Computers & Geosciences 164 (2022) 105123

Contents lists available at ScienceDirect

Computers and Geosciences

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

Improving remote sensing classification: A deep-learning-assisted model☆

Tsimur Davydzenka [a], Pejman Tahmasebi [a,*], Mark Carroll [b,c]

[a] College of Engineering and Applied Science, University of Wyoming, Laramie, WY, 82071, USA
[b] Biospheric Sciences Laboratory, University of Maryland, College Park, MD, 20742, USA
[c] Information Science and Technology Office (CISTO), NASA Goddard Space Flight Center, Greenbelt, MD, 20771, USA

ARTICLE INFO

Keywords:
Data augmentation
Image classification
Machine learning
Remote sensing. introduction

ABSTRACT

In many industries and applications, obtaining and classifying remote sensing imagery plays a crucial role. The accuracy of classification, in particular the machine learning methods, mainly depends on a multitude of factors, among which one of the most important ones is the amount of training data. Obtaining sufficient amounts of training data, however, can be very difficult or costly, and one must find alternative ways to improve the accuracy of predictions. To this end, a possible solution that we provide in this study is to use a stochastic method for producing variations of the training images that will retain the important class-wide features and thereby enrich the machine learning's "understanding" of the variabilities. As such, we applied a stochastic algorithm to produce additional realizations of the limited input imagery and thereby significantly increase the final overall accuracy in a deep learning method. We found that by enlarging the initial training set by additional realizations, we are able to consistently improve classification accuracy, compared with generic image augmentation approaches. The results of this study show that there is a great opportunity to increase the accuracy of predictions when enough data are not available.

1. Introduction

Obtaining frequent updates on land use and land cover information is crucial in many industries such as natural resources, vegetation mapping, geospatial data modeling, land management, agriculture, national security, environmental, regional, and urban planning, vegetation mapping, and (Liu et al., 2019; L. Zhang et al., 2016). Updated information can be used, for example, for estimating the quality of water in an aquifer or the state of the crops in a field, preventing natural hazards, or assessing levels of pollution. To accomplish that, researchers use remote sensing (RS) imagery more and more frequently, which is provided by orbiting satellites. Once obtained, the imagery can then be processed and classified to provide insight into a specific problem.

Classification methods used in remote sensing imagery can be mainly categorized as follows: pixel-based, sub-pixel-based, and object-based techniques. In the first category, each pixel is generally classified as a single land cover/land use type. Examples of this category include k-means, ISODATA, Support Vector Machine (Unsupervised), Minimum distance-to-means, Maximum likelihood, Mahalanobis distance (Supervised), and artificial neural networks (e.g., classification tree, random forests). In the second category, each pixel is based on the proportions of several classes. The most prominent methods in this category are neural networks, fuzzy classification, regression tree analysis, and fuzzy-spectral mixture analysis. Lastly, in the object-based techniques, single units are represented by geographical objects, and not by individual pixels. Examples of these methods are E-cognition, ArcGIS, and Feature Analyst. However, to represent the required images, most of these techniques rely on low-level visual features, which can be global or local. Such features like the color (Bosilj et al., 2016; Sebai et al., 2015), texture (Shao et al., 2014), and shape (Scott et al., 2011) are considered global, where the entire image is used for feature extraction. Other features, such as scale-invariant feature transform (SIFT) (Lowe, 2004), are considered local and are extracted from patches around a point of interest. Since classification local features are very important in remote sensing, there have been attempts to develop methods for analyzing such images (Shao et al., 2018). However, the global and local features in these methods are hand-crafted and their development is time-consuming. Therefore, the extraction of low-level features is not

☆ Tsimur Davydzenka and Dr. Pejman Tahmasebi together conceived the problem and developed the method. T Davydzenka performed the computations. T Davydzenka, P Tahmasebi, and M Carroll contributed to analyzing the results and writing the paper.

* Corresponding author.
E-mail address: ptahmase@uwyo.edu (P. Tahmasebi).

https://doi.org/10.1016/j.cageo.2022.105123
Received 20 December 2021; Received in revised form 23 March 2022; Accepted 25 April 2022
Available online 29 April 2022

ideal for certain classification problems. In other words, most of the approaches for the classification and segmentation of RS data that rely on low-level global and local features are not capable of producing sufficiently powerful feature representations for RS images (de Lima and Marfurt, 2020; Hu et al., 2015). Thus, the performance of these methods towards RS scene classification only marginally improved in recent years. Hu et al. (2015) have shown that higher-level features are more dominant in scene classification problems. Therefore, extracting high-level features has been shown to be one of the most important aspects and this is what deep-learning methods offer.

In this study, we aim to examine how Convolutional Neural Networks (CNN), as one of the successful deep learning methods, can be applied towards the classification of land cover land use maps, in particular when enough data is not available. The architecture of CNNs aims to mimic the connectivity of neurons in the human brain and was initially inspired by the visual cortex. The classification is accomplished by assigning importance (e.g., weights and biases) to different objects in the image to later differentiate one image from another. Due to the outstanding performance of CNN in terms of classification accuracy in many benchmarks and datasets with millions of natural images (e.g., MNIST database of handwritten digits (Russakovsky et al., 2015) and the ImageNet (W. Zhang et al., 2019) dataset) it is the most widely used method for image classification, detection, and segmentation. Recent developments in deep-learning algorithms, deep CNNs in particular, have further advanced areas like visual object detection, speech recognition, etc. (Lecun et al., 2015). Well known in this field, the AlexNet model has been one of the major breakthroughs and has affected the rapid adoption and development of deep learning in computer vision. With such significant progress, the architecture of CNN is continuously optimized, and the application area is consistently extended.

More importantly, due to its superior feature learning and extraction abilities, CNNs have been successfully applied for classification and object detection of remote sensing data. Maggiori et al. (2017) have developed a framework for pixel-wise classification of RS data with CNNs, where the latter were directly trained to output classification maps from the input data. In another study, Zhong et al. (2019) have developed deep CNNs for classification of crop maps using Landsat Enhanced Vegetation Index (EVI) time series and demonstrated the robustness of Conv1D-based framework in multi-temporal classification tasks for representation of time series. Zhang et al. (2019) proposed a Joint Deep Learning (JDL) for classification of RS data, which makes use of pixel-based MLP, patch-based CNN, mutual complementarity, and join reinforcement. The model was implemented involving iterative updating through the Markov process. The performance of their model was demonstrated to have a higher average classification accuracy compared with Markov process models and object-based CNN (OCNN). Among other recent research in classification of remote sensing data, Gu et al. (2019) have used an algorithm based on Squeeze and Excitation Networks (SENet's) to improve the quality of remote sensing images. Another advance in the RS classification accuracy has been with the introduction of EfficientNet-B3-Attn2 (Alhichri et al., 2021). In this study, a baseline EfficientNet-B3 has been enhanced with an attention mechanism, which enables the network to emphasize relevant regions of the image and suppress the regions that are not important for the accurate classification. The performance of the network has been shown to be superior compared with most popular CNN architectures, including the baseline EfficientNet-B3.

