© June 2024 | IJIRT | Volume 11 Issue 1 | ISSN: 2349-6002

Semantic Segmentation in VHR Remote-Sensing Images

AVDHUT SATISH PATIL[1], PRUTHVIRAJ KISAN ZINJADE[2], NIKHIL VIJAY DONGRE[3], SANKET SANDESH RATHOD[4]

[1, 2, 3, 4] B.E. Student, Dept. Computer Engineering, Sinhgad Institute of Technology, Lonavala (MH), India.

Abstract— The interactions between target features are more intricate in high-resolution remote sensing images, which encompass more feature in- formation such as texture, structure, geometry, and other geometric elements. These characteristics impede the process by which stan- dard convolutional neural networks execute feature classification on remote-sensing images and achieve optimal results. We suggested DMAU-Net, an attention-based multiscale maxpooling dense net- work based on U-Net for ground object categorization, as a solutionto this problem. The network is built with an inbuilt max-pooling module that uses dense connections in the encoder section to im- prove the feature map's quality and, consequently, the network's ca- pacity to extract features. Similarly, we present the Efficient Chan- nel Attention (ECA) module in the decoding process, which can am-plify the useful elements and stifle the superfluous ones.

Index Terms—High-resolution remote-sensing images, ground object classification, dense connections, Multiscale maximum pooling, Semantic segmentation, CNN algorithm

I. INTRODUCTION

Objective

Accurate and efficient semantic segmentation of Very High-Resolution (VHR) remote-sensing images is a critical task in applications such as urban planning, environmen- tal monitoring, and disaster management. VHR images contain rich details, and tra-ditional convolutional neural networks (CNNs) often struggle to capture fine-grainedinformation and spatial context while maintaining computational efficiency. The objectives of an attention-based multiscale max-pooling dense network for semantic segmentation in very high-resolution (VHR) remote-sensing images can be multifaceted, addressing challenges specific to high-resolution remote sensing data Improved Semantic Segmentation Accuracy.Enhance the accuracy of semantic segmentation in VHR remote-sensing imagesby everaging attention mechanisms and multiscale features. Effective Handling of HighResolution Data: Develop techniques to handle the challenges posed by high-resolution remote sens- ing data, such as capturing fine-grained details and handling large input sizes effi- ciently. Attention Mechanisms for Relevance: Integrate attention mechanisms to allow the network to focus on relevant regions of the image, improving the model's ability to distinguish between different classes in complex scenes.

Overview

For monitoring geographic national circumstances, detecting changes in land cover, and monitoring changes in land use dynamics, remote sensing image change detec-tion is an essential tool. observing. Three types exist: semi-automatic, automatic, and manual. techniques. The manual approach depends on the translator's aware- ness. It offers the benefits of increased precision and precision, as well as its draw- backs of poor efficiency additionally large labour expenses. The approaches that aresemi-automated and automaticare the developing ones, separated into two groups: Both object- and pixel-level.

The generalized prediction model can be given as:

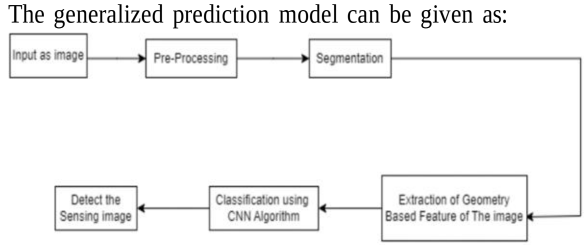


Fig 1: Prediction Model

Input (Image)

VHR remote sensing images are usually captured by satellites or aerial platforms. These images contain a large number of pixels, with each pixel representing a tiny area on the Earth's surface. The images often cover large geographic areas and capture fine details due to their high spatial resolution.

Data pre-processing

Preprocessing refers to all the transformations on the raw data before it is fed to the machine learning or deep learning algorithm. For instance, training a convolutional neural network on raw images will probably lead to bad classification performances. Before performing semantic segmentation, preprocessing steps might be applied to the input images. This can include tasks like radiometric correction, geometric correction, and normalization to enhance the quality and consistency of the images.

Training
The semantic segmentation model is trained using a dataset of labeled VHR images. During training, the model learns to recognize patterns and features associated with different classes in the images. This involves optimizing the model's parameters to minimize a loss function that measures the difference between the predicted segmentation and the ground truth labels.

Inference
Once the model is trained, it can be used to perform semantic segmentation on new, unseen VHR images. During inference, the model processes each pixel in the input image and assigns it a class label based on the learned patterns and features.

Processing
Post-processing techniques may be applied to the segmentation results to refine and improve their accuracy. This can include methods like smoothing, morphological operations, and spatial filtering to remove noise and artifacts from the segmentation output.

