Jurnal Ilmu Komputer dan Informasi (Journal of Computer Science and Information)
12/2 (2019), 103-112. DOI:_http://dx.doi.org/10.21609/jiki.v12i2.752

ANALYZING DEPTHWISE CONVOLUTION BASED NEURAL NETWORK: STUDY CASE IN SHIP DETECTION AND LAND COVER CLASSIFICATION

Kuntoro Adi Nugroho[1] and Yudi Eko Windarto[2]

Department of Computer Engineering, Diponegoro University
Jl.Prof.H.Soedarto S.H, Semarang, 50275, Indonesia

Email: kuntoro@live.undip.ac.id [1] *yudi@live.undip.ac.id* [2]

Abstract

Various methods are available to perform feature extraction on satellite image. Among the available alternatives, deep convolutional neural network (ConvNet) is the state of the art method. Although previous studies have reported successful attempts on developing and implementing ConvNet on remote sensing application, several issues are not well explored, such as the use of depthwise convolution, final pooling layer size, and comparison between grayscale and Red Green Blue (RGB) settings. The objective of this study is to perform analysis to address these issues. Two feature learning algorithms were proposed, namely ConvNet which represents the current state of the art for satellite image classification and Gray Level Co-occurence Matrix (GLCM) which represents a classic unsupervised feature extraction method. The experiment demonstrated consistent result with previous studies that ConvNet is superior in most cases compared to GLCM, especially with 3x3xn final pooling. The performance of the learning algorithms are much higher on features from RGB channels, except for ConvNet with relatively small number of features.

Keywords: *GLCM, Convolutional Neural Network, Satellite Image*

Abstrak

Banyak metode yang dapat digunakan untuk melakukan ekstraksi ciri pada citra satelit. Diantara banyaknya alternatif, *deep convolutional neural network (ConvNet)* adalah metode terbaru yang paling efektif. Meskipun telah banyak penelitian yang sukses dalam mengembangkan dan mengimplementasikan metode *ConvNet* untuk citra satelit, banyak hal yang belum dieksplorasi seperti *depthwise convolution*, ukuran lapisan *pooling* akhir, dan perbandingan aras keabuan dan *Red Green Blue (RGB)*. Tujuan dari penelitian ini adalah untuk melakukan analisis mengenai hal-hal tersebut. Dua metode yang digunakan dalam eksperimen adalah *ConvNet* sebagai metode handal berdasarkan penelitian sebelumnya dan *Gray Level Co-occurence Matrix (GLCM)* sebagai metode klasik ekstraksi fitur tanpa supervisi. Hasil penelitian menunjukkan konsistensi dengan penelitian sebelumnya, bahwa *ConvNet* unggul dibandingkan *GLCM* pada banyak parameter, terutama *ConvNet* dengan *pooling* berukuran 3x3xn. Peningkatan performa diperoleh cukup tinggi dengan *RGB*, kecuali pada *ConvNet* dengan jumlah fitur yang relatif lebih kecil.

Kata Kunci: *GLCM, Convolutional Neural Network, Citra Satelit*

1. Introduction

The use of earth surface photograph which obtained via satellite serves a wide range of applications. For instance, in meteorology, satellite image provide useful information for analyzing cloud cover [1]. In oceanography, some examples are coastal hazard and sea surface temperature estimation [2].

Among other examples are ship detection and land-use recognition.

In land usage identification, many feature extraction methods are available. The first example is dictionary learning with mutual incoherence K-Singular Value Decomposition [3]. Secondly, texture feature extractors serve as a useful predictor as shown by several studies. An experiment to compare Gray Level Co-ocurrence Matrix (GLCM) and other texture feature extractions was performed on inhabited region identification. The study shows that GLCM is comparable to Gabor and wavelet with compact feature vector [4]. GLCM combined with object-based classification was proposed to analyze TerraSAR-X satellite images and superior to the texture followed by pixel based classification [5].

Several studies on satellite image ship detection also demonstrated that texture feature provides useful information. Incorporation of gray level non-uniformity as a result of feature selection was proposed on the first stage of small ship classification [6]. Texture based ship representation using GLCM was used post fuzzy c-means based segmentation for classification [7].

Although low level representation such as texture is useful in practice, efforts have been carried to lower the gap between low and high level representation. An example of successful attempt is object detectors based on histogram of oriented gradient, which succesfully outperformed other methods [8].

