Deep Curriculum Learning for PolSAR Image Classification

Hamidreza Mousavi, Maryam Imani, Hassan Ghassemian

Image processing and Information Analysis Lab
Faculty of Electrical and Computer Engineering, Tehran, Iran
{h.mousavi, maryam.imani, ghassemi}@modares.ac.ir

Abstract—Following the great success of curriculum learning in the area of machine learning, a novel deep curriculum learning method proposed in this paper, entitled DCL, particularly for the classification of fully polarimetric synthetic aperture radar (PolSAR) data. This method utilizes the entropy-alpha target decomposition method to estimate the degree of complexity of each PolSAR image patch before applying it to the convolutional neural network (CNN). Also, an accumulative mini-batch pacing function is used to introduce more difficult patches to CNN. Experiments on the widely used data set of AIRSAR Flevoland reveal that the proposed curriculum learning method can not only increase classification accuracy but also lead to faster training convergence.

Index Terms—Polarimetric SAR image classification, Curriculum learning, Deep learning, Convolutional neural network.

I. INTRODUCTION

Deep neural networks have been extensively employed in the area of PolSAR image classification. Specially convolutional neural networks (CNNs) have proved to be a useful automatic feature extraction tool for PolSAR image classification in recent years [1]. The main focus in this area of research is on developing practical CNN architectures in terms of classification accuracy and model complexity. For instance, Zhang et al. [2] extended a real-valued CNN (RV-CNN) in a complex domain to exploit the potential of phase information in PolSAR data and their proposed complex-valued CNN (CV-CNN) could remarkably increase the classification accuracy over RV-CNN. The first architecture utilizing 3D convolution and 3D pooling layers for PolSAR image classification was introduced by Zhang et al. [3] and achieved a classification accuracy better than conventional 2D CNNs. Lightweight 3D convolutions introduced by Dong et al. [4] to reduce the redundancy of 3D convolutions and take advantage of their excellent ability in spatial-polarimetry feature extraction. Considerable attention has been focused on the CNN architecture evolution over the last few years for increasing the feature extraction ability. In contrast, to reduce the number of model parameters, prevent from the overfitting, and even to eliminate the fully connected layers, less attention has been devoted to improving the training process. Since neural network architectures are inspired by the human brain, it seems reasonable that studying the learning process should also be inspired by the biological processes of the human brain. Starting from easy tasks before progressing to high and complicated levels generally provides better learning results for Humans. Using a similar approach for training a machine learning model, we can obtain great benefits including faster convergence speed of the training process and better accuracy. This strategy has been formalized by Bengio et al. [5] and called curriculum learning in the context of machine learning. Curriculum learning approaches have been successfully applied in various fields of machine learning [6]. Each curriculum learning strategy needs a proper complexity measurement criterion of training samples and a pacing function for progressively introducing more complex samples to the machine learning model [5], [6]. To our knowledge, there isn’t a criterion to measure the difficulty of PolSAR patches before applying them to a deep learning model. This motivates us to formalize this idea. PolSAR data analysis is rich in accurate pixel information extraction thanks to the publication of a large number of proposed methods in the target decomposition theory. The Cloude-Pottier target decomposition method proposed parameters such as entropy (H), mean alpha angle (\(\bar{\alpha}\)), and anisotropy (AA), which provide the physical interpretation of the scattering mechanisms of objects. The most attractive and practical application of the H/\(\bar{\alpha}\) method is an unsupervised classification of PolSAR data by subdividing the H/\(\bar{\alpha}\) plane into nine zones correspond to various physical scattering characteristics. In this study, we use the H/\(\bar{\alpha}\) plane to introduce a novel patch complexity criterion (PaCC) for PolSAR patches. The basic idea is that each patch is considered more difficult for neural network to learn if it consists of pixels with a more complex scattering process. Therefore, two criteria are needed to first measure the complexity of each pixel in the patch and then determine the complexity of the patch with the obtained values for pixels. Two parameters of entropy H and mean alpha angle \(\bar{\alpha}\), which yield direct information about the scattering mechanism are used to measure the scattering complexity of pixels and a criterion named pixel complexity criterion (PiCC) is defined. Finally, by calculating the average of the obtained PiCC values, the PaCC parameter is determined. The rest of this paper is structured as follows. Section II presents the methodology. Experimental results are shown in Section III, which illustrate the effectiveness of the proposed method. Finally, we draw some conclusions in Section IV.
II. METHODOLOGY