However, despite powerful classification capabilities, the performance of CCN is largely dependent on the amount of training data, as demonstrated in (Hu et al., 2015). As a matter of fact, lack of training data is still one of the prominent issues in remote sensing data classification and researchers must find other ways to improve the accuracy of deep learning methods when they are used on such data. One way this has been addressed is through transfer learning. This approach allows researchers to reuse a model trained on a similar dataset towards new but similar data, which is very effective when the amount of training data for the new task is very limited. This technique has been successfully applied towards remote sensing applications like airplane detection (G. Chen et al., 2018), extracting socioeconomic indicators (Xie et al., 2016), change detection (Larabi et al., 2019), etc. Another approach that has been applied in limited training data cases is data augmentation. This group of methods attempts to improve the network's classification performance by enriching original training data and providing additional variations of that same data so that the network ultimately has access to a larger training set. One of the most common data augmentation methods is applying image transformation (rotation, zooming, shifting, etc.) to original images, which already has been widely applied to remote sensing data (Hirahara et al., 2020; Shawky et al., 2020; Yu et al., 2017). Another powerful approach for image augmentation is Generative Adversarial Nets (GANs) where the neural net generates better counterfeit samples from original data to mislead the other network, where the latter is trained to distinguish the counterfeits more accurately. This approach has been proven to be very effective in limited data scenarios (Gurumurthy et al., 2017), as well as in increasing image resolution (Marchesi, 2017). Another creative image augmentation approach is using data augmentation during the run time of the neural network, to find which specific data transformation leads to the least classification error (Hirahara et al., 2020). These methods, however, either provide unrealistic alternatives or images with less variation.

A possible solution to the above issues can be through enriching the training set with stochastic realizations of the original images such that the crucial features of the respective classes are retained and, therefore, the network classification accuracy can be improved by having access to larger training data. In this study, we apply a cross-correlation based algorithm, which has been proven to be very effective for producing complex 2D and 3D images (Kamrava et al., 2019; Tahmasebi and Sahimi, 2012, 2013, 2016a, 2016b), to improve the classification accuracy of remote sensing data using CNN.

The remainder of this paper is organized as follows: In Section 2 we discuss the methodology of convolutional neural networks and the data augmentation algorithm that was used for producing realizations of the training images. In Section 3 we discuss CNN architectures that were used for the classification of SAT-6 and UCMerced imagery. In Section 4 we discuss the results of model classification performance for the two above-mentioned datasets. Finally, Section 5 summarizes and concludes the paper.

2. Methodology

In the first part of this section, we will discuss the methodology of the convolutional neural networks and its most important concepts, and the second part of this paper will be focused on describing the cross-correlation based simulation (CCSIM) algorithm that was used for producing realizations of the training images that ultimately helped improve the predictive capabilities of the CNN model.

2.1. CNN

In recent years Convolutional Neural Networks (CNNs) have gained remarkable popularity in their use towards classification and segmentation applications. The year 2012 is considered a milestone in machine learning, as computer vision made giant progress in classification accuracy by introducing the award-winning architecture AlexNet, which surpassed the second-best classification algorithm by a significant margin. Ever since CNNs have been the preferred classification/segmentation tool of choice for most computer vision researchers. Regardless of the network architecture, all CNNs operate by convolutions, where each operation is applied twice, on the input and on the filter. The output is produced by applying filters on the input, which can range in size. In most cases, the convoluted output is later altered by an activation function. Based on the input and underlying mathematical equations, the activation function determines the output and accuracy of

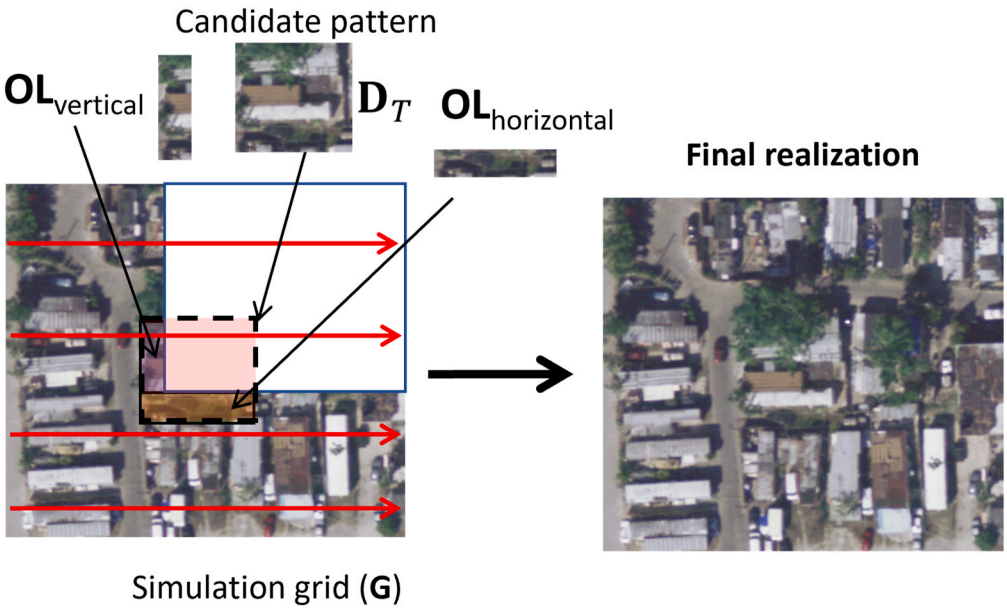

Fig. 1. Schematic demonstration of the proposed augmentation algorithm in the middle of *G*. The raster path based on which the new image is generated is shown in red arrows. The *OL* regions along with a pattern that is selected from the *DI* are also shown. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

the model, as well as the computational efficiency. The function is applied to each node (i.e., neuron) of the network and determines whether the neuron should be activated, depending on if the input of the neuron is relevant to the prediction. In this study, we apply ReLU (Rectified Linear Unit) activation function for all convolutional layers in the network and softmax for the final (output) layer and use Categorical Cross-entropy (CE) as the loss function. To find the optimal values for kernels/filters and eventually minimize the loss function, we use back-propagation which is a widely used algorithm in machine learning. Here, the gradients of the activation functions in each successive unit are tracked, to optimize the filters and reduce the loss function. The gradients are then used by an optimization algorithm (Adam (Kingma and Ba, 2015) in our case) to update the necessary parameters. It ultimately calculates to what degree the output values are affected by each individual weight of the model, by going back from the error function to a specific weight.

2.2. Data augmentation

The rationale behind data augmentation is to expand the training set and diversify the patterns. In this study, we apply cross-correlation based simulation (CCSIM) algorithm as the image augmentation method. The main advantage of this approach compared to the standard augmentation methods is that this algorithm allows producing novel image features that are impossible to produce with generic approaches. In this algorithm, the digital input image (*DI*) is modeled on a computational grid (*G*), partitioned into several overlapping blocks, where both the computational grid and *DI* are equal in size. It should be noted that one can use a smaller image and produce larger images as well. The overlap size (*OL*) between neighboring blocks is $l_x \times l_y$, which is considered to make sure the blocks will have a smooth transition. The algorithm is first applied to the corner block of *G* and then proceeds to every single grid block alongside a one-dimensional raster path, namely a raster path as in remote sensing. A pattern of heterogeneity (data event D_T) is randomly selected from the *DI* and inserted in the visiting block. It is important to note that each pattern of heterogeneity can change again for the next model as such a pattern is selected randomly. Then, a small overlap is selected from the existing pattern and its similarity is calculated with all the patterns in the *DI*, which is achieved using Eq. (4). Therefore, at this stage, the algorithm does not consider all the previously constructed blocks. One can sort the similarities from the best to the worst matching pattern. However, a specific number of the best patterns are chosen and one of them is selected randomly and inserted in the next block in *G*. The following equation, which represents the convolution of matrices, is used for quantifying the distances/similarities:

$$\psi(i,j,x,y) = -\sum_{x=0}^{l_x-1}\sum_{y=0}^{l_y-1} DI(x+i,y+j)D_T(x,y) \tag{4}$$

where $i \in [0, T_x + l_x - 1)$ and $j \in [0, T_y + l_y - 1)$,

The patterns in the digital image are matched via the overlap region (OR) between the two neighboring blocks and the data event. The OR consists of pixels selected from blocks simulated earlier and is used in Eq. (4) to identify the next data event. All the blocks comprising the grid are constructed by repeating the procedure described above. As mentioned, by starting a new simulation one can produce a totally different model and a countless number of realizations can be produced in this manner; see Fig. 1.