Besides object detectors, a method that systematically learn from low to high level representation is deep convolutional neural network. In deep neural network, the first layer learn simple low level representation of the image. The following layer incorporate information from the previous layer to learn higher level feature representation.

Deep convolutional neural networks (ConvNet) have been studied for many applications on satellite image analysis. Evaluated on two remote sensing land use datasets, a study confirmed that fine-tuned GoogLeNet outperformed CaffeNet and other learning algorithms [9]. Other study also confirmed that ConvNet outperformed other methods such as Spatial Pyramid Matching Kernel (SPMK) , Sparse Coding, and Bag of Visual Words (BoVW) [10]. ConvNet was proposed to identify terrains and structures which is useful for poverty mapping [11]. In synthetic aperture radar (SAR) based maritime target detection, ConvNet is useful for land masking [12] and object detection (such as cargo, harbor, and tanker) [13].

Most previous studies presented convolutional neural network as a robust methods for segmenting and classifying satellite image. Despite of the success, certain issues have not been addressed.

Firstly, there are many methods which performance have not been reported. For example, despite [14] and [9] discussed popular architecture such as Xception, DenseNet, and ResNet, other network has not been studied. One example is MobilNet, an architecture which utilized depthwise separable convolution to improve computation efficiency [15]. Other example is Gray Level Co-occurence Matrix, which performance has not been discussed such as in [9] and [10].

Secondly, although reducing the feature into 1x1xn-channels before classification layer is an option for ConvNet implementation, the impact of maintaining some spatial resolution before classification layer is still unknown.

Finally, learning on multiband / multichannel image resulted on better model performance in most cases. However, previous studies have not been specifically discussed how learning on multichannel image affect model performance compared to single grayscale image.

The objective of this study is to perform experiments and analysis to address these issues. This paper is organized as follows. Section 1 presented background and objective. The methodology is explained in Section 2. Next, the experiment results are presented and discussed in Section 3. Finally, the conclusion is mentioned in Section 4.

2. Methodology

2.1. Dataset

Two datasets were used for evaluation. The problem proposed by the first dataset is about recognizing an object, while the second dataset address a more general image classification of earth surface photograph.

2.1.1. Ship Detection. The task on ship detection is to detect the presence of ship on an image patch. The photos were taken from Planet Open California satellite imagery, depicting area of San Fransisco Bay and San Pedro Bay. The dataset is available via Kaggle dataset repository. With PlanetScope visual scene, the image spatial resolution is 3 meters [16] [17].

The images are classified into positive and negative samples. The first 1000 images, identified by its ID, are ship images. The rest 3000 samples are negative class images which divided equivalently into (1) landcover (such as building and water), (2) partially captured ship, and (3) previously misclassified instance by machine learning algorithms. A few samples are shown in Figure 1.

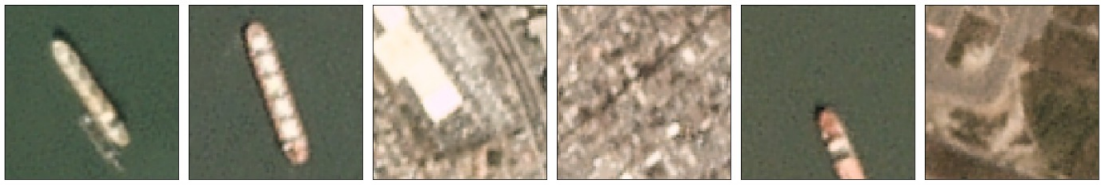

Fig. 1: Several samples from Ship Detection Dataset

2.1.2. EuroSAT Land Cover Recognition. The second task is to classify land cover given its photograph. The images were taken by Sentinel-2A satellite from EU Copernicus Programme with spatial resolution up to 10 meters. Two versions are available, RGB and multispectral. In this study we only utilized the RGB version. The ten categories of area are summarized in Table 1. Figure 2 [18] shows one sample for each class.

TABLE 1
EUROSAT DATASET SUMMARY

Index	Label	Quantity
0	Permanent crop	2500
1	Sea lake	3000
2	Highway	2500
3	Residential	3000
4	Annual crop	3000
5	Industrial	2500
6	River	2500
7	Herbaceous vegetation	3000
8	Forest	3000
9	Pasture	2000

2.2. Experiment

2.2.1. Research flow. The experiment began with randomly sampling the dataset into three subsets: training, validation, and testing. The distribution of sample for both dataset are summarized in Table 2. After splitting the dataset, the process is continued with feature extraction and image classification. The images were evaluated in grayscale and RGB.

