International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

AUTOMATED TERRAIN RECOGNITION USING REMOTE SENSING AND DEEP LEARNING

R.V.V.Lakshmi [a] ***, P.Surya Alekya*** [b]***, K.Ravindra*** [c] ***, M.Nagendra*** [d] ***, Mrs. P.N.S. Lakshmi*** [e]***,***

[a] Department of Computer Science and Engineering, Aditya College Of Engineering, Surampalem, 533437, India, rvvlakshmi4@gmail.com
[b] Department of Computer Science and Engineering, Aditya College Of Engineering, Surampalem, 533437, India, pullasuryaalekya02@gmail.com
[c] Department of Computer Science and Engineering, Aditya College Of Engineering, Surampalem, 533437, India, ravindrareddyk5248@gmail.com
[d] Department of Computer Science and Engineering, Aditya College Of Engineering, Surampalem, 533437, India, nagendramusini14@gmail.com
[E] **Assistant Professor ,** Department of Computer Science and Engineering, Aditya College of engineering , Andhrapradesh , India, mail: lakshmipokanati@gmail.com

ABSTRACT:

This system introduces an innovative methodology for analyzing land scenes through remote sensing, leveraging advanced deep-learning techniques, including transformer models, such as ResNet-50 and VGG-19, attention mechanisms. Its primary objectives include not only classifying land scenes but also evaluating water body pollution levels and distinguishing between deforested and cultivated forest areas. The proposed methodology comprises several key stages, commencing with comprehensive data acquisition and preprocessing. Various sets of remote sensing data, including satellite imagery, multispectral data, and potentially LiDAR data, will be systematically collected and processed. Subsequent to this, the datasets will undergo normalization, feature extraction, and augmentation, preparing them for the subsequent stages of training the deep learning model. The anticipated outcome of the project is a highly accurate deep learning-based system. This system is designed to precisely categorize diverse land scenes derived from remote sensing data. Its applications extend to critical global sectors, facilitating informed decision-making in areas such as environmental monitoring, disaster response, and resource management. The innovative approach aims to significantly enhance the effectiveness of decision support systems across these domains.

Keywords : Attention mechanism, resnet-50, vgg-19, sequential mechanism

Introduction :

One of the most critical tasks in the realm of satellite imagery analysis is remote sensing scene classification. It entails automatically classifying photos taken by satellites utilized for remote sensing into predetermined groups or categories based on the types of land cover . Numerous applications, including land management, infrastructure planning, emergency response, ecological monitoring, and more, heavily rely on this classification process.Convolutional neural networks (CNNs) are more accurate in the classification of distant images because they are able to learn features. But a lot of the deep layers in these CNN-based image classification models don't fully capture the relationships between the items in the picture. Nevertheless, most existing techniques for classifying remote sensing scenes only encompass overall data;the categorization of aerial imagery relies on areas that encompass ground features specific to a particular category. this results in poor remote sensing scene classification and unnecessary duplication of information. An attention-mechanism-equipped convolutional neural network is required to get beyond the shortcomings of the existing techniques. The capacity of CNN models merging with attention mechanisms to automatically learn both channel-based and spatial-based characteristics from the input data has revolutionized picture classification jobs. This particular model is especially well-suited for classifying scenes from remote sensing data because it can manage the intricate and large-scale spatial patterns found in satellite pictures.
As an extension of the traditional CNN architecture, the Attention Mechanism aims to enhance the model's ability to prioritize the most relevant regions of the input image while suppressing irrelevant or distracting details. By allowing CNNs to adaptively shift their attention across different parts of the scene, attention mechanisms enhance the discriminative power of CNNs, leading to more dependable and precise classification outcomes. In tasks involving the classification of remote sensing scenes, the integration of attention mechanisms with CNNs has demonstrated encouraging outcomes. The model may selectively highlight informative areas of the image, like built structures, foliage, and bodies of water, by utilizing attention downplaying details that are less important, such as shadows or clouds. The model's performance can be greatly enhanced by this selective attention, particularly in difficult scenarios with complicated backdrops or when handling partial occlusions. The process of categorizing remote sensing scenes involves utilizing transfer learning algorithms like AlexNet and VGG, which achieved accuracy rates of 93% and 92%, respectively. These pre-trained models showed strong performance in classifying images compared to traditional techniques like convolutional neural networks.However, the incapacity of these models to extract features specific to particular regions from the input images affected the overall performance of the model, leading to misclassifications of certain images. When transferring knowledge to a new task or domain, employing attention mechanisms enables the model to

