Deep Learning for Land Cover Classification Using Only a Few Bands

Chiman Kwan
Bulent Ayhan
Bence Budavari
Yan Lu
Daniel Perez

See next page for additional authors

Follow this and additional works at: https://digitalcommons.odu.edu/ece_fac_pubs

Part of the Remote Sensing Commons
Authors
Chiman Kwan, Bulent Ayhan, Bence Budavari, Yan Lu, Daniel Perez, Jiang Li, Sergio Bernabe, and Antonio Plaza
Deep Learning for Land Cover Classification Using Only a Few Bands

Chiman Kwan 1,*, Bulent Ayhan 1, Bence Budavari 1, Yan Lu 2, Daniel Perez 2, Jiang Li 2, Sergio Bernabe 3 and Antonio Plaza 4

1 Applied Research LLC, Rockville, MD 20850, USA; bulent.ayhan@signalpro.net (B.A.); bencebudavari@gmail.com (B.B.)
2 Department of Electrical and Computer Engineering, Old Dominion University, Norfolk, VA 23529, USA; yxxlu003@odu.edu (Y.L.); dpere013@odu.edu (D.P.); jli@odu.edu (J.L.)
3 Department of Computer Architecture and Automation, Complutense University of Madrid, 28040 Madrid, Spain; sebernab@ucm.es
4 Department of Technology of Computers and Communications, University of Extremadura, 10003 Cáceres, Spain; aplaza@unex.es

* Correspondence: chiman.kwan@signalpro.net

Received: 26 May 2020; Accepted: 18 June 2020; Published: 22 June 2020

Abstract: There is an emerging interest in using hyperspectral data for land cover classification. The motivation behind using hyperspectral data is the notion that increasing the number of narrowband spectral channels would provide richer spectral information and thus help improve the land cover classification performance. Although hyperspectral data with hundreds of channels provide detailed spectral signatures, the curse of dimensionality might lead to degradation in the land cover classification performance. Moreover, in some practical applications, hyperspectral data may not be available due to cost, data storage, or bandwidth issues, and RGB and near infrared (NIR) could be the only image bands available for land cover classification. Light detection and ranging (LiDAR) data is another type of data to assist land cover classification especially if the land covers of interest have different heights. In this paper, we examined the performance of two Convolutional Neural Network (CNN)-based deep learning algorithms for land cover classification using only four bands (RGB+NIR) and five bands (RGB+NIR+LiDAR), where these limited number of image bands were augmented using Extended Multi-attribute Profiles (EMAP). The deep learning algorithms were applied to a well-known dataset used in the 2013 IEEE Geoscience and Remote Sensing Society (GRSS) Data Fusion Contest. With EMAP augmentation, the two deep learning algorithms were observed to achieve better land cover classification performance using only four bands as compared to that using all 144 hyperspectral bands.

Keywords: land cover classification; hyperspectral; EMAP; augmented bands; LiDAR; data fusion

1. Introduction

Hyperspectral data have been used for chemical agent detection and classification [1,2], small target detection [3,4], fire damage assessment [5,6], anomaly detection [7–13], border monitoring [14], change detection [15–18], and mineral map abundance estimation on Mars [19,20]. Land cover classification is another application area where hyperspectral data are used extensively [21]. To improve land cover classification, the use of light detection and ranging (LiDAR) data and the fusion of LiDAR and hyperspectral data have been considered by several works [22–27]. Because some land covers differ among themselves with respect to their height, height could be a valuable piece of information.

There is also an increasing interest in adapting deep learning methods for land cover classification after several breakthroughs have been reported in a variety of computer vision tasks such as image
classification, object detection and tracking, and semantic segmentation. The authors of [28] described an unsupervised feature extraction framework, with the name “patch-to-patch convolutional neural network (PToP CNN)”, which was used for land cover classification using hyperspectral and LiDAR data. The authors of [29] introduced a spectral-spatial attention network for image classification using hyperspectral data. In this network, while the recurrent neural network (RNN) part with attention can learn inner spectral correlations within a continuous spectrum, the CNN part is for the saliency features and aims to extract spatial relevance between neighboring pixels in the spatial dimension. The authors of [30] used convolutional networks for multispectral image classification. The authors of [31] provided a comprehensive review on land cover classification and object detection approaches using high-resolution imagery. The performances of deep learning models were evaluated against traditional approaches, and the authors concluded that the deep-learning-based methods provided an end-to-end solution and showed better performance than the traditional pixel-based methods by utilizing both spatial and spectral information, whereas traditional pixel-based methods resulted in salt-and-pepper-type noisy land cover map estimates. A number of other works have also shown that semantic segmentation classification with deep learning methods are quite promising in land cover classification [32–35].

Researchers also utilized spatial information in hyperspectral data via the use of Extended Multi-attribute Profiles (EMAPs) [36]. EMAPs, which are based on morphological attribute filters, generate augmented bands from hyperspectral image bands and have found a considerable interest in land cover classification [37–42]. In the event that limited image bands are available, EMAP augmented bands from these limited image bands could be beneficial. A recent application of using EMAP-augmented bands for soil detection due to illegal tunnel excavation can be found in [43]. Because the number of image bands is too high in hyperspectral data, the number of EMAP augmented bands becomes too high as well. For this reason, works utilizing EMAP have generally applied dimension reduction techniques to extract the most useful information from the EMAP augmented bands [38]. The fusion of hyperspectral and LiDAR data was proposed by the authors [44] and was applied to the 2013 IEEE Geoscience and Remote Sensing Society (GRSS) Data Fusion Contest dataset. All 144 hyperspectral bands were used in EMAP augmentation and the overall accuracy was found to be 90.65%.

