IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium | 979-8-3503-6032-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/IGARSS53475.2024.10641895

VISUAL QUESTION ANSWERING FOR WISHART H-ALPHA CLASSIFICATION OF POLARIMETRIC SAR IMAGES

Hossein Aghababaei and Alfred Stein
Faculty of Geo-Information Science and Earth Observation, University of Twente, The Netherlands

ABSTRACT

Polarimetric Synthetic Aperture Radar (PolSAR) images offer a rich repository of information, crucial for diverse applications ranging from classification to target identification. In the domain of PolSAR image classification, the Wishart classifier emerged as a prominent and widely employed technique. This classifier is often used to articulate the properties of polarimetric scattering types in images, providing valuable insight into various types of targets. With the growing interest in multidisciplinary Artificial Intelligence (AI) research, especially in computer vision and Natural Language Processing (NLP), our goal is to integrate this enthusiasm into polarimetric image analysis. We propose extending the Wishart classifier framework to include a free-form and open-ended Visual Question Answering (VQA) model. This model is designed to answer natural language questions related to PolSAR images, covering pixel details and scattering patterns. The objective is to provide accurate natural language responses that reflect real-world scenarios, such as assisting the visually impaired. Both questions and answers in this context are intentionally left open-ended to capture the complexity of inquiries in the polarimetric SAR images domain.

***Index Terms*—** Polarimetric SAR image, Classification, Visual Question Answering, Deep Learning, Natural Language, Remote Sensing.

1. INTRODUCTION

Synthetic Aperture Radar (SAR) imaging is now a widely utilized remote sensing technique, providing valuable physical information from Earth's surface irrespective of weather conditions. Polarimetric SAR (PolSAR) enhances this capability by employing multiple orthogonal polarizations to extract additional information [1]. The classification of PolSAR images holds significance for ecological and earth observation applications, with various methods proposed for this purpose. Among these methods, the unsupervised clustering scheme introduced by Cloude and Pottier [2] is widely recognized as a commonly employed approach in various applications of PolSAR images. The approach hinges on extracting two widely recognized features, namely entropy (H) and scattering angle (α), for the identification of the clusters. Utilizing these two features (H-α) enables the comprehensive representation of all random scattering mechanisms [1]. Notably, entropy serves as a natural metric for assessing the inherent reversibility of the scattering data, while the alpha angle (α) proves instrumental in discerning the underlying average scattering mechanisms. In this classification, 8 distinct class boundaries within the H-α plane have been established to effectively discriminate among surface reflection, volume diffusion, and double bounce reflection along the α axis. Simultaneously, the boundaries are strategically defined to differentiate between low, medium, and high degrees of randomness along the entropy axis. This segmentation within the H-α plane facilitates the macroscopic identification of the prevailing scattering mechanisms [2].

Later on, Lee et al. [3] introduced a method that utilized the two-dimensional H-α classification plane for estimating the initial centers of distinct clusters in polarimetric SAR images. These initial clustering centers define the training sets for classification based on the Wishart distribution. Afterward, these initial centers are utilized as training sets for consecutive iterations employing maximum likelihood classifier based upon Wishart distribution. Notably, each iteration demonstrates a significant enhancement.

The method was also expanded to include 16 clusters by incorporating the anisotropy feature (third feature), as reported in [4]. This extension was undertaken to enhance the capability of distinguishing between various classes, especially when their cluster centers converge within the same zone.

In recent years, the advancement of machine learning techniques, notably deep learning and Convolutional Neural Networks (CNNs), has prompted numerous studies to explore their application in classifying polarimetric SAR images [5-7]. Despite the prevalence of supervised methods primarily targeting land use and land cover classification, a commonality exists in the reliance on standard metrics such as root mean square error or simple absolute value error for evaluating the agreement between the network output and the reference labels of clusters during the training process. Recognizing the unique characteristics of polarimetric imagery, the incorporation of polarization properties and scattering mechanisms and consideration of statistical distributions could potentially provide a more robust and insightful method, presenting advantages over error-based assessments.

