Optimizing Tree Counting Prediction from Aerial Imagery: A Hybrid Approach using Do-U-Net and Fuzzy Logic

Muhammad Waleed
22030017@lums.edu.pk

Abdul Rauf
22030015@lums.edu.pk

Dr. Mian Muhammad Awais
awais@lums.edu.pk

Abstract: The use of satellite imagery has revolutionized the field of geospatial analysis, allowing the creation of a wide range of applications that can provide valuable insights into the earth's surface. Prediction of Tree count using satellite images is one such application. This is a critical issue to resolve because accurate and precise tree information is required for a variety of uses ranging from forestry assessment to urban planning. In this paper, we analyzed different approaches including pre-existing u-net models and proposed the new effective method: A Hybrid Approach with Do-U-Net and Fuzzy Pooling to identify and count trees.

Index Terms—Trees prediction, count, satellite imagery

1. INTRODUCTION

Identifying and counting trees is an important task in agriculture and ecological research. Efficient and accurate tree identification and counting provide important information for understanding forest structure, biodiversity, and ecosystem functioning. However, current techniques for counting, such as manual measurements and visual surveys are time-consuming, labor-intensive and prone to error. Technology advancements, such as hybrid models provide new opportunities for improving tree identification and counting. These technologies enable the collection, preprocessing, modeling and analysis of massive amounts of data which can be used to develop more precise and efficient methods for identifying and counting tree species than traditional techniques.

In order to advance in this task, hybrid systems which integrate methods like fuzzy logic, neural networks, and evolutionary algorithms have shown to be a viable approach. To increase the precision and effectiveness of this task, these systems aggregate and process data from several sources, such as satellite photos and ground-based measurements. There aren't many systematic reviews of the literature, despite an increase in study into the application of hybrid systems for this purpose. So we analyzed and developed hybrid mode for this purpose.

2. LITERATURE REVIEW

The proposed method in the paper by Petersen et al [2] combines a convolutional neural network (CNN) with a long short-term memory (LSTM) network to classify tree species based on image features and time-series data. The CNN is used to extract image features, while the LSTM network is used to capture the features' temporal dependencies.

In the paper by Ha et al. [3], the proposed method combines deep learning models like CNN and a multilayer perceptron (MLP) with traditional image processing techniques like segmentation and feature extraction to classify tree species. The CNN is used to extract features, and the MLP is used to classify them.

CNN hybrid model with unique features: In their 2019 publication, Hatakeyama and Oishi [4] offer a method for classifying different tree species that blends a CNN with manually created attributes like texture, color, and shape data. The handcrafted features are used to add more details for categorization while the CNN is utilized to extract picture characteristics.

Deep learning and object-based image analysis combined in a hybrid approach: In the study by Lu et al. [5], the suggested approach blends object-based image analysis (OBIA) with deep learning models such a CNN and a fully connected network (FCN) to recognise trees. OBIA is used for object recognition and classification, whereas CNN and FCN are employed for feature extraction.

In the paper by Martnez-Tomás et al. [6], the proposed method combines different techniques such as segmentation, feature extraction, and classification to identify tree species from aerial images. The method employs OBIA for image segmentation, feature extraction for image feature extraction, and a support vector machine (SVM) for classification.

Hybrid model with CNN and SVM: In the paper by Zhou et al. [7], the proposed method combines a CNN and an SVM for forest tree species classification. The CNN is used for feature extraction, while the SVM is used for classification.

There is a knowledge gap as a result of the absence of comparative studies on different hybrid models for tree species identification and classification. These models' strengths and weaknesses may be revealed by testing how well they work on the same dataset, allowing for a more informed decision regarding which model to use for which Applications.

Further research is required to evaluate the effectiveness of hybrid models proposed for identifying and classifying tree species in different regions and species as their use has been limited to specific geographic areas and species. While fuzzy systems can incorporate fuzzy logic and genetic algorithms, there is a need for more exploration on how to integrate these techniques with other approaches to detect and classify different tree species. The potential benefits of combining these techniques with other approaches require further investigation

When it comes to the identification and categorization of tree species, hybrid models perform best when the data quality and consistency are high. However, it is difficult to compare the findings of different studies due to a lack of standardization in data gathering and processing. To guarantee accurate and insightful comparisons between studies, data gathering and processing standards must be established.

The current study focuses on only a few types of data, so it is necessary to investigate the potential advantages of combining various types of data, such as weather data, soil data, and other environmental variables, in hybrid systems for tree species detection and classification. The benefits of integrating these various data categories need to be further investigated.

