

RGB UAV Imagery Segmentation: A Comparative Study

Mathews Jere^{1*}, Esnart Sandikonda¹, Kondwani G. Munthali¹, Yamikani D.J. Phiri¹, Augustine Simwela¹, Ramsey Njema¹

¹Department of Computing, University of Malawi, Zomba, Malawi
{mjere, esandikonda, kmunthali, yphiri, asimwela, rnjema}@unima.ac.mw

Abstract

Semantic segmentation of remote sensing images is vital in computer vision for detailed scene understanding, with applications spanning agriculture, environmental monitoring, and urban planning. In this study, we conduct a comparative analysis of advanced RGB image segmentation methods using various Fully Convolutional Neural Network (FCN) architectures. The architectures evaluated include Attention U-Net, Deeplabv3+, U-Nets with backbones such as ResNetv2, MobileNetv2, VGG19, and EfficientNet. Our experiments on UAV-captured images from the Mpokwe Extension Planning Area (EPA) in Zomba, Malawi, reveal that the ResNetv2-based U-Net achieves the highest Intersection over Union (IoU) score of 0.91. This model effectively segments classes like maize, water, trees, bare land, built-up areas, and other vegetation types, facilitating applications in agricultural yield estimation, land use analysis, and environmental monitoring. The MobileNetv2-based U-Net provides a good balance between performance and efficiency, making it suitable for real-time applications. This work demonstrates the potential of leveraging cost-effective RGB sensors to achieve competitive performance in image segmentation, contributing to the Sustainable Development Goals (SDGs) by informing agricultural practices and environmental stewardship.

1 Introduction

Image segmentation is a fundamental task in computer vision, involving the partitioning of a digital image into multiple meaningful segments or regions to extract regions of interest. It is broadly categorized into semantic, panoptic, and instance segmentation. Semantic segmentation assigns a class label to every pixel in an image, enabling detailed scene understanding, which is crucial for various applications such as autonomous driving, medical diagnosis, and land-use analysis [Yu *et al.*, 2023]. Despite significant advancements, semantic segmentation remains a challenging task due to the complexity of accurately identifying and labeling objects within an image [Hao

et al., 2023].

In remote sensing, the challenge intensifies when dealing with RGB images. Unlike multispectral or hyperspectral images that provide rich spectral information, RGB images have limited spectral separability among different features, making it difficult to distinguish between objects with similar color characteristics. However, RGB sensors are more cost-effective and widely available, prompting the need for advanced algorithms capable of extracting meaningful information from RGB imagery.

These challenges directly relate to addressing several of the United Nations Sustainable Development Goals (SDGs), particularly SDG 2 (Zero Hunger) and SDG 15 (Life on Land). High-resolution segmentation of agricultural and natural landscapes can inform yield estimation, resource allocation, and environmental monitoring [Del Moro and others, 2022]. While freely available satellite data (e.g., Sentinel-2) offers moderate resolutions, UAV-captured imagery provides higher spatial detail. This high resolution is especially crucial for smallholder farms, where parcel sizes can be too small or heterogeneous for medium-resolution satellite imagery to be effective. By leveraging cost-effective RGB sensors on UAVs, stakeholders can obtain timely and fine-scale land cover maps essential for precision agriculture, food security interventions, and habitat conservation efforts.

In this paper, we conduct a comparative study of advanced image segmentation techniques using Fully Convolutional Neural Network (FCN) architectures on UAV-captured RGB imagery. We evaluate several state-of-the-art models, including Attention U-Net, Deeplabv3+, and U-Nets with various backbone networks such as ResNetv2, MobileNetv2, VGG19, and EfficientNet. Our goal is to identify models that can achieve high segmentation accuracy using RGB images, thereby leveraging cost-effective sensors for applications in agriculture, environmental monitoring, and urban planning while contributing to SDG-related outcomes.

