Feature and Deep Learning in Remote Sensing Applications

John E. Ball
Derek T. Anderson
Chee Seng Chan

Feature and Deep Learning in Remote Sensing Applications

John E. Ball,[a],* **Derek T. Anderson,[a]** and **Chee Seng Chan[b]**
[a]Mississippi State University, Department of Electrical and Computer Engineering,
Mississippi State, Mississippi, United States
[b]University of Malaya, Faculty of Computer Science and Information Technology,
Kuala Lumpur, Malaysia

Remote sensing is an extremely active area of research that impacts global topics like agriculture, disaster monitoring and response, defense and security, weather, and non-earth observations. The technologies that power remote sensing—i.e., allow us to observe the universe—include hyperspectral imaging, synthetic aperture radar (SAR), electro-optical, thermal, light detection and ranging (LiDAR), etc. However, while we have advanced optical tools to sense the universe, we lack in computational sophistication to automatically transform this objective data to human-centric decisions. Specifically, humans have been the architects to date of features, algorithms (e.g., classifiers) and their fusion within and across sensors and platforms (e.g., satellites, UAVs, etc.). In recent times, it has become clear that even the best experts are not always able to decide what set of transformations (features, classifiers, etc.) is sufficient for a given problem. The last two decades have represented an uprising against "hand-crafted solutions" in areas like signal/image processing, computer vision, and machine learning. The most famous of these revolts is deep learning, a resurrection of neural networks. The crux of this approach is that machines are better than humans at tasks like those outlined above. This special section is centered on recent advancements in deep learning (and just feature learning in general) in the area of remote sensing.

Deep learning has become the de facto for tasks like detection in computer vision on RGB imagery. However, it has not yet made the same impact on remote sensing. In part, this is because remote sensing has many unique challenges. For example, geospatial systems are plagued by factors like lack of (spatial, spectral, and temporal) labeled training data, high (spatial, spectral, and temporal) dimensionality, domain constraints (e.g., physics), and the need to integrate multiple sources (humans, machines, and sensors), to name a few. Whereas we are excited about the potential of deep learning for remote sensing, we are equally nervous about whether this technology can deliver. Furthermore, deep learning typically results in black-box solutions that give us little to no insight into how they are working and why we should trust them. Regardless of its fate, it is an analytics tool to help us better understand these sensors, platforms, and applications.

In this special section, we requested a combination of theory and applications papers on a variety of topics in remote sensing to showcase what has been done, what is being done, and what big questions remain and need to be tackled by the community. The special section encompassed twenty papers, which included one survey paper; three SAR papers; two papers on ocean remote sensing; four papers on classification and labeling; two papers using multi-modal processing; two papers utilizing spectral-spatial processing for hyperspectral image analysis; three papers on object tracking and recognition; one paper studying how deep networks need to be for remote sensing; one paper on domain adaptation; and one paper on feature extraction methods. These papers are discussed briefly below, where we highlight the main contributions and how certain challenges are overcome in the proposed methods.

A common theme encountered was the use of nonremote sensing pretrained networks and transfer learning. Most articles used or extended convolutional neural networks (CNNs) and were application oriented, with a few providing new deep learning models and modules. Most papers exploited electro-optical data, but there were some SAR, hyperspectral, and multitemporal

*John E. Ball, jeball@ece.msstate.edu

modalities. Many diverse methods were used to combat sparse training data: dilated convolutions, which allowed more shallow networks but still provided large receptive fields; using multitemporal data to augment training; and using an inception module with parallel 1×1 convolution layers. Articles in this special section also highlighted the need for more labeled community benchmark data sets—to train the networks but to also facilitate comparisons between methods and reproducible research—and new theory is needed to fuse (and understand) single and multisensor data.