Even though using hyperspectral data provides rich and distinctive information that can point subtle differences for land cover classification that would otherwise be impossible to identify using RGB image bands only or even using multispectral image bands, there is a price that comes with it. The high dimensionality in hyperspectral data due to the presence of hundreds of bands could result in the curse of dimensionality phenomena [37,45], especially if there is limited training data when training models with the classifiers for land cover classification. In this situation, the available training data could look sparse relative to the large space volume formed by the high dimensional data. This could potentially result in lower classification performances than anticipated when all hyperspectral bands are used. To mitigate this issue, several works have applied dimension reduction techniques to hyperspectral data [45,46], as well as to the EMAP augmented bands [36,38]. Moreover, hyperspectral sensors are expensive and demand large data storage and computational resources and times. Considering that some applications, such as pest damage assessment by farmers, have limited budgets and the use of hyperspectral sensors could be costly and off-limits for them, using fewer image bands like RGB and near infrared (NIR) for land cover classification via advanced algorithms that can utilize both spatial and spectral information from these fewer image bands would be invaluable.

In this paper, we summarized the evaluation of land cover classification performance of two CNN-based deep learning algorithms when only a few bands, namely RGB+NIR bands and RGB+NIR+LiDAR bands, were available, and when EMAP was applied to those limited bands to generate the augmented bands. This is different from the EMAP application described by the authors of [36–43] since, in those works, all hyperspectral image bands were augmented followed by dimension reduction of the features extracted from the augmented bands. Extensive performance comparisons
were conducted for the cases when all hyperspectral bands were used for land cover classification in contrast to EMAP-augmented bands only and RGB+NIR bands alone. Additionally, the impact of adding LiDAR on the classification performance of the three cases was examined. These three cases are: (a) Limited number of bands (RGB+NIR) vs. (RGB+NIR+LiDAR), (b) EMAP-augmentation of RGB+NIR vs. EMAP augmentation of RGB+NIR+LiDAR, and (c) hyperspectral bands vs. hyperspectral bands+LiDAR. We used the 2013 IEEE GRSS Data Fusion Contest dataset [44] in the investigations. Two conventional classification algorithms, Joint Sparse Representation (JSR) [14] and Support Vector Machine (SVM) [47], were also applied to the same dataset for the EMAP-augmented cases. The contributions of this paper are as follows:

- We provided a comprehensive performance evaluation of two CNN-based deep learning methods for land cover classification when a limited number of bands (RGB+NIR and RGB+NIR+LiDAR) were used for augmentation with EMAP. The evaluation also included detailed classification accuracy comparisons when limited bands were used only and when all hyperspectral bands were used without EMAP.
- We showed that, with deep learning methods using fewer number of bands and utilizing EMAP-based augmentation, it is quite possible to arrive at highly decent accuracies for land cover classification. This eliminates the need for hundreds of hyperspectral bands and reduces it to four bands.
- We demonstrated that, even though adding LiDAR band to RGB+NIR bands for EMAP augmentation made a significant impact with conventional classifiers, JSR and SVM, for land cover classification, no considerable impact was observed with the deep learning methods since their classification performances were already good when using the EMAP-augmented RGB+NIR bands.

The rest of this paper is organized as follows. In Section 2, we review the two deep learning methods, EMAP, and the 2013 IEEE GRSS Data Fusion Contest Data. In Section 3, we summarize our findings. Finally, we conclude our paper with a few remarks.

2. Methods

2.1. Our Customized CNN Method

For this CNN-based method, we used the same structure that was used in our previous work [39]. Only the filter size in the first convolution layer was changed accordingly to be consistent with the input patch sizes. This CNN model has four convolutional layers, with various filter sizes and a fully connected layer with 100 hidden units as shown in Figure 1. When we designed the network, we tried different configurations for the number of layers and the size of each layer and selected the one that provided the best results. We did this for all the layers (convolutional and FC layers). The choice of “100 hidden units” was the outcome of our design studies. This model was also used for soil detection due to illegal tunnel excavation, as described in [48]. Each convolutional layer utilizes the Rectified Linear Unit (ReLU) as the activation function. The last fully connected layer uses the SoftMax function for classification. We added dropout layer for each convolutional layer with a dropout rate of 0.1 to mitigate overfitting [49]. It should be noted that the network size changes depending on the input size. In Figure 1, the network should be the same for $5 \times 5$ as for $7 \times 7$. The only difference would be that the first layer is deleted. For $3 \times 3$, the first two layers are deleted. The number of bands (N) in the input image can be any integer number.
where the hyperspectral bands of interest were only RGB+NIR instead of all the hyperspectral bands available. EMAP was directly applied to these limited number of bands (RGB+NIR) with the selected attribute profiles, which were the ‘area (a)’ and ‘length of the diagonal of the bounding box’ (d) [43]. The lambda parameters in EMAP for both ‘area’ and ‘length of the diagonal

2.2. CNN-3D

CNN-3D is a CNN-based method [50] with two convolutional layers and two fully connected layers. The two convolutional layers have pooling units and all of its four layers have ReLu units. Figure 2 shows the architecture of this second CNN-based method that is used for land cover classification. This network was used in hyperspectral pixel classification before. The details have been described by the authors of [50]. Due to the use of pooling layers, the patch size cannot be too small. Otherwise, the later layers will not have any impacts.

