

Vision based crown loss estimation for individual trees with remote aerial robots

Boon Ho^{a,1}, Basaran Bahadir Kocer^{a,1}, Mirko Kovac^{a,b,*}

^a Aerial Robotics Laboratory, Imperial College London, London SW7 2AZ, UK

^b Materials and Technology Centre of Robotics, Swiss Federal Laboratories for Materials Science and Technology, 8600 Dübendorf, Switzerland

ARTICLE INFO

Keywords:

Aerial robots
Unmanned aerial vehicles
Crown loss estimation
Convolutional neural network
Variational autoencoder
Foliar sampling

ABSTRACT

With the capability of capturing high-resolution imagery data and the ease of accessing remote areas, aerial robots are becoming increasingly popular for forest health monitoring applications. For example, forestry tasks such as field surveys and foliar sampling which are generally manual and labour intensive can be automated with remotely controlled aerial robots. In this study, we propose two new online frameworks to quantify and rank the severity of individual tree crown loss. The real-time crown loss estimation (RTCLE) model localises and classifies individual trees into their respective crown loss percentage bins. Experiments are conducted to investigate if synthetically generated tree images can be used to train the RTCLE model as real images with diverse viewpoints are generally expensive to collect. Results have shown that synthetic data training helps to achieve a satisfactory baseline mean average precision (mAP) which can be further improved with just some additional real imagery data. We showed that the mAP can be increased approximately from 60% to 78% by mixing the real dataset with the generated synthetic data. For individual tree crown loss ranking, a two-step crown loss ranking (TSCLR) framework is developed to handle the inconsistently labelled crown loss data. The TSCLR framework detects individual trees before ranking them based on some relative crown loss severity measures. The tree detection model is trained with the combined dataset used in the RTCLE model training where we achieved an mAP of approximately 95% suggesting that the model generalises well to unseen datasets. The relative crown loss severity of each tree is estimated, with deep representation learning, by a probabilistic encoder from a fully trained variational autoencoder (VAE) model. The VAE is trained end-to-end to reconstruct tree images in a background agnostic way. Based on a conservative evaluation, the estimated crown loss severity from the probabilistic encoder generally showed moderate agreement with the expert's estimation across all species of trees present in the dataset. All the software pipelines, the dataset, and the synthetic dataset generation can be found in [the GitHub link](#).

1. Introduction

1.1. Motivation

Forests contribute to a wide range of economical and environmental benefits for humans. They ensure the biodiversity of our ecosystem by providing natural habitat, shelter and food to wildlife ([Sustainable Forestry Social and environmental benefits of forestry, 2004](#)). Furthermore, forests are also critical in regulating water and air supply ([Sustainable Forestry Social and environmental benefits of forestry, 2004](#)). As such, it is of paramount importance to continuously monitor forest

health so that quick actions can be taken to rectify and control any arising forest disturbance. Among many tree health indicators, crown defoliation is one of the most frequently investigated parameters ([Dobbertin and Brang, 2001](#)). Many studies have been carried out to investigate the relationship between crown loss and the severity of various tree health issues caused by tree diseases, environmental induced stress and pest infestation ([Raison et al., 1992](#)). There are different ways of assessing forest tree health. For example, according to [Kälin et al. \(2019\)](#), large scale field surveys are carried out at specific sites where the percentage of crown loss would be estimated visually by trained arborists. Such field surveys are manual, time-consuming and

* Corresponding author.

E-mail address: m.kovac@imperial.ac.uk (M. Kovac).

¹ Authors contributed equally to this work.

costly (Kälin et al., 2019).

The adoption of aerial robots or unmanned aerial vehicles (UAVs) for forest health monitoring and data acquisition purposes is a more cost-effective alternative to manual field survey (Gray and Ewers, 2021). Recent advances in platform development (Zheng et al., 2020), adaptation in control algorithms (Kocer et al., 2021a), and the cheap operational and material costs of UAVs made them an appealing tool for monitoring forest health (Torresan et al., 2017). Also, UAVs are capable of acquiring high spatial resolution data. This is especially important for a more granular level of forest health monitoring, e.g. on an individual tree stand level. Depending on the UAV's cruising altitude, even specific tree structures can be seen from the acquired imagery data (Waite et al., 2019). Most UAVs can also be easily equipped with different types of sensors (Kocer et al., 2019a; Kocer et al., 2019b), mechanisms and cameras (Bayraktar et al., 2020; Xiao et al., 2021), thus adapting them for more specific use cases such as foliar sampling (Charron et al., 2020). UAVs are also suitable to be used for close inspection of vegetation in areas that are inaccessible to humans.

These advantages suggested that UAVs can be used to fully or partially automate many forestry-related applications. For example, the UAV pilot can stay at a base station, remotely control a first-person-view (FPV) UAV via teleoperation instead of being physically present in the forest, when carrying out various forestry tasks. In that context, we propose a few tools to visually assist virtual reality (VR) pilots in estimating and comparing individual tree health more efficiently.

1.2. Contribution

In this paper, our contribution is twofold. We have developed a novel methodology consisting of:

1. a real-time crown loss estimation (RTCLE) tool that estimates the crown loss percentage bins and
2. a two-step crown loss ranking (TSCLR) framework which ranks the trees based on some relative crown loss measure.

Here, we framed the below canopy crown loss estimation (CLE) task as an object detection problem where each tree would be either categorised into a crown loss percentage bin or ranked relative to all other detected trees based on their crown loss severity. We have adopted lightweight deep learning models for these detection, classification and ranking tasks. From the vision side, the virtual reality pilot would see the predicted bounding boxes and labels drawn on each tree. The input of these visual tools consists of RGB images or video frames. Hence, they are platform agnostic and compatible with any off-the-shelf consumer-grade camera including the ones on most UAVs. No specific sensors or additional channels such as LiDAR or Infrared are needed. To the best of our knowledge, this is the first implementation that leverages virtual reality aerial robots and object detection models for crown loss estimation tasks in real-time.

1.3. Overview

Tree health data sources mostly consist of satellite imagery obtained using multispectral sensors (Eitel et al., 2011; Torres et al., 2021). Depending on the considered problem, the choice of data acquisition platform can be different, e.g. satellite (Zarco-Tejada et al., 2018; Bhattarai et al., 2021), UAV (Chianucci et al., 2016; Dash et al., 2017; Zheng et al., 2021), aircrafts (Näsi et al., 2018), terrestrial (Huo and Zhang, 2019; Abbas et al., 2021), or a combination of these (Sankey et al., 2017; Campbell et al., 2021). Forest tree health indicators can be exhibited as a spectrum of plant and site characteristics, phenological and biophysical attributes, soil chemistry, air quality and the presence or absence of bioindicator plants (Shendryk et al., 2016). To obtain the measurements of these indicators, relevant sensing approaches such as radar, microwave, LiDAR, thermal, hyperspectral and multispectral

(Roth et al., 2015; Webster et al., 2018; Goodbody et al., 2018; Wagner et al., 2018; Yin and Wang, 2019; Puliti et al., 2020) are frequently used in remote sensing operations. However, the data collection with satellite or aircraft might be affected by air conditions leading to the degradation in the quality of the collected data (Dainelli et al., 2021). The use of UAVs is explored in Brovkina et al. (2018) to separate a few different forest tree species and identify dead trees. The top view and near-infrared images are used for this classification problem. Similarly, 100 m above the ground flights are programmed in Buras et al. (2018) with a multispectral camera. These infrared ranges allowed the computation of normalized difference vegetation index and the investigation of the impact of forest edge distance on pine forest's drought-induced mortality. In Safonova et al. (2019), a set of RGB images are collected 120 m above the ground to identify bark beetle infested fir trees. These efforts can be also used to generate point cloud data using structure from motion and 3D segmentation approaches (Ferraz et al., 2012; Yurtseven et al., 2019). Another advantage of UAVs usage is the ability to collect data on-demand (Guimar Aes et al., 2020). This flexibility allows the collection of temporal and spatial data for phenology monitoring and the development of microclimate models (Berra et al., 2019). Therefore, learning models can be developed with the collected data (Ma et al., 2019; Hao et al., 2021; La Rosa et al., 2021).

The choice of views and instruments used for crown and canopy related measurements often depend on the scale of interest and use cases (Leckie et al., 2005; Waser et al., 2011; Ardila et al., 2012; Duncanson et al., 2014; Blomley et al., 2017). Starting from the below canopy view, there are two types of canopy related measurements, namely the canopy closure and canopy cover (Jennings et al., 1999). Both methods involved measurements of light penetrated through the canopy. These measurements can be performed via simple visual assessment or with the aid of instruments such as the spherical densiometer and canopy-scope (Hale and Brown, 2005). However, these manual methods all suffer from systematic differences between observers and the lack of repeatability in measurements to varying degrees (Cook et al., 1973; Brown et al., 2000).

Large scale forest health monitoring is usually performed above the canopy. For example, in Chan et al. (2020), above canopy hyperspectral imagery data was used to detect ash trees and classify ash dieback severity. In Michez et al. (2016), multi-temporal high-resolution near-infrared (NIR) and visible red, green and blue channels (VIS-RGB) data were used to differentiate deciduous riparian forest species and their health conditions. Multi-spectral tree canopy aerial images are also captured and used for urban and forest tree species classification in Gini et al. (2014), Zhang et al. (2020). In Huang et al. (2018), the bias field estimation applied on the high-resolution images captured from the top of the canopy improves the segmentation of tree crowns. A structure from motion approach with above canopy photogrammetry is utilized in Khokthong et al. (2019) where an RGB camera is used to analyse oil palm canopy cover on tree mortality. A similar approach to estimate the segmented crown diameter is proposed in Yilmaz and Güngör (2019) with photogrammetric point clouds. A significant cost benefit is also reported in Navarro et al. (2020) for the UAV-based high-resolution imagery as compared to the ground-based measurements. Its LiDAR counterpart is also given in Wu et al. (2019) for the canopy cover. In this context, a comparison study is proposed in González-Jaramillo et al. (2019) to identify the accuracy of the RGB and multispectral cameras. The results suggested that above ground biomass estimation is more reliable with RGB cameras when compared to the LiDAR measurements. Furthermore, individual tree crown segmentation methods are compared in Gu et al. (2020) showing the accurate estimations with the use of spectral lightness. Another study reported in Brede et al. (2019) in favour of UAV-based LiDAR scanning as compared to the terrestrial measurements due to its faster data acquisition speed for tree volume estimation.