3. Model setup

To study the effectiveness of the utilized augmentation algorithm at improving the CNN performance by producing realistic realizations, we have chosen two RS datasets that are different in terms of resolution, the complexity of the patterns, and the number of classes. First, the algorithm was tested on the simpler dataset (SAT-6), and then after obtaining successful results, it was applied on the more complex one (UCMerced), with some of the classes not being included.

3.1. SAT-6 dataset

Originally introduced by Basu et al. (2015), the SAT-6 dataset is composed of 405,000 non-overlapping image patches produced from the National Agriculture Imagery Program (NAIP) spanning the entirety of the Continental United States (approximately 800 square kilometers in total), each with the size 28 × 28 pixels and with 4 bands (red, green, blue, and Near Infrared). The images cover 6 landcover classes: water

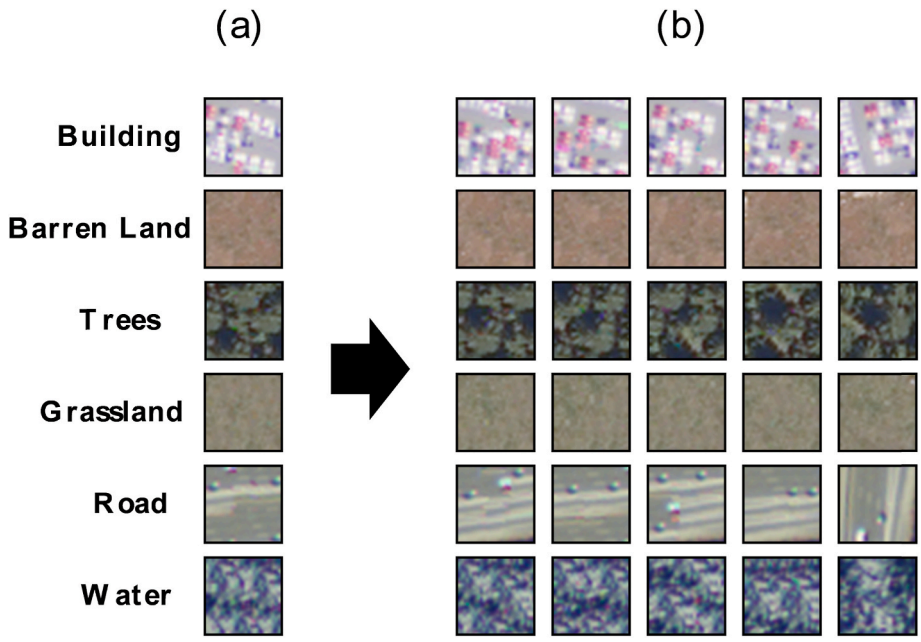

Fig. 2. Demonstration of the realizations produced by CCSIM of the first 6 images of the selected from SAT-6 dataset 60 training images.

bodies, trees, buildings, roads, grassland, and barren land. The original images are acquired at 1-m ground sample distance (GSD) with horizontal accuracy around 6 m of ground control points identifiable from the acquired imagery, covering different landscapes such as urban areas, rural areas, densely forested, water bodies, mountainous terrain, agricultural areas, etc. This dataset has been widely used to benchmark accuracies of novel RS classification methods (Albert et al., 2017; Ball et al., 2017; Cheng et al., 2020; Isikdogan et al., 2017).

First, to quantify the model performance using the original data, a limited number of images were selected randomly (10 images per class) from the dataset to be further used as training images. The choice of this limited sample of images stems from assuming a limited data scenario, and 10 images per class allow us to better highlight improvement in model performance when following augmentation workflow. This situation can be found in many remote sensing studies as the existing images are often dropped due to incomplete scenes or weather conditions. Here, 20% of the data are used as validation to prevent overfitting. To test the performance, 40 images per class are further randomly selected from the remaining dataset. Consequently, to observe the effectiveness of the proposed data augmentation method at increasing the accuracy of model predictions, we applied the algorithm to produce 5 realizations of every single image in the training set. Fig. 2 demonstrates the produced realizations.

As can be seen from Fig. 2, when applied to relatively simple, small, and stationary images, the proposed method is capable of producing images that are visually very similar to the class they pertain to. In the case of the SAT-6 dataset, the class on which applying the proposed method produced relatively inferior results is the Road class, which is

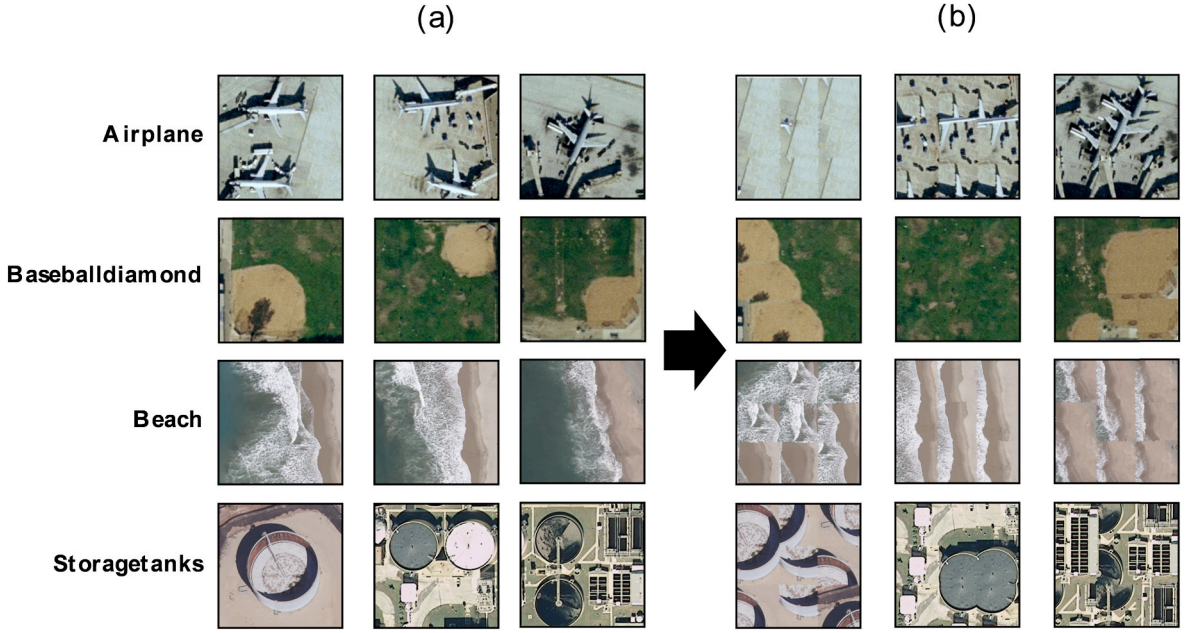

Fig. 3. Demonstration of the problematic realizations produced by CCSIM using 4 of the 12 complex UCMerced classes.

