However, VHR images may be limited in their ability to support large-scale urban areas due to data storage load and data acquisition cost.

Method 2: Only Google Earth Pro data are available. Google Earth Pro provides essential information to calculate the solar elevation of a VHR image. It was used to access solar and sensor azimuth, as well as the length of the line segments AO₂ and BO₂ in Figure 1(D). Qi and Wang (2014) and Qi, Zhai and Dang (2016) report the strategy to use the information derived from Google Earth to predict building height. As the lengths of line segments L_{AB} and L_{AO_2} shown in Figure 1(D) were manually measured by the Google Earth Pro tool, precision is a major concern in building height estimation using Google Earth.

Method 3: Both Google Earth Pro and commercial VHR imagery are available. Under this condition, Google Earth Pro provides essential parameters, including the solar and sensor elevation, the solar and sensor azimuth, and the length of line segments AO₂ and BO₂. The shadow shape derived from commercial VHR images can help to promote the precision of building height estimation.

2.3 Methods

2.3.1 Shadow Pattern Determination.

Although shadows provide useful information to characterize the building structure, the variation of their shapes poses a great challenge to determine the shadow-derived line useful in predicting building height. Therefore, I propose a shadow pattern classification system (*ShadowClass*), not only to determine the pattern category to which

a shadow area belongs, but also to draw the corresponding shadow-derived line useful for building height estimation.

2.3.1.1 *Shadow pattern classification system (ShadowClass).*

Figure 2 shows the 10 basic classes of shadow patterns and the patterns that mix multiple basic classes of shadow patterns. Gray polygons represent the shadow areas, and the red and orange dotted lines with round nodes denote the length of a shadow-derived line useful for building height estimation.

The shape of each type of shadow pattern may be influenced by a variety of factors, such as building roof, building structure, sun azimuth, sensor azimuth, and neighboring land cover, among others. Buildings of contemporary architecture, such as skyscrapers and landmarks, always comprise varying and complex structures, thus making it impossible to assign their shadow shapes to a simple pattern category. In this case, complicated shadows generally encompassed multiple basic shadow pattern classes. Therefore, *ShadowClass* provides a complex pattern to determine the shadow patterns that are combined to produce a complicated shadow pattern. For example, three complex shadow examples comprise (1) the individual patterns of Pattern 2 and Pattern 4, the individual patterns of Patterns 4 and Pattern 5, and the individual patterns of Patterns 1, Pattern 2, and Pattern 4.

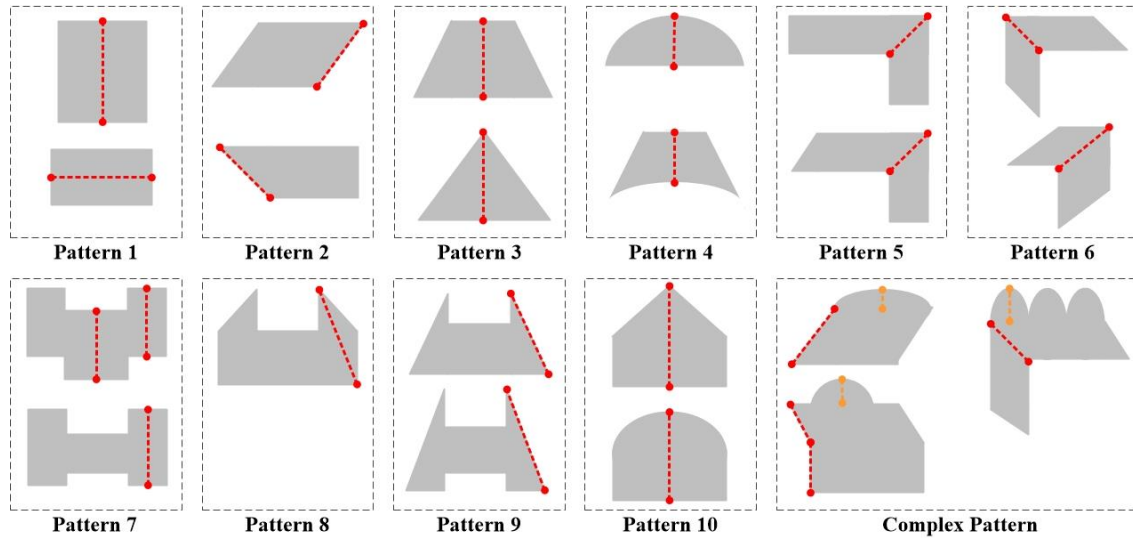


Figure 2. Illustration of the pattern categories in *ShadowClass*.

2.3.1.2 CNNs-based shadow pattern determination.

(1) Data augmentation.

A convolutional neural network (CNN)-based binary classifier was developed to perform shadow pattern determination. Shape binary classification is an important concern for computer vision analysis, the research progress of which has been reported in the last two decades (Zhang and Lu, 2004; Shen, 2015). Generally, shape classification faces several challenges, including scale variation, rotation, and affine transformation (Krizhevsky, Sutskever and Hinton, 2012). Recently, the evolution of deep neural network algorithms and powerful computing hardware has reinforced a promising way for an efficient shape binary classification. Moreover, to enhance the robustness and transferability of CNN-powered shape classification, data augmentation is needed to increase the diversity of intraclass training samples and the similarity of interclass training samples (Taylor and Nitschke, 2017; Hernández-García and König, 2018). This study applies three strategies for performing data augmentation:

Strategy 1: Flipping. Two new images are created by flipping the original training image over the horizontal and vertical dimensions.

Strategy 2: Rotation. Seventy-two new images are generated by rotating the original training image and its corresponding two flipped images every 5° , respectively.

Strategy 3: Scaling. Four new images are generated by rotating the original training image and its corresponding flipped and rotated images. Assuming that the dimensionality of an image generated is $x \times y$, then the scaled images have the dimensionality of $4x \times 4y$, $2x \times 2y$, $x/2 \times y/2$, and $x/4 \times y/4$, respectively.

The data augmentation generated an additional 876 images for every original training image (original image + 2 flipped images + $(1+2) \times 72$ rotated images + $(1+2+(1+2) \times 72) \times 4$ scaled images).

(2) CNN-reinforced shape classification.

The Inception_ResNet_V2 is a cutting-edge CNN for scene classification that achieved state-of-the-art accuracy in image scene classification in the prestigious benchmark dataset called ILSVRC (Szegedy, et al., 2017). Therefore, this CNN model was used to classify the extracted shadow areas into a pattern defined in *ShadowClass*. Inception_ResNet_V2 is a systematic neural network integrating the architecture of two CNNs, namely, Inception V3 and ResNet, to take advantage of network depth in the classification while reducing a large load in terms of time and computation. The method also controls the effects of vanishing gradient and a degradation problem. Figure 3 shows the architecture of Inception-ResNet-V2.

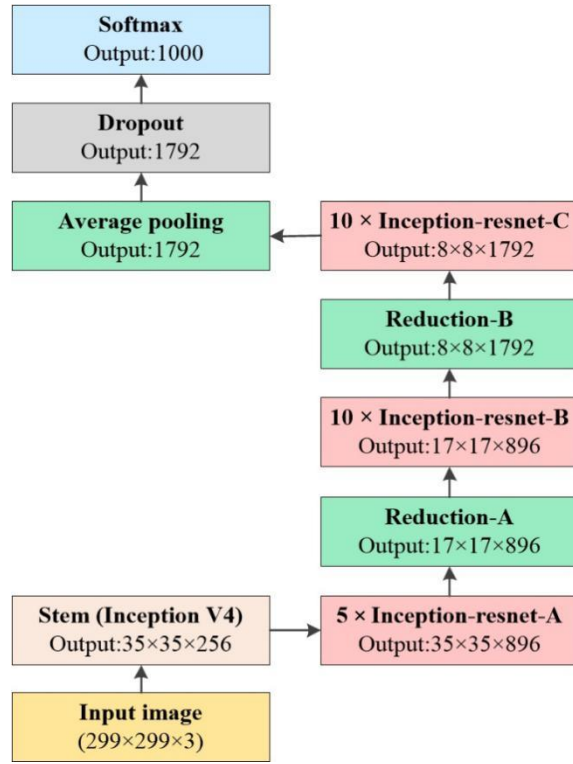


Figure 3. Illustration of Inception-resnet-V2 (Szegedy et al., 2017).

The dimensions of the input image required by Inception_ResNet_V2 are $299 \times 299 \times 3$. In the first state, Inception-ResNet-V2 processes the input image with the Stem of Inception V4 to generate a feature map with dimensions of $35 \times 35 \times 256$. The feature map is then processed consecutively by five independent blocks of Inception-ResNet-A and compressed by the block of Reduction-A. The results obtained by Reduction-A are further processed consecutively by 10 independent blocks of Inception-ResNet-B and compressed by the block of Reduction-B. The results generated by Reduction-B are processed consecutively by 10 independent blocks of Inception-ResNet-C and compressed by average pooling. Furthermore, the 1,792 features generated by average pooling are processed with Dropout. Finally, the Softmax classifier produces the top three scores of shadow pattern classes to which every input image belongs.

2.3.2 Height Calculation with Google Earth Pro.

2.3.2.1 Solar declination.

The angle between the ground surface and sunlight varies in different places on Earth because it is a spherical planet. Solar declination is the angle that specifies the solar position between the incident orientation of sunlight and the Earth's equator. Coper (1973) and Bourges (1985) proposed algorithms to calculate solar declination considering the date when the corresponding satellite image was produced. The algorithm is expressed as follows:

$$\omega = 0.3723 + 23.2567 \sin \delta + 0.1149 \sin 2\delta - 0.1712 \sin 3\delta - 0.758 \cos \delta + 0.3656 \cos 2\delta + 0.0201 \cos 3\delta \quad (1)$$

where δ is calculated by the following equation:

$$\begin{cases} \delta = \frac{360(y-y_0-0.5)}{365.2422} \\ y_0 = 78.801 + 0.2422(Y - 1969) - \text{int}(0.25(Y - 1969)) \end{cases} \quad (2)$$

where Y is the year when the VHR image was created, and y is the n th day of Y .

2.3.2.2 Solar elevation.

Solar elevation, which is the angle β in Figure 1(D), specifies the angle of sunlight over the horizontal dimension on the ground surface. Solar elevation is calculated by the following equation:

$$\beta = \arcsin(\sin \sigma \sin \delta + \cos \sigma \cos \delta \cos \phi) \quad (3)$$

where σ is the latitude of a building, and ϕ is the solar hour angle, which is obtained by the following equation:

$$\phi = \begin{cases} \min(\arccos\left(\frac{-b+\sqrt{b^2-4ac}}{2a}\right), \arccos\left(\frac{-b-\sqrt{b^2-4ac}}{2a}\right)) \\ -\min(\arccos\left(\frac{-b+\sqrt{b^2-4ac}}{2a}\right), \arccos\left(\frac{-b-\sqrt{b^2-4ac}}{2a}\right)) \end{cases} \quad (4)$$

where $\min()$ and $-\min()$ are used in the morning and in the afternoon, respectively.

Moreover, a , b and c in Equation (4) are expressed as follows:

$$\begin{cases} a = (\tan \theta_2)^2 (\sin \sigma)^2 + 1 \\ b = -\sin 2\sigma \tan \delta (\tan \theta_2)^2 \\ c = (\tan \theta_2)^2 (\sin \sigma)^2 (\tan \theta_2)^2 - 1 \end{cases} \quad (5)$$

where θ_2 is the sensor azimuth, as shown in Figure 1(D). θ_2 can be accessed by the Ruler tool in Google Earth Pro. The details of the sensor azimuth are presented in the following subsection.

2.3.2.3 Solar and sensor azimuth

The solar (sun) and sensor (satellite) azimuth, solar and sensor elevation, and building position are crucial concerns associated with the geometrical relationship between building height and shadow length. As shown in Figure 1(A), when the sun and the sensor are on the same side, line segment AB is the shadow visible from the VHR image. The length of this line segment is measured by the following equation:

$$L_{AB} = \frac{\tan \alpha \tan \beta}{\tan \alpha - \tan \beta} \times L_{AB} \quad (6)$$

As shown in Figure 1(B), when the sun and the sensor are on opposite sides, line segment AO₂ is the shadow visible from the VHR image. The length of this line segment is measured by the following equation:

$$L_{AO_2} = \tan \beta \times L_{AC} \quad (7)$$

Without metadata defining the position and the altitude of the satellite, the sensor elevation α cannot be specified in Google Earth. The solar elevation (β) can be obtained by Equation (3). Moreover, sensor azimuth (θ_1) and solar azimuth (θ_2) are accessed by Google Earth Pro, as shown in Figure 4.

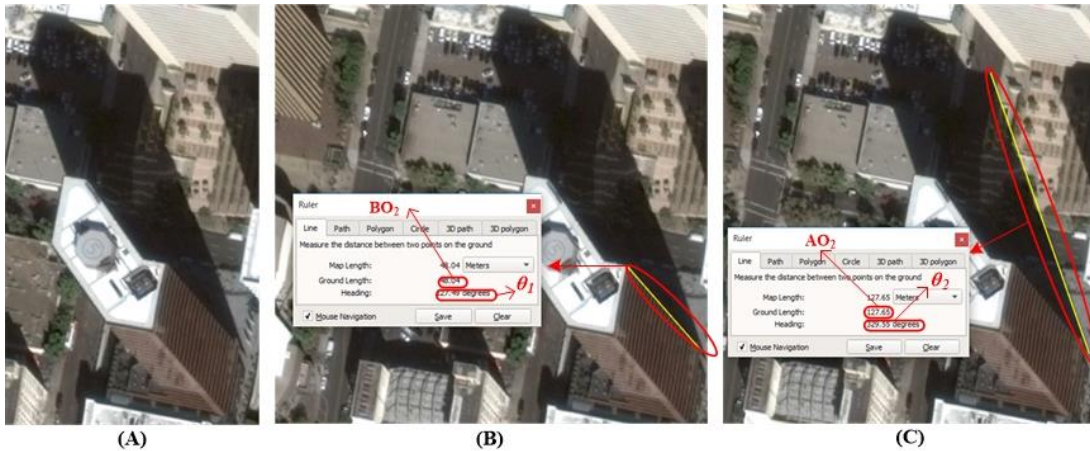


Figure 4. Illustration of accessing the solar and the sensor azimuth from Google Earth Pro (Qi and Wang, 2014; Qi, Zhai and Dang, 2016). (A) Original building image. (B) Accessing the length of BO_2 and the sensor azimuth (θ_2) using the Ruler tool. (C) Accessing the length of AO_2 and the solar azimuth using the Ruler tool (θ_1).

Figure 4(A) shows a building in a Google Earth satellite image. Based on this image, Qi and Wang (2014) and Qi, Zhai and Dang (2016) reported an approach to access the essential parameters useful for building height estimation using the Ruler tool in Google Earth Pro. Figure 4(B) illustrates the Ground Length and the Heading of the yellow line drawn on the satellite image corresponding to the sensor azimuth and the length of BO_2 in Figure 1(D), respectively. Figure 4(C) shows the Ground Length and the Heading of another yellow line drawn on the satellite image corresponding to the solar azimuth and the length of AO_2 in Figure 1(D), respectively..

When receiving the values of the sensor and the solar azimuth and the lengths of AO₂ and BO₂, I calculate the building height using the approaches presented in Subsection 3.3.4.

2.3.3 Height Calculation with Shadow Area Extraction

The proposed approach for building height estimation with shadow area extraction consists of four sections. The first section converts an RGB color image into a grayscale one and then performs image enhancement using an algorithm called adaptive gamma correction with weighted distribution (AGCWD) (Huang, Cheng and Chiu, 2013; Rahman et al., 2016). Grayscale conversion and contrast enhancement have been reported to make shadows distinct from other land cover features (Liu, Fang and Li, 2011; Liasis and Stavrou, 2016), facilitating efficient shadow extraction. The second section exploits an algorithm called simple linear iterative clustering (SLIC) to segment the image preprocessed by the previous section into multiple superpixels and then classifies these superpixels into shadow and non-shadow areas. The third section detects and simplifies the contour of each shadow area. The last section calculates the shadow length defined in ShadowClass. I predict the building height by considering the shadow length and the relationship among the solar position, sensor position, and building position.

2.3.3.1 Image pre-processing.

AGCWD enhances the contrast of an image by dynamically applying the parameters derived from the whole image content (Huang, Cheng and Chiu, 2013). In this study, I use the AGCWD to process the contrast of VHR images to draw shadow areas distinct from other land covers through three steps (Huang, Cheng and Chiu, 2013): histogram analysis, weighting distribution adjustment, and gamma correction.

The cumulative density function (cdf) and the probability density function ($pdf(i)$) of a VHR image are expressed as follows:

$$\begin{cases} pdf(i) = num_i/num_{all} \\ cdf = \sum_{i_{min}}^{i_{max}} pdf(i) \end{cases} \quad (8)$$

where num_i is the frequency of intensity i , num_{all} is the total number of an image and i_{min} and i_{max} are the maximal and minimal intensity in this VHR image, respectively.

Based on the cumulative density function and the probability density function in Equation (8), the adaptive gamma correction (AGC) converts the original intensity i into a new value i_{agc} by the following expression, where pdf_{max} is the maximal probability density function,

$$i_{agc} = pdf_{max} \times \left(\frac{pdf(i)}{pdf_{max}} \right)^{1-cdf} \quad (9)$$

To further correct the result of the histogram analysis in AGC, AGCWD uses a weighting distribution function to process the probability density function and the cumulative density function expressed in Equation (8). The probability density function with weighting distribution (pdf_{wd}) is expressed as follows:

$$pdf_{wd} = pdf_{max} \times (pdf(i)_{norm})^\sigma \quad (10)$$

where σ is the user-defined parameter to control the distribution of histogram statistics, and $pdf(i)_{norm}$ is the normalized $pdf(i)$. Accordingly, the cumulative density function with weighting distribution (cdf_{wd}) is expressed as follows:

$$cdf_{wd} = \sum_{i_{min}}^{i_{max}} (pdf_{wd}(i) \times \sum_{i_{min}}^{i_{max}} pdf(i)) \quad (11)$$

where C_{low} and C_{high} are the low-contrast or high (or moderate)-contrast image, respectively, and τ is the threshold used for a binary contrast classification.

Based on Equations (10) and (11), the original intensity i in the VHR image becomes i_{agcwd} after AGCWD by the following expression:

$$i_{agcwd} = pdf_{max} \times \left(\frac{pdf_{wd}(i)}{pdf_{max}} \right)^{1-cdf_{wd}} \quad (12)$$

2.3.3.2 Shadow area extraction

(1) Superpixel-based segmentation

I apply a popular object-based segmentation called SLIC (Achanta et al., 2012) to segment the input contrast-enhanced VHR image. Compared with the graph-based segmentation approaches, superpixel-based segmentation such as SLIC is more efficient to group the connected pixels into meaningful sub-regions. Moreover, SLIC speeds up the process of clustering multiple pixels by measuring the distance over spatial space and intensity (color) differences between.... In SLIC, the image space, including intensity and spatial space, is represented as (L, A, B, X, Y) , where L , A , and B denote the three channels of image color space, and X and Y denote the distance over the horizontal and vertical dimensions, respectively. As the image enhanced by AGCWD only contains one channel, intensity space ($D_{intensity}$) and spatial space ($D_{spatial}$) in the enhanced VHR image are represented as (I, X, Y) , where I is the intensity of one channel. Then, every pixel in an image joins the nearest cluster center pixel. The “nearest” is measured by the distance of image space, which is expressed as follows:

$$D_{total} = D_{intensity} + \frac{\theta}{\sqrt{N}} \times D_{spatial} \quad (13)$$

where θ is the ratio between spatial distance and intensity difference. A higher θ generates a result that contains superpixels within a larger size, and vice versa. N denotes the approximate number of superpixels after segmentation. Moreover, for a pixel located at position (x_0, y_0) , the image gradient is computed by the L_2 norm, which is shown in the following equation:

$$g(x_0, y_0) = L_2(I(x_0 + 1, y_0) - I(x_0 - 1, y_0))^2 + L_2(I(x_0, y_0 + 1) - I(x_0, y_0 - 1)) \quad (14)$$

where $L_2()$ is the L_2 norm, and $I(x, y)$ is the intensity vector of a pixel at the coordinate (x, y) .

Assuming that the result of segmentation is $Seg_{num} \ni \{s_1, s_2, \dots, s_{num}\}$, where num is the number of superpixels or segmented image regions, s_i is the i th superpixel. I calculate the histogram for each superpixel to obtain the result: $Hist_{num} \ni \{h_1, h_2, \dots, h_{num}\}$, where h_i is the histogram corresponding to s_i . Then, I generate a feature vector including the density of every bin to determine whether a superpixel belongs to the shadow class. Finally, I fuse the shadow superpixels into a new image.

(2) Vegetable area removal

In a VHR image covering urban areas, the shadow of trees may overlap with artificial architectures. Therefore, I attempt to detect trees from the VHR image and remove their shadows. Whereas the normalized difference vegetation index is a popular parameter to determine whether a pixel contains plant content, a near-infrared waveband is not available in the majority of VHR images. Therefore, I apply an algorithm called the triangular greenness index (TGI) to detect vegetated areas with RGB channels (Hunt, 2013). Equation (15) shows the TGI value of a pixel:

$$TGI = w_{gree} - 0.39 \times w_{red} - 0.61 \times w_{blue} \quad (15)$$

where w_{gree} , w_{red} , and w_{blue} refer to the intensity of the green, red, and blue wavelengths of a pixel, respectively.

Then, I set a threshold to select the pixels with a high TGI value and remove these pixels and their connected shadow areas from the original result of the shadow area extraction.

2.3.3.3 Contour detection and edge simplification

The result of shadow extraction is a binary image that assigns every pixel to two values: shadow or non-shadow. However, the shadow area extracted from a VHR image encompasses rough edges. Moreover, the shadow area may carry complicated shapes mainly because of the structure of a building roof and the neighboring land cover on the ground surface. Therefore, raw shadow areas fail to precisely characterize the form of a building body, let alone to model the relationship between shadow shape and building height. To address the challenges mentioned above, this study performs contour detection and edge simplification to smooth and straighten the rough shadow edges.

I use the marching squares algorithm (MSA) to detect the contour from the rough shadow edges. MSA aims to find the edge in every 2×2 pixels' window based on a two-dimensional image array. Figure 5(A) shows all 16 possible configurations possibly observed in the 2×2 pixels' window. Each circle denotes a pixel that only has either of two binary values, namely, shadow and non-shadow, which are represented by black and white, respectively. The red line is the edge interpolated in the 2×2 pixels' window to divide the shadow and the non-shadow area, that is, the edge of the shadow area. Except for cases 1 and 2 in which no edge exists, an edge can be drawn from other cases. I use

all the 16 cases to generate an approximate contour from the rough shadow edges, maintaining the trade-off between the number of vertices remaining and the similarity of the approximate contour and the original shadow shape.

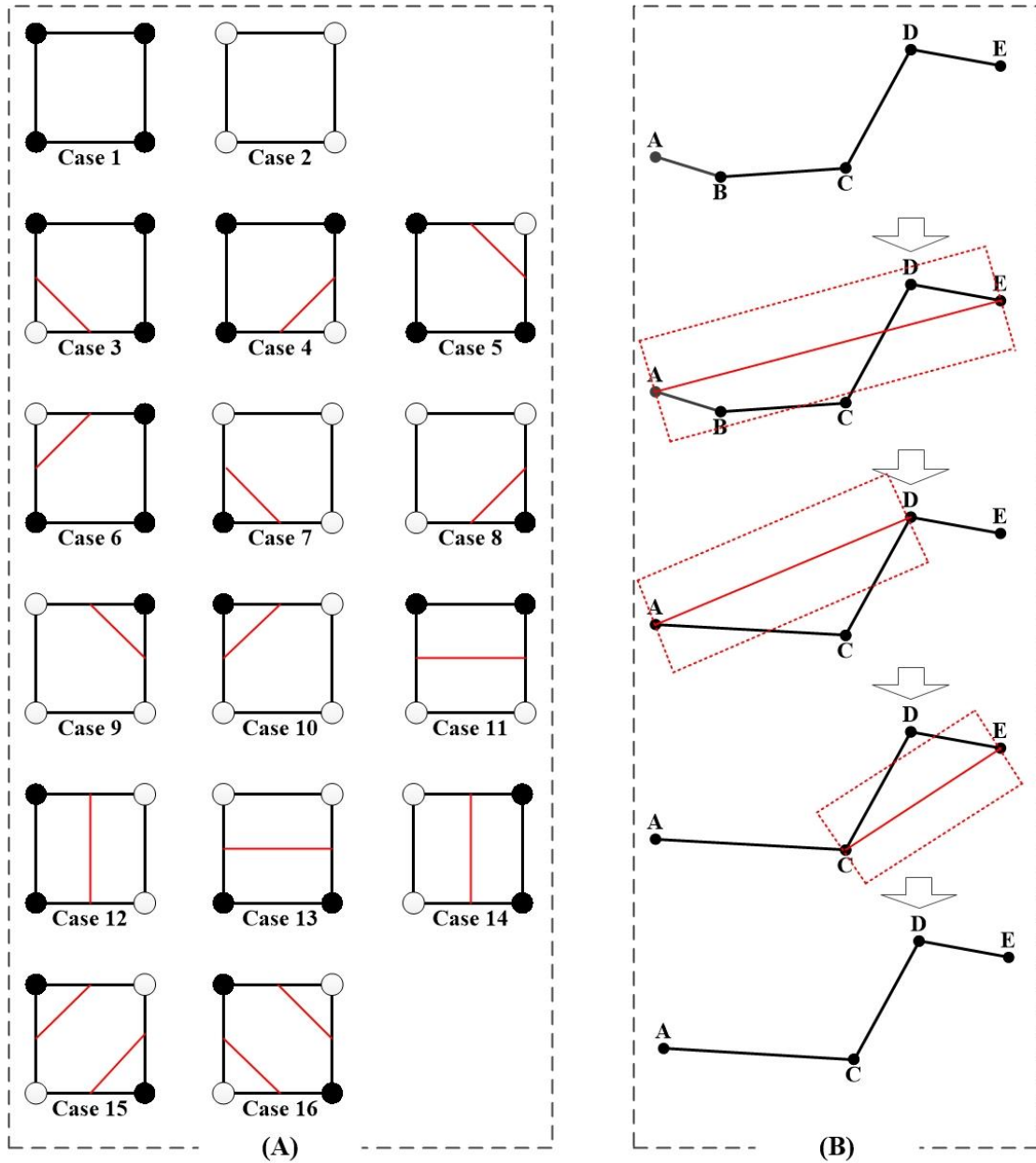


Figure 5. Illustration of (A) the 16 possible configurations of the MSA, and (B) the sample workflow of the Ramer–Douglas–Peucker algorithm.

