# A novel spatial recurrent neural network for hyperspectral imagery classification

Andong Ma [a, b, *], Anthony M. Filippi [a, b]

[a] *Department of Geography, College of Geosciences, Texas A&M University, College Station, TX, 77843, USA, adma@tamu.edu*
[b] *Center for Geospatial Sciences, Applications and Technology, Texas A&M University, College Station, TX, 77843, USA, filippi@tamu.edu*

* Corresponding author

Abstract:

Hyperspectral images (HSIs) contain hundreds of spectral bands, providing high-resolution spectral information pertaining to the Earth's surface. Additionally, abundant spatial contextual information can also be obtained simultaneously from a HSI. To characterize the properties of ground objects, classification is the most widely-used technology in the field of remote sensing, where each pixel in a HSI is assigned to a pre-defined class. Over the past decade, deep learning has attracted increasing attention in the machine-learning and computer-vision domains, due to its favourable performances for various types of tasks, and it has been successfully introduced to the remote-sensing community. Instead of utilizing the shallow features within in a given image, which is the approach that is generally adopted in other conventional classification methods, deep-learning algorithms can extract hierarchical features from raw HSI data. Within the deep-learning framework, recurrent neural networks (RNNs), which are able to encode sequential features, have exhibited promising capabilities and have achieved encouraging performances, especially for the natural-language processing and speech-recognition communities. As multi-temporal remote-sensing images can be readily obtained from increasing numbers of satellite and unmanned aircraft systems, and since analysis of such multi-temporal data comprises a critical issue within numerous research subfields, including land-cover and land-change analyses, and land-resource management, RNNs have been applied in recent studies in order to extract temporal sequential features from multi-temporal remote-sensing images for the purpose of image classification. Apart from using multi-temporal image datasets, RNNs can also be utilized on a single image, where the spectral feature/band of each individual pixel can be taken as a sequential feature for the input layer of RNNs. However, the application of such sequential feature extraction that relies on a single image still needs to be further investigated since applying RNNs to spectral bands will directly introduce more parameters that need to be optimized, consequently increasing the total training time.

In this study, we propose a novel RNN-based HSI classification framework. In this framework, unlabelled pixels obtained from a single image are considered when constructing sequential features. Two spatial similarity measurements, referred to as pixel-matching and block-matching, respectively, are employed to extract pixels that are "similar" to the target pixel. Then, the sequential feature of the target pixel is constructed by exploiting several of the most "similar" pixels and ordering them based on their similarities to the target pixel. The aforementioned two schemes are advantageous, as unlabelled pixels within the given HSI are taken into consideration for similarity measurement and sequential feature construction for the RNN model. Moreover, the block-matching scheme also takes advantage of spatial contextual information, which has been widely utilized in spatial-spectral-based HSI classification methods. To evaluate the proposed methods, two benchmark HSIs are used, including a HSI collected over Pavia University, Italy by the airborne Reflective Optics System Imaging Spectrometer (ROSIS) sensor, and an image acquired over the Salinas Valley, California, USA via the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor. Spatio-temporally coincident ground-reference data accompanies each of these respective HSIs. In addition, the proposed methods are compared with three state-of-the-art algorithms, including support vector machine (SVM), the 1-dimensional convolutional neural network (1DCNN), and the 1-dimensional RNN (1DRNN).

Experimental results indicate that our proposed methods achieve markedly better classification performance compared with the baseline algorithms on both datasets. For example, for the Pavia University image, the block-matching based RNN achieves the highest overall classification accuracy, with 94.32% accuracy, which is 9.87% higher than the next most accurate algorithm of the aforementioned three baseline methods, which in this case is the 1DCNN, with 84.45% overall accuracy. More specifically, the block-matching method performs better than the pixel-matching method in terms of both quantitative and qualitative assessments. Based on visual assessment/interpretation of the classification maps, it is apparent that "salt-and-pepper" noise is markedly alleviated; with block-matching, smoother classified images are generated compared with pixel-matching-based methods and the three baseline algorithms. Such results demonstrate the effectiveness of utilizing spatial contextual information in the similarity measurement.