Remote Sensing of Environment 214 (2018) 73–86

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery

Bo Huang[a,b,c], Bei Zhao[a,*], Yimeng Song[a]

[a] Department of Geography and Resource Management, The Chinese University of Hong Kong, Hong Kong
[b] Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, PR China
[c] Institute of Space and Earth Information and Science, The Chinese University of Hong Kong, Hong Kong

ARTICLE INFO

Keywords:
Deep convolutional neural networks
Transfer learning
Skeleton extraction
Street block
Land-use mapping
Classification
High spatial resolution
Multispectral remote sensing image

ABSTRACT

Urban land-use mapping is a significant yet challenging task in the field of remote sensing. Although numerous classification methods have been developed for obtaining land-use information in urban areas, the accuracy and efficiency of these methods are insufficient to meet the requirements of real-world applications such as urban planning and land management. In recent years, deep learning techniques, especially deep convolutional neural networks (DCNN), have achieved an astonishing level of performance in image classification. However, the traditional DCNN methods do not focus on multispectral remote sensing images with more than three channels, and they are limited by their training samples. In addition, these methods uniformly decompose large images into small processing units, which chop up the land-use patterns and produce land-use maps with obvious "blocks". In this study, a semi-transfer deep convolutional neural network (STDCNN) approach is proposed to overcome these weaknesses. The proposed STDCNN has three parts: one part involves a transferred DCNN with deep architecture; another part is designed to analyze multispectral images; and the final part fuses the first two parts into a classification layer. Moreover, a skeleton-based decomposing method using street block data is devised to maintain the integrity of the land-use patterns. In two case studies, the proposed method is used to generate urban land-use maps from a WorldView-3 image of a 143 km^2 area of Hong Kong and a WorldView-2 image of a 25 km^2 area of Shenzhen. The results show that the proposed STDCNN obtains an overall accuracy (OA) of 91.25% and a Kappa coefficient (Kappa) of 0.903 for Hong Kong land-use classification, and an OA of 80% and a Kappa of 0.780 for Shenzhen land-use classification. In addition, due to the proposed skeleton-based decomposition method, the proposed method can produce better land-use maps for real-world urban applications.

1. Introduction

Urban land-use mapping is a fundamental method for recognizing and locating land-uses for different purposes, such as industrial, residential, institutional and commercial areas. Urban land-use maps have great value for urban environment monitoring, planning and designing (Voltersen et al., 2014; Wu et al., 2017), In particular, they are useful for the study of phenomena, such as urban heat island effects (Chen et al., 2006), urban transport (Geurs and Van Wee, 2004) and house rents (Fujita, 1989). At present, methods for updating of urban land-use maps depend on the interpretation of aerial photos and field surveys, both of which are laborious and time consuming. With the development of remote sensing technologies, a large number of high spatial resolution (HSR) remote sensing images covering an urban area can be obtained by sensors installed on aircraft or satellites. Although it would be useful to extract land-use information from such HSR remote sensing imagery, a single land parcel used for one purpose (e.g. a residential, commercial, or industrial area) often contains multiple types of land-cover with distinct spatial/spectral/geometric characteristics, e.g. a single residential area may contain trees, buildings, and waterbodies, which makes the automatically mapping of land usage more challenge (Zhao et al., 2016b).

Traditionally, land-use classification methods have operated at the pixel level and assessed the geometrical, textural, and contextual features surrounding the focal pixels. However, such methods are not suitable for urban land-use classification based on HSR imagery. As urban land-use class labels are usually assigned at the land parcel and each land parcel contains a variety of land-covers. HSR images of land parcels are too complex to be categorized at the pixel level (Wu et al., 2009; Zhao et al., 2016b). Object-based classification methods

* Corresponding author.
E-mail address: zhaoys@whu.edu.cn (B. Zhao).

https://doi.org/10.1016/j.rse.2018.04.050
Received 29 October 2017; Received in revised form 20 April 2018; Accepted 28 April 2018

(Blaschke et al., 2014) can have similar problems, as they derive land-use descriptions through the application of co-occurrence (Aksoy et al., 2005), neighborhood-graph-based (Voltersen et al., 2014; Walde et al., 2014) or geometric measure (Huang et al., 2015) methods. These methods can incorporate land-cover information and are compatible with many existing land-cover classification methods. The performance of these methods, however, depends heavily on the selected land-cover classification system and on the accuracy of the land-cover classification. These methods also require knowledge of land-covers and information on land-uses.

Another approach is to use per-field classification methods to directly extract and classify the low-level features of the fields (e.g., spectral, textural, geometrical and contextual features) with predetermined boundaries. This approach has many advantages over the per-pixel or object-based methods of urban land-use classification (Hu and Wang, 2013; Wu et al., 2009). Due to the complexity of the land-use images, per-field classification methods should extract numerous features and exhaustively select an optimal subset of features to obtain the satisfied classification process. To reduce the difficulty and complexity of the feature extraction, a bag-of-visual-words (BOVW) model can be used, which views each land-use image as a bag of "visual terms" (Quelhas et al., 2007; van Gemert et al., 2010), where each term identifies a small aspect of the overall land-use, and captures a simple biophysical characteristic. Instead of the low-level features, the BOVW model represents the land-use images by mid-level features which are obtained by coding the low-level features with a learned "dictionary" of visual terms. The "dictionary" is often from the low-level features through an unsupervised learning algorithm such as the *k*-means (Chen and Tian, 2015; Yang and Newsam, 2011; Zhao et al., 2014), Gaussian mixture model (Perronnin et al., 2010), spectral clustering (Hu et al., 2015a), part-lets detector (Cheng et al., 2015) and sparse dictionary learning (Yang et al., 2009) algorithms. The commonly used feature coding methods (Bosch et al., 2008; Huang et al., 2013) include hard-voting (Zhao et al., 2016b), spare coding (Cheriyadat, 2014; Zheng et al., 2013), fisher coding (Zhao et al., 2016c), and the probabilistic topic models (Bahmanyar et al., 2015; Lienou et al., 2010; Luo et al., 2014; Zhao et al., 2016a, 2013; Zhong et al., 2015). Due to the adoption of the BOVW model and the effective organization of low-level features, the mid-level-feature-based methods can obtain better classification accuracy than low-level-feature-based methods.