I then apply the Ramer–Douglas–Peucker algorithm to simplify the contours generated by the MSA to represent the shadow shape with approximate line segments containing fewer vertices. I assume that I have an original curve $\{v_1, v_2, \dots, v_k\}$, where v_k is the sequentially numbered vertex in this curve. The Ramer–Douglas–Peucker algorithm simplifies this curve using the following steps:

Step 1. Create a line segment connecting the starting point v_1 and the ending point v_k and then define the distance to generate a buffer zone around the line segment. The buffer size is given based on specific applications.

Step 2. Remove all vertices inside the buffer zone. The remaining vertices are $\{v_1, v'_1, v'_2, \dots, v_k\}$.

Step 3. Create a new line segment connecting v_1 and v'_2 and then define the distance to generate a buffer zone around the line segment.

Step 4. Examine whether v'_2 is located in the buffer zone, and remove or maintain v'_2 by following the rules defined in Step 2.

Step 5. Repeat Steps 3 and 4 using other vertices until the line segment is connected to the end point.

Figure 5(B) shows an example of how this algorithm works. I have an original shadow edge A-B-C-D-E. The first step creates a line segment connecting point A and point E and generates a buffer zone around the line segment AE. Following the rule defined in Step 2, I remove point B from the line segment AE as this point is located inside the buffer zone. The second step creates a new line segment connecting point A and point D and generates a buffer zone around the line segment AD. In this case, I

maintain point C as it is outside of the buffer zone. I then create a new line segment connecting point C and point E and generate a buffer zone around the line segment CE. In this case, I still maintain point D. Overall, the Ramer–Douglas–Peucker algorithm simplifies the original shadow edge to a new line: A-C-D-E.

2.3.4 Building Height Estimation.

(1) Building height estimation using Google Earth Pro

When the building wall is visible in the VHR image:

As shown in Figure 4(B), I can obtain the solar azimuth and the length of BO_2 when the building wall is visible in the VHR image. On the basis of the geometrical relationship shown in Figure 4(D), I have two approaches to calculate the building height. The first approach uses solar elevation, which was introduced in Equation (3), and the length of AO_2 , which is expressed as follows:

$$H = L_{AO_2} \tan \beta \quad (16)$$

where H is the building height.

Another approach uses the solar and sensor azimuth and the length of AO_2 and BO_2 . The following equations are based on the geometrical relationship shown in Figure 4(D):

$$\left\{ \begin{array}{l} L_{AO_2} = \frac{H}{\tan \beta} \\ L_{BO_2} = \frac{H}{\tan \alpha} \\ L_{AB}^2 = L_{AO_2}^2 + L_{BO_2}^2 - 2L_{AO_2}L_{BO_2} \cos \theta \end{array} \right. \quad (17)$$

In Equation (17), $\tan \alpha$ and L_{AB} are the unknown parameters. By consolidating the three expressions in Equation (17), building height is expressed as follows:

$$H = \tan\beta(L_{BO_2}\cos\theta + \sqrt{L_{AO_2}^2 + L_{BO_2}^2 - 2L_{AO_2}L_{BO_2}\cos\theta + L_{BO_2}^2\sin^2\theta}) \quad (18)$$

When the building wall is not visible in the VHR image:

As shown in Figure 4(B), the solar azimuth and the length of BO_2 are not available when the building wall is invisible in the remote sensing image. Therefore, the expression in Equation (16) is the only approach that can be used to calculate building height from shadow length.

(2) Building height estimation using shadow area extraction

Assume that the length of an extracted shadow area is p_{num} pixels in the VHR image and that the spatial resolution is sr . The length of this shadow (L_{AO_2}) is computed by the following equation:

$$L_{AO_2} = p_{num} \times sr \quad (19)$$

Then, I substitute the L_{AO_2} obtained by Equation (19) into Equation (16) to obtain the height of the building associated with this shadow.

2.4 Experiments

This section collects a number of VHR images to test the performance of the two approaches for building height estimation: 1) building height estimation using Google Earth Pro and 2) building height estimation using shadow extraction. Subsection 3.4.1 introduces the dataset used for the experiment. Subsection 3.4.2 compares the results of shadow extraction by various methods, as the shadow extraction result is critical to the precision of building height estimation. Subsection 3.4.3 presents the results of building height estimation using the two approaches.

2.4.1 Experimental dataset

The experimental dataset consisted of training images included in the *ShadowClass* and test VHR images. First, I created five original binary images corresponding to each basic pattern of the *ShadowClass* shown in Figure 2 and prepared 43,300 training samples, including the original images and those generated by data augmentation. I then collected 18 test images covering Los Angeles, San Diego, and Las Vegas from Google Earth Pro (Figure 6). The shadows in these collected test images varied in scale, size, orientation, and shape and were located in different landscape scenario and contexts, including downtown, dense residential, sparse residential, and industrial areas.

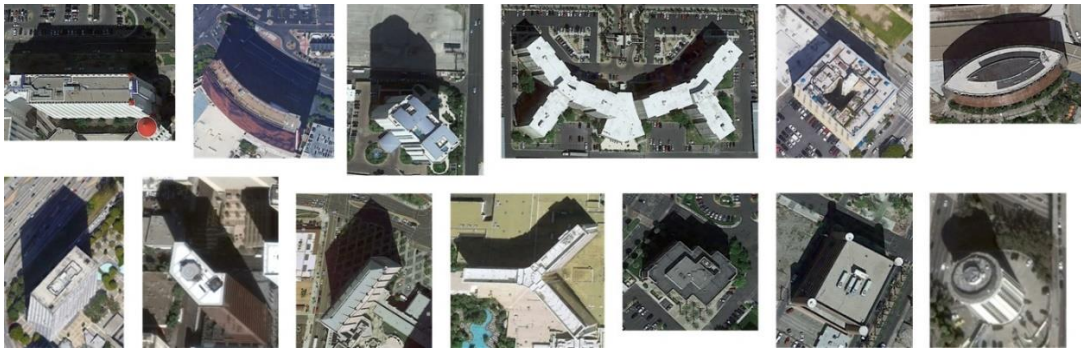


Figure 6. *Illustration of the selected test VHR images.*

2.4.2 Shadow Extraction Results.

As mentioned above, the quality of the resulting shadow extraction plays a decisive role in building height estimation. Figure 7 compares the image enhancement results with a Gaussian filter, histogram equalization, traditional AGC, and AGCWD. The result generated by histogram equalization could not support shadow extraction because the objects were obviously visible in the shadow areas. Even some parts of the shadow areas seemed brighter than the other land cover surrounding them. The Gabor

filter outperformed histogram equalization in differentiating between shadows and other land covers. However, the distribution of grayscale, or the intensity histogram, seemed imbalanced in the results produced by the Gabor filter. This result might pose a challenge to illuminating the shadow areas to make them more distinct from other dark land cover such as asphalt. Both AGC and AGCWD generated a new image, making the shadow areas much more distinguishable. Generally, only a few differences were observed in the results processed by AGC and AGCWD. In some cases, the shadow areas in the results generated by AGCWD remained the intensity of high frequencies more complete. Reported as a state-of-the-art technique for image contrast processing, AGCWD can effectively be used to generate an image-making shadow and other land cover distinguishable (Liu, Fang and Li, 2011; Liasis and Stavrou, 2016).

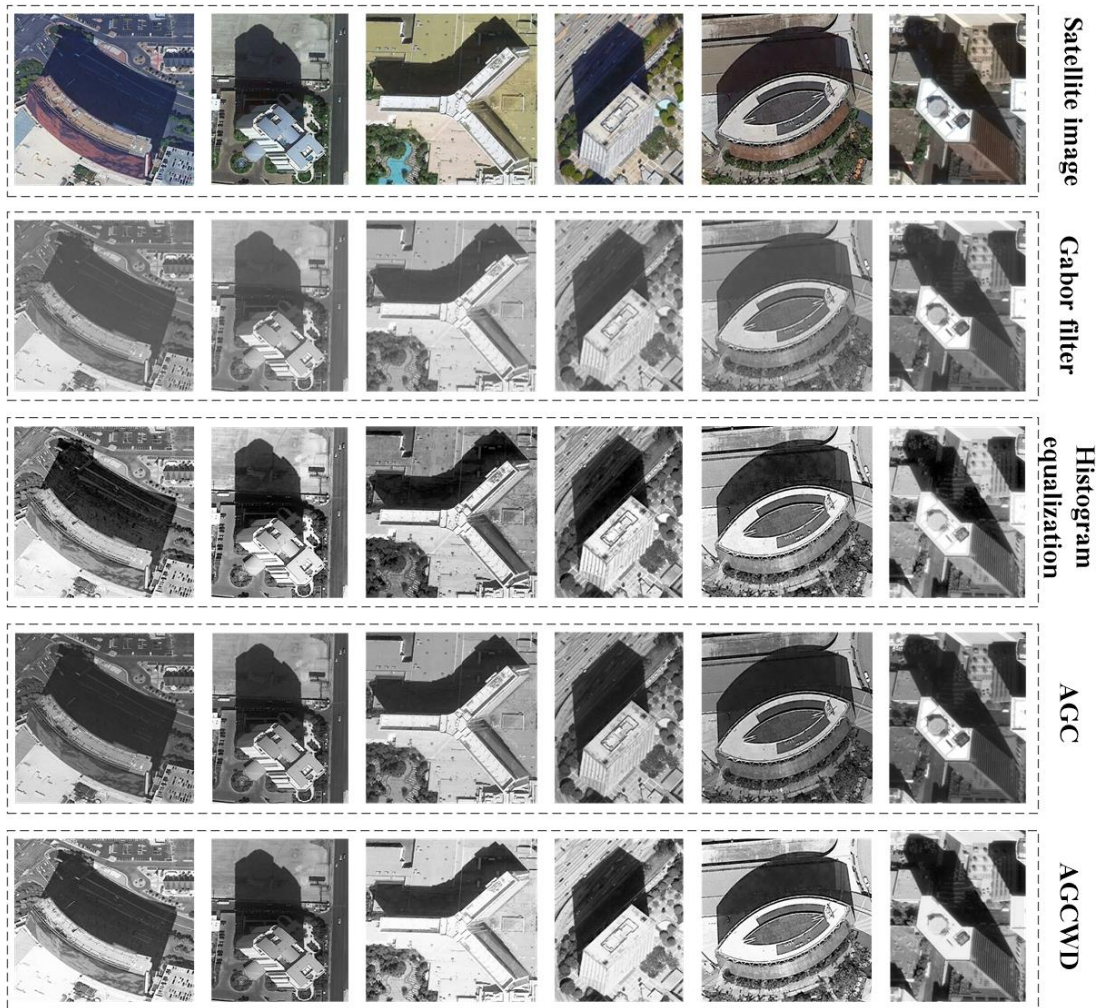


Figure 7. Illustration of buildings visualized by VHR images and image processing results by the Gabor filter, histogram equalization, traditional AGC, and AGCWD (Liu, Fang and Li, 2011; Liasis and Stavrou, 2016).

Figure 8 shows the results of raw shadow extraction, shadow contour detection, and shadow area extraction after polygon simplification of the image processed by AGCWD. As the VHR image presented the detailed shapes of the majority of land covers, the results of the raw shadow extraction contained tree shadows, roads, and other dark land covers within the RGB channels. Moreover, a variety of objects visible in the shadow areas led to an incomplete and fragmented extracted shadow. These two factors

caused the raw shadow extraction to appear with salt and pepper noises and stripe noises. Raw shadow extraction from VHR images could not obtain a precise shadow area for building height estimation.

I removed the tree shadows using the method expressed in Equation (8), along with other kinds of noise. As roads can connect to building shadow areas in some cases, I had to visually remove the dark area belonging to roads. I then detected the primary contour associated with each shadow area. Although the results of contour detection created concise shapes of shadows, these shapes still carried rough edges, posing a challenge in precisely measuring the length of a line segment.

Thus, I further simplified the shadow shapes using the Ramer–Douglas–Peucker algorithm. Comparing the results of shadow contour extraction and those generated by shadow polygon simplification, many rough shadow edges were smoothed and became straight. Straight lines useful for length calculation were available in the results of shadow polygon simplification.

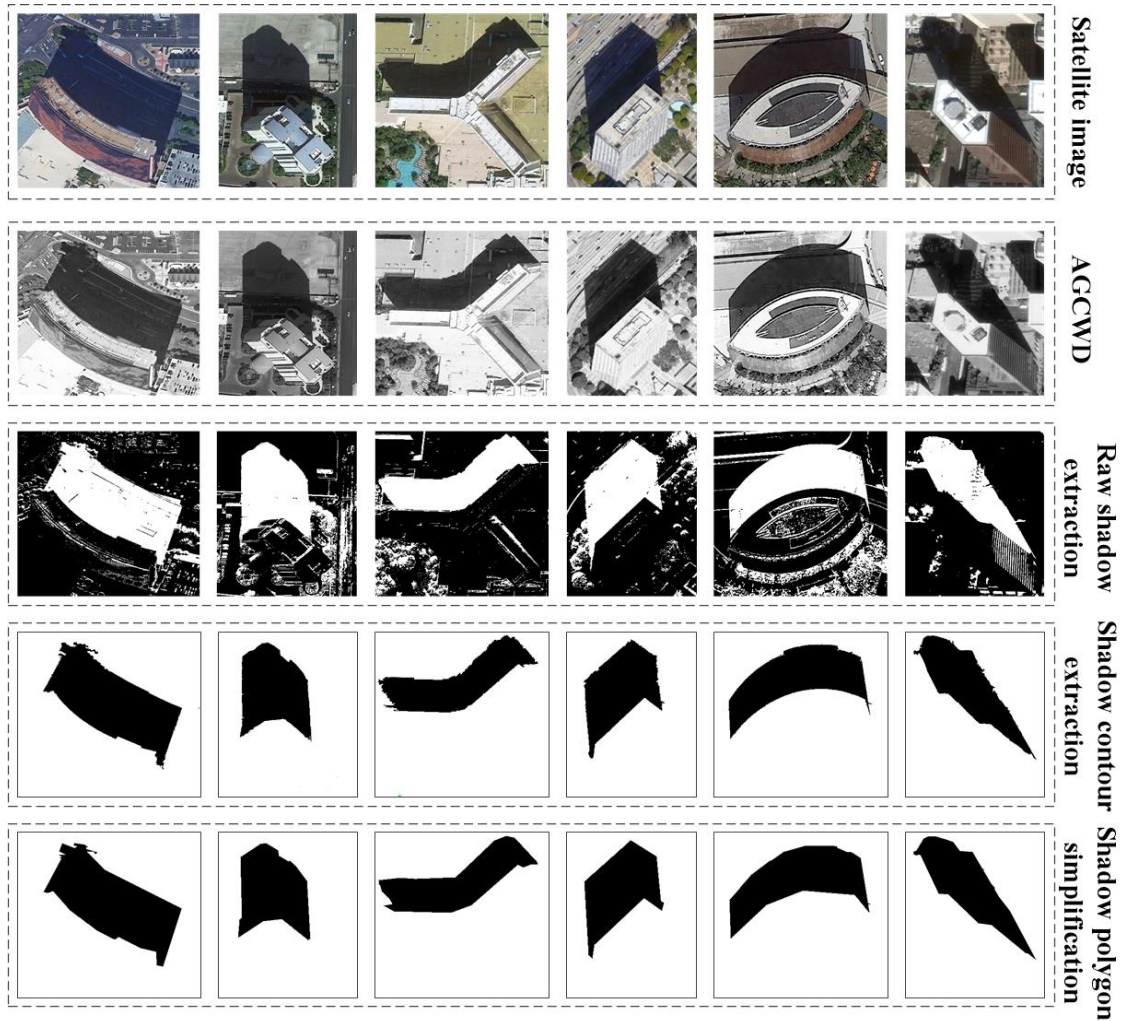


Figure 8. *Illustration of buildings visualized by VHR images, images enhanced by AGCWD, and results of raw shadow extraction, shadow contour extraction, and shadow polygon simplification.*

2.4.3 Accessing Parameters from Google Earth Pro.

I predicted the building height from the collected test VHR image using two approaches. The first approach performed building height estimation using Google Earth Pro. I accessed the solar and sun azimuth, image date, geographical coordinates, and altitude using the Ruler tool in Google Earth Pro and measured the length of a selected line segment useful for building height estimation on Google Earth Pro. I applied the

methods expressed in Subsection 3.2 to obtain the building height values. Figure 9 shows the line segment I selected from Google Earth for building height estimation. The red line is the length of BO_2 and the solar azimuth (θ_2) shown in Figure 4(B), and the yellow line is the length of AO_2 and the sensor azimuth measured with the Ruler tool (θ_1) shown in Figure 4(C).



Figure 9. Illustration of buildings visualized by the VHR image, the line segment for accessing the length of AO_2 and the sensor azimuth (θ_1) shown in Figure 4(C), and the length of BO_2 and the solar azimuth (θ_2) shown in Figure 4(C).

The second approach focuses on predicting building height with the extracted shadows. Considering the cost of accessing VHR images, I used the altitude and the width of roads to estimate the spatial resolution of every test VHR image derived from Google Earth Pro. On the basis of the results of shadow polygon simplification, I fine-tuned an ImageNet-Pretrained Inception_ResNet_V2 model, which was accessed from the Tensorflow Github repository, with the dataset prepared in *ShadowClass*. I used this

fine-tuned CNN model to classify every simplified shadow polygon into a predefined basic shadow pattern of *ShadowClass*.

Figure 10 shows the clarification result of the shadow pattern for the selected buildings shown in Figure 9. In the shadow pattern demonstrations and simplified shadow polygons, I drew the position of the line segment from which I calculated the length for building height estimation. The gray polygons refer to the basic pattern in *ShadowClass*, and the red dotted lines denote the position of the line segment used for building height estimation. I then calculated the length of these line segments using the method expressed in Equations (19) and (16).

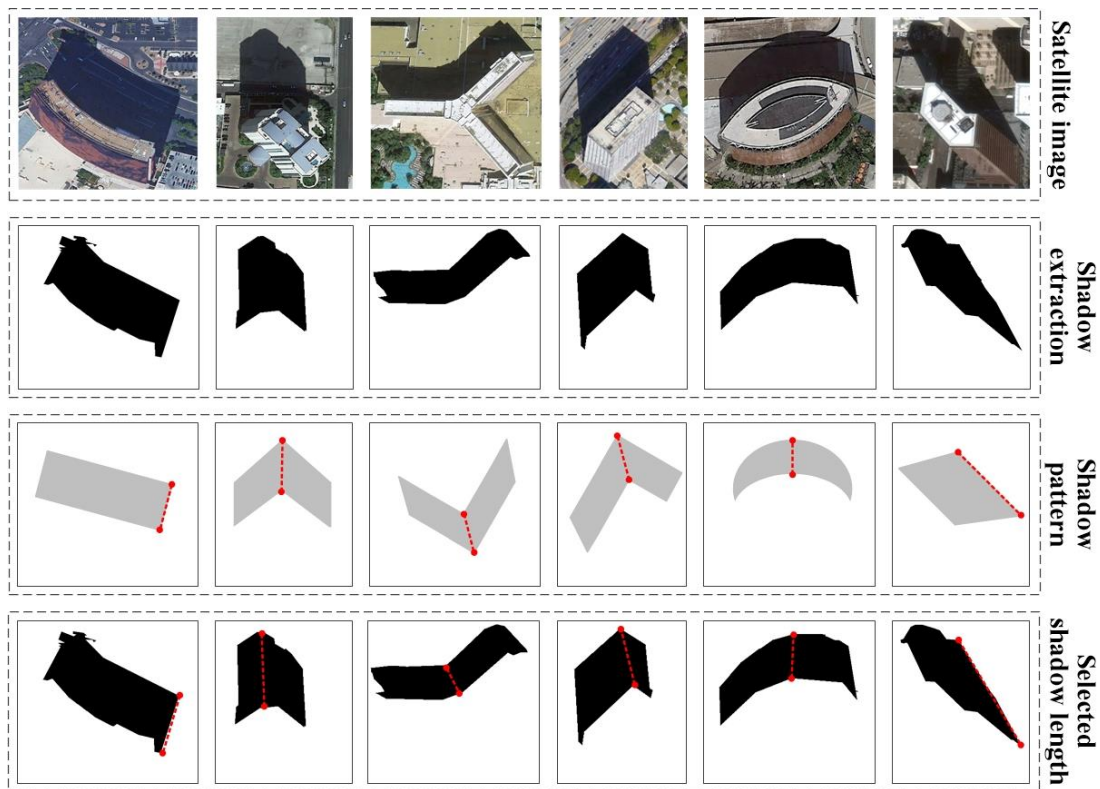


Figure 10. Illustration of the buildings visualized by VHR images, the results of shadow extraction, the shadow pattern to which an extracted shadow was assigned, and the selected shadow length for building height estimation.

2.4.4 Results of Building Height Estimation.

Table 3 lists the detailed information related to the testing of VHR images accessed from Google Earth Pro, including image data, city where the building is located, approximate geographical coordinates, and altitude. The 3D Buildings layer in Google Earth Pro enables the measurement of the height of buildings through the 3D path function of the Ruler. For buildings lower than around eight floors, for which the corresponding 3D models were not created, I used the Google street view photos to measure their height and compared the results of ground truth building height to predictive building height.

Method 1: building height estimation using Google Earth Pro

Method 2: building height estimation using shadow area extraction

Table 3

Building height estimation using two different methods

Buildings	Image date	City	Approximate geographical coordinate	Altitude	Ground truth height	Estimated height (Method 1)	Estimated height (Method 2)
> 10 floors							
Building 1	11/18/2017	LV*	(36°05'59.91"N, 115°10'28.25"W)	479m	75m	72.8±5m	73.5m
Building 2	11/18/2017	LV	(36°06'37.41"N, 115°10'08.70"W)	677m	192m	200.1±5m	190.8m
Building 3	11/18/2017	LV	(36°08'16.95"N, 115°09'16.68"W)	481m	145m	139.5±5m	147.1m
Building 4	9/3/2016	LV	(36°07'22.54"N, 115°10'27.46"W)	764m	90m	90.5±5m	88m
Building 5	10/19/2016	LA* *	(34°03'14.53"N, 118°15'21.09"W)	546m	148m	147.2±5m	148.4m
Building 6	3/23/2016	SD* **	(32°42'37.48"N, 117°10'07.09"W)	314m	135m	132.8±5m	133.5m
Building 7	2/11/2015	SD	(32°43'05.26"N, 117°09'37.08"W)	321m	100m	105.4±5m	96.2m
Building 8	4/15/2015	SD	(32°42'31.06"N, 117°09'54.82"W)	276m	110m	108.9±5m	109m

5 - 10 floors							
Building 1	3/28/2017	LA	(34°03'04.41"N, 118°14'40.21"W)	213m	42m	40.5±5m	40.3m
Building 2	4/15/2015	SD	(32°42'34.93"N, 117°12'59.64"W)	452m	37m	37.2±5m	42.5m
Building 3	4/15/2015	SD	(32°43'04.69"N, 117°09'37.37"W)	309m	27m	30.1±5m	25.3m
Building 4	3/23/2015	LV	(36°09'50.33"N, 115°08'40.56"W)	358m	32m	31±5m	33.8m
Building 5	3/23/2015	LV	(36°07'03.52"N, 115°09'24.84"W)	346m	38m	36.5±5m	40m
Building 6	3/25/2014	LV	(36°06'48.99"N, 115°08'22.29"W)	178m	26m	27.9±5m	27.2m
1 - 5 floors							
Building 1	8/14/2018	SD	(32°41'54.77", 117°07'33.86")	180m	—	—	3.6m
Building 2	3/25/2014	LA	(33°53'20.98", 118°09'33.87")	104m	—	—	4m
Building 3	3/16/2015	LV	(36°06'19.96", 115°06'31.95")	108m	—	—	3.2m
Building 4	10/19/2016	LV	(36°06'09.19", 115°07'14.10")	93m	—	—	5.5m

LV*: Las Vegas.

LA**: Los Angeles.

SD***: San Diego.

From the Google street view photos, I found the precise prediction of the height of buildings lower than five floors to be difficult. Moreover, the building wall or the line segment BO₂ was generally invisible from the VHR image. Thus, I only provided the results of building height estimation with shadow area extraction. As shown in Table 3, the errors produced by the first approach were mainly from the position of the line segment selected from Google Earth Pro. A tiny offset in the line selection would lead to a great difference in the result of building height estimation. Conversely, imprecise shadow area extraction mainly resulted in the errors observed in the results generated by the second approach. Moreover, the objects touching a shadow with similar intensity accounted for the major challenge in precise shadow area extraction. However, the offset

of the shadow area had less influence on the final product of building height estimation, as pixel length was relatively small in the VHR image.

As shown in Table 3, the results of building height estimation were close to the ground truth values. Without the support of elevation products (e.g., LiDAR, DTM, etc.), the shadow in both oblique and orthorectified VHR images could support the height prediction for low-, mid-, and high-floor buildings. The results in Table 3 also justify the conclusion that shadow-based building height estimation has good transferability in predicting the height of various types of buildings, such as apartments, houses, stores, tank, and skyscrapers. The results in Table 2 and the illustrations in Figure 10 confirm the practicality of the shadow patterns defined in *ShadowClass* in dealing with a variety of shadows of different shapes, orientations, scales, and sizes to calculate their length for building height prediction.

2.5 Conclusions

Shadows visible in a VHR image have been discovered to offer an economic solution to support large-scale building height estimation. Previous work proposed a number of approaches to represent the geometrical relationship between building positions, shadow shapes, and the sun and solar positions. Given the overpriced VHR imagery products used for large-scale urban areas, open VHR images available from Google Earth Pro can provide a potential data resource for investigating shadow-based building height estimation. Moreover, previous methods that use shadow to support building height estimation performed well only in specific data sources and image conditions. The information required by these approaches, such as the sensor and solar azimuth, may not be available in some VHR images. Therefore, a methodological

framework that provides various solutions according to the availability of data sources is essential to promote shadow-based building height estimation.