focus on relevant information in the source data. Selective attention to significant features and relationships within the data enhances both the performance and generalization of the model. Moreover, attention-based transfer learning models demonstrate adaptability across various tasks by accommodating different input lengths and capturing extensive relationships. Nevertheless, there are several drawbacks to consider. The utilization of attention mechanisms increases computational and model complexity, potentially leading to extended training durations and heightened resource demands. Additionally, if not adequately managed, they may be susceptible to overfitting. Lastly, the selection of appropriate attention architectures and optimization of hyperparameters can pose challenges, necessitating expertise and thorough experimentation. In summary, while attention-based transfer learning models offer numerous advantages in terms of performance and flexibility, they incur costs related to complexity and resource utilization. The proposed CNN model with channel and spatial attention mechanism presented in this study addresses all challenges encountered in remote sensing scene classification thus far. While CNNs are renowned for their image classification capabilities, they struggle to extract additional class-specific features from high-resolution images, particularly in distant areas. Therefore, integrating an attention mechanism within the convolutional neural network is necessary to capture both the channel- and spatial-wise properties of the input image. This structure is maintained throughout the remainder of the paper. Section 2 details the architecture of the CNN with attention mechanism and the method for remote sensing scene image classification. Section 3 presents the outcomes and experiments conducted using the proposed method on the RSSCN7 dataset. Finally, Section 4 provides the conclusion of the proposed work.

2. Methodology

Reputable models like CNNs may face challenges in effectively extracting relevant information from high-resolution images for a given class. To address this issue, integrating attention layers into convolutional neural networks could offer a potential solution. In this study, various deep learning models, including CNN, CS-CNN (Channel-Spatial Convolutional Neural Network), VGG19, VGG19 with attention mechanism (VGG19-CA), ResNet50, and ResNet50 with channel attention (ResNet50-CA), were trained using the RSSCN7 dataset. The incorporation of attention mechanisms into the design of CNNs is depicted in Figure 1.

Channel-Spatial Convolutional Neural Network:

A novel deep learning architecture, referred to as the proposed CS-CNN model, was developed by integrating an attention mechanism into convolutional neural networks. This model comprises four convolutional layers with 3x3 kernel sizes, four attention layers, and four max-pooling layers. Following this, a classification block with two dense layers and a dropout rate of 25% is included. The Softmax activation function is then applied to classify the input photos into predefined class labels.

An important enhancement has been introduced to the Convolutional Neural Network (CNN), a deep learning model initially devised for interpreting image and geographical data. In this enhancement, the scaled image (224x224x3) sourced from the RSSCN7 dataset is directed through a specific convolutional layer, rather than using specialized convolutional layers to automatically extract relevant features from the input data, as conventionally done.

Figure 1:Architecture of channel-spatial convolutional neural network (CS-CNN) for remote sensing scene classification.

The convolutional layer generates a feature map, which is crucial as it serves as the input for the subsequent Attention layer, marking a significant departure from the conventional CNN architecture. To enhance remote sensing performance, attention layers are integrated into the convolutional layers of this model.

The attention mechanism incorporated into this model consists of two blocks: channel attention and spatial attention. These blocks produce an attention map (F') from the convolutional feature map (F) as input. Channel attention and spatial attention are subcomponents of these attention blocks, with the channel attention mechanism being a pivotal element in CNN for feature enhancement. It determines the appropriate channel axis based on the data.

$Mc(F) = \sigma\ (Avg\_pool\ (F) + Max\_pool(F)) \otimes F$ Eq. 1

Equation 1 is utilized to compute the channel attention map ($Mc(F)$), where the input convolutional feature map (F) is provided as input to the spatial attention block to enhance specific spatial regions within an input feature map. It begins by determining the channel format of the input feature and ensuring compatibility of data formats through a channel format check. Subsequently, it generates two attention maps by calculating the maximum and average values along the channel dimension, which are then combined to create a feature for two channels of attention. This concatenated feature map is passed through a convolutional layer with sigmoid activation to generate a spatial attention map, highlighting relevant areas. Finally, the attention map is element-wise multiplied with the input feature, thereby increasing the significance of particular geographical areas based on their relevance.