Output
Semantic segmentation in very high-resolution (VHR) remote sensing output refers to the result of classifying each pixel in a VHR image into specific categories or classes, such as roads, buildings, vegetation, water bodies, and so on.

Material
- VHR Remote Sensing Imagery
- Ground Truth Data
- Programming Language: Python

- Methods
- Deep learning techniques (particularly convolutional neural networks (CNNs)).
- Transfer Learning
- Data Augmentation
- Post-processing

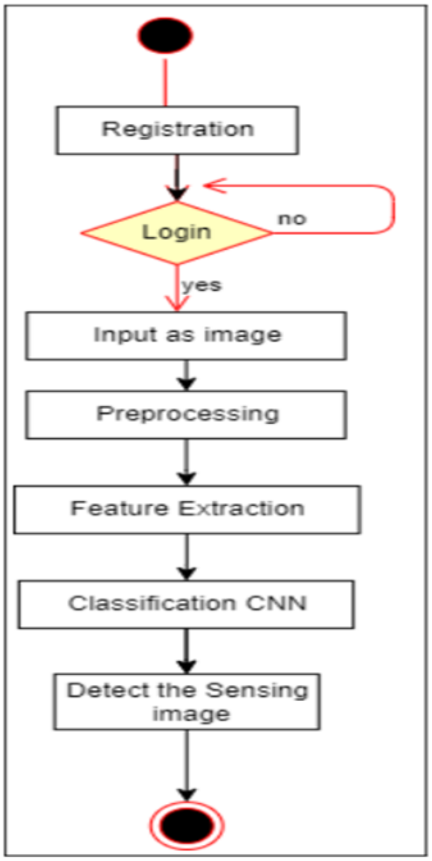
Tools
- Deep Learning Frameworks
- Remote Sensing Software

Proposed work: Remote Sensing Image
Clearly define the problem of semantic segmentation in VHR remote sensing. Specify the research objectives, such as improving accuracy, exploring novel architectures, or addressing specific challenges in the field. Conduct an extensive review of existing literature related to semantic segmentation in VHR remote sensing. Identify state-of-the-art methods, key challenges, and recent advancements in the field. Analyze the strengths and limitations of previous approaches and identify gaps for potential contributions.

Acquire VHR remote sensing datasets suitable for semantic segmentation. Preprocess the datasets, including radiometric correction, geometric correction, and normalization. Annotate the datasets with ground truth labels for training and evaluation.
Select appropriate deep learning architectures for semantic segmentation, considering factors such as model complexity, computational efficiency, and suitability for VHR imagery. Develop and implement the selected models using deep learning frameworks like TensorFlow or PyTorch. Explore techniques such as transfer learning, data augmentation, and regularization to enhance model performance.

Workflow implementation
The flowchart depicts a process for an image-based system, likely for biometric recognition. It begins with user registration, where data is captured and stored. Upon subsequent login attempts, the system checks if the user is registered. If not, it loops back to registration. If the user is registered, their input is taken as an image, which undergoes preprocessing to enhance quality or normalize the data. The preprocessed image is then subjected to feature extraction, isolating critical characteristics necessary

for identification. These features are passed through a Convolutional Neural Network (CNN) for classification, determining if the input matches the registered data, ultimately detecting and verifying the sensing image, indicating successful recognition or access.



Machine Learning

Annotated datasets containing VHR remote sensing images paired with pixel-level labels indicating the class of each pixel (e.g., road, building, vegetation) are essential for training a semantic segmentation model. Data preprocessing steps such as normalization, augmentation, and cropping are often applied to the training data to enhance model performance and generalization. Convolutional Neural Networks (CNNs) are the most commonly used architectures for semantic segmentation due to their ability to automatically learn hierarchical features from input data. Many semantic segmentation models follow an encoder-decoder architecture, where the encoder extracts features from the input image, and the decoder generates a pixel-wise segmentation map based on the extracted features. Once the model is trained, it can be used to predict the class labels for unseen VHR remote sensing images. During inference, the model processes each pixel in the input image and assigns it a class label based on the learned patterns and features.

Post-processing techniques such as conditional random fields (CRFs), morphological operations, and spatial filtering may be applied to the segmentation results to refine and improve their accuracy. Thresholding techniques can be used to convert pixel-wise probability maps into binary segmentation masks by selecting a threshold value above which a pixel is considered to belong to a certain class.