A. Deep learning

Deep learning (DL) attempts to develop suitable models by mimicking the structure and functioning of the human brain in processing data. From the viewpoint of automation, the evolution of DL has transformed the handcrafted design of features in machine learning to facilitate automatic feature extraction from high-dimensional raw data. This remarkable ability of high-dimensional data processing and automatic feature extraction has made breakthroughs in pattern recognition. Most recently, the huge success of DL in the computer vision area has motivated the remote-sensing community to this technology and DL algorithms bring huge advancements in many remote sensing image interpretation tasks including Image fusion, scene classification, object detection, surface terrain classification, and other tasks. Convolutional neural networks (CNN) is a specific type of DL model which is particularly useful for recognizing visual patterns in images. It is inspired by the structure of the human and animal visual system and designed to automatically extract low, mid, and high-level features from multi-dimensional data. Because of this characteristic, it is applicable for processing multi-channel remote sensing images.

CNN has proven to have an exemplary performance to fill the gaps in conventional machine learning algorithms. Inspired by the lightweight operations for CNNs in the PolSAR image classification field, in this paper, pseudo-3D(P3D) CNN introduced by [4] is used in the experiments to investigate the performance of the proposed curriculum learning algorithm. Lightweight P3D CNN was introduced to not only perform feature learning in both the spatial and polarimetric dimensions but also reduce network parameters and high computational costs of 3D CNNs [4], [15].

B. Curriculum Learning

Curriculum Learning (CL) can be considered as an optimization technique, which can improve optimization performance by discovering better relative minima of a non-convex cost function [5], [7]. It is inspired based on imitating the learning behaviour of humans when trying to learn concepts. They can master a task when try to learn easy parts of a task and gradually increase complexity level. Using a similar approach for training a machine learning model, we can obtain great benefits including faster convergence speed of the training process and better accuracy. Since that samples are considered in random order during training of machine learning model, the hard or noisy samples, especially in the initial stage of the training process may result in over-fitting the train set and leading to poor generalization. This learning strategy has been empirically proved to be robust to noisy samples and be effective in avoiding getting stuck in poor local minima [14]. PolSAR images also are heavily affected by speckle noise. This noise diminishes the visual quality of PolSAR images for interpretation and limits image processing algorithms during information extraction from PolSAR data. Hence, CL can effectively help machine learning algorithms to suppress the effect of noisy PolSAR samples during the training process and achieve more reliable results. However, the implementation of CL in machine learning is subject to finding a proper computational complexity measurement criterion of training samples and a pacing function for progressively introducing more complex samples to the machine learning model. This is the principal motivation behind the design and implementation of our CL framework to improve the classification performance of SAR polarimetry data.

The complexity of the scattering process complicates the direct analysis of coherency matrix T. Hence, the ultimate purpose of the incoherent decompositions tools is to simplify the interpretation of the scattering process by expanding the second-order matrices into a weighted addition of basis matrices describing scattering responses of simpler objects. Two important average parameters (the entropy H and mean angle $\bar{\alpha}$) were introduced by Cloude and Pottier [7] using the eigenvalue/eigenvector analysis of T. Entropy H is defined as a function of the eigenvalues:

$$H = -\sum_{i=1}^{3} p_i \log_3(p_i), \quad p_i = \frac{\lambda_i}{\sum_{j=1}^{3} \lambda_j} \quad (1)$$

Entropy H can go above 0 and below 1. $H = 0$ corresponds to only one nonzero eigenvalue in T. It physically indicates that there is only a single scattering mechanism caused by the surface scattering. When $H$ equals to 1, T has three equal nonzero eigenvalues, which statically means three scattering mechanisms occur simultaneously. From a physical standpoint, this shows that the behaviour of scatter is completely random and this mechanism occurs over the forest and vegetated surfaces. As $H$ increases from 0 to 1, the randomness of the scattering mechanism increases, which means the ability to distinguish between scattering mechanisms decreases [8], [9]. In short, the complexity of every pixel in the PolSAR image is proportional to entropy H.

