ENHANCING TREE COUNTING ACCURACY USING MACHINE LEARNING MODELS

Niranjana Shaji, Abhinaya J and Saravanan S
niranjana.shaji2021@vitstudent.ac.in, abhinaya.2021@vitstudent.ac.in, saravanan.s2021@vitstudent.ac.in

Abstract— In order to improve tree counting accuracy, this study investigates the use of Convolutional Neural Networks (CNNs) for tree detection and classification in satellite data. We compared several CNN architectures and preprocessing methods (RGB vs. grayscale pictures) while training models on tree images using TensorFlow and Keras. Our findings provide insights into the model complexity needed for efficient tree detection in satellite data by showing that particular configurations increase accuracy and training time. Applications for scalable environmental monitoring and forest management may benefit from this strategy.

Keywords— CNN, TensorFlow, Kera's, Greyscale,Convo-2D,Greyscale,Ensembling

1. INTRODUCTION

With the importance of resource management and environmental monitoring growing, precise tree counting from satellite imagery offers useful information for the agricultural, urban planning, and conservation industries. The goal of this project is to use the TensorFlow and Kera libraries to create an optimal CNN model for recognizing and counting trees in satellite photos. We seek to strike a balance between model accuracy and computing efficiency by experimenting with various CNN architectures and training setups.
In order to identify a model configuration that offers the best balance between accuracy and computing efficiency, this study looks at many CNN architectures, building on earlier work in remote sensing and picture categorization. We applied different model configurations with variations in color processing (grayscale versus RGB) and architecture depth using a collection of annotated satellite pictures. By assessing precision,

2. MOTIVATION

Traditional manual tree-counting techniques are time-consuming and ineffective when applied to wide geographic areas. Although model quality and processing speed are still difficult issues, automated tree detection using satellite data and deep learning offers a scalable approach. This study aims to enhance these models by examining and refining CNN layers and color modes, offering useful information for practical applications where resource limitations might exist.

3. LITERATURE SURVEY

1. Tong, P. et al. (2021) presented an innovative method known as the Point-Supervised Segmentation Network (PSS Net), which is tailored for the counting and localization of trees through weakly supervised deep learning techniques. This network integrates a unique segmentation framework that requires minimal annotations, thereby offering an efficient and cost-effective solution. The results indicate that this approach surpasses both fully supervised and weakly supervised techniques, especially in intricate settings characterized by overlapping canopies and diverse tree shapes, leading to enhanced accuracy in tree counting.

2. Moharram, D. et al. (2022) developed a model based on YOLOv5s for the detection and counting of tree seedlings, specifically targeting three species: dragon spruce, black chokeberries, and Scots pine. The selection of the YOLOv5s network was based on its rapid processing capabilities and effectiveness in identifying small objects. The model exhibited high precision in recognizing and enumerating tree seedlings, successfully detecting the three specified species with commendable performance metrics.

3. Sun, Y. et al. (2022) introduced a framework utilizing high-resolution aerial imagery alongside the Cascade Mask R-CNN, which has been augmented with three attention modules to facilitate the automatic detection and counting of individual tree crowns in subtropical Guangzhou. This enhanced Cascade Mask R-CNN successfully identified over 112 million trees in the area, achieving remarkable accuracy metrics of 88.32% (R^2) and an F1-score of 82.56%.

4. Putra, Y. C. et al. (2023) proposed two distinct models for tree detection and enumeration: an Image Processing Threshold method and an Object-Based Deep Learning architecture utilizing YOLO. Both approaches leverage Very High Resolution (VHR) imagery from Microsoft Bing Maps and UAV satellite sources. A range of evaluation metrics, including precision, recall, F1-Score, and Intersection over Union (IoU), were employed to measure the efficacy of these methods. The results indicate that the automated approach

significantly outperforms the semi-automated model across all three performance metrics—precision, recall, and F1-score—demonstrating superior percentage values.

4. METHODOLOGY

4.1 Convolutional Neural Networks

The architecture of Convolutional Neural Networks (CNNs) is a specialized model within deep learning, specifically tailored for the analysis of structured grid data, including images. This architecture is distinguished by its capacity to automatically and adaptively learn spatial hierarchies of features via convolutional layers. A standard CNN comprises several essential components: convolutional layers that utilize filters to extract local features from the input data, pooling layers that diminish the dimensionality of feature maps while preserving critical information, and fully connected layers that facilitate class predictions. The operations of pooling and convolution enable the network to effectively identify patterns and spatial hierarchies, rendering it particularly adept at various computer vision tasks, such as image classification, object detection, and segmentation. CNNs capitalize on extensive labeled datasets and robust computational resources to enhance their performance through backpropagation, ultimately achieving strong results across a range of visual recognition challenges.