Thanks to recent advancements in deep learning, predicting answers to questions related to images has become achievable. This represents a relatively new task within the computer vision community, where the goal is to respond to free-form and open-ended questions formulated in natural language about a given image. This task is commonly referred to as Visual Question Answering (VQA) [8]. Through VQA, individuals can pose questions about the image they observe, seeking clarifications to aid in navigation or decision-making. Implementing this scheme for earth observation data could greatly enhance the accessibility and comprehensibility of remote sensing images for non-specialists. Many studies [9-11] have focused on adapting VQA for remote sensing images. These studies are however typically confined to optical images and supervised frameworks that necessitate reference labels. In this study, our objective is to extend visual question answering to the unsupervised classification problem of polarimetric SAR images. We specifically concentrate on developing a deep learning model grounded in the polarimetric scattering properties and statistical distribution of polarimetric data. To do this, we aim to extend the Wishart classifier framework to include a free-form and open-ended visual question answering. This model is designed to answer natural language questions related to PolSAR images, covering pixel details and scattering patterns. Next section introduces the details of the proposed methodology.

2. METHODOLOGY

Figure 1 presents an overview of the proposed VQA model for PolSAR classification. In its basic setup, the model extracts visual features from polarimetric patches and textual features from questions independently. These features are subsequently combined to streamline the classification of potential answers. VQA, commonly addressed through deep learning methods, places significant emphasis on datasets comprising polarimetric image patches and diverse questions. Such datasets play a pivotal role in both training and evaluating the methods. For our training dataset, we utilized a RadarSAT 2 polarimetric image with a spatial resolution of 5.2 × 7.6 m, acquired in April 2009 over the Flevoland area in the Netherlands. Subsequently, the state-of-the-art despeckling method, Mulog [12], was employed to generate and denoise the 3×3 fully polarimetric sample coherence matrix (**T**). Then, the coherence matrix undergoes partitioning into 16×16 pixel patches, resulting in 106250 patches for training and 18750 patches for testing.

An open-ended question is formulated for each image patch, designed to explore specific polarimetric scattering mechanisms. In this study, questions related to scattering are answered within the H-α plane using one of the 8 types of mechanisms, namely: 1) complex structures, 2) random anisotropic scatterers, 3) double reflection propagation effects, 4) anisotropic particles, 5) random surface, 6) dihedral reflector, 7) dipole, and 8) bragg surface. The word

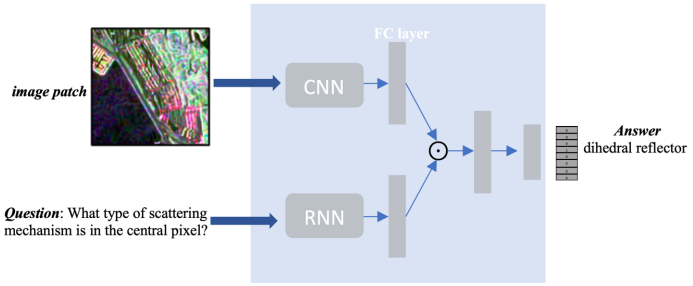

Figure 1. Visual Question Answering model for PolSAR classification.

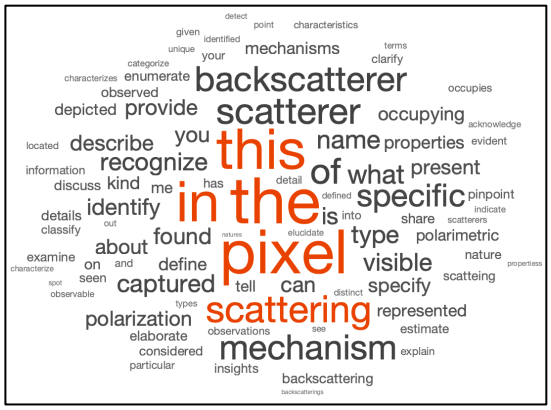

Figure 2. Visual Question Answering model for PolSAR classification.

cloud chart of the questions is shown in Figure 2, where the word sizes correspond to the count variables.

