POLSAR IMAGE CLASSIFICATION USING ATTENTION BASED SHALLOW TO DEEP CONVOLUTIONAL NEURAL NETWORK

Mohammed Q. Alkhatib, Mina Al-Saad, Nour Aburaed, M. Sami Zitouni and Hussain Al-Ahmad

College of Engineering and IT, University of Dubai, Dubai, UAE

ABSTRACT

This paper proposes a novel multi-branch feature fusion network for PolSAR image classification and interpretation. It is built using Complex-valued Convolutional Neural Networks (CV-CNNs). The proposed approach utilizes extraction of polarimetric features at each branch to achieve high classification accuracy. Moreover, Squeeze and Excitation (SE) is also introduced within the model's architecture. SE block improves channel interdependencies with almost no additional computational cost. The proposed approach is tested and evaluated using Flevoland benchmark dataset. Experiments demonstrate the effectiveness of the proposed attention based shallow to deep CV-CNN model for PolSAR image classification in terms of Kappa Coefficient (k), Overall Accuracy (OA), and Average Accuracy (AA) metrics.

The project can be accessed at https://github.com/mqalkhatib/PolSAR_CV-CNN-SE

***Index Terms*—** PolSAR, Complex-Valued CNN, Classification, Squeeze and Excitation Networks

1. INTRODUCTION

Polarimetric Synthetic Aperture Radar (PolSAR) image classification plays a crucial role in remote sensing by providing detailed information about Earth's surface characteristics. These images offer insights into vegetation, water bodies, and man-made structures, and their phase information is valuable for applications, such as resource management and land use mapping. However, analyzing PolSAR images is challenging due to high dimensionality, data complexity, and interdependencies between polarimetric channels.

Recently, Deep Learning (DL) technology has proven to be effective in the field of PolSAR image classification [1, 2]. In particular, Convolutional Neural Networks (CNNs) have demonstrated extremely powerful performance in such tasks [3, 4]. Phase information is unique to SAR imagery, and it is a crucial component in many SAR applications, such as object classification and recognition. Many research studies discussed utilizing Complex-valued Convolutional Neural Networks (CV-CNN) for PolSAR data classification [5, 6, 7]. Compared to traditional CNNs, CV-CNNs can handle complex-valued input data by using complex-valued filters and activations in the network. This allows the network to effectively capture the phase and amplitude information of the data, which is important for accurate PolSAR data classification.

Training CNNs typically requires a large number of samples, but collecting sufficient ground reference data can be expensive and complex. Limited training samples can lead to unreliable parameters and overfitting. Therefore, achieving good classification performance with less training data is crucial. Shallow networks capture shallow features and struggle with complex ones, while deep structures excel at extracting complex features. Integrating information from different depths enables effective learning from a small number of training samples, improving the network's ability to understand dataset characteristics.

Additionally, Squeeze and Excitation (SE) networks [8] have been shown to enhance CNN performance by selectively emphasizing crucial features in the input data. This is particularly beneficial for PolSAR data classification, given its multidimensional nature and redundant information [9]. By highlighting the most important features, SE can improve CV-CNN performance by reducing spatial dimensions and eliminating redundancy in the feature maps of the input data.

This paper presents a novel method for PolSAR image classification based on multi-branch feature fusion network that is built using CV-CNNs. The proposed approach utilizes extraction of polarimetric features at each branch to achieve high classification accuracy. The model performance is further enhanced by adding an SE block. The SE block improves channel interdependencies with almost no additional computational cost. This block consists of two steps: a "squeeze" step, which compresses global spatial information into a channel descriptor, and an "excitation" step, which takes the channel descriptor and applies a fully-connected layer to generate a vector of activations, which is then used to rescale the feature maps. The proposed model is tested and evaluated on the Flevoland benchmark dataset.

Overall, this framework demonstrates the effectiveness of the proposed attention based shallow to deep CV-CNN model for PolSAR image classification and provides a strong foundation for further research in this area.

