How to Cite:
Dutta, S. R., Ullah, H. M. H., Quader, M. A., & Akhtar, M. S. (2022). Remote sensing scene classification using an attention consistent network. *International Journal of Health Sciences*, *6*(S8), 844–852. https://doi.org/10.53730/ijhs.v6nS8.9810

Remote sensing scene classification using an attention consistent network

Dr. Sushama Rani Dutta
Associate Professor, Department of Information Technology, Lords Institute of Engineering &Technology, Jawaharlal Nehru Technological University, Hyderabad (JNTUH)

Habeebulla Mir Hashmat Ullah
B.Tech. Student, Department of Information Technology, Lords Institute of Engineering &Technology, Jawaharlal Nehru Technological University, Hyderabad (JNTUH)

Mohammed Abdul Quader
B.Tech. Student, Department of Information Technology, Lords Institute of Engineering &Technology, Jawaharlal Nehru Technological University, Hyderabad (JNTUH)

Mohammed Sayeed Akhtar
B.Tech. Student, Department of Information Technology, Lords Institute of Engineering &Technology, Jawaharlal Nehru Technological University, Hyderabad (JNTUH)

***Abstract*---**Image scene classification in the remotely sensed (RS) society is an interesting subject that aims to allocate land use / cover semantic information. A big amount of Convolutional neural classification models for RS images have been proposed by the authors due to the massive behaviour of CNNs in features extracted. Despite their impressive results, there are still opportunities for advancement. To begin, local characteristics are just as important as global ones in identifying RS images. The CNNs' hierarchical organizational structure and multidimensional suitable capabilities make them good at trying to capture spatial information [1]. It's not uncommon for the feature maps to be overlooked, moreover. First and foremost, the ranges among RS image pairs must be minimised or maximised in order to achieve satisfying classifier performance. Despite this, the importance of these focuses in classification tasks is undervalued. We propose a new CNN called focus high quality infrastructure (ACNet) depending on the Siamese network in the this paper to improve the robustness discussed previous section. [3]To begin, the input data for ACNet are the picture groups achieved

International Journal of Health Sciences ISSN 2550-6978 E-ISSN 2550-696X © 2022.
Manuscript submitted: 27 March 2022, Manuscript revised: 9 May 2022, Accepted for publication: 18 June 2022

through spatial rotation as a result of the network's the double framework. For example, this enables our model to explore all the feature descriptors of RS images. Second, we use a variety of attention techniques to extract all of the relevant data from the RS images. Another consideration is how spatial rotation affects RS images, as well as how RS pictures from that very same or separate definitional groups impact or are separated by an attention-consistent model. Ultimately, the main differentiator can be used to derive recognition accuracy. The ACNet has been tested against three well-known RS scene datasets. The proposed technique has the potential to outperform several chance to participate. As an outcome of the promising outcomes, ACNet can be used to classify RS images refer (https://github.com/TangXuGroup/Remote-Sensing) imageClassification/tree/main/ GLCnet for the source code of this technique