TABLE 2
INSTANCE DISTRIBUTION

Dataset	Training	Validation	Testing
Ship	1333	1333	1334
EuroSAT	9000	9000	9000

The feaures from each image were extracted with three methods. The first two are convolutional neural networks (ConvNet-1 and ConvNet-2). The last is GLCM.

The ConvNets used the training subset for training and the validation subset to validate the model. The weights were obtained by learning only from the datasets without any pre-training process. Models with the lowest validation error were selected for feature extraction purpose. No data augmentation performed on the training and validation process. In case of GLCM, the training, validation, and testing subsets were directly processed because GLCM does not require any supervised training process.

After the features had been extracted, classifiers were trained to evaluate each feature extractor. The features were first normalized with mean normalization before applying learning algorithm as shown in Equation 1. The normalization was performed feature-wise. The $\epsilon = 10^{-30}$ was added to avoid division by zero.

$$x_{i,j} = \frac{x_{i,j} - mean(x_j)}{stdev(x_j) + \epsilon} \tag{1}$$

The training and validation subsets were joined to train a linear support vector machine (SVM) model. Then, the SVM model was evaluated on the testing subset.

Several metrics were evaluated, namely acuracy, precision, recall, and F1-score. Besides performance evaluation, principal component analysis (PCA) one the feature was also performed to visualize the test results. Besides that, the principal component histograms were observed.

Both ConvNet-1 and ConvNet-2 used the same architecture. Their difference is on the final pooling layer. The input image batch is first processed by four convolutional blocks. After convolution, the process continued with adaptive pooling and flattening to obtain feature vector. The feature then used to predict class label by the fully connected layer. The architecture is illustrated in Figure 3. The networks were implemented using PyTorch deep learning library [19].

Among the four convolutional blocks, only Block 1 is different, as shown in Figure 4. Block 1 begin with convolution layer followed by ReLU activation and max-pooling layer. Block 2, 3, and 4

Fig. 2: Several samples from EuroSAT Dataset

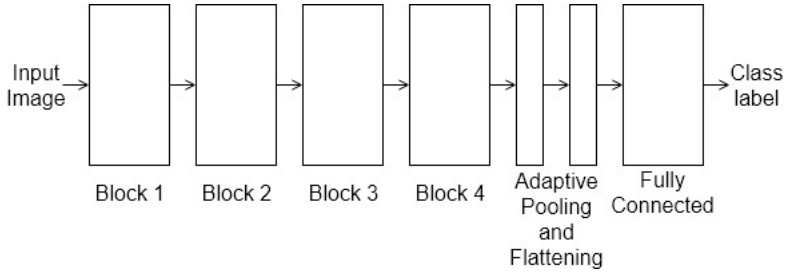

Fig. 3: Convolutional Network Architecture

are identical. These blocks consists of two convolutional layers followed by ReLU activation, max-pooling, and dropout layer. The first convolution is a grouped convolution, which also called depthwise convolution. The number of group is equal to the number of input channel. The second convolution is a 1x1 convolution which applied to the entire channel (non-grouped). The scheme utilized in Block 2, 3, and 4 is inspired by MobileNet [15]. The difference among Block 2, 3, and 4 is merely the number of convolutional filter.

The parameter used for adaptive maximum pooling is the only factor that contrasts ConvNet-1 and ConvNet-2. Adaptive pooling on ConvNet-1 reduces the n-channel output from Block 4 into 3 x 3 x n tensor. In contrast to the first, adaptive pooling on ConvNet-2 is computed entirely per channel, which resulted into 1 x 1 x n tensor. Consequently, ConvNet-1 still retains some spatial location information of the feature (in 3x3 size) while ConvNet-2 does not. Besides, the former has nine times more features than the later.

Different parameter settings were applied for Ship recognition and EuroSAT dataset. Nevertheless, some parameters are identical accross accross networks in this experiment. Dropout probability is set to 10 %. The maximum pooling size is set to 2 x 2. The detail of the parameters are shown in Table 3.

The third feature extraction method is gray level co-occurence matrix (GLCM). GLCM was selected because according to previously discussed studies, texture is a reliable predictor and GLCM was one among the presented texture extractors. GLCM works by constructing co-occurence matrix which values represent spatial relationship among pixel values. Several features could be obtained from the co-occurence matrix [20].