This paper's contributions include: developing a new DL

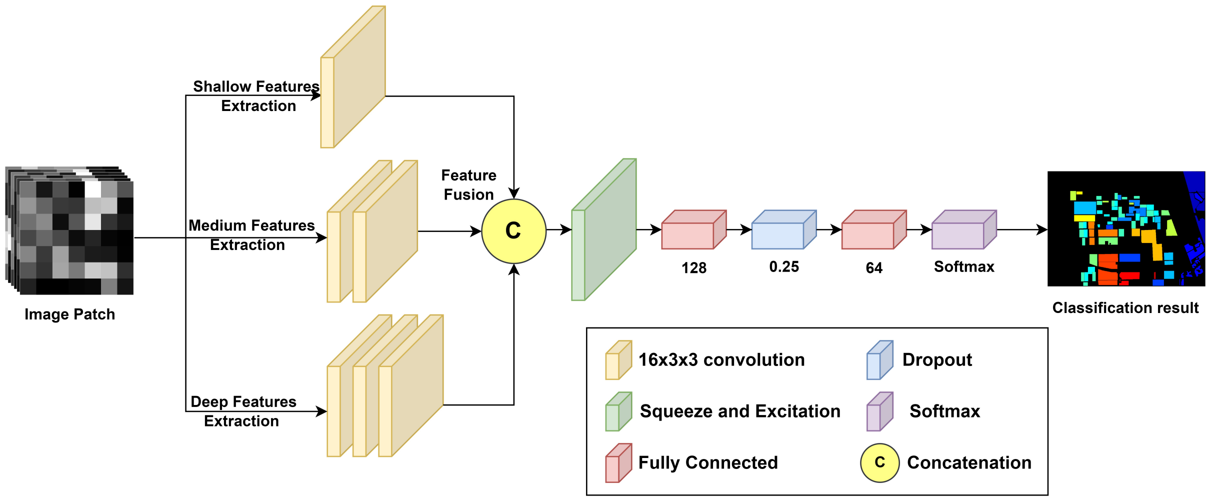

Fig. 1: Proposed CV-CNN-SE model for PolSAR image classification.

model based on multi-branch CV-CNN for PolSAR image classification, demonstrating the effectiveness of using CV-CNN compared to Real-valued CNN (RV-CNN), enhancing the model's performance with complex SE block, which is referredto as CV-CNN-SE, and evaluating classification results quantitatively using Kappa Coefficient, Overall Accuracy, and Average Accuracy metrics, while also providing qualitative results for visual comparison.

The rest of the paper is organized as follows: Section 2 explains the overall architecture of the proposed model, Section 3 describes the dataset and demonstrates the experimental results, and finally, Section 4 summarizes and concludes this paper.

2. NETWORK ARCHITECTURE

The proposed CV-CNN-SE model for PolSAR image classification employing CV-CNNs, multi-branch feature fusion, and channel attention using SE is shown in Figure 1. The model feeds the data through a three-branch network to extract features at various levels before concatenating the three features. The SE block is then applied to the concatenated features to further enhance channel dependencies. Two fully connected layers are used to perform classification at the end, with dropout used to reduce over-fitting and a softmax layer used to produce the final prediction. Due to the fact that the pooling layer causes loss of information, it has not been used within the network architecture.

Each complex convolution layer consists of 16 filters with kernel size 3×3. However, for the RV-CNN, the number of filters is multiplied by $\sqrt{2}$ and rounded to the closest integer as described in [10]. It is demonstrated that both RV-CNN and CV-CNN models have the same capacity in terms of real-valued trainable parameters, indicating that a fair comparison can be made between them.