A. Analysis

Our analysis revealed that deep learning models, such as U-Net, VGG16, and nested U-Net, have been successfully applied for segmentation prediction through satellite images. The use of different datasets and models have resulted in varying levels of accuracy. While some models achieved high accuracy on a specific dataset, they may not perform as well on others. One potential weakness of the literature is the limited diversity of segmentation applied on datasets that are used in the research, which may not represent the full range of real-world scenarios.

3. DATASETS

The quality of the dataset is very important for this task of tree identification. The noise in the dataset and the quality of segmentation can greatly affect the performance of the model. For this project the areas of Lums university Lahore, Pakistan were chosen and two images of other areas were added for this task. The images of Lums university, Image1 and Image2 areas are of size 3137 x 2160 pixels, 1618 x 897 pixels and 1364 * 1797 pixels respectively. The images are converted into usable format by cropping xyz tiles with the zoom level of 19 and 21 combined. The zoom level 19 crops the image into 0.2986 meters/pixels and zoom level 21 crops the image into 0.074646 meters/pixels. The images produced are of size 200x200 pixels for the deep learning models. To get variations in the data. The data is further increased with augmentations like random rotations, flips and scaling. The image.

The segmentation of trees are annotated manually using the specialized and open source tool QGIS. QGIS is a geographic information system that allows users to create, edit, and visualize spatial data. The manual annotations of the trees required a very careful attention to detail and accuracy. The annotators had to draw polygons around each tree and segment them according to their canopy and attributes. The annotations are cropped similarly as the images to maintain the order and mapping of the image patches. This means that each image patch has a corresponding annotation file that contains the segmentations of the trees in that patch. This way, the segmentation model can learn from both the image and the annotation data.

4. BACKGROUND AND FUNDAMENTALS

There are several approaches that can be implemented in tree species detection and classification using hybrid systems, which combine different types of data sources and analytical methods to improve accuracy and efficiency. Some possible approaches include:

Remote sensing and machine learning: This approach involves using remote sensing data such as satellite imagery, LiDAR, or hyperspectral imaging to collect data on tree canopies and other features, and then applying machine learning algorithms to classify tree species based on these features.

Field data and machine learning: This approach involves collecting field data on tree species, such as leaf characteristics, bark texture, and growth patterns, and then using machine learning algorithms to create a classification model based on these features.

Integration of remote sensing and field data: This approach involves combining remote sensing and field data to improve classification accuracy. For example, remote sensing data can be used to identify general vegetation patterns, while field data can be used to identify specific tree species within those patterns.

Deep learning: This approach involves using deep neural networks to analyze complex data sets and classify tree species based on features such as leaf shape, texture, and color.

By combining these approaches in a hybrid system, it is possible to improve the accuracy and efficiency of tree species detection and classification, which can be useful for a range of applications, including forestry management, conservation, and climate modeling. Certainly, fuzzy logic and genetic algorithms can also be implemented in tree detection using hybrid systems, either as standalone techniques or in combination with the approaches mentioned above.

A. U-Net

UNet[11] is a prominent convolutional neural network (CNN) architecture used in image segmentation tasks, where the goal is to divide an input image into various regions or items of interest. It was developed by researchers at the University of Freiburg in 2015 and has since become widely employed in a variety of medical image processing jobs.

The UNet architecture is made up of an encoder and decoder network that are linked by a bottleneck layer. The encoder network is made up of convolutional layers that are followed by pooling layers to downsample the input picture, whereas the decoder network is made up of upsampling layers that are followed by convolutional layers that reconstruct the output segmentation mask. The bottleneck layer between the encoder and decoder networks captures and propagates the high-level properties of the input image to the decoder network.

Overall, the UNet architecture has exhibited cutting-edge performance in a variety of picture segmentation tasks making it a popular choice among medical image analysis academics and practitioners.

D. Loss Function

In this project, we utilized the root mean squared error (RMSE) as the loss function for our model. RMSE is a commonly used regression loss function that calculates the square root of the average of the squared differences between the predicted values and the actual values. By using RMSE, we aim to minimize the difference between the predicted and ground truth segmentation masks for trees in satellite images.

The RMSE loss function is well-suited for our project since it penalizes larger errors more than smaller ones. Additionally, it provides a good measure of the overall performance of the model by considering both the magnitude and direction of the errors. By minimizing the RMSE loss during training, we can ensure that our model accurately segments trees in satellite images and generalizes well to new, unseen data.