2.3. EMAP

For a grayscale image \( f \) and a sequence of threshold levels \( \{Th_1, Th_2, \ldots, Th_n\} \), the attribute profile (AP) of \( f \) was obtained by applying a sequence of thinning and thickening attribute transformations to every pixel in \( f \) as follows:

\[
AP(f) = \{\phi_1(f), \phi_2(f), \ldots, \phi_n(f), f, \gamma_1(f), \gamma_2(f), \ldots, \gamma_n(f)\}
\] (1)

where \( \phi_i \) and \( \gamma_i \) (\( i = 1, 2, \ldots, n \)) are the thickening and thinning operators at threshold \( Th_i \), respectively. The EMAP of \( f \) was then acquired by stacking two or more APs using any feature reduction technique on multispectral/hyperspectral data, such as purely geometric attributes (e.g., area, length of the perimeter, image moments, shape factors), or textural attributes (e.g., range, standard deviation, entropy) [36,39–41] as shown in (2). Further technical details about EMAP can be found in [36,39–41].

\[
EMAP(f) = \{AP_1(f), AP_2(f) \ldots AP_m(f)\}
\] (2)

When generating the EMAP-augmented bands in the conducted investigations, no feature dimension reduction process was applied to the hyperspectral data (such as Principal Component Analysis) because the hyperspectral bands of interest were only RGB+NIR instead of all the hyperspectral bands available. EMAP was directly applied to these limited number of bands (RGB+NIR) with the selected attribute profiles, which were the ‘area (a)’ and ‘length of the diagonal of the bounding box’ (d) [43]. The lambda parameters in EMAP for both ‘area’ and ‘length of the diagonal
of the bounding box’ attributes were arbitrarily chosen. For the ‘area’ attribute, which is related to modeling spatial information, the higher number of extrema (lambda parameters) contains more detail, whereas the smaller number smooths out the input data. In this work, smoothing was favored and the lambda parameters for the area attribute of EMAP, which is a sequence of thresholds used by the morphological attribute filters, were set to 10 and 15, respectively. For the ‘length of the diagonal of the bounding box’ attribute, which is related to the shape of the regions, the lambda parameters were set to 50, 100, and 500. With these two attributes and their parameter settings, 10 augmented bands were generated for a single band image. The total number of image bands becomes 11 when including the original single band image, Because the resultant EMAP augmented bands were few in number, no feature dimension reduction process was applied to the augmented EMAP bands and all of the EMAP augmented bands were used with the deep learning classifiers.

2.4. Dataset

The hyperspectral image dataset for the University of Houston area and the corresponding LiDAR data were used in this paper. This dataset, together with its ground truth land cover maps, were obtained from the IEEE GRSS Data Fusion package [44] and were used in the 2013 IEEE Geoscience and Remote Sensing Society Data Fusion Contest. The hyperspectral data in this dataset contain 144 bands ranging in wavelength values from 380 mm to 1050 nm with a spectral width of 4.65 nm and a spatial resolution of 2.5 m per pixel. The LiDAR data contain the height information and have a resolution of 2.5 m per pixel. Table 1 displays the number of training and test data pixels per land cover class provided in this dataset. There were 15 land cover classes which were identified and named by the contest. These classes and their corresponding training and test datasets (each pixel in the training and test dataset had its own land cover class label) were fixed. The training data set included 2832 pixels and the test dataset includes the remaining 12,197 labeled pixels. Figure 3a shows the color image, and Figure 3b,c show the color image with ground truth land cover annotations for both the training and test dataset. The brightness and contrast of the color image in Figure 3 was adjusted for better visual assessment.

<table>
<thead>
<tr>
<th>Class</th>
<th>Number</th>
<th>Color Legend</th>
<th>Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy grass</td>
<td>1</td>
<td></td>
<td>198</td>
</tr>
<tr>
<td>Stressed grass</td>
<td>2</td>
<td></td>
<td>190</td>
</tr>
<tr>
<td>Synthetic grass</td>
<td>3</td>
<td></td>
<td>192</td>
</tr>
<tr>
<td>Tree</td>
<td>4</td>
<td></td>
<td>188</td>
</tr>
<tr>
<td>Soil</td>
<td>5</td>
<td></td>
<td>186</td>
</tr>
<tr>
<td>Water</td>
<td>6</td>
<td></td>
<td>182</td>
</tr>
<tr>
<td>Residential</td>
<td>7</td>
<td></td>
<td>196</td>
</tr>
<tr>
<td>Commercial</td>
<td>8</td>
<td></td>
<td>191</td>
</tr>
<tr>
<td>Road</td>
<td>9</td>
<td></td>
<td>193</td>
</tr>
<tr>
<td>Highway</td>
<td>10</td>
<td></td>
<td>191</td>
</tr>
<tr>
<td>Railway</td>
<td>11</td>
<td></td>
<td>181</td>
</tr>
<tr>
<td>Parking lot 1</td>
<td>12</td>
<td></td>
<td>192</td>
</tr>
<tr>
<td>Parking lot 2</td>
<td>13</td>
<td></td>
<td>184</td>
</tr>
<tr>
<td>Tennis court</td>
<td>14</td>
<td></td>
<td>181</td>
</tr>
<tr>
<td>Running track</td>
<td>15</td>
<td></td>
<td>181</td>
</tr>
<tr>
<td>1–15</td>
<td></td>
<td></td>
<td>2832</td>
</tr>
</tbody>
</table>

