Hybrid Deep Learning Framework for Real-Time Border Surveillance using Satellite Image Intelligence and Geospatial Monitoring

Devanathan Muthukrishnan[1], Sai Kiran Oruganti[2]

[1]Postdoctoral Research Fellow, Lincoln University College, Petaling Jaya, Selangor, Malaysia.

[2]Faculty of Engineering and Built Science, Lincoln University College, Petaling Jaya, Selangor, Malaysia. saisharma@lincoln.edu.my

Corresponding Email: pdf.Devanathan@lincoln.edu.my

Abstract

Border security systems are significantly supported by geospatial intelligence. Recently developed border monitoring systems have utilized some specifics to analyze the images from satellites located in different places of the planet. Traditional means of image analysis may fail to deliver the required information in due time with due regard to what is going on. In the paper, a hybrid approach is proposed, which combines Convolutional Neural Networks and attention mechanisms of transformers for border monitoring. The major aim of the proposed approach is to identify border breaches, changes in the topology, suspicious activities, and any alterations of the environment. The approach will rely on deep learning techniques and transfer learning, and use some specific models, such as ResNet, EfficientNet, and ViT. The hybrid approach will include the use of a few techniques aimed at improving the awareness of using the models in the process. National and international benchmark datasets of satellite images will be used, including Sentinel-2 dataset and EuroSAT dataset. The focus of the present work is to establish a satellite surveillance framework that is intelligent enough and is scalable and can work along with the border, military and real-time geospatial security systems.

Keywords: Border surveillance, Satellite image intelligence, Deep learning, Geospatial intelligence, remote sensing.

I. INTRODUCTION

The clever evaluation of the satellite images received through remote sensing has undergone significant transformation due to the advanced technologies such as Artificial Intelligence and deep learning technology. In the context of research, deep learning technologies are highly beneficial in many ways; however, the technology most commonly used for the interpretation of satellite imagery is the so-called Convolutional Neural Network which makes it possible to achieve very good outcomes in the field of geospatial technologies [1]. The CNN technology is advantageous as it does not require any manual segmentation of the image data too. The role of satellite systems in the field of resource intelligence in border monitoring is important because of the importance of the field in general. One of the major reasons for the significance of the technology is its application for managing the possible threats and crimes at borders.

Conventional instruments of border control are based on manual surveillance and radar systems as well as weather surveillance systems. However, while such methods have been effective for many years, one of their main problems is that they do not respond well to challenges arising due to big amount of data and the ever-changing environment. This paper aims at creating smart geospatial monitoring tools

that are able to support border security activities and conduct operations in real time. Moreover, it fills existing knowledge gap and creates a framework for further development of satellite-based border monitoring and geospatial intelligence [2].

II. LITERATURE REVIEW

The emergence of deep learning has led to its adoption in the world of analysis of satellite images due to its success in obtaining the needed information through the large quantities of data which remote-sensing provides. The capabilities of this method allowed its usage in various fields ranging from studying ecosystems to military surveillance, disaster assessment and exploration [5]. Before deep learning era, remote-sensing technologies mainly relied on traditional techniques in machine learning, such as SVM (Support vector machines), Random forests and K-nearest neighbors. While their use in many cases has been effective, they do not perform well with regard to satellite images.

The emergence of deep CNNs has greatly influenced the sphere of satellite image processing. A number of architectures are known to succeed in resolving problems in this area, namely AlexNet, ResNet, DenseNet, EfficientNet, and U-Net. The issues with performance deterioration that generally occurs in deep networks have been addressed by ResNet through employing the concept of residual learning. EfficientNet has managed to achieve higher performance based on the application of a well-balanced scaling of neural networks, which allowed it to provide better results in terms of accuracy and efficiency than other types of architectures. Over the last few years, a number of important advances have been made in the field of transformers [3]. Transformers differ from CNNs, as they possess self-attention mechanisms, which enables them to recognize long-term dependencies within images and the wider context. Examples of transformers include Vision Transformer (ViT) and Swin Transformer, which are both suitable for land use assessment and other applications.

Transfer learning is already widely employed in the remote sensing sector nowadays. Since obtaining high-quality labeled datasets from satellite images is both an extremely costly and time-consuming process, researchers familiar with deep learning methods tend to apply pre-trained algorithms to simplify the training process. The methods of multi-modal data integration are active areas of research, where satellite imagery is integrated with other forms of data such as temporal, environmental, and geospatial data.

There are still many issues present in this area, although growth has been achieved. Although CNNs can be successfully used to gather information about the local spatial features, they neglect the use of contextual information [4]. At the same time, transformers help drawing out global features, yet are far from being cheap computationally. Furthermore, few publications are dedicated to border surveillance in the context of satellite imagery. Also, it is hard to exploit the said method in real time due to insufficient computing capacities and the lack of effectiveness of existing methods in different countries and regions. This paper presents an integrated approach that combines both CNN and transformer in terms of completing border monitoring with the help of satellite imagery. The objective of the proposed solution is to make the most of both methods for effective and precise surveillance during different weather conditions and in diverse geographical contexts.