1.4. Related work

Ground-level tree images with associated crown loss estimation are usually produced manually from field survey work. As such, there is always inherent subjectivity of the experts' evaluations for the crown loss values as studied in [Solberg and Strand \(1999\)](#). In [Lee et al. \(2011\)](#), imagery data of Douglas-fir was collected to record the progression of fir crown loss (CL) due to Douglas-fir tussock moth infestation. There are also various works, such as [Mizoue \(2002\)](#), [Dobbertin et al. \(2004\)](#), which aimed to partially automate the CLE task performed on ground-level images. These implementations have commonly taken a multi-step approach where crown images are pre-processed, delineated and have relevant features extracted before computing for crown loss measurements. These multi-step approaches would likely impose a relatively large processing time penalty. Hence, they are deemed unsuitable for real-time and lightweight applications involving UAVs. There are a few of the under-canopy studies presented for the tree diameter measurements [Chisholm et al. \(2013\)](#), [Krisanski et al. \(2018\)](#), [Kuzelka and Surový \(2018\)](#), [Krisanski et al. \(2020\)](#) instead of crown loss estimations. Recent work such as [Kälin et al. \(2019\)](#) has adopted an end-to-end convolutional neural network to fully automate the CLE task on ground level tree images. Despite the relative success, the model can only output a single estimated crown loss value for the entire scene, which would be less appropriate if there are multiple trees in the image. Conversely, our proposed frameworks can estimate the crown loss, either in absolute crown loss percentage bin or relative rankings, of multiple trees present in a real-time video stream. Despite that, these proposed frameworks do have multiple limitations and disadvantages that are summarized in [Table 1](#).

2. Methods

An object may be perceived differently when seen from different angles or viewpoints. Similarly, trees may appear denser when viewed from certain viewpoints. We aim to develop object detection models which are robust to viewpoint changes due to UAV's flight. To achieve that, a dataset with diverse viewpoints would be needed for the training process. To the best of our knowledge, there is currently no open-source imagery dataset with diverse viewpoints for crown loss estimation purposes ([Kälin et al., 2019](#)). Manually collecting and annotating the images would be very tedious and prone to mistakes. Hence, we decided to experiment with the usage of synthetic data to increase the viewpoint diversity of training data. Also, synthetic data generation allows a more standardised and precise control on the crown loss ground truth, which would be needed for the training of the RTCLE model. This helps to reduce the amount of manual work involved in the data collection stage

Table 1
Disadvantages and Limitations of the Proposed Methodology.

RTCLE	TSCLR
Both approaches work best for sparse and individual trees.	
Model training is not trivial.	Individual tree detection with dense background is challenging.
Both models are not tree species specific.	
Training and validation involve collecting CL ground truth which is inherently subjective and may result in inconsistencies in dataset and during inference time.	Unable to predict the absolute crown loss percentage bin - relative ranking is estimated instead.
Other aspect such as tree height, lighting, species were not explored in synthetic dataset. This may limit the generalising capability of RTCLE.	If the pilot wants to optimise flight path for leaf sampling (e.g. prioritise trees with high CL ranking), this may not be suitable for field surveys where CL bins are needed for large scale comparison.

of a usual machine learning project workflow. However, training the models solely on the synthetic dataset would likely cause some serious over-fitting issues where the detection model fails to generalise beyond the dataset that was used for training. [Dwivedi et al. \(2017\)](#) has shown that a model trained on a combined (real and synthetic) dataset would outperform the same models trained on a pure synthetic dataset alone. We, therefore, gained the inspiration of mixing a small proportion (10% by dataset size relative to synthetic data) of real annotated crown loss images from the Swiss Federal Research Institute (WSL) into the synthetic dataset to build a robust detection model which is able to generalise well into real-world tree images.

2.1. WSL data

The field survey real-world images from WSL mostly consist of well-centred trees whose corresponding crown loss value has been estimated by WSL experts. The field surveys were conducted in forests across Switzerland. The captured forest scenes are complex by nature. For example, most images also captured the surrounding trees as illustrated in [Fig. 1](#). Unavoidably, there are also scenarios where the main tree of interest is partially occluded by its neighbouring tree(s). The label of each image consists of the tree species and their corresponding crown loss percentage estimated by WSL experts. The original dataset was labelled at 5% crown loss intervals. However, due to the inherent subjectivity of the CLE task, some of the labels appeared to be visually inconsistent with their respective labelling.

2.2. Synthetic data

For the generation of synthetic images, we explored the usage of a game engine, Blender for the generation of photo-realistic tree images. A Blender add-on, Modular Tree (or MTree) was used to generate the tree objects. MTree is a flexible tool that allows a procedural generation of photorealistic trees with diverse properties. The main tree nodes consist of the trunk node, branch node and tree parameters. With the trunk node, the length and the breast diameter can be changed to generate the trees in various sizes. Similarly, the curve of the body and the shape can be controlled to generate more specific trees. The branch node can define the number of branches, their angles, length, radius, and split probabilities. The shapes of the branches can be also controlled together with the gravity effect on each branch. The growing node can also provide a more distributed branch that can represent real species. After defining the basic shape of the tree, the next step is to add the leaf configuration. This process starts with configuring the twig node. This defines the shape of the leaf, its length, radius and the foliage density per twig. The last stage is to combine the twigs with the tree parameters. This effectively spawns trees with different crown loss severities. To generate a sufficiently large synthetic dataset, we first create a number of principal trees for each of the 5 crown loss bins. These crown loss bins include 0–20% (healthy tree), 20–40%, 40–60%, 60–80% and 80–100% (dead trees). These principal tree objects were created by referring to the real tree images from WSL data. Afterwards, we pan the virtual camera around the principal tree of interest and render the tree into PNG format images to generate the dataset automatically. To ensure viewpoint diversity, we considered panning the camera at different heights h , revo-



Fig. 1. Some examples of trees from WSL Dataset illustrating the complex scene in the forest.

lution α and depression β angles as well as zoom-in distance r from the tree object, as illustrated in Fig. 2. The viewpoints are defined considering the drone trajectories that we tested in the field. Lastly, we manually removed rendered images with viewpoints that do not capture the tree crown well enough. The rendered images of trees are categorised automatically based on their respective crown loss bins. Some examples of rendered trees are shown in Fig. 3.

2.3. Data annotation and pre-processing

To train the detection models in a supervised way, both the real and synthetic data have to be labelled with bounding boxes coordinates and crown loss bins. For the real dataset, we labelled some images based on the expert's estimation as shown in Fig. 4. In complex cases where multiple trees are present in the scene, we approximately identified their crown loss bins based on the knowledge we have acquired from the entire WSL dataset. For the synthetic dataset, we employed the cut and paste method as presented in Dwivedi et al. (2017) to simulate a more realistic scenario where trees with different levels of crown loss can all be present in the view. Since all images were rendered in the PNG format, we thus have an additional alpha channel on top of the usual red, green and blue (RGB) channels. This allowed us to delineate and paste the individual trees into various background images.

To automatically label the trees with bounding boxes, the region props function from Python's skimage measure module was utilised to compute the individual tree's bounding box coordinates, width and height based on the alpha channel. With this approach, we managed to completely automate the synthetic data generation and labelling process.

2.4. Detection task with you only look once model

You Only Look Once (YOLO) is a one-step object detection model Redmon et al. (2016-December (2015)). YOLO frames object detection as a regression problem where features of the entire image are used for the prediction of bounding boxes and classes probability in a single stage. This approach is different to classifiers that have been re-purposed for detection problems through a sliding window approach, such as the deformable parts model (DPM) (Felzenszwalb et al., 2009) or the two-stage object detectors such as Faster-RCNN (Ren et al., 2016) with the region proposal and the refinement of candidate bounding boxes steps happening as two stages in the framework. The one-step object detection approach gives YOLO its speed advantage making it a suitable model for real-time object detection applications. YOLO first divides a given input image into an $S \times S$ grid cells. Each grid cell is responsible for detecting an object with a centre that falls within it. The grid cell detection comprises bounding boxes properties, object confidence scores and conditional class probabilities. Concretely, B bounding boxes and their respective confidence scores are predicted in each grid cell. Each bounding box consists of five predictions, i.e. the x , y coordinates, width w , height h and the confidence c of the box itself. The object confidence score encodes the confidence that the box predicted contains an object in it, $\Pr(\text{Object})$ and the relative amount of overlapping between the

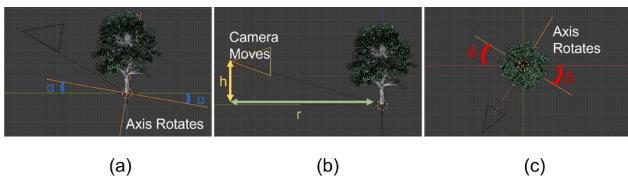


Fig. 2. The virtual camera rotates in accordance of the axes rotations i.e. the camera is fixed relative to the axes. (a) Side View: Axis rotates to adjust x - y plane's angle of inclination, α (b) Side View: Camera moves to adjust height, h and zoom-in distance, r , the axes remain fixed when the camera translates (c) Top down View: Axis rotates to adjust rotation around z axis, β .

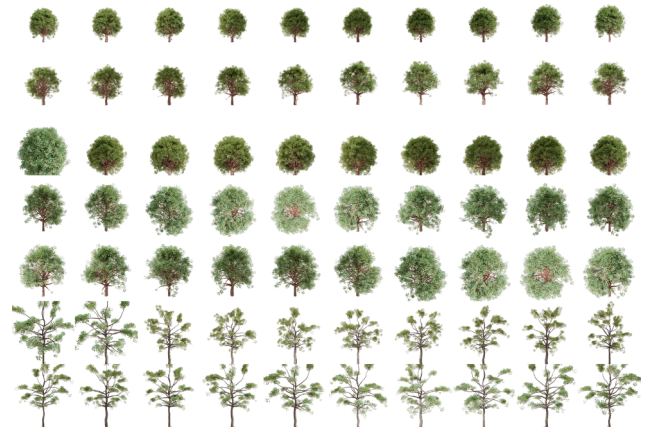


Fig. 3. Examples of individual trees rendered at different viewpoints from the virtual camera in Blender.

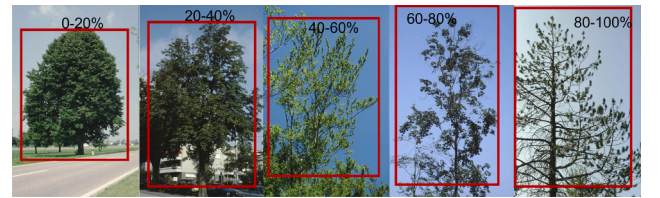


Fig. 4. Examples of trees manually labelled based on WSL expert's annotations.

predicted box and the ground truth bounding box annotation, $\text{IOU}_{\text{pred}}^{\text{truth}}$. Formally, the object confidence score is defined as

$$\Pr(\text{Object}) * \text{IOU}_{\text{pred}}^{\text{truth}} \quad (1)$$

The intuition behind this is that the object confidence score should be close to zero if there is no object in the cell, otherwise, the score will be closed to the intersection over union (IOU) score between the predicted box and the ground truth. The centre coordinates and dimensions x , y , w and h are predicted relative to the whole image's dimensions. Each grid cell also predicts C conditional class probabilities, $\Pr(\text{Class}_i|\text{Object})$, where C is the number of classes in the training dataset of YOLO and is independent of B . This means that the final output tensor of YOLO will have dimensions of $S \times S \times (5 \times B + C)$.