2 Related Work

Image segmentation has been a vital area of research in computer vision for decades. Early approaches relied on classical methods such as edge detection, region growing, clustering, and graph-based algorithms [Yu *et al.*, 2023; Freedman and Zhang, 2005; Chuang *et al.*, 2006]. These

*Contact Author

methods often utilized low-level features like color, intensity, and texture to partition images into meaningful regions.

With the advent of machine learning, classical techniques like Support Vector Machines (SVM), Random Forests, and Markov Random Fields were employed for segmentation tasks [Seo *et al.*, 2020]. These methods provided improved performance by learning from data but were limited by their reliance on handcrafted features.

The emergence of deep learning revolutionized image segmentation, with Convolutional Neural Networks (CNNs) demonstrating remarkable capabilities in learning hierarchical features directly from raw data. Fully Convolutional Networks (FCNs) [Long *et al.*, 2015] were introduced to adapt CNNs for pixel-wise prediction tasks, leading to significant improvements in segmentation accuracy.

The U-Net architecture [Ronneberger *et al.*, 2015], initially proposed for biomedical image segmentation, introduced a symmetric encoder-decoder structure with skip connections, enabling precise localization and context integration. Variations of U-Net have since been developed, incorporating different backbone networks and attention mechanisms to enhance performance.

Attention mechanisms, as incorporated in Attention U-Net [Oktay *et al.*, 2018], allow models to focus on relevant parts of the image, improving segmentation of complex structures. Deeplabv3+ [Chen *et al.*, 2018] introduced atrous spatial pyramid pooling and encoder-decoder structures to capture multi-scale context.

In remote sensing, segmentation of UAV and satellite imagery is critical for applications such as land cover mapping, agricultural monitoring, and environmental assessment [Del Moro and others, 2022]. While multispectral and hyperspectral images provide rich spectral information, they are costly and not always accessible. Therefore, leveraging RGB images for segmentation tasks is of significant interest.

Previous studies have applied deep learning techniques to UAV RGB imagery for tasks like crop type classification [Safonova *et al.*, 2021] and vegetation analysis [Hunt *et al.*, 2013], but challenges remain due to limited spectral information. Our work contributes to this area by evaluating advanced FCN architectures on RGB UAV imagery, aiming to achieve competitive performance with more expensive multispectral data while supporting agricultural productivity, environmental conservation, and ultimately advancing SDGs related to food security and sustainable land use.

3 Materials and Methods

3.1 Overall Methodology Flow

Figure 1 provides an overview of our methodology, from data collection through annotation and model evaluation.

3.2 Study Area and Data Acquisition

The study was conducted in the Mpokwe Extension Planning Area (EPA) located in Zomba district, Malawi. This region is characterized by a mix of settlements, water bodies, forested areas, and smallholder agricultural fields, primarily cultivating maize and other crops. The diverse land cover types present in

this area make it an ideal location for evaluating segmentation models targeting multiple classes.

Data collection took place during the mid-rainy season on February 26, 2024. During this period, Malawi experiences lush vegetation growth due to the subtropical climate, with most crops and natural vegetation at peak greenness [Services,]. This seasonal timing ensured the dataset captured a wide range of vegetation types and conditions, enhancing the robustness of the segmentation models.

Imagery was acquired using a DJI Matrice 300 RTK drone equipped with a Zenmuse H20N camera [DJI,]. The drone was flown at an altitude of 120 meters above ground level, capturing high-resolution RGB images with a ground sampling distance (GSD) of 3.6 cm per pixel. A total of 2,780 images were collected, covering approximately 214 hectares of the EPA.

To ensure complete coverage and sufficient image overlap for accurate mosaicking, the drone flight was planned with a side overlap of 70% and a frontal overlap of 80%. The high overlap ratios also enhance the quality of the orthomosaic and facilitate accurate feature extraction during segmentation. Sample images from the dataset are shown in Figure 2.