1 SAR Processing

In SAR processing, traditional results mostly utilize hand-crafted features for feature extraction. The following papers utilized deep CNNs to extract higher-quality features for classification and change detection in SAR imagery analysis. Amrani et al. in "Deep feature extraction and combination for synthetic aperture radar target classification" utilized a pretrained VGG-S net that was fine-tuned on MSTAR SAR data to extract features and a K-nearest neighbor algorithm was used to classify the results. Liu et al. in "Change detection in multitemporal synthetic aperture radar images using dual-channel convolutional neural network" utilized dual CNN channels to extract deep features from SAR imagery for change detection. It is worth noting that their algorithm required no preprocessing or presegmentation. Quach in "Convolutional networks for vehicle track segmentation" overcame the problems of continuity and parallelism in detecting vehicle tracks using current methods by applying small-sized dilated convolutional networks, which exponentially increase the network's receptive field size in a small number of layers. Dilated convolution places spaces between the pixels.

2 Ocean Processing

The oceans cover roughly 71% of the Earth's surface, making ocean remote sensing a very important task. Yao et al. in "Ship detection in optical remote sensing images based on deep convolutional neural networks" is a complicated problem due to the small size of ships and interference from clouds, waves, etc. A deep CNN extracts features and a region proposal network discriminates ships and provides accurate detection bounding boxes. Lima et al. in "Application of deep convolutional neural networks for ocean front recognition" investigated AlexNet, CaffeNet, GoogLeNet, and VGGNet and then developed a custom CNN with fewer layers for this task. In addition to detecting ocean fronts, they also classified them into weak and strong fronts (based on gradient intensity). The reduced model had the smallest learning time and all networks achieved high results despite the small number of training samples.

3 Classification and Labeling

Maskey et al. in "Deep learning for phenomena-based classification of Earth science images" utilized AlexNet and fine-tuned it on remote sensing imagery in order to classify imagery based on Earth science phenomena, such as dust, hurricane, smoke, etc. The CNN-based approach provided superior results and demonstrated that transfer learning works from AlexNet trained on the ImageNet database. Ha et al. in "Deep convolutional neural network for classifying Fusarium wilt of radish from unmanned aerial vehicles" utilized unmanned aerial vehicles (UAVs) to detect Fusarium wilt in radishes. The field are segmented into radish, ground and mulch, then the deep learning system identifies infected radishes with much higher accuracies than conventional methods. Sun et al. in "Semantic labeling of high-resolution aerial images using an ensemble of fully convolutional networks" used a fully convolutional networks (VGG and ResNet) were augmented with cross-scene learning and fused the results with a conditional random field graph method. To handle the large size of aerial imagery, a split-and-merge method was employed which split the image into tiles, and each tile creates a belief map, which are then merged to form an overall belief map.

Maltezos et al. in "Deep convolutional neural networks for building extraction from orthoimages and dense image matching point clouds" uses height information derived from the a dense

image matching algorithm (which does not require LIDAR data or LIDAR/image co-registration) as additional inputs to the deep learning system to identify buildings. The system outperforms shallow methods that attempt LiDAR/image fusion.

4 Multimodal

Chen et al. in "Knowledge-guided golf course detection using a convolutional neural network fine-tuned on temporally augmented data" use a knowledge-driven region proposal, a CNN detector and knowledge-driven postprocessing. Knowledge-derived rules are applied to propose candidate golf regions. Temporal data augmentation is used to enhance the training data. A final postprocessing step removed errors.

Zhang et al. in "Collaborative classification of hyperspectral and visible images with convolutional neural network" utilized a CNN to extract deep spectral features. Next, effective binarized statistical image features are learned as contextual basis vectors for the high-resolution VIS image, followed by a classifier. Then decision fusion unites the spectral and spatial data and statistical data together.

5 Spectral-Spatial Processing

Zhao et al. in "Hyperspectral anomaly detection based on stacked denoising autoencoders" uses both spectral features and fused features extracted via a stacked denoising autoencoder from clustered data to detect anomalies is hyperspectral data.

Abdi et al. in "Spectral-spatial feature learning for hyperspectral imagery classification using deep stacked sparse autoencoder" uses an unsupervised stacked sparse autoencoder to extract high-level feature representations of joint spectral-spatial information. A soft classifier is then used to fine-tune the deep learning architecture.