For this experiment, six features were selected, namely contrast, dissimilarity, homogeneity (inverse difference moment), energy, correlation, and homogenity (angular second moment).

GLCM has several parameters that must be set on the algorithm. The parameter of GLCM was set

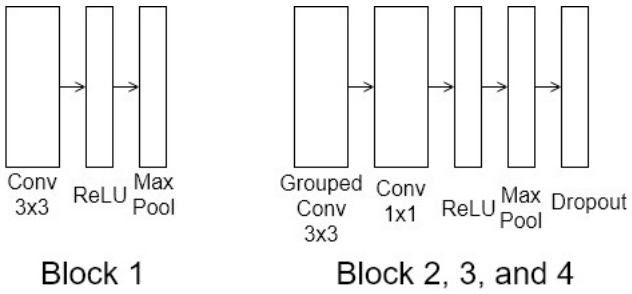

Fig. 4: Convolutional Network Block Detail

TABLE 3
PARAMETER DETAILS

Block	Layer	Parameter	Dataset	
			Ship	EuroSAT
1	Convolution	input	3	3
		output	12	16
		filter size	3x3	3x3
2	Depthwise Convolution	input	12	16
		output	72	96
		group	12	16
		filter size	3x3	3x3
	Convolution	input	72	96
		output	24	32
		filter size	1x1	1x1
3	Depthwise Convolution	input	24	32
		output	144	192
		group	24	32
		filter size	3x3	3x3
	Convolution	input	144	192
		output	36	48
		filter size	1x1	1x1
4	Depthwise Convolution	input	36	48
		output	216	288
		group	36	48
		filter size	3x3	3x3
	Convolution	input	216	288
		output	48	64
		filter size	1x1	1x1
Fully Connected (ConvNet-1)		input	432	576
		output	2	10
Fully Connected (ConvNet-2)		input	48	64
		output	2	10

to be identical accross datasets. First, the image pixels were converted from 256 levels into 4 levels of intensity per channels. The co-occurence were computed for pixels with distance of 1, 2, 4, and 8. The angles which considered for co-occurence are 0, $\frac{\pi}{2}$, $\frac{3\pi}{4}$ and π. The order of value pair was ignored (resulted in symmetric matrix) and the matrix was normalized before feature computation.

With four values of pixel distance, four values of angles, and six types of features, there are 96 features extracted for a single channel. For experiment with RGB image, the number of evaluated feature is 96 x 3 = 288. GLCM library from scikit-learn library to implement the method [21].

3. Result and Discussion

3.1. Results

Both Ship and EuroSAT dataset were evaluated in grayscale and RGB. In each case, three feature extraction methods were tested. Thus, there are twelve models in total. The accuracy of each model is summarized in Table 4.

Principal component visualization for EuroSAT dataset is provided in Figure 5. The figure depicts

TABLE 4
SUMMARY OF MODEL ACCURACY

Feature from	Dataset			
	Ship		EuroSAT	
	Gray	RGB	Gray	RGB
ConvNet-1	0.9640	0.9708	0.7162	0.786
ConvNet-2	0.9190	0.9258	0.7124	0.7268
GLCM	0.8950	0.9250	0.5014	0.6318

16 samples per class on the first two principal components. The principal component histograms are shown in Figure 6 and 7 for Ship and EuroSAT RGB dataset respectively. PC0 denotes the first principal component while PC1 the second.

Besides accuracy, other model performance indicator were also evaluated, namely precision, recall, and F1-score. Table 5 shows model recall, precision, and F1-score on grayscale samples of Ship Dataset. The results on RGB samples are shown in Table 6.

TABLE 5
SHIP DATASET (GRAYSCALE)

	Class	Precision	Recall	F1-score
ConvNet1	0	0.98	0.97	0.98
	1	0.91	0.94	0.93
ConvNet2	0	0.94	0.95	0.95
	1	0.84	0.82	0.83
GLCM	0	0.92	0.94	0.93
	1	0.81	0.75	0.77

TABLE 6
SHIP DATASET (RGB)

	Class	Precision	Recall	F1-score
ConvNet1	0	0.98	0.98	0.98
	1	0.94	0.94	0.94
ConvNet2	0	0.96	0.94	0.95
	1	0.83	0.87	0.85
GLCM	0	0.95	0.95	0.95
	1	0.85	0.84	0.84

Table 7 and Table 8 summarize recall, precision, and F1-score on grayscale and RGB version of EuroSAT dataset respectively. On the table, C1, C2, GM denotes ConvNet-1, ConvNet-2, and GLCM.