3.3 Data Processing

Image Mosaicking and Patch Extraction

The raw UAV images were processed using Pix4DMapper [Pix4D,], a photogrammetry software widely used for processing drone imagery. Pix4DMapper was utilized to stitch the individual images into a seamless orthomosaic of the study area. The software performs aerial triangulation, generates a dense point cloud, and produces high-resolution orthorectified images, which are essential for accurate spatial analysis.

From the orthomosaic, image patches of size 224×224 pixels were extracted for model training and testing. The choice of patch size was influenced by:

1. Compatibility with the input dimensions of popular deep learning models (often 224×224 due to ImageNet pre-training).
2. Computational resource constraints, allowing efficient training on available hardware.

Data Annotation

Despite having a large number of images, we performed detailed pixel-wise annotations on a curated subset of 200 image patches. Pixel-level annotation is labor-intensive, so we selected 200 patches to ensure balanced coverage of each land cover type while keeping the workload feasible. This subset was representative of the various land cover classes:

- **Maize:** Representing maize crops in various growth stages.
- **Trees:** Including forested areas and isolated trees.
- **Other Vegetation:** Encompassing other crop types and grasslands.
- **Water:** Covering rivers, streams, and ponds.
- **Built-up Areas:** Including buildings, roads, and other man-made structures.

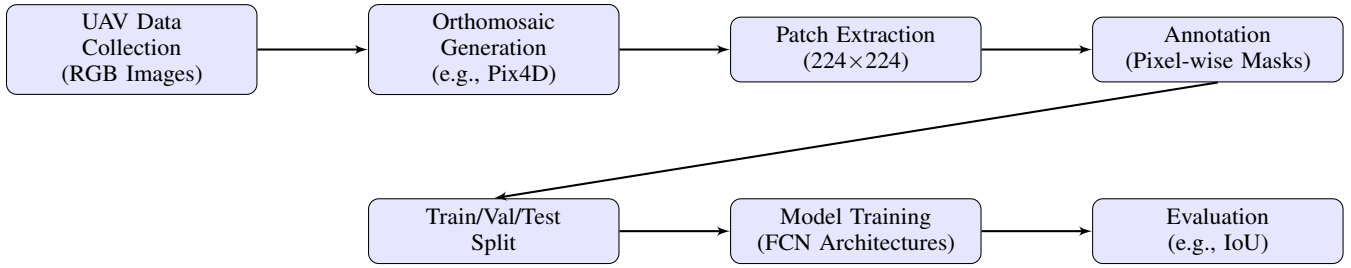


Figure 1: Methodology overview. UAV RGB imagery is processed into an orthomosaic, from which patches are extracted. Pixel-wise masks are manually annotated, and the dataset is split into training, validation, and test sets. FCN-based segmentation models are then trained and evaluated (e.g., via IoU).

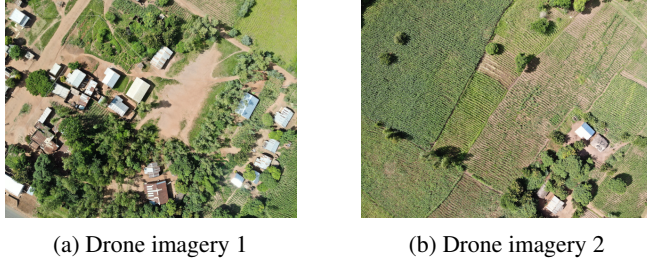


Figure 2: Sample UAV imagery from the study area.

- **Bare Land:** Areas with exposed soil and minimal vegetation.

These classes were selected due to their predominance in the region and relevance to agricultural and environmental analyses. Manual annotation was performed in Label Studio [Heartex,], resulting in precise pixel-wise masks. Figure 3 illustrates examples of annotated image patches and corresponding masks.

Training, Validation, and Test Sets

The 200 annotated patches were split into training (70%), validation (10%), and test (20%) sets. The training set was used to optimize model weights, the validation set guided hyperparameter tuning and early stopping, and the test set provided an unbiased evaluation of final model performance.