Mean alpha angle $\bar{\alpha}$ is obtained from parameterizing of the eigenvectors of T versus some of the angle parameters which $\alpha$ is one of them and given by

$$\bar{\alpha} = \sum_{i=1}^{3} p_i \alpha_i \quad (2)$$

The mean alpha angle $\bar{\alpha}$ indicates the type of dominant scattering mechanism and lies within the range of $0^\circ$ to $90^\circ$. $\bar{\alpha} = 0^\circ$ corresponds to the surface scattering caused by water body or bare soil, $\bar{\alpha} = 45^\circ$ to the volumetric scattering arisen from woodlands, and $\bar{\alpha} = 90^\circ$ to double-bounce scattering as a result of interaction between the incident wave and street/building classes in urban areas. By continuously increasing $\bar{\alpha}$ from 0 to 90, the scattering mechanism changes from isotropic simple scattering($\bar{\alpha} = 0$) to random volume scattering ($\bar{\alpha} = 60$) and finally $\bar{\alpha} = 90$, is related to double bounce scattering [7], [8]. For the measurement of complexity based on the alpha angle, if we consider complexity range from 0 to 1, by increasing $\bar{\alpha}$ from 0 to 60, complexity
increases to 1 and by changing alpha from 60 to 90, the complexity decreases from the maximum value to the median one, which equals to 0.5. Based on the above analysis, the PaCC parameter is proposed to measure the complexity of each PolSAR image patch by averaging the complexity of whole pixels exist in that patch. The complexity of each pixel is calculated by using Euclidean distance in the two-dimensional (2D) \( H/\bar{\alpha} \) plane. The details of the patch-ranking algorithm are shown in Algorithm 1 and the details of the method are expressed as follows.

First, we introduce a metric, called pixel complexity criterion (PiCC) to measure the complexity of every pixel in a patch using Euclidean distance as follows:

\[
P_{iCC}(v(i)) = \sqrt{H_i^2 + (1 - \frac{\bar{\alpha}_i - 60}{60})^2} \quad i = 1, 2, ..., m_1m_2
\]  

(3)

Where \( v(i) \) is i-th pixel in a PolSAR patch, which defined as a vector as shown in Equation (4):

\[
v = \{t_{11}, t_{22}, t_{33}, \Re \{t_{12}\}, \Im \{t_{12}\}, \Re \{t_{13}\}, \Im \{t_{13}\}, \Re \{t_{23}\}, \Im \{t_{23}\}\}
\]  

(4)

In equation (3), \( H_i \) and \( \bar{\alpha}_i \) are the entropy and mean alpha angle of pixel \( v(i) \), which exist in a PolSAR image patch with \( m_1 \times m_2 \times 9 \) size. Finally, by averaging on all the calculated PiCCs for the patch, the patch complexity criterion (PaCC) for arbitrary PolSAR image patch \( X \in \mathbb{R}^{m_1 \times m_2 \times 9} \) is defined by the following formula:

\[
PaCC(X) = \frac{1}{m_1m_2} \sum_{i=1}^{m_1m_2} P_{iCC}(v(i)) = \frac{1}{m_1m_2} \sum_{i=1}^{m_1m_2} \sqrt{H_i^2 + (1 - \frac{\bar{\alpha}_i - 60}{60})^2}
\]  

(5)

Any patch with a higher PaCC value is considered a more difficult patch for training. Therefore, after calculating the PaCC value for all the patches in the training set \( L \), it is necessary to arrange the PaCC values in ascending order. Finally, the training samples are reordered to form the sorted training set \( L^1 \). The \( \text{argsort} \) operator in the eighth line of Algorithm 1 returns indices of the training samples after sorting PaCC values of them in an ascending order.