***Keywords*---**classify scenes, convolutional neural networks, remote sensing, RS.

Introduction

Remotely sensed (RS) technologies have made it possible for a wide variety of satellites to acquire a growing amount of RS photographs each day. Scholars can use these photographs to obtain a greater understanding of our globe. Since there are so many photographs, organising them is becoming a pressing issue [13]. The RS society relies heavily on RS picture classification and segmentation as a foundational and practical technology. Pictures from the RS can be divided into many categories by using semantically tags (e.g., "airport" and "beach"). In this way, investigators could determine the correct RS images based on a variety of definitions to accomplish their objectives. It is for this reasons that RS picture situation identification is widely used in a variety of relevant fields, including agribusiness, water resources, and forestry. Many effective RS images image retrieval algorithms are proposed in the decades past [4]. The 2 different structures first dominated the profession of image captioning. Learning algorithms are then used to finish the categorisation process after investigators have developed methods for extracting or learning the RS pictures' visual properties. When using the SVM [9] to produce possibility pictures with various handcrafted features, Sheng et al. [5] presented a two classifiers method that combines the probabilities generated by the SVM with confidence intervals produced by the produced possibility pictures with various feature extraction. [10] Another method for classifying scenes has been proposed in the literature. Handcrafted visual characteristics are extracted at first. This is followed by the development of a totally fragmented semantics theme strategy that incorporates the contributions of different features. The SVM classifier then assigns a classification to the combined features. Image elements (e.g. Gabor features [9], bag of words) are famous in this time because they are easy to implement and have a high degree of stability. Deep learning classifications statistical & Probabilistic concepts (such as random forest and SVM) are preferred by academics for classification tasks. Due to its high feature learning abilities and end-to-end classifications mechanism, convolution neural networks (CNNs) are becoming increasingly popular among researchers. RS scene categorization approaches based on CNNs proliferate [11]. Unaccompanied deep feature approach was introduced by Lu et al. for scene classification. The weighting nonlinear networks and the hierarchical feature model, both based on the convolutional networks are coupled to extract the most useful functions from RS images [6]. Then, scene categorization results can be derived from these characteristics. An algorithm called SRSCNN by Liu et al. involves random stretching regions cut from an input image to training a CNN for resolving objects in a scene that exhibit scale fluctuation. Lu et al. [12] suggested an end-to-end characteristic aggregating CNN to examine semantic label information (FACNN). Feature encoding and aggregation are proposed in FACNN using supervised convolutional characteristics' recording and a progressing aggregates technique that leverages semantic labelled data. To get decent results with the approaches described above, you need a large number of labelled samples as well as CNN's hierarchical feature learning structure. [8] However, they fail to take into account an important factor in the pattern categorization, namely the importance of maximising interclass differences while minimising infraclass variations. The Siamese networking is used in the RS pictures classification problem to overcome the constraint described above. Siamese network, on the other hand, is able to extract high-level semantic information from the RS images using the basic CNN architecture. The Siamese network, on the other hand, is able to mine similarities between RS picture pairings because of its dual-channel structure and particular loss function. By applying the metric learning regularisation term to the initial Siam networks, Liu et al. have made the Siam network more robust. [15] There is still room for improvement in RS scene classification using the Siamese network-based methods outlined above. When learning features for the RS pictures, it is important that both global and local information be taken into consideration. In general, the common CNN may examine the global information in an RS image. However, it is difficult for a conventional CNN to mining the RS pictures because they are so different and large in number. For the inter - class examples and intra class photos to be compacted and separated, the similarity connections among RS images should indeed be quantified in different ways. We developed an attention consistent approach that relies on the Siam networks in response to the foregoing debate and use it to classify RS picture scenes (ACNet). The intermediary extracted features (global information) are first learned from the RS pairs of images using a famous CNN (VGG16Net [13]). An object's specific information can be mined from the feature maps using a parallel-attention model, which takes advantage of the visual attention process in order to do so. Third, an awareness consistency model is developed here in order to limit the impact of discrepancies across concentration maps relating to input pairs. For the classification job, this model is useful because it can reduce intra - subject variations and increase interclass differences in RS images from a specific region perspective. Finally, utilising the learnt deep features, RS video classification outcome can be obtained. The following is a list of the article's main contributions. 1) An end-to-end RS picture scene categorization proposed based mostly on Siamese network is proposed. Additionally, by using spatial rotation, we may enhance our model's training data as well as highlight intraclass commonalities. 2) The parallel-attention model was created to collect the local data from the spectral and spatial elements of RS images. The final feature discrimination can be greatly enhanced by incorporating the global knowledge gained by a successful CNN. The attention consistent structure is intended to unify multiple types of concentration levels and evaluate the consistency among RS images for scene classification. However, the deleterious impacts of attention map disparities [11] between image pairs can be avoided. It's also possible to lessen the differences between classes and enhance the variability within classes with this strategy. Extensive trials on three real - world data revealed that our ACNet is successful for the RS picture scene classification problem, and the overall findings confirm this.

