AIQTrees: A Drone Imagery Dataset for Tree Segmentation

Joseph Chai 123119637@UMAIL.UCC.IE
University College Cork, Ireland

Alex To DUTO3894@UNI.SYDNEY.EDU.AU
University of Sydney, Australia

Barry O' Sullivan B.OSULLIVAN@CS.UCC.IE
University College Cork, Ireland

Hoang D. Nguyen HN@CS.UCC.IE
University College Cork, Ireland

Abstract

The reliability of AI models typically depends on the data they are trained with, and accurate interpretations require large amounts of data. The scarcity of publicly available datasets is typically encountered for specific small-scale sustainability projects, making data accessibility a limiting factor for developing AI models for semantic segmentation tasks. In sustainability and forestry applications, the usage of UAVs is common due to their lightweight nature and the ability to provide a huge variety of data. In this paper, we present a new dataset of realistic and high-quality drone images taken around sites in Ireland. The images encompass temporal, spatial, and seasonal dimensions, which could alter the tree appearance or illumination conditions of the images and have to be taken into consideration. We also included a baseline benchmark for the semantic segmentation task along with the dataset. It can be accessed at: https://github.com/ReML-AI/AIQTrees.

Keywords: public dataset; semantic segmentation; sustainability; drone imagery

1. Introduction

In the field of image processing and computer vision, deep learning (DL) techniques such as convolution neural network (CNN) for image analysis gained great traction due to increased computational power and larger datasets (Ma et al. (2019); Isola et al. (2017)). Image segmentation has been receiving a lot of attention lately due to the wide range of practical applications it can have in different industries such as agriculture, robotics, and healthcare (Duckett et al. (2018); Gao et al. (2018); Jarvis (1982)). The task of semantic segmentation is to classify each pixel in an image into a predefined set of labels. The accurate semantic segmentation of high vegetation plays a huge role in our goal of progressing towards sustainable development, it aids in practical applications such as carbon stock estimation and land optimization for sustainable activities. High vegetation in our case is defined as any kind of vegetation that stands out prominently due to height. Trees and shrubs fall into this category, but grasses do not. Figure 1 shows an overview of our goals towards sustainable development. Data accessibility is also a common issue faced due to the reliance on quality data when training an AI model. This makes it challenging for niche, small-scale projects to find a suitable dataset that is publicly available. To mitigate these issues, we published a high-resolution aerial image dataset in Cork, Ireland. Our main contribution in this paper includes the release of a new publicly available dataset, detailed documentation of our data collection and processing, and a benchmark for the image segmentation of high vegetation.