The class-specific confidence score for each box is calculated at test time as in

$$\Pr(\text{Class}_i) * \text{IOU}_{\text{pred}}^{\text{truth}} = \Pr(\text{Class}_i|\text{Object}) * \Pr(\text{Object}) * \text{IOU}_{\text{pred}}^{\text{truth}} \quad (2)$$

The class-specific confidence score encodes both the probability of that class appearing in the box and how well the predicted box fits the object. With multiple bounding boxes predicted, the non-maximal suppression algorithm is then applied to remove duplicated boxes based on the class-specific confidence score. A multi-part loss function is optimised during the training of the YOLO model as presented in (3) Redmon et al., 2016-December (2015):

$$\begin{aligned}
& \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{obj} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\
& + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{obj} \left[(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \right] \\
& + \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{obj} (C_i - \hat{C}_i)^2 + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{noobj} (C_i - \hat{C}_i)^2 \\
& + \sum_{i=0}^{S^2} \mathbb{1}_i^{obj} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2
\end{aligned} \quad (3)$$

where $\mathbb{1}_i^{obj}$ denotes if an object appears in grid cell i and $\mathbb{1}_{ij}^{obj}$ denotes that the j bounding box predictor in cell i is responsible for that prediction, C is the object confidence score and $p(c)$ represents the class specific confidence. Also, $\lambda_{coord} = 5.0$ and $\lambda_{noobj} = 0.5$.

(3) can be broken down into multiple parts, each of these optimises either localisation or classification. To explain the relative weightings λ_{coord} and λ_{noobj} , the majority grid cells of every image do not contain any object, this pushes (Redmon et al., 2016, 2015) the confidence scores of those cells towards zero, causing the gradient from grid cells with objects in it to be less significant. This can lead to instability in training. The model would prefer to not predict any bounding boxes as most grid cells are empty. As such, λ_{coord} and λ_{noobj} are used to increase the error weights of localisation relative to error from empty grid cell. The first two terms in (3) penalises detection error i.e. they encode the error due to the position and dimensions of the bounding boxes, from the grid cells containing an object, respectively. The third term penalises the model for detecting an object with low confidence when there is indeed an object in the corresponding grid cell. The fourth term represents the error due to misdetection of the object when there is not. Finally, the last term is the classification error. $\mathbb{1}_i^{obj}$ in the last term means that the classification error will only be considered in the optimisation process if and only if there is an object (based on ground truth) in the grid cell.

For all the object detection models used in this work, we employed a lightweight variant of YOLO, tiny YOLO v3. The architecture of tiny YOLO v3 is as shown in Fig. 5 (Gong et al., 2020). Several improvements have been implemented in tiny YOLO v3. Firstly, instead of predicting the positions and dimensions directly, YOLO v3 predicts the offsets of bounding boxes from some predefined prior bounding boxes' properties instead. These prior bounding boxes are also known as anchor boxes. The object confidence score is also predicted for each bounding box using logistic regression. The predictions are then transformed to

bounding boxes coordinates b_x and b_y , dimensions b_w and b_h and object confidence score, $\text{Pr}(\text{object})$ based on the formulae as presented, from (4)–(8)

$$b_x = \sigma(t_x) + c_x \quad (4)$$

$$b_y = \sigma(t_y) + c_y \quad (5)$$

$$b_w = p_w \exp(t_w) \quad (6)$$

$$b_h = p_h \exp(t_h) \quad (7)$$

$$\text{Pr}(\text{object}) = \sigma(t_o) \quad (8)$$

where t_x , t_y , t_w , t_h and t_o are predictions from tiny YOLO v3. c_x and c_y are the offset of the top left corner of the image. p_w and p_h are the original width and height of the prior bounding box of concerned, calculated from the training data with k -means clustering. Next, each grid cell can now predict multiple labels as each bounding box is associated with C conditional class probabilities. Instead of a softmax function, independent logistic classifiers are used to allow for multi-label prediction. During training, binary cross-entropy loss is used for the class predictions. Finally, YOLO v3 predictions happen at three different scales. The concept of how this works is similar to that of feature pyramid networks (Dollar et al., 2014). Hence, the final output tensor has dimensions of $S \times S \times (B \times (5 + C))$, where the number of bounding boxes B is chosen to be 3.

2.5. Real-time crown loss estimation model

The RTCLE model takes in an RGB video frame and outputs bounding boxes and associated crown loss percentage bin as predictions as shown in Fig. 6. For the training of the RTCLE model, we took some suggested modifications from Redmon (2013) and set the hyperparameters for tiny YOLO v3 as in Table 2.

2.6. Deep representation learning with variational autoencoder

CLE tasks are inherently very subjective. In this work, we have utilised an unsupervised deep representation learning method to learn an objective crown loss scale automatically without relying on any human estimation. A variational autoencoder (VAE) is a probabilistic variant of the vanilla autoencoder (Bank et al., 2020). The usage of a VAE (Kingma and Welling, 2019) assumes that there are several unobserved data generative factors, each controlling different aspect(s) of a given observation. The goal here is to infer the data generative factors from the observations. In the context of individual tree CLE tasks, examples of data generative factors could be tree species and foliage density. In theory, these factors control how a tree would look in the real world. Hence, the objective of using a VAE here is to deduce some relative measure of foliage density for any given tree image.

As shown in Fig. 7, the generic architecture of a VAE consists of two main components i.e. the probabilistic encoder and probabilistic decoder. In between these components, there is a latent vector which typically has a smaller dimension than the input and output. Concretely, the statistical motivation can be inferred as follow: it is supposed that z is some hidden variable that controls certain aspect(s) of observation x . The goal here is hence to compute $p(z|x)$. By Bayes' theorem,

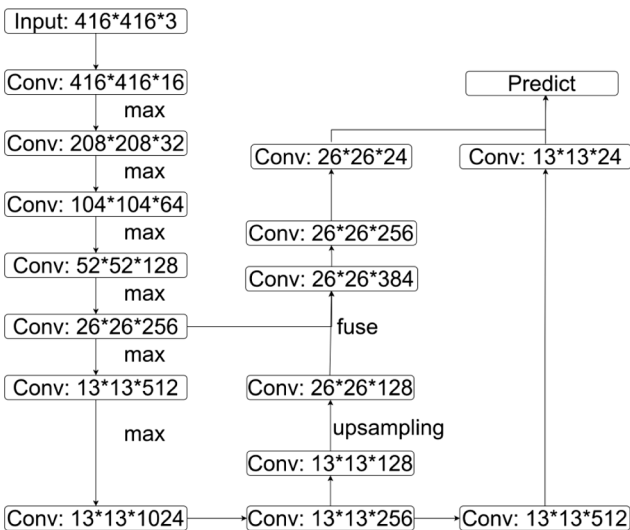


Fig. 5. Tiny YOLO v3 architecture.



Fig. 6. RTCLE model and the predicted bounding boxes.

Table 2
Hyperparameters for RTCLE and TSCLR Models.

Hyperparameter	RTCLE Value	TSCLR Value
Batch size	16	8
Subdivisions	8	8
Momentum	0.9	0.9
Decay	0.0005	0.0005
Angle	5	5
Saturation	1.5	1.5
Exposure	1.5	1.5
Hue	0.1	0.1
Learning rate	0.0001	0.0001
Burn in	1000	1000
Max batches	10000	6000
Steps	8000, 9000	4800, 5400
Scales	0.1, 0.1	0.1, 0.1

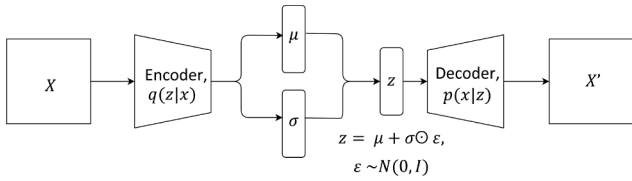


Fig. 7. The generic architecture of a variational autoencoder.

$$p(z|x) = \frac{p(x|z) \times p(z)}{p(x)} \quad (9)$$

where

$$p(x) = \int p(x|z)p(z) dz \quad (10)$$

From (10), the computation of $p(x)$ is intractable in many cases. However, variational inference can be applied for the estimation of $p(x)$. Firstly, $p(z|x)$ is approximated by $q(z|x)$ such that $q(z|x)$ has a tractable distribution. For this approximation purpose, the KL divergence between $q(z|x)$ and $p(z|x)$ distributions has to be minimised i.e.

$$\min \text{KL}(q(z|x)||p(z|x)) \quad (11)$$

This is equivalent to maximising (12)

$$E_{q(z|x)}(\log p(x|z)) - \text{KL}(q(z|x)||p(z)) \quad (12)$$

where the first term is the reconstruction likelihood (generate x based on z) and the second term is the KL divergence between $q(z|x)$ and $p(z)$. Based on the statistical motivation laid out, $q(z|x)$ can be used to infer the data generative factor, z based on the observation x . A neural net can be used to construct the probabilistic encoder $q(z|x)$ and decoder $p(x|z)$. The loss function to be minimised during the training process is given as (13)

$$-E_{q(z|x)}(\log p(x|z)) + \text{KL}(q(z|x)||p(z)) \quad (13)$$

where $p(z) \sim N(0, 1)$ is assumed. Minimising the reconstruction loss, $-E_{q(z|x)}(\log p(x|z))$ ensures that the reconstructed data is similar to the input while $\text{KL}(q(z|x)||p(z))$ imposes a penalty to the model if it encodes the input into a dense region in the latent space. Minimising (13) is equivalent as maximising the lower bound of the data log-likelihood, $p(x)$ also known as evidence lower bound. The probabilistic encoder $q(z|x)$ will output a pair of (mean μ and variance σ) parameters describing the distribution of each latent component. The decoder $p(x|z)$ will then sample from these distributions with the defined parameters for input reconstruction. It is noted that sampling is needed because z is stochastic and its true value can only be analysed statistically. Following the re-parameterisation trick (Kingma and Welling, 2019), $z = \mu +$

$\sigma \odot \epsilon$, a VAE can be optimised via back-propagation while still allowing for random sampling. A generic VAE architecture is shown in Fig. 7.

2.7. Two-step crown loss ranking framework

In the two-step crown loss ranking (TSCLR) framework, individual trees are first cropped out from a tree detector. These individual tree images are then resized and passed as an input to an encoder from a fully trained VAE. The individual crown loss severity ranking is then obtained by sorting the cropped tree images based on the latent representation predicted by the encoder. The TSCLR framework is illustrated in Fig. 8.