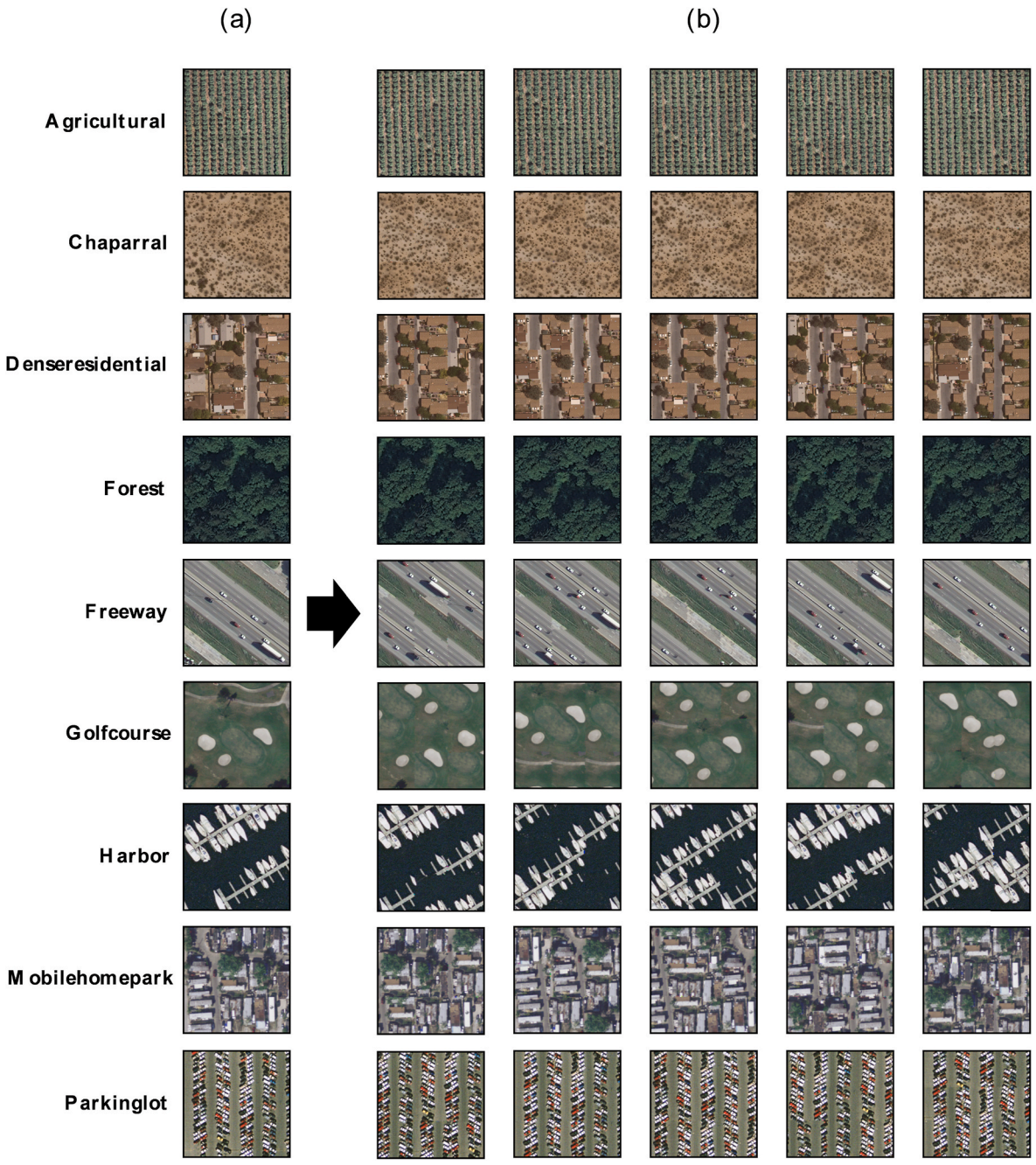

Fig. 4. Demonstration of the realizations produced by CCSIM of 9 random images in the original training set (UCMerced dataset).

due to presenting non-stationary behavior. Among the other 5 classes, the road class is the most sophisticated in terms of special relationships and connectivity of the patterns, therefore some of the realizations of this class are far from perfect.

By appending the original training images with the produced realizations, the size of the training set is increased by 500%. Similarly, 20% of the images are assigned to validation and are chosen randomly. The model was trained for 800 epochs. For both the original and enriched dataset, the training batch size was 20 images per step. To conduct the model training, we used Keras (Chollet, 2013) framework, and Adam optimization algorithm with the following hyperparameters: Learning rate $= 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 1 \times 10\text{-}7$ for both SAT-6 and UCMerced datasets.

University of California Merced Land Use Dataset (UCMerced).

Introduced by Yang and Newsam (2010), UCMerced dataset contains 2,100 images pertaining to 21 classes. Most images have a resolution of 256 × 256 pixels with rare exceptions. The images of various urban areas in the US were extracted and manually labeled from large aerial orthoimagery with RGB color space and a spatial resolution of 0.3 m per pixel, provided originally by USGS National Map Urban Area Imagery open-source image depository. The original maps covered the following US regions: Birmingham, Boston, New York, Buffalo, Dallas, Columbus, Harrisburg, Jacksonville, Houston, Las Vegas, Miami, Los Angeles, Napa, San Diego, Reno, Santa Barbara, Tampa, Seattle, Ventura, and Tucson. The dataset can be characterized by relative complexity and high spatial resolution, where images share similar features with general-purpose optical images. Furthermore, the variety of spatial patterns, where are some homogeneous with respect to color, some homogeneous with respect to texture, others lack any homogeneity makes it suitable for further benchmarking. The dataset has been widely

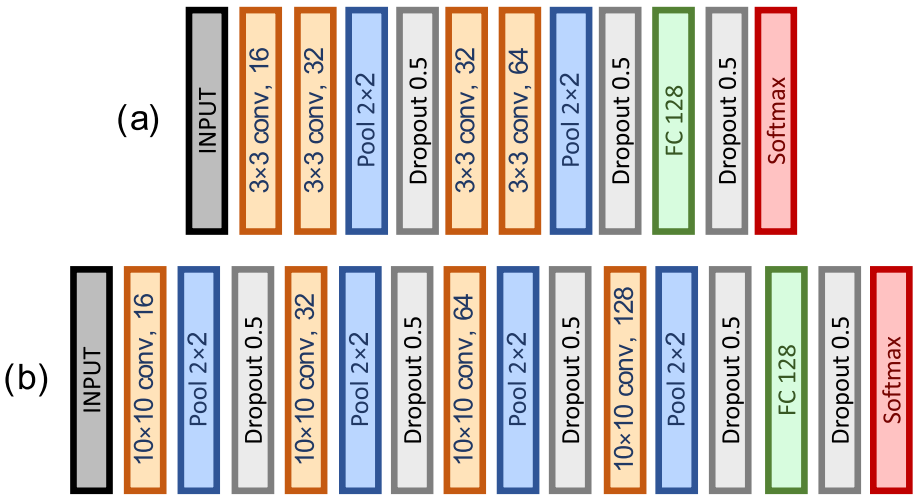

Fig. 5. The utilized CNN model architectures with the SAT-6 (a) and UCMerced (b) datasets.

used for image classification tasks (Castelluccio et al., 2015; Chaib et al., 2017; G. Chen et al., 2018; J. Chen et al., 2018; Gong et al., 2018; Hu et al., 2015; Liu et al., 2018; Marmanis et al., 2016; Yu and Liu, 2018; Zeng et al., 2018; F. Zhang et al., 2016; W. Zhang et al., 2019; Zhu et al., 2016; Zou et al., 2016).