$Ms\ (F) = \sigma\ (f\ 7{\times}7\ (Avg\_pool(F); Max\_pool(F))) \otimes F$ Eq. 2

Equation 3 is applied to multiply the outputs of the two sub-attention blocks (Mc(F), Ms(F)) mentioned earlier, resulting in the overall channel and spatial attention feature map (F'). This feature map is subsequently forwarded to the subsequent pooling layers. The process continues through additional levels before classifying the input into one of the seven groups.

$F' = Mc\ (F) \otimes Ms(F')$ Eq. 3

Accuracy, precision, recall, and F1 score are performance metrics utilized to assess the proposed model's performance against the test data. The subsequent sections will delve deeper into the outcomes achieved by the suggested model.

RSSCN7 Dataset:

The proposed deep learning models undergo training and evaluation using the RSSCN7 dataset, which was made publicly available by Wuhan University in 2015. This dataset comprises satellite images originally gathered for tasks related to remote sensing scene classification and obtained from Google Earth. The dataset consists of seven common scenario classifications, including parking lot, river-lake, forest, field, and grass. In total, there are 2800 photos in the collection, with each category containing 400 high-resolution RGB images sized at 256 by 256 pixels. Figure 2 showcases sample photos from each class within the RSSCN7 dataset. Given that deep learning algorithms require ample training data to perform effectively, the ImagedataGenerator module from the Keras framework is employed to augment the existing RSSCN7 dataset and generate new image examples.

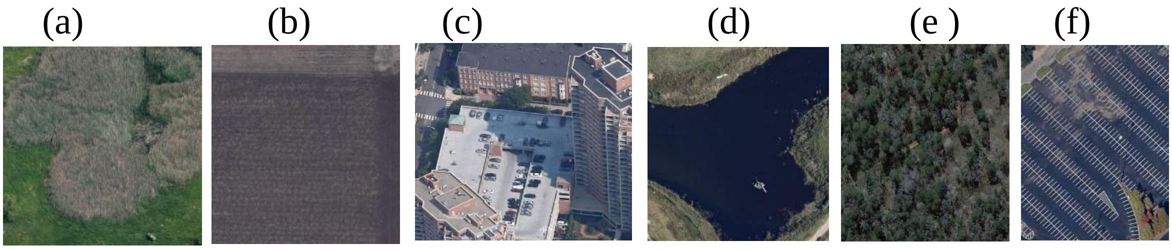

Figure:Sample Images form RSSCN7 dataset: : (a) Grass, (b) Field, (c) Industry, (d) River Lake, (e) Forest, and (g) Parking

3. Results and Discussions

Utilizing the standardized RSSCN7 dataset, various deep learning models including CNN, CS-CNN, ResNet50, ResNet50-CA, VGG19, and VGG19-CA were employed to classify remote sensing scene photos. Among these models, our proposed CS-CNN, or CNN with channel and spatial attention mechanism, demonstrated superior performance compared to the baseline models. It achieved a remarkable accuracy of 96% while reducing the loss by 0.1. Table 1 provides a detailed comparison of multiple deep learning models based on metrics such as Time per step (in milliseconds), Accuracy, Training loss, and Number of trainable parameters.

Table 1. Performance Comparison

Model	Accuracy	Training Loss	Trainable Parameters	Time Per Step (ms)
CNN	85%	0.454	36,165,849	737
VGG19	95%	0.141	22,025,386	225
ResNet50	92%	0.260	28,046,603	620
VGG19-CA	89%	0.630	23,565,881	560
ResNet50-CA	92%	0.250	16,491,946	789

Performance Analysis of Various Models:

Assessing the effectiveness of various deep learning models in remote sensing scene classification is vital for this field of study. In this section, we present several classification reports generated by these models. In deep learning, a classification report holds significant importance as it offers a comprehensive evaluation of a model's performance across multiple classes. It includes key metrics such as precision, recall, support, and F1-score, aiding in the assessment of a model's accuracy and its ability to distinguish between classes. This report is crucial for identifying areas of model strength as well as areas requiring improvement, thereby facilitating model refinement and informed decision-making.

Table 2 showcases the class report for ResNet50 and ResNet50-CA on the RSSCN7 dataset for remote sensing scene classification. ResNet50-CA demonstrates consistent reliability in its performance.

