Multi-Band Feature Fusion in Satellite Images for Land Cover Classification

Russo Mohammad Ashraf Uddin
Department of Electrical, Electronic and Computer Engineering
University of Ulsan
Ulsan, Korea
ashrafrusso@islab.ulsan.ac.kr

Xuan-Thuy Vo
Department of Electrical, Electronic and Computer Engineering
University of Ulsan
Ulsan, Korea
xthuy@islab.ulsan.ac.kr

Adri Priadana
Department of Electrical, Electronic and Computer Engineering
University of Ulsan
Ulsan, Korea
priadana3202@mail.ulsan.ac.kr

Kang-Hyun Jo
Department of Electrical, Electronic and Computer Engineering
University of Ulsan
Ulsan, Korea
acejo@ulsan.ac.kr

***Abstract*—Satellite imagery plays a crucial role in land cover classification for various applications, including urban planning, environmental monitoring, and land resource management. This paper presents a novel approach for land cover classification using satellite images by leveraging multi-band feature fusion. The proposed method aims to enhance the representation of land cover classes by performing fusion operations on the nine spectral bands beyond RGB and NIR, effectively condensing them into three fused bands. To accomplish this, we propose the EfficientNet_mb-B0 architecture as the backbone, by integrating the Band-Fusion Module (BFM) into the backbone 9 spectral bands are fused into 3 feature channels and concatenated with RGB and NIR which enables the CNN network to accurately capture the class variations using multi-spectral image features. Our, results show the effectiveness of our idea in the EuroSat dataset.**

***Index Terms*—remote sensing, satellite images, multi-spectral imaging, feature-fusion**

I. Introduction

Satellite images provide a valuable source of information for land cover classification, enabling the identification and characterization of various land cover classes across wide areas [1]. Accurate and efficient land cover classification plays a crucial role in numerous applications, including urban planning, environmental monitoring, and land resource management. Over the years, advancements in remote sensing technology have enabled the acquisition of multispectral imagery, capturing the Earth's surface in multiple spectral bands beyond the conventional RGB channels [2].

The task of land cover classification using satellite images is of significant importance due to its potential for providing detailed and comprehensive information about land cover classes [3]. By analyzing the spectral signatures present in satellite imagery, it becomes possible to differentiate between various land cover types such as residential areas, forests, agricultural fields, and water bodies [4]. The classification of land covers assists in understanding and monitoring land use patterns, assessing environmental changes, and aiding decision-making processes for sustainable development.

However, there are inherent challenges in accurately classifying land covers from satellite images. One major challenge is the presence of complex and diverse spectral characteristics that differ among land cover classes. Furthermore, variations in atmospheric conditions, sensor characteristics, and illumination conditions introduce noise and uncertainties into the acquired imagery [5]. Additionally, the vast amount of information contained in the multispectral data poses computational challenges for efficient analysis and classification.

To address these challenges, this paper proposes a novel approach for land cover classification using satellite images. The key idea is to leverage multi-band feature fusion to enhance the representation of land cover classes. Specifically, we aim to perform fusion operations on the ten spectral bands beyond RGB, effectively condensing them into three fused bands. By modifying EfficientNet-B0 [6] to process multi-band images, as the backbone architecture, which has demonstrated effectiveness in image classification tasks, we expect to achieve improved accuracy and computational efficiency in land cover classification.

In the subsequent sections of this paper, we will present the experimental analysis conducted on the EuroSAT [7] dataset, which comprises a diverse collection of satellite images with labeled land cover classes. We will discuss the fusion operation applied to the nine bands excluding RGB and NIR, the training procedure using our proposed EfficientNet_mb-B0, and the evaluation of our proposed approach against existing methods. The experimental results will highlight the efficacy of the multi-band feature fusion in enhancing land cover classification accuracy and shed light on its potential for various remote sensing applications.