In this paper, the UCMerced dataset was used to examine the capabilities of the proposed algorithm when applied to more complex and higher resolution images. As the algorithm was developed to recreate pattern-based stationary images, it was not effective when applied to classes where the patterns are distributed in a non-stationary fashion. Among the total of 21 classes, the following 12 were found to be unfit for successful image augmentation: baseball diamond, airplane, buildings, beach, intersection, medium residential, river, overpass, runway, storage tanks, sparse residential, and tennis court. Fig. 3 demonstrates the first 4 classes with unsuccessful models. Here, on the left side, we have included three unique images pertaining to the problematic classes, and on the right side, we have shown three realizations produced using such images. As can be seen, although the algorithm is capable of constructing an image that contains features relevant to the class, their special distribution is not realistic. For example, in the airplane class, we note the presence of the wings, tails, and bodies of the airplane, however, their order on the image does not constitute an adequate representation of the airplane class. For baseball diamond class we encounter another limitation, where the realization of the second image is missing the crucial component of the class entirely (the diamond part) and the only feature of the class included in the image is grass.

By visually inspecting the realizations of the total 21 classes, we selected the remaining 9 classes, where the spatial relationship between the image features was optimal for the proposed algorithm to produce adequate realizations. Similar to the SAT-6 dataset, 540 images were randomly selected (30 images per class) from the dataset to be later used for training the CNN models. Further, 20% of the training images were randomly selected at each epoch and used as validation. To test model performance on data that the network had no access to during the training, 1260 of the remaining images (140 images per class) were used as the test set. The proposed data augmentation method was applied to each image in the training set to produce five realizations. An example of the produced realizations from the images in 9 selected classes can be found in Fig. 4.

By adding the 1350 realizations to the original training images, the

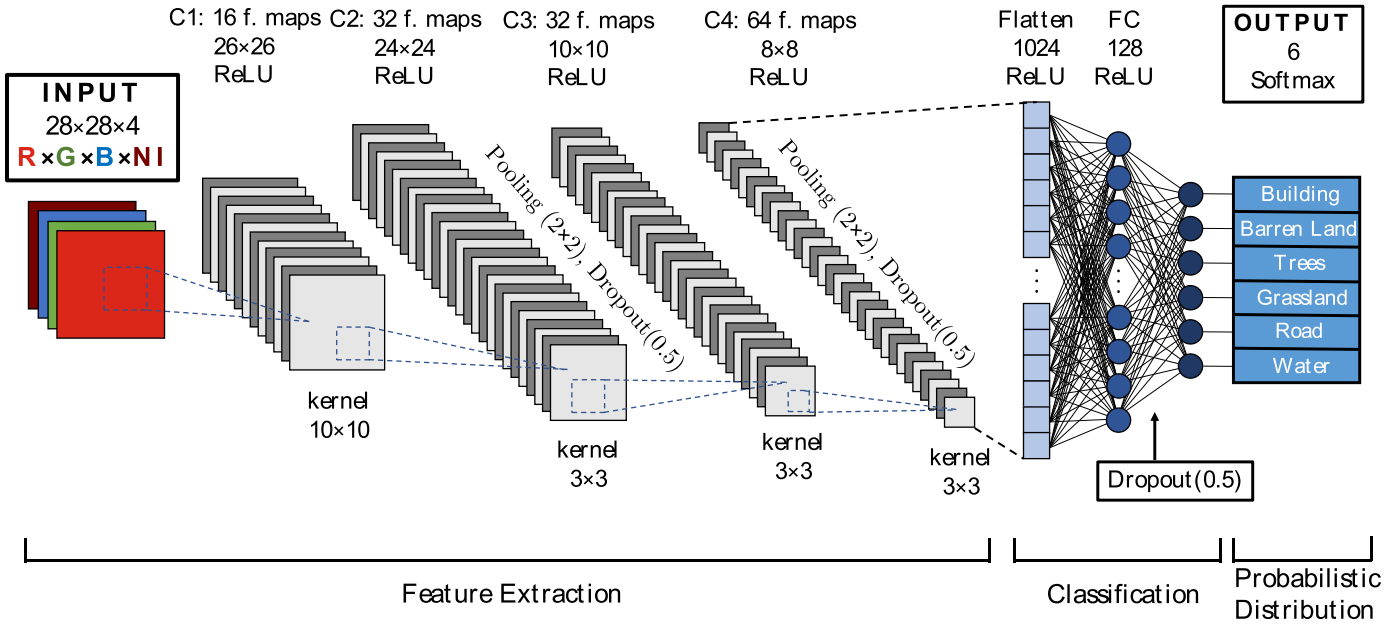

Fig. 6. The CNN model architecture used with SAT-6 dataset (Fig. 5 – a).

Table 1
Classification accuracies (SAT-6 dataset).

Classification Method	Augmentation method		
	None	Generic	Our method
Simple	67.9%	80.0%	83.7%
VGG-19	68.3%	73.5%	82.5%
Inception-v3	87.1%	88.7%	89.3%
ResNet-50	58.9%	79.2%	82.5%

Table 2
Classification accuracies (UCMerced dataset).

Classification Method	Augmentation method		
	None	Generic	Our method
Simple	41.9%	48.1%	55.9%
VGG-19	50.0%	80.8%	82.5%
ResNet-50	44.2%	87.5%	90.7%

training set increased by 500%. To avoid overfitting, 20% of the training images were randomly chosen at each epoch and used as validation. The models were trained for 200 epochs. For both cases, the training batch size was 50.

To compare the performance gains of our data augmentation technique using different CNN architectures, we selected 3 very popular and well-established CNN models, namely VGG-19, Inception-v3, ResNet-50. Additionally, for each of the datasets, the training was conducted using architectures provided in Fig. 5.

Since the resolution of the SAT-6 images is only 28×28, VGG-19 was used without the last MaxPool layer. Furthermore, Inception-v3 network was used only with SAT-6 dataset due to encountered computational limitations when using it with UCMerced dataset.

4. Results and discussions

In this section, we present the results of two CNN networks trained using the SAT-6 and UCMerced datasets: a network having access only to the original images, and a model based on the original images and the produced realizations. Below, we will discuss the evaluation protocol that was followed when comparing the results.

In this work, the experimental results between the original and augmented datasets are evaluated by calculating the overall accuracy (OA) and confusion matrix. The overall accuracy is defined as the ratio of the number of accurately classified images to the total number of images. This metric is generally used as the main measure to quantify classification performance on the images that the network had no access to during the training step (i.e., test images). The value of OA ranges from 0 to 1 where higher values signify better performance of the model in terms of classification accuracy. The confusion matrix (error matrix), as our secondary measure, is a specific informative matrix that allows convenient visualization of the accuracy of an algorithm and is often used for analyzing the performance (errors and confusion) in classification problems. Each matrix row denotes the instances of an actual class and each column indicates the instances of a predicted class. As a result, each value in the matrix is the ratio of the images in a respective class that is predicted accurately to the total number of images in that class, which all represent the test images here.