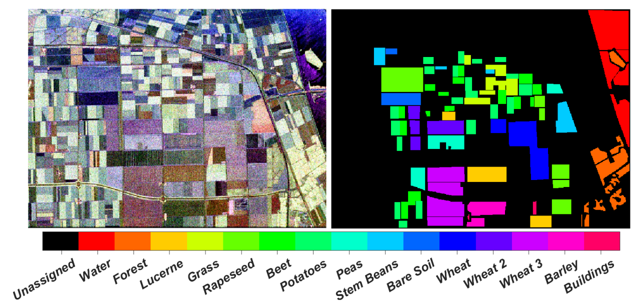

Fig. 2: Flevoland PolSAR data (left) Pauli RGB composite (right) Reference class map.

With regards to the activation function, the Complex-valued ReLU ($\mathbb{C}$ReLU(.)) was used. It is obtained by applying the well-known ReLU(.) to both the real and imaginary parts separately. So that $\mathbb{C}\text{ReLU}(x) = \text{ReLU}(\Re(x)) + j\text{ReLU}(\Im(x))$

The loss function employed in this study is categorical cross-entropy, which is computed separately for the real and imaginary parts of the prediction result in the complex network. The two error values are then averaged, and the resulting value is optimized using the Adam optimizer with a learning rate of 1×10^{-3}.

The CVNN library, developed by [11], is utilized to define a CV-CNN for PolSAR data classification. The CVNN library is based on the TensorFlow library in python environment.

3. EXPERIMENTS AND ANALYSIS

3.1. Dataset

The electromagnetic scattering characteristics of ground objects can be completely described by a polarized scattering matrix. The scattering matrix S is defined as:

$$S = \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix}, \quad (1)$$

where $S_{AB}(A, B \in H, V)$ represents the backscattering coefficient of the polarized electromagnetic wave in emitting A direction and receiving B direction. H and V represent the horizontal and vertical polarization, respectively.

The S matrix is usually transformed into the polarization coherency matrix or polarization covariance matrix. The polarization vector based on Pauli decomposition are expressed as [12]:

$$\vec{k} = -\frac{1}{\sqrt{2}}[S_{HH} + S_{VV}, S_{HH} - S_{VV}, 2S_{HV}]^T. \quad (2)$$

The Hermitian coherency matrix is expressed as:

$$T = \frac{1}{n}\sum_{j=1}^{n} k_j k_j^H, \quad (3)$$

where the operator ^{H} stands for complex conjugate operation and n is the number of pixels chosen in a boxcar located in each local area. It is worth mentioning that the the diagonal elements of T are real-valued, and the rest of the elements are complex-valued and conjugated at symmetric positions of the main diagonal. As a result, the three real-valued and three complex-valued elements of the upper triangle of the coherency matrix (i.e. $T11$, $T22$, $T33$, $T12$, $T13$, $T23$) are used as the input features of the models.

To validate this statement, the performance of the proposed network is evaluated on the Felovland PolSAR image. The scene was acquired by the NASA/JPL AirSAR system over the agricultural area in Netherlands with a size of 750×1024. The Pauli RGB image (left) and the corresponding Ground Truth (right) map are shown in Figure 2.

3.2. Simulation Results

Three experiments were conducted using RV-CNN, CV-CNN and CV-CNN-SE architectures with coherency matrix as input. The image data was first split into patches of size $11 \times 11 \times 6$. The number of labeled samples is 207,832 which gave the same number of patches. 1% of the image patches was used for training and the remaining 99% were used for testing. To minimize the impact of sample selection randomness on classification results, 10 repetitions of experiments were conducted and the average value was calculated as the final result. The classification results for each category were also provided.

RV-CNN, CV-CNN and CV-CNN-SE achieved accuracy of 93.68%, 94.52% and 95.65% Overall Accuracy (OA), respectively. Figure 3 illustrates the classification maps obtained in this research.

Since all of the accuracy values are high and it is difficult to observe the difference in the resulting class maps, the area in the white box is zoomed for a better visualization, as shown in Figure 4. It can be observed that the results of CV-CNN are better than RV-CNN. Furthermore, the use of (SE) block enhanced the resulting output, especially when looking at the lower region of the selected area.