Dataset:

This dataset is utilized for the classification of land based on the presence or absence of trees in geospatial images. The structure of the content is straightforward. Each data point has a resolution of 64x64 pixels and is organized into "tree" and "notree" folders. Each folder, representing a class, contains 5,200 images. Consequently, the entire dataset comprises a total of 10,400 images.

PREPROCESSING

Data preprocessing plays an essential role in the preparation of input data for machine learning models, especially in the realm of computer vision tasks. In this research, the preprocessing stage encompasses several critical activities:

Format Conversion: All images are transformed into standardized formats, including RGB and grayscale. This standardization is vital for maintaining consistency throughout the dataset, which is crucial for effective model training.

Data Augmentation: Methods such as rotation, flipping, and scaling are utilized to artificially increase the size of the dataset. This process enhances the model's robustness by presenting it with diverse representations of the training data**.**

By methodically implementing these preprocessing techniques, we improve the quality of the input data, which subsequently leads to enhanced model performance in later stages of the workflow.

4.2 CNN BASE ARCHITECTURE

The Convolutional Neural Network (CNN) architecture is a specialized type of deep learning model primarily designed for processing structured grid data, such as images. It consists of several layers, including convolutional layers that apply filters to extract spatial features, activation functions (typically ReLU) that introduce non-linearity, pooling layers that down-sample feature maps to reduce dimensionality and computational complexity, and fully connected layers that culminate in classification or regression outputs. This hierarchical structure enables CNNs to capture intricate patterns and hierarchical features in data, making them particularly effective for image recognition, object detection, and other visual tasks.

4.3 CNN GREYSCALE ARCHITECTURE

A Convolutional Neural Network (CNN) architecture tailored for grayscale images is specifically constructed to handle single-channel inputs. It comprises multiple convolutional layers that employ filters to identify significant features, succeeded by pooling layers that reduce the dimensionality of the feature maps, and fully connected layers that execute classification functions. This architecture generally features a sequence of convolutional and pooling layers to detect local patterns, while activation functions such as ReLU introduce non-linearity, allowing the model to acquire intricate representations from the grayscale image data.

4.4 CONVO-2D LAYER WITH 16 OUTPUT FILTERS

A Convolutional Neural Network (CNN) designed for grayscale images is specifically engineered to process single-channel data. This architecture includes several convolutional layers that utilize filters to extract important features, followed by pooling layers that decrease the dimensionality of the feature maps. Additionally, fully connected layers are employed to perform classification tasks. Typically, this architecture consists of a series of convolutional and pooling layers aimed at recognizing local patterns, while activation functions like ReLU introduce non-linearity, enabling the model to learn complex representations from the grayscale image data.

5. RESULTS AND DISCUSSIONS

Below shown are the results of Accuracies shown by 3 different CNN Architectures

- Base
- Greyscale
- Convo-2d Layers with 16 output Filters

I. BASE MODEL

Model 1 (Base): Highest training time but retains good validation accuracy

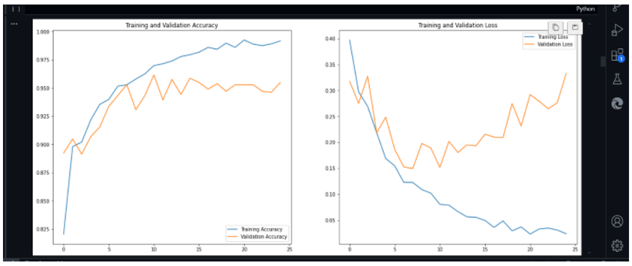

Fig1

Model 1 (BASE) shown in Fig1 gives an Accuracy of 95%

II. GREYSCALE MODEL

Model 2(Greyscale): Fastest training time comparable to Model 1 with similar accuracy.