3.2. Discussion

3.2.1. Model Performance. The experiment results indicate consistency accross datasets. First of all, features extracted from the ConvNets are roughly more predictive than GLCM as indicated by accuracy, precision, recall, and F1-score shown from Table 4 to Table 8. The only case where GLCM performed nearly as good as convolutional neural network is on Ship recognition dataset, where the resulting accuracy is approximately equal to ConvNet2. Although the performance is similar, ConvNet-2 utilized much smaller number of feature (48) compared to GLCM (96).

Besides measuring performance, principal component analysis was also performed for both visualization and observing feature value. Because the model performance are relatively high on Ship dataset, there is no interesting pattern to be presented and discussed. Figure 5 depicts the first two principal components of the feature learned on EuroSAT dataset. The visualization provided in the figure clearly shows that a more separable pattern is created by convolutional neural network compared to GLCM. The separability is consistent with the performance measure, where convolutional neural network based methods performed better than GLCM.

The distribution of feature principal components is shown by histograms on Figure 6 and 7 for Ship and EuroSAT dataset respectively. There is an interesting pattern visualized by the histograms. A very high zero frequency is shown by principal component of ConvNet features. On the other hand, there is no clear distribution shape in GLCM. Possible cause of the distribution shape could be the use of ReLU activation (which set negative values to zero) or the ability convolutional neural network to learn features efficiently (represent the pattern with minimum number of non-zero component). Nonetheless, these possibilities needs verification by further study with more datasets and network architectures.

3.2.2. The effect of Adaptive Pooling Output. Compared to ConvNet-2, ConvNet-1 achieved better performance as indicated by higher score on a lot of cases. For example, the precision and recall of ConvNet-1 is higher in Table 5 and 6. ConvNet-2 only outperformed ConvNet-1 slightly at some metrics of some classes on EuroSAT dataset, as indicated by Table 7.

Specifically on both grayscale and RGB Ship Dataset, ConvNet-1 outperformed other method significantly. In relation to spatial information, this result is rational because the positive sample must be a full ship object, as shown in Figure 1 with ID 131 and 889. A partial ship object, such as ID 2434, is classified as negative sample. Therefore, spatial information is useful to detect ship boundary.

The reason for difference in performance is difficult to be explained given the limited number of experiment. However, by considering the model and the case, there are several possibilities. First, spatial resolution does matter. This implies removing spatial