This study provides two approaches for shadow-based building height estimation: building height estimation using Google Earth Pro and building height estimation using shadow area extraction. The approach using Google Earth Pro focuses on the use of open data when metadata (e.g., spatial resolution) are not available. This approach is a low-cost, quick strategy for updating building height attributes in a large urban area. However, the precision of prediction results may vary. By contrast, the approach using shadow area extraction focuses on using commercial VHR imagery to produce precise building height information. However, this strategy is expensive and time consuming for a very large urban area.

This study also proposes a classification system called *ShadowClass* to categorize building shadow patterns. The patterns defined in *ShadowClass* are valuable in determining the shadow length useful for building height estimation. In the future, efforts in accurate shadow extraction from VHR images can be valuable. Moreover, the framework that integrates state-of-the-art CNNs into the process of shadow extraction and building height estimation is worthy of considerable attention.

CHAPTER 3

INTEGRATING DEEP LEARNING AND SEMANTIC ANALYSIS TO SUPPORT HUMAN-LEVEL DIGITAL MAP RECOGNITION

3.1 Introduction

A map is an essential medium for providing symbolic representation and geographical information about the characteristics of a place in terms of georeferenced location, distribution of patterns over space, the configuration of cultural and natural elements, and the relationships between a variety of objects, areas, and phenomena. In comparison with other georeferenced data—including remote sensing imagery and LiDAR data, which have been gaining popularity—the benefits of maps have been identified in many geospatial interpretations, analyses, visualizations, and communications (Crampton, 2001; Perkins, 2003a; Monmonier, 2006; Konecny, 2011). Over the last two decades, the progress of surveying, mapping, and web-service techniques have facilitated a significant portion of the efficiency of map generation and map sharing. Many digital maps are available from miscellaneous sources, such as scanned paper maps, online map services (e.g., Google Earth and Google Maps) (Li, 2007; Kobayashi, et al., 2010), data repositories of volunteered geographical information (e.g., OpenStreetMap) (Neis and Zielstra, 2014), and georeferenced cyberinfrastructures (Wright and Wang, 2011). The huge number of maps currently available have encouraged researchers to focus on the efficiency of map retrieval and discovery, which are significant aspects of the productivity and efficiency of digital maps. A foremost challenge associated with critical techniques for map discovery is how to conduct automatic interpretations of map content, since the capabilities of labor and traditional

interactive tools for map interpretation cannot meet the qualifications for processing the massive quantity and diverse nature of digital maps currently available.

An approach that supports automatic raster-map interpretation must consider several challenges. First, in comparison to the traditional viewpoint that users' needs are the most important factor in improving map design, many researchers have found that exactly defining user needs is close to impossible (Carter, 2005; Perkins, 2013b). Currently, cyber technology has led maps servers—such as Google Maps and Bing Maps—to evolve into an essential part of a person's daily routine. This means that the same map might be used in different ways according to specific objectives and backgrounds (White, 2006; Foody, 2007). Second, maps are never the product used and created by professional agents alone. Maps with similar configurations and themes might be designed and depicted in various ways. Third, the emergence of web services, volunteered geographical information, and cyberinfrastructure provide comfortable platforms on which people can publish, edit, and share their maps that they created with web resources, making it impossible to establish a unified standard for map design. Today, the roles and distinctions between map producers and map users are much vaguer in the era of big data, volunteered geographical information and spatial cyberinfrastructure. For example, a person who seeks maps that regard specific interests could also publish individually designed map creations (Hurst and Clough, 2013).

In the context of supporting on-demand and historical map discovery, a number of methods based on metadata or annotations have been reported (Li, Yang and Yang, 2010; Li, et al., 2011). Though these methods incorporate cutting-edge techniques in terms of semantic analyses, data mining, and machine learning, metadata- and

annotation-based map interpretation cannot support map understanding because of three limitations. First, most of the map resources available on the Internet lack fundamental map elements, including map content like map titles, legends, and descriptive text. Moreover, the degree of detail in the metadata and annotation methods has a considerable influence on results generated with metadata- and annotation-based raster-map interpretation. However, the quality of the metadata and annotation always varies considerably for maps with a similar theme that are available from web map services (WMS), map repositories, and map services (Wu et al., 2011; Gui et al., 2013). For example, a map annotated as a road map might be a map that depicts only linear street networks or instead illustrates detailed information, including street levels and street names. Finally, map metadata and annotation are generated on the basis of individual viewpoints and understanding, meaning that the theme and content of similar maps might be annotated differently. Thus, metadata and annotation are not useful for explicitly representing the content of a map, nor do they support automatic map content understanding (ref).

To address the limits of metadata and annotation-based approaches for map interpretation, map content-based approaches have become a primary area of investigation over the last couple decades (Chiang et al., 2013). Since the layouts and configurations of maps are varied and complicated, not all features derived from a map are useful. In addition to metadata, map annotation, and map elements, the features useful for obtaining map content include map text, map symbols, and map type (Pezeshk and Tutwiler, 2013; Chiang, et al, 2016; Li, Liu and Zhou, 2018). Map symbols provide key graphical features for map object identification. Though the principles of cognition are

significant in map symbol design, the configuration of map symbols for a similar geographical object might still be different in maps that are accessible from widely available resources. Thus, map symbols are not used as a fundamental map feature in the proposed method. Map-text recognition is a branch of optical character recognition (OCR) (Mithe, Indalkar and Divekar, 2013) in cartography and GIS data mapping, which attempts to convert the map text within a variety of printed media likescanned papers, PDF files, and images into a machine-readable format. Chiang et al. (2016) evaluated the state-of-the-art method for map-text recognition on the basis of a variety of criteria. Because of the distinction of map fonts, styles of map characters, printing quality, map resolution, and map complexity, the previous approaches for map-text recognition, which apply techniques from image processing, clustering analysis, object-based image analysis (OBIA), and machine learning, had limitations in automatic map-text recognition in high-resolution maps. Meanwhile, few works have reported investigations into automatically classifying map type; however, these investigations are valuable for people trying to understand the content of a map. For example, DEMs and topographic maps, rather than orthophoto maps, are useful to a query that retrieves maps regarding elevation information (Zhou, et al., 2012). The state-of-the-art approaches to map-text recognition and map-type classification are discussed in the next section. Considering the power of deep-learning techniques for scene classification and object recognition (Bengio, Courville and Vincent, 2013; LeCun, Bengio and Hinton, 2015), the first section of the proposed method describes a methodological framework for implementing a convolutional neural network (CNN) that supports map-text recognition and map-type classification.

However, the machine-readable information obtained via map-text recognition and map-type classification falls far short of supporting human-level map understanding (Li, Liu and Zhou, 2018). In addition to map-text recognition and map-type classification, the transformation from map information into explicit knowledge, which enables explicitly conceptualizing and semantically organizing the content of a map, is another critical stage in realizing automatic map discovery. Up to now, no literature has comprehensively reported studies to convert map information into map knowledge. At the same time, the conversion of georeferenced information into geosemantics and spatial knowledge has been a primary research focus in the community of geographical information science (Janowicz et al., 2012). The relevant technologies—including geontologies (Fonseca, et al., 2002), geosemantic queries (Battle and Kolas, 2012), and the geospatial semantic web (Becker and Bizer, 2009)—have been developed to facilitate georeferenced data interoperation, spatio-temporal pattern discovery, geospatial knowledge discovery, and so on. In regard to the potential of geospatial semantic analyses to facilitate georeferenced information analyses, the second section of the proposed method focuses on advancing the transformation from plentiful map information into explicit map knowledge.

This chapter reports an integrated framework to support automatic human-level map understanding and map discovery, and the framework includes map-text recognition and map-type classification with deep-learning techniques, as well as the discovery of map semantics via the techniques of semantic analyses. The remainder of this chapter is organized as follows. Section 2 reviews the literature on map-text localization, map-text recognition, and map-type classification. Section 3 sketches the framework of the

proposed method, and then introduces the technical details of each part of the proposed method. Section 4 describes and discusses the experimental results of map discovery with the proposed method. Section 5 summarizes the proposed method, highlights of the proposed method, and prospective work in map-content recognition.

3.2 Methodology

3.2.1 Map-text Recognition.

Map-text recognition attempts to detect the location of text units in a digital map and then converts the detected map text into machine-encoded documents. The architecture of the methodological framework for map-text recognition comprises three parts: map-text detection, map-text unit separation, and map-text classification (see Figure 11). Three independent CNN models were developed for each part. A fine-tuned and Faster R-CNN is used for map-text detection, a fine-tuned DeepLab V3+ is used for map-text segmentation, and the Tesseract OCR engine (Smith, 2007) is used for map-text classification. The following subsections introduce the details of each part. In particular, an approach for map-text straightening was developed to further process the raw segmentation result.

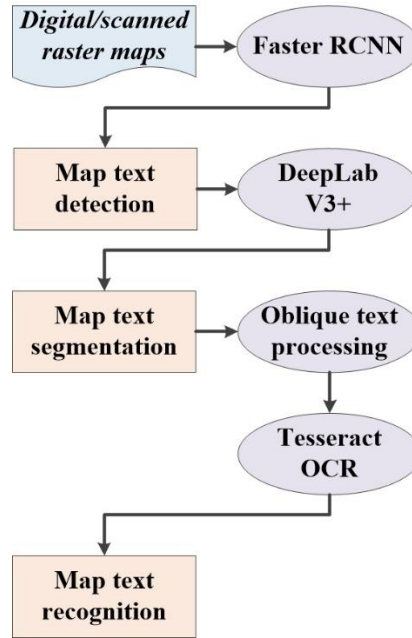


Figure 11. *Methodological framework of map-text recognition.*

3.2.2 Map-text Detection.

The CNN for object recognition can effectively deal with the tasks in terms of optical character recognition-OCR (Liao, et al., 2017; Liu and Jin, 2017). Thus, the proposed method conducts text detection with the CNN for object detection, viewing text units as objects in a digital map. Faster R-CNN is a state-of-the-art CNN, which was developed on the basis of the architecture of Fast R-CNN. The main component of the conventional Faster R-CNN comprises the region proposal network (PRN) for generating region proposals from an image, as well as a network architecture of Fast R-CNN (Girshick, 2015) for classifying the region proposals into a predefined category. To further improve the power of the conventional Faster R-CNN for object recognition, the following methods exploited strategies to increase the classification accuracy and to reduce the overlap areas between a bounding box and the detected object. The attempt to gain classification accuracy involves replacing resnet (He, Zhang, Ren et al., 2016) with the

cutting-edge CNN models for classification, such as the Inception series (Szegedy, Ioffe and Vanhoucke, et al., 2017), Mobilenet (Howard, Zhu and Chen et al., 2017), NASNet (Zoph, Vasudevan and Shlens, 2017), and PNASNet (Liu, Zoph and Shlens, et al., 2017). Atrous convolution is the strategy for adapting the field of view to control the overlap between a bounding box and the detected object. Thus, the present study uses the Faster R-CNN incorporated with Atrous convolution and PNASNet to detect the bounding box of the map-text unit.

Though the accuracy of map-text detection remains stable to a high degree within the Faster R-CNN, a challenge is commonly reported in the process of recognizing every character in the detected map text (ref). Unlike the characters in a photo, the orientation of characters in a map-text string might vary because of the map configuration and generalization. Moreover, map-text units might be overlapped with other map features or a complicated map background, which poses a significant challenge for an OCR engine to efficiently recognize and convert the text in a digital map into machine readable format (Pezeshk and Tutwiler, 2011; Li, Liu and Zhou, 2018). As mentioned in Subsection 2.2, the operation for setting map text upright plays a key role in map-text classification, and an operation for oblique map-text straightening is always required in advance.

3.2.3 Map-text Unit Separation and Classification.

Using the results of map-text detection, the form of which includes the boxes that contain map-text units, map-text unit separation next extracts map-text units from its context and other overlapping features. Considering the potential of the semantic segmentation technique on the image analysis reported previously (Chen et al., 2017), the proposed method develops a CNN for semantic segmentation for map-text unit separation.

Semantic segmentation studies the object and scene in an image at a pixel-level resolution, depending on the contextual information around each pixel. Thus, semantic segmentation can effectively avoid the limitations of pixel-level image-analysis techniques while extracting the precise shape of an object at the pixel level. As a cutting-edge technique for pixel-wise semantic segmentation, DeepLab V3+ (Chen, Papandreou and Schroff, et al., 2017) was implemented as the CNN for semantic segmentation to segment the detected map feature into various map characters or to separate map characters from a detected bounding box.

In regard to the loss of spatial details during the end-to-end learning process, including an encoder module and a decoder module, DeepLab 3+ proposes a new approach called Residual Block for fine-feature learning over multiple scales. The strategies associated with this approach include two substantial components: atrous convolution and atrous spatial pyramid pooling (ASPP).

After map-text unit separation, the method for straightening the map text that was available in the previous version of Intelligent Map Reader was improved (Li, Liu and Zhou, 2017), which might have limitations in dealing with upright, curved text strings. The proposed method provides two separate strategies for processing aligned and curved map-text strings; the workflow of the two strategies are shown in Figure 12.

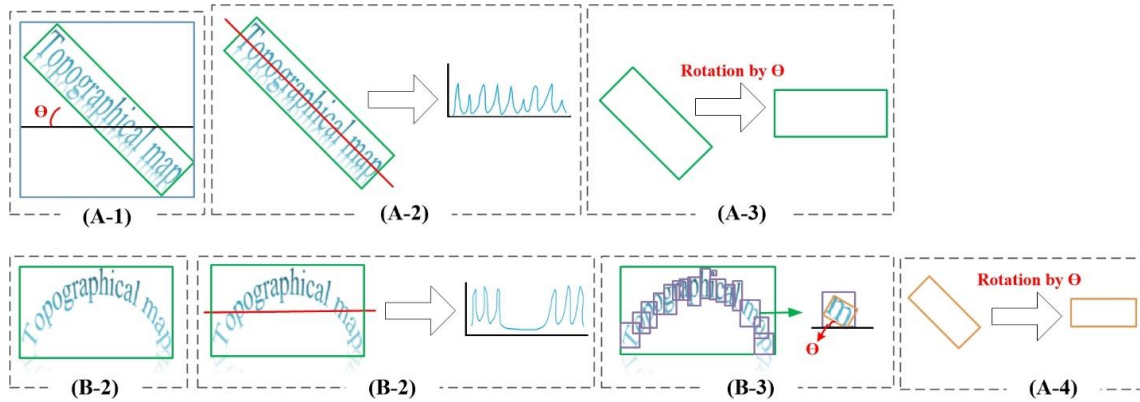


Figure 12. (A) Aligned map-text string straightening and (B) curved map-text string straightening.

In Figure 12, the blue and the green rectangles refer to the minimal bounding box (MBB) and the rotated MBB for a map-text string, respectively. The purple and the orange rectangles refer to the MBB and the rotated MBB for a single map-text character, respectively. Both of these strategies employ similar operations during the first two steps, which are shown in Figure 12(A-1 and A-2) and Figure 12(B-1 and B-2). First, the MBB and the rotated MBB for an oblique map-text string are created, respectively. If this map-text string is upright, the MBB and the rotated MBB overlay each other overall. Otherwise, an intensity histogram of the cross-profile section of the rotated MBB, which is shown as the red line in Figure 12(A-2) and Figure 12(B-2), is calculated. The intensity histogram of the curved map-text string is significantly different from this aligned map-text string: a floor is seen in the middle of the histogram, because no map text exists in the middle section of the red line. Thus, the result of the histogram determines which strategy is applied in the following steps. The aligned map-text string is directly straightened by rotating the intersection angle between the rotated MBB and the MBB—namely, the green box and the blue box in Figure 12(A-1). Alternatively, a MBB and a

rotated MBB are created for each character in the curved map-text string. Next, every character is straightened individually by rotating the intersection angle between its rotated MBB and MBB—namely, the orange box and the purple box in Figure 12(B-3).

Intelligent Map Reader exploits the Tesseract OCR engine (Smith, 2007; Patel, Patel, and Patel, 2012) to recognize the text obtained via the map-text unit separation. Tesseract OCR is an open-source OCR engine that supports 116 languages in the newest version, and is available at this link: <https://github.com/tesseract-ocr>. The previous version of Intelligent Map Reader (Li, Liu, and Zhou, 2018) reports that the accuracy of text recognition with Tesseract OCR reached 100% for the map texts, which were extracted well and without noise.

3.2.4 Map-type Classification.

A methodical classification system of map category is challenging because of established bias. Moreover, classification systems of map category are difficult to unify because of the diversity in map theme and map visualization. However, map types might have essential effects on map-content understanding. For example, though the place name “Phoenix” can be seen in a street map and a topographic map, elevation information involving contour lines, spot elevation, and terrain features are only available in the topographic map. In this case, map-type information is integral to effectively conducting an example search request: “a map depicting the topographical features in Phoenix.” Moreover, place names printed in diverse digital maps differ because of varying topics, scales, readership, and visualization. For example, school building names depicted in a campus map might not be printed on a topographical map. This means that a search

request like “a map including detailed campus information” might overlook the maps with respect to topography, terrain, or elevation.

Considering the power of CNNs for image-scene classification and remote-sensing imagery (LeCun, Yoshua and Hinton, 2015; Szegedy, et al., 2017), this research attempts to exploit the techniques of CNN for image classification to conduct map-type classification. Zhou et al. (2018) reported a comprehensive evaluation of the performance of a variety of CNNs in map-type classification. The results showed that with the support of a systematically prepared training data set, the state-of-the-art CNNs for scene classification could produce a satisfactory result for map-type classification within the accuracy range of 93% to 99%. The proposed method applies PNASNet to conduct map-type classification, with a reported efficiency in image-scene classification based on a variety of benchmark data sets.

3.3 Human-level Map Understanding.

Machine-readable map-text and map-type are fragmentary, unsystematic, and disconnected, and such qualities hinders users from understanding the contents of maps. For example, because Arizona State University (ASU) is located in the city of Tempe, both the Tempe city map and ASU campus map contain map names such as Arizona State University and Tempe Downtown Lake. These map names produce unreliable results that render them useless to determine which is the true ASU campus map. Thus, the proposed method uses a framework that integrates ontology, semantic queries, and semantic reasoning to transform extracted map text and the confirmed map type into explicitly meaningful descriptions of map content.

The varied techniques of geo-semantic analyses facilitate the systematic organization of the relationships between - and explicitly represents the knowledge of - objects, events, and phenomena. Georeferenced semantic analysis has been exploited into four parts: (1) relevant and valuable information discovery (Yue, et al., 2011), (2) inherent meaning mining (Bogorny et al., 2011), (3) heterogeneous data interpretation (Fonseca et al, 2002), and (4) reasoning-driven automatic semantic query (Battle and Kolas, 2012). These four parts are useful in the proposed method to establish an integrated ontology for map information organization, developing a GeoSPARQL-enabled geosemantic reasoning system, and building a semantic query to facilitate the representation of map characteristics.

Figure 13 illustrates the methodological framework for supporting human-level map understanding. USTopographic (Tambassi, 2018) and GeoNames (Ballatore, Bertolotto and Wilson, 2014) are the ontology and taxonomy already established. In addition to these two ontologies, the proposed method uses a MapType Ontology to semantically organize the description of each map type accessed from online dictionaries and Wikipedia. Next, a GeoSPARQL-enabled semantic query is used to access the hidden information.

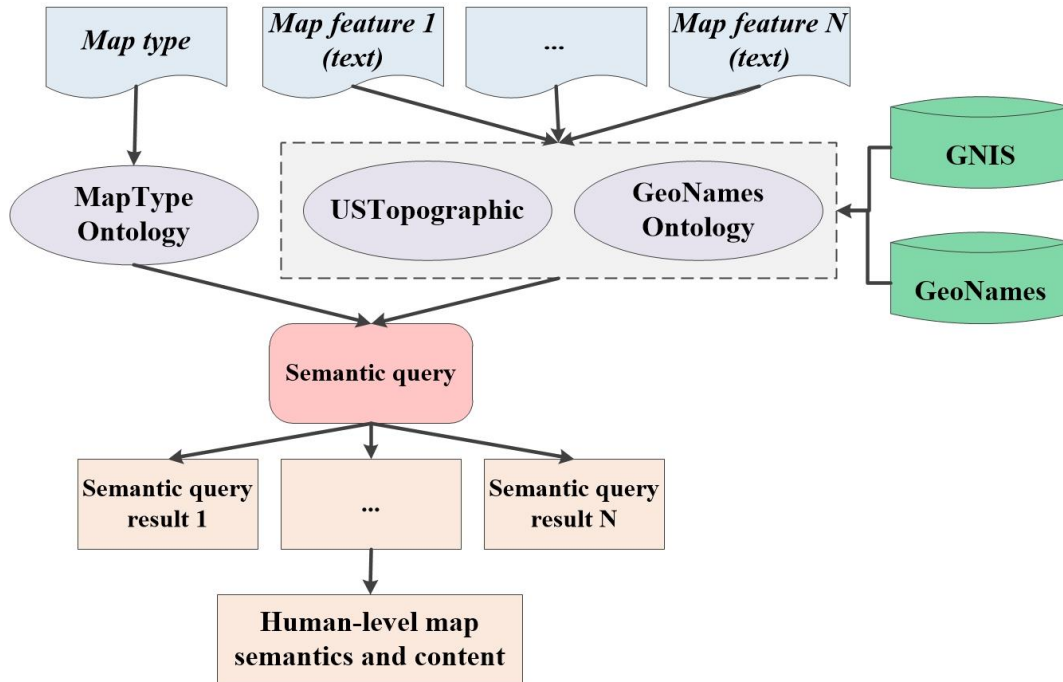


Figure 13. From extracted map features to map content with semantic analysis.

3.3.1 Ontology Development and Integration.

Ontology is a systematic model that includes the formal definitions, semantic categories, properties, and relationships among various classes, entities, and data. In terms of georeferenced information, the ontological model has been viewed as an important tool for dealing with geodata interoperability (Fonseca et al., 2002). This paper exploits a state-of-the-art ontology and a geospatial taxonomy for map information—GeoNames Ontology (Wick, Vatant and Christophe, 2015) and USTopographic (Usery and Varanka, 2012)—to semantically organize the conceptual hierarchy of map names. GeoNames Ontology provides a fully organized conceptual hierarchy that includes all map feature classes available in the OpenStreet Map products. USTopographic ontology is for semantically organizing the names available in the National Map products, and it includes six subcategories: Built-Up Area, Division, Ecological Regime, Surface Water,

and Terrain. In addition, an ontology is developed for the formal representation of map-type information: MapType ontology. MapType ontology was created with an ontology-editing framework called Protege that supports Ontology Web Language (OWL) and Resource Description Framework (RDF) languages.

Considering the heterogeneity of terminology in these three ontologies, an integrated ontology that merges these three ontologies is integral to effectively supporting semantic query. First, based on the categorical system in USTopographic, Geonames ontology and GNIS gazetteer two ontologies were integrated depending on their similar classes. Next, the terminology of MapType ontology was joined to the ontology integrating Geonames ontology and USTopographic as a new category called “Map type.” However, a number of classes were still not fused into the preliminary integration result because of the polysemy and synonymy. For example, the “Administrative area” class in GeoNames ontology and the “Division” class in USTopographic ontology are semantically similar but literally different. Thus, a further integration aims at fusing the remaining classes on the basis of the synonym list and geographical knowledge.

Moreover, as shown in Figure 4(A), individuals are missing in the ontology integrating GeoNames and USTopographic ontologies. Classes alone cannot support to map content identification, since place names and other names shown in a digital map are individual features of a class. As shown in Figure 4(B), both classes and their individuals were essential to supporting an efficient semantic query. The proposed method uses the GeoNames gazetteer and the GNIS gazetteer to enrich the individuals in the ontological model integrated into the GeoName and USTopographic ontologies. Figure 14 shows an

example solution to add individual instances to an integrated Arizona State University ontology.

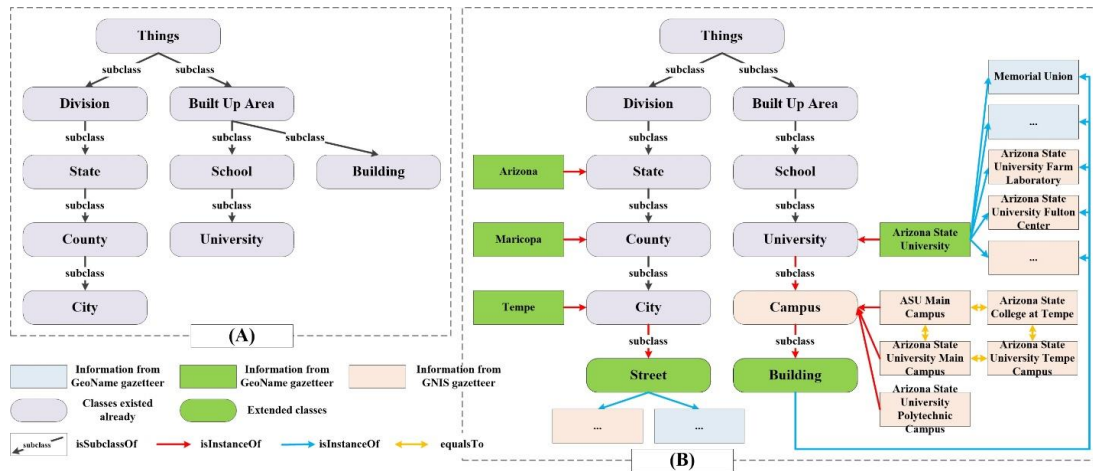


Figure 14. Adding individual instances to an integrated Arizona State University ontology.

3.3.2 Semantic Query and Reasoning.

Semantic query attempts to automatically provide answers to a question with logical reasoning using an existing knowledge graph. In comparison to an information query, a semantic query can mine the hidden meaning by discovering the architecture of knowledge graphs. For example, it is easy to literally deduce which state Arizona State University is located in. However, determining which city Arizona State University is located in, as well as the neighboring cities, requires a knowledge graph on the spatial and topological relationships between Arizona cities and Arizona State University.