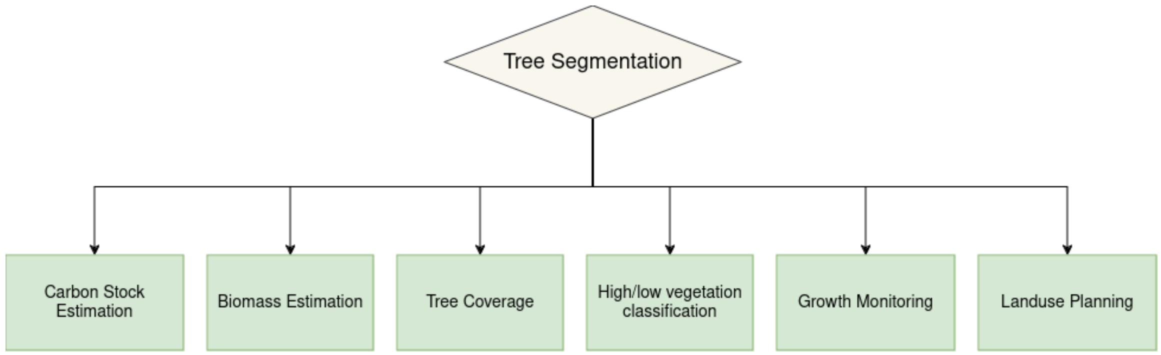

Figure 1: Applications of High-Resolution Drone Imagery Dataset

2. Related Work

2.0.1. Aerial Image Datasets

Several computer vision datasets exist for comparison among semantic segmentation methods in various scenes. ReForestree is a dataset for estimating tropical forest carbon stock with DL and Aerial Imagery (Reiersen et al. (2022)), focused on 6 agroforestry sites in Ecuador. However, the six sites are similar in nature, containing a majority of banana trees (Musaceae) and cacao plants. In a forest environment, the tree structure differs greatly from open spaces such as parks. (Scher et al. (2019); Hirons and Thomas (2017); Sterck et al. (2011)). The difference is due to the responses of their surroundings; for example, trees in closed forests have to compete with surrounding trees for resources such as light and water. In order to provide insight into the intrinsic biology of the particular tree, where growth pattern is more inherent to their genetic traits, the absence of competition is preferred.

The TUGRAZ (2019) dataset consists of 400 annotated images with 20 different classes such as building parts, vehicles, ground, vegetation, and humans. Another similar dataset, UAVid (Lyu et al. (2020)), includes 300 annoated images with 8 different categories along with 4K resolution UAV videos. While these datasets consist of vegetations, it is not the focus of the dataset, and some the images in the dataset do not consist of any form of vegetation. The VALID dataset by Chen et al. (2020) shows the strength of a virtual dataset, the usage of Unreal Engine to reconstruct real-world scenes, along with the AirSim simulator to simulate a drone, offers great flexibility and control for their data collection. However, the issue with synthetic data is that it lacks realism, the generation of synthetic datasets still relies on real-world data, making the algorithms or models susceptible to statistical noise and sampling biases, which could possibly lead to inaccurate results. Therefore, data collection is necessary.

3. Data collection

Being based in Ireland, we considered many different counties in Ireland to collect the dataset. We eventually settled on Cork for its abundance of green landscapes. We decided to go mainly with parks, as not only is there an abundance of parks to choose from, but there is a good mix of high vegetation, as well as other features such as walkways, shelters, ponds, and grass. We also exercise data protection by trying our best to avoid capturing cars, human subjects, and houses. If any of our images are found to have potential privacy concerns, we will perform additional image processing steps described in the data processing section to comply with the General Data Protection Regulation (GDPR) (Ducato (2020)).

Figure 2 shows an ideal drone image taken in Fitzgerald Park, Cork.

Figure 2: Example of an ideal drone image

The images were taken during different seasons with varying lighting conditions too. This is important for a dataset as trees during the winter usually do not have leaves, making their appearance differ greatly from the summer. Figure 3 shows an example of the different appearance of trees due to seasonal aspacts. Data were collected during different weather conditions and times of the day to vary the lighting conditions of the images as those conditions altered the appearance of the image. For example, during a cloudy weather condition, the lighting is more evenly distributed on the subject. The time of day alters the images in terms of where the sunlight is coming from, causing bright lighting or strong cast shadows in certain parts of the images. Our data also includes images taken under bright sunlight coming from an angle, casting a strong shadow. Including these variations in the dataset ensures a generalized machine-learning model that is able to identify and perform accurate image segmentation despite the altered appearance of trees due to the aforementioned conditions.

Various drone models were also considered to be used for capturing the images. We decided to go with commercial drones as they come fully-featured out of the box, which simplifies the usage of the drone. We considered various models, such as the DJI Fly, DJI Mavic 3, and DJI Mini 3 Pro, but eventually decided on the DJI Mini 3 Pro. Despite the superior sensor size of the Mavic 3 and DJI Fly, the DJI Mini 3 Pro is the cheapest and the only drone that weighs below 250 g, omitting the need to register drones with the Federal Aviation Administration (FAA) as well as being remote ID compliant. It also comes with

Figure 3: Seasonal (winter/summer) and lighting variations of drone images for reliable semantic segmentation

a 48 MP camera that is able to capture images with a high resolution of 8k (8064 $\times$ 6048), which is good enough for our use case. It has a folded dimension of 145$\times$90$\times$62 mm, which makes it portable and easy to bring around.