Related work

Deep learning-based tree categorization in complicated forest point clouds

In current surveying & forest management researches, the identification of trees and shrubs utilizing 3-D point cloud data has gotten a lot of interest. [3] The purpose of this letter is to introduce a new photorealistic rendering computational intelligence approach to classify species of trees in 3-D 3d point cloud obtained from complicated forest landscapes. The proposed methodology consists of three main steps: 1) different vegetation extract using point cloud density; 2) easy and accessible representation using photorealistic rendering developed a comprehensive; and 3) tree species categorization using a deep convolutional neural network [2]. The suggested technique is tested using training dataset of 3-D forested point clouds obtained by airborne laser scanning equipment. On the different datasets, the method produces an average precision of 93.1 percent and 95.6 percent. Additionally, in comparison to other 3-D fruiting trees categorisations, the presented scheme outperforms them in comparison tests.

Using VHR satellite photos to evaluate local patterns for scene classification

We evaluate the results of 13 semantic segmentation parameters, include architecture, texturing, and colour. Firstly, a single instance is used to classify images, then the functionality of other characteristics is evaluated. The support vector machine (SVM) and the k-nearest-neighbor (KNN) algorithms were all used. We employ 4 distinct measurements again for Knn method [6]. Subsequently, depending on the test outcomes, 3 among those high-performing characteristics are joined using simple concatenated. The merged feature is then utilised to classify the data. This results in an evaluation of the 13 characteristics as a whole. [7] Investigations on exceptionally high satellite images show that now the merged characteristic regularly beats the specific elements and increases the outcomes.

Sparse coding-based multiple feature combination for high-resolution satellite scene categorization

In this paper, a new method for classifying high-resolution satellite scenes is presented. The following are the three key contributions we make: Firstly, we start introducing the scarcity discussing the research methodology for [8] remote sensing analysis and classification; secondly, we introduce local texture sequence frequency distribution Fourier (LTP-HF) features, a continued to improve scale - invariant feature creaminess descriptor based on LTPs; and thirdly, we efficaciously merge a variety of input and mutually reinforcing features to further enhance the performance. A two - dimensional support vector machine (SVM) classification was performed for this purpose. [11] During first phase, the SVM is required to develop probabilistic pictures with the scale invariant transformation (SIFT), LTP-HF and coloured spectrum characteristics, correspondingly. The classification process parameters are validated by combining the produced probabilities images in different features in the second stage. Experimental results reveal that the recommended categorization algorithm produces highly exciting performances.

Methodology

It is possible to learn a great deal regarding with us biosphere through analysis of Remotely Sensed Satellite pictures because of the massive number of pictures that can be accessed, but previous approaches only analyse worldwide [14] segmentation method and don't take into account local (fractal dimension or location) characteristics. With the use of a novel process known as 'CNN dubbed awareness consistency networks (ACNet)', the researcher of this study can evaluate both properties. The training set could be used to categorise SCENES in new Remotely Sensed photos using both domestic and international features extracted by ACNet.

A Space Rotating method is used to select global information from images, followed by the application of attentiveness approaches like as VGG16 (to retrieve spatial information) and the Focused Asian Modelling to collect local patterns, in the proposed research. It will be possible to categorise fresh Remotely Sensed SCENES using the classification method that is created in the end [15].

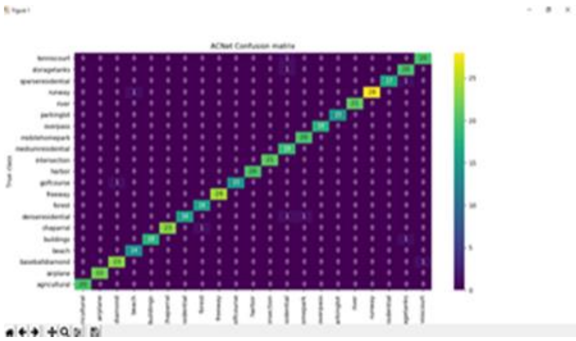
Modules of the proposed work include the following:

Overall Framework: In process of extracting global information from photos, we'll use techniques for adjusting the illumination as well as rotations of the pictures.
Intermediate Feature Extraction: To capture spatial information from rotating pictures, our approach would use an integrated and from before the VGG16 CNN architecture.
Parallel-Attention Models: A Siamese-based CNN architecture would be used to automatically extract characteristics from VGG16, including such Channels (images RGB formats) & SPATIAL (location information).
Attention Consistent Model: For the purposes of training a classification method, pairs of images with similar qualities will be combined throughout this component.
Classification Model: A sample image would be taken as well as the scenario it depicts would be predicted utilising attentiveness categorization algorithm.

Result and Discussion

The project can be started by double-clicking on the Run tab. To get the following result

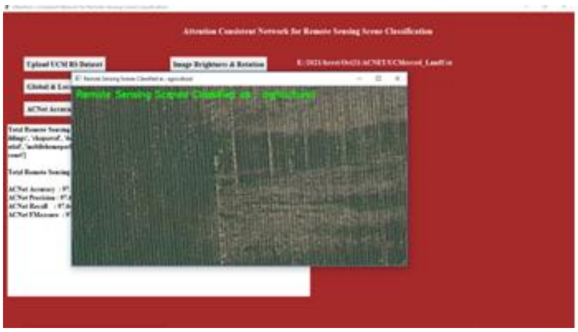
The database has a total of 2100 pictures, as seen in the results applications. Now that the datasets has been cleaned and turned, you may choose the 'Global & Locally VGG16 & ACNet Characteristics Extractor' tab to begin training ACNet with both of these attributes. The following result will appear after training ACNet with the both capabilities.



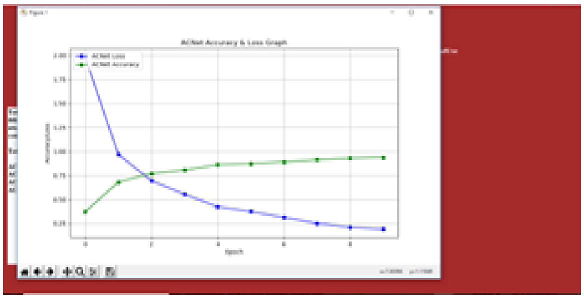
We can see in the above confusion matrix that the ACNet developed model classifies the entire test photos, as the identities of events on the X and Y axes matches the estimated couple of scenes; hence ACNET accurately classifies all image features.

To obtain the classifications seen below, I first loaded the '3.tiff' images from my computer and then clicked the 'Open' tab.

We could post different photographs then see if ACNet correctly identified the situation as a 'Airplane' like in the example above.



Efficiency and damage data can be seen by clicking the ACNET Precision and Loss graph tab



EPOCH is depicted on the x-axis, while the y-axis shows accuracy/loss levels. The green line depicts accuracy, while the blue line depicts loss values. As the epoch increases, so do the accuracy values. An effective learning algorithm is one that improves accuracy and reliability loss over time.

Conclusion

Multi networks (ACNet) classifications of RS scenes are proposed here. Models for extracting the features, concentration consistency, as well as categorization are all included. A intermediates extraction of features algorithm achieves the global information from the source picture sets (created by spatially rotations). The localized data in the RS photos is then thoroughly explored using 2 attentiveness approaches running in parallel [9]. The concentration consistency classifier is trained using reversing rotations as well as a unique transfer functions to remove the effect of spatially rotations upon that creation of visual features. [13] This phase has the potential to affect examples in the very same category and segregate examples from distinct groups, depending on how it is implemented. Finally, the classification method yields the findings. RS space transformation can benefit from our approach, as evidenced by the high scores it received on 3 well-known benchmark.