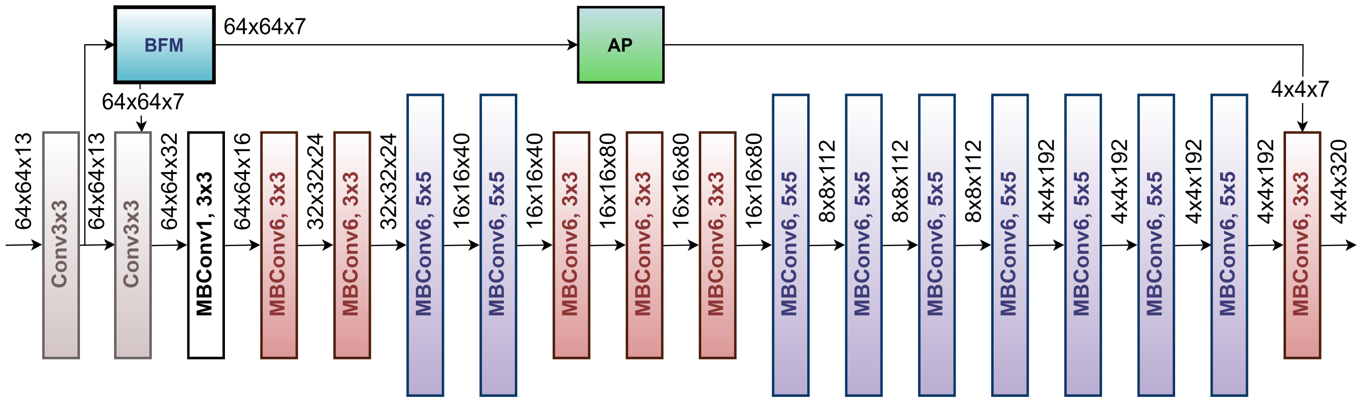

Fig. 1. EfficientNet_mb-B0 Architecture with Band-Fusion Module(BFM). BFM process the 13bands to 7-channels and facilitates the Multi-Band processing.

II. Related Works

Land cover classification using satellite images has been extensively explored in the field of remote sensing and computer vision. Various methods have been proposed to extract discriminative features from satellite imagery for accurate classification. One common approach is the utilization of spectral indices, such as the Normalized Difference Vegetation Index (NDVI) and the Enhanced Vegetation Index (EVI), which capture vegetation-related information [8] [9]. While these indices have been successful in distinguishing vegetation classes, they often rely solely on a limited set of spectral bands and may not fully exploit the richness of the available multispectral data.

To overcome these limitations, researchers have explored the use of deep learning techniques for land cover classification. Convolutional Neural Networks (CNNs) have demonstrated remarkable performance in image classification tasks and have been applied to satellite imagery as well. For instance, Zhang et al. proposed a CNN-based method that incorporates both spectral and spatial information for land cover classification [10]. Their approach achieved competitive accuracy by jointly considering multiple spectral bands and their spatial relationships.

Feature fusion has also been investigated as a means to enhance land cover classification. Zhong et al. proposed a method that combines both spectral and texture features for improved classification performance [11]. By fusing information from different feature representations, their approach achieved more robust discrimination among land cover classes. Similarly, Zhang et al. employed a fusion strategy to combine the outputs of multiple CNN models, each trained on a specific spectral band [12]. Their method effectively captured complementary information from different bands and achieved higher accuracy compared to single-band classification.

In the context of our proposed approach, we focus on the fusion of spectral bands to enhance land cover classification. While previous studies have explored feature fusion, there is limited research specifically addressing the fusion of multiple spectral bands in satellite images. Our approach aims to leverage the benefits of multi-band feature fusion by performing fusion operations on the ten additional bands beyond RGB, effectively reducing their dimensionality while preserving discriminative information. We employ the EfficientNet-B0 architecture as the backbone, which has shown promising results in image classification tasks. By integrating multi-band feature fusion with EfficientNet-B0, we expect to achieve improved accuracy and computational efficiency in land cover classification.