Object Detected

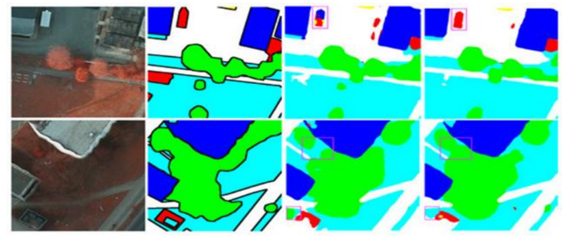
In semantic segmentation of remote sensing images, CNNs (Convolutional Neural Networks) play a critical role by automatically learning and extracting hierarchical features from the input images. These features include low-level attributes like edges and textures, and high-level representations such as shapes and object parts. By applying multiple convolutional layers, the CNN can capture spatial and contextual information, enabling it to distinguish between different classes of land cover or objects in the image. This process results in a pixel-wise classification, where each pixel in the remote sensing image is labeled with a specific category, facilitating detailed and accurate analysis of the geographic area. Semantic segmentation in remote sensing images involves the use of CNNs to achieve detailed, pixel-level classification of various elements in a scene, such as buildings, roads, vegetation, and water bodies. The process starts with the CNN taking an input image and passing it through several convolutional layers, each of which applies filters to detect specific patterns or features. Early layers might focus on basic features like edges and corners, while deeper layers identify more complex structures and contextual relationships. By combining information from multiple layers, the CNN constructs a comprehensive understanding of the image, assigning a class label to each pixel. This hierarchical feature extraction enables precise delineation of different land cover types, aiding in tasks such as urban planning, environmental monitoring, and disaster management. The output is a segmented map where each pixel's classification is

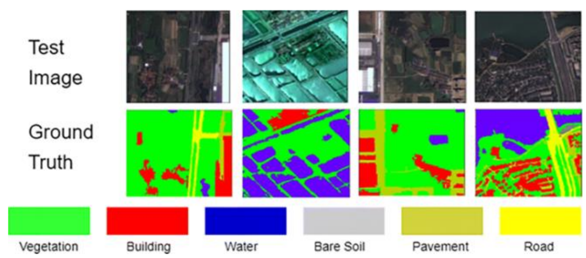
informed by both local details and broader contextual information captured by the CNN.
The CNNs model used to detect the following objects from the input image and set different colours to it

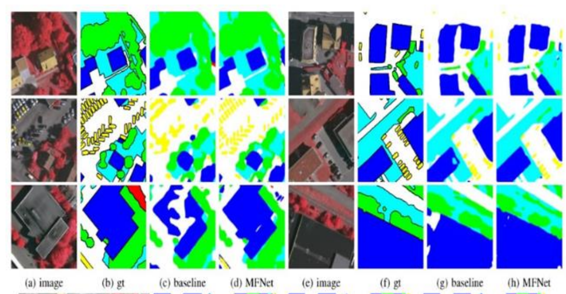
- Buildings
- Roads
- Trees
- River
- Cars

Implementation and Results
In the labeled image, each pixel is colored according to its assigned class label. Common classes include roads, buildings, vegetation, water bodies, vehicles, and other objects of interest. Implementing a project using CNNs for semantic segmentation of remote sensing images involves several key steps. First, you'll need a labeled dataset, where each pixel in the images is annotated with a class label. Popular datasets like ISPRS or custom annotated data can be used. Preprocess the images to ensure consistent input size and normalize pixel values. Choose a CNN architecture suitable for segmentation tasks, such as U-Net, SegNet, or DeepLab. These architectures typically include an encoder-decoder structure where the encoder captures spatial features through convolutional layers and the decoder reconstructs the pixel-wise classification map. Implement the chosen architecture using a deep learning framework like TensorFlow or PyTorch. Train the model on your dataset, utilizing techniques such as data augmentation to improve robustness and regularization methods like dropout to prevent overfitting. During training, use a loss function appropriate for segmentation tasks, such as cross-entropy loss or Dice coefficient loss, to guide the optimization process. Monitor performance metrics like Intersection over Union (IoU) to evaluate accuracy. Once the model is trained, validate it on unseen data to ensure generalization. Finally, deploy the trained model to perform semantic segmentation on new remote sensing images. This involves integrating the model into a pipeline that handles image preprocessing, segmentation, and post-processing steps like smoothing the output to reduce noise, providing a practical tool for applications such as urban planning. Different colors or shades are used to represent different classes, making it easy to visually interpret the segmentation output. The accuracy and quality of the segmentation result depend on factors such as the choice of model architecture, the quality of training data, and the complexity of the scene being segmented. Evaluation metrics such as pixel accuracy, intersection over union (IoU), and F1 score are used to quantify the performance of the segmentation model and assess the similarity between the predicted segmentation and ground truth labels.