Data Preprocessing and Augmentation

Before training, the images and masks underwent several preprocessing steps to enhance image quality and normalize the data:

- **Contrast Enhancement:** Images were converted to the LAB color space, and the L (lightness) channel was enhanced using Contrast Limited Adaptive Histogram Equalization (CLAHE) [Pisano *et al.*, 1998]. This step improved the contrast and visibility of features.
- **Histogram Equalization:** The images were also processed in the HSV color space, applying histogram equalization to the V (value) channel to adjust the brightness distribution [Xiong *et al.*, 2021].
- **Normalization:** Pixel values were normalized to the range [0, 1] to standardize the input for the neural networks [Goodfellow *et al.*, 2016].

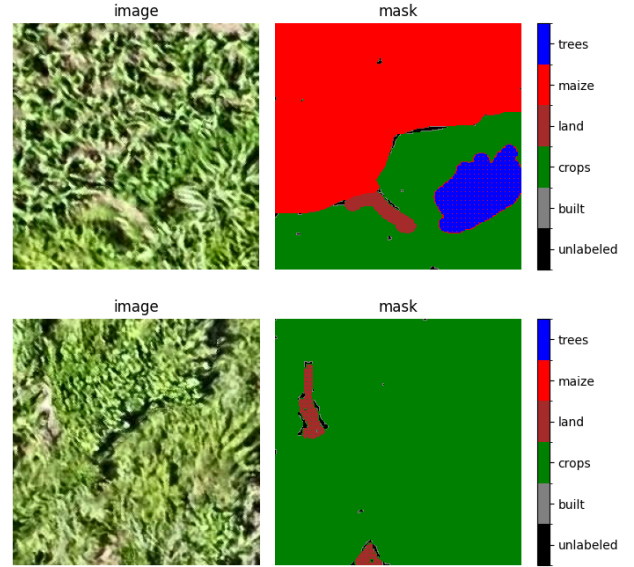


Figure 3: Examples of annotated image patches and masks for six land cover classes.

- **Data Augmentation:** To increase the diversity of the training data and prevent overfitting, random horizontal and vertical flips were applied during training. Additional techniques like random rotations and brightness adjustments were considered but only selectively applied to avoid distorting the natural appearance of crops [Shorten and Khoshgoftaar, 2019].

4 Proposed Method

4.1 Model Architectures

We evaluated six state-of-the-art FCN architectures for semantic segmentation:

1. **Attention U-Net** [Oktay *et al.*, 2018]: An extension of the U-Net architecture incorporating attention gates to focus on relevant spatial regions.
2. **Deeplabv3+** [Chen *et al.*, 2018]: An advanced model utilizing atrous spatial pyramid pooling and an encoder-decoder structure for capturing multi-scale context.

3. **U-Net with ResNetv2 Backbone** [He *et al.*, 2016]: Integrates a pre-trained ResNetv2 model as the encoder to leverage learned features from large-scale image classification tasks.
4. **U-Net with MobileNetv2 Backbone** [Sandler *et al.*, 2018]: Uses MobileNetv2 for the encoder, offering a balance between performance and computational efficiency.
5. **U-Net with VGG19 Backbone** [Simonyan and Zisserman, 2014]: Employs VGG19 as the encoder, known for its simplicity and effectiveness in image classification.
6. **U-Net with EfficientNet Backbone** [Baheti *et al.*, 2020]: Incorporates an EfficientNet encoder, designed with compound scaling to balance depth, width, and resolution.

All models follow the typical encoder-decoder structure of U-Net, with skip connections to preserve spatial information. Pre-trained weights on ImageNet were utilized for the encoder parts to benefit from transfer learning.

4.2 Training Procedure

Training was performed on a workstation equipped with an NVIDIA Quadro P1000 GPU. We adopted a two-stage training strategy:

- **Stage 1 (Fine-tuning Encoders):** Models were trained with an initial learning rate of 0.001 to gradually adapt the pre-trained encoder weights to the UAV imagery.
- **Stage 2 (Focusing on Decoders):** The learning rate was reduced to 0.0001, allowing the decoder layers to refine segmentation boundaries more precisely.