This research exploits the conceptual class hierarchy integrating GeoNames ontology and USTopographic ontology, conceptual hierarchy of classes, and the individual instances of each class to support the semantic query. The semantic query was conducted on the basis of a GeoSPARQL framework (Li, et al., 2016). Query conditions and goals were defined in the form of a semantic triple: Subject—Property/Predicate—Object,

where Property/Predicate denote a relationship between the subject and object. The semantic triple supports the explicit representation of the relationship between two objects or between an entity and an attribute in a machine-readable way. Because the relationship is modeled in an explicit and unambiguous manner, the semantic query enables a precise result to be produced via processing and inferring the semantic relationships represented by the semantic triples. For example, the unstructured text description “Arizona State University is located in Tempe” could be organized as two structured triples: “Arizona State University—isIn—Tempe” and “Tempe—encloses—Arizona State University.” In addition to the document-based semantic triple, GeoSPARQL enables the use of geographical coordinates to create a geometry-based semantic triple that represents the topographic relationship “contained” between Tempe and Arizona State University.

The semantic query available in the proposed method was developed according to the semantic triple, which supports queries with a subject, a predicate, and an object. Example queries based on the semantic triple “Arizona State University—isIn—Tempe” are listed in Table 4.

Table 4.

Comparison of the previous version of Intelligent Map Reader and the proposed method.

	Query codes	Query result
	Query type 1	
Query form	?who/?what/?which/?where Predicate Object	Subject
Practice	?what isIn Tempe	Arizona State University
	Query type 2	
Query form	Subject ?relationship Object	Predicate
Practice	Arizona State University ?how Tempe	isIn

	Query type 3	
Query form	Subject Predicate ?whom/?what/?where	Object
Practice	Arizona State University isIn ?where	Tempe

3.4 Experiments

3.4.1 deepMap: A Benchmark Data Set for Map-text Recognition.

It is critical to prepare large-scale, well-labeled data to feed a neural network to enhance its capability of distinguishing different classes (Bengio, Courville and Vincent, 2012). Thus, a benchmark data set was created for map-type classification (Zhou, et al, 2018). The dataset includes data collected from online ArcGIS maps, Google Maps, the USGS' US Topo and historical DRGs, and Bing Maps. Figure 14 shows three types of data available in *deepMap*: map characters, labeled map text, and map samples that belong to various types of map.

- The map character dataset was used to fine-tune the CNNs for map-character classification and semantic segmentation. There were 43 categories in the map-character data set, because the capital case and lower case of “C,” “U,” “V,” “W,” and “Z” are similar in the majority of maps. Each category had 100 samples, and the dimensionality of each image that included a map character varied from $10 \times 10 \times 3$ to $40 \times 40 \times 3$.
- The dataset of labeled map text was used for fine tuning the CNN for map-text detection, the dimensionality of which was $64 \times 64 \times 3$. There were 300 maps labeled with text, and around 1,000 were labeled as map-text samples.
- The map-type data set was used for map-type classification with CNN, and the data set included 10 categories: topographical map, transportation map, 3-D map,

nighttime imagery map, ortho imagery map, land cover map, DEM, boundary map, comic map, and sketch map. Figure 15 illustrates the selected maps in terms of the 10 categories. Each category had 250 map samples, and the dimensionality of each map sample was $256 \times 256 \times 3$.



Figure 15. Illustration of selected samples from the deepMap benchmark data set.

3.4.2 Experimental Design.

Two hundred maps for each category were downloaded from Google search engine via a crawler called Google Image Downloader. The keywords used for the search were “Arizona State University map,” “Yellowstone National Park map,” “Gulf of Mexico map,” “San Francisco map,” and “Newark Airport map.” These five themes denoted the maps in terms of small-scale urban areas, popular areas of interest (AOI), large-scale natural regions, large-scale urban regions, and landmarks, respectively, covering a variety of aspects possibly seen by people in their daily lives. However, the quality and reliability of data resources on the Internet were not evaluated comprehensively. Each map category included some maps that were useless because of unrecognizable file formats and irrelevant map content. Thus, valid maps were manually selected from the results generated by Google Image Downloader. Moreover, true and incorrect samples were manually selected and labeled in the valid maps. Table 5 lists the numbers of downloaded maps, valid maps, and useful maps in terms of each category. The dimensionality of the collected maps varied from around $300 \times 300 \times 3$ to around $3000 \times 3000 \times 3$.

Table 5.

Statistics of the experiment.

Map theme	Total number of downloaded maps	Total number of readable maps	Total number of useful maps
Arizona State University	250	182	65
Yellowstone	250	226	132
Gulf of Mexico	250	231	98
San Francisco	250	222	140

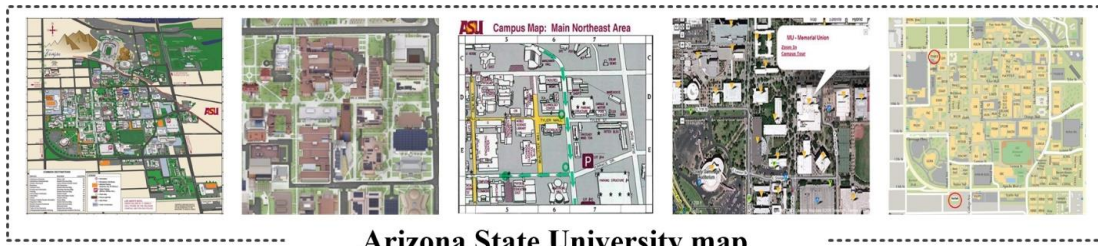
Newark Airport	250	210	85
----------------	-----	-----	----

The significant difference between the number of useful maps and the number of downloaded maps shows that a majority of digital maps accessed from web sources were irrelevant and useless. Many downloaded maps are not useful were due to four main reasons:

- First, two maps that had approximate themes might be classified as a similar category, such as Arizona State University and University of Arizona.
- Second, two maps might cover neighboring locations, such as Yellowstone National Park and Grand Teton National Park. Yellowstone National Park and Grand Teton National Park are neighboring areas. Thus, some maps titled Yellowstone National Park also covers Grand Teton National Park. Otherwise, some maps titled Grand Teton National Park may also cover Yellowstone National Park.
- Third, two maps might have incomplete, semantically confusing annotations, such as Yellowstone National Park and Grand Canyon National Park maps. Yellowstone National Park and Grand Teton National Park are neighboring areas. Thus, some maps titled Yellowstone National Park are the map about Grand Teton National Park.
- Fourth, the theme or spatial coverage of one map might contain the theme or spatial coverage of another map—for example, a Yellowstone National Park map and a US state map that highlights the location of Yellowstone National Park.

Moreover, the significant difference between the number of useful maps and the number of downloaded maps also indicated that an additional operation should be included to support accurate and efficient map discovery from web sources.

Figure 16 displays map samples collected via the Google search engine. The maps accessed from web sources were created with varied themes, styles, configurations, scales, and readership orientations. Many maps lacked fundamental map elements, such as map titles, legends, and scales.



Arizona State University map



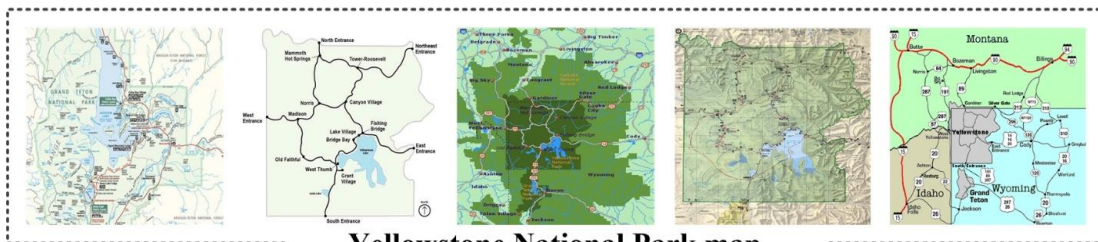
Gulf of Mexico map



Newark Airport map



San Francisco map



Yellowstone National Park map

Figure 16. *The selected maps accessed via Google image search.*

To select the appropriate and representative maps, the following three approaches were evaluated: a metadata/annotation-based approach, a map element-based approach, and the proposed method. The metadata/annotation-based approach used the file name

and metadata of a digital map accessed via the Google search engine to determine whether it met the criteria. The map element-based approach discovered maps based on the fundamental map elements of map title, map legend, and citations. The proposed method attempted to discover the useful maps via a sequence of steps mentioned in the proposed method based on a NVidia 1070 GPU-enabled computer.

The workflow of the proposed method comprised four steps. The first step finely tuned the architecture of an imagenet-pretrained Faster RCNN with atrous convolution, a COCO-pretrained DeepLab V3+, and an imagenet-pretrained PNASNet. To objectively evaluate the performance of the CNNs for map-type classification, map-text detection, and map-character segmentation, the proposed method used the Faster R-CNN default value for the learning rate, patch size, and anchor ratio, and the default kernel size, stride, pad, and rate of DeepLab 3+. The results of the first step included the bounding boxes that enclosed map-text units.

In the second step, map-text detection was conducted with the fine-tuned Faster R-CNN, and then the fine-tuned DeepLab V3+ was used to segment the map-text units from the bounding boxes generated by detection. Next, the separated map units were straightened using the proposed method presented in Subsection 3.3.1.2, and it was recognized with the Tesseract OCR engine. The results of the second step contained a list of machine-readable map texts.

The third step applied the fine-tuned PNASNet to classify the maps collected from web resources. Because it was impossible for the nighttime categories shown in Figure 15 to cover all map types, a new category called “unclassified” was added, in case a map could not be assigned to any predefined category. If the classification scores of all

night categories were low for a map, this map was viewed as unclassified. The results of the third step were machine-readable map-type information.

In the last step, the map-type information and map-text information were integrated to conduct the semantic query, with the main goal of gaining the implicit semantics and hidden content of every map. As mentioned in Subsection 3.3.3.2, a semantic query was implemented on the basis of the platform developed with Protege. All triples potentially used for semantic reasoning and query were organized according to an ontology integrating the GeoNames, USTopographic, and MapType ontologies, and the triples were correspondingly derived from the GNIS gazetteer and the GeoNames gazetteer. In the last step, a conceptual graph of the content of every map was generated.

3.4.3 Demonstrative Results.

Table 6 lists the precision and recall of the map recognition for each category, according to the metadata/annotation-based approach, map element-based approach, and the proposed method. Few maps were retrieved by the map element-based approach, meaning that a majority of maps already on the Internet and web sources were created with no professional standards. In addition, this result indicates that map elements might not be the ideal features for the representation of map content.

Table 6.

Evaluation of map recognition.

	Precision			Recall			F-scores		
	1*	2*	3*	1*	2*	3*	1*	2*	3*
ASU	0.6735	0.28	0.9184	0.5893	0.3214	0.8036	0.3143	0.1497	0.4286
Yellowstone	0.7179	0.9333	0.8980	0.6364	0.3182	0.6818	0.3374	0.2372	0.3876
Gulf of	0.4048	0.2727	0.84	0.5102	0.1633	0.4286	0.2257	0.1021	0.2838

Mexico									
San Francisco	0.3590	0.2643	0.9756	0.5357	0.9487	0.5857	0.2149	0.2067	0.366
Newark Airport	0.4264	0.1882	0.9516	0.8706	1	0.7294	0.2862	0.1584	0.4129

*1: Metadata/annotation-based approach

*2: Map element-based approach

*3: The proposed method

Figure 17 illustrates an example semantic query of an Arizona State University campus map and a Gulf of Mexico map, as well as the conceptual graph developed for maps of Arizona State University and the Gulf of Mexico. Figure 17(A) lists the entities that belong to Arizona State University, as well as those located in the Tempe campus or near the Tempe campus. Figure 17(B) lists the entities that were near the Gulf of Mexico. Figure 17(C) shows classes in the conceptual hierarchy and their corresponding individuals, which were developed on the basis of the first five maps accessed via the Google search engine in terms of “Arizona State University map” and “Gulf of Mexico map.”

occurred because of a semantic heterogeneity (Lutz et al., 2009) in map decryption and annotation, including polysemy, homophony, and word context.

Similar to the metadata/annotation-based approach, the results of the map element-based approach were generated assuming that all map elements in a map were recognizable. However, the map element-based approach retrieved few useful maps. The results of the map element-based approach showed that a majority of maps accessible from multiple sources were not created with professional standards.

The proposed method generated the results shown in Table 6 without setting the IoU parameter, because the result of the map-text detection was true only when the whole body of map text was included in a bounding box. For example, the IoU of straight map text should be close to 100%, but the IoU of oblique map text might be 60%. The precision shown in Table 6 was influenced by the precision of the map-text detection, the precision of the map-text segmentation, the precision of the map-text unit straightening, and the precision of the map-text classification.

- The precision of map-text detection was heavily affected by the spatial resolution of a map.
- Few losses were seen in the results of the map-text segmentation and map-text unit straightening. Moreover, it was found that the results of the conventional object-based segmentation (Achanta, et al., 2012) were similar to those produced by semantic segmentation. This might have occurred because the color and texture features of the map text were clean and uncomplicated.

- If the text in an image was relatively straight and without noise, the precision of the map-text classification reached 100% via Google Tesseract OCR, which was reported in the previous version of the proposed method.

Although the proposed method performed well in map-text recognition for topographical maps, the variation and complexity of digital maps from web sources are much more critical, which poses substantial challenges for state-of-the-art CNNs. Three main issues affected the performance of the proposed method in map text recognition. First, some maps lacked map text, making it impossible for Intelligent Map ReaderV2 to recognize the content of these maps. Second, the resolution of the maps significantly influenced the effectiveness of CNNs for text detection. The size of map text beyond the limit of field of receptive view (FoV) poses a great challenge for a CNN to generate precise region proposals. Last, more training samples in terms of different types of map text were needed to further improve the CNNs for map text recognition.

As indicated by the results listed in Table 6, the traditional approaches for studying metadata, annotations, and map elements have limitations in terms of supporting efficient map recognition, because of numerous factors—such as unstructured map data, limited map quality, diverse map configurations, and unprofessional map generation. The insufficient performance of the metadata/annotation-based approach confirmed the importance of using map recognition within content-based map analyses, which is similar to the standpoint substantiated by content-based image retrieval (Smeulders, et al, 2000; Liu, 2007). Moreover, the results generated by the proposed method indicate that it would be valid and feasible to develop efficient map recognition utilizing the advantages of deep-learning techniques.

In map-type classification, the previous version of the *deepLab* benchmark data set was updated to increase intraclass variations and decrease interclass dissimilarities. The-state-of-the-art CNNs for classification, including Inception_Resnet and PNASNet, can obtain accuracies higher than 90% in the top two classification. In the new version of the *deepMap* benchmark dataset, the same configuration of CNNs obtained accuracies ranging from 88% to 97%. Moreover, the accuracies varied on the basis of different types of maps, which are summarized as follows,

- The classification accuracies of topographic maps and orthophoto maps remained high. These two types of maps were clearly distinguishable from other types of maps.
- The classification accuracy of transportation maps remained high for urban scene maps and large-scale natural scene maps.
- 3-D maps were likely to be misclassified into comic maps.
- There were no nighttime maps, DEMs, or sketch maps accessed via the Google image search engine.
- Land cover maps were easily confused with boundary maps.

3.5 Conclusions

Maps are significant in terms of representing the natural characteristics and human-made components of a place. Because of the rapid development of earth observation systems and cyberinfrastructure and the spread of Internet techniques, considerable amounts of digital maps can now be accessed from multifarious sources. This new phenomenon not only reforms the traditional manner and viewpoint of map generation and map usage but also poses two major data-associated challenges. First, a

number of maps that people access are not created well, nor do they precisely fit the relevant demands and criteria. Moreover, the number of available maps far exceeds the capability of map storage, retrieval, and analysis. Thus, though countless digital maps are available and generated by a variety of sources, automatic map retrieval, map discovery, and map content understanding still face difficulties.

Traditional ways of map creation—including metadata-based approaches, map element-based approaches, and OGC standard-based approaches—have limitations that hinder efficient access, discovery, and comprehension of map content. To address these challenges, the techniques that enable the extraction of map content-based map information has been reported in the previous two decades, ranging from image processing techniques to machine-learning algorithms. Today, the potential power of deep-learning techniques in image analysis is attracting much attention from the cartography and GIS communities. The proposed method was founded on the strategy of exploring digital map content with deep-learning techniques, and it employs state-of-the-art CNNs to facilitate map-text and map-type recognition.

Moreover, the proposed method focuses on not only the results of map-feature extraction, but also the implicit meaning of each map feature and the relationship among various map features. Recently, the conversion of map-feature information into map semantics and knowledge has been an unexplored research area. The second part of the proposed method provided a framework for map semantics and knowledge discovery, in an attempt to bridge the gap between map-feature recognition and map-knowledge discovery.

Several areas are still worthy of further investigation. First, useful input data are the essence of machine learning and AI techniques. In addition to map text and map type, it would be beneficial to explore other features that could be developed to help computers learn the representation of a map. Moreover, it has been reported that reinforcing learning techniques helps the cutting-edge CNNs to be more efficient in scene classification, object recognition, and semantic segmentation. Last, spatial knowledge and models are integral to geospatial analyses and applications, so developing an approach that incorporates spatial thoughts into machine-learning and AI techniques would be a worthwhile topic for future research.

CHAPTER 4

DOMAIN KNOWLEDGE-ENHANCED LAND-COVER/LAND-USE

CLASSIFICATION WITH IMAGE-SEMANTIC MODEL

4.1 Introduction

Land-cover display characteristics of Earth's surface that include the physical appearance of natural materials and the places where artificial activities occur. Mapping and analyzing land cover are processes aimed at efficiently identifying specific objects or events on Earth's surface over a specific period of time. However, field investigation for land-use/land-cover (LULC) classification in a large-scale area is laborious and time consuming. Frequently, updated remote sensing images provide the potential to support large-scale LULC classification and reinforce researchers' competence in understanding, estimating, and predicting the influences of natural forces and artificial activities. A brief workflow of approaches for land-cover classification mainly includes extracting land-cover features on global and local scales and designing an integrated classifier to label land-cover classes revealed in a remote sensing image. Traditional approaches to land-cover classification, such as spatio-contextual approaches (Li et al., 2014), geographical object-based image analysis (GEOBIA) approaches (Blaschke, 2010), machine learning approaches (Maxwell, Warner and Fang, 2018), and rules-based approaches (Zhang and Zhu, 2011), present several limitations for land-cover feature learning and classification (Zhang, Zhang and Du, 2016; Cheng, Han and Lu, 2017). Many approaches might require a great amount of manual effort for thresholding, setting a representative scale of segmentation, and selecting the representative training samples and features. Robustness is another major concern because these approaches possess restrictions in dealing with

noise and diversely complicated patterns in LULC scenarios. Lastly, and most importantly, traditional approaches are insufficient for handling the hyper-dimensional data feature space being derived from high numbers of remote-sensing images available today.

The rapid progress of convolutional neural networks (CNNs) provides a significant opportunity to extract high-level abstract features from a remote sensing image for characterization of complicated LULC types. In other words, a CNN model supports to discover the high-level features of remote sensing images, which is useful for understanding the nature of various LULC types (Bengio, Courville and Vincent, 2013; LeCun, Bengio and Hinton, 2015). In the remote sensing community, the earliest attempts at using deep learning to facilitate land-cover classification might be traced back to Dai and Yang's work (2011), wherein they designed a two-layer sparse coding system to detect the features useful for representing the image content. Some deep learning-based approaches include stacked sparse autoencoder (Li et al., 2016; Li et al., 2016; Zhang et al., 2017), deep belief net (Zou et al., 2015), CNNs (Zhang, Zhang and Du, 2016; Zhu et al., 2017), and the derivative models within a deep neural network architecture. Considering the extensive computing power required for creating a completely fresh CNN model, fine-tuned CNN models, such as classical CNNs (Scott et al., 2017; Wang, 2017), ImageNet, and COCO pretrained CNNs models (Marmanis et al., 2016), have been widely used for land-cover classification. These efforts deduced that the CNN models—pretrained with digital photos—remain effective in land-cover classification from remote-sensing imagery. However, the representation of LULC

scenarios in a remote sensing image is different from those of objects and phenomena depicted in a normal photograph.

Some studies paid attention to the improvement of the architecture of CNN models designed to represent the content in photos. In these studies, robust classifiers (Weng et al., 2017) and feature post-processing mainly attempted to reduce the dimensionality of the extracted features before applying them into a classification layer (Wang, 2017; Xiao et al., 2017; Zeggada et al., 2017). Although creating an extra step for processing features and replacing the classifier seemed to improve the results of land-cover classification, these strategies did not significantly influence the feature extraction process, which is the critical component of a CNN model.

The next efforts concentrated on improving CNN models by two strategies: re-designing the architecture of convolution and pooling layers, and appending additional processing to the workflow to generate a high-level feature map (Zhao, Du and Emery, 2016; Fu et al., 2017; Geng et al., 2017; Zhou et al., 2017). From the ideas proposed by random forest and adaptive boosting, Zhang, Du, and Zhang (2016) reported the strategy associated with assembling multiple CNN models. Taking advantage of the integration of multiple CNN models, a resultant CNN model, with a similar architecture, was expected to outperform a single CNN model in land-cover classification. However, Bergstra and Bengio (2012) and Liu et al. (2018) reported that the integration of different CNN architectures might be challenging and involve intensive computing.

Although previous work using CNNs has demonstrated a powerful capability to perform land-cover classification, further research is needed to support LULC classification based on the results of land-cover scene mapping. *Land-cover* is the

representation of physical land types while *land-use* indicates the interaction between natural elements and human activities. Thus, LULC type attributes derived from remote sensing images are insufficient for directly predicting the functionality and organization of different land parcels.

Benchmark datasets and land-cover classification systems provide a solution for incorporating the representation of different land-cover categories into a CNN model to support LULC classification. A number of benchmark remote-sensing datasets have been developed in recent years (Cheng, Han and Lu, 2017; Xia et al., 2017; Zhou, et al., 2018; Shen et al., 2018). These benchmark datasets have organized and labeled aerial and satellite images as land-cover classes associated with LULC properties. Aside from the development of benchmark datasets, several techniques, such as data augmentation (Yu et al., 2017; Scott et al., 2017), stochastic large-patch sampling (Zhong, Fei and Zhang, 2016), and land-cover feature refinement (Zhong et al., 2017), have been reported to effectively generate the training samples that fit the representation of different LULC scenarios.

However, several limits are still observed in those labeled samples. First, the classification systems of LULC types are defined differently for specific departments, organizations, and institutes (Tchuenté, Roujean and De Jong, 2011; Shen et al., 2018). Considering the complexity of LULC shown in the remote sensing image, it is difficult to create a unified taxonomy associated with LULC. Moreover, semantic heterogeneity, including polysemy and synonymy, are commonly observed in different versions of land-cover classification systems. For example, the *residential area* in the UC Merced system (Yang and Newsam, 2010) may correspond to *low-density residence*, *medium-density*

residence, and *high-density residence* in the US. Geological Survey (USGS) land-cover classification system. Lastly, but most importantly, data labeling is always a time-consuming process (Zhou, 2018); it is tough and expensive to update and manage a large-scale benchmark dataset constructed for all possible LULC scenarios.

Recently, discovering how to convert from “big data, small task” to “small data, big task” has become a major concern in deep learning investigations. A number of papers have reported efforts to raise the transferability of high-level features generated by CNNs to moderate the need for tremendous amounts of labeled data (Nogueira, Penatti and dos Santos, 2015; Gu et al., 2018). The concept of these algorithms (e.g., zero-shot learning and one-shot-learning, among others) motivated the development of weakly supervised and rule-enhanced deep learning algorithms. In the remote sensing community, few investigations have combined a CNN model with semantic analysis to support unknown LULC classification. Jean et al. (2016) used poverty survey data to guide a CNN model to quantitatively predict a ‘poor’ class from land-cover classification. Yao et al. (2016) proposed a unified annotation system to assign concepts to the content of a remote sensing image. Cheng et al. (2017) created a visual bag model called Bag of Visual Words (BoVW) to semantically organize convolutional features in support of LULC classification. BoVW is a technique for image classification based on middle-level features. In BoVW, visual words refer to the vector of local image features derived from an image. Chai et al. (2017) proposed a new model called “visual geometry group network” (VGG-Net) to extract the representative features from various geographical scenes, and then performed a discriminant correlation analysis to fuse the extracted features for unknown LULC classification.

This chapter proposes an integrated framework called “image-semantic model” to identify LULC types from remote sensing images using a CNN-powered multi-label classification system with spatial weights, and then to convert the land-cover categories into an overall interpretation of the target LULC scenarios via the vector space model (VSM). The proposed framework comprises three sections: (1) building a benchmark remote sensing dataset within limited LULC categories, (2) performing multi-label LULC classification with a pretrained CNN based on the training images available in the benchmark remote sensing dataset, and (3) recognizing target LULC types by organizing the classified LULC categories and measuring the similarity between different images with the VSM.

The remainder of this chapter is organized as follows. Section 4.2 presents the benchmark data used for target LULC classification in coastline territories. Section 4.3 presents the details of the proposed image-semantic model, which include data preparation, multi-label land-cover classification, and semantically aware target LULC understanding. Section 4.4 reports the experimental results using the coastlines in California as the study area, while Section 4.5 summarizes the contributions of this chapter and provides related perspectives.

4.2 Benchmark Dataset for Coastal Scene Recognition

This research developed the benchmark remote-sensing dataset based on two stages: it creates a land-cover classification system, and then collects images corresponding to each land-cover category in the classification system. In the last three decades, the pattern recognition community witnessed a variety of algorithms designed for classification using low-level features, mid-level features, and high-level abstract

features (Schmidhuber, 2015; Guo et al., 2016). Meanwhile, it is commonly acknowledged that the content included in an image places a considerable influence on the performance of cutting-edge algorithms for image classification (Fei-Fei, Fergus and Perona, 2006; Russakovsky et al., 2015). If the training data may not reflect a landscape scene's characteristics, well-designed or fully fine-tuned models that densely rely on training data would be insufficient for precise classification (Shen et al., 2018). Thus, large-scale benchmark datasets such as Caltech (Fei-Fei, Fergus and Perona, 2006), (Everingham et al., 2010), ImageNet (Russakovsky et al., 2015), and Visual Genome (Krishna et al., 2017) have been established to organize these datasets to facilitate visual recognition, including classification, object detection, three-dimensional pose recognition, and semantic segmentation.

However, photos in these popular benchmark datasets are insufficient for land-cover classification. The taxonomies of these benchmark datasets were created from daily activities, a majority of which are irrelevant to the characteristics of land cover. A benchmark remote sensing dataset concerning specific LULC needs a new taxonomy that enables the description of land-cover systems. Moreover, scale is a significant attribute in both remote-sensing data processing and the definitions of LULC systems. The scale of similar LULC scenarios might vary significantly in the remote sensing image. For example, a place might be defined as *downtown* at one scale, but *commercial area* at a different scale. Lastly, although most importantly, the earth observation system has always confronted mediocre imaging conditions, resulting in noise such as clouds, fog, and abnormal contrast.