6 Object Tracking and Recognition

Zheng et al. in "Object tracking by transitive learning using perspective transformation with asymptotic stability" employs multiple frames of local invariant features on and around the object are used to build an object and context template. A nonparametric learning algorithm using transitive matching perspective transformation. The asymptotic stability is shown to be drift-free in terms of long-term tracking. Wang et al. in "DeepPlane: a unified deep model for aircraft detection and recognition in remote sensing images" proposes a model with dual correlative deep networks: the first generates object proposals as well as feature maps; the second network is cascaded upon the first to perform classification and box regression in one shot. The "inception module" which consists of parallel layers of convolutions and pooling and a concatenation layer. This module contains several 1×1 convolution layers that provide feature dimensionality reduction.

Marcum et al. in "Rapid broad area search and detection of Chinese surface-to-air missile sites using deep convolutional neural networks" put forth a deep CNN-based chip detection followed by spatial clustering to rapidly narrow down very large areas for identifying surface-to-air missile sites. The search times were sped up by about 81 times.

7 Architectural Studies

Luo et al. in "Do deep convolutional neural networks really need to be deep when applied for remote scene classification?" analyze five pre-trained networks (AlexNet, CaffeNet, VGG-VD16, GoogLeNet, Resnet) and show that the features learned in shallow layers of deep CNNs are not general enough for remote scenes and that the depth of CNNs seemingly enhances the generalization power of learned features and is essential for remote scene classification. They also provide a history showing how networks have become deeper over time, and provide a nice

chronologic history over time of different approaches classification accuracies on the UC Merced.

8 Domain Adaptation

Ma et al. in "Deep neural network-based domain adaptation for classification of remote sensing images" utilizes class centroid alignment is used for unsupervised domain adaptation (assuming that the class labels are only available in the source domain). Hyperion, NCALM, and WorldView-2 imagery were analyzed. Also, different network configurations were studied and optimization equations are provided in appendices.

9 Feature Extraction

Karim et al. in "Comparative analysis of feature extraction methods in satellite imagery" examined several different feature extraction methods for their ability to discriminate in shadowed regions of the image.

10 Survey Paper

Ball, Anderson, and Chan contributed a survey paper, "Comprehensive Survey of Deep Learning in Remote Sensing: Theories, Tools and Challenges for the Community," which provides a list of challenges and open problems in deep learning for remote sensing, discusses modifications of DL architectures for remote sensing, provides an overview of deep learning tools, and gives an extensive summary of remote sensing datasets. Challenges related to (i) inadequate data sets, (ii) human-understandable solutions for modeling physical phenomena, (iii) big data, (iv) non-traditional heterogeneous data sources, (v) DL architectures and learning algorithms for spectral, spatial, and temporal data, (vi) transfer learning, (vii) an improved theoretical understanding of DL systems, (viii) high barriers to entry, and (ix) training and optimizing the deep learner were discussed in detail. We sincerely thank the numerous reviewers and the authors for their hard work. We also recognize that several papers were submitted that were not a good match to this special section, and these papers were resubmitted to the *Journal of Applied Remote Sensing* as regular papers.

John E. Ball is an assistant professor of electrical and computer engineering at Mississippi State University (MSU), USA. He received a PhD degree from MSU in 2007. He is a codirector of the Sensor Analysis and Intelligence Laboratory (SAIL) at MSU. He has authored 45+ articles, and 22 technical and reports. His research interests are deep learning, remote sensing, and signal/ image processing. He is an associate editor for the *Journal of Applied Remote Sensing*.

Derek T. Anderson received his PhD in 2010. He is currently an associate professor in the Electrical Engineering and Computer Science Department at the University of Missouri-Columbia, Missouri, USA, and an intermittent faculty member with the Naval Research Laboratory. His research is data/information fusion for machine learning and automated decision making in signal/image processing and computer vision. He has published 110+ articles, he is an AE for IEEE TFS and program co-chair of FUZZ-IEEE 2019.

Chee Seng Chan received his PhD from the University of Portsmouth, UK, in 2008. He is currently an associate professor with the Department of Artificial Intelligence, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia. His research interest is computer vision with an emphasis on image/video understanding. He has published 80+ articles and he is an AE for IEEE/CAA JAS.