Table 1. Number of pixels per class in the IEEE GRSS data. Total number of unlabeled pixels is 649,816, and total number of pixels is 664,845.
Figure 3. Color image for the University of Houston area with ground truth land cover annotations of test data.

In the hyperspectral image, the investigations with the four bands (RGB+NIR) correspond to the narrow hyperspectral bands, which were directly retrieved from hyperspectral data. These bands are Red (R), Green (G), Blue (Blue), and NIR, and the band numbers in the hyperspectral data for RGB and NIR bands are (R), #30 (G), #22 (B), and #103 (NIR).

We would like to mention that there is another dataset known as Trento data that contains both hyperspectral and LiDAR images. However, the Trento dataset is no longer publicly available to researchers.

2.5. Performance Metrics

For classification performance evaluation of the deep learning methods with EMAP-augmented bands and all other band combinations, we used overall accuracy, OA, which is the ratio of the sum of correctly classified pixels from all classes to the total number of pixels in the test data. In addition to OA, we also generated average accuracy, AA, value which corresponds to the average of individual class accuracies, which is also known as ‘balanced accuracy’. The last performance metric was the well-known Kappa (K) coefficient [51].

3. Results

3.1. Our Customized CNN Results

Table 2 shows our customized CNN model [48] classification results for the test dataset with six different sets of image bands. These six different sets of image bands are: (a) RGB+NIR (four bands), (b) RGB+NIR+LiDAR (5 bands), (c) 44 EMAP-augmented bands (RGB+NIR bands augmented to 44 bands with EMAP, $11 \times 4 = 44$), (d) 55 EMAP-augmented bands (RGB+NIR+LiDAR bands augmented to 55 bands with EMAP), (e) 144 hyperspectral bands, and (f) 144 hyperspectral bands+LiDAR (145 bands). The results in Table 2 consist of two parts. The first part corresponds to the resultant performance metrics (OA, AA, and Kappa), and the second part corresponds to the correct classification
accuracy for each land cover type using the investigated band combinations with our customized CNN model. Figures 4 and 5 show these results using bar chart type of plots. When applying the customized CNN model, two different patch sizes were investigated, $3 \times 3$ and $5 \times 5$, respectively. In Figures 4 and 5, the value in parenthesis after the band combination type corresponds to patch size. As an example, (p3) indicates a patch size of $3 \times 3$.

**Figure 4.** Performance metrics for the investigated band combinations with our customized CNN model.

**Figure 5.** Classification accuracy for each land cover type with our customized CNN model using the investigated band combinations.

The highest classification accuracy was found to be with the 44 EMAP-augmented bands case using the $3 \times 3$ patch size, followed by the 55 EMAP-augmented bands case with $3 \times 3$ patch size. Among the 15 different land covers, two land covers, synthetic grass and tennis court, had perfect correct classification followed by close-to-perfect classification accuracy for soil land cover. The lowest classification accuracy was found to be for highway class. Figure 6 shows the estimated land cover maps for the whole image with our customized CNN model. The estimated land cover map with the highest overall accuracy for each image band set is shown in Figure 6.
Table 2. Classification results obtained by our customized CNN model for the considered test dataset with six different sets of image bands. Bold numbers indicate best performing datasets for each row.

