Remote Sensing of Environment 237 (2020) 111593

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Scale Sequence Joint Deep Learning (SS-JDL) for land use and land cover classification

Ce Zhang[a,b,*], Paula A. Harrison[b], Xin Pan[c,d], Huapeng Li[e], Isabel Sargent[f], Peter M. Atkinson[g,h,i,j,*]

[a] Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK
[b] Centre for Ecology & Hydrology, Library Avenue, Bailrigg, Lancaster LA1 4AP, UK
[c] School of Computer Technology and Engineering, Changchun Institute of Technology, 130012 Changchun, China
[d] The Key Laboratory of Changbai Mountain Historical Culture and VR Technology Reconfiguration, Changchun Institute of Technology, 130012 Changchun, China
[e] Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130102, China
[f] Ordnance Survey, Adanac Drive, Southampton SO16 0AS, UK
[g] Faculty of Science and Technology, Lancaster University, Lancaster LA1 4YR, UK
[h] School of Natural and Built Environment, Queen's University Belfast, Belfast BT7 1NN, Northern Ireland, UK
[i] Geography and Environmental Science, University of Southampton, Highfield, Southampton SO17 1BJ, UK
[j] Institute of Geographic Science and Natural Resources Research, Chinese Academy of Sciences, 11A Datun Road, Beijing 100101, China

ARTICLE INFO

Edited by Emilio Chuvieco

Keywords:
Multi-scale deep learning
Optimal scale selection
Convolutional neural network
Joint classification
Hierarchical representations

ABSTRACT

Choosing appropriate scales for remotely sensed image classification is extremely important yet still an open question in relation to deep convolutional neural networks (CNN), due to the impact of spatial scale (i.e., input patch size) on the recognition of ground objects. Currently, the optimal scale selection processes are extremely cumbersome and time-consuming requiring repetitive experiments involving trial-and-error procedures, which significantly reduce the practical utility of the corresponding classification methods. This issue is crucial when trying to classify large-scale land use (LU) and land cover (LC) jointly (Zhang et al., 2019). In this paper, a simple and parsimonious Scale Sequence Joint Deep Learning (SS-JDL) method is proposed for joint LU and LC classification, in which a sequence of scales is embedded in the iterative process of fitting the joint distribution implicit in the joint deep learning (JDL) method, thus, replacing the previous paradigm of scale selection. The sequence of scales, derived autonomously and used to define the CNN input patch sizes, provides consecutive information transmission from small-scale features to large-scale representations, and from simple LC states to complex LU characterisations. The effectiveness of the novel SS-JDL method was tested on aerial digital photography of three complex and heterogeneous landscapes, two in Southern England (Bournemouth and Southampton) and one in North West England (Manchester). Benchmark comparisons were provided in the form of a range of LU and LC methods, including the state-of-the-art joint deep learning (JDL) method. The experimental results demonstrated that the SS-JDL consistently outperformed all of the state-of-the-art baselines in terms of both LU and LC classification accuracies, as well as computational efficiency. The proposed SS-JDL method, therefore, represents a fast and effective implementation of the state-of-the-art JDL method. By creating a single, unifying joint distribution framework for classifying higher order feature representations, including LU, the SS-JDL method has the potential to transform the classification paradigm in remote sensing, and in machine learning more generally.

1. Introduction

Land use and land cover (LULC) information is essential for diverse applications in geospatial domain, such as urban and regional planning, environmental monitoring and management (Liu et al., 2017; Zhang et al., 2019). LULC information can also provide insights to tackle a multitude of socioeconomic and environmental challenges, including food insecurity, poverty, climate change and disaster risk (Stürck et al., 2015). Recent advances in sensor technologies have led to a constellation of satellite and airborne platforms, from which a large

* Correspondence to: C. Zhang, Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK.
** Faculty of Science and Technology, Lancaster University, Lancaster LA1 4YR, UK
E-mail addresses: c.zhang9@lancaster.ac.uk (C. Zhang), pma@lancaster.ac.uk (P.M. Atkinson).

https://doi.org/10.1016/j.rse.2019.111593
Received 23 April 2019; Received in revised form 24 September 2019; Accepted 2 December 2019
Available online 13 December 2019
0034-4257/

amount of very fine spatial resolution (VFSR) remotely sensed imagery is available commercially. While great opportunities are offered by VFSR imagery to capture fine-grained LULC detail, information extraction and retrieval is still immature and inefficient, primarily undertaken by means of traditional field survey and manual interpretation (Hu and Wang, 2013). Such routine tasks are labour-intensive and time-consuming. At the same time, our environment is constantly changing requiring frequent updates of LULC information to support scientific decision-making. It is, therefore, of paramount importance to develop highly efficient and effective techniques to derive LULC information in an automatic and intelligent fashion.