All of the aforementioned per-pixel, object-based and per-field land-use classification methods are based on shallow architectures and hand-craft feature descriptors, which fail to capture the fine features of the complex land-use images used for generalization. Consequently, none of thee methods achieve the level of accuracy required by practical applications. In an urban land-use scheme, land-use can be described at many levels, including pixel intensities, edges, object parts, objects (building, trees, roads, etc.), and land parcels, all of which can be represented efficiently with deep architectures. Deep learning is a process through which a set of machine learning algorithms attempt to model high-level abstractions of data by using deep architectures composed of multiple nonlinear transformations (LeCun et al., 2015). As deep learning is able to model the hierarchical representations of features and as urban land-use schemes can be described by such features, the deep learning model is a very promising avenue to address urban land-use classification problems. Among the various deep learning techniques, the deep convolutional neural networks (DCNN) method has achieved an astonishing level of performance in the land-use classification of HSR images (Jia et al., 2015; Zhang and Du, 2016). DCNNs are composed of multiple convolutional layers, and are able to learn high-level abstract features from the original pixel values of land-use images. However, the increase in the number of layers increases the number of parameters in the DCNN, which creates the requirement for a large amount of training samples. To reduce the required number of training land-use samples, some researchers have proposed using transfer DCNNs (Castelluccio et al., 2015; Hu et al., 2015b; Marmanis et al., 2016; Penatti et al., 2015; Zhao et al., 2017) or small DCNNs with only a few layers (Zhang et al., 2016) to prevent the overfitting of trained networks. However, the traditional transfer DCNNs only incorporate the grey or RGB images and fail to adapt to HSR multispectral images (which have more than three channels), whereas the small DCNNs are unable to take advantage of deep architecture. Therefore, a new method is needed that can use transfer DCNNs and small DCNNs to determine land-use classification from HSR multispectral images. Moreover, existed developed land-use classification methods tend to evenly split the large HSR images into small processing units of fixed sizes through the uniform decomposition method (Zhang et al., 2014, 2016). This method chop up the patterns of land-use and generates a land-use map with obvious "blocks". As a result, the land-use maps obtained by traditional DCNNs do not meet the standards needed for practical applications.

To solve these problems, this study proposes a semi-transfer deep convolutional neural networks (STDCNN) method of land-use classification for urban land-use mapping based on HSR multispectral images. This STDCNN system consists of three parts. The first part of the system is a DCNN that is transferred from the pretrained AlexNet model. This model has been trained with a large set of natural images, including more than 1.2 million images and 1000 classes of topographical features (Krizhevsky et al., 2012), and is available for free on the Internet. This transferred DCNN allows the proposed STDCNN to acquire a deep architecture. The second part of the method is a small DCNN with only a few layers, which is designed for interpreting multispectral images. This small DCNN needs to be trained on the HSR multispectral images with randomly initialized parameters. The third part of the STDCNN method contains a fully connected layer and a softmax layer; the former layer fuses the first two parts, and the latter layer generates a final confidential vector of the land-use image. The proposed network can be trained through a semi-transfer process. Due to the small DCNN and the transfer DCNN, the proposed STDCNN can obtain good outcomes using the limited number of training samples derived from HSR multispectral imagery.

To obtain an urban land-use map, a street block is recommended as the minimum land-use mapping unit (Hu and Wang, 2013; Voltersen et al., 2014; Walde et al., 2014; Wu et al., 2009; Zhang and Du, 2015). However, street blocks are always irregular in shape and are not suitable for input into a DCNN model. Therefore, a skeleton-based decomposition method that incorporates the street block data is proposed to adaptively split every mapping unit (or street block) into processing units with regular shapes. The proposed decomposition method maintains a better integrity of the mapping units than the uniform decomposition method. Case studies with a WorldView-3 image covering 143 km^2 of Hong Kong and a WorldView-2 image covering 25 km^2 of Shenzhen demonstrate that the proposed STDCNN can obtain better land-use classification accuracies, and can produce more practical land-use maps.

2. Study area and classification system

2.1. Study area and data collection

Hong Kong is one of the world's most significant financial centers, and one of the most popular destinations for visitors. The city's service sector-dominated economy is characterized by free trade and low taxation, and Hong Kong has been consistently listed as the freest market economy in the world. The second study area is Shenzhen which is a major city in Guangdong Province, China and one of the five largest and wealthiest cities of China. The city is located immediately north of Hong Kong Special Administrative Region with more than 10 million population in 2015. Land-use mapping of these areas are valuable for better understanding and analysis of the cities. The complicated spatial arrangement and the various types of land-use in these areas make them worthwhile to generate the land-use map automatically. In this study, the area under investigation covers the metropolitan areas of both Hong