The tree detection model was trained on the synthetic and real combined datasets. However, the crown loss percentage bins label were simply replaced with a single label (labelled as "tree") as the tree detector's only task is to predict the bounding boxes for individual trees. The hyperparameters are summarised in Table 2.

The encoder part of a VAE is used for the relative crown loss measure estimation. The VAE was trained in an unsupervised way on the WSL dataset. For the number of dimensions on the latent variable, the initial intuition was that two latent components would be sufficient to capture the data generative factors (one for crown loss, the other for tree geometry or species (e.g. deciduous and coniferous trees)). Hence, a quick experiment was conducted with a VAE with 2-D latent space. Having trained for 300 epochs, a grid of 2-D latent vectors was passed to the decoder for the reconstruction of images to infer the data generative factor encoded in the latent representations. As shown in the reconstructed images from Fig. 9, the VAE has unexpectedly learned to encode the sky intensity, which is unimportant for the crown loss estimation task. Furthermore, the latent components are also entangled, in the sense that varying one latent component would cause changes in multiple aspects of the image. This causes the latent encoding to be less interpretable. However, by anchoring some training images on the latent space, the VAE model showed some signs that it has captured the crown loss aspect of the trees as shown in Fig. 10. Less dense trees are generally encoded with lower values of the second latent component, Z2 and vice versa. It should be emphasised that the reconstruction quality is not the main concern for this work, they are merely used to infer the data generative factor captured by the latent components.

A background subtraction training process was devised for the training of VAE to direct the model's focus on the vegetation instead of the background. During training, the VAE would be exposed to an RGB image as input and the same image, but with background artificially removed, as the reference image for reconstruction, as illustrated in Fig. 11. The motivation behind the background-subtracted away from the reference image is that the VAE would find it unimportant to encode the sky intensity in its latent representation since this information is no longer required for image reconstruction. The reconstruction loss term can hence be seen as a dissimilarity score between the reconstructed image and the reference image. For the same reason, a single node (1-D) was used as the latent component so that the encoder only encodes the most important data generative factor, the crown loss severity. During deployment, the latent component value will be used to measure the relative crown loss severity for the cropped trees from the detection model. The input image will still be an RGB image as running background subtraction algorithms on the cropped image will slow down the image processing speed which is undesirable. In this work, we have employed a VAE architecture as illustrated in Fig. 12.

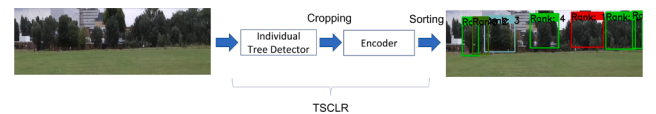


Fig. 8. TSCLR framework and the predicted crown loss severity ranking. Colour code: red (rank 1 i.e. most severe), dark green (rank 2), cyan (rank 3), bright green (rank 4 and onward i.e. less severe).

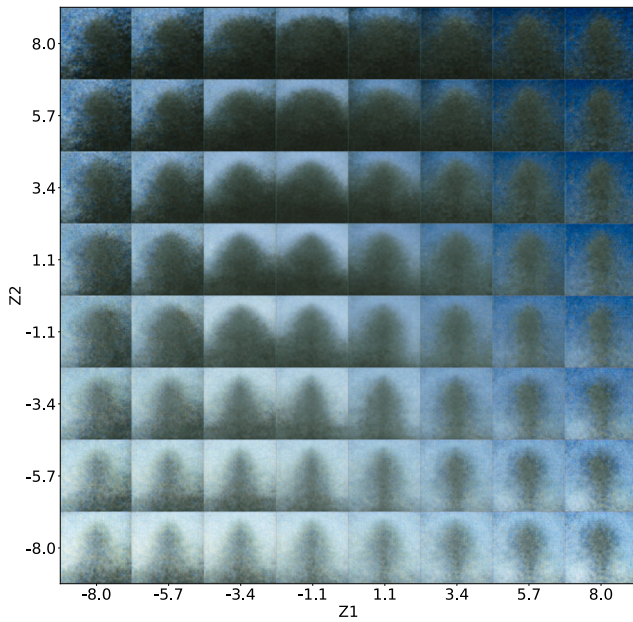


Fig. 9. Reconstruction grid of images in the latent space. Each reconstructed image corresponds to a 2D latent vector. The grid shows that both latent components are entangled. Data generative factors such as the crown loss severity and sky intensity are also captured.

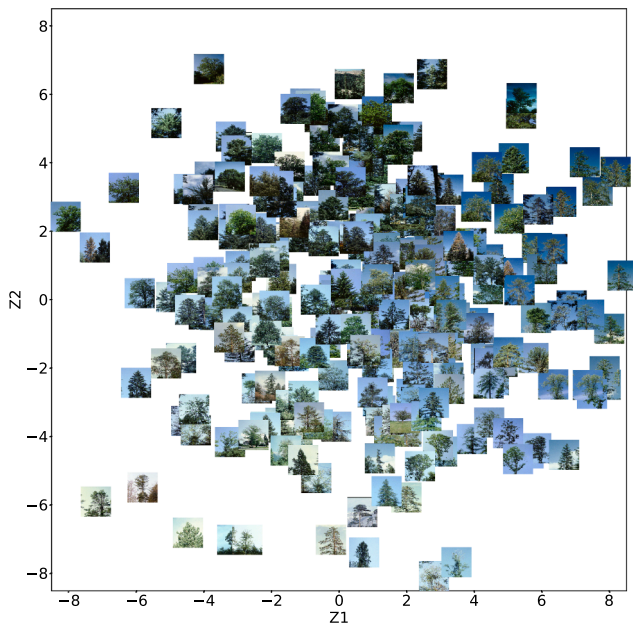


Fig. 10. Images of trees anchored onto the latent space. The second latent component Z2 does seem to capture some aspect of crown loss severity.

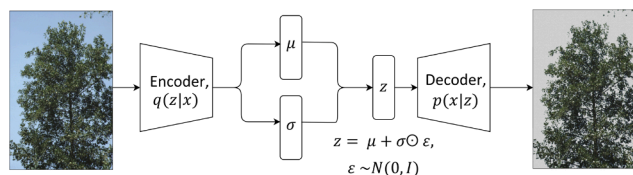


Fig. 11. Training of the VAE with an original RGB image as input and the image with background subtracted as reconstruction reference. (The background was made grey in colour for better visibility, some images do have clouds as part of the background).

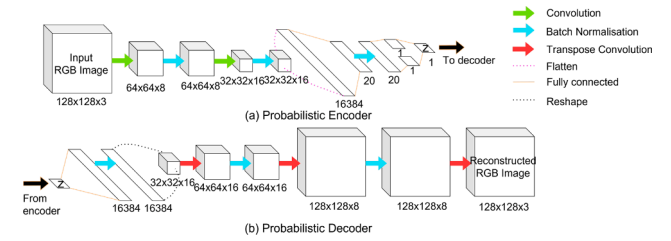


Fig. 12. Architecture of (a) probabilistic encoder (b) probabilistic decoder used in this work.

To produce the foreground mask, a few background subtraction methods were explored. These segmentation methods include *k*-means clustering colour quantisation and Otsu thresholding on various vegetation indices (VI) and colour space. The quality of these unsupervised segmentation masks was visually assessed over a large number of images. It was found that relatively consistent segmentation masks can be produced by applying *k*-means clustering on the RGB channels and Otsu thresholding on the Green Leaf Index (GLI) mapped image and the lightness channel (V channel of the hue, saturation and value (HSV) colour space). Hence, a few experiments were carried out to find out if the choice of background subtraction method affects the crown loss estimator performance in terms of how it ranks the trees present in the scene.

For evaluation of the crown loss rankings, ten random trees were repeatedly sampled (10,000 times) with each species from the unseen WSL dataset. We rank these trees based on the latent encoding node value obtained from the probabilistic encoder. This evaluation was repeated for the encoder from the VAE model trained with the aforementioned background subtractive methods. The Kendall rank correlation coefficient, τ metric between the encoder's ranking and the crown loss severity ranking based on WSL experts' estimation. Since the sampling was performed repeatedly, the result would be presented as a distribution of τ metric for each tree species.

2.8. Aerial robotic platforms

The purpose of the proposed RTCLE and TSCLR frameworks is to facilitate individual tree health comparison and quantification using crown loss severity as the health indicator. This provides visual clues to remote drone pilots who observe the scene from an FPV using a VR headset, signalling to them which individual tree is potentially unhealthy and requires a closer visual inspection. This can boost the efficiency of tasks such as foliage sampling from targeted trees at inaccessible areas. The collected samples from these trees can provide critical information regarding specific tree diseases and the tree immune system.

These proposed frameworks were integrated and experimented on the aerial robot platforms as depicted in Fig. 13. The first one is the small size system that weighs around 1 kg. The endowed passive mechanisms are designed and used to decrease the foliage sample collection time for forestry researchers. The second system is a palm-sized drone that weighs less than 250 g. It is also endowed with smaller sample collection



Fig. 13. Aerial robotic platforms: (i) left - palm-sized platforms; (ii) right - small size platforms. These platforms are designed to detect the crown loss and inform the user to approach the tree for the sampling on-site.

mechanisms to fly through the denser canopies to collect samples from various parts of the tree. The small size aerial robot is endowed with a USB stick which provides the internet connection to stream the data with GStreamer. Similarly, a ground computer is used for the palm-sized system to facilitate the internet connection to stream the data remotely.

3. Results and discussions

3.1. RTCLE

A few datasets are used for the experiments. We have trained the RTCLE model on purely synthetic data, a small amount of real WSL data and a combined dataset consisting of both the real and synthetic datasets. The training was conducted for 10,000 iterations, the training loss and validation mean average precision (mAP) values are kept track of during the course of training, as shown in Fig. 14 (a) and (b) respectively. The weights of Tiny YOLO v3 is saved into a backup folder only when there is an improvement on the mAP metric.

Based on Fig. 14 (a), the training losses fall sharply at an early stage of training, giving an early signal that the training process would eventually converge in the end. The mAP metric evaluated on the validation dataset rise rapidly from the start of training. Then, the mAP improvement slows down beyond 3000 iterations. Beyond 6000 iterations of training, the mAPs start to plateau implying that the training processes have converged.

On average, across all crown loss bins, the RTCLE model trained on the combined dataset was able to generalise better than the same model trained on the synthetic dataset alone, indicated by its higher validation mAP towards the end of training in Fig. 14 (b). On the other hand, training with a small amount of real tree images has resulted in over-fitting. Although the training loss converged at an early stage of the training, the mAP remains very low throughout the training, implying that the model has failed to generalise beyond the training data. This is unsurprising as the amount of data is simply too little for the training process.