4.1. SAT-6 dataset

Since the goal in this study is not to achieve the highest classification accuracies but to observe the improvement in model performance when the training set is enriched by a more realistic data augmentation method, we selected a relatively simple network architecture shown in (Fig. 5, – a) along with VGG-19, Inception v3, and ResNet-50. For better demonstration, CNN architecture provided in (Fig. 5, – a) is also included in Fig. 6. As one of the reasons for appending the training set is the scenario of limited training data, we reduce the dataset significantly to better demonstrate the effect of adding more data. To compare the model performance of these two scenarios, we first conducted 20 model training (for each of the models) for 800 epochs using a small dataset and recorded the overall accuracies. The mean values of OAs, for each respective CNN architecture, are listed in Table 1 (column Augmentation method - None). To compare these results with the scenario of augmented training, we kept all the CNN parameters constant and added the produced 5 additional realizations of every image in the training set. The new images are labeled accordingly and included in the training set to bring the total set to be 360 images (60 per class). Estimations of the overall accuracies are carried out using the same test set of 240 images. Results of the mean OAs presented in Table 1 (column Augmentation method – our method). Additionally, to compare the effectiveness of our augmentation method with another well-established method, we use generic image transformations (shifting, rotation, flipping, shearing, zooming, and channel shifting) and produce an equal amount of additional augmented images, as when using our augmentation method. Results of OAs using generic augmentation method are provided in Table 1 (column Augmentation method - Generic). The greatest accuracy improvement from non-augmentation to our method was achieved when using ResNet-50 architecture (23.6%), which is 3.3% better than generic augmentation. The smallest accuracy improvement was obtained using Inception-v3 architecture, where the non-augmentation approach yielded 87.1% accuracy, and the approach using our augmentation method – 89.3%.

4.2. UCMerced dataset

In this study, UCMerced dataset was used in addition to the SAT-6 dataset to observe the ability of our proposed method to produce realistic realizations of larger and more complex images. To evaluate the performance using only original images, we selected 270 images randomly from the dataset for training (30 images per class from 100) and conducted CNN training using the architecture presented in (Fig. 5 – b), along with VGG-19 and ResNet-50. The remaining 630 images (70 images per class) were used for testing. The number of classes that are used for training and testing was reduced to 9 from the original 21 due to the complexity of the images in the 12 classes that are not included. This complexity was discussed earlier. The training is conducted for 200 epochs with a batch size of 50 images per step.

To observe if the model performance can be improved by appending the training set using the proposed data augmentation method, we applied the algorithm to every 270 images in the training set to produce 5 realizations of each image. Consequently, the number of images in the training set is enlarged to be 1620 (500% larger than the original). The validation ratio in both cases is kept constant at 20% of the training set and the images are chosen randomly for each epoch. Other parameters in the networks are also remained unchanged, including the test set, batch size, and the number of epochs. Similar to SAT-6, generic augmentation was also used with UCMerced dataset, to produce the same number of realizations as when using our augmentation method.

Furthermore, results of classification accuracies for all three scenarios and three networks are presented in Table 2. For this dataset, the greatest accuracy improvement from non-augmentation to our method was achieved with the Simple architecture (Fig. 5 – b) - 14%. Using the same architecture, but with the generic image augmentation, the improvement in accuracy constituted only 6.2%. We also note that using VGG-19 and ResNet-50 architectures with both data augmentation methods resulted in very substantial accuracy improvements, where our proposed method consistently outperformed generic image augmentation.

The results of training and validation accuracy evolutions for the three cases (using best performing network – ResNet-50) are presented in Fig. 7. We found, that although generic augmentation has

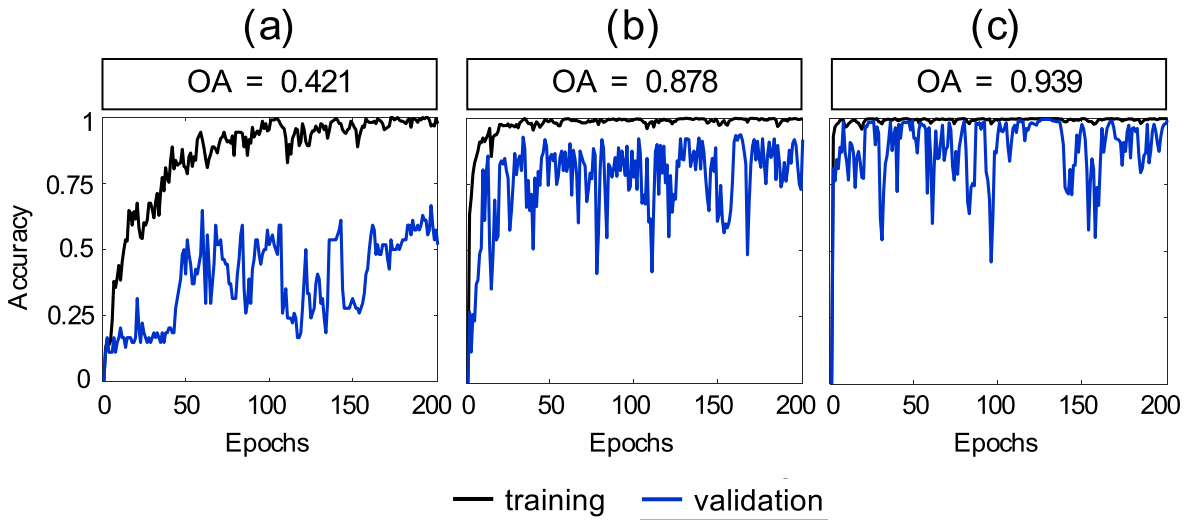

Fig. 7. Comparison between training and validation accuracy evolutions, for (a): using original images, (b) using the generically enriched dataset, and (c) using the dataset augmented by our method (UCMerced dataset).