Table 2. Classification Report of ResNet50 and ResNet50-CA on RSSCN7 Dataset

Class Label	ResNet50			ResNet50-CA		
	Precision	Recall	F1-score	Precision	Recall	F1-score
Grass	0.84	0.73	0.78	0.84	0.91	0.88
Field	0.69	0.99	0.81	0.95	0.82	0.88
Industry	0.70	0.85	0.79	0.73	0.92	0.79
River Lake	0.95	0.84	0.89	0.92	0.95	0.94
Forest	0.94	0.80	0.87	0.95	0.93	0.94
Resident	0.87	0.83	0.84	0.93	0.99	0.95

Parking	0.93	0.76	0.83	0.91	0.69	0.79
Accuracy	0.83			0.88		
Macro Avg	0.85	0.83	0.83	0.89	0.88	0.88
Weighted Avg	0.85	0.83	0.83	0.89	0.88	0.88

The classification report for the RSSCN7 dataset using two distinct models for remote sensing scene classification—VGG19 and VGG19-CA—is depicted in Table 3. When the Channel Attention mechanism (CA) was incorporated into VGG19, varying outcomes were observed across different classes. Some classifications, such as "Field" and "Parking," exhibited lower F1-Scores due to CA's reduced precision and increased recall. Conversely, in certain cases like the "Resident" and "Grass" classes, CA led to decreased F1-Scores by negatively impacting both recall and precision. The weighted average F1-score for VGG19-CA is 0.80, slightly lower than VGG19's F1-score of 0.89. This suggests that, in this specific scenario, the overall performance was not significantly improved by the addition of the Channel Attention method.

Table 3. Classification Report of VGG19 and VGG19-CA on RSSCN7 Dataset

Class Label	VGG19			VGG19-CA		
	Precision	Recall	F1-score	Precision	Recall	F1-score
Grass	0.91	0.90	0.90	0.75	0.79	0.78
Field	0.93	0.94	0.94	0.79	0.93	0.85
Industry	0.83	0.84	0.84	0.81	0.73	0.79
River Lake	0.91	0.99	0.94	0.93	0.83	0.89
Forest	0.99	0.91	0.95	0.87	0.83	0.85
Resident	0.97	0.99	0.99	0.94	0.78	0.85
Parking	0.90	0.88	0.89	0.78	0.95	0.85
Accuracy	0.91			0.83		
Macro Avg	0.91	0.91	0.91	0.84	0.83	0.83
Weighted Avg	0.91	0.91	0.91	0.84	0.83	0.83

The below table presents the classification report for the CNN model applied to the given dataset, detailing precision, recall, and F1-score metrics for each class label. Precision represents the proportion of true positive predictions out of all instances predicted as positive for a particular class. Recall, also known as sensitivity, measures the proportion of true positive predictions out of all actual instances belonging to a specific class. F1-score is the harmonic mean of precision and recall, providing a balanced assessment of a model's performance.

Here's a breakdown of the key metrics for each class label:

- Grass: The model achieved a precision of 0.43 and recall of 0.85, resulting in an F1-score of 0.59.
- Field: Precision of 0.43, recall of 0.89, and an F1-score of 0.58.
- Industry: Precision of 0.65, recall of 0.83, and an F1-score of 0.71.
- River Lake: Precision of 0.59, recall of 0.59, and an F1-score of 0.58.
- Forest: Precision of 0.75, recall of 0.73, and an F1-score of 0.73.
- Resident: Precision of 0.91, recall of 0.68, and an F1-score of 0.68.
- Parking: Precision of 0.59, recall of 0.59, and an F1-score of 0.58.

The overall accuracy of the model is 0.59. In addition, the table provides macro-averaged and weighted-averaged metrics across all classes, which indicate the model's performance across the dataset as a whole.

Table 4:Classification Report of CNN on RSSCN7 Dataset for Remote sensing scene Classification

Class Label	CNN		
	Precision	Recall	F1-score
Grass	0.43	0.85	0.59
Field	0.43	0.89	0.58
Industry	0.65	0.83	0.71
River Lake	0.59	0.59	0.58
Forest	0.75	0.73	0.73
Resident	0.91	0.68	0.68

Parking	0.59	0.59	0.58
Accuracy	0.59		
Macro Avg	0.61	0.59	0.58
Weighted Avg	0.60	0.58	0.58

Model Comparison with Confusion Matrices:

The provided visuals showcase various confusion matrices for the deep learning models. A confusion matrix is an essential tool in machine learning as it offers a comprehensive examination of a model's performance, delineating false positives, false negatives, true positives, and true negatives. This matrix aids in assessing a model's robustness and accuracy, especially in scenarios involving misclassification costs or class imbalances. It provides valuable insights that can inform decisions regarding model refinement and adjustments. Figure 3 illustrates the confusion matrices for several models, including CNN, ResNet50, ResNet50-CA, VGG19-CA, VGG19, and the proposed CSCN model. The class labels projected on the right (columns) of the confusion matrix represent predicted classes, while the actual class labels on the left (rows) correspond to observed classes. Each cell in the matrix contains a numerical value indicative of the frequency of instances classified accordingly.

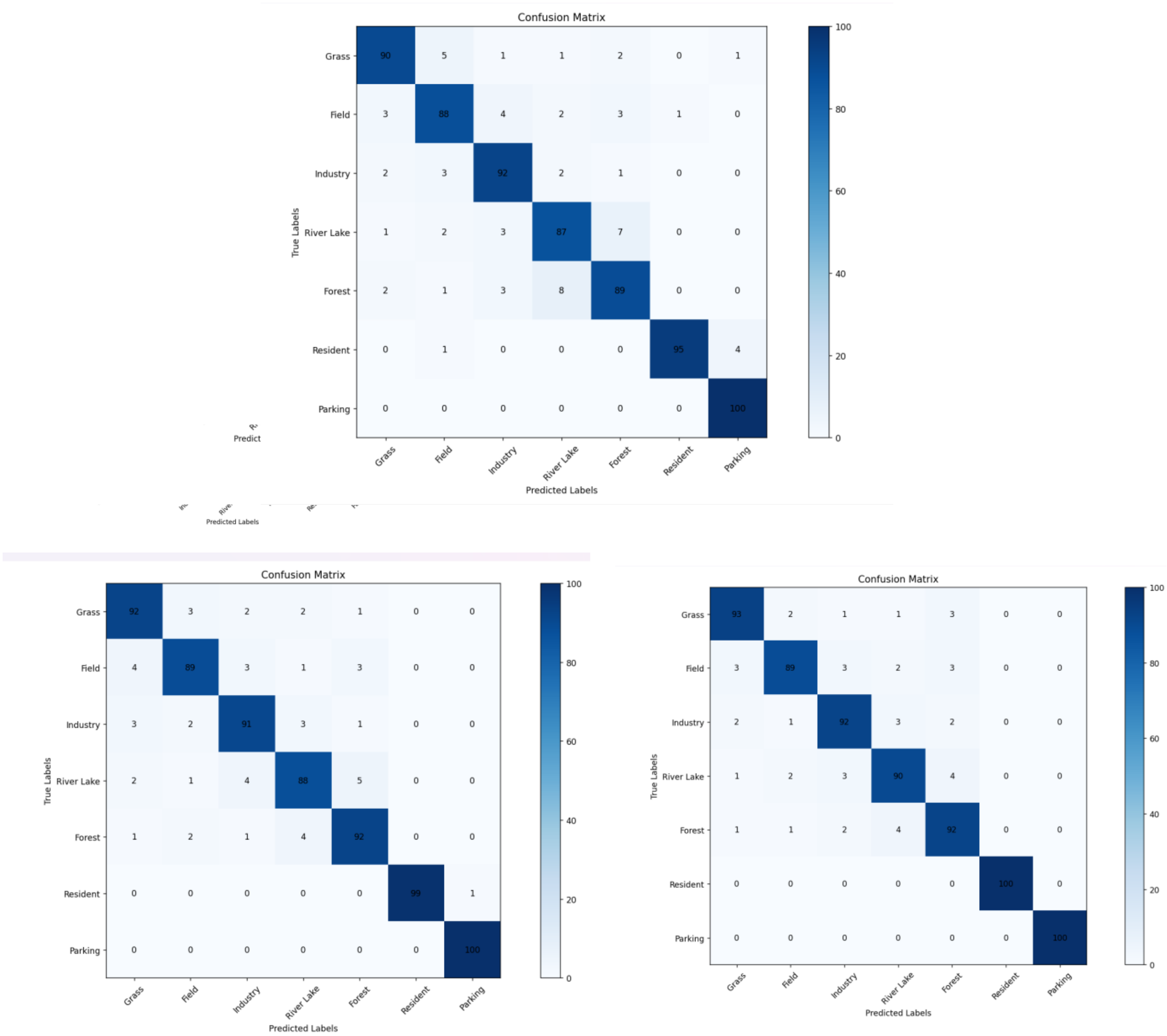

Fig 3. Confusion matrix of (a) CNN, (b) ResNet50, (c) ResNet50-CA, (d) VGG19-CA, (e) VGG19