A few insights are gained from the training process. Firstly, the synthetic dataset alone has contributed substantially to producing a detection model with satisfactory performance. Secondly, a small amount of real-world dataset alone is insufficient to build a robust RTCLE model, but a substantial increase in performance can be achieved when we combine it with even more synthetic data. Hence, annotated real-world data is still necessary to produce a robust model and could be the key to improving the performance. However, these experiments have showcased that the manual effort for tree images collection and annotation could be reduced significantly with procedurally generated synthetic data.

Testing the RTCLE models on an independent dataset does further verified the intuition developed from the training loss plots. Fig. 15 (a) highlights the low average precision particularly at lower crown loss percentage bins for training on a pure synthetic dataset. This may be

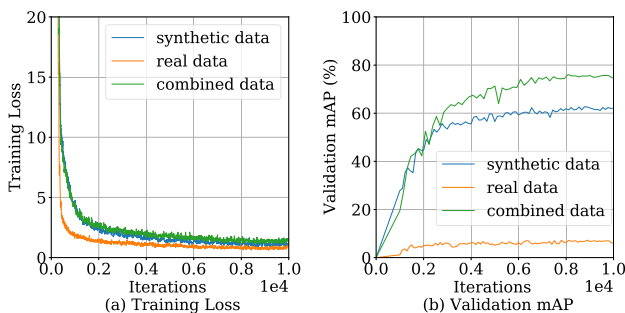


Fig. 14. Synthetic dataset: Tiny YOLO v3 training loss and mean average precision(mAP) curves.

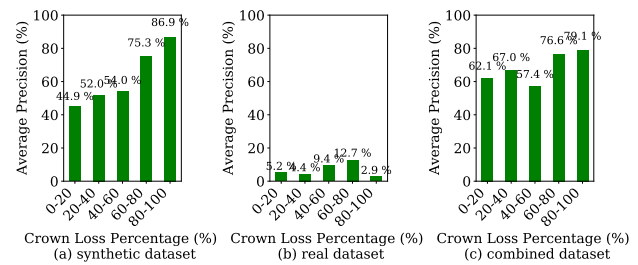


Fig. 15. Test average precision of RTCLE model on each crown loss bin (a) pure synthetic dataset, (b) real WSL dataset and (c) combined dataset.

attributed to synthetic trees appearing to be denser at certain view-points. Based on Fig. 15 (a) and (c), on average the combined dataset achieved better average precision for all crown loss bins except for 80–100% bin when compared to a pure synthetic dataset. This can be attributed to the lack of completely defoliated trees in the real world data included in the combined dataset. The real dataset, as expected, does not perform well due to its quantity which was insufficient for training.

3.2. TSCLR

The first step of the TSCLR framework is to detect individual trees. The individual tree detection model was trained on the combined dataset containing real and synthetic images of trees. The training loss and validation mAP were kept track and shown in Fig. 16. Based on the validation mAP, the model has reached convergence at around 2000 iterations of training, although subsequent iterations of training did result in a diminishing decrease in training loss. The validation mAP was around 95–96% after the model has converged. This is close to the mAP from the test dataset evaluation, 96.33%, suggesting that the model generalises well to unseen datasets. This huge leap in performance compared to the RTCLE model also showed that crown loss severity classification was responsible for the relatively low mAP in previous RTCLE related experiments.

The VAE was instantiated and trained with background subtractive methods for 100 epochs, the validation loss was kept tracked and weights are saved whenever the minimum validation loss decreases. Fig. 17 (a), (b) and (c) show the reconstructed images by the probabilistic decoder trained with background subtracting methods with V-

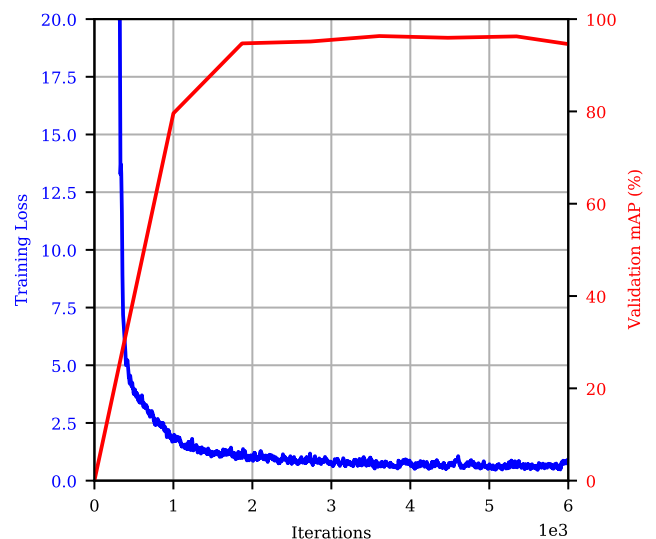


Fig. 16. Training loss and validation mean average precision (mAP) for individual tree detectors.

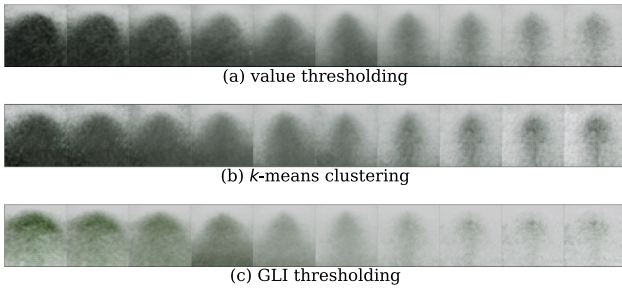


Fig. 17. Reconstruction of images, by the probabilistic decoder, based on equally spaced latent encodings ranging from -1.5 to 1.5 . The respective decoder was obtained by background subtraction training of VAE based on (a) V-channel masking (b) k -means clustering colour masking and (c) GLI masking.

channel masking, k -means clustering and GLI masking respectively. Generally, one can infer from the reconstructions that the crown loss severity aspect has indeed been encoded in the latent space. Going from left to right of the reconstructed images, foliage density decreases with the increasing value of the latent component.

The unseen images of trees from WSL data are passed to the probabilistic encoder for the computation of the latent component. Repeated sampling of tree images was carried out 10,000 times (with 10 trees per sampling) for each tree species. The τ metric was computed based on the ranking by latent component and ranking by WSL ecologist's estimation. It is possible to have ties from the WSL dataset since the crown loss percentages are in 5% intervals, so the comparisons performed here are conservative ones. Fig. 18 shows the overlay Kendall Tau's distribution

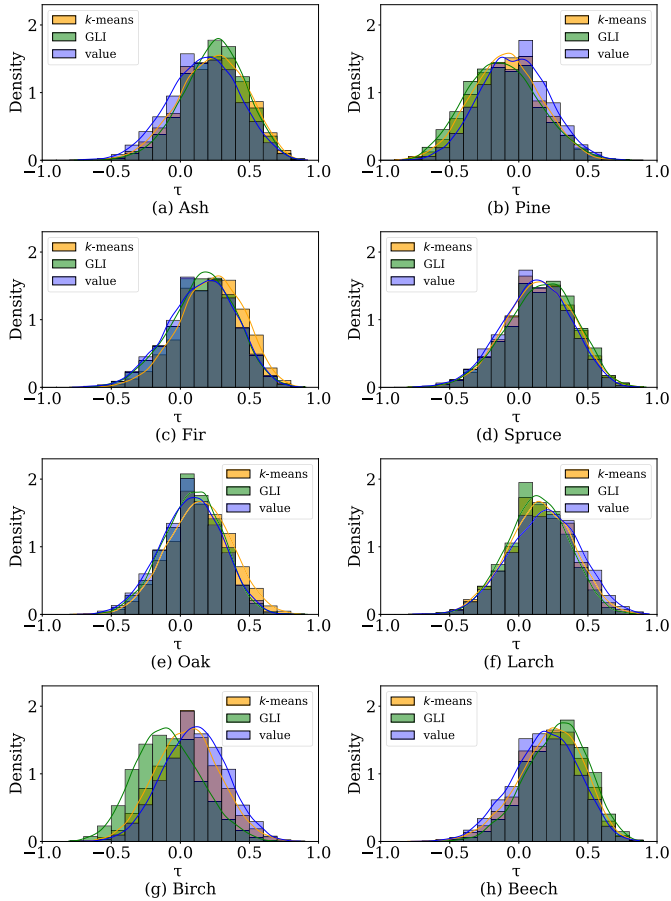


Fig. 18. Overlay of Kendall Tau. (a) Ash, (b) Pine, (c) Fir, (d) Spruce, (e) Oak, (f) Larch, (g) Birch and (h) Beech.

from repeated sampling and ranking comparisons for the considered background subtraction methods for various tree species such as ash, pine, fir, spruce, larch and birch. The crown loss rankings by the probabilistic encoder are generally in positive association with the ranking by WSL experts, that is, most distributions peak at positive τ . Tree species such as ash, fir, spruce and beech peaked at around $\tau = 0.25$ exhibiting moderate association with their respective ground truth rankings. In general, it was also observed that k -Means and GLI masking result in slightly better agreement than V channel masking as these distributions can be seen having a higher density of positive τ in Fig. 18 (a), (c), (e), (g) and (h) with the exception that GLI masking's moderate disagreement with the ground truth ranking for Birch tree ash shown in Fig. 18 (g). It should be emphasised that the ground truth crown loss estimation refers to a specific candidate tree in the image but there are images with multiple trees in it, resulting in potential crown loss severity underestimation by the probabilistic encoder. For example, in Fig. 19, the trees are sorted in increasing crown loss severity by the latent encoding estimation (GLI masking). Visually, such crown loss ranking seems appropriate. However, in the first few images, the ground truth labelling specifically refers to one of the trees which was highly occluded by its neighbouring trees. The encoder does not take this into account, hence it underestimated the crown loss severity, causing a major ranking disagreement with the expert ($\tau = -0.459$). This is illustrated by the reconstructed images which suggested that the first few images were encoded as low crown loss severity.

3.3. Field deployment and discussions

We have collected some field data video streams with our proposed systems, tested them on both RTCLE and TSCLR frameworks and presented some snapshots side-by-side in this section for visual comparisons. Note for TSCLR, the ranking is presented in descending order of relative crown loss severity.

To illustrate that the frameworks are platform-agnostic, we have inspected some video frames from our two platforms. Example frames overlaid with bounding boxes predictions are presented in Fig. 20 and Fig. 21 which were captured by Intel RealSense D455 and a cheaper Tello drone camera respectively. As it can be seen from these images, the predictions are visually representative of an expert view and our methodology is independent of platform and camera settings. Field data is inherently complex due to different outdoor illuminations and dense backgrounds. We have chosen some cases where the background is dense with varying illumination conditions. Fig. 22 illustrates one of these cases. Furthermore, we searched for the trees that have a lower density of crown. Fig. 23 shows one of the snapshots from our online video stream. The RTCLE model could detect the severe crown loss and the TSCLR framework have successfully produced a visually coherent relative crown loss ranking.