III. Methodology

A. *EfficientNet_mb-B0*

EfficientNet-B0 is a convolutional neural network (CNN) architecture designed to achieve high performance while being computationally efficient. It employs a compound scaling method to balance model size and accuracy. The architecture of EfficientNet-B0 involves three main components: depth-wise separable convolutions, point-wise convolutions, and mobile inverted residual blocks. Depth-wise separable convolutions reduce computational complexity by applying separate convolutional filters to each input channel, followed by point-wise convolutions to combine the outputs. This combination is known as a depth-wise separable convolution. Mobile inverted residual blocks capture and reuse feature maps effectively. They consist of an expansion convolution, a depth-wise convolution, and a projection convolution. Residual connections enable better gradient flow and facilitate the training of deeper networks. EfficientNet-B0 stacks these building blocks to create a deep and efficient network architecture. It has been shown to achieve competitive accuracy in image classification tasks while being smaller and faster than other popular CNN architectures. We modify the EfficientNet-B0 architecture to process 13 bands of multi-spectral input from the first conv. layer, for this we increase one layer in the stem block and remove the down-sampling process by changing the stride from 2 to 1. Fig.1 shows the detailed architecture of our EfficientNet_mb-B0 model, which processes 64x64x13 input data. The two stem blocks have input, and output channels of 13, 13, and 20, 32 respectively, the rest of the model is

TABLE I
EVALUATION OF MULTI-SPECTRAL LAND CLASSIFICATION IN EUROSAT DATASET.

Bs/Wr	Model Name	Params. (M)	GFlops	Accuracy(%)	mIoU(%)	Epochs	TrainingTime(s)
-	EfficientNet-B0 (RGB) [13]	4.032	0.034	97.80	-	-	-
16/4	EfficientNet-B0 (M-Band)	4.023	0.037	98.29	96.55	100	7775
16/4	**EfficientNet_mb-B0 (ours)**	4.062	0.162	**98.7**	**97.36**	100	8170
16/4	EfficientNet_mb-B0 (k5)	4.065	0.173	98.5	96.96	100	8174
16/4	EfficientNet_mb-B0 (k7)	4.069	0.19	98.35	96.67	100	8228
16/4	EfficientNet_mb-B0 (k13)	4.089	0.273	98.29	96.56	100	8298
256/4	EfficientNet-B0 (M-Band)	4.023	0.037	95	90.3	200	1381
256/4	**EfficientNet_mb-B0 (ours)**	4.062	0.162	**97.64**	**95.27**	200	2417

similar to the original, except for the addition of our Band-Fusion Module.

B. Band-Fusion Module

Band-Fusion Module or BFM is our core contribution in this work to process the 13bands of multi-spectral satellite images and produce high-quality features for the CNN network. In Fig.1, the input of the First Conv. layer is 64x64x13, and the first conv. layer has a kernel of 3x3, stride=1, and outputs a 64x64x13 feature map. This feature map will be input for the band-fusion module. Fig.2 shows the detailed BFM structure. BFM reduces the 13 spectral bands to 7 channels based on the spatial resolution(s.r.). R, G, B and NIR bands have s.r. 10m per pixel, these 4 bands are kept intact. Red Edge bands (B05, B06, B07, B08A) have s.r. of 20m/p are fused into one channel. Here we use sum-fusion as our fusion method. Similarly, SWIR 1 and SWIR 2 (B11, B12) are fused to the 6th channel. And Aerosols, Water vapor and Cirrus (B01, B09, B10) which have s.r. 60m/p are fused into the 7th channel. Finally, these 7 channels are concatenated to produce the output of the BFM. Thus, the 13bands of multi-spectral features are fused into the size 64x64x7. This output concatenates with the First conv. layer output and inputs to the Second Conv3x3 layer as a feature map of 64x64x20. Additionally, the BFM output is average-pooled into size 4x4x6 and concatenated with the final layer input as 4x4x199.

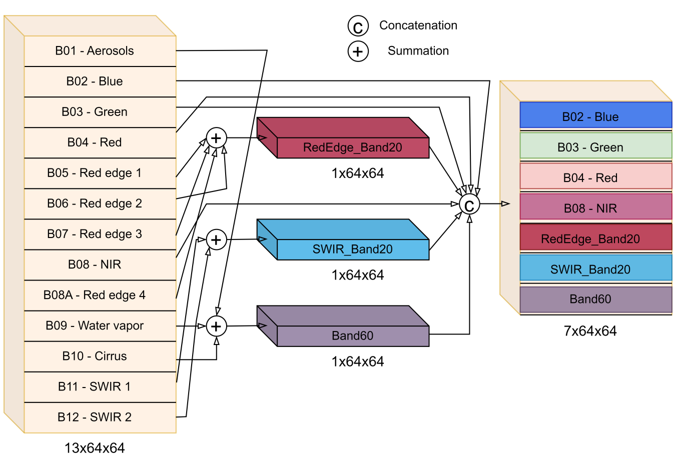

Fig. 2. Band-Fusion Module(BFM).