4. Data Processing and Method

In preparation for the AI Quest competition, we reviewed all the photos taken and ensured that every image had a good balance of high vegetation and background. Due to the privacy-invading nature of drone photography, we have taken some several steps to ensure privacy: (i) review all the drone images captured, (ii) identify any images that may raise privacy concerns, such as images containing cars, private housing, or human subjects, and (iii) perform a crop on the original images, leaving out the area with privacy concerns.

The labeling of the ground truth data in the dataset was done manually with the Computer Vision Annotation Tool (CVAT), a free and open source web-based image and video annotation tool that is used for labeling data for computer vision algorithms.

Figure 4: Example of an annotated mask.

Since the images were annotated manually by humans, they are not 100 percent accurate, and the masks were roughly drawn around the high vegetation. Figure 4 shows an example of a mask annotated. These human factors with noisy masks also pose an extra challenge to the machine learning tasks working with the dataset, prompting robust and reliable segmentation methods.

After annotation, the images are carefully cropped into various smaller images, making sure each image subset contains a good balance of high vegetation and background. This is done to ensure we have a large enough meaningful dataset for the participants to utilize in the training of their AI models to perform semantic segmentation. Figure 5 shows an example of a smaller image.

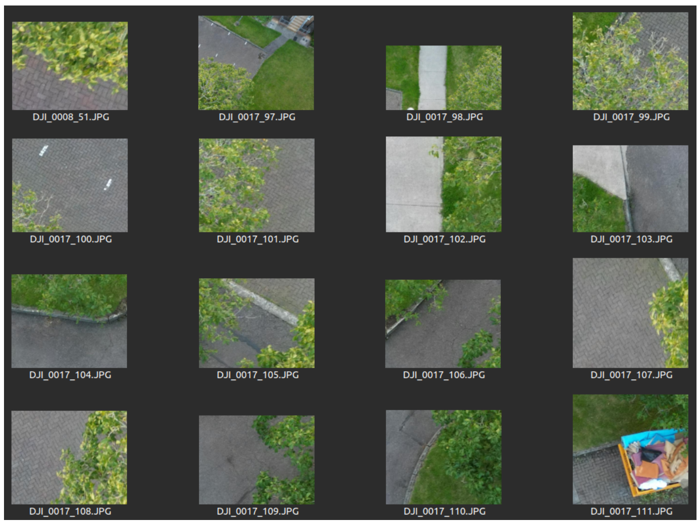

Figure 5: Various smaller image cropped from the larger drone images.

These images are then released as different sets for public and private benchmarking. The public dataset is used to train and fine-tune their models, whereas the private dataset is used to evaluated against the trained model.

In each set, the images are further split into training, validation, and testing sets. A full breakdown of the images is shown in Table 1. At each set, we ensure that the images released are fair and have a good mix of the different image variances, such as trees taken during winter vs. summer and time of day, as well as including images from all the sites.

Table 1: Breakdown of images

Set	Split	Original images	Number of crops
Private	Test	12	783
Public	Test	12	682
Public	Train	93	4322
Public	Val	15	689

5. Baseline benchmark

To validate the quality of our dataset, a baseline benchmark was performed. The ResNet50 model architecture, a popular image classifier model, was employed for the training with a batch size of 8. Intersection over union (IoU) is used as an evaluation metric for the image segmentation task as it provides a quantitative measure of the model's accuracy in segmenting the high vegetation from the background. A higher IoU score indicates better model performance. The evaluation equation is as follows:

$$IoU_i = \frac{C_{i,i}}{C_{i,i} + \sum_{j \neq i}(C_{i,j} + C_{j,i})} = \frac{TP}{TP + FP + FN} \tag{1}$$

Table 2 shows the baseline benchmark of the dataset.

Table 2: Baseline result

category	score
aAcc	80.58
mIoU	67.48
mAcc	80.98
background IoU	67.47
background IoU	86.6
high vegetation IoU	67.49
high vegetation Acc	75.35

Benchmark datasets facilitates the development and testing of new methods or models to create a baseline for a particular area of interest. Applications include forestry, tree/species monitoring, agriculture natural resource management, ecological and landscape planning.