Considering the significant differences between photos and remote-sensing images, a number of influential benchmark datasets were developed to focus on land-cover classification, including UC Merced (Yang and Newsam, 2010), SAT-4/SAT-6 (Basu et al., 2015), SIRI-WHU (Zhao et al., 2016), RSSCN7 (Zou et al, 2015), RSC11 (Zhao, Tang and Huo, 2016), RSI-CB (Li et al., 2017), WHU-RS19 (Shao, Yang and Xia, 2013), AID (Xia et al., 2017), PatternNet (Zhou, et al., 2018), and NWPU-RESISC45 (Cheng, Han and Lu, 2017). Shen et al. (2018) provide a detailed table describing these benchmark datasets.

Table 7

Brief information on remote-sensing datasets' existing benchmarks (Shen et al., 2018).

Benchmark name	Total images	Total class	Average images per class	Image size	Spatial resolution	Sources	Scale variation	Complex scene	SOC
UC Merced dataset	2100	21	100	256*256	0.3m	U.S. Geological Survey	No	Yes	No
SAT-4/SAT-6	405000	6		28*28		USDA Farm Service Agency	No	No	No
SIRI-WHU	2400	12	200	200*200	2m	Google Earth	No	No	No
RSSCN7 dataset	2800	7	400	400*400		Google Earth	No	No	No
RSC11 dataset	1232	11	100	512*512	0.2m	Google Earth	No	No	No
RSI-CB	36707 24747	45 35	800 690	128*128 256*256	0.22m-3m	Google Earth	No	No	No
WHU-RS19 dataset	1005	19	≈ 52	600*600	0.5m	Google Earth	No	Yes	No
AID	10000	30	220~420	600*600	0.5m-8m	Google Earth	No	Yes	No
PatternNet	30400	38	800	256*256	0.062m-4.69m	Google Earth	No	Yes	No
NWPU-RESISC45 dataset	31500	45	700	256*256	0.2m-30m	Google Earth	No	Yes	No
CSRS-SIAT	70000	70	1000	512*512	Varied	Google Earth	Yes	Yes	Yes

*SOC: Semantically-organized category

To support target LULC classification in coastal areas, I reports the development of a new land-cover classification system for coastal scenes by unifying both the existing land-cover classification systems, including the USGS Land Use and Land Cover Classification System and the NOAA Coastal Change Analysis Program (C-CAP), as well as geo-spatial query interface and gazetteers including USTopographic, USGS (U.S. Geological Survey) Geographical Name Information System (GNIS). The land-cover categories selected in the benchmark dataset include airport, beach, circular farmland, cloud, commercial area, dense residential area, desert, forest, freeway, golf course, running track, harbor, industrial area, transportation intersection, island, meadow, mountain, palace, parking lot, pier, railway, rectangular farmland, river, runway, sea ice, ship, iceberg, sparse residential area, stadium, storage tank, tennis court, terrace, thermal power station, wharf, water, and wetland. Accordingly, the research collected 700 images for each land-cover category from the benchmarks listed in Table 1. These 700 images represent the *wave* land-cover type from Google Earth Pro because wave might affect the recognition of water. Figure 18 illustrates selected image samples for each land-cover category in the new benchmark dataset.

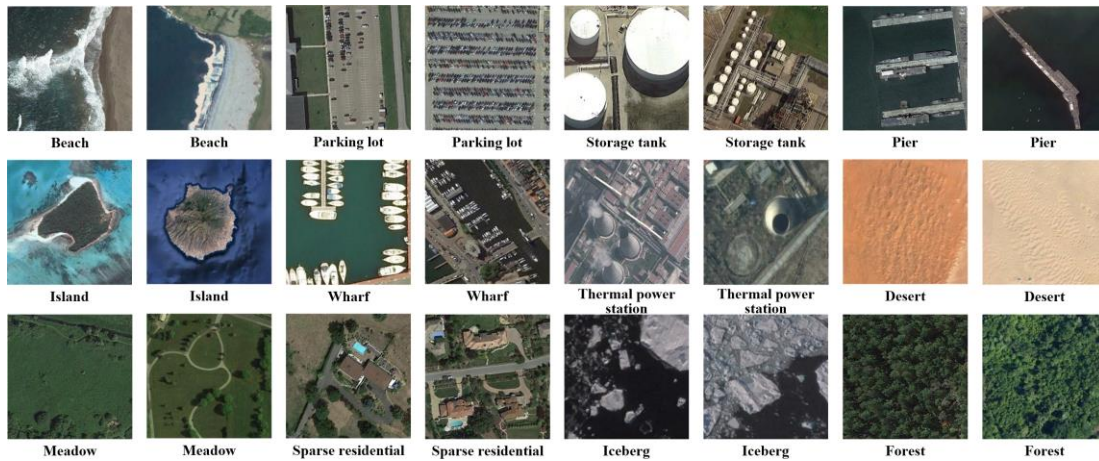


Figure 18. *Illustration of selected image samples in the benchmark dataset.*

4.3 CNNs-enhanced Image-semantic Model

4.3.1 Multi-label Land-cover/land-use Classification.

Multi-label LULC classification aims to label multiple LULC classes to a remote sensing image. In this chapter, I propose several steps to conduct multi-label LULC classification

Numerous flat or planar surface assessments demonstrate the consequences of the composition and configuration of urban land-cover (i.e., land system architecture [Turner, 2017]) on land surface temperature (LST) and above-ground air temperature for Phoenix, Arizona (Li et al., 2016; Myint et al., 2013; Kamarianakis et al., 2017). For the most part, these works demonstrate that the compactness of individual land-cover patch and the clustering of the same patch type can increase or decrease diurnal temperatures, depending on the land-cover type. The built urban environment, however, is repeat with vertical structures (i.e., buildings, trees) that affect climate within the canopy layer in various ways, such as shading, wind tunnels, and sky view. This vertical dimension is

central to research on turbulence and flux dynamics as undertaken in urban climatology (Unger, 2004, 2009; Coseo & Larsen, 2014).

4.3.1.1 Multi-label classification

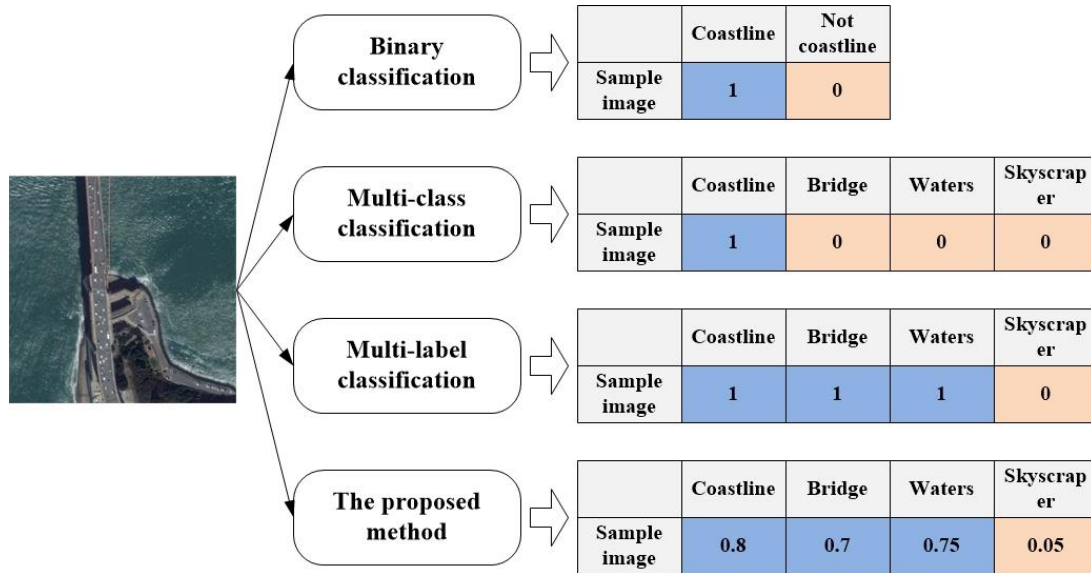


Figure 19. Comparison of binary classification, multi-class classification, multi-label classification, and multi-label classification with spatial weights.

Figure 19 illustrates four types of land-cover classification, including binary classification, multi-class classification, multi-label classification, and multi-output classification assuming I have C total land-cover classes and L total labels (or variables) assigned to the content of a remote-sensing image. These four classifications are discussed below.

Binary classification addresses the classification problem when $C = 2$ and $L = 1$. For example, the sample image in Figure 2 is labeled merely as either coastline or non-coastline. Multi-class classification deals with the classification problem when $C \geq 2$ and $L = 1$. For example, the sample image in Figure 2 is classified into one of four land-cover types.

Multi-label classification supports the assignment of the content of a remote-sensing image to multiple classes with no restrictions on the total number of land-cover classes to which it can be assigned. Multi-label classification addresses the classification problem when $C \geq 2$ and $L \geq 1$; in other words, both land-cover classes are recognized in the image. Multi-label classification with spatial weights addresses the classification problem when $C \geq 2$ and $L \geq 1$ and each L is weighted. In Figure 19, all generated land-cover types are quantitatively normalized as scores within the interval 0~1.





An improvement in spatial resolution increases the number of LULC types visible in one remote-sensing image. For example, sand might be the only land-cover type recognizable along the coastline in a medium-resolution image, whereas, in a high-resolution version of the same image, a variety of LULC types (e.g., roads, piers, and buildings) may be visible, meaning neither binary classification nor multi-class classification is fitting for the high-resolution remote-sensing image. Binary classification and multi-class classification can only generate one LULC type for a remote sensing image, the scene in which may contain multiple LULC classes. Additionally, LULC might vary according to different portions of similar land-cover types. For example, a coastline with few palms would be different from another coastline scene that contains mangrove would be different although these two coastline scenes contain other similar LULC elements, including forest, sand, and water. Thus, compared with the multi-label classification that equally weights every land-cover element, multi-label classification with spatial weights becomes a decisive strategy for studying complicated target LULC scenarios in a high-resolution remote-sensing image.

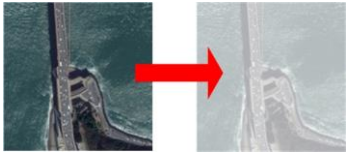

4.3.1.2 Data augmentation

Data augmentation is an important strategy for enhancing the robustness and transferability of a CNN model for feature extraction and classification (Taylor and Nitschke, 2017; Hernández-García and König, 2018). However, some data augmentation methods might not be helpful for land-cover classification. Considering the characteristics of LULC in remote-sensing imagery, this chapter applies six data augmentation methods: rotation, flip, scale, contrast, brightness, and the introduction of cloud and fog noise, the details of which are listed in Table 3.

Table 8

Brief information on data augmentation.

Data Augmentation	Methods	Demo
Rotation	Generating 36 new images through rotating the original image every 10 degree.	
Flip	Generating 2 new images through flipping the original image over horizontal dimension and vertical dimension.	
Scale	Generating 4 new images through scaling the original image by 1:4, 1:2, 2:1, and 4:1.	
Contrast	Generating 4 new images through modifying the original image with separate weighting parameters of AGCWD*: 0.2, 0.4, 0.6, and 0.8.	

Brightness	Generating 4 new images through randomly modifying the brightness of the original image.	
Cloud and fog noises	Generating 4 new images through randomly adding different types of cloud and fog noise to the original image.	
Total number of images after augmentation	$36 \text{ rotated images} \times 4 \text{ scaled images} \times 4 \text{ contrast enhanced images} \times 4 \text{ brightness modified images} \times 4 \text{ noises incorporated images} = 9,216 \text{ images}$ $2 \text{ flipped images} \times 4 \text{ scaled images} \times 4 \text{ contrast enhanced images} \times 4 \text{ brightness modified images} \times 4 \text{ noises incorporated images} = 512 \text{ images}$ Total images: $9,216 + 512 = 9,728 \text{ images}$	

*AGCWD: Adaptive Gamma Correction with Weighting Distribution (Huang, Cheng and Chiu, 2013)

Classical visual feature descriptors such as scale-invariant feature transform (SIFT) (Lowe, 2004) and speeded-up robust features (SURF) (Bay, Tuytelaars and Van Gool, 2006) reported the significance of rotation and scale variation on visual recognition. Moreover, scale is a fundamental attribute in a remote-sensing data analysis. Thus, the research reported in this chapter generated new training samples through rotating, flipping, and scaling the original image. Atmospheric conditions and the distance between ground surfaces and satellite sensors may lead to some loss in the electromagnetic energy obtained by sensors. The loss of electromagnetic energy results in changes in the image contrast, brightness, and intensity distribution. Thus, this work generated new images through modifying the brightness and changing the contrast of each original image. Finally, to raise the robustness of a CNN model to deal with noise, the introduction of pepper noise, Gaussian noise, and mosaics is commonly used for data augmentation. However, these types of noise are rarely observed in the rectified remote-sensing images

sent to end-users, whereas clouds and fog are. Thus, cloud noise and fog noise, rather than signal noise and mosaics, were randomly added to the original remote-sensing image.

4.3.1.3 PNASNet

This research applied a state-of-the-art CNN called Progressive Neural Architecture Search-PNAS (Liu et al., 2018) to conduct land-cover classification. Advanced CNNs such as Inception-ResNet and DenseNet are inefficient to automate hyperparameters (e.g. learning rate, the number of filter, convolutional size, etc.) during neural network architecture optimization (Liu et al., 2018; Liu, Simonyan and Yang, 2018). Hyperparameters, which were manually designed with expert experiences, heavily influenced the learning rate scheduling as well as determining and optimizing the neural network architecture. To save the labor and time required to design a complex neural network architecture, the AutoML project proposed a new strategy (PNAS Progressive) to automatically generate an optimized neural network architecture. The architecture search space designed in PNAS works similar to that proposed in Neural Architecture Search-NAS (Zoph and Le, 2016).

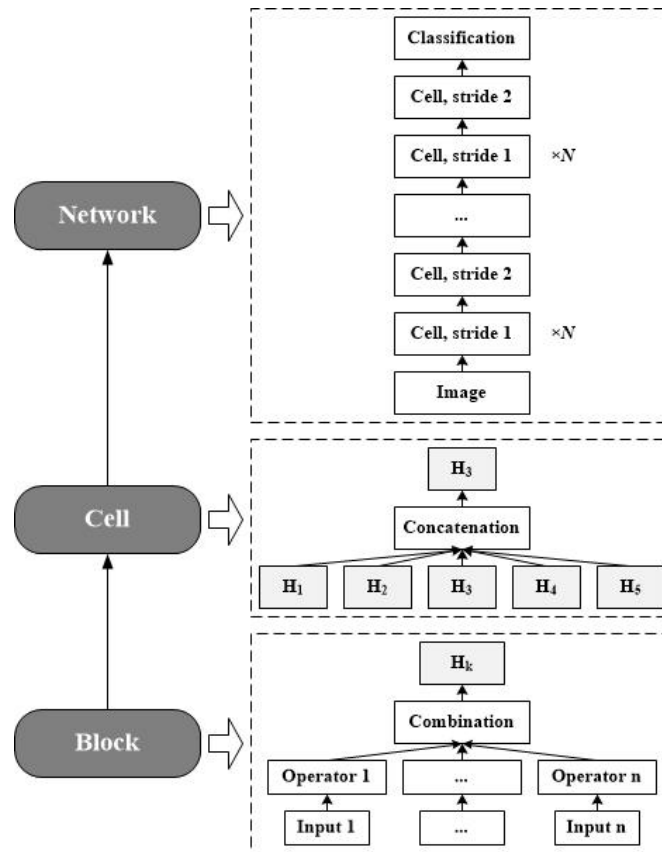


Figure 20. *Illustration of NAS strategy (Zoph and Le, 2016).*

Figure 20 illustrates the general structure of the architecture search space that comprises networks, cells, and blocks. The network is a stacked layer that includes a normal cell and a reduction cell. Each cell possesses a different structure that includes five blocks, meaning a cell’s output integrates these five blocks’ outputs. The block is a box that comprises multiple inputs and their corresponding operators (e.g., filters), and a block’s input may be the outputs generated by a cell or a block. A block’s output combines outputs from multiple operators. Moreover, the operators in a block may be a convolution with predefined dimensions and dilation rates, a pooling designed with different downsampling strategies, or a classifier. In the process of feature learning, NAS automatically modifies the network architecture constructed by cells and blocks to

identify the one that produces the best-fitting results. The following are the available operators:

- 3x3, 5x5, and 7x7 depth wise separable convolution (Chollet, 2017)
- 1x7 followed by 7x1 convolution
- 3x3 average pooling, max pooling, and dilated pooling

The CNN architecture is optimized by architecture search space, and the NASNet outperforms humans' ability to design CNN architecture. However, intensive computation is a major concern when extensively implementing these strategies for a variety of applications, and thus, PNASNet aims to speed up the computational load using a progressive and iterative solution. In detail, this solution's workflow is described below.

Step 1. Train all N cells that only include one block and select the most promising $N_1 (N_1 \leq N)$ cells based on their training scores;

Step 2. Expand the N_1 selected cells into two-block cells;

Step 3. Train all N_1 two-block cells and select the most promising $N_2 (N_2 \leq N_1)$ cells based on their training scores;

Step 4. Expand the N_2 selected cells into three-block cells;

Step 5. Training all N_2 three-block cells and selecting the most promising $N_3 (N_3 \leq N_2)$ cells based on their training scores;

Step 6. Expanding the N_3 selected cells into four-block cells;

Step 7. Train all N_3 four-block cells and select the most promising $N_4 (N_4 \leq N_3)$ cells based on their training scores;

Step 8. Expand the N_4 selected cells into five-block cells;

Step 9. Train all N_4 five-block cells, selecting the best five-block $N_5 (N_5 \leq N_4)$ cells, and then increase the number of cells in the stacked normal cell.

Compared to five-block cell training, the initiation of one-block cell training proposed by PNASNet gradually removes the valueless cells to more significantly decrease the computing complexity. Moreover, a number of efforts reported the classification performance of PNASNet based on two well-known, large-scale benchmark datasets: CIFAR-10 and ImageNet. Their experimental results acknowledge the significance of PNASNet in maintaining state-of-the-art classification accuracy while decreasing five to eight computational times the cost of feature learning. PNASNet can remain a tradeoff between computational load and classification accuracy.

4.3.1.4 Image gridding

Before performing land-cover classification with PNASNet, the research applied image gridding to preprocess the remote sensing image because of two major reasons. First, varied portions of different land-cover in a remote-sensing image may considerably affect land-classification results. Recently, the techniques of semantic segmentation and instance segmentation have dealt with the variations of position and portion of every object included in an image (Long, Shelhamer and Darrell, 2015; Dai, He and Sun, 2016). However, a well-developed benchmark dataset that supports land-cover semantic segmentation has not yet been made available. Thus, the limitation of datasets labeled at the pixel level inspired adoption of the classical strategy for classification, which involves dividing the whole image into multiple sub-regions (segmentation) and then classifying each sub-region. Figure 4 displays our proposed image gridding strategy, which follows the workflow detailed below.

Assuming an original image is $Img(s, \frac{x}{2^k}, \frac{y}{2^k})$, x and y refer to the horizontal and vertical dimensions, s refers to the index of an image sub-region, and k refers to the image gridding scale; the original image has $s = 0$ and $k = 0$.

Step 1. Classify $Img(s, \frac{x}{2^k}, \frac{y}{2^k})$ and select all promising categories that have a classification score higher than θ , where $\theta = 0.6$;

Step 2. If the number of selected categories is greater than 1, skip to Step 3; otherwise, skip to Step 4.

Step 3. Divide the previous image region into four sub-regions: $Img(1, \frac{x}{2^{k+1}}, \frac{y}{2^{k+1}})$, $Img(2, \frac{x}{2^{k+1}}, \frac{y}{2^{k+1}})$, $Img(3, \frac{x}{2^{k+1}}, \frac{y}{2^{k+1}})$, and $Img(4, \frac{x}{2^{k+1}}, \frac{y}{2^{k+1}})$; for each sub-region, return to Step 1.

Step 4. Combine the categories generated from the whole image with those generated from the multiple sub-regions.

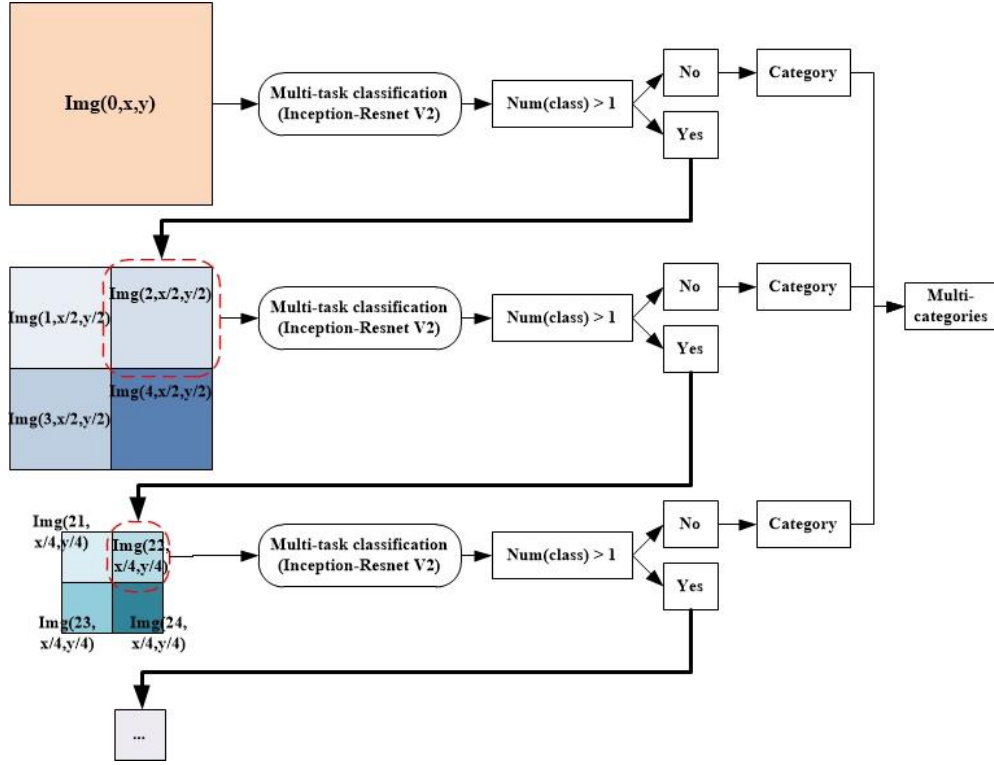


Figure 21. *Image gridding.*

In the example shown in Figure 21, the method first performs multi-label classification on the original image, $Img(0, x, y)$. If the amount of promising categories equals 1, it becomes the land-cover class assigned to the image $Img(0, x, y)$; otherwise, the research divides the original image into four parts: $Img(1, \frac{x}{2}, \frac{y}{2})$, $Img(2, \frac{x}{2}, \frac{y}{2})$, $Img(3, \frac{x}{2}, \frac{y}{2})$, and $Img(4, \frac{x}{2}, \frac{y}{2})$. For the second part, $Img(2, \frac{x}{2}, \frac{y}{2})$, the process performs multi-label classification and selects all promising categories. Then, the research classifies the second sub-region, $Img(2, \frac{x}{4}, \frac{y}{4})$, into various promising categories. If the amount of promising category equals 1, it becomes the land-cover class assigned to sub-region $Img(4, \frac{x}{4}, \frac{y}{4})$; otherwise, the research divides this sub-region into four sub-regions, and so on. In the end, the research compiles all promising categories into a land-cover

scene feature vector. The details for processing this feature vector are presented in the following section.

4.3.2 Image-semantic Model.

A semantic analysis in this research studies and discovers the meaning of textual information. In a remote-sensing image that contains multiple LULC types, an image's content or land-use might be hidden in the LULC classification. For example, land-cover that includes land uses like buildings and piers may be classified as an industrial harbor or an entertainment beach.

4.3.2.1 Land-cover category frequency and inverse image frequency

To answer the question about target LULC in a remote-sensing image, this section focuses on transforming the land-cover categories derived from the remote-sensing image into a feature vector space to support a context-based semantic analysis model.

A VSM identifies each text term or individual as a vector in a multi-dimensional space and then measures the similarity between each set of two terms or individuals (Turney and Pantel, 2010; Mikolov, Yih and Zweig, 2013). The multi-dimensional space comprises a set of linearly independent vectors, and each vector denotes one dimension in the vector space. State-of-the-art VSMs were designed differently according to three representations of text information: term-document, word-context, and pair-pattern. Typically, the term-document model detects the meaning of each document, the term-context model evaluates the meaning of each term, and the pair-pattern model assesses the similarity between two complicated patterns. In this chapter, the research focuses on the content (or target land-cover/land-use) of a remote sensing image. The hypothesis is that the relationship between land-cover categories and target LULC type is analogous to

that of the term and document in a term-document VSM. Figure 22(A) illustrates this mapping between the term-document VSM and the image land-cover category-image.

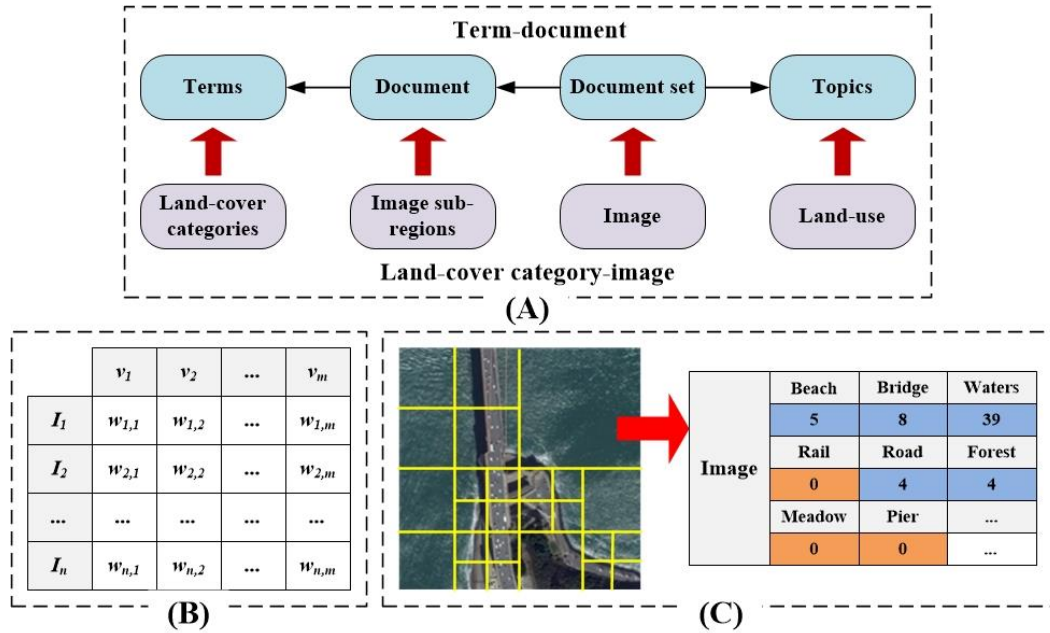


Figure 22. Matrix of VSM; (A) land-cover category frequency and inverse image frequency; (B) matrix structure; (C) an example of a VSM matrix.