Our frameworks only require RGB input channels. Hence, their cost requirements are lower compared to similar counterparts for CLE tasks including LiDAR and infrared sensors. Furthermore, these detection frameworks operate on individual tree crown levels. Individual crown loss bins or rankings are predicted for each detected tree instead of outputting a single regression value for the whole scene with potentially



Fig. 19. Upper row: Trees sorted in order of increasing latent encoding from left to right (increasing relative crown loss severity). Latent representation is estimated from the probabilistic encoder trained with GLI masking method. Bottom row: Respective reconstruction of the trees based on the encoded latent representation.

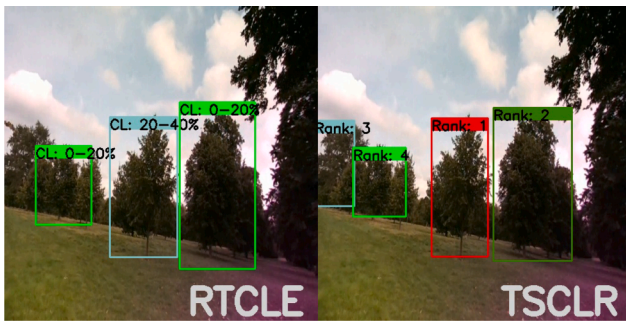


Fig. 20. Field data with D455 camera.

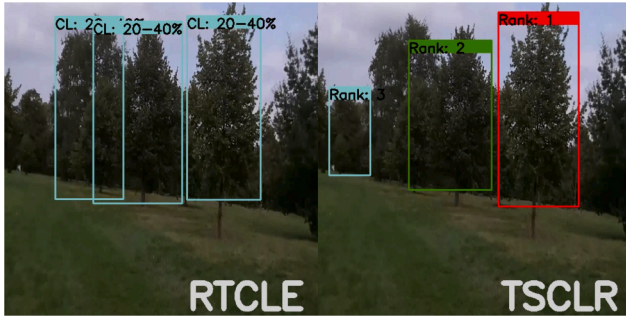


Fig. 21. Field data with Tello camera.

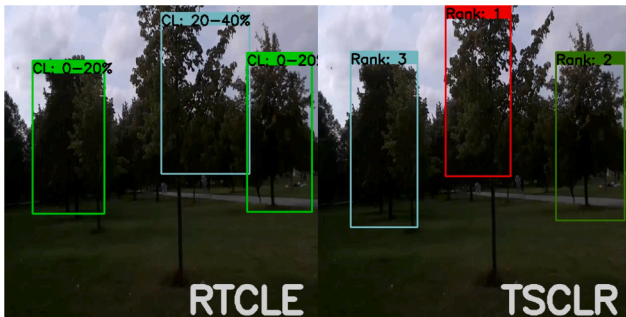


Fig. 22. Field data with denser background and various illumination conditions.



Fig. 23. Field data with sparse crown cases.

many trees or no tree at all as presented in Kälén et al. (2019). From the prediction speed aspect, both frameworks achieved near real-time 20 to 25 FPS for frame-by-frame prediction without leveraging a GPU.

To the best of our knowledge, we are the first to approach the online CLE task by formulating it as an object detection problem. Hence, it is very challenging to compare our methodology with the other works

since the labelling and the problem formulations are different. For example, the same WSL dataset used in Kälén et al. (2019) formulates the CLE problem as a regression problem where a single crown loss ground truth is assigned to each image with potentially more than one tree. Direct comparisons between our work and theirs cannot be conducted due to different problems formulation and the unavailability of bounding boxes labelling ground truth.

Furthermore, we proposed the use of a synthetic dataset to account for viewpoint variations and to ease the manual labelling workload. It can be seen in Fig. 23 that the tree detector in the TSCLR framework was able to localise trees at varying distances. Besides, the inherent CLE subjectivity has proved to be a challenging problem for many supervised machine learning approaches due to the resulting inconsistency present in the ground truth labelling. In this context, our contribution with TSCLR autonomously learns a scale for CLE tasks without human bias since it does not rely on any expert's input for training.

In order to leverage the remote operation, we used the pipeline illustrated in Fig. 24. This framework can be enriched with multiple sensors as desired but it might require the use of a higher payload platform as detailed in Orr et al. (2021).

3.3.1. Synthetic dataset evaluations

To test the proposed methodology in a more controlled environment, we expanded evaluations considering various factors. The first one is based on the viewpoint variation on a single tree that is illustrated in Fig. 25. For example, this result shows consistency in our approach with different view angles, distances and changes in backgrounds. Since the aerial images with a front view might include unstructured elements as well, we also generated multiple tree views to test the algorithm. Fig. 26 shows such a case where the camera view is blocked with trunks and there are other trees next to the single tree. Our approach may not capture all the trees in this type of environment but the results of the estimation values are sufficiently accurate and consistent. To illustrate how our methodology can fail, we have tested more scenarios considering occluded trees on top of each other. Fig. 27 shows the case of two trees with different interlocking degrees. The left side of the image still shows a sufficiently good estimation for this case but the right side only detects a value while missing the right estimation on the tree.

3.3.2. Field dataset evaluations

We have selected some representative images to show the performance of our approach on the real dataset. Fig. 28 shows our estimation values which are representative and these values can provide the user to select which tree to approach for a foliage sample collection case. These

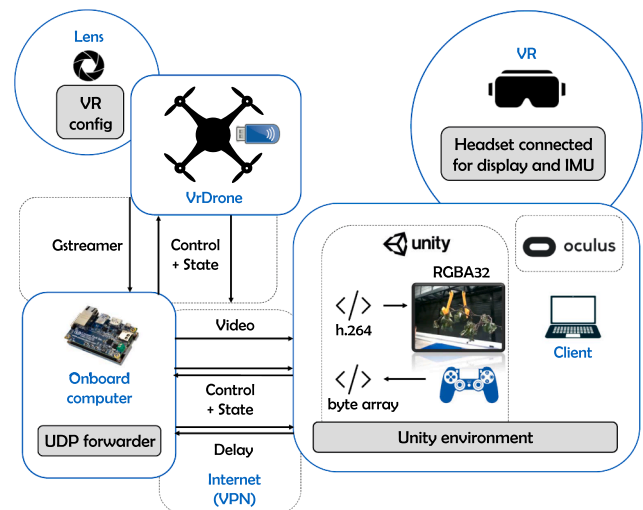


Fig. 24. Teleoperation pipeline: the user can see the video stream from the operated location to conduct the sample collection task (Kocer et al., 2021b).



Fig. 25. Various viewpoints in Blender.

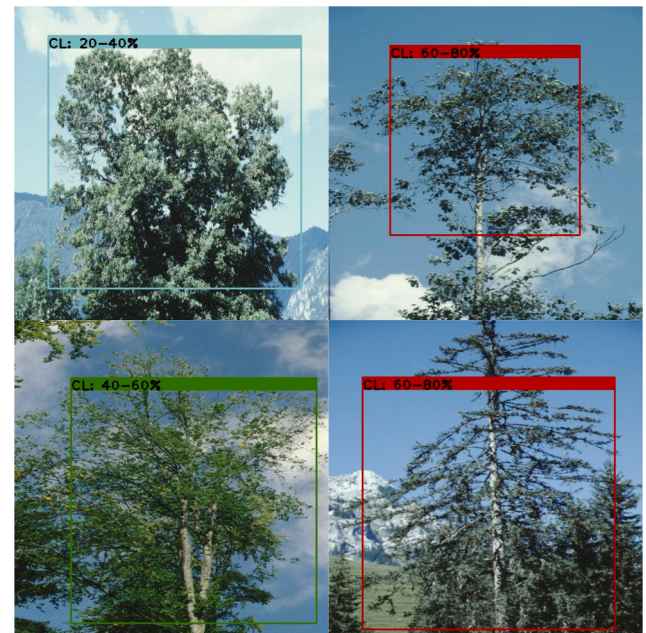


Fig. 28. Selected WSL field data.

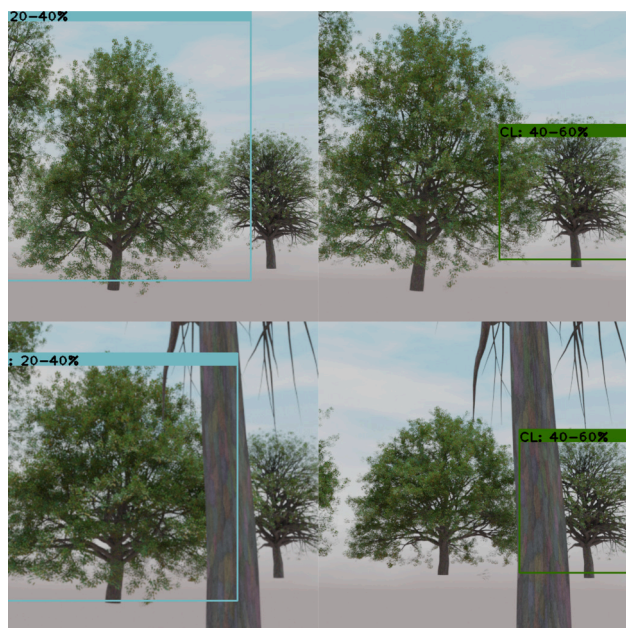


Fig. 26. Unstructured environment in Blender.



Fig. 27. Interlocking trees in Blender.

tests revealed that the estimations output from our framework indeed capture the crown loss values defined by experts and our approach can serve as a stepping stone for aerial robot deployment for remote tree health operations.

Additional tests carried out with rendered scene and field scenes showed that the RTCLE model is able to generalise to a satisfactory level on a real scenario. However, when the scene is unstructured, the model struggled to contain all individual trees in their respective bounding boxes. The estimated crown loss bins are visually consistent across many different images in the synthetic scene and real scenario field tests. The model also showed great generalisation capability in terms of view angles and distances variation.

In terms of usability, the model performs best when the background is clear. This property is similar to that of a human field expert. For example, most tree images in WSL field survey data has a clear background despite being located in a forest. This is because even a human expert failed to estimate crown loss objectively when the background of the candidate tree is dense and complex. Thus, the RTCLE model is on par with a human expert's judgement. Such tool shows great potential to be used for field surveys automation.

4. Conclusions

Forestry related crown loss estimated tasks are mostly manual and time-consuming. In this work, we have presented two object detection based machine learning approaches for real-time crown loss estimation applications suitable to be used on any drone teleoperation platform.