Another key part of any curriculum learning method is task scheduling, which specifies how to gradually increase the complexity of tasks and update the training process. Several scheduling strategies have been proposed, where batching, sampling and weighing are the most extensively employed ones [6]. In this study, we introduce a new batching strategy, called batch accumulation scheduler function. The main idea is that in each iteration after random sampling, we sort the training samples from easy to hard. Then, the sorted training set is split into \( n \) same size batches. The fine-tuning process starts on the easiest batch (\( k = 1 \)), the second batch is concatenated with the previous one and applied to the deep learning model for updating (\( k = 2 \)). This process continues until all mini-batches are concatenated together. Eventually, the deep learning model is updated with all training samples existing in the training set (\( k = n \)). In step \( k \) for gradient updating, the sliced batch from the sorted training set \( L^1 \) is sliced as follows:

\[
\mathbb{B}_k = \{(X_i, y_i)\}_{i=1}^{m \times n}, m = k \frac{N}{n}, k = 1, 2, ..., n
\]  

(6)

where \( \mathbb{B}_k \) and \( m \) are sliced batch and batch size at \( k \)-th step, respectively. This pacing function is also inspired by curricula taught in education systems around the world that usually enables learning concepts in progressively increasing levels of complexity while considering and reviewing previously learned concepts.

### C. Deep Curriculum Learning (DCL)

Algorithm 2 presents the pseudo-code of the DCL method. First, the weights of P3D CNN initialized by random values. In the iterative phase, \( N_t \) samples randomly chose from all labeled training set \( U \) and these samples are added to the training set \( L \). Training samples in the training set sorted from easy to hard based on Algorithm 1. Then training set divided into \( n \) batches and we fine-tune P3D CNN by incrementally accumulating batches.

### III. EXPERIMENTAL RESULTS

In this section, to analyze the performance of the algorithms mentioned in this paper, we conduct experiments on the publicly available PolSAR data set, which is collected by the AIRSAR sensor over the Flevoland area. The overall accuracy (OA) and runtime are also used to evaluation of classification performance.

#### A. Experimental data and parameter setting

The AIRSAR data is a four-look full polarized image acquired by NASA/JPL using L-band PolSAR sensor was mounted on a NASA DC-8 aircraft on 16 August 1989. It reflects an agricultural area of Flevoland in the Netherlands and it is widely used as a benchmark data for PolSAR image...
Algorithm 2: Deep Curriculum learning (DCL) for PolSAR Image Classification

Input:
\( U \): All labeled patch set.
\( N_t \): The number of patches to be selected in each stage of DCL, where \( t = 0, 1, \ldots, T \), \( T \) is number of stages.
\( n \): Training set splitting parameter

Output: Trained P3D CNN model.

Initialization:
Initalize weights of P3D CNN by random values.
\( S_0 = {} \).
\( L = {} \).

1. for \( t \leftarrow 0 \) to \( T \) do
2. Select \( N_t \) patches randomly from \( U \) and build \( S_t \).
3. Remove Selected patched from \( U \), i.e, \( U \leftarrow U \setminus S_t \).
4. \( L \leftarrow L \cup S_t \).
5. Sort patches in \( L \) using Algorithm 1 to obtain sorted patch set \( L^* \).
6. for \( k \leftarrow 1 \) to \( n \) do
7. Slice Batch \( B_k \) by Eq. (6).
8. P3D CNN training (\( c = 1 \)) or fine-tuning (\( c > 1 \)) using batch \( B_k \).
9. return Trained P3D CNN model.

Fig. 1. AIRSAR Floevoland data set adopted in the experiments.