<table>
<thead>
<tr>
<th>Band Type</th>
<th>Bands</th>
<th>Patch Size</th>
<th>4 Band</th>
<th>4 Band + Lidar</th>
<th>44 Band EMAP</th>
<th>55 Band EMAP</th>
<th>144 Band</th>
<th>144 Band + Lidar</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>OA (%)</td>
<td>0.7844</td>
<td>0.8108</td>
<td>0.7990</td>
<td>0.7990</td>
<td><strong>0.8792</strong></td>
<td>0.8602</td>
<td>0.8602</td>
<td>0.8429</td>
</tr>
<tr>
<td>AA (%)</td>
<td>0.8103</td>
<td>0.8364</td>
<td>0.8224</td>
<td>0.8264</td>
<td><strong>0.8982</strong></td>
<td>0.8830</td>
<td>0.8840</td>
<td>0.8670</td>
</tr>
<tr>
<td>K</td>
<td>0.7661</td>
<td>0.7947</td>
<td>0.7818</td>
<td>0.7825</td>
<td><strong>0.8694</strong></td>
<td>0.8487</td>
<td>0.8492</td>
<td>0.8304</td>
</tr>
<tr>
<td>(ID)-Class Type (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-Healthy grass</td>
<td>0.8224</td>
<td>0.8234</td>
<td>0.8205</td>
<td>0.8224</td>
<td>0.8310</td>
<td>0.8310</td>
<td>0.8234</td>
<td>0.8310</td>
</tr>
<tr>
<td>2-Stressed grass</td>
<td>0.8449</td>
<td>0.8449</td>
<td>0.8421</td>
<td>0.8459</td>
<td>0.8393</td>
<td>0.8205</td>
<td>0.8487</td>
<td>0.7162</td>
</tr>
<tr>
<td>3-Synthetic grass</td>
<td>0.9426</td>
<td>0.9109</td>
<td>0.9485</td>
<td>0.8792</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9941</td>
<td>0.9980</td>
</tr>
<tr>
<td>4-Trees</td>
<td>0.9252</td>
<td>0.9195</td>
<td>0.8958</td>
<td>0.9223</td>
<td>0.9896</td>
<td>0.9271</td>
<td>0.9574</td>
<td>0.9261</td>
</tr>
<tr>
<td>5-Soil</td>
<td>0.9792</td>
<td>0.9934</td>
<td>0.9735</td>
<td>0.9915</td>
<td>0.9943</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9991</td>
</tr>
<tr>
<td>6-Water</td>
<td>0.8462</td>
<td>0.9301</td>
<td>0.8462</td>
<td>0.9231</td>
<td>0.9510</td>
<td>0.9580</td>
<td>0.9650</td>
<td>0.9021</td>
</tr>
<tr>
<td>7-Residential</td>
<td>0.8647</td>
<td>0.8918</td>
<td>0.8909</td>
<td>0.8386</td>
<td>0.8647</td>
<td>0.9347</td>
<td>0.8750</td>
<td>0.8675</td>
</tr>
<tr>
<td>8-Commercial</td>
<td>0.5404</td>
<td>0.6277</td>
<td>0.5850</td>
<td>0.6068</td>
<td>0.7873</td>
<td>0.6021</td>
<td>0.6135</td>
<td>0.5470</td>
</tr>
<tr>
<td>9-Road</td>
<td>0.6959</td>
<td>0.7866</td>
<td>0.7592</td>
<td>0.7460</td>
<td>0.8659</td>
<td>0.8234</td>
<td>0.8527</td>
<td>0.8036</td>
</tr>
<tr>
<td>10-Highway</td>
<td>0.5531</td>
<td>0.5010</td>
<td>0.6120</td>
<td>0.5154</td>
<td>0.6844</td>
<td>0.6400</td>
<td>0.6371</td>
<td>0.8070</td>
</tr>
<tr>
<td>11-Railway</td>
<td>0.6803</td>
<td>0.7021</td>
<td>0.6774</td>
<td>0.6812</td>
<td>0.8634</td>
<td>0.8966</td>
<td>0.8681</td>
<td>0.8662</td>
</tr>
<tr>
<td>12-Parking lot 1</td>
<td>0.7166</td>
<td>0.8444</td>
<td>0.7301</td>
<td>0.8309</td>
<td>0.9193</td>
<td>0.9462</td>
<td>0.9452</td>
<td>0.8674</td>
</tr>
<tr>
<td>13-Parking lot 2</td>
<td>0.8035</td>
<td>0.8737</td>
<td>0.8316</td>
<td>0.8596</td>
<td>0.8842</td>
<td>0.8702</td>
<td>0.8877</td>
<td>0.8737</td>
</tr>
<tr>
<td>14-Tennis Court</td>
<td>0.9798</td>
<td>0.9919</td>
<td>0.9676</td>
<td>0.9838</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>15-Running Track</td>
<td>0.9598</td>
<td>0.9049</td>
<td>0.9556</td>
<td>0.9493</td>
<td>0.9979</td>
<td>0.9958</td>
<td>0.9915</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Table 3 shows the classification results of the CNN-3D method [49] for the test dataset with six different sets of image bands. With this method, two different patch sizes, $13 \times 13$ and $17 \times 17$, were investigated for three of the five sets of bands. As mentioned earlier, the patch sizes cannot be too small due to the pooling layers in C3D. For instance, if the patch size is $7 \times 7$, the patch size will drop to single digits after the first two layers, hence the network cannot function as desired. For the two

**Figure 6.** Estimated land cover maps by our customized CNN method for the whole image.
were investigated for three of the five sets of bands. As mentioned earlier, the patch sizes cannot be too small due to the pooling layers in C3D. For instance, if the patch size is 7 × 7, the patch size will drop to single digits after the first two layers, hence the network cannot function as desired. For the two remaining band sets, which were 144 hyperspectral bands and the 144 hyperspectral bands + LiDAR, the GPU memory was not sufficient for CNN-3D to run. Consequently, only the patch size of 7 × 7 was considered for 144 hyperspectral bands and the 144 hyperspectral bands + LiDAR cases. Even though comparing the 7 × 7 patch size results of these two image band sets with the 13 × 13 and 17 × 17 patch size results of the other three image band sets would not be fair, we still included these results in Table 3 to provide some rough idea about the impact of adding LiDAR band to the hyperspectral bands.