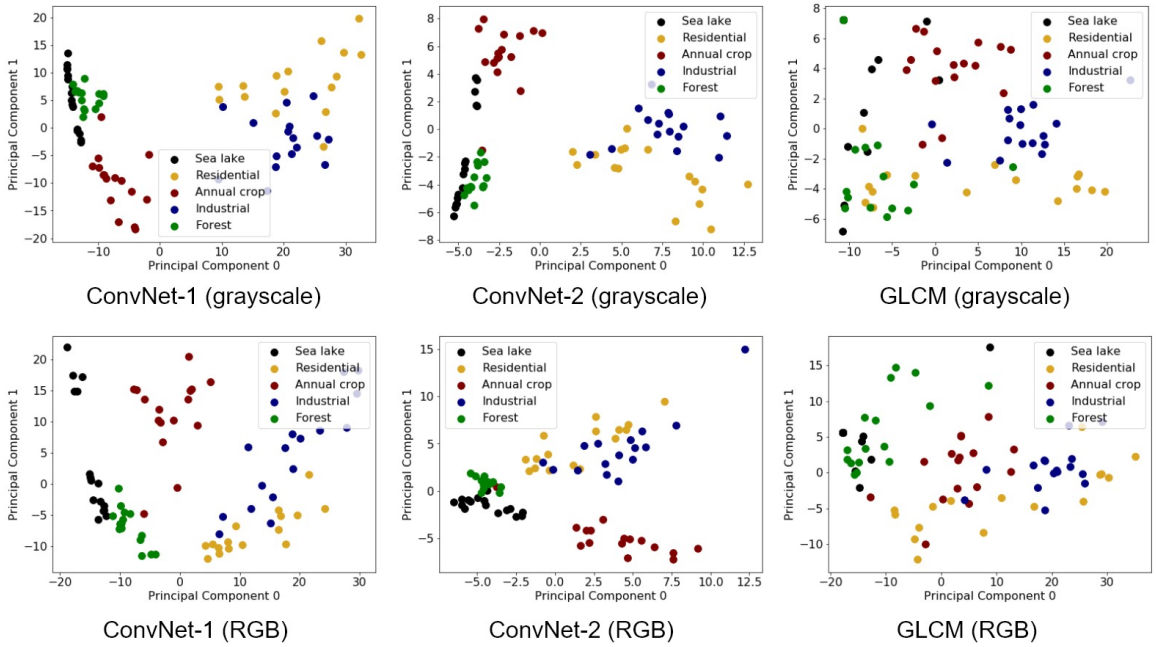

Fig. 5: 2-Components PCA visualization for EuroSAT dataset

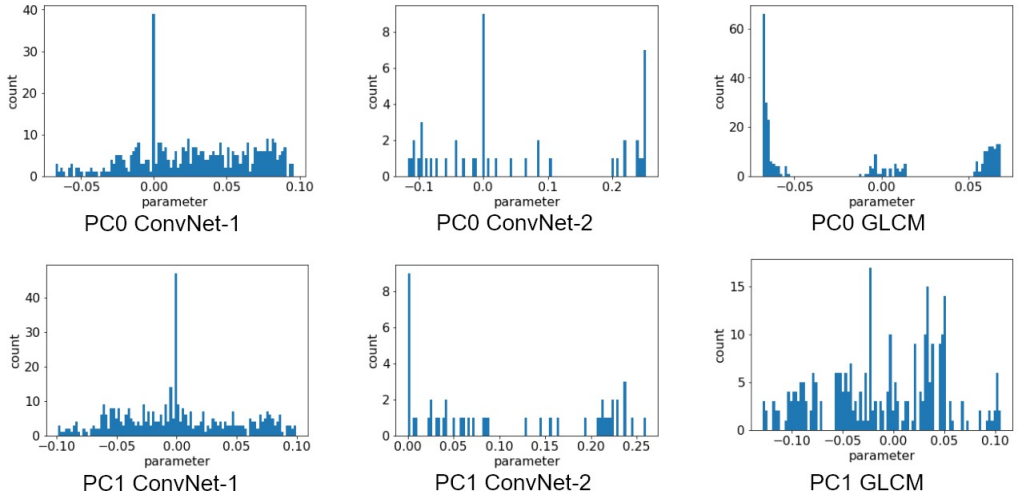

Fig. 6: Principal Component Value Histogram (Ship RGB Dataset)

information completely with global maximum pooling (adaptive pooling with 1x1xn output) resulted in less performance compared to retaining spatial information with 3x3xn adaptive pooling. Second, with 3x3 pooling output, ConvNet-1 has nine times more feature than ConvNet-2. Therefore, a more complex pattern could be learned.

3.2.3. RGB and Grayscale Performance. Models trained on RGB version of the dataset performed better than the grayscale counterparts. The difference is on the gain of performance.

For example, on Ship dataset, the improvement gained from training on RGB with respect to grayscale is small for convolutional neural network (0.964 to 0.971 for ConvNet-1 and 0.919 to 0.925 for ConvNet-2) compared to GLCM (0.89 to 0.92). On grayscale alone, the smallest accuracy of all methods is 0.89. This indicates that a single channel texture