In our VQA model, the CNN to extract visual feature consists of eighteen convolutional layers with a kernel size of 3×3. All the layer output 64 feature. Batch normalization and ReLU activation function follow each layer for contrasting internal covariate shift and vanishing gradient issues. Skip connections between every three convolutional layers ensure that the input of layer number $3n+1$ is the average of the outputs of layer numbers $3n$ and $3n-2$. The last layer of the CNN is connected to a fully connected (FC) layer with size of 1024.

The utilized Recurrent Neural Network (RNN) is a bidirectional long short-term memory (BiLSTM) network. It begins with a sequence input layer, followed by a word embedding layer of dimension 25. Subsequently, a BiLSTM layer with 40 hidden units is employed. Finally, a fully connected layer with 1024 units is added, followed by batch normalization and a ReLU activation function.

Then, the fusion is executed through a straightforward point-wise multiplication between the two feature vectors. Finally, classification is carried out via two fully connected layer, transitioning from a size of 1024 to 256 and from 256 to the number of classes, i.e. 8 scattering mechanisms in the *H-α* plane.

Within the training process, we define the loss function as below:

$$L = L_{wishart} + \lambda L_{entropy} \quad (1)$$

Question	What kind of scattering pattern exist in central pixel?	Identify the backscattering mechanism in this pixel	Tell me what you see in the central pixel of this patch in terms of scattering mechanisms.
Answer	Random surface	Dihedral reflector	Random anisotropic scatterers
Wishart classifier	Class #5	Class #6	Class #2

Figure 3. Examples demonstrate the VQA model's ability to address queries using RadarSAT-2 image.

Question	Name the scatterer present in this pixel.	Categorize the scatterer represented in this pixel.	Can you recognize the backscatterer occupying this pixel.?
Answer	Double reflection propagation effects	Random anisotropic scatterers	Random surface
Wishart classifier	Class #3	Class #2	Class #8

Figure 4. Examples demonstrate the VQA model's ability to address queries using ALOS-PALSAR image.

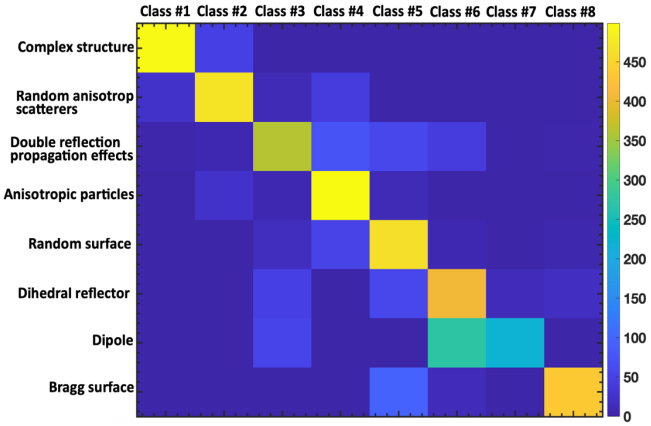

Figure 5. Confusion matrix of the answers provided by the proposed model against the results from conventional Wishart classifier. (vertical axis: the proposed VQA model, horizontal axis: Wishart classifier)

where $L_{Wishart}$ is the minimum Wishart distance [1, 3] between the central pixel in the patch and the initial centers of 8 clusters, originally derived from the H-α plane. Subsequently, a reference label is generated for each patch based on this minimum Wishart distance and then the cross-entropy loss function ($L_{entropy}$) is estimated. Throughout the iteration process, the centers of clusters are updated based on the network's output. This iterative process continues until the network is well-trained (in our case, 132 iteration).

In equation (1), the parameter λ represents the weight that balances the importance of the two terms in the loss function. In our experiments, it is set to λ = 1.

3. EXPRIMENTAL RESULT AND DISCUSSION

We present our model's accuracy based on two distinct fully polarimetric datasets: 1) RadarSAT2 images over Floveland (a test region excluded from training) and 2) ALOS-PALSAR

images over the San Francisco Bay area. Figure 3 and 4 display outcomes for various image patch categories and non-trivial questions. Visual examples demonstrate our VQA model's ability to address queries, yielding logical and visually anticipated results.