IV. EXPERIMENTS

A. Dataset

The EuroSAT dataset is a valuable resource for our paper's land cover classification research. It provides a comprehensive collection of satellite images specifically designed for land use and land cover classification tasks. The dataset consists of 27,000 labeled images covering ten distinct land cover classes found in Europe. The ten land cover classes in the EuroSAT dataset include residential areas, industrial areas, pastures, annual crop fields, permanent crop fields, forests, herbaceous vegetation, river, sea, or lake, and bare soil. Each class contains 2,700 images, making it a well-balanced and diverse dataset. These satellite images were acquired from the Sentinel-2 satellite, which offers high-resolution multispectral imagery. The dataset includes 13 spectral bands, ranging from bands 2 to 9, B11, and B12, with a spatial resolution of 10 meters per pixel. The images are provided in RGB format, where the three spectral bands (R, G, B) are combined to form the color images. The EuroSAT dataset has gained popularity in the remote sensing community as a benchmark for land use and land cover classification tasks. It allows researchers and practitioners to evaluate the performance of various machine learning and deep learning models in classifying different land cover types accurately. Moreover, the dataset's wide coverage of land cover classes and realistic satellite imagery makes it suitable for diverse applications in agriculture, urban planning, and environmental monitoring. In our paper, we employ the EuroSAT dataset for experimental analysis to validate the effectiveness of our proposed multi-band feature fusion approach using EfficientNet-B0. By conducting experiments on this dataset, we can assess the classification accuracy and compare our approach against existing methods in the field. The EuroSAT dataset serves as a reliable and standardized dataset for evaluating the performance of land cover classification algorithms and allows us to make meaningful comparisons with other state-of-the-art techniques.

B. Implementation Details

We use the AITLAS toolbox for EO from [13] to conduct our experiments. NVIDIA Tesla V-100 GPU with a memory size of 32GB was used as the main hardware for training. We use two different batch size/workers combinations (Bs/Wr in Table I), 16/4 and 256/4. While 16/4 gives consistent better results it takes a long time to train, 256/4 setting can train very fast but cannot achieve the best results from our experience. The learning rate was set to 0.0001 and reduced by a factor of 0.1 when the loss doesn't improve for more than 5 epochs. Each model was trained for a total of 100 epochs for batch size 16 and 200 epochs for batch size 256. The default image size is 64x64. Only two simple data augmentations- Random Horizontal Flip and Random Vertical Flip were used.

C. Evaluation on EuroSat Dataset

The detailed experimental results on the EuroSat dataset are shown in Table. I, we use Accuracy and mean IoU as our evaluation metrics following [13]. EfficentNet-B0 (M-Band) is the original EfficientNetb0 model which can process multi-band (in this case 13) input, it can already achieve 98.29% accuracy, 0.5% increase from the regular RGB input reported by the aitlas authors. Our model EfficientNet_mb-B0 further improves the accuracy by 0.41% to 98.7%, which clearly demonstrates the effectiveness of the BFM. The increase of mIoU is also consistent across the models. We further investigate, with different kernel sizes for the first conv. block, kernel size = 5, 7, and 13 was used in our experiments, but the results doesn't improve further. We also show the results using batch size 256. Our EfficientNet_mb-B0 has almost similar parameters to the original model, while the Flop increase is negligible. The training time increases in our model by around 400s for 100 epochs at batch size 16.