We used the Adam optimizer [Kingma and Ba, 2014] due to its adaptive learning rate capabilities, with a batch size of 8 chosen to balance computational efficiency and memory constraints. Class imbalance in the dataset was addressed by computing class weights and incorporating them into the loss function, ensuring that minority classes (e.g., water, built-up areas) were not overlooked.

4.3 Evaluation Metrics

We evaluated the models using standard segmentation metrics:

- **Intersection over Union (IoU):** Measures the overlap between predicted and ground-truth segments.
- **Recall (Sensitivity):** Indicates the model’s ability to identify all relevant pixels of a class.
- **Precision:** Measures the accuracy of the model’s positive predictions.
- **Categorical Accuracy:** Represents the overall fraction of correctly classified pixels.

5 Results

5.1 Hyperparameter Tuning

We experimented with different configurations for learning rates, batch sizes, and optimizer settings. Key findings include:

Model	IoU	Recall	Precision	Accuracy
Attention U-Net	0.38	0.69	0.78	0.74
Deeplabv3+	0.08	0.70	0.17	0.21
ResNetv2 U-Net	0.91	0.94	0.95	0.95
MobileNetv2 U-Net	0.52	0.69	0.70	0.70
VGG19 U-Net	0.50	0.66	0.72	0.70
EfficientNet U-Net	0.36	0.56	0.62	0.60

Table 1: Overall Performance Metrics (Aggregated Across Classes)

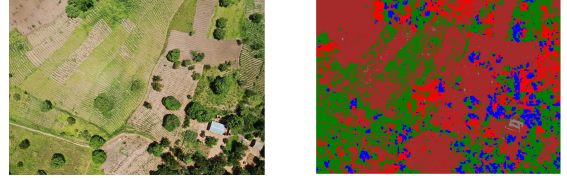


Figure 4: Segmentation results by ResNetv2-based U-Net.

Key: Red = Maize, Blue = Trees, Green = Other Vegetation, Brown = Bare Land, Grey = Built-up Areas, (Water not visible in this patch)

- **Learning Rate:** An initial learning rate of 0.001 was optimal for most architectures, with a reduction to 0.0001 in later epochs improving fine-grained segmentation quality.
- **Batch Size:** A batch size of 8 was chosen to avoid exceeding GPU memory while ensuring stable gradient estimates.
- **Optimizer:** Adam [Kingma and Ba, 2014] outperformed other optimizers such as RMSProp, offering faster convergence in our experiments.

5.2 Overall Model Performance

Table 1 summarizes the overall performance metrics (aggregated across classes) for each model. The ResNetv2-based U-Net achieved the highest IoU (0.91) and accuracy (0.95), while the MobileNetv2-based U-Net provided a strong trade-off between accuracy and efficiency. Deeplabv3+ struggled under limited training data, yielding the lowest overall metrics.

5.3 Per-Class Performance

To better understand how each model handles class imbalance and diverse land covers, Table 2 reports per-class IoU. Notably, minority classes such as *Water* and *Built-up Areas* had lower IoUs, indicating the challenge of segmenting underrepresented classes. The ResNetv2-based U-Net again outperformed others in most classes, illustrating its robustness to class imbalance.

5.4 Qualitative Assessment

Figure 4 showcases an example of segmentation outputs from the ResNetv2-based U-Net. The model effectively segmented fine details in various land cover types, although slight confusion remained between *Other Vegetation* and *Maize* at field boundaries.