The land-cover category and the image sub-regions are analogous to the term and document in the term-frequency and inverse document frequency model, which are depicted in Figure 22(A). Based on the image gridding results, for remote-sensing image I_i , the x th (land-cover) category frequency ($cf_{i,x}$) is calculated by the following equation:

$$cf_{i,x} = \begin{cases} 1, & k = 0 \\ \sum_{c=1}^k \frac{freq_c(w_{i,x})}{2^{c+1}}, & k \geq 1 \end{cases} \quad (1)$$

where k is the image gridding scale shown in Figure 21. $freq_c(w_{i,x})$ denotes the number of land-cover classes, while $w_{i,x}$ appears in the sub-regions of scaled image c .

Specifically, the category frequency produced by Equation (1) is locally normalized.

Then, the process converts the category frequency to $cf_{i,x} \times 2^{k_{max} \times 2}$, with k_{max} being the largest image gridding scale.

Inverse image frequency is calculated by the following equation:

$$imf_{i,x} = \log\left(\frac{N}{cf_{i,x} \times 2^{k_{max} \times 2}}\right), N = \begin{cases} 1, k_{max} = 0 \\ 2^{k_{max} \times 2}, k_{max} \geq 1 \end{cases} \quad (2)$$

where N is the total number of sub-regions in remote-sensing image I_i . Then, to quantitatively weight the importance of each category ($w_{i,x}$) in remote-sensing image I_i , the approach calculates the category frequency and inverse image frequency.

$$w_{i,x} = cf_{i,x} \times imf_{i,x} \quad (3)$$

4.3.2.2 Semantically aware land-use similarity analysis

Assuming the n -dimension vector space of classification categories and images are $V \subseteq \{v_1, v_2, \dots, v_m\}$ and $I_i = (w_{i,1}, w_{i,2}, \dots, w_{i,m})$, respectively, the category (term)-image (document) model creates a matrix that represents land-cover categories and images as rows and columns (see Figure 22(B)). The similarity among LULC scenarios in two images, I_i and I_j , is computed by the following five methods: Euclidean distance, inner product, cosine similarity, dice similarity, and Jaccard similarity. The similarity measured by Euclidean distance (and inner product) is expressed by the following equation:

$$\begin{cases} sim(I_i, I_j) = \sqrt{\sum_{x=1}^m (w_{i,x} - w_{j,x})^2} \\ sim(I_i, I_j) = |I_i \cap I_j| = I_i \times I_j = \sum_{x=1}^m w_{i,x} \times w_{j,x} \end{cases} \quad (4)$$

Neither Euclidean distance nor inner product consider the number of the number of category, nor the total number of land-cover categories included in a remote sensing image. For example, inner product may measure a five-category image that contains three common classes equal to a ten-category image that contains three common classes. Therefore, cosine similarity, Jaccard similarity, and dice similarity were proposed to supportively measure the similarity between two images while considering the total number of classes embodied in an image.

Cosine similarity is computed as follows,

$$sim(I_i, I_j) = \frac{|I_i \cap I_j|}{|I_i| \times |I_j|} = \frac{\sum_{x=1}^m w_{i,x} \times w_{j,x}}{\sqrt{\sum_{x=1}^m (w_{i,x})^2} \times \sqrt{\sum_{x=1}^m (w_{j,x})^2}} \quad (5)$$

Jaccard similarity is computed as follows,

$$sim(I_i, I_j) = \frac{|I_i \cap I_j|}{|I_i \cup I_j|} = \frac{\sum_{x=1}^k w_{i,x} \times w_{j,x}}{\sum_{x=1}^k (w_{i,x})^2 + \sum_{x=1}^k (w_{j,x})^2 - \sum_{x=1}^k w_{i,x} \times w_{j,x}} \quad (6)$$

Dice similarity is computed as follows,

$$sim(I_i, I_j) = 2 \frac{|I_i \cap I_j|}{|I_i| + |I_j|} = 2 \frac{\sum_{x=1}^k w_{i,x} \times w_{j,x}}{\sqrt{\sum_{x=1}^k (w_{i,x})^2} + \sqrt{\sum_{x=1}^k (w_{j,x})^2}} \quad (7)$$

The similarity between two LULC classifications in remote-sensing images is measured differently from the similarity between two documents. If two remote-sensing images contain different LULC, their land scenarios are distinct. Thus, the cosine, Jaccard, and dice similarity methods cannot distinctly, adaptively identify two images when they have different LULC types—in other words, when $w_{j,x} \neq 0$ and $w_{j,x} = 0$. Thus, the approach modified Equation (5) to fit for image similarity:

$$sim(I_i, I_j) = \begin{cases} \frac{\Theta_x \times \sum_{x=1}^m w_{i,x} \times w_{j,x}}{\sqrt{\sum_{x=1}^m (w_{i,x})^2} \times \sqrt{\sum_{x=1}^m (w_{j,x})^2}} \\ \Theta_x = \begin{cases} 1, w_{i,x} \times w_{j,x} \neq 0 \\ 0, w_{i,x} \times w_{j,x} = 0 \end{cases} \end{cases} \quad (8)$$

In the example shown in Figure 22(C), the image gridding scale is 2. Blue boxes denote the sub-region that contains the *sand* (land-cover) class, and the normalized sand category frequency ($1/2^{2+1} + 2/2^{1+1} = 5/8$). Assuming the k_{max} is 2, the new sand category frequency ($cf_{i,x}$) is $5/8 \times 2^{2 \times 2} = 10$; then, the inverse image frequency ($imf_{i,x}$) for the sand class is $\log(2^{2 \times 2} / 10) = 0.2041$. Thus, the importance weight of category i,x ($w_{i,x}$) equals 10×0.2041 , or 2.041.

4.4 Experiments

Human populations deform a coastline's natural landscape, while artificially constructed barriers distributed along the coastline deposit an unnatural footprint on the coastal environment. Coastline pollution and the loss of biodiversity have increased alongside the rapid development of energy, commercial manufacturing, and transportation industries, among others. Coastal territories' LULC is a key indicator for measuring economic values, biodiversity, and the ecosystem (Martínez et al., 2007; Murray et al., 2013). Thus, a number of studies report efforts to power coastline landscape analysis with GIS and remote-sensing imagery (Gens, 2010; Alexakis et al., 2011). These studies might be insufficient to support a semantically aware coastline land-use classification due to the limited spatial resolution of multispectral imagery and the lack of domain knowledge concerning LULC. This experiment evaluated the proposed image-semantic model for land-use image classification based on a high spatial-resolution image.

4.4.1 Study Area.

The red polyline in Figure 23(A) is the coastline in California—the study area used for evaluating the proposed methodology—which runs from the US-Mexico border to the California-Oregon border. The coastal images included a variety of urban, rural, and natural land-cover/land-use, such as downtown, sand, shore, pier, road, and so on. Figure 23(B) illustrates selected image samples from the 1,000 high-resolution images (around 5 meter) the research collected from Google Earth Pro. Although all the images contain the coastline, their content comprises different land-cover scenarios, meaning the land-use represented by each image might vary accordingly.

As mentioned in Section 4.2, the benchmark dataset includes 37 land-cover categories. Moreover, a new category called waves were created to support land-cover classification along a coastline because waves were observed in a significant number of the collected images. Based on the training images in the benchmark dataset, the methodology extended the number of training images though the data augmentation methods introduced in sub-section 4.3.2.

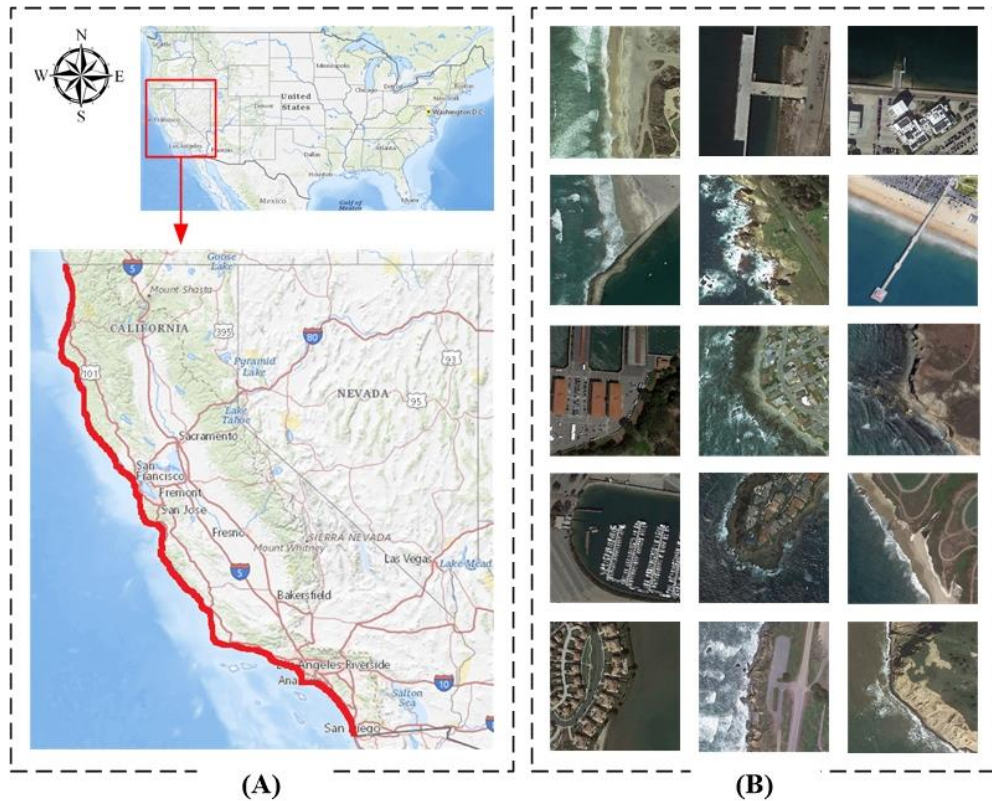


Figure 23. Visualization of study area; (A) study area map; (B) samples of coastal LULC.

Current research on fine-tuning deep neural networks report that a constant learning rate schedule may be tough to manually design for hyperparameter optimization since the performance of this schedule heavily relies on the representation of data and the classification problem itself (Hoffer, Hubara and Soudry, 2017). Adaptive learning rate methods schedule learning rates for each parameter to obtain a perfect tradeoff between feature loss and overfitting. This chapter reports the use of a state-of-the-art method called Adaptive Moment Estimation (Adam) to schedule adaptive learning rates because Adam outperformed other similar methods overall, including RMSprop, Adadelata, and Adagrad (Ruder 2016).

Two separate CNN models were created for binary image classification and multi-image classification. In binary classification, the process fine-tuned the PNSANet

with the training images, solely to include the coast category in the benchmark dataset. Then, fine-tuned model classified each collected image into a binary result—coastline or non-coastline.

Multi-image classification includes multi-class classification, multi-label classification, and multi-label classification with spatial weights. The method fine-tuned the PNSANet with all categories' training images in the benchmark dataset. Then, I selected the top scoring class as the output of the multi-class classification; otherwise, the process selected the top-t scoring classes as the output of multi-label classification, and multi-label classification with spatial weights. Finally, the research used the proposed VSM to convert the output of the multi-label classification into a land-use classification.

4.4.2 Coastline Land-use Similarity Analysis.

This subsection describes the measurement of the similarity of LULC classifications between two collected images. The process selected 500 odd-numbered images from the total 1,000 collected images and sequentially combined each two images into a set, resulting in 200 sets, each containing two images. Then, the approach computed the similarity among those two images in every group using binary classification, multi-class classification, multi-label classification, and multi-label classification with spatial weights. Table 8 presents the comparison of the coastline scene similarity analysis by various PNASNet-enhanced classification strategies. The column *Precision* assessed how many of the 500 selected images were correctly recognized as coastlines by various classification strategies. The column *Similarity* evaluated the similarity analysis between two images included in each group. The process used the

modified cosine distances expressed in Equation (8) to measure the similarity between land-use classifications of two separate images.

The similarity of these results between those generated by other classification strategies is expressed in the following equation:

$$\begin{cases} sim_{bin}(I_i, I_j) = L_{bin}(I_i) - L_{bin}(I_j) \\ sim_{cla}(I_i, I_j) = L_{cla}(I_i) - L_{cla}(I_j) \\ sim_{lab}(I_i, I_j) = \sum(L_{lab}(I_i)_k - L_{lab}(I_j)_k) \end{cases} \quad (9)$$

where $sim_{bin}(I_i, I_j)$, $sim_{cla}(I_i, I_j)$, and $sim_{lab}(I_i, I_j)$ denote the similarity of two images, I_i and I_j , generated by binary classification, multi-class classification, and multi-label classification, respectively. $L_{bin}(I_i)$ denotes the class number assigned to I_i by binary classification, equaling either 1 or 0. $L_{cla}(I_i)$ denotes the class number assigned to I_i by multi-class classification, which ranges from 1 to 37 due to our total 37 land-cover categories. The technique recognized two images as similar when $sim_{bin}(I_i, I_j) = 0$ or $sim_{cla}(I_i, I_j) = 0$.

In Equation (9), $L_{lab}(I_i)_k$ denotes the k th ($k \leq 5$) vector of the feature vector assigned to I_i by multi-label classification. Similar to multi-label classification, the approach also ignored the distance of the *coast* class vector between two images.

Table 9

Comparison of coastline scene analysis by various PNASNet-enhanced classification strategies.

	Precision				Recall				F-scores			
	1*	2*	3*	4*	1*	2*	3*	4*	1*	2*	3*	4*
1-100 groups	0.91	0.72	0.94	0.94	0.50	0.54	0.78	0.83	0.32 27	0.30 86	0.42 63	0.44 08

101-200 groups	0.93	0.76	0.93	0.93	0.48	0.57	0.88	0.92	0.3166	0.3257	0.4522	0.4625
1-200 groups	0.92	0.74	0.935	0.935	0.49	0.555	0.83	0.875	0.3197	0.3174	0.4397	0.452

1* Binary classification

2* Multi-class classification

3* Multi-label classification

4* Multi-output classification (Proposed method)

Binary classification can only determine whether an image’s content contains a coastline without considering the details of coastal scenes. PNASNet-powered binary classification reached 98% in the precision of coastline similarity measurement for the selected 200 groups. However, although the process discovered the components of each image, binary classification could not support the characterization of the intra-class diversity apparent in coastline images. For example, binary classification could not distinguish the image groups in Figure 24(B), although it correctly recognized that all images in these three groups contained coastlines. In Figure 24(B), each image group was enclosed by a purple box.

As opposed to binary classification, multi-class classification assigned an image to one of the 37 land-cover categories in the benchmark dataset, thus leading to a significant decrease in the precision of the coastline image classification generated by multi-class classification. The difference in the precision generated by both binary and multi-class classification revealed some disadvantages of traditional CNNs. First, the features obtained by a CNN model designed for a small number of land-cover classifications could not transfer to a new classification task on a large number of land-cover classes. In Table 8, the classification precision for coastlines alone is reduced by

multi-class classification. Second, semantic annotation becomes a critical challenge for data preparation. For example, binary classification identifies the images shown in Figure 24(A) as the *coast* type. However, the multi-class classification classified these images from left to right as *coast, pier, coast, dense residential, harbor, and harbor*.

The images in Figure 24(B) explain why the multi-label classification (or the proposed image-semantic model) was largely competent in supporting land-use recognition. In this experiment, 3 was set as the default image gridding scale. Although those two images in the left and middle groups have approximately similar land-cover/land-uses on a global scale, they represent two separate land-use scenarios. In the left group, PNASNet recognized a sparse residential area in the right image that was not pictured in the left image. Thus, the LULC scenario in the left image is a purely natural landscape with no components of residential objects. Similarly, in the middle group, PNASNet recognized a bridge in the left image that was not included in the right one, meaning the land-use of the left image might contain some transportation functions.

In Table 6, multi-label classification with spatial weights outperformed multi-label classification to some extent because of the quantitative weighting of each land-cover class with the proposed VSM. In the right group of Figure 24(B), multi-label classification recognized and equally weighted the impact of the sparse residential area and sand. Thus, multi-label classification identified these two images as mutually similar. The proposed image-semantic model used the VSM to quantitatively weight the impact of each LULC category and then differentiated these two images according to the LULC scenario.

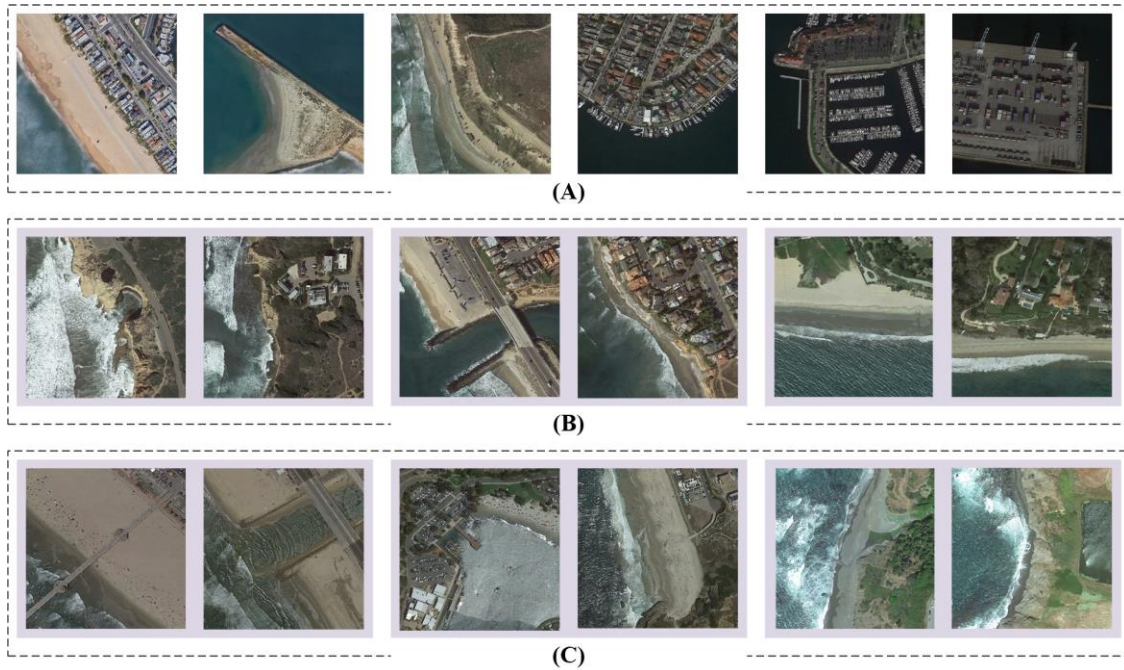


Figure 24. *Illustration of the selected image pairs used for scenic similarity evaluation. (A) Selected examples on different coastline LULC types. (B) Selected examples on coastline LULC type measurement. (C) Illustration on challenges for the proposed method.*

Moreover, the proposed image-semantic model presents several challenges that should be addressed. The first challenge is the limited capability of distinguishing an interclass that carries similar features. The left pair in Figure 24(C) illustrate that PNASNet generally could not effectively differentiate a bridge and pier when they comprise an incomplete structure. Noise accounted for the second reason, making the proposed method limited in terms of LULC classification. As depicted in the middle pair of Figure 24(C), PNASNet mistakenly recognized the abnormal water surface as sand. The last challenge is transferability; in the right pair of Figure 24(C), the land-cover type

in the left pair's beaches was not defined in the benchmark data. Strategies for recognizing these unknown LULC types are a new topic worthy of future investigation.

4.4.3 Coastline land-use scene retrieval.

This subsection evaluates the proposed image-semantic model based on the application of image retrieval. Image retrieval aims to search and retrieve the needed images from a large-scale image dataset (Liu et al., 2007). Based on the remaining 500 even-numbered images, twelve images were selected, and the rest were left as samples in the image database. For each selected image, the process retrieved all images relevant to its content based on binary classification, multi-class classification, multi-label classification, and the proposed image-semantic model.



Figure 25. *The selected images for image retrieval.*

Figure 25 displays the eleven images used in image retrieval-based evaluation. The target LULC scenarios portrayed in these eleven images are different, and each image represents an independent class. Table 9 lists the target LULC classes shown in each image.

Table 10

Target LULC in the selected images.

Images	Target LULC type	LULC annotations	Amount of GTS*
Image 1	Natural sand coastline without vegetation	Water, sand	90
Image 2	Industrial harbor	Commercial area (or industrial area), harbor, waters	20
Image 3	Entertainment harbor	Waters, wharf	15
Image 4	Residential coastline with vegetation	Waters, meadow, forest, sparse/dense residential area (or commercial area)	40
Image 5	Industrial coastline	Waters, industrial area, parking lot	11
Image 6	Residential coastline without vegetation	Waters, sand, meadow, forest, sparse/dense residential area	35
Image 7	Residential coastline with a pier	Waters, pier, sand, sparse/dense residential area (or commercial area)	11
Image 8	Residential coastline without sand or vegetation	Waters, sparse/dense residential area (or commercial area)	6
Image 9	Natural coastline with a freeway	Waters, sand, (or meadow/forest), freeway	56
Image 10	Natural sand coastline with vegetations	Waters, sand, meadow (or forest)	58
Image 11	Natural coastline without sand	Waters, meadow (or forest)	30

*GTS: ground truth samples.

Table 10 illustrates the retrieval results for the selected images shown in Figure 25 with various classification strategies.

Binary classification identified all images in the database as *beach* or *non-beach* classes. Thus, this classification strategy produced the same precision, recall, and accuracy among all twelve selected images. Above all, the results in Tables 8 and 10 acknowledge that binary classification may not support the representation of the land-use scenario included in a remote-sensing image.

Multi-class classification has been commonly reported in previous works related to content-based image retrieval-CBIA (Liu et al., 2007). The top class generated by PNASNet categorized the selected eleven images into the following classes: *beach*,

harbor, harbor, beach, harbor, beach, pier, dense residential area, freeway, beach, and beach. Thus, the retrieval accuracy varied significantly for each image. For the first, fourth, tenth, and eleventh images displaying similar content, multi-class classification could not distinguish their detailed land-cover differences. Otherwise, although multi-class image classification reached a relatively higher precision and recall on the ninth image, the classification strategy only recognized these images' content as *freeway* rather than other LULC types.

The results generated by binary classification and multi-class classification prove that merely one land-cover annotation or merely one land-cover class is not viable to characterize complicated land-cover in a remote sensing image—much less hidden land-use. These results also justify the significance of semantics in LULC recognition from high-resolution remote-sensing images.

During the image retrieval experiment, the multi-label classification—the proposed image-semantic model—collected the top five classes rather than the top class selected by multi-class classification. The *beach* class was included in the top-five class results generated by PNASNet for the selected eleven images. From the results shown in Tables 3 and 5, multi-label classification significantly outperformed the above two classification strategies in retrieving the required images. However, PNASNet's performance in multi-label classification were much poorer than were those previously reported—a difference that might have two main causes. First, the LULC diversities were tiny among the collected images. For example, the first, ninth, tenth, and eleventh images were generally labeled as the same LULC type in the existing benchmark dataset. If the retrieved images related to these four images were grouped into one LULC category, the

precision of multi-label classification would reach $90\pm 3\%$. Moreover, some LULC types occupied a small portion of the whole image, making them tough to be recognized by PNASNet based on the whole image.

The multi-label classification with spatial weights, or the proposed image-semantic model, outperformed multi-label classification, thus confirming the value of an operation that weighted different features based on image content. Moreover, the disparities in results generated by these two classification strategies varied significantly based on different LULC categories. Generally speaking, the proposed image-semantic model performed more effectively on images that contained more complicated LULC configurations. Finally, the process of semantic analysis in the proposed image-semantic model might have led to unwanted complexity when the LULC in a remote-sensing image was unmixed. For example, the proposed image-semantic model produced many false-negative results due to its overestimating the role of other non-freeway land-cover/land-uses. In these images, although *freeway* was the critical element of the land-use characteristics, its coverage seemed relatively more restricted compared to other land-cover/land-uses.

Table 11

Comparison of coastline retrieval by various PNASNet-enhanced classification strategies.

	Precision				Recall				F-scores			
	1*	2*	3*	4*	1*	2*	3*	4*	1*	2*	3*	4*
Image 1	0.17 70	0.43 65	0.75 21	0.80 73	0.95 56	0.95 56	0.97 78	0.97 78	0.14 93	0.29 96	0.42 51	0.44 22
Image 2	0.04 11	0.20 97	0.46 88	0.62 5	1	0.65	0.75	0.75	0.03 95	0.15 85	0.28 85	0.34 09
Image 3	0.03 09	0.20 97	0.73 68	0.73 68	1	0.86 67	0.93 33	0.93 33	0.03	0.16 88	0.41 18	0.41 18
Image 4	0.08 23	0.17 26	0.64 15	0.78 26	1	0.85	0.87 5	0.9	0.07 6	0.14 35	0.37 01	0.41 86
Image	0.02	0.09	0.61	0.73	0.84	0.54	1	1	0.02	0.08	0.37	0.42

5	26	68	11	33	62	55			2	22	93	3
Image 6	0.0679	0.1472	0.5636	0.7857	0.9429	0.5714	0.8857	0.9429	0.063	0.117	0.3444	0.4285
Image 7	0.0226	0.6667	0.5238	0.6875	1	0.3636	1	1	0.0221	0.2352	0.3437	0.4074
Image 8	0.0103	0.0617	0.75	0.7056	0.8333	0.8333	1	0.8333	0.0101	0.0574	0.4286	0.382
Image 9	0.1152	0.6053	0.8644	0.9	1	0.8214	0.9808	0.9643	0.1033	0.3484	0.4595	0.4655
Image 10	0.1111	0.2843	0.8281	0.8438	0.931	0.9655	0.9138	0.931	0.0992	0.2196	0.4344	0.4426
Image 11	0.0598	0.1472	0.5455	0.7429	0.9667	0.9667	0.8	0.8667	0.0563	0.127703243	0.3243	0.4

1* Binary classification

2* Multi-class classification

3* Multi-label classification

4* Multi-output classification (Proposed method)

The results generated by the proposed image-semantic model present several new phenomena and challenges that remain unsolved. The results exhibit the major disadvantages of deep learning models recently claimed in scene classification. CNN models rely heavily on the training features and data, which intensely restrict these models' transferability. In the experiment, when fusing the *harbor* and *wharf* into one LULC class, the research saw a dramatic decrease in the precision of retrieving images relevant to the second and third images. Moreover, the results generated by multi-label classification suggest a pressing need to integrate semantics into the CNN-powered land-use recognition process.