Therefore, at the end of the training process, a total of 3000 training samples are used. In the proposed curriculum, two hyper-parameters, including splitting parameter \( n \), and fine-tuning iterations need to be chosen. The mentioned two hyper-parameters experimentally are set to 25 and 5, respectively.

IV. CONCLUSION

In this paper, a novel deep curriculum learning method called DCL has been designed for PolSAR image classification. Besides, a patch complexity criterion, which is called classification. The size of this image is 1024 × 750 pixels and contains 15 categories, which are building, water, forest, bare soil, lucerne, beet, potatoes, grass, rapeseed, barley, three different types wheat, steam beans, peas. The RGB image formed by the Pauli decomposition and the corresponding ground truth image, containing a total of 157296 samples and legend of this data set are shown in Fig. 1. Each colour denotes one category and white region stands for unknown land covers. multi-look full polarized image We set the experimental settings in this work as follows: The lightweight P3D CNN is implemented by Keras with the TensorFlow backend. During the CL cycle, each data set is randomly divided into three sets. We randomly select 100 labeled samples in each iteration \( (N_t = 100) \). To evaluate a model fit while tuning network hyper-parameters, 10000 samples are selected as a validation set and the remaining samples are used for further random sampling. We conduct 30 iterations of network training for the experiments. In each iteration, we join 100 randomly selected training samples to the original training samples to fine-tune the network using our proposed curriculum learning strategy.

From Table I, It is shown that the proposed DCL has higher classification accuracy and lower runtime, which illustrates the strength of the curriculum learning approach and its positive impact on promoting the CNNs. Finally, for comparative purposes, Fig. 3 (a) and (b) show the classification maps by curriculum learning and no-curriculum learning, respectively. It can be observed that the result obtained by curriculum learning is remarkable, which is better than that obtained by no curriculum learning.

B. Classification Results

To completely study the influence of curriculum learning on the training process, two groups of experiments are conducted, which are training P3D CNN using curriculum and no-curriculum learning. We have monitored the values of OA during training according to the number of training samples as shown in Fig. 2. The main observations from this figure are as follows:

(a) The OA values with the curriculum, except in a few initial steps are higher than without curriculum. This reflects that curriculum learning provides better generalization to the unseen examples.

(b) With attention to the zoomed area in Fig. 2, the OA value curve for a curriculum learning experiment is smoother than the no-curriculum experiment, which shows that the output results of curriculum-based learning are more reliable for PolSAR data classification.

The obtained OA and running time are given in Table I. From Table I, It is shown that the proposed DCL has higher classification accuracy and lower runtime, which illustrates the strength of the curriculum learning approach and its positive impact on promoting the CNNs. Finally, for comparative purposes, Fig. 3 (a) and (b) show the classification maps by curriculum learning and no-curriculum learning, respectively. It can be observed that the result obtained by curriculum learning is remarkable, which is better than that obtained by no curriculum learning.

| Method       | OA (%) | run time (sec) |
|--------------|--------|----------------|
| curriculum   | 97.64  | 623            |
| no-curriculum| 96.42  | 818            |

TABLE I: OA VALUES (%) AND RUN TIMES (SEC) WITH CURRICULUM AND NO-CURRICULUM STRATEGY ON FLEVOLAND AREA DATA SET.
Fig. 2. The obtained overall accuracies (OAs, averaging from 10 independent runs) using two different strategies for the AIRSAR Flevoland data set. In each step, the original training samples are joined with 100 randomly selected training samples. Therefore, at the end of the training process, a total of 3000 training samples are used.

Fig. 3. Classification results of DCL on AIRSAR Flevoland data set. (a) Result of whole classification map using curriculum learning. (b) Result of whole classification map using no curriculum learning.

PuCC, is introduced to estimate the learning difficulty of the PolSAR patches before applying them to a CNN. The experiments are conducted on benchmark AIRSAR Flevoland data set and the experimental results show that the DCL is fast and can achieve better accuracy than the methods which consider samples in a random order during training.