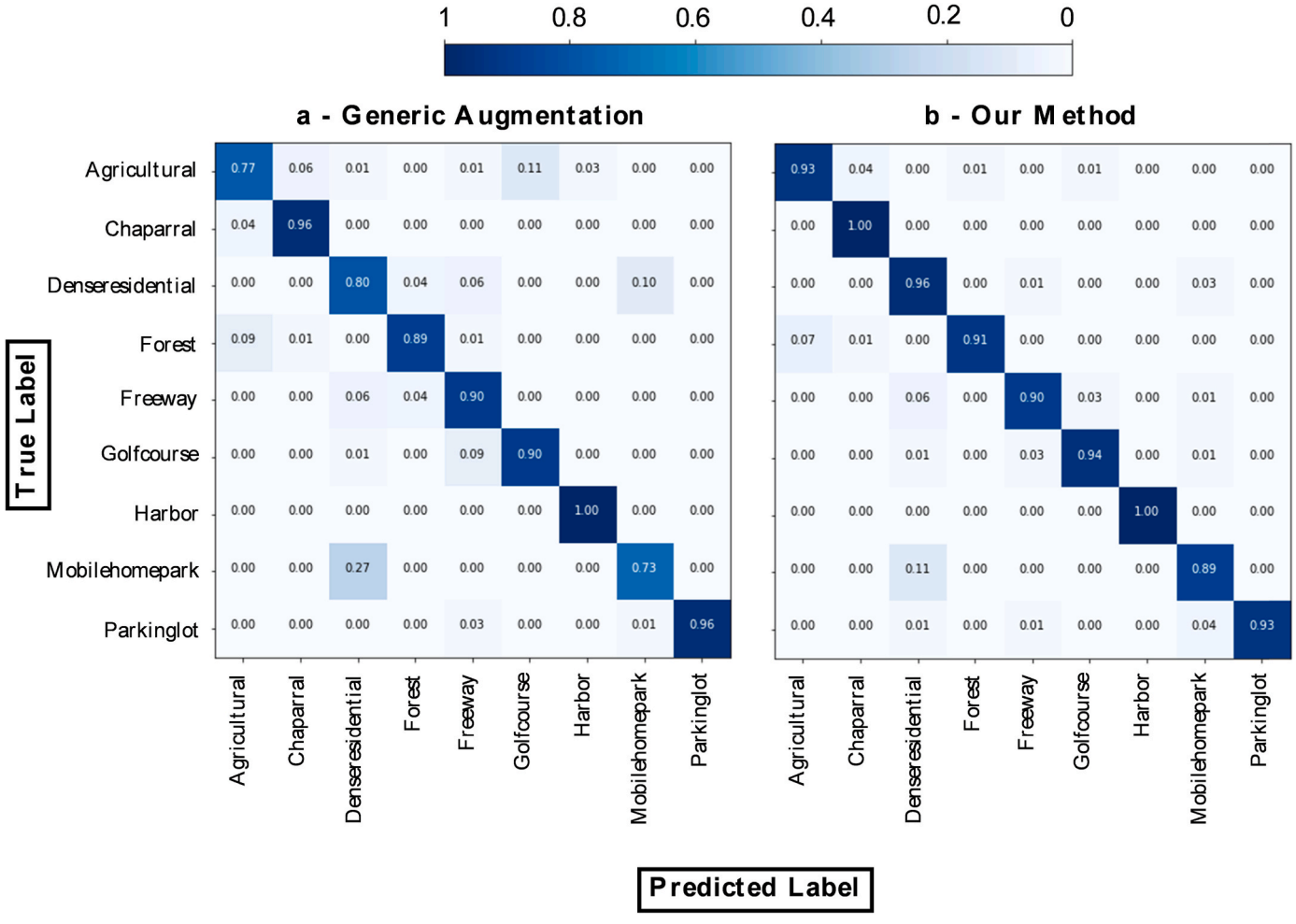

Fig. 8. Comparison between confusion matrices obtained from training on (a) dataset that included images produced from generic augmentation, and (b) dataset, that included images produced using our augmentation method.

demonstrated an improvement in terms of overfitting and accuracy evolution (Fig. 7 - b), the training and validation curves are further improved when our proposed augmentation method is used. To compare the prediction capabilities of the same network on each separate class using (a) generic augmentation and (b) our method, we include confusion matrixes in Fig. 8. These matrices are constructed from the same model results used for Fig. 7. From per-class classification accuracies included in the matrices, we find that our method outperformed generic

augmentation in six out of 9 classes, demonstrated equal performance in two, and was less effective in only one. The greatest improvement was achieved for Agriculture, Denseresidential, and Mobilehomepark classes (16% accuracy difference).

5. Conclusions

The importance of classification of remote sensing imagery is encountered in a large number of industries and applications. It can affect the predictions of phenomena that are of medical, financial, or natural concerns. It is, therefore, very important that the accuracy of the land use and land cover classification technique is as high as possible. To be able to increase the accuracy of predictions when enough data is not available, we applied a data augmentation algorithm to produce 5x additional training images that are stochastic realizations of the initial training data. Unlike the previous methods that are mostly based on implementing image processing operations (e.g., cropping, rotation, flipping), we proposed a method for producing new features by combining the existing images. This particular quality is what makes the proposed method more effective and different from the available data augmentation methods. The algorithm was tested on two popular remote sensing datasets, namely SAT-6 and UCMerced, to observe its effectiveness at different resolutions and the complexity of the images. Using four different CNN architectures in the case of SAT-6 dataset and three architectures with UCMerced dataset, we demonstrated that our augmentation approach has consistently outperformed generic image augmentation, up to 23.6% with SAT-6 and 14% with UCMerced datasets. The findings of this study show that there is a great opportunity to increase the accuracy of classification of remote sensing data using machine learning when the training data is very limited. Our results show that one can even start with a small dataset and take advantage of the recent advanced machine learning methods as more alternatives can be generated by extrapolating the existing dataset.

6. Computer code availability

The implementation of the networks used in this study with original and augmented images can be found at: https://github.com/tdavydze/Data-Augmentation. The input data is accessible with the same link, and the instructions for downloading are given in README.MD.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Albert, A., Kaur, J., Gonzalez, M.C., 2017. Using convolutional networks and satellite imagery to identify patterns in urban environments at a large scale. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. https://doi.org/10.1145/3097983.3098070.

Alhichri, H., Alswayed, A.S., Bazi, Y., Ammour, N., Alajlan, N.A., 2021. Classification of remote sensing images using EfficientNet-B3 CNN model with attention. IEEE Access 9, 14078–14094. https://doi.org/10.1109/ACCESS.2021.3051085.

Ball, J.E., Anderson, D.T., Chan, C.S., 2017. Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. J. Appl. Remote Sens. https://doi.org/10.1117/1.jrs.11.042609.

Basu, S., Ganguly, S., Mukhopadhyay, S., DiBiano, R., Karki, M., Nemani, R., 2015. DeepSat - a learning framework for satellite imagery. In: GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems. https://doi.org/10.1145/2820783.2820816.

Bosilj, P., Aptoula, E., Lefèvre, S., Kijak, E., 2016. Retrieval of remote sensing images with pattern spectra descriptors. ISPRS Int. J. Geo-Inf. https://doi.org/10.3390/ijgi5120228.

Castelluccio, M., Poggi, G., Sansone, C., Verdoliva, L., 2015. Land Use Classification in Remote Sensing Images by Convolutional Neural Networks.

Chaib, S., Liu, H., Gu, Y., Yao, H., 2017. Deep feature fusion for VHR remote sensing scene classification. IEEE Trans. Geosci. Rem. Sens. https://doi.org/10.1109/TGRS.2017.2700322.

Chen, G., Zhang, X., Tan, X., Cheng, Y., Dai, F., Zhu, K., Gong, Y., Wang, Q., 2018. Training small networks for scene classification of remote sensing images via knowledge distillation. Rem. Sens. https://doi.org/10.3390/rs10050719.

Chen, J., Wang, C., Ma, Z., Chen, Jiansheng, He, D., Ackland, S., 2018. Remote sensing scene classification based on convolutional neural networks pre-trained using attention-guided sparse filters. Rem. Sens. https://doi.org/10.3390/rs10020290.

Cheng, G., Xie, X., Han, J., Guo, L., Xia, G.S., 2020. Remote sensing image scene classification meets deep learning: challenges, methods, benchmarks, and opportunities. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. https://doi.org/10.1109/JSTARS.2020.3005403.

Chollet, F., 2013. Keras. J. Chem. Inf. Model.

de Lima, R.P., Marfurt, K., 2020. Convolutional neural network for remote-sensing scene classification: transfer learning analysis. Rem. Sens. https://doi.org/10.3390/rs12010086.

Gong, X., Xie, Z., Liu, Y., Shi, X., Zheng, Z., 2018. Deep salient feature based anti-noise transfer network for scene classification of remote sensing imagery. Rem. Sens. https://doi.org/10.3390/rs10030410.