4.5 Conclusions

The availability of deep learning models and varied high-resolution remote-sensing images significantly facilitates the mapping LULC scenarios within large-scale areas. However, a LULC classification map cannot help individuals understand an area's

functionality and organization indicated by multiple LULC types. The conversion from a LULC classification map to a land-use classification map remains a major geospatial concern yet unsolved. A majority of previous work applied CNN models to a study case, for which the amount and term of LULC classes were fixed and generally defined. Here, when the study centered on coastal areas, which may possess greater LULC definition and detail, state-of-the-art CNNs were faced with a decline in classification accuracy. Specifically, it was difficult to prepare a useful massive training dataset when the distinction between two land-use classes was very limited.

The cost of large-scale data preparation urges researchers to reinforce CNNs' power in land-use image classification with a limited amount of data. This chapter presented an image-semantic model that integrates domain knowledge into the process of a CNN model for converting LULC attributes into meaningful land-use information. The research evaluated the proposed model after choosing coastal scenario in California to study. The results support the hypothesis that the CNN model might be insufficient at classifying land-use in remotely-sensed imagery without the support of domain knowledge. The proposed image-semantic model outperformed other CNN models that exclusively focus on the features derived from remote sensing images.

The research performed an investigation on a semantic-aware deep learning approach for land-use classification, but there nevertheless remains room for improvement. The modified VSM created by this chapter might accurately weight the priority of the LULC that belongs to a critical element of the land-use scene. In the future, it may be valuable to explore how the semantics may be integrated into the features generated by the feature extraction layer in a CNN model. Furthermore, the techniques

associated with transfer learning and reinforcement learning are fields worthy of considerable attention.

Planar surface assessments account for a full array of land-covers (e.g., building, tree, and impervious and soil cover). Significantly, though, common vertical indicators often do not discriminate among different land-covers. Buildings and tree canopies affect temperature through different mechanisms, for example, but are not necessarily made distinct in vertical dimension assessments (Unger, 2009). Interestingly, Google Street View possesses an immense collection of street panoramas, providing information on surface properties that include the differences in the vertical dimension objects, the heterogeneity of which is large in an urban context (Carrasco-Hernandez, 2015, Middel et al., 2017; Li et al., 2018).

CHAPTER 5

CONCLUSIONS

5.1 Conclusions

The dissertation presents a variety of strategies that exploits various deep learning models to support geospatial applications, which are summarized as follows:

Useful input data are significant to extend the power of a deep learning model in dealing with data classification problem. In case of building height estimation with shadows, this research develops a unified system that organizes a great number of shadow shapes into limited number of categories, and determines the shadow edge useful for building height estimation. The unified shadow pattern identifies the useful input data to promote deep learning to support efficient shadow-based building height estimation.

The significance of open geospatial data and volunteered geographical information has been reported in previous decades. However, open geospatial data may contain neither incomplete metadata nor quality control, limiting their use. This research focused on big data stored in two data-rich platforms—Google Earth Pro and Openstreet Map. Chapter 2 proposed a methodological framework to effectively deal with building height estimation with metadata and remote sensing images in Google Earth Pro. Moreover, Chapter 3 collects place names and their attributes to develop a ontological model for semantic query, which supports discovery of hidden knowledge in the map text and map features derived from raw digital maps with the deep learning model.

Optical character recognition is a key research focus of computer vision that aims to identify text information from various media, such as photos, videos, and digital documents. Map text is often rotated or curved relative to the map feature it represents,

which might limit state-of-the-art deep learning models for optical character recognition. The research proposed a methodological framework for detecting map text from digital maps, separating text units from the rotated or curved map text, and identifying every map unit with an advanced optical character recognition platform.

Multi-label image classification aims to assign spatial weights to an image with multiple predefined labels. Compared to multi-class image classification, multi-label classification can derive more details from a remote sensing image. However, multi-label classification is limited when the research needs to describe the spatial coverage of each LULC class. Although semantic segmentation can deal with the variations of position and portion, a well-developed benchmark dataset that supports LULC semantic segmentation has not yet been made available. This research divided the image into multiple sub-regions and then conducted multi-label classification on each sub-region.

5.2 Future Works

Although deep learning facilitates various dimensions of geography, several challenges remain unsolved. First, although substantial amount of data is generated, a majority of these data might be valueless for specific data analysis task. The value of geospatial data is much more important than the amount. Big geospatial data changes the way that researchers perform geospatial computing and analysis. However, big geospatial data cannot directly be connected to a “big task.” A number of papers propose concerns regarding the value of big data (L’heureux et al., 2017; Lv et al., 2017; Zhuang et al., 2017; Zhou, 2018). In some geospatial applications, the value of similar geospatial data might be varied according to the goal of specific tasks. For example, although an RGB remote sensing image is significant to support build-up area changes, it is useless for a

CNN model to distinguish natural and man-made lawn. Thus, a strategy for selecting appropriate input data is important to implement varied GeoAI techniques.

Moreover, although a variety of CNNs obtain promising results in computer vision, speech recognition, natural language processing, etc., these methods are challenged upon transferring their learning process into a new field. Thus, how to help the machine adaptively understand unknown or unlabeled input is a compelling task. The research in Chapter 5 proves that training data and samples cannot cover all phenomena, objects and events on the Earth's surface. Although deep learning has been found to outperform humans in object detection from remote sensing images (Chen and Gong, 2016), these state-of-the-art approaches can only identify a limited number of object classes. Training data cannot be prepared for all situations possible occurred in reality. For example, the accident involving a self-driving car that occurred in the last year revealed that crowdsourcing data collected and stored still could not lead to a safe driving task under complex and unlimited traffic conditions. Other techniques such as semantic analysis and heuristic reasoning have been reportedly useful to facilitate the power of deep learning models in data processing and analysis (Pan and Yan, 2010; Lu et al., 2015). Geospatial domain knowledge are essential sources to promote the transferability of GeoAI techniques in dealing with new geospatial data analysis tasks.

REFERENCES

Abohela, I., Hamza, N., & Dudek, S. (2013). Effect of roof shape, wind direction, building height and urban configuration on the energy yield and positioning of roof mounted wind turbines. *Renewable Energy*, 50, 1106-1118.

- Ablameyko S, Beveisbik V, Homenko M, et al. Interpretation of colour maps. A combination of automatic and interactive techniques. *Computing & Control Engineering Journal*, 2001, 12(4): 188-196.
- Ablameyko S, Bereishik V, Homenko M, et al. A complete system for interpretation of color maps. *International Journal of Image and Graphics*, 2002, 2(3): 453-479.
- Acker, J. G., & Leptoukh, G. (2007). Online analysis enhances use of NASA earth science data. *Eos, Transactions American Geophysical Union*, 88(2), 14-17.
- Adam, S., Ogier, J. M., Cariou, C., Mullot, R., Labiche, J., & Gardes, J. (2000). Symbol and character recognition: application to engineering drawings. *International Journal on Document Analysis and Recognition*, 3(2), 89-101.
- Ahmed, E., Jones, M., & Marks, T. K. (2015). An improved deep learning architecture for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3908-3916).
- Alexakis, D., Sarris, A., Astaras, T., & Albanakis, K. (2011). Integrated GIS, remote sensing and geomorphologic approaches for the reconstruction of the landscape habitation of Thessaly during the Neolithic period. *Journal of Archaeological Science*, 38(1), 89-100.
- Arevalo V, González J, Ambrosio G, 2005. Detecting Shadow QuickBird satellite images. ISPRS Commission VII Mid-term Symposium 'Remote Sensing: From Pixels to Processes'. Enschede, the Netherlands, 8–11 May.
- Arévalo, V., González, J., & Ambrosio, G. (2008). Shadow detection in colour high-resolution satellite images. *International Journal of Remote Sensing*, 29(7), 1945-1963.
- Ballatore, A., Bertolotto, M., & Wilson, D. C. (2014). Linking geographic vocabularies through WordNet. *Annals of GIS*, 20(2), 73-84.
- Banks, D. C., Linton, S. A., & Stockmeyer, P. K. (2004). Counting cases in subitope algorithms. *IEEE Transactions on Visualization and Computer Graphics*, 10(4), 371-384.
- Basu, S., Ganguly, S., Mukhopadhyay, S., DiBiano, R., Karki, M., & Nemani, R. (2015, November). DeepSat: a learning framework for satellite imagery. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems* (p. 37). ACM.
- Battle, R., & Kolas, D. (2012). Enabling the geospatial semantic web with parliament and geosparql. *Semantic Web*, 3(4), 355-370.

- Bay, H., Tuytelaars, T., & Van Gool, L. (2006, May). Surf: Speeded up robust features. In *European conference on computer vision* (pp. 404-417). Springer, Berlin, Heidelberg.
- Becker, C., & Bizer, C. (2009). Exploring the geospatial semantic web with dbpedia mobile. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(4), 278-286.
- Bengio, Y., Courville, A., & Vincent, P. (2012). Representation Learning: A Review and New Perspectives. *Pattern Analysis and Machine Intelligence, IEEE Transactions*, Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8), 1798-1828.
- Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8), 1798-1828.
- Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(Feb), 281-305.
- Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS journal of photogrammetry and remote sensing*, 65(1), 2-16.
- Bogorny, V., Avancini, H., de Paula, B. C., Kuplich, C. R., & Alvares, L. O. (2011). Weka-STPM: a Software Architecture and Prototype for Semantic Trajectory Data Mining and Visualization. *Transactions in GIS*, 15(2), 227-248.
- Bourges, B. (1985). Improvement in solar declination computation. *Solar Energy*, 35(4), 367-369.
- Boutell, M. R., Luo, J., Shen, X., & Brown, C. M. (2004). Learning multi-label scene classification. *Pattern Recognition*, 37(9), 1757-1771.
- Budig, B., Dijk, T. C. V., & Wolff, A. (2016). Matching labels and markers in historical maps: an algorithm with interactive postprocessing. *ACM Transactions on Spatial Algorithms and Systems (TSAS)*, 2(4), 13.
- Cao, R., & Tan, C. L. (2001a). Separation of overlapping text from graphics. *International Conference on Document Analysis and Recognition, 2001. Proceedings* (pp.44-48). IEEE.
- Cao, R., & Tan, C. L. (2001b). Text/Graphics Separation in Maps. *Graphics Recognition Algorithms and Applications, International Workshop, Grec 2001, Kingston*,

- Ontario, Canada, September 7-8, 2001, Selected Papers(Vol.18, pp.167-177). DBLP.
- Caprioli, M., & Gamba, P. (2015). Detecting and grouping words in topographic maps by means of perceptual concepts. *Signal Processing Conference, 2000, European* (pp.1-4). IEEE.
- Chaib, S., Liu, H., Gu, Y., & Yao, H. (2017). Deep Feature Fusion for VHR Remote Sensing Scene Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(8), 4775–4784.
- Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information sciences*, 275, 314-347.
- Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4), 834-848.
- Chen, L. C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.
- Chen, L. H., & Wang, J. Y. (1997, August). A system for extracting and recognizing numeral strings on maps. In *Document Analysis and Recognition, 1997., Proceedings of the Fourth International Conference on* (Vol. 1, pp. 337-341). IEEE.
- Cheng, F., & Thiel, K. H. (1995). Delimiting the building heights in a city from the shadow in a panchromatic SPOT-image—Part 1. Test of forty-two buildings. *Remote Sensing*, 16(3), 409-415.
- Cheng, G., & Han, J. (2016). A survey on object detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 117, 11-28.
- Cheng, G., Han, J., & Lu, X. (2017). Remote sensing image scene classification: Benchmark and state of the art. *Proceedings of the IEEE*, 105(10), 1865-1883.
- Chiang Y Y, Knoblock C A. Recognition of multi-oriented, multi-sized, and curved text. International Conference on Document Analysis and Recognition, IEEE, 2011: 1399-1403.
- Chiang Y Y, Knoblock C A. Recognizing text in raster maps. *GeoInformatica*, 2015, 19(1): 1-27.

- Chiang, Y.-Y., Leyk, S., Honarvar Nazari, N., Moghaddam, S., & Tan, T. X. (2016). Assessing the impact of graphical quality on automatic text recognition in digital maps. *Computers & Geosciences*, 93, 21–35.
- Chiang, Y. Y., Leyk, S., & Knoblock, C. A. (2014). A survey of digital map processing techniques. *ACM Computing Surveys (CSUR)*, 47(1), 1.
- Chiang, Y. Y., Moghaddam, S., Gupta, S., Fernandes, R., & Knoblock, C. A. (2014, November). From map images to geographic names. In *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 581-584). ACM.
- Chollet, F. (2017, July). Xception: Deep Learning with Depthwise Separable Convolutions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1800-1807). IEEE.
- Comber, A., Umezaki, M., Zhou, R., Ding, Y., Li, Y., Fu, H., ... Tewkesbury, A. (2012). Using shadows in high-resolution imagery to determine building height. *Remote Sensing Letters*, 3(7), 551–556.
- Cooper, P. I. (1973). The maximum efficiency of single-effect solar stills. *Solar Energy*, 15(3), 205-217.
- Crampton, J. W. (2001). Maps as social constructions: power, communication and visualization. *Progress in human Geography*, 25(2), 235-252.
- Dai, D., & Yang, W. (2011). Satellite image classification via two-layer sparse coding with biased image representation. *IEEE Geoscience and Remote Sensing Letters*, 8(1), 173-176.
- Dai, J., He, K., & Sun, J. (2016). Instance-aware semantic segmentation via multi-task network cascades. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3150-3158).
- Dare, P. M. (2005). Shadow analysis in high-resolution satellite imagery of urban areas. *Photogrammetric Engineering & Remote Sensing*, 71(2), 169-177.
- Deng, X., Zhu, Y., & Newsam, S. (2018, November). What is it like down there?: generating dense ground-level views and image features from overhead imagery using conditional generative adversarial networks. In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 43-52). ACM.
- Derekenaris, G., Garofalakis, J., Makris, C., Prentzas, J., Sioutas, S., & Tsakalidis, A. (2001). Integrating GIS, GPS and GSM technologies for the effective

- management of ambulances. *Computers, Environment and Urban Systems*, 25(3), 267-278.
- Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., & Wimmer, W. (2012). The operational sea surface temperature and sea ice analysis (OSTIA) system. *Remote Sensing of Environment*, 116, 140-158.
- Dori, D., & Velkovitch, Y. (1999). Segmentation and recognition of dimensioning text from engineering drawings. *Computer Standards & Interfaces*, 20(s 6–7), 416.
- Emani, C. K., Cullot, N., & Nicolle, C. (2015). Understandable big data: a survey. *Computer science review*, 17, 70-81.
- Estoque, R. C., Murayama, Y., & Myint, S. W. (2017). Effects of landscape composition and pattern on land surface temperature: An urban heat island study in the megacities of Southeast Asia. *Science of the Total Environment*, 577, 349-359.
- Faghmous, J. H., & Kumar, V. (2014). A big data guide to understanding climate change: The case for theory-guided data science. *Big data*, 2(3), 155-163.
- Franklin, Carl and Paula Hane, "An introduction to GIS: linking maps to databases," Database. 15 (2) April, 1992, 17-22.
- Frischknecht, S., Kanani, E., & Carosio, A. (1998). A raster-based approach for the automatic interpretation of topographic maps. *International Archives of Photogrammetry and Remote Sensing*, 32, 523-530.
- Fonseca, F. T., Egenhofer, M. J., Agouris, P., & Câmara, G. (2002). Using ontologies for integrated geographic information systems. *Transactions in GIS*, 6(3), 231-257.
- Foody, G. M. (2007). Map comparison in gis. *Progress in Physical Geography*, 31(4), 439-445.
- Fu, G., Liu, C., Zhou, R., Sun, T., & Zhang, Q. (2017). Classification for high resolution remote sensing imagery using a fully convolutional network. *Remote Sensing*, 9(5), 1–21.
- Gelbukh, A., Levachkine, S., & Han, S. Y. (2003, July). Resolving ambiguities in toponym recognition in cartographic maps. In *International Workshop on Graphics Recognition* (pp. 75-86). Springer, Berlin, Heidelberg.
- Gens, R. (2010). Remote sensing of coastlines: detection, extraction and monitoring. *International Journal of Remote Sensing*, 31(7), 1819-1836.

- Gholizadeh, M., Melesse, A., & Reddi, L. (2016). A comprehensive review on water quality parameters estimation using remote sensing techniques. *Sensors*, 16(8), 1298.
- Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440-1448).
- Gómez, C., White, J. C., & Wulder, M. A. (2016). Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 116, 55-72.
- Goodchild, M. F. (2013). The quality of big (geo) data. *Dialogues in Human Geography*, 3(3), 280-284.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18-27.
- Graham, M., & Shelton, T. (2013). Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography*, 3(3), 255-261.
- Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., ... & Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, 77, 354-377.
- Gui, Z., Yang, C., Xia, J., Liu, K., Xu, C., Li, J., & Lostritto, P. (2013). A performance, semantic and service quality-enhanced distributed search engine for improving geospatial resource discovery. *International Journal of Geographical Information Science*, 27(6), 1109-1132.
- Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, 187, 27-48.
- Gupta, P., Christopher, S. A., Wang, J., Gehrig, R., Lee, Y. C., & Kumar, N. (2006). Satellite remote sensing of particulate matter and air quality assessment over global cities. *Atmospheric Environment*, 40(30), 5880-5892.
- Hall, D. K., Comiso, J. C., DiGirolamo, N. E., Shuman, C. A., Box, J. E., & Koenig, L. S. (2013). Variability in the surface temperature and melt extent of the Greenland ice sheet from MODIS. *Geophysical Research Letters*, 40(10), 2114-2120.
- Han, X., Zhong, Y., & Zhang, L. (2017). Spatial-spectral unsupervised convolutional sparse auto-encoder classifier for hyperspectral imagery. *Photogrammetric Engineering & Remote Sensing*, 83(3), 195-206.

- Hang, J., Li, Y., Sandberg, M., Buccolieri, R., & Di Sabatino, S. (2012). The influence of building height variability on pollutant dispersion and pedestrian ventilation in idealized high-rise urban areas. *Building and Environment*, 56, 346-360.
- Hansen, M. C., & Loveland, T. R. (2012). A review of large area monitoring of land cover change using Landsat data. *Remote sensing of Environment*, 122, 66-74.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup, D., & Meger, D. (2018, April). Deep reinforcement learning that matters. In *Thirty-Second AAAI Conference on Artificial Intelligence*.
- Hernández-García, A., & König, P. (2018). Data augmentation instead of explicit regularization. *arXiv preprint arXiv:1806.03852*.
- Hodul, M., Knudby, A., & Ho, H. C. (2016). Estimation of continuous urban sky view factor from landsat data using shadow detection. *Remote Sensing*, 8(7), 568.
- Hoffer, E., Hubara, I., & Soudry, D. (2017). Train longer, generalize better: closing the generalization gap in large batch training of neural networks. In *Advances in Neural Information Processing Systems* (pp. 1731-1741).
- Houston, J. B., Hawthorne, J., Perreault, M. F., Park, E. H., Goldstein Hode, M., Halliwell, M. R., ... & Griffith, S. A. (2015). Social media and disasters: a functional framework for social media use in disaster planning, response, and research. *Disasters*, 39(1), 1-22.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.
- Hu, X., Liu, S., Chen, R., Wang, W., & Wang, C. (2018). A Deep Reinforcement Learning-Based Framework for Dynamic Resource Allocation in Multibeam Satellite Systems. *IEEE Communications Letters*, 22(8), 1612-1615.
- Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W., & Prasad, S. (2015). Extracting and understanding urban areas of interest using geotagged photos. *Computers, Environment and Urban Systems*, 54, 240-254.
- Huang, S. C., Cheng, F. C., & Chiu, Y. S. (2013). Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE Transactions on Image Processing*, 22(3), 1032-1041.

- Huang, S. C., Cheng, F. C., & Chiu, Y. S. (2013). Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE Transactions on Image Processing*, 22(3), 1032-1041.
- Hunt Jr, E. R., Doraiswamy, P. C., McMurtrey, J. E., Daughtry, C. S., Perry, E. M., & Akhmedov, B. (2013). A visible band index for remote sensing leaf chlorophyll content at the canopy scale. *International Journal of Applied Earth Observation and Geoinformation*, 21, 103-112.
- Hurst, P., & Clough, P. (2013). Will we be lost without paper maps in the digital age?. *Journal of Information Science*, 39(1), 48-60.
- Ishtiaque, A., Myint, S. W., & Wang, C. (2016). Examining the ecosystem health and sustainability of the world's largest mangrove forest using multi-temporal MODIS products. *Science of the Total Environment*, 569, 1241-1254.
- Irvin, R. B., & McKeown, D. M. (1989). Methods for exploiting the relationship between buildings and their shadows in aerial imagery. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(6), 1564-1575.
- Izadi, M., & Saeedi, P. (2012). Three-Dimensional Polygonal Building Model Estimation From Single Satellite Images. *Geoscience and Remote Sensing, IEEE Transactions*, 50(6), 2254–2272.
- Janowicz, K., Scheider, S., Pehle, T., & Hart, G. (2012). Geospatial semantics and linked spatiotemporal data—Past, present, and future. *Semantic Web*, 3(4), 321-332.
- Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301), 790–794.
- Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301), 790-794.
- Jiang, S., Alves, A., Rodrigues, F., Ferreira Jr, J., & Pereira, F. C. (2015). Mining point-of-interest data from social networks for urban land use classification and disaggregation. *Computers, Environment and Urban Systems*, 53, 36-46.
- Jiang, S., Kong, Y., & Fu, Y. (2017). Deep Geo-constrained Auto-encoder for Non-landmark GPS Estimation. *IEEE Transactions on Big Data*.

- Kasun, L. L. C., Zhou, H., Huang, G. B., & Vong, C. M. (2013). Representational learning with extreme learning machine for big data. *IEEE intelligent systems*, 28(6), 31-34.
- Kesorn, K., & Poslad, S. (2012). An enhanced bag-of-visual word vector space model to represent visual content in athletics images. *IEEE Transactions on Multimedia*, 14(1), 211-222.
- Khatab, Z. E., Hajihoseini, A., & Ghorashi, S. A. (2018). A fingerprint method for indoor localization using autoencoder based deep extreme learning machine. *IEEE sensors letters*, 2(1), 1-4.
- Khotanzad, A., & Zink, E. (2003). Contour line and geographic feature extraction from USGS color topographical paper maps. *IEEE transactions on pattern analysis and machine intelligence*, 25(1), 18-31.
- Kim, T., Javzandulam, T., & Lee, T. Y. (2007). Semiautomatic reconstruction of building height and footprints from single satellite images. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 4737-4740.
- Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in human geography*, 3(3), 262-267.
- Knoblock C A, Chiang Y Y. An approach for recognizing text labels in raster maps. *International Conference on Pattern Recognition, IEEE, 2010: 3199-3202.*
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).
- Krishna, R., Zhu, Y., Groth, O., Johnson, J., Hata, K., Kravitz, J., ... & Bernstein, M. S. (2017). Visual genome: Connecting language and vision using crowdsourced dense image annotations. *International Journal of Computer Vision*, 123(1), 32-73.
- Kobayashi, S., Fujioka, T., Tanaka, Y., Inoue, M., Niho, Y., & Miyoshi, A. (2010). A geographical information system using the Google Map API for guidance to referral hospitals. *Journal of medical systems*, 34(6), 1157-1160.
- Kommareddy, A. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160), 850-853.
- Konecny, M. (2011). Cartography: challenges and potential in the virtual geographic environments era. *Annals of GIS*, 17(3), 135-146.

- Lutz, M., Sprado, J., Klien, E., Schubert, C., & Christ, I. (2009). Overcoming semantic heterogeneity in spatial data infrastructures. *Computers & Geosciences*, *35*(4), 739-752.
- LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT sloan management review*, *52*(2), 21.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436.
- Lee, J. G., & Kang, M. (2015). Geospatial big data: challenges and opportunities. *Big Data Research*, *2*(2), 74-81.
- Lee, T., & Kim, T. (2013). Automatic building height extraction by volumetric shadow analysis of monoscopic imagery. *International Journal of Remote Sensing*, *34*(16), 5834–5850.
- Leyk, S., & Boesch, R. (2010). Colors of the past: color image segmentation in historical topographic maps based on homogeneity. *GeoInformatica*, *14*(1), 1.
- Leyk, S. (2009, July). Segmentation of colour layers in historical maps based on hierarchical colour sampling. In *International Workshop on Graphics Recognition* (pp. 231-241). Springer, Berlin, Heidelberg.
- L'heureux, A., Grolinger, K., Elyamany, H. F., & Capretz, M. A. (2017). Machine learning with big data: Challenges and approaches. *IEEE Access*, *5*, 7776-7797.
- Li, S., Dragicevic, S., Castro, F. A., Sester, M., Winter, S., Coltekin, A., ... & Cheng, T. (2016). Geospatial big data handling theory and methods: A review and research challenges. *ISPRS journal of Photogrammetry and Remote Sensing*, *115*, 119-133.
- Li, T., Shen, H., Yuan, Q., Zhang, X., & Zhang, L. (2017). Estimating ground-level PM_{2.5} by fusing satellite and station observations: A geo-intelligent deep learning approach. *Geophysical Research Letters*, *44*(23).
- Li, H., Liu, J., & Zhou, X. (2018). Intelligent Map Reader: A Framework for Topographic Map Understanding With Deep Learning and Gazetteer. *IEEE Access*, *6*, 25363-25376.
- Li, H., Tao, C., Wu, Z., Chen, J., Gong, J., & Deng, M. (2017). Rsi-cb: A large scale remote sensing image classification benchmark via crowdsourced data. *arXiv preprint arXiv:1705.10450*.