Figure 7 corresponds to the bar chart of the resultant performance metrics (OA, AA, and Kappa), and Figure 8 shows the bar chart plot of the correct classification accuracy for each land cover type using the investigated band combinations with CNN-3D model in Table 3.

Table 3. CNN-3D classification results for the test dataset with six different sets of image bands. Bold numbers indicate best performing datasets for each row.

<table>
<thead>
<tr>
<th>Bands</th>
<th>4 Band</th>
<th>4 Band + Lidar</th>
<th>44 Band EMAP</th>
<th>55 Band EMAP</th>
<th>144 Band</th>
<th>144 Band + Lidar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patch Size</td>
<td>13</td>
<td>17</td>
<td>13</td>
<td>17</td>
<td>13</td>
<td>17</td>
</tr>
<tr>
<td>OA (%)</td>
<td>0.7536</td>
<td>0.6179</td>
<td>0.7843</td>
<td>0.7066</td>
<td>0.8764</td>
<td>0.8634</td>
</tr>
<tr>
<td>AA (%)</td>
<td>0.7503</td>
<td>0.6767</td>
<td>0.7825</td>
<td>0.6933</td>
<td>0.8816</td>
<td>0.8530</td>
</tr>
<tr>
<td>K</td>
<td>0.7332</td>
<td>0.6634</td>
<td>0.7663</td>
<td>0.6824</td>
<td>0.8657</td>
<td>0.8515</td>
</tr>
</tbody>
</table>

(ID)-Class Type (%)

1-Healthy grass
2-Stressed grass
3-Synthetic grass
4-Trees
5-Soil
6-Water
7-Residential
8-Commercial
9-Road
10-Highway
11-Railway
12-Parking lot 1
13-Parking lot 2
14-Tennis Court
15-Running Track

Figure 7. Performance metrics for the investigated band combinations with the CNN-3D model.
Figure 7. Performance metrics for the investigated band combinations with the CNN-3D model.

Figure 8. Classification accuracy for each land cover type with CNN-3D model using the investigated band combinations.

With CNN-3D, the highest classification accuracy was obtained with the 55 EMAP-augmented bands and 44 EMAP-augmented bands cases using a patch size of $13 \times 13$. The 55 EMAP-augmented bands case was observed to have slightly higher accuracy than the 44 EMAP-augmented bands case. The classification results with all other image band sets had relatively lower values. For the 55 EMAP-augmented bands case, the top three close-to-perfect correct classifications were for the soil, synthetic grass, and tennis court land covers, whereas the lowest classification accuracy was for highway class. In our customized CNN model results, similar observations were made with respect to the land covers with best top three and the worst classification accuracies. Figure 9 shows the estimated land cover maps with this method for the whole image. For image band sets with more than one investigated patch window size, the one providing better classification performance is shown in Figure 9.

Figure 9. Cont.
Remote Sens. 2020, 12, x FOR PEER REVIEW 12 of 17

(d) 55 EMAP-augmented bands from RGB+NIR+LiDAR; patch size 13 × 13

(e) 144 hyperspectral bands; patch size 7 × 7

(f) 144 hyperspectral bands+LiDAR (145 bands); patch size 7 × 7

(g)

Figure 9. CNN-3D method estimated land cover maps for the whole image.

3.3. Performance Comparison of Deep Learning Methods

Table 4 shows the highest classification accuracy values obtained with our customized CNN model and CNN-3D method for each of the six image band sets. In five of these six cases, our customized CNN method was observed to perform better than CNN-3D method. The CNN-3D performed better than our customized CNN method only in the 55 EMAP-augmented bands case.

Considering the best classification results from any of the two deep learning methods in each image band set, one interesting observation from Table 5 is that using RGB+NIR bands (four bands) without LiDAR and using hyperspectral bands (144 bands) without the LiDAR resulted in slightly better classification accuracies than the same cases which included LiDAR. For the 44 EMAP-augmented bands case and 55 EMAP-augmented bands case, a similar observation was made when the average accuracy and Kappa metrics were taken into consideration instead of the overall accuracy metric, since these two metrics were found to be slightly higher in the 44 EMAP-augmented bands case in comparison to the 55 EMAP-augmented bands case. For the overall accuracy metric, they were also very close to each other in value. Considering the results with and without LiDAR in these three cases, it was observed that including LiDAR did not seem to have a considerable impact with the deep learning methods. This is somewhat not intuitive, since LiDAR was expected to bring valuable information with respect to height. This could be because several of the 15 land covers in this dataset are surface-level type land covers, meaning they have similar heights with the exception of trees, residential, and commercial land covers. When looking at the classification accuracies of these three land cover classes for the LiDAR-added cases to see whether LiDAR had any impact, mixed results were also seen. It was anticipated that no significant impact of LiDAR would be observed with the two deep learning methods because their classification performances were already good when using RGB+NIR bands.
By comparing the classification maps of Figures 6 and 9, it can be seen that our customized CNN results look crisper due to the use of small patch sizes. In a nutshell, the two CNNs differed only in the pooling layers. In our investigations, we found that the pooling layers did not help the overall classification performance.

### Table 4. Classification performance comparison of the two deep learning methods for the test dataset with six different image band sets (best performances are highlighted in bold for each image band set).