The current baseline benchmark may not be suitable for real-world application, as it does not have a high enough accuracy. Analyzing some of the errors reveals certain challenging cases, particularly images with the leaves/tree being similar to the background, making it almost indistinguishable. Another difficult case is images with heavy lighting conditions, for e.g. strong sunlight coming from an angle, and hence altering much of the appearance. However, much could be done in the future to improve the result with different machine learning techniques, such as changing the base model, as well as data augmentation techniques.

6. Conclusions

In this work, we introduce a new dataset of high-resolution drone images with varying spatial, temporal, and seasonal conditions, along with a benchmark for reliable semantic segmentation, in hopes of tackling the issues of data accessibility and fostering sustainable development. The dataset can be used by the machine learning community to further improve the accuracy of their AI model for semantic segmentation tasks. The baseline benchmark serves as a point of reference for other trained models to be evaluated against;

this encourages the community to continue pushing its boundaries, providing more improved solutions for image segmentation. This dataset also enables the development of a generalized model that can be fine-tuned on a local scale. These improvements would eventually lead to the progress of sustainable development with practical usage such as carbon stock estimation and land optimization for carbon activities.

Acknowledgments

This work was conducted with the financial support of Taighde Éireann – Research Ireland, formerly Science Foundation Ireland, under Grant No. 12/RC/2289-P2 at Insight Research Ireland Centre for Data Analytics and Grant No. 18/CRT/6223 at Research Ireland Centre for Research Training in Artificial Intelligence.

References

Lyujie Chen, Feng Liu, Yan Zhao, Wufan Wang, Xiaming Yuan, and Jihong Zhu. Valid: A comprehensive virtual aerial image dataset. In *2020 IEEE international conference on robotics and automation (ICRA)*, pages 2009–2016. IEEE, 2020.

Rossana Ducato. Data protection, scientific research, and the role of information. *Computer Law & Security Review*, 37:105412, 2020.

Tom Duckett, Simon Pearson, Simon Blackmore, Bruce Grieve, Wen-Hua Chen, Grzegorz Cielniak, Jason Cleaversmith, Jian Dai, Steve Davis, Charles Fox, et al. Agricultural robotics: the future of robotic agriculture. *arXiv preprint arXiv:1806.06762*, 2018.

Junfeng Gao, Yong Yang, Pan Lin, and Dong Sun Park. Computer vision in healthcare applications. *Journal of healthcare engineering*, 2018, 2018.

Andrew Hirons and Peter Thomas. *Applied Tree Biology*. 12 2017. ISBN 9781118296400. doi: 10.1002/9781118296387.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

Ray A Jarvis. A computer vision and robotics laboratory. *Computer*, 15(06):8–24, 1982.

Ye Lyu, George Vosselman, Gui-Song Xia, Alper Yilmaz, and Michael Ying Yang. Uavid: A semantic segmentation dataset for uav imagery. *ISPRS journal of photogrammetry and remote sensing*, 165:108–119, 2020.

Lei Ma, Yu Liu, Xueliang Zhang, Yuanxin Ye, Gaofei Yin, and Brian Alan Johnson. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing*, 152:166–177, 2019.

Gyri Reiersen, David Dao, Björn Lütjens, Konstantin Klemmer, Kenza Amara, Attila Steinegger, Ce Zhang, and Xiaoxiang Zhu. Reforestree: A dataset for estimating tropical forest carbon stock with deep learning and aerial imagery. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 12119–12125, 2022.

C Lane Scher, Emily Griffoul, and Charles H Cannon. Drone-based photogrammetry for the construction of high-resolution models of individual trees. *Trees*, 33:1385–1397, 2019.

Frank Sterck, Lars Markesteijn, Feike Schieving, and Lourens Poorter. Functional traits determine trade-offs and niches in a tropical forest community. *Proceedings of the National Academy of Sciences*, 108(51):20627–20632, 2011.

TUGRAZ. Tugraz. semantic drone dataset., 2019. URL http://dronedataset.icg.tugraz.at/.