The RTCLE model is a one-step object detection approach that performs object detection in near real-time. For the RTCLE model, we have trained a Tiny YOLO v3 model on purely synthetic data, pure real data and a combination of both to explore the potential of replacing manual data collection from field surveys. From the training loss and test results, synthetic data alone has shown great potential in getting crown loss estimation to a satisfactory level of performance. With just a small amount of manually labelled data, the RTCLE model managed to generalise better beyond the training dataset.

Crown loss estimation by visual assessment often results in very subjective annotations. Training a machine learning model on an inconsistently labelled dataset may do more harm than good to the model. This motivated us to come out with an unsupervised learning

method that does not need to rely on any human's annotated ground truth in the training process. As a result, the crown loss estimation by the machine learning model also happens on a relative scale. Hence, we compared the crown loss severity ranking from the unsupervised prediction with the ranking by human's experts. To achieve this, we have formulated a two-step crown loss ranking framework. The first step of this framework is to detect and crop the individual trees present in an image. The tree detection model was trained on a mixture of real and synthetic tree images. The tree detection model achieved a test mAP of 96.33%. Afterwards, the crown loss severity ranking step involves sorting the trees detected from the first step based on their estimated crown loss severity. Background subtraction methods were employed to train the variational autoencoder. Our proposed solutions generally showed moderate agreement with human experts' estimation.

To improve the test performance, the RTCLE model may need a more representative training dataset, either via synthetic generation or manually collected from the real world. We have explored the generation of synthetic data with diverse viewpoints via Blender. Similarly, the real dataset to be collected would also ideally account for the different viewpoints where an aerial robot would encounter while approaching trees. Semi-supervised learning approaches (Sohn et al., 2020) could be employed, using the trained model as the teacher model, to ease the manual labelling workload. As for the synthetic data, more principal trees can be generated with the guidance of ecologist experts to introduce a more variety of tree images in the training dataset. Furthermore, more time can also be devoted to generating different tree species and geometries so that the RTCLE model can learn a joint distribution for different tree species for the crown loss estimation task. For the crown loss ranking model, other generative models such as bidirectional generative adversarial network (BiGAN) (Donahue et al., 2016) can be experimented to obtain an alternative latent representation for crown loss prediction. One of the extensions of this study will include crown loss estimation for different tree species as this is a more accurate representation of most multi-species forest (Hastings et al., 2020).

We will also extend our work for forestry mission planning considering battery lifetime (Kocer et al., 2019c) and the information-rich trajectories (Jeon et al., 2020).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was partially supported by funding from EPSRC (award No. EP/N018494/1, EP/R026173/1, EP/R009953/1, EP/S031464/1, EP/W001136/1), NERC (award No. NE/R012229/1) and the EU H2020 AeroTwin project (grant ID 810321). Mirko Kovac is supported by the Royal Society Wolfson fellowship (RSWF/R1/18003). The visual evaluations were based on data from the Swiss Long-term Forest Ecosystem Research programme LWF (www.lwf.ch), which is part of the UNECE Co-operative Programme on Assessment and Monitoring of Air Pollution Effects on Forests ICP Forests (www.icp-forests.net). We are in particular grateful to C Hug for providing WSL dataset. We also thank Dr. Richard Buggs for valuable comments and discussions on the associated problem.

References

Abbas, S., Peng, Q., Wong, M.S., Li, Z., Wang, J., Ng, K.T.K., Kwok, C.Y.T., Hui, K.K.W., 2021. Characterizing and classifying urban tree species using bi-monthly terrestrial hyperspectral images in hong kong. *ISPRS J. Photogramm. Remote Sens.* 177, 204–216. <https://doi.org/10.1016/j.isprsjprs.2021.05.003>.

Ardila, J.P., Bijker, W., Tolpekin, V.A., Stein, A., 2012. Quantification of crown changes and change uncertainty of trees in an urban environment. *ISPRS J. Photogramm. Remote Sens.* 74, 41–55. <https://doi.org/10.1016/j.isprsjprs.2012.08.007>.

Bank, D., Koenigstein, N., Giryas, R., 2020. Autoencoders, arXiv preprint arXiv: 2003.05991.

Bayraktar, E., Basarkan, M.E., Celebi, N., 2020. A low-cost uav framework towards ornamental plant detection and counting in the wild. *ISPRS J. Photogramm. Remote Sens.* 167, 1–11. <https://doi.org/10.1016/j.isprsjprs.2020.06.012>.

Berra, E.F., Gaulton, R., Barr, S., 2019. Assessing spring phenology of a temperate woodland: A multiscale comparison of ground, unmanned aerial vehicle and landsat satellite observations. *Remote Sens. Environ.* 223, 229–242. <https://doi.org/10.1016/j.rse.2019.01.010>.

Bhattarai, R., Rahimzadeh-Bajgiran, P., Weiskittel, A., Meneghini, A., MacLean, D.A., 2021. Spruce budworm tree host species distribution and abundance mapping using multi-temporal sentinel-1 and sentinel-2 satellite imagery. *ISPRS J. Photogramm. Remote Sens.* 172, 28–40. <https://doi.org/10.1016/j.isprsjprs.2020.11.023>.

Blomley, R., Hovi, A., Weinmann, M., Hinz, S., Korpela, I., Jutzi, B., 2017. Tree species classification using within crown localization of waveform lidar attributes. *ISPRS J. Photogramm. Remote Sens.* 133, 142–156. <https://doi.org/10.1016/j.isprsjprs.2017.08.013>.

Brede, B., Calders, K., Lau, A., Raunonen, P., Bartholomeus, H.M., Herold, M., Kooistra, L., 2019. Non-destructive tree volume estimation through quantitative structure modelling: Comparing uav laser scanning with terrestrial lidar. *Remote Sens. Environ.* 233, 111355. <https://doi.org/10.1016/j.rse.2019.111355>.

Brovkina, O., Cienciala, E., Surový, P., Janata, P., 2018. Unmanned aerial vehicles (uav) for assessment of qualitative classification of norway spruce in temperate forest stands. *Geo-spatial Inform. Sci.* 21 (1), 12–20. <https://doi.org/10.1080/10095020.2017.1416994>.

Brown, N., Jennings, S., Wheeler, P., Nabe-Nielsen, J., 2000. An improved method for the rapid assessment of forest understorey light environments. *J. Appl. Ecol.* 37 (6), 1044–1053. <https://doi.org/10.1046/j.1365-2664.2000.00573.x>.

Buras, A., Schunk, C., Zeiträg, C., Herrmann, C., Kaiser, L., Lemme, H., Straub, C., Taeger, S., Gößwein, S., Klemm, H.-J., et al., 2018. Are scots pine forest edges particularly prone to drought-induced mortality? *Environ. Res. Lett.* 13 (2), 025001. <https://doi.org/10.1088/1748-9326/aa0b04>.

Campbell, M.J., Dennison, P.E., Kerr, K.L., Brewer, S.C., Anderegg, W.R., 2021. Scaled biomass estimation in woodland ecosystems: Testing the individual and combined capacities of satellite multispectral and lidar data. *Remote Sens. Environ.* 262, 112511. <https://doi.org/10.1016/j.rse.2021.112511>.

Chan, A.H., Barnes, C., Swinfield, T., Coomes, D.A., 2020. Monitoring ash dieback (*Hymenoscyphus fraxineus*) in British forests using hyperspectral remote sensing. *Remote Sens. Ecol. Conserv.* <https://doi.org/10.1002/rse2.190>.

Charron, G., Robichaud-Courteau, T., La Vigne, H., Weintraub, S., Hill, A., Justice, D., Bélanger, N., Desbiens, A.L., 2020. The Deleaves: A UAV device for efficient tree canopy sampling. *J. Unmanned Veh. Syst.* 8 (3), 245–264. <https://doi.org/10.1139/juvs-2020-0005>.

Chianucci, F., Disperati, L., Guzzi, D., Bianchini, D., Nardino, V., Lastrì, C., Rindinella, A., Corona, P., 2016. Estimation of canopy attributes in beech forests using true colour digital images from a small fixed-wing uav. *Int. J. Appl. Earth Observ. Geoinform.* 47, 60–68. <https://doi.org/10.1016/j.jag.2015.12.005>.

Chisholm, R.A., Cui, J., Lum, S.K., Chen, B.M., 2013. Uav lidar for below-canopy forest surveys. *J. Unmanned Veh. Syst.* 1 (01), 61–68. <https://doi.org/10.1139/juvs-2013-0017>.

Cook, J.G., Stutzman, T.W., Bowers, C.W., Brenner, K.A., Irwin, L.L., 1973. Spherical Densimeters Produce Biased Estimates of Forest Canopy Cover. *Bulletin* 23 (4), 711–717.

Dainelli, R., Toscano, P., Gennaro, S.F.D., Matrese, A., 2021. Recent advances in unmanned aerial vehicles forest remote sensing—a systematic review. part ii: Research applications. *Forests* 12 (4), 397. <https://doi.org/10.3390/f12040397>.

Dash, J.P., Watt, M.S., Pearse, G.D., Heaphy, M., Dungey, H.S., 2017. Assessing very high resolution uav imagery for monitoring forest health during a simulated disease outbreak. *ISPRS J. Photogramm. Remote Sens.* 131, 1–14. <https://doi.org/10.1016/j.isprsjprs.2017.07.007>.

Dobbertin, M., Brang, P., 2001. Crown defoliation improves tree mortality models. *For. Ecol. Manage.* 141 (3), 271–284. [https://doi.org/10.1016/S0378-1127\(00\)00335-2](https://doi.org/10.1016/S0378-1127(00)00335-2).

Dobbertin, M., Hug, C., Mizoue, N., 2004. Using slides to test for changes in crown defoliation assessment methods. Part I: Visual assessment of slides. *Environ. Monit. Assessm.* 98 (1–3), 295–306. <https://doi.org/10.1023/B:EMAS.00000038192.84631.B6>.

Dollar, P., Appel, R., Belongie, S., Perona, P., 2014. Fast feature pyramids for object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 36 (8), 1532–1545. <https://doi.org/10.1109/TPAMI.2014.2300479>.

Donahue, J., Krähenbühl, P., Darrell, T., 2016. Adversarial feature learning, arXiv preprint arXiv:1605.09782.

Duncanson, L., Cook, B., Hurr, G., Dubayah, R., 2014. An efficient, multi-layered crown delineation algorithm for mapping individual tree structure across multiple ecosystems. *Remote Sens. Environ.* 154, 378–386. <https://doi.org/10.1016/j.rse.2013.07.044>.

Dwivedi, D., Misra, I., Hebert, M., 2017. Cut, paste and learn: Surprisingly easy synthesis for instance detection. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1301–1310. <https://doi.org/10.1109/ICCV.2017.146>.