REFERENCES

[1] Zhu, Xiao Xiang, Devis Tuia, Lichao Mou, Gui-Song Xia, Liangpei Zhang, Feng Xu, and Friedrich Fraundorfer. "Deep learning in remote sensing: A comprehensive review and list of resources." IEEE Geoscience and Remote Sensing Magazine 5, no. 4 (2017): 8-36.

[2] Zhang, Zhimian, Haipeng Wang, Feng Xu, and Ya-Qiu Jin. "Complex-valued convolutional neural network and its application in polarimetric SAR image classification." IEEE Transactions on Geoscience and Remote Sensing 55, no. 12 (2017): 7177-7188.

[3] Zhang, Lamei, Zexi Chen, Bin Zou, and Ye Gao. "Polarimetric SAR terrain classification using 3D convolutional neural network." In IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 4551-4554. IEEE, 2018.

[4] Dong, Hongwei, Lamei Zhang, and Bin Zou. "PolSAR image classification with lightweight 3D convolutional networks." Remote Sensing 12, no. 3 (2020): 396.

[5] Bengio, Yoshua, Jérôme Louradour, Ronan Collobert, and Jason Weston. "Curriculum learning." In Proceedings of the 26th annual international conference on machine learning, pp. 41-48. 2009.

[6] Soviany, Petru, Radu Tudor Ionescu, Paolo Rota, and Nicu Sebe. "Curriculum learning: A survey." arXiv preprint arXiv:2101.10382 (2021).

[7] Ren, Zhongle, Biao Hou, Zaidao Wen, and Licheng Jiao. "Patch-sorted deep feature learning for high resolution SAR image classification." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11, no. 9 (2018): 3113-3126.

[8] Clande, Shane R., and Eric Pottier. "An entropy based classification scheme for land applications of polarimetric SAR." IEEE transactions on geoscience and remote sensing 35, no. 1 (1997): 68-78.

[9] Lee, Jong-Sen, and Eric Pottier. Polarimetric radar imaging: from basics to applications. CRC press, 2017.

[10] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." nature 521, no. 7553 (2015): 436-444.

[11] Sajjad Eghbalian and Hassan Ghassemian, "Multi Spectral Image Fusion by deep convolutional neural network and new Spectral loss function", International Journal of Remote Sensing, vol. 39, no. 12, pp. 3983–4002, 2018.

[12] Adelipour, Sadjad, and Hassan Ghassemian. "Building Detection in Very High Resolution SAR Images Via Sparse Representation Over Learned Dictionaries." IEEE Journal of Selected Topics in Applied Earth Observations, 11, no. 12 (2018): 4808-4817.

[13] Jiang, Bitao, Xiaobin Li, Lu Yin, Wenzhen Yue, and Shengjin Wang. "Object recognition in remote sensing images using combined deep features." In 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), pp. 606-610. IEEE, 2019.

[14] Braun, Stefan, Daniel Neil, and Shih-Chii Liu. "A curriculum learning method for improved noise robustness in automatic speech recognition." In 2017 25th European Signal Processing Conference (EUSIPCO), pp. 548-552. IEEE, 2017.

[15] Imani, Maryam, and Hassan Ghassemian. "An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges." Information fusion 59 (2020): 59-83.

[16] Adelipour, Sadjad, and Hassan Ghassemian. "Building extraction from very high-resolution synthetic aperture radar images based on statistical and structural information fusion." International Journal of Remote Sensing 40, no. 18 (2019): 7113-7126.

[17] Maggiori, Emmanuel, Yuliya Tarabalka, Guillaume Charpiaf, and Pierre Alliez. "Convolutional neural networks for large-scale remote-sensing image classification." IEEE Transactions on Geoscience and Remote Sensing 55, no. 2 (2016): 645-657.

[18] Marmanis, Dimitrios, Mihai Datcu, Thomas Esch, and Uwe Stilla. "Deep learning earth observation classification using ImageNet pretrained networks." IEEE Geoscience and Remote Sensing Letters 13, no. 1 (2015): 105-109.