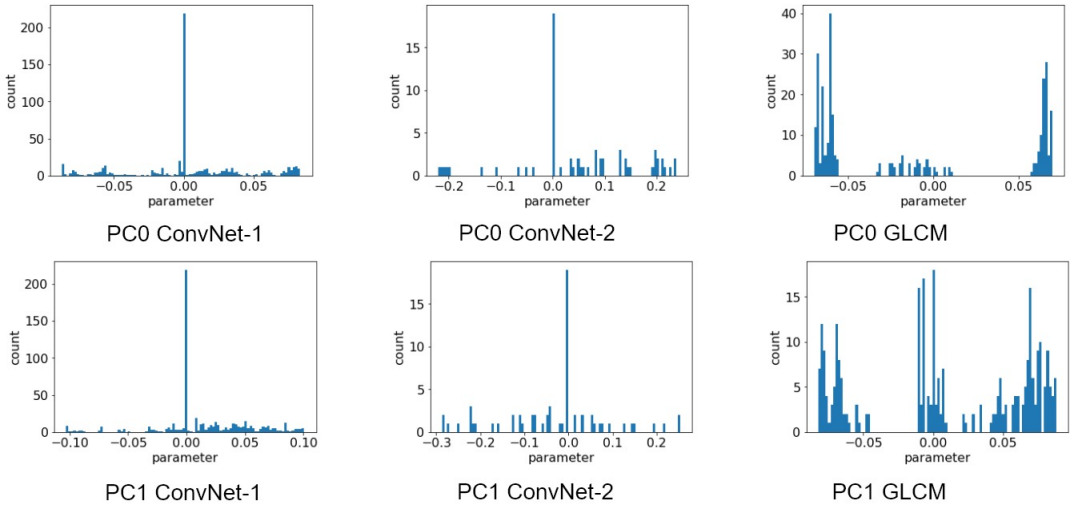

Fig. 7: Principal Component Value Histogram (EuroSAT RGB Dataset)

TABLE 7
EUROSAT DATASET (GRAYSCALE)

Class Index	Precision			Recall			F1-score		
	C1	C2	GM	C1	C2	GM	C1	C2	GM
0	0.59	0.58	0.45	0.43	0.51	0.41	0.5	0.54	0.43
1	0.93	0.94	0.71	0.98	0.96	0.69	0.95	0.95	0.7
2	0.54	0.56	0.31	0.47	0.31	0.09	0.5	0.4	0.14
3	0.92	0.85	0.63	0.92	0.96	0.76	0.92	0.9	0.69
4	0.74	0.72	0.48	0.78	0.8	0.69	0.76	0.76	0.57
5	0.81	0.79	0.72	0.86	0.9	0.88	0.84	0.84	0.79
6	0.58	0.56	0.36	0.54	0.54	0.17	0.56	0.55	0.23
7	0.5	0.52	0.37	0.44	0.45	0.19	0.47	0.48	0.25
8	0.91	0.87	0.39	0.93	0.95	0.82	0.92	0.9	0.53
9	0.47	0.47	0.19	0.71	0.6	0.09	0.56	0.53	0.12

feature already provide useful information for this task. Because ConvNets already learned the texture optimally, there is not much improvement by adding color (RGB) information.

In EuroSAT dataset, the result is rather different. ConvNet-2 gained small improvement, which likely caused by the use of identical network architecture for both grayscale and RGB. With identical architecture, the number of extracted feature is equal. As shown in Table 4, only a very small improvement was gained (from 0.712 to 0.726) because ConvNet-2 provides only 64 number of features for both grayscale and RGB. ConvNet-1 improves quite significantly (from 0.716 to 0.786) possibly because ConvNet-1 has significantly more feature (9x64). GLCM features also gained significant improvement on RGB dataset likely with the same reason as ConvNet-1. For the GLCM experiment, as feature is extracted per channel, RGB has three times features than grayscale.

4. Conclusion

The study presented performance evaluation of models which learned features produced by ConvNets and GLCM. In contrast to previous study, the proposed network utilized depthwise separable convolution and was trained with no transfer learning. The result is consistent with previous studies that convolutional neural network is superior to classic method such as GLCM in terms of most of the metrics (accuracy, recall, precision, and F1-score) for both of the evaluated datasets.

Two similar ConvNets are evaluated. The difference between both networks is on the final adaptive pooling layer. The result shows that the network with 3x3xn pooling output demonstrated better performance compared to the network with 1x1xn.

Training on RGB image improved model performance on most of the evaluated cases. However, the amount of improvement is varied accross all

TABLE 8
EUROSAT DATASET (RGB)