References

1. N. Zhu et al., "Deep learning for smart agriculture: Concepts, tools, applications, and opportunities," Int. J. Agricultural Biol. Eng., vol. 11, no. 4, pp. 32–44, 2018.
2. J. Marçais and J.-R. de Dreuzy, "Prospective interest of deep learning for hydrological inference," Ground Water, vol. 55, pp. 688–692, 2017.
3. X. Zou, M. Cheng, C. Wang, Y. Xia, and J. Li, "Tree classification in complex forest point clouds based on deep learning," IEEE Geosci. Remote Sens. Lett., vol. 14, no. 12, pp. 2360–2364, Dec. 2017.
4. L. Chen, W. Yang, K. Xu, and T. Xu, "Evaluation of local features for scene classification using VHR satellite images," in Proc. IEEE Joint Urban Remote Sens. Event, 2011, pp. 385–388.
5. G. Sheng, W. Yang, T. Xu, and H. Sun, "High-resolution satellite scene classification using a sparse coding based multiple feature combination," Int. J. Remote Sens., vol. 33, no. 8, pp. 2395–2412, 2012.
6. L. Jiao, X. Tang, B. Hou, and S. Wang, "SAR images retrieval based on semantic classification and region-based similarity measure for earth observation," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 8, no. 8, pp. 3876–3891, Aug. 2015.
7. X. Tang and L. Jiao, "Fusion similarity-based reranking for SAR image retrieval," IEEE Geosci. Remote Sens. Lett., vol. 14, no. 2, pp. 242–246, Feb. 2017.
8. X. Tang, L. Jiao, W. J. Emery, F. Liu, and D. Zhang, "Two-stage reranking for remote sensing image retrieval," IEEE Trans. Geosci. Remote Sens., vol. 5, no. 10, pp. 5798–5817, Oct. 2017.
9. X. Tang, L. Jiao, and W. J. Emery, "SAR image content retrieval based on fuzzy similarity and relevance feedback," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 10, no. 5, pp. 1824–1842, May 2017.
10. Q. Zhu, Y. Zhong, L. Zhang, and D. Li, "Scene classification based on the fully sparse semantic topic model," IEEE Trans. Geosci. Remote Sens., vol. 55, no. 10, pp. 5525–5538, Oct. 2017.
11. R. Xu, Y. Tao, Z. Lu, and Y. Zhong, "Attention-mechanism-containing neural networks for high-resolution remote sensing image classification," Remote Sens., vol. 10, no. 10, p. 1602, 2018.
12. Q. Zhu, Y. Zhong, L. Zhang, and D. Li, "Adaptive deep sparse semantic modeling framework for high spatial resolution image scene classification,"

IEEE Trans. Geosci. Remote Sens., vol. 56, no. 10, pp. 6180–6195, Oct. 2018.
13. Q. Zhu, Y. Zhong, S. Wu, L. Zhang, and D. Li, "Scene classification based on the sparse homogeneous-heterogeneous topic feature model," IEEE Trans. Geosci. Remote Sens., vol. 56, no. 5, pp. 2689–2703, May 2018.
14. V Jothsna, Ibrahim Patel, K Raghu, P Jahnavi, K Nagi Reddy, K Saikumar, "A Fuzzy Expert System for The Drowsiness Detection from Blink Characteristics", in proceedings of 7th International Conference on Advanced Computing and Communication Systems (ICACCS) held during 19-20 March 2021, pp.1976-1981
15. Thammana Ajay, K Nagi Reddy, Dasari Anantha Reddy, Pattem Sampath Kumar, K Saikumar, "Analysis on SAR Signal Processing for High-Performance Flexible System Design using Signal Processing" in proceedings of 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA) held during 02-04 December 2021, pp.30-34.
16. Widana, I.K., Sumetri, N.W., Sutapa, I.K., Suryasa, W. (2021). Anthropometric measures for better cardiovascular and musculoskeletal health. *Computer Applications in Engineering Education*, *29*(3), 550–561. https://doi.org/10.1002/cae.22202
17. Susilo, C. B., Jayanto, I., & Kusumawaty, I. (2021). Understanding digital technology trends in healthcare and preventive strategy. *International Journal of Health & Medical Sciences*, *4*(3), 347-354. https://doi.org/10.31295/ijhms.v4n3.1769