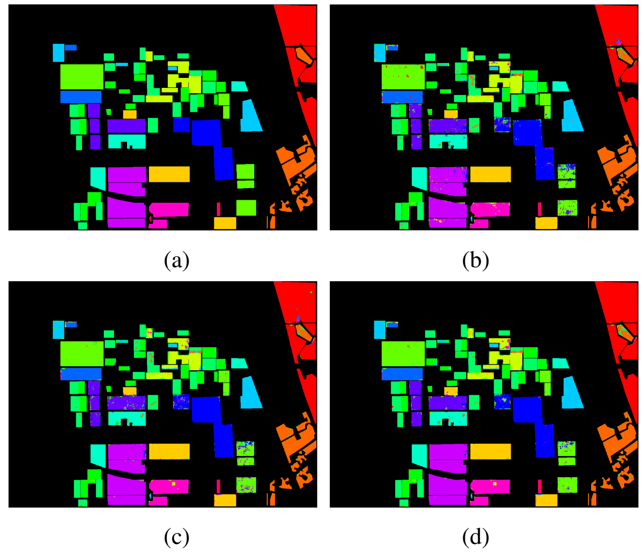

Fig. 3: Classification Maps (a) Ground Truth (GT) (b) RV-CNN (c) CV-CNN (d) CV-CNN-SE

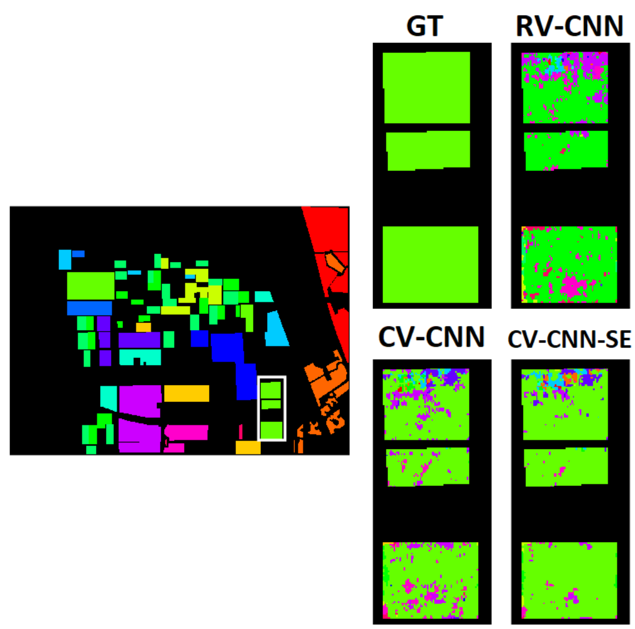

Fig. 4: Zoomed region of the area in the white box.

The classification accuracy results for the 15 classes are shown in Table 1. Only in one class (Forest) the RV-CNN performed better than the CV-CNN models. CV-CNN produced the highest accuracy values at Lucerne and Beet classes.

Table 1: This table summarizes the classification performance for each network. Bold indicates the best result.

Class #	Train	Test	RV-CNN	CV-CNN	CV-CNN-SE
Water	293	28956	98.28%	98.56%	**99.52%**
Forest	159	15696	**98.12%**	97.61%	95.88%
Lucerne	112	11088	95.17%	**95.22%**	94.46%
Grass	103	10098	86.39%	81.73%	**89.39%**
Rapeseed	219	21636	90.45%	91.96%	**95.11%**
Beet	148	14559	91.84%	**92.77%**	90.85%
Potatoes	214	21130	89.12%	92.40%	**94.52%**
Peas	104	10292	95.05%	94.90%	**95.22%**
Stem Beans	85	8386	94.16%	94.61%	**96.42%**
Bare Soil	64	6253	93.67%	94.08%	**94.20%**
Wheat	177	17462	91.92%	94.24%	**96.57%**
Wheat 2	107	10522	93.25%	93.31%	**95.88%**
Wheat 3	221	21801	95.89%	98.41%	**99.31%**
Barley	74	7295	96.16%	96.93%	**97.56%**
Buildings	6	572	85.31%	86.19%	**89.66%**
AA			$93.01 \pm 1.06\%$	$93.53 \pm 0.53\%$	$\mathbf{94.44 \pm 0.37\%}$
OA			$93.68 \pm 0.93\%$	$94.52 \pm 0.50\%$	$\mathbf{95.65 \pm 0.36\%}$
Kappa			$93.10 \pm 1.01\%$	$94.02 \pm 0.57\%$	$\mathbf{94.98 \pm 0{,}41\%}$