Model	Maize	Trees	Other Veg	Water	Built-up	Bare Land
Attention U-Net	0.51	0.34	0.46	0.21	0.39	0.38
Deeplabv3+	0.10	0.12	0.09	0.06	0.07	0.05
ResNetv2 U-Net	0.93	0.90	0.92	0.84	0.88	0.89
MobileNetv2 U-Net	0.61	0.54	0.49	0.40	0.42	0.50
VGG19 U-Net	0.58	0.47	0.52	0.38	0.44	0.42
EfficientNet U-Net	0.42	0.36	0.38	0.25	0.30	0.35

Table 2: Per-Class IoU for Each Model

6 Discussion

The ResNetv2-based U-Net’s superior performance can be attributed to its deeper encoder with residual connections, which mitigates vanishing gradients and captures complex spatial features effectively. However, this comes with higher computational costs, potentially limiting applicability for real-time or edge-device scenarios.

MobileNetv2-based U-Net, on the other hand, offers a solid balance between performance and efficiency, making it a strong candidate for real-world deployments where GPU resources are limited (e.g., on-board UAV processing). Its lighter architecture also suggests faster inference times, valuable for quick decision-making in fields such as precision agriculture.

Deeplabv3+ exhibited the poorest performance among the tested architectures, which may stem from the relatively small training set (200 annotated patches). Its sophisticated atrous convolution and decoder structure likely overfitted to limited training data. Increasing the annotated dataset size, simplifying the network, or employing stronger regularization may address these issues.

The annotated data’s class imbalance particularly affected minority classes (e.g., water and built-up areas). Although class weighting helped, future work could employ focal loss or generate synthetic training samples to further reduce imbalance. Integrating multispectral data or other sensing modalities (e.g., LiDAR) could also enhance the separability of tricky classes.

7 Applications and Relevance to SDGs

Accurate segmentation of land cover in smallholder farming regions has direct implications for SDGs, particularly in zero hunger (SDG 2), sustainable cities and communities (SDG 11), and life on land (SDG 15). Potential applications include:

- **Precision Agriculture:** Timely identification of crop health and type supports improved yield estimation, resource allocation, and monitoring of pest or disease outbreaks.
- **Natural Resource Management:** Delineation of forested areas and water bodies aids in conservation efforts, reforestation planning, and watershed management.
- **Urban and Rural Planning:** Mapping built-up areas informs land-use policies, infrastructure development, and disaster risk reduction strategies.

- **Real-time Monitoring:** Models like the MobileNetv2 U-Net can facilitate on-board UAV analytics, enabling immediate insights and interventions.

8 Conclusion and Future Work

In this study, we evaluated several advanced Fully Convolutional Neural Network (FCN) architectures for semantic segmentation of RGB UAV imagery in the Mpokwe EPA of Zomba, Malawi. Our experiments revealed that the ResNetv2-based U-Net delivers the highest accuracy (IoU of 0.91), effectively segmenting critical classes such as maize, trees, other vegetation, water, bare land, and built-up areas. The MobileNetv2-based U-Net provides a promising alternative for resource-limited settings, maintaining competitive performance with significantly reduced computational cost.

By focusing on cost-effective RGB sensors, this work demonstrates a pathway for broader adoption of high-resolution remote sensing solutions that address SDGs related to agriculture and environmental management. Future directions include:

- **Synthetic Data and Augmentation:** Employing generative models to create synthetic samples for minority classes could improve segmentation performance.
- **Focal Loss Adoption:** Further mitigating class imbalance via focal loss or similar strategies.
- **Multimodal Fusion:** Integrating LiDAR or SAR data to better discriminate overlapping spectral signatures.
- **Deployment Optimization:** Exploring model compression or pruning, enabling real-time analytics on low-power UAVs.

Such advancements will bolster the accuracy and utility of UAV-based remote sensing, enhancing decision-making in agriculture, urban planning, and environmental conservation.