Class Index	Precision			Recall			F1-score		
	C1	C2	GM	C1	C2	GM	C1	C2	GM
0	0.6	0.62	0.55	0.63	0.55	0.34	0.61	0.58	0.42
1	0.95	0.94	0.7	0.97	0.94	0.88	0.96	0.94	0.78
2	0.6	0.53	0.34	0.5	0.29	0.47	0.55	0.37	0.4
3	0.9	0.82	0.82	0.97	0.95	0.93	0.93	0.88	0.87
4	0.83	0.77	0.65	0.79	0.79	0.73	0.81	0.78	0.69
5	0.89	0.83	0.92	0.91	0.87	0.85	0.9	0.85	0.88
6	0.72	0.61	0.54	0.62	0.61	0.33	0.67	0.61	0.41
7	0.64	0.54	0.49	0.63	0.57	0.36	0.64	0.55	0.42
8	0.92	0.83	0.81	0.95	0.96	0.77	0.94	0.89	0.79
9	0.68	0.55	0.43	0.82	0.61	0.55	0.75	0.58	0.48

models. The ConvNet with 1x1xn polling, which consequently has the smallest number of features, exhibited the smallest improvement on both datasets. The improvement also depends on the complexity of the pattern. In our experiment on Ship Dataset, for example, ConvNets gained small improvement as the method learned the pattern from single channel texture optimally.

References

[1] MAJOR QUENTEN L WILKES. Meteorology applications of satellite imagery. *PHOTOGRAMMETRIC ENGINEERING AND REMOTE SENSING*, 40(10):1165–1172, 1974.

[2] R Rajeesh and GS Dwarakish. Satellite oceanography–a review. *Aquatic Procedia*, 4:165–172, 2015.

[3] Jiayuan Fan, Tao Chen, and Shijian Lu. Unsupervised feature learning for land-use scene recognition. *IEEE Transactions on Geoscience and Remote Sensing*, 55(4):2250–2261, 2017.

[4] Shawn D Newsam and Chandrika Kamath. Retrieval using texture features in high-resolution multispectral satellite imagery. In *Data Mining and Knowledge Discovery: Theory, Tools, and Technology VI*, volume 5433, pages 21–33. International Society for Optics and Photonics, 2004.

[5] Biswajeet Pradhan, Ulrike Hagemann, Mahyat Shafapour Tehrany, and Nikolas Prechtel. An easy to use arcmap based texture analysis program for extraction of flooded areas from terrasar-x satellite image. *Computers & geosciences*, 63:34–43, 2014.

[6] A Makedonas, C Theoharatos, V Tsagaris, V Anastasopoulos, and S Costicoglou. Vessel classification in cosmoskymed sar data using hierarchical feature selection. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 2015.

[7] Haitao Lang, Jie Zhang, Ting Zhang, Di Zhao, and Junmin Meng. Hierarchical ship detection and recognition with high-resolution polarimetric synthetic aperture radar imagery. *Journal of Applied Remote Sensing*, 8(1):083623, 2014.

[8] Gong Cheng, Junwei Han, Lei Guo, Zhenbao Liu, Shuhui Bu, and Jinchang Ren. Effective and efficient midlevel visual elements-oriented land-use classification using vhr remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 53(8):4238–4249, 2015.

[9] Marco Castelluccio, Giovanni Poggi, Carlo Sansone, and Luisa Verdoliva. Land use classification in remote sensing images by convolutional neural networks. *arXiv preprint arXiv:1508.00092*, 2015.

[10] Dimitrios Marmanis, Mihai Datcu, Thomas Esch, and Uwe Stilla. Deep learning earth observation classification using imagenet pretrained networks. *IEEE Geoscience and Remote Sensing Letters*, 13(1):105–109, 2016.

[11] Michael Xie, Neal Jean, Marshall Burke, David Lobell, and Stefano Ermon. Transfer learning from deep features for remote sensing and poverty mapping. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[12] Quanzhi An, Zongxu Pan, and Hongjian You. Ship detection in gaofen-3 sar images based on sea clutter distribution analysis and deep convolutional neural network. *Sensors*, 18(2):334, 2018.

[13] Carlos Bentes, Domenico Velotto, and Björn Tings. Ship classification in terrasar-x images with convolutional neural networks. *IEEE Journal of Oceanic Engineering*, 43(1):258–266, 2018.

[14] Masoud Mahdianpari, Bahram Salehi, Mohammad Rezaee, Fariba Mohammadimanesh, and Yun Zhang. Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery. *Remote Sensing*, 10(7):1119, 2018.

[15] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[16] Planet Team. Planet application program interface: In space for life on earth. *San Francisco, CA*, 2017.

[17] Ships in satellite imagery. _https://www.kaggle.com/rhammell/ships-in-satellite-imagery. Accessed: 2019-03-13.

[18] Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. Introducing eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 204–207. IEEE, 2018.

[19] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017.

[20] Robert M Haralick, Karthikeyan Shanmugam, et al. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6):610–621, 1973.

[21] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.