However, The CV-CNN-SE scored higher in the remaining 12 classes. The highest classification accuracy for one class is achieved by the CV-CNN-SE model for water class, with more than 99% accuracy, about 1% higher than CV-CNN and RV-CNN. The 'Buildings' class is of interest since it has only 6 training samples, and yet, the proposed model achieved a classification score of 89.66%. This shows the ability of the proposed CV-CNN-SE to yield decent results despite being trained on datasets of limited size.

4. CONCLUSIONS

In this work, it has been demonstrated using the well-known Flevoland PolSAR dataset that the CV-CNN architecture achieves better performance than its equivalent RV-CNN for PolSAR image classification. Also, by incorporating SE block into the model, referred to as CV-CNN-SE, an improvement in the overall performance has been observed.

In the future, the impact of each stream and the SE block in the model will be thoroughly examined. By selectively disabling or removing each stream, the relative importance and effectiveness of each component can be determined through an ablation study. the model can be also tested on various datasets to assess its generalizability and robustness. Comparative analysis with state-of-the-art methodologies will highlight the model's performance, strengths, and weaknesses.

5. REFERENCES

[1] H. Parikh, S. Patel, and V. Patel, "Classification of sar and polsar images using deep learning: A review," *International Journal of Image and Data Fusion*, vol. 11, no. 1, pp. 1–32, 2020.

[2] H. Wang, F. Xu, and Y.-Q. Jin, "A review of polsar image classification: From polarimetry to deep learning," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 3189–3192.

[3] Y. Zhou, H. Wang, F. Xu, and Y.-Q. Jin, "Polarimetric sar image classification using deep convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 12, pp. 1935–1939, 2016.

[4] R. Shang, J. Wang, L. Jiao, X. Yang, and Y. Li, "Spatial feature-based convolutional neural network for polsar image classification," *Applied Soft Computing*, vol. 123, p. 108922, 2022.

[5] R. Hänsch and O. Hellwich, "Complex-valued convolutional neural networks for object detection in polsar data," in *8th European Conference on Synthetic Aperture Radar*. VDE, 2010, pp. 1–4.

[6] J. Barrachina, C. Ren, G. Vieillard, C. Morisseau, and J.-P. Ovarlez, "Real-and complex-valued neural networks for sar image segmentation through different polarimetric representations," in *2022 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2022, pp. 1456–1460.

[7] R. M. Asiyabi, M. Datcu, H. Nies, and A. Anghel, "Complex-valued vs. real-valued convolutional neural network for polsar data classification," in *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2022, pp. 421–424.

[8] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141.

[9] R. Qin, X. Fu, and P. Lang, "Polsar image classification based on low-frequency and contour subbands-driven polarimetric senet," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 4760–4773, 2020.

[10] J. A. Barrachina, C. Ren, G. Vieillard, C. Morisseau, and J.-P. Ovarlez, "About the equivalence between complex-valued and real-valued fully connected neural networks-application to polinsar images," in *2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2021, pp. 1–6.

[11] J. A. Barrachina, C. Ren, C. Morisseau, G. Vieillard, and J.-P. Ovarlez, "Complex-valued vs. real-valued neural networks for classification perspectives: An example on non-circular data," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 2990–2994.

[12] J.-S. Lee and E. Pottier, *Polarimetric radar imaging: from basics to applications*. CRC press, 2017.