- Li, L., Nagy, G., Samal, A., Seth, S., & Xu, Y. (2000). Integrated text and line-art extraction from a topographic map. *International Journal on Document Analysis & Recognition*, 2(4), 177-185.
- Li, M., Zang, S., Zhang, B., Li, S., & Wu, C. (2014). A review of remote sensing image classification techniques: The role of spatio-contextual information. *European Journal of Remote Sensing*, 47(1), 389-411.
- Li, W., Fu, H., Yu, L., & Cracknell, A. (2016). Deep Learning Based Oil Palm Tree Detection and Counting for High-Resolution Remote Sensing Images. *Remote Sensing*, 9(1), 22.
- Li, W., Yang, C., & Yang, C. (2010). An active crawler for discovering geospatial web services and their distribution pattern—A case study of OGC Web Map Service. *International Journal of Geographical Information Science*, 24(8), 1127-1147.
- Li, W., Zhou, X., & Wu, S. (2016). An integrated software framework to support semantic modeling and reasoning of spatiotemporal change of geographical objects: A use case of land use and land cover change study. *ISPRS International Journal of Geo-Information*, 5(10), 179.
- Li, Z. (2007). Digital map generalization at the age of enlightenment: a review of the first forty years. *The Cartographic Journal*, 44(1), 80-93.
- Li, Z. and Huang, P. (2002). Quantitative measures for spatial information of maps. *International Journal of Geographical Information Science*, 16(7), 699-709.
- Li, Z., Yang, C. P., Wu, H., Li, W., & Miao, L. (2011). An optimized framework for seamlessly integrating ogc web services to support geospatial sciences. *International Journal of Geographical Information Science*, 25(4), 595-613.
- Liao, M., Shi, B., Bai, X., Wang, X., & Liu, W. (2017, February). TextBoxes: A Fast Text Detector with a Single Deep Neural Network. In *AAAI* (pp. 4161-4167).
- Liasis, G., & Stavrou, S. (2016). Satellite images analysis for shadow detection and building height estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 119, 437–450.
- Lin, T. Y., Cui, Y., Belongie, S., & Hays, J. (2015). Learning deep representations for ground-to-aerial geolocalization. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5007-5015).

- Liu, J., Fang, T., & Li, D. (2011). Shadow detection in remotely sensed images based on self-adaptive feature selection. *IEEE Transactions on Geoscience and Remote Sensing*, 49(12), 5092-5103.
- Liu, Y., & Jin, L. (2017, March). Deep matching prior network: Toward tighter multi-oriented text detection. In *Proc. CVPR*(pp. 3454-3461).
- Liu, F., Li, S., Zhang, L., Zhou, C., Ye, R., Wang, Y., & Lu, J. (2017). 3DCNN-DQN-RNN: a deep reinforcement learning framework for semantic parsing of large-scale 3D point clouds. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 5678-5687).
- Liu, T., Miao, Q., Tian, K., Song, J., Yang, Y., & Qi, Y. (2016). SCTMS: Superpixel based color topographic map segmentation method. *Journal of Visual Communication and Image Representation*, 35, 78-90.
- Liu, H., Simonyan, K., & Yang, Y. (2018). Darts: Differentiable architecture search. *arXiv preprint arXiv:1806.09055*.
- Liu, T., Xu, P., & Zhang, S. (2018). A Review of Recent Advances in Scanned Topographic Map Processing. *Neurocomputing*.
- Liu, Y., Zhang, D., Lu, G., & Ma, W. Y. (2007). A survey of content-based image retrieval with high-level semantics. *Pattern recognition*, 40(1), 262-282.
- Liu, C., Zoph, B., Neumann, M., Shlens, J., Hua, W., Li, L. J., ... & Murphy, K. (2018). Progressive neural architecture search. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 19-34).
- Liu, C., Zoph, B., Shlens, J., Hua, W., Li, L. J., Fei-Fei, L., ... & Murphy, K. (2017). Progressive neural architecture search. *arXiv preprint arXiv:1712.00559*.
- Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431-3440).
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91-110.
- Lu, Z. (1998). Detection of text regions from digital engineering drawings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(4), 431-439.
- Luo, W., Tan, H., Chen, L., & Ni, L. M. (2013, June). Finding time period-based most frequent path in big trajectory data. In *Proceedings of the 2013 ACM SIGMOD international conference on management of data* (pp. 713-724). ACM.

- Lv, Z., Song, H., Basanta-Val, P., Steed, A., & Jo, M. (2017). Next-generation big data analytics: State of the art, challenges, and future research topics. *IEEE Transactions on Industrial Informatics*, 13(4), 1891-1899.
- Majid, A., Chen, L., Chen, G., Mirza, H. T., Hussain, I., & Woodward, J. (2013). A context-aware personalized travel recommendation system based on geotagged social media data mining. *International Journal of Geographical Information Science*, 27(4), 662-684.
- Maple, C. (2003, July). Geometric design and space planning using the marching squares and marching cube algorithms. In *Geometric Modeling and Graphics, 2003. Proceedings. 2003 International Conference on* (pp. 90-95). IEEE.
- Marmanis, D., Datcu, M., Esch, T., Stilla, U., Tang, J., Deng, C., ... Maharaj, B. T. J. (2016). Deep Learning Earth Observation Classification Using ImageNet Pretrained Networks. *IEEE Geoscience and Remote Sensing Letters*, 13(3), 1174-1185.
- Martínez, M. L., Intralawan, A., Vázquez, G., Pérez-Maqueo, O., Sutton, P., & Landgrave, R. (2007). The coasts of our world: Ecological, economic and social importance. *Ecological Economics*, 63(2-3), 254-272.
- Massalabi, A., & He, D. (2004). Detecting information under and from shadow in panchromatic Ikonos images of the city of Sherbrooke. *Geoscience and Remote Sensing Symposium, 00(c)*, 2000-2003.
- Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International journal of remote sensing*, 39(9), 2784-2817.
- Miao, Q., Liu, T., Song, J., Gong, M., & Yang, Y. (2016). Guided superpixel method for topographic map processing. *IEEE Transactions on Geoscience and Remote Sensing*, 54(11), 6265-6279.
- Miller, H. J., & Goodchild, M. F. (2015). Data-driven geography. *GeoJournal*, 80(4), 449-461.
- Miller-Rushing, A. J., Gallinat, A. S., & Primack, R. B. (2019). Creative citizen science illuminates complex ecological responses to climate change. *Proceedings of the National Academy of Sciences*, 116(3), 720-722.
- Mikolov, T., Yih, W. T., & Zweig, G. (2013). Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 Conference of the North*

American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 746-751).

- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Petersen, S. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529.
- Monmonier, M. (2006). Cartography: uncertainty, interventions, and dynamic display. *Progress in Human Geography*, 30(3), 373-381.
- Morais, C. D. (2012). Where is the Phrase “80% of Data is Geographic” From? . Retrieved May 2, 2015, from <http://www.gislounge.com/80-percent-data-is-geographic/>
- Mordret, A., Mikesell, T. D., Harig, C., Lipovsky, B. P., & Prieto, G. A. (2016). Monitoring southwest Greenland’s ice sheet melt with ambient seismic noise. *Science Advances*, 2(5), e1501538.
- Murray, A. B., Gopalakrishnan, S., McNamara, D. E., & Smith, M. D. (2013). Progress in coupling models of human and coastal landscape change. *Computers & geosciences*, 53, 30-38.
- Myers, G. K., Mulgaonkar, P. G., Chen, C. H., Decurtins, J. L., & Chen, E. (1995). Verification-based approach for automated text and feature extraction from raster-scanned maps. *Selected Papers from the First International Workshop on Graphics Recognition, Methods and Applications* (Vol.1072, pp.190-203). Springer-Verlag.
- Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2(1), 1.
- Neis, P., & Zielstra, D. (2014). Recent developments and future trends in volunteered geographic information research: The case of OpenStreetMap. *Future Internet*, 6(1), 76-106.
- Nogueira, K., Penatti, O. A., & dos Santos, J. A. (2017). Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recognition*, 61, 539-556.
- Patel, C., Patel, A., & Patel, D. (2012). Optical character recognition by open source OCR tool tesseract: A case study. *International Journal of Computer Applications*, 55(10).

- Peng, X. B., Berseth, G., Yin, K., & Van De Panne, M. (2017). Deeploco: Dynamic locomotion skills using hierarchical deep reinforcement learning. *ACM Transactions on Graphics (TOG)*, 36(4), 41.
- Perini, K., & Magliocco, A. (2014). Effects of vegetation, urban density, building height, and atmospheric conditions on local temperatures and thermal comfort. *Urban Forestry & Urban Greening*, 13(3), 495-506.
- Perkins, C. (2003a). Cartography: mapping theory. *Progress in Human Geography*, 27(3), 341-351.
- Perkins, C. (2013b). Cultures of map use. *Cartographic Journal*, 45(2), 150-158.
- Pezeshk, A., & Tutwiler, R. L. (2008). Text segmentation and reorientation from scanned color topographic maps. In *Proc. ICSIP* (pp. 94-97).
- Pezeshk, A., & Tutwiler, R. L. (2011). Automatic feature extraction and text recognition from scanned topographic maps. *IEEE Transactions on Geoscience & Remote Sensing*, 49(12), 5047-5063.
- Pouderoux, J., Gonzato, J. C., Pereira, A., & Guitton, P. (2007). Toponym Recognition in Scanned Color Topographic Maps. *International Conference on Document Analysis and Recognition* (Vol.1, pp.531-535). IEEE.
- Power, C., Simms, A., & White, R. (2001). Hierarchical fuzzy pattern matching for the regional comparison of land use maps. *International Journal of Geographical Information Science*, 15(1), 77-100.
- Qi, F., & Wang, Y. (2014). A new calculation method for shape coefficient of residential building using Google Earth. *Energy and Buildings*, 76, 72–80.
- Qi, F., Zhai, J. Z., & Dang, G. (2016). Building height estimation using Google Earth. *Energy and Buildings*, 118, 123–132.
- Rahman, S., Rahman, M. M., Abdullah-Al-Wadud, M., Al-Quaderi, G. D., & Shoyuib, M. (2016). An adaptive gamma correction for image enhancement. *EURASIP Journal on Image and Video Processing*, 2016(1), 35.
- Raju, P. L. N., Chaudhary, H., & Jha, A. K. (2014). Shadow analysis technique for extraction of building height using high resolution satellite single image and accuracy assessment. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 40(8), 1185–1192.

- Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (6), 1137-1149.
- Roy, P. P., Vazquez, E., Lladós, J., Baldrich, R., & Pal, U. (2007). A System to Segment Text and Symbols from Color Maps. *Graphics Recognition. Recent Advances and New Opportunities, International Workshop, Grec 2007, Curitiba, Brazil, September 20-21, 2007. Selected Papers* (Vol.5046, pp.245-256). DBLP.
- Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Berg, A. C. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3), 211-252.
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4510-4520).
- Schimel, D., Pavlick, R., Fisher, J. B., Asner, G. P., Saatchi, S., Townsend, P., ... & Cox, P. (2015). Observing terrestrial ecosystems and the carbon cycle from space. *Global Change Biology*, 21(5), 1762-1776.
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, 61, 85-117.
- Scott, G. J., England, M. R., Starms, W. A., Marcum, R. A., & Davis, C. H. (2017). Training Deep Convolutional Neural Networks for Land-Cover Classification of High-Resolution Imagery. *IEEE Geoscience and Remote Sensing Letters*, 14(4), 549-553.
- Shao, W., Yang, W., & Xia, G. S. (2013). Extreme value theory-based calibration for the fusion of multiple features in high-resolution satellite scene classification. *International journal of remote sensing*, 34(23), 8588-8602.
- Shao, Y., Taff, G. N., & Walsh, S. J. (2011). Shadow detection and building-height estimation using IKONOS data. *International Journal of Remote Sensing*, 32(22), 6929-6944.
- Shen, W., Wang, X., Wang, Y., Bai, X., & Zhang, Z. (2015). Deepcontour: A deep convolutional feature learned by positive-sharing loss for contour detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3982-3991).

- Shen, Y., Zhou, X., Liu, J., & Chen, J. (2018, July). CSRS-SIAT: A benchmark remote sensing dataset to semantic-enabled and cross-scales scene recognition. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 1780-1783). IEEE.
- Shettigara, V. K., & Sumerling, G. M. (1998). Height determination of extended objects using shadows in SPOT images. *Photogrammetric Engineering and Remote Sensing*, *64*(1), 35-43.
- Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, *22*(12), 1349-1380.
- Smith, R. (2007, September). An overview of the Tesseract OCR engine. In *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on* (Vol. 2, pp. 629-633). IEEE.
- Spangenberg, T. (2014). Development of a mobile toolkit to support research on human mobility behavior using GPS trajectories. *Information Technology & Tourism*, *14*(4), 317-346.
- Srivastava, P. K., Singh, S. K., Gupta, M., Thakur, J. K., & Mukherjee, S. (2013). Modeling impact of land use change trajectories on groundwater quality using remote sensing and GIS. *Environmental Engineering & Management Journal (EEMJ)*, *12*(12).
- Steiger, E., Resch, B., & Zipf, A. (2016). Exploration of spatiotemporal and semantic clusters of Twitter data using unsupervised neural networks. *International Journal of Geographical Information Science*, *30*(9), 1694-1716.
- Su, Y., Li, J., Plaza, A., Marinoni, A., Gamba, P., & Chakravorty, S. (2019). DAEN: Deep Autoencoder Networks for Hyperspectral Unmixing. *IEEE Transactions on Geoscience and Remote Sensing*.
- Sui, D., & Goodchild, M. (2011). The convergence of GIS and social media: challenges for GIScience. *International Journal of Geographical Information Science*, *25*(11), 1737-1748.
- Syamkumar, M., Durairajan, R., & Barford, P. (2016, October). Bigfoot: A geo-based visualization methodology for detecting bgp threats. In *2016 IEEE Symposium on Visualization for Cyber Security (VizSec)* (pp. 1-8). IEEE.
- Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. In *AAAI* (Vol. 4, p. 12).

- Tambassi, T. (2018). From a geographical perspective: spatial turn, taxonomies and geontologies. In *The Philosophy of Geo-Ontologies* (pp. 27-36). Springer, Cham.
- Tan, C. L., & Ng, P. O. (1998). Text extraction using pyramid. *Pattern Recognition*, 31(1), 63-72.
- Taylor, L., & Nitschke, G. (2017). Improving deep learning using generic data augmentation. *arXiv preprint arXiv:1708.06020*.
- Tchuenté, A. T. K., Roujean, J. L., & De Jong, S. M. (2011). Comparison and relative quality assessment of the GLC2000, GLOBCOVER, MODIS and ECOCLIMAP land cover data sets at the African continental scale. *International Journal of Applied Earth Observation and Geoinformation*, 13(2), 207-219.
- Tian, Y., Chen, C., & Shah, M. (2017). Cross-view image matching for geo-localization in urban environments. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3608-3616).
- Tombre, K., Tabbone, S., Lamiroy, B., & Dosch, P. (2002). Text/Graphics Separation Revisited. *International Workshop on Document Analysis Systems V* (Vol.2423, pp.200-211). Springer-Verlag.
- Tsoumakas, G., & Katakis, I. (2007). Multi-label classification: An overview. *International Journal of Data Warehousing and Mining*, 3(3), 1-13.
- Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research*, 37, 141-188.
- Uhl, J. H., Leyk, S., Chiang, Y. Y., Duan, W., & Knoblock, C. A. (2018). Map Archive Mining: Visual-Analytical Approaches to Explore Large Historical Map Collections. *ISPRS International Journal of Geo-Information*, 7(4), 148.
- Usery, E. Lynn, and Dalia Varanka. "Design and development of linked data from the national map." *Semantic web* 3, no. 4 (2012): 371-384.
- Velázquez, A., & Levachkine, S. (2003). Text/Graphics Separation and Recognition in Raster-Scanned Color Cartographic Maps. *Graphics Recognition, Recent Advances and Perspectives, Internationalworkshop, Grec 2003, Barcelona, Spain, July 30-31, 2003, Revised Selected Papers* (Vol.3088, pp.63-74). DBLP.
- Velázquez, A., & Levachkine, S. (2004). *Text/Graphics Separation and Recognition in Raster-Scanned Color Cartographic Maps. Graphics Recognition. Recent Advances and Perspectives*. Springer Berlin Heidelberg.

- VoPham, T., Hart, J. E., Laden, F., & Chiang, Y. Y. (2018). Emerging trends in geospatial artificial intelligence (geoAI): potential applications for environmental epidemiology. *Environmental Health*, 17(1), 40.
- Wang, A. H., & Yan, H. (1994). Text extraction from color map images. *Journal of Electronic Imaging*, 3(4), 390-396.
- Wang, J., Luo, C., Huang, H., Zhao, H., & Wang, S. (2017). Transferring Pre-Trained Deep CNNs for Remote Scene Classification with General Features Learned from Linear PCA Network. *Remote Sensing*, 9(3), 225.
- Wang, J., & Wang, X. (2009). Information extraction of building height and density based on quick bird image in Kunming, China. *2009 Joint Urban Remote Sensing Event*.
- Wang, J., Yang, Y., Mao, J., Huang, Z., Huang, C., & Xu, W. (2016). Cnn-rnn: A unified framework for multi-label image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2285-2294).
- Wang, L., Zhao, X., & Shen, Y. (2012). Coupling hydrodynamic models with GIS for storm surge simulation: application to the Yangtze Estuary and the Hangzhou Bay, China. *Frontiers of Earth Science*, 6(3), 261-275.
- Wang, Q., Yan, L., Yuan, Q., & Ma, Z. (2017). An automatic shadow detection method for VHR remote sensing orthoimagery. *Remote Sensing*, 9(5), 13–20.
- Wang, X., Yu, X., & Ling, F. (2014). Building Heights Estimation using ZY3 Data- A case study of Shanghai , China. *Igrass*, 1749–1752.
- Weng, Q., Mao, Z., Lin, J., & Guo, W. (2017). Land-Use Classification via Extreme Learning Classifier Based on Deep Convolutional Features. *IEEE Geoscience and Remote Sensing Letters*, 14(5), 704–708.
- White, R. (2006). Pattern based map comparisons. *Journal of Geographical Systems*, 8(2), 145-164
- Wick, Marc, B. Vatant, and B. Christophe. "Geonames ontology." URL <http://www.geonames.org/ontology> (2015).
- Wright, D. J., & Wang, S. (2011). The emergence of spatial cyberinfrastructure. *Proceedings of the National Academy of Sciences*, 108(14), 5488-5491.

- Wu, H., Li, Z., Zhang, H., Yang, C., & Shen, S. (2011). Monitoring and evaluating the quality of Web Map Service resources for optimizing map composition over the internet to support decision making. *Computers & Geosciences*, 37(4), 485-494.
- Wu, X., Zurita-Milla, R., & Kraak, M. J. (2015). Co-clustering geo-referenced time series: exploring spatio-temporal patterns in Dutch temperature data. *International journal of geographical information science*, 29(4), 624-642.
- Xia, G. S., Hu, J., Hu, F., Shi, B., Bai, X., Zhong, Y., ... & Lu, X. (2017). AID: A benchmark data set for performance evaluation of aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(7), 3965-3981.
- Xiao, Z., & Long, Y. (2017). High-Resolution Remote Sensing Image Retrieval Based on CNNs from a Dimensional Perspective. *Remote Sensing*, 9(7), 725.
- Xiong, Y., & Zuo, R. (2018). GIS-based rare events logistic regression for mineral prospectivity mapping. *Computers & Geosciences*, 111, 18-25.
- Xu, P., Miao, Q., Liu, R., Chen, X., & Fan, X. (2016). Dynamic character grouping based on four consistency constraints in topographic maps. *Neurocomputing*, 212, 96-106.
- Yamada, H., Yamamoto, K., & Hosokawa, K. (1993). Directional mathematical morphology and reformalized hough transformation for the analysis of topographic maps. *IEEE Trans Pattern Analysis & Machine Intelligence*, 15(4), 380-387.
- Yang, C., Raskin, R., Goodchild, M., & Gahegan, M. (2010). Geospatial cyberinfrastructure: past, present and future. *Computers, Environment and Urban Systems*, 34(4), 264-277.
- Yang, Y., & Newsam, S. (2010, November). Bag-of-visual-words and spatial extensions for land-use classification. In *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems* (pp. 270-279). ACM.
- Yao, X., Han, J., Cheng, G., Qian, X., & Guo, L. (2016). Semantic Annotation of High-Resolution Satellite Images via Weakly Supervised Learning. *IEEE Transactions on Geoscience and Remote Sensing*, 54(6), 3660-3671.
- Yin, J., Soliman, A., Yin, D., & Wang, S. (2017). Depicting urban boundaries from a mobility network of spatial interactions: A case study of Great Britain with geo-located Twitter data. *International Journal of Geographical Information Science*, 31(7), 1293-1313.

- Yu, R., Luo, Z., & Chiang, Y. Y. (2016, December). Recognizing text in historical maps using maps from multiple time periods. In *Pattern Recognition (ICPR), 2016 23rd International Conference on* (pp. 3993-3998). IEEE.
- Yu, X., Wu, X., Luo, C., & Ren, P. (2017). Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework. *GIScience & Remote Sensing*, 54(5), 741-758.
- Yue, P., Gong, J., Di, L., He, L., & Wei, Y. (2011). Integrating semantic web technologies and geospatial catalog services for geospatial information discovery and processing in cyberinfrastructure. *GeoInformatica*, 15(2), 273-303.
- Zeggada, A., Melgani, F., & Bazi, Y. (2017). A Deep Learning Approach to UAV Image Multilabeling. *IEEE Geoscience and Remote Sensing Letters*, 14(5), 694–698.
- Zhang, C., Zhang, K., Yuan, Q., Zhang, L., Hanratty, T., & Han, J. (2016, August). Gmove: Group-level mobility modeling using geo-tagged social media. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1305-1314). ACM.
- Zhang, D., & Lu, G. (2004). Review of shape representation and description techniques. *Pattern Recognition*, 37(1), 1-19.
- Zhang, D., Zhang, W., Huang, W., Hong, Z., & Meng, L. (2017). Upscaling of Surface Soil Moisture Using a Deep Learning Model with VIIRS RDR. *ISPRS International Journal of Geo-Information*, 6(5), 130.
- Zhang, F., Du, B., & Zhang, L. (2016). Scene classification via a gradient boosting random convolutional network framework. *IEEE Transactions on Geoscience and Remote Sensing*, 54(3), 1793–1802.
- Zhang, H., Sun, K., & Li, W. (2014). Object-oriented shadow detection and removal from urban high-resolution remote sensing images. *IEEE Transactions on geoscience and remote sensing*, 52(11), 6972-6982.
- Zhang, L., Zhang, L., & Du, B. (2016). Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, 4(2), 22-40.
- Zhang, R., & Zhu, D. (2011). Study of land cover classification based on knowledge rules using high-resolution remote sensing images. *Expert Systems with Applications*, 38(4), 3647-3652.
- Zhao, B., Zhong, Y., Xia, G. S., & Zhang, L. (2016). Dirichlet-derived multiple topic scene classification model for high spatial resolution remote sensing

- imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 54(4), 2108-2123.
- Zhao, L., Tang, P., & Huo, L. (2016). Feature significance-based multibag-of-visual-words model for remote sensing image scene classification. *Journal of Applied Remote Sensing*, 10(3), 035004.
- Zhao, W., Du, S., & Emery, W. J. (2017). Object-Based Convolutional Neural Network for High-Resolution Imagery Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(7), 3386–3396.
- Zhong, D. X. (2002). Extraction of embedded and/or line-touching character-like objects. *Pattern Recognition*, 35(11), 2453-2466.
- Zhong, Y., Fei, F., & Zhang, L. (2016). Large patch convolutional neural networks for the scene classification of high spatial resolution imagery. *Journal of Applied Remote Sensing*, 10(2), 25006.
- Zhou, W., Newsam, S., Li, C., & Shao, Z. (2017). Learning low dimensional convolutional neural networks for high-resolution remote sensing image retrieval. *Remote Sensing*, 9(5).
- Zhou, W., Newsam, S., Li, C., & Shao, Z. (2018). Patternnet: a benchmark dataset for performance evaluation of remote sensing image retrieval. *ISPRS Journal of Photogrammetry and Remote Sensing*. 145(A), 197-209.
- Zhou, X., Li, W., Arundel, S., et al. (2018). Deep Convolutional Neural Networks for Map-Type Classification. In Proceedings of AutoCafto/UCGIS 2018. (pp. 147-155).
- Zhou, Y., Zhang, F., Du, Z., Ye, X., & Liu, R. (2017). Integrating cellular automata with the deep belief network for simulating urban growth. *Sustainability*, 9(10), 1786.
- Zhu, Q., Zhong, Y., Zhao, B., Xia, G. S., & Zhang, L. (2016). Bag-of-visual-words scene classifier with local and global features for high spatial resolution remote sensing imagery. *IEEE Geoscience and Remote Sensing Letters*, 13(6), 747-751.
- Zhu S.C. Dark, Beyond Deep. Keynote at Int'l Workshop on Vision Meets Cognition, at CVPR, Hawaii, June, 2017.
- Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: a comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8-36.

- Zhu, Y., Zhu, A. X., Feng, M., Song, J., Zhao, H., Yang, J., ... & Yao, L. (2017). A similarity-based automatic data recommendation approach for geographic models. *International Journal of Geographical Information Science*, 31(7), 1403-1424.
- Zhuang, Y. T., Wu, F., Chen, C., & Pan, Y. H. (2017). Challenges and opportunities: from big data to knowledge in AI 2.0. *Frontiers of Information Technology & Electronic Engineering*, 18(1), 3-14.
- Zoph, B., & Le, Q. V. (2016). Neural architecture search with reinforcement learning. *arXiv preprint arXiv:1611.01578*.
- Zoph, B., Vasudevan, V., Shlens, J., & Le, Q. V. (2017). Learning transferable architectures for scalable image recognition. *arXiv preprint arXiv:1707.07012*, 2(6).
- Zou, Q., Ni, L., Zhang, T., & Wang, Q. (2015). Deep Learning Based Feature Selection for Remote Sensing Scene Classification. *IEEE Geosci. Remote Sensing Lett.*, 12(11), 2321-2325.