V. Conclusion

In this paper, we presented a novel approach for land cover classification using satellite images, leveraging multi-band feature fusion and the EfficientNet-B0 architecture. Through experimental analysis of the EuroSAT dataset, we demonstrated the effectiveness of our proposed approach in enhancing classification accuracy for multi-spectral images. We integrated EfficientNet-B0 as the backbone architecture and modified it into EfficientNet_mb-B0 which is able to process 13 bands of multi-spectral data. By performing fusion operations on the nine additional spectral bands beyond RGB and NIR, our approach effectively condensed them into three fused bands. This fusion process enhanced the representation of land cover classes, capturing diverse and discriminative information from the multi-band satellite imagery. The experimental results showcased significant improvements in accuracy compared to existing methods. Our approach leveraged the complementary information from multiple spectral bands, enabling the extraction of intricate land cover characteristics. The fusion of spectral bands effectively utilized the rich information available in satellite imagery, leading to more accurate and reliable land cover classification. The findings of this paper have several implications for remote sensing applications. The proposed approach can contribute to a better understanding of land use patterns, monitoring environmental changes, and supporting decision-making processes in various domains. The improved accuracy makes our approach suitable for large-scale land cover classification tasks. Future research directions can focus on exploring additional fusion strategies, such as incorporating spatial information or employing attention mechanisms, to further enhance classification performance. Moreover, the application of the proposed approach to other datasets and land cover classification scenarios would provide further insights into its generalizability and robustness. In conclusion, our proposed approach demonstrates the value of multi-band feature fusion in land cover classification using satellite images. The integration of EfficientNet_mb-B0 and the BFM led to improved accuracy, showcasing the potential of our approach for advancing remote sensing applications and facilitating a better understanding of the Earth's surface and its dynamic changes.

Acknowledgement

This result was supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea (NRF) funded by the Ministry of Education(MOE)(2021RIS-003).

References

[1] S. Abburu and S. B. Golla, "Satellite image classification methods and techniques: A review," *International journal of computer applications*, vol. 119, no. 8, 2015.

[2] F. Luo, T. Zhou, J. Liu, T. Guo, X. Gong, and J. Ren, "Multiscale diff-changed feature fusion network for hyperspectral image change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–13, 2023.

[3] E. J. Blankson and B. H. Green, "Use of landscape classification as an essential prerequisite to landscape evaluation," *Landscape and Urban planning*, vol. 21, no. 3, pp. 149–162, 1991.

[4] J. T. Perron, J. W. Kirchner, and W. E. Dietrich, "Spectral signatures of characteristic spatial scales and nonfractal structure in landscapes," *Journal of Geophysical Research: Earth Surface*, vol. 113, no. F4, 2008.

[5] S. Talukdar, P. Singha, S. Mahato, S. Pal, Y.-A. Liou, and A. Rahman, "Land-use land-cover classification by machine learning classifiers for satellite observations—a review," *Remote Sensing*, vol. 12, no. 7, p. 1135, 2020.

[6] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.

[7] P. Helber, B. Bischke, A. Dengel, and D. Borth, "Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2019.

[8] J. W. Rouse, R. H. Haas, J. A. Schell, D. W. Deering *et al.*, "Monitoring vegetation systems in the great plains with erts," *NASA Spec. Publ*, vol. 351, no. 1, p. 309, 1974.

[9] A. Huete, H. Liu, K. Batchily, and W. Van Leeuwen, "A comparison of vegetation indices over a global set of tm images for eos-modis," *Remote sensing of environment*, vol. 59, no. 3, pp. 440–451, 1997.

[10] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geoscience and remote sensing magazine*, vol. 4, no. 2, pp. 22–40, 2016.

[11] Y. Zhong, Q. Cao, J. Zhao, A. Ma, B. Zhao, and L. Zhang, "Optimal decision fusion for urban land-use/land-cover classification based on adaptive differential evolution using hyperspectral and lidar data," *Remote Sensing*, vol. 9, no. 8, p. 868, 2017.

[12] J. Zhang, "Multi-source remote sensing data fusion: status and trends," *International Journal of Image and Data Fusion*, vol. 1, no. 1, pp. 5–24, 2010.

[13] I. Dimitrovski, I. Kitanovski, P. Panov, A. Kostovska, N. Simidjievski, and D. Kocev, "Aitlas: Artificial intelligence toolbox for earth observation," *Remote Sensing*, vol. 15, no. 9, 2023.