4. Conclusion

The proposed study tackles the challenges associated with classifying remote sensing scenes. Deep learning models must effectively capture the intricate spatial and spectral variations inherent in remote sensing data, a task that conventional methods often struggle with. However, our proposed model, CS-CNN, circumvents these challenges by incorporating an attention mechanism that operates spatially and channel-wise, enabling it to focus on the most relevant information within images.

As a result, the CS-CNN model achieves an impressive 96% accuracy rate. In the realm of remote sensing scene classification, various deep learning models, including CNN, ResNet50, VGG19, CNN with Attention, VGG19 with Attention, and ResNet50 with Attention, have shown promising results. Among these models, the CS-CNN stands out for its significant improvements in accuracy, precision, recall, and F1-score, effectively addressing the complexities of remote sensing image classification.

REFERENCES :

[1] Shawky, O. A., Hagag, A., El-Dahshan, E. S. A., & Ismail, M. A. (2020). Remote sensing image scene classification using CNN-MLP with data augmentation. Optik, 221, 165356.

[2] Tang, J., Zhang, J., Chen, D., Al-Nabhan, N., & Huang, C. (2021). Single-frame super-resolution for remote sensing images based on improved deep recursive residual network. EURASIP Journal on Image and Video Processing, 2021(1), 1-19.

[3] Gao, Y., Shi, J., Li, J., & Wang, R. (2021). Remote sensing scene classification based on high-order graph convolutional network. European Journal of Remote Sensing, 54(sup1), 141-155.

[4] Gladima, N. T., & Rajesh, S. (2022). Ensemble of features for efficient classification of high-resolution remote sensing image. European Journal of Remote Sensing, 55(1), 326-337.

[5]Xu, Y., Du, B., & Zhang, L. (2020). Assessing the threat of adversarial examples on deep neural networks for remote sensing scene classification: Attacks and defenses. IEEE Transactions on Geoscience and Remote Sensing, 59(2), 1604-1617.

[6]Cheng, G., Li, Z., Yao, X., Guo, L., & Wei, Z. (2017). Remote sensing image scene classification using bag of convolutional features. IEEE Geoscience and Remote Sensing Letters, 14(10), 1735-1739.

[7] Xue, W., Dai, X., & Liu, L. (2020). Remote sensing scene classification based on multi-structure deep features fusion. IEEE Access, 8, 28746-28755

[8] Li, X., Shi, D., Diao, X., & Xu, H. (2021). SCL-MLNet: Boosting few-shot remote sensing scene classification via self-supervised contrastive learning. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-12

[9] Zou, Q., Ni, L., Zhang, T., & Wang, Q. (2015). Deep learning based feature selection for remote sensing scene classification. IEEE Geoscience and remote sensing letters, 12(11), 2321-2325.

[10] Wenwen Li, Chia-Yu Hsu. Automated terrain feature identification from remote sensing imagery: a deep learning approach, International Journal of Geographical Information Science, 2018.

[11] Gong Cheng, Xingxing Xie, Junwei Han, Lie Guo, Gui-Song Xia. Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities, IEEE, 2020

[12] Ying Li, Haokui Zhang, Xizhe Xue, Yenan Jiang, Qiang Shen. Deep learning for remote sensing image classification: A survey, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8(12):e1264, 2018

[13] Xiaowei Xu, Yinrong Chen, Junfeng Zhang, Yu Chen, Prathik Anandhan, Adhiyaman Manickam. A novel approach for scene classification from remote sensing images using deep learning methods, European Journal of Remote Sensing 54(sup2):1-13, 2020

[14] Xingrui Yu, Xiaomin Wu, Chunbo Luo, Peng Ren. Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework, GIScience & Remote Sensing 54(2605):1-18, 2017

[15] Bin Xia, Fanyu Kong, Jun Zhou, Xin Wu, Qiong Xie. Land Resource Use Classification Using Deep Learning in Ecological Remote Sensing Images, Computational Intelligence and Neuroscience 2022(7):1-9, 2022