References

- [Baheti *et al.*, 2020] Bhushan Baheti, Shubham Innani, Sarthak Gajre, and Sanjay Talbar. Eff-UNet: A novel architecture for semantic segmentation in unstructured environment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 1494–1503, 2020.
- [Chen *et al.*, 2018] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic

- image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.
- [Chuang *et al.*, 2006] Keh-Shih Chuang, Hong-Long Tzeng, Shyang Chen, Jay Wu, and Tzong-Jer Chen. Fuzzy c-means clustering with spatial information for image segmentation. *Computerized medical imaging and graphics*, 30(1):9–15, 2006.
- [Del Moro and others, 2022] Giacomo Del Moro *et al.* The benefits of dual-polarisation sar data for agricultural crop classification in the ukrainian carpathians. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-3-W2-2022:67–74, 2022.
- [DJI,] DJI. Matrice 300 rtk. <https://enterprise.dji.com/matrice-300>. Accessed: June 13, 2024.
- [Freedman and Zhang, 2005] Daniel Freedman and Tao Zhang. Interactive graph cut based segmentation with shape priors. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 755–762. IEEE, 2005.
- [Goodfellow *et al.*, 2016] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep learning. 2016. <https://www.deeplearningbook.org/>.
- [Hao *et al.*, 2023] Xuesong Hao, Lei Yin, Xia Li, Lei Zhang, and Ruixia Yang. A multi-objective semantic segmentation algorithm based on improved u-net networks. *Remote Sensing*, 15(7):1838, 2023.
- [He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [Heartex,] Heartex. Label studio: Open source data labelling tool. <https://labelstud.io/>. Accessed: June 13, 2024.
- [Hunt *et al.*, 2013] E Raymond Hunt, Prasad C Doraiswamy, James E McMurtrey, Craig ST Daughtry, Edward M Perry, and Boris Akhmedov. A visible band index for remote sensing leaf chlorophyll content at the canopy scale. *International Journal of Applied Earth Observation and Geoinformation*, 21:103–112, 2013.
- [Kingma and Ba, 2014] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [Long *et al.*, 2015] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [Oktay *et al.*, 2018] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, *et al.* Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.
- [Pisano *et al.*, 1998] E. D. Pisano, S. Zong, B. M. Hemminger, M. DeLuca, R. E. Johnston, K. Muller, M. P. Braeuning, and S. M. Pizer. Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. *Journal of Digital Imaging*, 11(4):193–200, November 1998.
- [Pix4D,] Pix4D. Pix4dmapper photogrammetry software. <https://www.pix4d.com/product/pix4dmapper-photogrammetry-software>. Accessed: June 13, 2024.
- [Ronneberger *et al.*, 2015] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241, 2015.
- [Safonova *et al.*, 2021] Anastasia Safonova, Eduardo Guirado, Yaroslav Maglinets, Domingo Alcaraz-Segura, and Siham Tabik. Olive tree biovolume from uav multi-resolution image segmentation with mask r-cnn. *Sensors*, 21(5):1617, 2021.
- [Sandler *et al.*, 2018] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018.
- [Seo *et al.*, 2020] Hyungseok Seo, Mojtaba Badiei Khuzani, Vijay Vasudevan, Chieh-Li Huang, Huaqing Ren, Ruogu Xiao, Xianrong Jia, and Lei Xing. Machine learning techniques for biomedical image segmentation: An overview of technical aspects and introduction to state-of-art applications. *Medical physics*, 47(5):e148–e167, 2020.
- [Services,] Malawi Meteorological Services. Climate of malawi. <https://www.metmalawi.com/climate/climate.php>. Accessed: June 13, 2024.
- [Shorten and Khoshgoftaar, 2019] Connor Shorten and Taghi M. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):60, 2019.
- [Simonyan and Zisserman, 2014] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [Xiong *et al.*, 2021] J. Xiong, D. Yu, Q. Wang, L. Shu, J. Cen, Q. Liang, H. Chen, and B. Sun. Application of histogram equalization for image enhancement in corrosion areas. *Shock and Vibration*, 2021:Article ID 8883571, 2021.
- [Yu *et al.*, 2023] Yixin Yu, Chen Wang, Qiang Fu, Ronghua Kou, Feng Huang, Bin Yang, Tian Yang, and Ming Gao. Techniques and challenges of image segmentation: A review. *Electronics*, 12(5):1199, 2023.