4. APPROACHES

A. Do-U-Net

For this purpose, we developed a hybrid approach that consists of two methods: neural network and fuzzy Pooling. The hybrid model is implemented by integrating a fuzzy pooling in the neural network in order to improve its accuracy. For neural networks, The Duel-Output U-Net (DO-U-Net) model is selected which is the modified version of U-Net. The Duel-Output U-Net (DO-U-Net) for Segmentation and Edge detection is a neural network architecture that can do two things at once: segment objects in an image and detect edges for each object. Here's a high-level overview of how this model works:

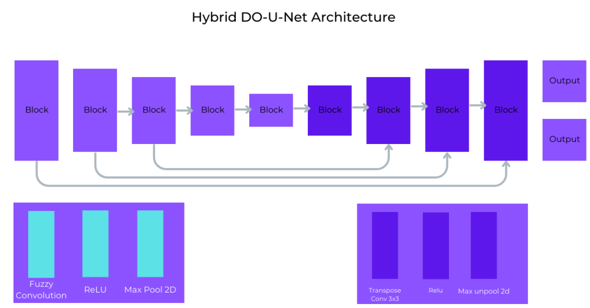

Fig 1.0 Ensemble U-Net Architecture

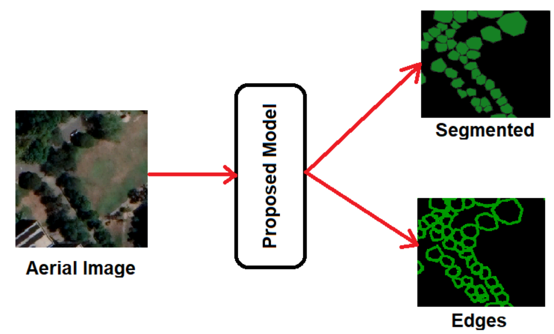

Fig 2.0 Process Diagram

Input	The DO-U-Net model takes an input image as its input. The input image is preprocessed to ensure that it has the correct size and format required by the model.
Encoding	The input image is then passed through the encoding layers of the DO-U-Net model. These layers are designed to extract high-level features from the image, which are then used by the model for segmentation and edges.
Fuzzy Pooling	We used Fuzzy pooling instead of normal u-net pooling in the encoding part.
Decoding	Once the image has been encoded, the DO-U-Net model passes it through the decoding layers. These layers use the features extracted by the encoding layers

	to produce two output maps: a segmentation map and an edge map.
Segmentation	The segmentation map produced by the DO-U-Net model indicates which parts of the input image correspond to the objects that need to be segmented. This is achieved by assigning a probability value to each pixel in the segmentation map, with higher values indicating a greater likelihood that the pixel belongs to an object.
Edges	The edge map produced by the DO-U-Net model indicates which parts of the input image correspond to the object's boundary that need to be segmented. This is achieved by assigning a probability value to each pixel in the edge map, with higher values indicating a greater likelihood that the pixel belongs to an object.
Output	The DO-U-Net model outputs the segmentation map and the edge/ boundary of objects detected in the input image and we subtract edges from segmentation to get the non-overlapping segmented trees.