Result







CONCLUSION

Deep learning-based change detection in the field of remote sensing has garnereda lot of interest lately and has shown promising results. Unlike traditional feature- based methods that require human intervention, deep learning techniques may au- tomatically identify intricate elements in remote sensing photos by utilising an ex- tensive array of hierarchical layers.This effort involved a thorough metaanalysis of literature pertaining to DL in remote sensingImage, The project focused on implementing

Convolutional Neural Networks (CNNs) for semantic segmentation of remote sensing images, aimed at achieving detailed, pixel-level classification for various applications such as urban planning, environmental monitoring, and disaster management. Through the systematic process of dataset preparation, model selection, training, and validation, the project demonstrated the efficacy of CNN-based approaches in accurately identifying and categorizing different elements within a geographic area. Initially, the project emphasized the importance of a well-labeled dataset, recognizing that the quality and quantity of annotated data are pivotal to training an effective segmentation model. Using datasets like ISPRS or custom-annotated data, the images were preprocessed to ensure consistency in input size and normalization of pixel values, laying a solid foundation for the subsequent modeling steps. The selection of a suitable CNN architecture, such as U-Net, SegNet, or DeepLab, was critical. These architectures, known for their encoder-decoder structures, are adept at capturing spatial features and reconstructing pixel-wise classification maps, which are essential for precise segmentation tasks.

The implementation phase leveraged a deep learning framework like TensorFlow or PyTorch to construct the chosen model. During training, techniques such as data augmentation were employed to enhance the model's robustness by simulating a variety of real-world conditions. Regularization methods like dropout were utilized to mitigate overfitting, ensuring the model's generalizability to new data. The training process was guided by appropriate loss functions, such as cross-entropy loss or Dice coefficient loss, and performance was monitored using metrics like Intersection over Union (IoU), which provided a quantitative measure of the model's accuracy.

Upon completion of the training phase, the model was validated on unseen data to assess its generalization capability. The validation results indicated that the CNN-based approach was effective in achieving high accuracy in pixel-wise classification, successfully distinguishing between different land cover types. The performance metrics demonstrated significant improvements over traditional image processing techniques, highlighting the potential of deep learning models in handling the complexities of remote sensing imagery.

The deployment phase involved integrating the trained model into a practical pipeline capable of processing new remote sensing images. This pipeline handled tasks ranging from image preprocessing and segmentation to post-processing, such as smoothing the output to reduce noise. The deployed system proved to be a valuable tool for various applications, offering precise and reliable segmentation maps that facilitate informed decision-making in urban planning, agricultural monitoring, and disaster response.

In conclusion, the project underscored the transformative impact of CNNs in the domain of remote sensing. The ability of these models to learn and extract hierarchical features from images enabled them to perform detailed and accurate semantic segmentation, surpassing the limitations of conventional methods. The success of the project not only demonstrated the feasibility of CNN-based segmentation for remote sensing images but also highlighted the broader potential of deep learning technologies in advancing the field of geographic information systems (GIS). Future work could explore the integration of additional data sources, such as LiDAR or multispectral imagery, to further enhance the segmentation accuracy and expand the range of applications. The continuous advancements in deep learning architectures and computational resources promise even greater improvements in the precision and efficiency of remote sensing analysis, paving the way for more sophisticated and scalable solutions in the years to come.

REFERENCES

[1] Dibs, H., Mansor, S., Ahmad, N. and Al-Ansari, N. (2023) Robust Radiometric Normalization of the near Equatorial Satellite Images Using Feature Extraction and Remote Sensing Analysis. Engineering, 15, 75-89. https://doi.org/10.4236/eng.2023.152007

[2] Dibs, H., Mansor, S., Ahmad, N., Pradhan, B. and Al-Ansari, N. (2020) Automatic Fast and Robust Technique to Refine Extracted SIFT Key Points for Remote Sensing Images. Journal of

Civil Engineering and Architecture, 14, 339-350. https://doi.org/10.17265/1934-7359/2020.06.005

[3] Hashim, F., Dibs, H. and Jaber, H.S. (2022) Adopting Gram-Schmidt and Brovey Methods for Estimating Land Use and Land Cover Using Remote Sensing and Satellite Images. Nature Environment and Pollution Technology, 21, 867-881.

[4] Su, Y.; Cheng, J.; Bai, H.; Liu, H.; He, C. Semantic Segmentation of Very-HighResolution Remote Sensing Images via Deep Multi-Feature Learning. Remote Sens. 2022, 14, 533. [CrossRef]

[5] S. Morgan, ―Cybercrime Damages Trillion By 2021‖, Cybercrime Magazine Available

[6] J. D'Souza, ―Let's learn about AUC ROC Curve!‖, Greyatom, [Online]. Avail-able:

[7] S. Gupta, ―Decision Tree‖, Hackerearth, [Online]. Available:

[8] Yeo, M., Koo, Y., Yoon, Y., Hwang, T., Ryu, J., Song, J., Park, C., ―Flow-based malware detection using convolutional neural network‖, 2018 International Conference on Information Networking, 2018.

[9] Rhode, M., Burnap, P., Jones, K., ―Early-stage malware prediction using re-current neural networks‖, computers 2018