<table>
<thead>
<tr>
<th>CNN</th>
<th>Bands</th>
<th>4 Band</th>
<th>4 Band+LiDAR</th>
<th>44 Band EMAP</th>
<th>55 Band EMAP</th>
<th>144 Band</th>
<th>144 Band+LiDAR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patch Size</td>
<td>C3D-13</td>
<td>C3D-13</td>
<td>CNN-5</td>
<td>CNN-5</td>
<td>C3D-13</td>
<td>CNN-3</td>
</tr>
<tr>
<td>OA (%)</td>
<td>0.7536</td>
<td>0.8108</td>
<td>0.7843</td>
<td>0.7990</td>
<td>0.8764</td>
<td>0.8792</td>
<td>0.8796</td>
</tr>
<tr>
<td>AA (%)</td>
<td>0.7503</td>
<td>0.8364</td>
<td>0.7825</td>
<td>0.8264</td>
<td>0.8816</td>
<td>0.8982</td>
<td>0.8918</td>
</tr>
<tr>
<td>K</td>
<td>0.7332</td>
<td>0.7947</td>
<td>0.7663</td>
<td>0.7825</td>
<td>0.8657</td>
<td>0.8694</td>
<td>0.8692</td>
</tr>
</tbody>
</table>

#### 3.4. Comparison with Conventional Methods

The deep learning methods’ best classification performances for 44 EMAP-augmented bands and 55 EMAP-augmented bands cases were compared with two conventional methods, Joint Sparse Representation (JSR) [14] and Support Vector Machine (SVM) [47], using the same two sets of EMAP-augmented bands (44 and 55). JSR is more computationally demanding, as it exploits neighborhood pixels for joint land type classification. In JSR, a $3 \times 3$ or $5 \times 5$ patch of pixels is used in the S target matrix. The parameter $s_0$ in Equation (13) of [52] is the design parameter, which controls the number of sparse elements. Details of the mathematics have been described by the authors of [52]. We chose that parameter to be $10^{-4}$. There were two SVM parameters, the penalty factor $C$ and the radial basis function (RBF) kernel width, $\gamma$, which were chosen to be 10 and 0.1, respectively. In addition to these two conventional methods, classification results reported by two papers, which used this same dataset with hyperspectral and LiDAR bands augmented with EMAP, were also considered for comparison as well. Details of the JSR and SVM results have been described by the authors of [52].

The resultant overall accuracies of these methods are shown in Table 5. The two conventional methods were observed to perform extremely close in accuracy to the deep learning methods for the 55 EMAP-augmented bands case. For the 44 EMAP-augmented bands case, on the other hand, the deep learning methods’ performances were considerably better than the conventional methods. From Table 5, it can be noticed that adding the LiDAR band to the RGB+NIR bands for EMAP augmentation made a significant impact in overall classification accuracy with the conventional classifiers, JSR and SVM, for land cover classification. However, no considerable impact was observed when adding the LiDAR band to the the EMAP augmentation with the deep learning methods as discussed before.

In [44], the authors investigated three different combinations of hyperspectral bands for land cover classification. Among these, the investigation which used the hyperspectral bands, additional bands from LiDAR data, and the EMAP-augmented hyperspectral bands resulted in an overall accuracy values of 90.65%, whereas the overall accuracy when using only the hyperspectral bands and EMAP-augmented hyperspectral bands was 84.40%. In another paper [53], a graph-based approach was proposed to fuse HS and LiDAR data for land cover classification. In [53], the authors used the same dataset with three different band combinations. Their investigation with hyperspectral bands only using SVM classifier had an accuracy of 80.72%. Our results using only hyperspectral bands are in line with this result, since we reached an accuracy of 81.48% using our customized CNN method, which provided slightly higher accuracy than their result. A second investigation described by the authors of [53] used the morphological profiles of hyperspectral and LiDAR bands and reached an accuracy of 86.39%.

In our case, even though only four bands (RGB+NIR) and five bands (RGB+NIR+LiDAR) were used with EMAP for augmentation, decent classification accuracies of 87.92% (Our CNN) and 87.96% (CNN-3D) were reached, respectively. These overall accuracy values are higher than the investigations.
described by the authors of [44,53], which used all hyperspectral bands and additional LiDAR bands together with EMAP augmentation. This is quite impressive considering that we augmented only four bands with EMAP instead of using all hyperspectral bands and additional bands from LiDAR and still reached an overall accuracy of 87.92%. Our work shows that with deep learning methods using a fewer number of bands and utilizing EMAP augmentation, it is possible to arrive at very decent correct classification accuracies for land cover classification, eliminating the need of using hundreds of hyperspectral bands. This could ultimately reduce computation and data storage needs.