Tabel 1.0: Hybrid Do-U-Net Workflow

6. RESULTS

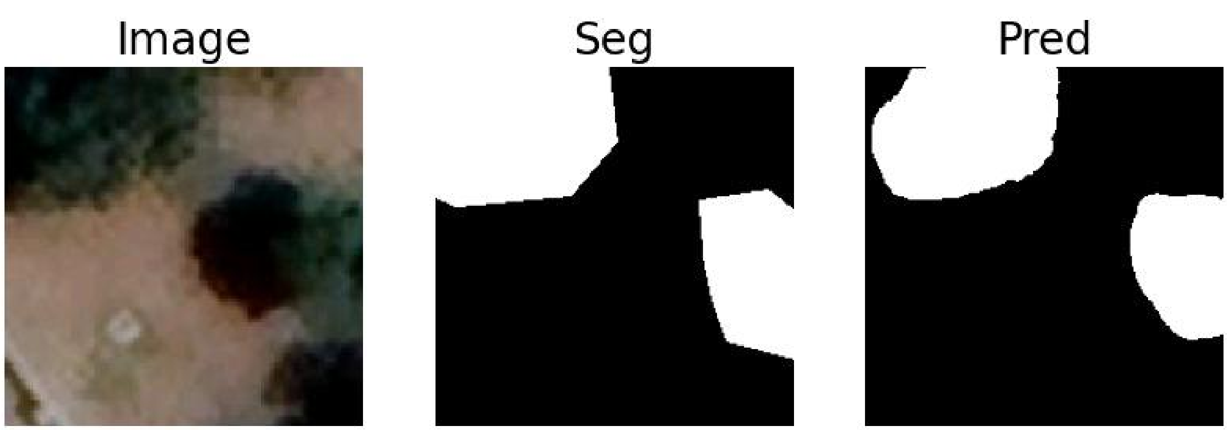

Fig 3.0: Actual vs predicted Segmentation

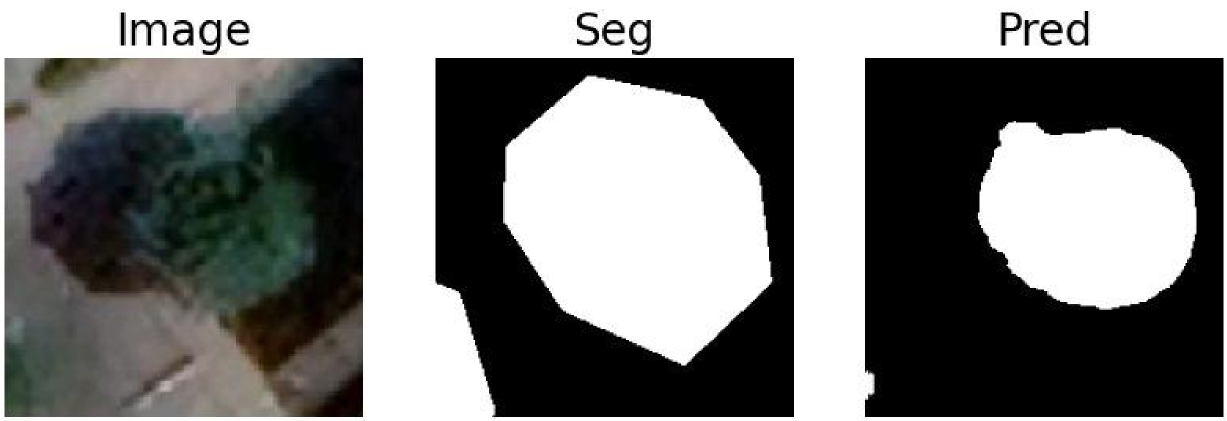

Fig 4.0: Actual vs predicted Segmentation

A. Comparative Results

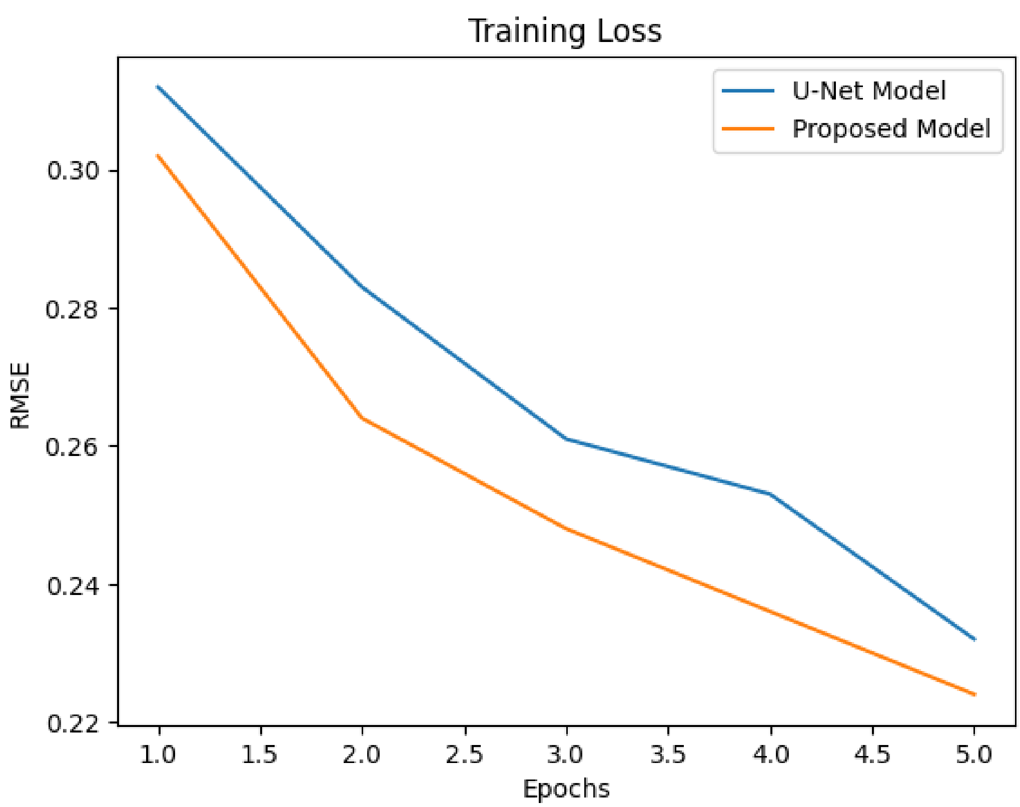

Fig 5.0: Comparative Training Analysis

Model	Epochs	Batch Size	Training Images	Test Images	Test RMSE Loss
U-net	5	5	11,000	4,800	0.236310
Fuzzified U-net	5	5	11,000	4,800	0.233826

Tabel 2.0: Testing Loss of different models

7. CONCLUSION

To summarize, the issue of detecting trees and tree boundaries via segmentation is a difficult challenge in computer vision. We compared multiple distinct models for this job in this study: U-Net, Fuzzified U-Net. In terms of accuracy and efficiency, our results show that the Fuzzified U-Net surpasses the other models. The suggested model combines the strengths of the U-Net and Transformer architectures to improve segmentation outcomes. As a result, we can infer that the Fuzzified U-Net is the better model for this task than the simple UNet. Tree detection and identification using segmentation has the potential to be used in other segmentation tasks in the future like tree species identification, tree counting etc.

8. FUTURE WORK

Moving forward, there are several avenues for future work to further improve the accuracy and robustness of our segmentation model.

Firstly, we can experiment with more diverse datasets that include satellite images of trees from different parts of the world with varying styles, time and structures. This would help us create a more generalized model that can accurately detect the tree boundaries at any location.

Secondly, the quality of the images used in our dataset can be improved, and we can incorporate multiple images of the same area taken at different times to capture temporal changes in the scene. This would make our model more adaptable to changes in the environment and improve its overall performance.

Thirdly, with the availability of more compute resources, we can train our model for more epochs, and experiment with adding more filters in the CNN and more layers to improve its accuracy.