Gu, J., Sun, X., Zhang, Y., Fu, K., Wang, L., 2019. Deep residual squeeze and excitation network for remote sensing image super-resolution. Rem. Sens. 11, 1817. https://doi.org/10.3390/rs11151817.

Gurumurthy, S., Sarvadevabhatla, R.K., Babu, R.V., 2017. DeLiGAN : Generative adversarial networks for diverse and limited data. In: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR.2017.525. CVPR 2017.

Hirahara, D., Takaya, E., Takahara, T., Ueda, T., 2020. Effects of data count and image scaling on Deep Learning training. PeerJ Comput. Sci. https://doi.org/10.7717/peerj-cs.312.

Hu, F., Xia, G.S., Hu, J., Zhang, L., 2015. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Rem. Sens. https://doi.org/10.3390/rs71114680.

Isikdogan, F., Bovik, A.C., Passalacqua, P., 2017. Surface water mapping by deep learning. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. https://doi.org/10.1109/JSTARS.2017.2735443.

Kamrava, S., Tahmasebi, P., Sahimi, M., 2019. Enhancing images of shale formations by a hybrid stochastic and deep learning algorithm. Neural Network. 118, 310–320. https://doi.org/10.1016/J.NEUNET.2019.07.009.

Kingma, D.P., Ba, J.L., 2015. Adam: a method for stochastic optimization. In: 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings. International Conference on Learning Representations. ICLR.

Larabi, M.E.A., Chaib, S., Bakhti, K., Hasni, K., Bouhlala, M.A., 2019. High-resolution optical remote sensing imagery change detection through deep transfer learning. J. Appl. Remote Sens. https://doi.org/10.1117/1.jrs.13.046512.

Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep Learning, Nature. Nature Publishing Group. https://doi.org/10.1038/nature14539.

Liu, X., Han, F., Ghazali, K.H., Mohamed, I.I., Zhao, Y., 2019. A review of convolutional neural networks in remote sensing image. In: ACM International Conference Proceeding Series. https://doi.org/10.1145/3316615.3316712.

Liu, Y., Zhong, Y., Fei, F., Zhu, Q., Qin, Q., 2018. Scene classification based on a deep random-scale stretched convolutional neural network. Rem. Sens. https://doi.org/10.3390/rs10030444.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. https://doi.org/10.1023/B:VISI.0000029664.99615.94.

Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. Convolutional neural networks for large-scale remote-sensing image classification. IEEE Trans. Geosci. Rem. Sens. https://doi.org/10.1109/TGRS.2016.2612821.

Marchesi, M., 2017. Megapixel Size Image Creation Using Generative Adversarial Networks.

Marmanis, D., Datcu, M., Esch, T., Stilla, U., 2016. Deep learning earth observation classification using ImageNet pretrained networks. Geosci. Rem. Sens. Lett. IEEE. https://doi.org/10.1109/LGRS.2015.2499239.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. https://doi.org/10.1007/s11263-015-0816-y.

Scott, G.J., Klaric, M.N., Davis, C.H., Shyu, C.R., 2011. Entropy-balanced bitmap tree for shape-based object retrieval from large-scale satellite imagery databases. IEEE Trans. Geosci. Rem. Sens. https://doi.org/10.1109/TGRS.2010.2088404.

Sebai, H., Kourgli, A., Serir, A., 2015. Dual-tree complex wavelet transform applied on color descriptors for remote-sensed images retrieval. J. Appl. Remote Sens. https://doi.org/10.1117/1.jrs.9.095994.

Shao, Z., Yang, K., Zhou, W., 2018. A benchmark dataset for performance evaluation of multi-label remote sensing image retrieval. Rem. Sens. https://doi.org/10.3390/rs10060964.

Shao, Z., Zhou, W., Zhang, L., Hou, J., 2014. Improved color texture descriptors for remote sensing image retrieval. J. Appl. Remote Sens. https://doi.org/10.1117/1.jrs.8.083584.

Shawky, O.A., Hagag, A., El-Dahshan, E.S.A., Ismail, M.A., 2020. Remote sensing image scene classification using CNN-MLP with data augmentation. Optik. https://doi.org/10.1016/j.ijleo.2020.165356.

Tahmasebi, P., Sahimi, M., 2016a. Enhancing multiple-point geostatistical modeling: 1. Graph theory and pattern adjustment. Water Resour. Res. 52, 2074–2098. https://doi.org/10.1002/2015WR017806.

Tahmasebi, P., Sahimi, M., 2016b. Enhancing multiple-point geostatistical modeling: 2. Iterative simulation and multiple distance function. Water Resour. Res. 52, 2099–2122. https://doi.org/10.1002/2015WR017807.

Tahmasebi, P., Sahimi, M., 2013. Cross-correlation function for accurate reconstruction of heterogeneous media. Phys. Rev. Lett. 110, 078002 https://doi.org/10.1103/PhysRevLett.110.078002.

Tahmasebi, P., Sahimi, M., 2012. Reconstruction of three-dimensional porous media using a single thin section. Phys. Rev. E - Stat. Nonlinear Soft Matter Phys. 85, 1–13. https://doi.org/10.1103/PhysRevE.85.066709.

Xie, M., Jean, N., Burke, M., Lobell, D., Ermon, S., 2016. Transfer learning from deep features for remote sensing and poverty mapping. In: 30th AAAI Conference on Artificial Intelligence. AAAI, 2016.

Yang, Y., Newsam, S., 2010. Bag-of-visual-words and spatial extensions for land-use classification. In: GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems. https://doi.org/10.1145/1869790.1869829.

Yu, X., Wu, X., Luo, C., Ren, P., 2017. Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework. GIScience Remote Sens. https://doi.org/10.1080/15481603.2017.1323377.

Yu, Y., Liu, F., 2018. A two-stream deep fusion framework for high-resolution aerial scene classification. Comput. Intell. Neurosci. https://doi.org/10.1155/2018/8639367.

Zeng, D., Chen, S., Chen, B., Li, S., 2018. Improving remote sensing scene classification by integrating global-context and local-object features. Rem. Sens. https://doi.org/10.3390/rs10050734.

Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., Atkinson, P.M., 2019. Joint Deep Learning for land cover and land use classification. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2018.11.014.

Zhang, F., Du, B., Zhang, L., 2016. Scene classification via a gradient boosting random convolutional network framework. IEEE Trans. Geosci. Rem. Sens. https://doi.org/10.1109/TGRS.2015.2488681.

Zhang, L., Zhang, Lefei, Du, B., 2016. Deep learning for remote sensing data: a technical tutorial on the state of the art. IEEE Geosci. Remote Sens. Mag. https://doi.org/10.1109/MGRS.2016.2540798.

Zhang, W., Tang, P., Zhao, L., 2019. Remote sensing image scene classification using CNN-CapsNet. Rem. Sens. https://doi.org/10.3390/rs11050494.

Zhong, L., Hu, L., Zhou, H., 2019. Deep learning based multi-temporal crop classification. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2018.11.032.

Zhu, Q., Zhong, Y., Zhao, B., Xia, G.S., Zhang, L., 2016. Bag-of-Visual-Words scene classifier with local and global features for high spatial resolution remote sensing imagery. Geosci. Rem. Sens. Lett. IEEE. https://doi.org/10.1109/LGRS.2015.2513443.

Zou, J., Li, W., Chen, C., Du, Q., 2016. Scene classification using local and global features with collaborative representation fusion. Inf. Sci. https://doi.org/10.1016/j.ins.2016.02.021.