Over the past twenty years, significant efforts have been made towards the automation of LULC classification methods using VFSR images. Traditional techniques can be categorised into pixel-based and object-based approaches. Pixel-based methods focus on classifying individual pixels based on spectral reflectance, which often result in speckle noise effects with limited classification accuracy, given the spectral and spatial complexity presented in VFSR remotely sensed imagery. Textures (Herold et al., 2003) and contextual information (Wu et al., 2009) can be integrated to characterise spatial patterns using moving kernels or windows. These approaches, however, are built on arbitrarily structured images (e.g., squares), whereas real world objects are often irregularly shaped and structured in specific patterns (Herold et al., 2003). Object-based methods are now adopted widely for LULC image classification based on segmented objects (group of pixels), thereby allowing the extraction of discriminative features (e.g., spectral, texture, shape) within the objects and contextual information between adjacent regions. However, those object-based approaches are often challenged by selecting appropriate segmentation scales to achieve meaningful objects (e.g., particular land cover categories), with under- and over-segmentation occurring within the single image (Ming et al., 2015). Besides, the extracted features that characterise the objects are essentially hand-coded via feature engineering, which is subject to individual user experience and expertise, making it difficult to achieve comparable results when transferring the classifier to other datasets. Additionally, the spatial configurations of land use objects can be extremely difficult to hand-code into explicit features, thus, limiting representation and discrimination through traditional methods. Moreover, traditional methods lack a clear definition of the classification hierarchy (i.e. the level of representations of the landscape) and LULC classes are often used interchangeably in remotely sensed image classification. Ontologically, however, land cover (LC) and land use (LU) are manifested at different levels of representation: LC represents low-level states whereas LU characterises high-level functions of the landscape.

Recently, deep learning-based methods have attracted enormous interest in the field of pattern recognition and computer vision, owing to their capability to learn the most representative and discriminative features hierarchically in an end-to-end fashion (Arel et al., 2010). Deep convolutional neural network (CNN), as a popular deep learning method, has achieved significant breakthroughs in image processing and analysis (Krizhevsky et al., 2012), with impressive results beyond the state-of-the-art in a variety of disciplines, not only in classical computer vision fields such as visual recognition, target detection and robotics, but also in many other practical applications (Hu et al., 2015; Nogueira et al., 2017). In the remotely sensed domain, the CNN has shown huge potential in diverse tasks through high-level feature representations, such as road extraction (Cheng et al., 2017), vehicle detection (Dong et al., 2015), scene classification (Liu et al., 2018), semantic segmentation (Wang et al., 2017), and LULC image classification (Zhang et al., 2018a, 2018b).

Within a CNN network, a patch-based architecture is used to learn and extract higher-level features in image patches autonomously through a hierarchy of filters. As a consequence, the choice of image patch size, as a key CNN parameter, has a significant influence on the scale of representations that are manifested over the landscape and, consequently, the accuracy of remotely sensed image classification. These scales are also dependent on the definition of the LULC classification hierarchy, which is unclear so far. Therefore, the determination of the CNN scale for a specific LULC classification task is still an open question in the remote sensing community, and a common approach is to consider scale variations, that is, not constrain to a single scale representation (Pan and Zhao, 2018). Previous research has attempted to incorporate multiple scales into the CNN network to improve spatial feature representations across different scales (e.g., Lv et al., 2018; Yang et al., 2018; Zhang et al., 2018b). For example, a set of CNNs with different patch sizes and scales were integrated by Deng et al. (2018) and Liu et al. (2018) to enhance feature representations across multiple scales, thereby achieving increased accuracy of scene classification. Yang et al. (2018) utilised multi-scale CNNs to differentiate complex scenes (e.g., airport, residential, commercial) in remotely sensed imagery, and demonstrated increased accuracy compared with single-scale CNN networks. Deep features at a range of scales have also been embedded into the CNN to identify vehicles (e.g., ships, cars) within remotely sensed scenes, leading to increased accuracy of target detection (Q. Li et al., 2018; Y. Li et al., 2018). In remotely sensed image classification, Lv et al. (2018) combined region-based CNNs at multiple scales to differentiate land cover objects with high accuracy and efficiency. In addition, object-based CNNs comprising of two distinctive scales were developed to solve the complex land *use* classification task (Zhang et al., 2018b). Finally, deep features at multiple scales were extracted through CNN networks, and used to boost land cover classification accuracy for hyperspectral images (He et al., 2019). A challenge for these multi-scale CNN techniques, however, is to determine the optimal scales (patch sizes) from a large sampling space that is extremely difficult to explore exhaustively across the full range of scales.