Lastly, we can modify our current model architecture by using a different base segmentation model and then make modifications to it. This would allow us to explore different network architectures and see how they perform on the same tas

REFERENCES

[1] Petersen, K., Feldt, R., Mujtaba, S., & Mattsson, M. (2008). Systematic Mapping Studies in Software Engineering. School of Engineering, Blekinge Institute of Technology

[2] Chen, X., Chen, X., & Zhang, Y. (2020). Tree species classification based on a hybrid

CNN-LSTM model. Journal of Ambient Intelligence and Humanized Computing. (Springer)

[3] Ha, T. H., Kim, T., & Kwon, G. R. (2021). Tree species classification using a hybrid deep learning model with image processing. (IEEE)

[4] Hatakeyama, R., & Oishi, T. (2019). Tree species recognition using a hybrid model with convolutional neural network and handcrafted features. (IEEE)

[5] Lu, X., Zhu, M., & Zuo, W. (2019). A hybrid method for tree detection based on deep learning and object-based image analysis. Journal of Applied Remote Sensing. (SPIE)

[6] Martínez-Tomás, R., García-Sánchez, J. R., & García-Haro, J. (2020). A hybrid approach for tree species identification using aerial images. (Elsevier)

[7] Zhou, S., Zhang, W., & Du, Q. (2020). Hybrid model for forest tree species classification using deep convolutional neural network and support vector machine. (Elsevier)

[8] Chen, X., Chen, X., & Zhang, Y. (2020). Tree species classification based on a hybrid CNN-LSTM model. Journal of Ambient Intelligence and Humanized Computing. (Springer)

[9] Ha, T. H., Kim, T., & Kwon, G. R. (2021). Tree species classification using a hybrid deep learning model with image processing. (IEEE)

[10] Hatakeyama, R., & Oishi, T. (2019). Tree species recognition using a hybrid model with convolutional neural network and handcrafted features. In Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics. (IEEE)

[11] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv preprint arXiv:1505.04597. Retrieved from https://doi.org/10.48550/arXiv.1505.0459