III. PROBLEM STATEMENT

Border security and border surveillance system has become complex, as there are a large number of satellite images received from border areas. In order to monitor the border effectively, it is needed to

analyze these images in live mode. Traditional monitoring methods in terms of manual processing, radar systems, and image processing techniques do not provide reliable information for suspiciousness identification. Numerous factors affect the reception of satellite images from the territories under investigation, such as geographical features, weather conditions, cloudiness, illumination, seasons, and resolution of the image.

With the advent of technological advancements in machine learning, specifically in Convolutional Neural Networks (CNN)s, improvements have been made in the area of satellite image classification and object detection. However, at present, there are still some hindrances in the way. Although CNNs are effective in learning spatial properties, they are not very good at the analysis of the large context, which is very important in case of border situations. On the contrary, the application of Transformer models gives the opportunity to machines to learn long-distance links, but they require powerful computers for this purpose. Furthermore, the small volume of satellite image databases belonging to different regions and times can often lead to excessive overtraining and low capacity of generalization of the model. Besides, it is worth mentioning that the majority of other studies focused on land monitoring and environmental monitoring while border monitoring area was underestimated. Given all the factors mentioned above, this study presents a hybrid model based on CNN and Transformer architectures for the purpose of the real-time border monitoring.

IV. RESEARCH GAP

Furthermore, despite numerous advancements achieved by deep learning techniques in the realm of handling satellite images and remote sensing, there are still many challenges that remain unresolved. Most of the books on this subject concentrate on such matters as environmental control, land classification, and emergency management while not much attention is paid to border security surveillance, although there are techniques that can be applied in border intelligence. The process of border monitoring itself is complicated and involves precise mapping along with constant monitoring and analyzing of the situation. One of the greatest limitations of existing solutions is a heavy dependence on the CNN architecture. It is widely believed that CNNs simplify the extraction of local characteristics, but, at the same time, they make it impossible to obtain the global context needed for understanding the terrain and following the events occurring on it.

Another problem associated with the existing methods is insufficient access to a sufficient pool of data that could be utilized for training the models. Additionally, the achievement of consistent operation regardless of the specific border conditions remains a challenge. Furthermore, only a limited number of studies has been conducted to evaluate the performance of the combination of CNNs and transformers in border security technologies. Consequently, the research into the hybrid technologies and potential benefits of effective monitoring while using deep learning is still under basic development.

On the whole, the research presented in this paper is devoted to the outline of advanced hybrid deep learning system for border security provided by the use of CNNs and transformers technologies. The system utilizes transfer learning and self-supervised learning together with geospatial data fusion in order to provide high performance in the process of border detection.

V. RESEARCH OBJECTIVES

Our research sheds light on new approaches to effective border monitoring through the already existing use of satellite data and geospatial tools. Hence, we plan to enhance the safety/control of borders

by timely identifying unusual movements along the border and being informed about what is going on in nearby territories. We are going to determine whether it is possible to successfully merge the architecture of CNN and transformer models in order to achieve both spatial understanding and local features' processing. Finally, various approaches of transfer learning will be elaborated, mainly regarding the utilization of some already trained models such as ResNet, EfficientNet, and Vision Transformers (ViT).

In addition to that, the research will point out prospects of enhancing the efficiency of models based on using multi-level feature-learning technologies, self-supervised learning, data preparation, and data augmentation techniques. The idea of utilizing multimodal content derived from geospatial and environmental data will also be presented in the paper. Finally, optimization techniques will be described allowing to reduce the time of calculations and memory load.

VI. CONCLUSION

The present research proposes a hybrid technique using CNN and transformer approaches to carry out real-time border monitoring based on satellite data, remote sensing, and other geospatial data. Various methods have been engaged, such as image processing, generating new images, transfer learning, understanding the context and optimizing models, which has resulted in the border monitoring system that uses the new approach. The importance of the proposed system is in using both CNN and transformer approaches for processing satellite images that are rather complex. The goal of the new technique is aimed at improving risk assessment, land monitoring, and geospatial intelligence acquiring with the help of satellite data. The system attempts to combine the effectiveness of detection and efficiency of computation.

REFERENCES

[1] S. Dong, L. Wang, B. Du, and X. Meng, "ChangeCLIP: Remote sensing change detection with multimodal vision-language representation learning," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 208, pp. 53–69, 2024, doi: 10.1016/j.isprsjprs.2024.01.004.

[2]Z.-H. You, S.-B. Chen, J.-X. Wang, and B. Luo, "Robust feature aggregation network for lightweight and effective remote sensing image change detection," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 215, pp. 31–43, 2024, doi: 10.1016/j.isprsjprs.2024.06.013.

[3] "ChangeRD: A registration-integrated change detection framework for unaligned remote sensing images," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 220, pp. 64–74, 2025, doi: 10.1016/j.isprsjprs.2024.11.019.

[4] P. Lv, W. Wu, Y. Zhong, F. Du, and L. Zhang, "SCViT: A spatial-channel feature preserving vision transformer for remote sensing image scene classification," IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–12, 2022, doi: 10.1109/TGRS.2022.3157671.

[5]X. Tang, M. Li, J. Ma, X. Zhang, F. Liu, and L. Jiao, "EMTCAL: Efficient multiscale transformer and cross-level attention learning for remote sensing scene classification," IEEE Trans. Geosci. Remote Sens., vol. 60, 5626915, 2022, doi: 10.1109/TGRS.2022.3194505.