In summary, current LULC classification approaches (both traditional and deep learning methods) suffer from two major issues: (1) definition of the classification hierarchy; and (2) definition of the optimal scale to represent the landscape. In terms of the classification hierarchy, land use (LU) and land cover (LC) are often defined interchangeably, without differentiating their intrinsic differences in semantic meaning. LC represents the physical characteristics of the Earth's surface, whereas LU is defined as a higher-order function within a particular space through a mosaic of different LC categories. The spatially nested and hierarchical relationships between LU and LC are given little consideration in LULC image classification, except for the recently proposed joint deep learning (JDL) method (Zhang et al., 2019). As for the choice of scale, it is challenging to determine an optimal scale that can represent the entire scene of a complex and heterogeneous landscape, and multi-scale feature representations are often incorporated to capture large or small land features over different scales. These multiple scales are searched exhaustively through trial and error and tested through extensive experiments with different combinations of candidate scales (Kim et al., 2011; Ming et al., 2015). For deep learning methods (e.g., CNN), such scale parameterisation processes are extremely time-consuming with a large amount of CNN model training. The process can be labour-intensive with repetitive experiments, especially for joint LU and LC classification such as through the JDL method. Furthermore, the selected multiple scales are considered independently as individual evidence to support integrated decisions, which do not capture the mutual connections among the different scales. As such, these scale selection processes are far from operational for deep learning in remotely sensed image classification.

The objective of this research was to develop an automatic approach that is applicable in engineering practices to model the nested relationships between LU and LC, with the ability to address scale issues effectively and efficiently in remotely sensed image classification. A novel Scale Sequence Joint Deep Learning (SS-JDL) method for LU and LC classification is proposed, in which, scales (input patch sizes) of the CNN networks are autonomously derived as a sequence of

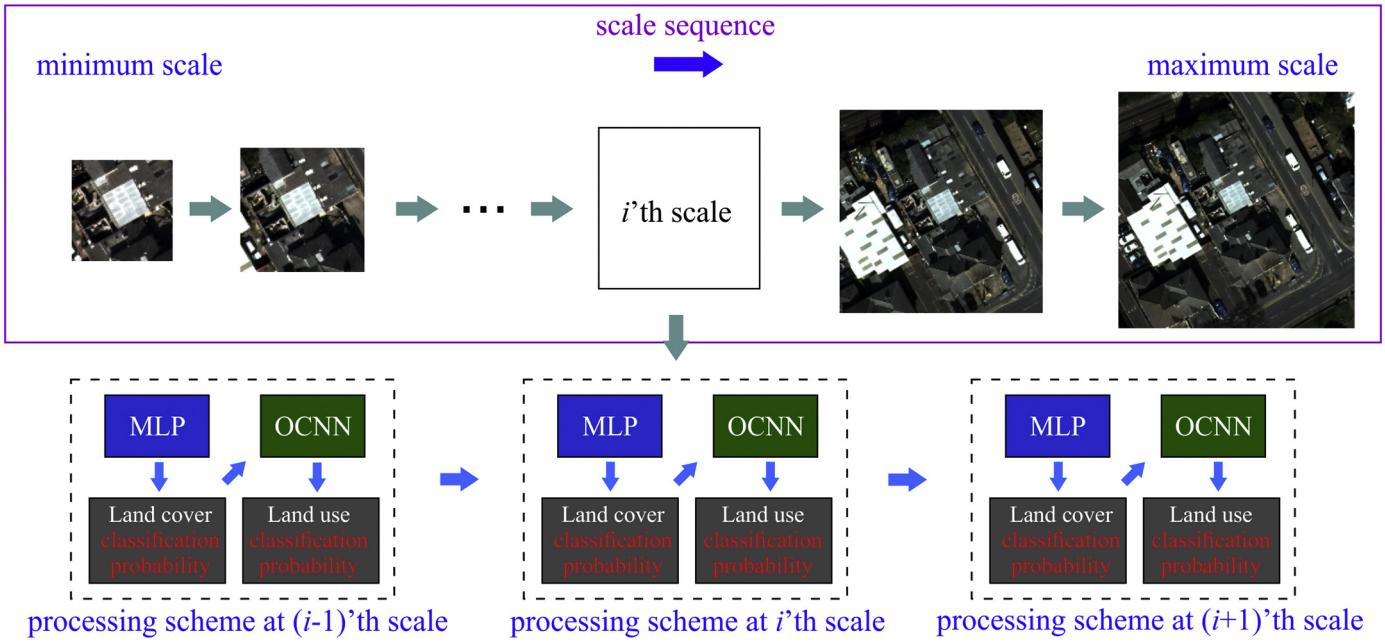

Fig. 1. The general workflow of Scale Sequence Joint Deep Learning (SS-JDL) for land cover and land use classification.

representations. The scale sequence is designed to mimic the human cognition of image pattern recognition through continuously increasing scales, with information transmission between neighbouring scales from small-scale features to large-scale visual representations. The SS-JDL has the key advantage that it is simple and parsimonious in the way that it constructs the sequence of scales and determines an efficient solution, such that the cumbersome and time-consuming process of optimal scale selection is avoided. The rest of the paper is organized as follows: the proposed method is detailed in Section 2; followed by experiments and results analysis in Section 3; discussions and conclusions are made in Sections 4 and 5, respectively.