### Table 5. Comparison of the overall accuracies (%) of several methods.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Dataset Adopted</th>
<th>Algorithm Adopted</th>
<th>Overall Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>[52]</td>
<td>44 EMAP (RGB+NIR)</td>
<td>JSR</td>
<td>80.77</td>
</tr>
<tr>
<td>[52]</td>
<td>44 EMAP (RGB+NIR)</td>
<td>SVM</td>
<td>82.64</td>
</tr>
<tr>
<td>This paper</td>
<td>44 EMAP (RGB+NIR)</td>
<td>Our CNN</td>
<td>87.92</td>
</tr>
<tr>
<td>This paper</td>
<td>44 EMAP (RGB+NIR)</td>
<td>CNN-3D</td>
<td>87.64</td>
</tr>
<tr>
<td>[52]</td>
<td>55 EMAP (RGB+NIR+LiDAR)</td>
<td>JSR</td>
<td>86.86</td>
</tr>
<tr>
<td>[52]</td>
<td>55 EMAP (RGB+NIR+LiDAR)</td>
<td>SVM</td>
<td>86.00</td>
</tr>
<tr>
<td>This paper</td>
<td>55 EMAP (RGB+NIR+LiDAR)</td>
<td>Our CNN</td>
<td>86.02</td>
</tr>
<tr>
<td>This paper</td>
<td>55 EMAP (RGB+NIR+LiDAR)</td>
<td>CNN-3D</td>
<td>87.96</td>
</tr>
</tbody>
</table>

Hyperspectral data; EMAP augmentation applied hyperspectral data ($X_h + \text{EMAP}(X_h)$) MLRsub 84.40

Hyperspectral data; Additional bands from LiDAR data ($X_h + \text{AP}(X_L)$) MLRsub 87.91

Hyperspectral data; EMAP augmentation applied hyperspectral data; Additional bands from LiDAR data ($X_h + \text{AP}(X_L) + \text{EMAP}(X_h)$) MLRsub 90.65

Hyperspectral data | SVM 80.72

Morphological Profile of hyperspectral and LiDAR data (MPHS Li) SVM 86.39

Here, we compared the computational costs of different methods. We used a desktop PC with i7 quad-core CPU and a GPU (NVIDIA GeForce Titan X Pascal 12GB GDDR5X). The JSR and SVM results only used CPU, and the CNN and C3D used GPU. Table 6 summarizes the computational times in minutes. It can be seen that SVM were much faster than JSR because JSR needed a lot more time in minimizing the sparseness constraint. The CNN and C3D times are similar and less than three minutes, which might be misleading. It should be noted that the deep learning methods all used GPU, hence the computational times are less than those conventional methods.

### Table 6. Computational times in minutes for different methods and datasets. CNN and C3D used GPU, whereas JSR and SVM used CPU only.

<table>
<thead>
<tr>
<th>Computational Times (min)</th>
<th>4 Band</th>
<th>4 Band+LiDAR</th>
<th>44 Band EMAP</th>
<th>44 Band EMAP</th>
<th>55 Band EMAP</th>
<th>144 Band EMAP</th>
<th>144 Band+LiDAR EMAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>~1.5</td>
<td>~1.5</td>
<td>~3</td>
<td>~3</td>
<td>~3</td>
</tr>
<tr>
<td>C3D</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>~1</td>
<td>~1.5</td>
<td>~3</td>
<td>~3</td>
<td>~3</td>
</tr>
<tr>
<td>JSR</td>
<td>629.23</td>
<td>891.71</td>
<td>2248.17</td>
<td>2198.56</td>
<td>2210.15</td>
<td>2310.42</td>
<td></td>
</tr>
<tr>
<td>SVM</td>
<td>5.32</td>
<td>3.76</td>
<td>0.69</td>
<td>0.47</td>
<td>1.30</td>
<td>1.41</td>
<td></td>
</tr>
</tbody>
</table>

4. Conclusions

In this paper, we investigated the classification performance of two deep learning methods for land cover classification, where only four bands (RGB+NIR) and five bands (RGB+NIR+LiDAR) were used in terms of EMAP augmentation. In the performance evaluations, the RGB+NIR bands and hyperspectral bands were used without EMAP as well. The results showed that, using deep learning methods and EMAP augmentation, RGB+NIR (four bands) or RGB+NIR+LiDAR (five bands) can produce very good classification performance, and these results are better than those obtained by conventional classifiers such as SVM and JSR. We observed that, even though adding LiDAR band to RGB+NIR bands for augmentation made a significant classification performance impact with respect
to conventional classifiers (JSR and SVM), no significant impact was observed with the deep learning methods, since their classification performances were already very good when using RGB+NIR bands. This work demonstrated that using RGB+NIR (four bands) or RGB+NIR+LiDAR (five bands) with EMAP augmentation is feasible for land cover classification, as the resultant accuracies are only a few percentage points lower than some of the best performing methods in the literature which used the same dataset and utilized all the available hyperspectral bands.

The fusion of different classifiers using decision level fusion [54] or pixel level fusion [55] could be a good future direction. Parallel implementation of JSR and SVM using parallel processing techniques could be another future direction.

Author Contributions: Conceptualization, C.K.; methodology, C.K., B.B., Y.L., D.P., J.L., S.B., A.P.; writing—original draft preparation, B.A. and C.K.; supervision, C.K.; project administration, C.K.; Writing—review & editing—B.A., C.K., S.B., A.P., J.L.; funding acquisition, C.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by US Department of Energy under grant # DE-SC0019936. The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

Conflicts of Interest: The authors declare no conflict of interest.

References
24. Wu, Y.; Zhang, X. Object-Based Tree Species Classification Using Airborne Hyperspectral Images and LiDAR Data. Forests 2020, 11, 32. [CrossRef]