Moreover, we expand our analysis by comparing our VQA model to the traditional Wishart classifier. The accuracy of our model across diverse question categories is outlined in Figure 5, along with the corresponding confusion matrix utilizing 2850 patches of ALOS-PALSAR images. The proposed model demonstrates approximately 88% agreement with the conventional Wishart, signifying the feasibility of visually answering questions based on PolSAR data, while acknowledging room for improvement.

This observation is further underscored by evaluating the confusion matrix across different classes, revealing a high level of agreement between the standard Wishart and our proposed method. Examining the agreement for various question types in Figure 5, it becomes evident that our model tends to provide logical answers concerning the images. There remains an opportunity for improvement and generalization by expanding the training dataset. Addressing a wider variety of question types within the dataset could significantly enhance the accessibility and comprehensibility of PolSAR image classification for both experts and novice learners in the SAR community.

Lastly, it's important to note that the 88% agreement with Wishart classifier is not a direct measure of accuracy but rather a comparison with one of the common standard methods.

4. CONCLUSION

This paper introduced a new unsupervised Visual Question Answering (VQA) approach tailored for polarimetric synthetic aperture radar images. Our initial analyses have yielded promising results, indicating the potential application of such systems across various domains involving polarimetric SAR images. Future endeavors should center on refining and expanding the methodology. Specifically, the methodology can be extended to encompass 16 clusters, mirroring entropy-alpha angle-anisotropy decomposition, or adopting analogous techniques from Freeman-Durden or other decomposition methods.

Furthermore, there is room to enhance the variety of questions posed, incorporating inquiries that solicit binary responses (YES/No) regarding the presence of scattering. Notably, there exists room for improvement in the model itself; one avenue involves employing more sophisticated and deep networks to extract both visual and textual features. Subsequent phases of this research can concentrate on improving the model and evaluating the outcomes against real ground truth information, derived either from simulations or external data sources like high-resolution optical images.

REFERENCES

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

2. Cloude, S.R. and E. Pottier, An entropy based classification scheme for land applications of polarimetric SAR. IEEE transactions on geoscience and remote sensing, 1997. 35(1): p. 68-78.

3. Lee, J.-S., et al., Unsupervised classification using polarimetric decomposition and the complex Wishart classifier. IEEE Transactions on Geoscience and Remote Sensing, 1999. 37(5): p. 2249-2258.

4. Pottier, E. Unsupervised classification scheme of Polsar images based on the complex Wishart distribution and H/A/α polarimetric decomposition theorems. in Proc. 3^< rd> European Conf. on Synthetic Aperture Radar: EUSAR, 2000. 2000.

5. Hua, W., et al., PolSAR Image Classification Based on Relation Network with SWANet. Remote Sensing, 2023. 15(8): p. 2025.

6. Mousavi, H., M. Imani, and H. Ghassemian. Deep curriculum learning for polsar image classification. in 2022 International Conference on Machine Vision and Image Processing (MVIP). 2022. IEEE.

7. Zhang, L., et al. Deep learning based classification using semantic information for polsar image. in IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium. 2020. IEEE.

8. Antol, S., et al. Vqa: Visual question answering. in Proceedings of the IEEE international conference on computer vision. 2015.

9. Lobry, S., et al., RSVQA: Visual question answering for remote sensing data. IEEE Transactions on Geoscience and Remote Sensing, 2020. 58(12): p. 8555-8566.

10. Hackel, L., et al. LiT-4-RSVQA: Lightweight transformer-based visual question answering in remote sensing. in IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium. 2023. IEEE.

11. Feng, J., et al., Improving visual question answering for remote sensing via alternate-guided attention and combined loss. International Journal of Applied Earth Observation and Geoinformation, 2023. 122: p. 103427.

12. Deledalle, C.-A., et al., MuLoG, or how to apply Gaussian denoisers to multi-channel SAR speckle reduction? IEEE Transactions on Image Processing, 2017. 26(9): p. 4389-4403.