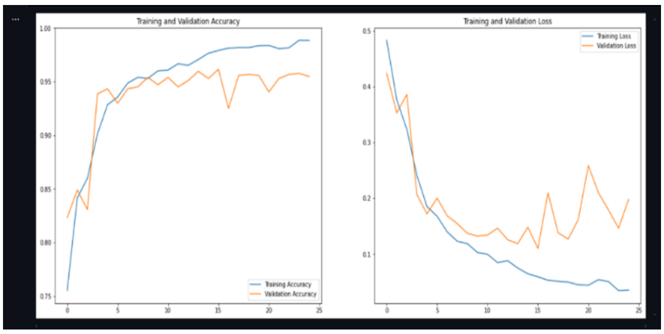

Fig2

Model 2(GREYSCALE) shown in Fig2 gives an accuracy of 95% but is faster

III. CONVO-2D LAYER WITH 16 OUTPUT FILTERS

Model 3 (Convo-2D layer with 16 output Filters): The fastest in terms of training time despite having the most parameters, but slightly lower validation accuracy compared to the other models.

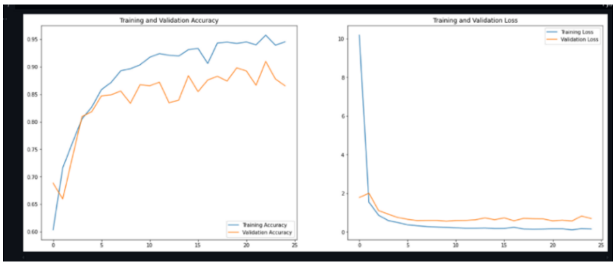

Fig3

Model 3(Convo-2D layer with16 output filters) shown in Fig3 gives an accuracy of 94%

5.1 MODEL ENSEMBLING

In our study, we employed three models: a Convolutional Neural Network (Conv2D layer) featuring 16 output filters, a base model, and a Greyscale model. We utilized two different ensemble techniques—majority voting and prediction averaging—to combine the Conv2D model with 16 output filters and the Greyscale model. Initially, the accuracy of the majority voting ensemble method was recorded at 77%. However, we observed a significant enhancement in accuracy, which rose to 96%, following the transition to the prediction averaging method. This considerable improvement indicates that the averaging approach effectively bolsters the robustness of tree counting predictions. The results suggest that averaging predictions facilitates a more efficient integration of the strengths of each individual model, leading to overall more reliable performance.

```python
# Generate predictions for each model using test_data
#pred1 = model1.predict(test_data)
pred2 = model2.predict(test_data)
pred3 = model3.predict(test_data)

# Ensemble by averaging predictions
ensemble_pred = ( pred3 + pred2) / 2

# Get final predictions for classification tasks
final_pred = np.argmax(ensemble_pred, axis=1)
true_labels = np.concatenate([labels for _, labels in test_data], axis=0)

# Calculate and print accuracy
accuracy = accuracy_score(true_labels, final_pred)
print("Ensemble Accuracy:", accuracy)
```

```
Ensemble Accuracy:0.9681538461538462
```

Fig4

Fig4 Shows 96% accuracy obtained by Ensembling model2 and model3 using averaging predictions method.

5. CONCLUSIONS

The findings suggest that the prediction averaging approach not only successfully consolidates the strengths of each model but also yields more reliable performance in predictions. This highlights the potential benefits of using ensemble techniques, particularly in scenarios where model diversity can contribute to improved outcomes. Future work may explore further refinements in both model architectures and ensemble strategies to enhance predictive performance even further.

REFERENCES

[1] Liu, Z., Song, S., & Xu, **S.** (2020). *Tree detection and counting using deep learning techniques: A review and perspective*. Remote Sensing, 12(16), 2675.

[2] Yuan, J., Liu, Y., & Wang, X**.** (2019). *Improving tree crown detection in high-resolution satellite images using convolutional neural networks and transfer learning*. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(8), 2910-2921.

[3] Weinstein, B. G., & *Marconi, S.* (2019). *Individual tree-crown detection in RGB imagery using deep learning and local maxima filtering*. Remote Sensing, 11(10), 1123.

[4] *residual learning for image recognition*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770–778.

[5] Li, C., Fang, L., & Li**,** S. (2020). *Tree counting using a hybrid CNN-R-CNN model on satellite images*. IEEE Geoscience and Remote Sensing Letters, 17(12), 2108-2112.

[6] Zhu, H., Tian, Y., & Qin, R. (2022). *Ensemble learning approaches for improved tree detection in satellite imagery: A comparative study*. International Journal of Applied Earth Observation and Geoinformation, 109, 102720

.