[Deep Learning for Aerial Image Classification of Cars and Trees in Distinct Geographic Regions: A Case Study of Addis Ababa and Stirling]

Michael Bogale
Addis Ababa, Ethiopia
michaelabebaw1@gmail.com

I Introduction

This project focuses on developing deep learning models [1] to classify aerial images containing cars and trees in two distinct geographic regions: Addis Ababa, Ethiopia, and Stirling, Scotland. I used a dataset containing 2,400 labeled images sourced from Google Earth Pro, captured at an eye altitude of 200 meters above the Earth's surface. The images were collected in March 2023 and evenly distributed between the two regions. The main objective was to find the best deep learning model for classifying images in the Stirling and Addis Ababa datasets and assess the transferability of the best models.

II Proposed Solution

To achieve the objectives, various techniques were applied:

Data augmentation [2]: horizontal/vertical flip, colour jitter, random rotation, and Gaussian blur were applied to both datasets.

Pre-processing: Stirling dataset was resized to (224,224), Addis dataset was sharpened and resized to (224,224)

CNN architectures [1]: ResNet18 and ResNet50 were experimented with, chosen for their well-established performance in image classification tasks.

Grid search for optimizers (SGD and Adam [3]), learning rates (0.001 and 0.01), and weight decay (0.001) was performed to evaluate model performance based on test accuracy.

Training settings: 50:50 train-test split with 10 epochs

III Results

Table 1: Best models with respective parameter and metrics

Dataset/Model	Test Accuracy	F1-score	AUC
Best Stirling Model (ResNet50, SGD)	0.9883	0.9884	0.9998
Best Addis Model (ResNet18, SGD)	0.9550	0.9549	0.9966
Stirling Model on Addis dataset	0.4500	0.3818	0.7655
Addis Model on Stirling dataset	0.8983	0.8990	0.9789

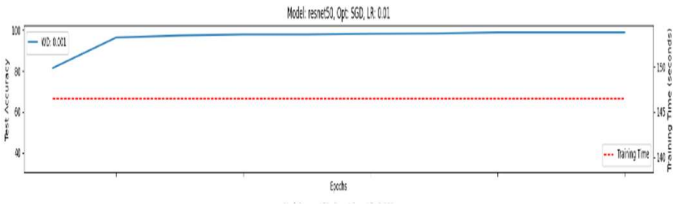

Fig. 1: Test accuracy and Training time for Best Stirling Model (ResNet50, SGD)

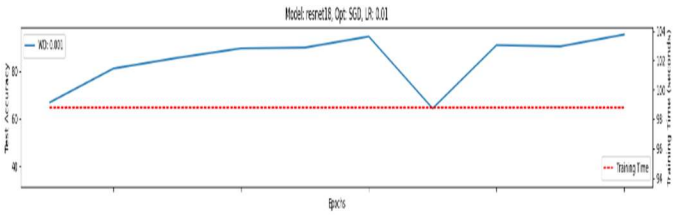

Fig. 2: Test accuracy and training time for Best Addis Model (ResNet18, SGD)

IV Discussion

The grid search identified ResNet50 [1] with the SGD optimizer as the best combination for the Stirling dataset and ResNet18 [1] with the SGD optimizer for the Addis dataset.

Data augmentation [2] and pre-processing techniques improved the model's performance by increasing the diversity of the training data, reducing the risk of overfitting, and enhancing generalizability.

The transferability of the models was assessed by testing them on the other dataset, revealing a drop in performance. The Stirling model performed poorly on the Addis dataset, while the Addis model achieved relatively better performance on the Stirling dataset. This suggests that the Addis model might have better generalization capabilities than the Stirling model.

Limitations and Future Work

Some limitations of the current approach could be addressed in future work:

• Increase dataset size and diversity for improved performance and generalizability for example if we mix Stirling and Addis datasets, we will have a better performance.

• Explore alternative model architectures [4] and hyperparameters for optimization.

• Modify the training data split ratio from 0.5 for better learning.

To improve transferability, consider:

• Fine-tuning the models using a smaller subset of the other dataset, allowing the models to adapt to the specific characteristics of the other dataset, potentially improving their performance and generalizability. I have checked this and it works but we have to make our models perform on unseen datasets.

Further research could focus on:

• Exploring alternative CNN architectures like EfficientNet, DenseNet, or MobileNet [4] for valuable insights into the models' performance.

• Incorporating additional evaluation metrics, such as precision, recall, and specificity, to provide a more comprehensive assessment of the models' performance, allowing for a better understanding of the models' strengths and weaknesses.

V Conclusion

The project developed deep learning models using ResNet18 and ResNet50 architectures [1] to classify cars and trees in aerial images from Addis Ababa and Stirling. The models were enhanced through systematic pre-processing, data augmentation [2], and hyperparameter optimization [3]. While transferability assessments showed a performance drop, the models maintained relatively high accuracies. Future work could focus on increasing dataset size, exploring alternative architectures [4], and employing additional evaluation metrics to further improve the models' performance.

References

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems (NIPS), 2012, pp. 1097-1105.

[3] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

[4] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proceedings of the International Conference on Machine Learning (ICML), 2019, pp. 6105-6114.