Eitel, J.U., Vierling, L.A., Litvak, M.E., Long, D.S., Schulthess, U., Ager, A.A., Krofcheck, D.J., Stoscheck, L., 2011. Broadband, red-edge information from satellites improves early stress detection in a new mexican conifer woodland. *Remote Sens. Environ.* 115 (12), 3640–3646. <https://doi.org/10.1016/j.rse.2011.09.002>.

Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D., 2009. Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (9), 1627–1645. <https://doi.org/10.1109/TPAMI.2009.167>.

- Ferraz, A., Bretar, F., Jacquemoud, S., Gonçalves, G., Pereira, L., Tomé, M., Soares, P., 2012. 3-d mapping of a multi-layered mediterranean forest using als data. *Remote Sens. Environ.* 121, 210–223. <https://doi.org/10.1016/j.rse.2012.01.020>.
- Gini, R., Passoni, D., Pinto, L., Sona, G., 2014. Use of unmanned aerial systems for multispectral survey and tree classification: A test in a park area of northern Italy. *Eur. J. Remote Sens.* 47 (1), 251–269. <https://doi.org/10.5721/EuJRS20144716>.
- Gong, X., Ma, L., Ouyang, H., 2020. An improved method of tiny yolov3. In: *IOP Conference Series: Earth and Environmental Science*, vol. 440. IOP Publishing, p. 052025. <https://doi.org/10.1088/1755-1315/440/5/052025>.
- González-Jaramillo, V., Fries, A., Bendix, J., 2019. Agb estimation in a tropical mountain forest (tmf) by means of rgb and multispectral images using an unmanned aerial vehicle (uav). *Remote Sens.* 11 (12), 1413. <https://doi.org/10.3390/rs11121413>.
- Goodbody, T.R., Coops, N.C., Hermosilla, T., Tompalski, P., McCartney, G., MacLean, D. A., 2018. Digital aerial photogrammetry for assessing cumulative spruce budworm defoliation and enhancing forest inventories at a landscape-level. *ISPRS J. Photogramm. Remote Sens.* 142, 1–11. <https://doi.org/10.1016/j.isprsjprs.2018.05.012>.
- Gray, R.E., Ewers, R.M., 2021. Monitoring forest phenology in a changing world. *Forests* 12 (3), 297. <https://doi.org/10.3390/f12030297>.
- Gu, J., Grybas, H., Congalton, R.G., 2020. A comparison of forest tree crown delineation from unmanned aerial imagery using canopy height models vs. spectral lightness. *Forests* 11 (6), 605. <https://doi.org/10.3390/f11060605>.
- Guimar Aes, N., Pádua, L., Marques, P., Silva, N., Peres, E., Sousa, J.J., 2020. Forestry remote sensing from unmanned aerial vehicles: A review focusing on the data, processing and potentialities. *Remote Sens.* 12 (6), 1046. <https://doi.org/10.3390/rs12061046>.
- Hale, S.E., Brown, N., 2005. Use of the canopy-scope for assessing canopy openness in plantation forests. *For. Int. J. For. Res.* 78 (4), 365–371. <https://doi.org/10.1093/FORESTRY/CP043>.
- Hao, Z., Lin, L., Post, C.J., Mikhailova, E.A., Li, M., Chen, Y., Yu, K., Liu, J., 2021. Automated tree-crown and height detection in a young forest plantation using mask region-based convolutional neural network (mask r-cnn). *ISPRS J. Photogramm. Remote Sens.* 178, 112–123. <https://doi.org/10.1016/j.isprsjprs.2021.06.003>.
- Hastings, J.H., Ollinger, S.V., Ouimette, A.P., Sanders-DeMott, R., Palace, M.W., Ducey, M.J., Sullivan, F.B., Basler, D., Orwig, D.A., 2020. Tree species traits determine the success of lidar-based crown mapping in a mixed temperate forest. *Remote Sens.* 12 (2), 309. <https://doi.org/10.3390/rs12020309>.
- Huang, H., Li, X., Chen, C., 2018. Individual tree crown detection and delineation from very-high-resolution uav images based on bias field and marker-controlled watershed segmentation algorithms. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 11 (7), 2253–2262. <https://doi.org/10.1109/JSTARS.2018.2830410>.
- Huo, L., Zhang, X., 2019. A new method of equiangular sectorial voxelization of single-scan terrestrial laser scanning data and its applications in forest defoliation estimation. *ISPRS J. Photogramm. Remote Sens.* 151, 302–312. <https://doi.org/10.1016/j.isprsjprs.2019.03.018>.
- Jennings, S.B., Brown, N.D., Sheil, D., 1999. Introduction Assessing forest canopies and understorey illumination: canopy closure, canopy cover and other measures. *Tech. Rep.* 1.
- Jeon, B.-F., Shim, D., Kim, H.J., 2020. Detection-aware trajectory generation for a drone cinematographer. In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 1450–1457. <https://doi.org/10.1109/IROS45743.2020.9341368>.
- Kälin, U., Lang, N., Hug, C., Gessler, A., Wegner, J.D., 2019. Defoliation estimation of forest trees from ground-level images. *Remote Sens. Environ.* 223, 143–153. <https://doi.org/10.1016/j.rse.2018.12.021>.
- Khokhthong, W., Zemp, D.C., Irawan, B., Sundawati, L., Kreft, H., Hölscher, D., 2019. Drone-based assessment of canopy cover for analyzing tree mortality in an oil palm agroforest. *Front. For. Global Change* 2, 12. <https://doi.org/10.3389/ffgc.2019.00012>.
- Kingma, D.P., Welling, M., 2019. An introduction to variational autoencoders. *arXiv preprint arXiv:1906.02691*.
- Kocer, B.B., Tjahjowidodo, T., Pratama, M., Seet, G.G.L., 2019a. Inspection-while-flying: An autonomous contact-based nondestructive test using uav-tools. *Autom. Constr.* 106, 102895. <https://doi.org/10.1016/j.autcon.2019.102895>.
- Kocer, B.B., Tiryaki, M.E., Pratama, M., Tjahjowidodo, T., Seet, G.G.L., 2019b. Aerial robot control in close proximity to ceiling: A force estimation-based nonlinear mpc. In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 2813–2819. <https://doi.org/10.1109/IROS40897.2019.8967611>.
- Kocer, B.B., Kumtepli, V., Tjahjowidodo, T., Pratama, M., Tripathi, A., Lee, G.S.G., Wang, Y., 2019c. Uav control in close proximities-ceiling effect on battery lifetime. In: *2019 2nd International Conference on Intelligent Autonomous Systems (ICoIAS)*. IEEE, pp. 193–197. <https://doi.org/10.1109/ICoIAS.2019.00041>.
- Kocer, B.B., Hady, M.A., Kandath, H., Pratama, M., Kovac, M., 2021a. Deep neuromorphic controller with dynamic topology for aerial robots. In: *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 110–116. <https://doi.org/10.1109/ICRA48506.2021.9561729>.
- Kocer, B.B., Ho, B., Zhu, X., Zheng, P., Farinha, A., Xiao, F., Stephens, B., Wiesemüller, F., Orr, L., Kovac, M., 2021b. Forest drones for environmental sensing and nature conservation. In: *2021 Aerial Robotic Systems Physically Interacting with the Environment (AIRPHARO)*, pp. 1–8. <https://doi.org/10.1109/AIRPHARO52252.2021.9571033>.
- Krisanski, S., Del Perugia, B., Taskhiri, M.S., Turner, P., 2018. Below-canopy uas photogrammetry for stem measurement in radiata pine plantation. In: *Remote Sensing for Agriculture, Ecosystems, and Hydrology XX*, vol. 10783. International Society for Optics and Photonics, p. 1078309. <https://doi.org/10.1117/12.2325480>.
- Krisanski, S., Taskhiri, M.S., Turner, P., 2020. Enhancing Methods for Under-Canopy Unmanned Aircraft System Based Photogrammetry in Complex Forests for Tree Diameter Measurement. *Remote Sens.* 12 (10), 1652. <https://doi.org/10.3390/rs12101652>.
- Kuzelka, K., Surový, P., 2018. Mapping forest structure using uas inside flight capabilities. *Sensors* 18 (7), 2245. <https://doi.org/10.3390/s18072245>.
- La Rosa, L.E.C., Sothe, C., Feitosa, R.Q., de Almeida, C.M., Schimanski, M.B., Oliveira, D. A.B., 2021. Multi-task fully convolutional network for tree species mapping in dense forests using small training hyperspectral data. *ISPRS J. Photogramm. Remote Sens.* 179, 35–49. <https://doi.org/10.1016/j.isprsjprs.2021.07.001>.
- Leckie, D.G., Gougeon, F.A., Tinis, S., Nelson, T., Burnett, C.N., Paradine, D., 2005. Automated tree recognition in old growth conifer stands with high resolution digital imagery. *Remote Sens. Environ.* 94 (3), 311–326. <https://doi.org/10.1016/j.rse.2004.10.011>.
- Lee, Y.-J., Alfaro, R.I., Sickle, G.A.V., 2011. Tree-crown defoliation measurement from digitized photographs. *Can. J. For. Res.* 13 (5), 956–961. <https://doi.org/10.1139/X83-127>.
- Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* 152, 166–177. <https://doi.org/10.1016/j.isprsjprs.2019.04.015>.
- Michez, A., Piégay, H., Lisein, J., Claessens, H., Lejeune, P., 2016. Classification of riparian forest species and health condition using multi-temporal and hyperspatial imagery from unmanned aerial system. *Environ. Monit. Assess.* 188 (3), 1–19. <https://doi.org/10.1007/s10661-015-4996-2>.
- Mizoue, N., 2002. CROCO: Semi-automatic Image Analysis System for Crown Condition Assessment in Forest Health Monitoring. *J. For. Plan.* 8 (1), 17–24. <https://doi.org/10.20659/jfp.8.1.17>.
- Näsi, R., Honkavaara, E., Blomqvist, M., Lyytikäinen-Saarenmaa, P., Hakala, T., Viljanen, N., Kantola, T., Holopainen, M., 2018. Remote sensing of bark beetle damage in urban forests at individual tree level using a novel hyperspectral camera from uav and aircraft. *Urban For. Urban Green.* 30, 72–83. <https://doi.org/10.1016/j.ufug.2018.01.010>.
- Navarro, A., Young, M., Allan, B., Carnell, P., Macreadie, P., Ierodiaconou, D., 2020. The application of unmanned aerial vehicles (uavs) to estimate above-ground biomass of mangrove ecosystems. *Remote Sens. Environ.* 242, 111747. <https://doi.org/10.1016/j.rse.2020.111747>.
- Orr, L., Stephens, B., Kocer, B.B., Kovac, M., 2021. A high payload aerial platform for infrastructure repair and manufacturing. In: *2021 Aerial Robotic Systems Physically Interacting with the Environment (AIRPHARO)*, pp. 1–6. <https://doi.org/10.1109/AIRPHARO52252.2021.9571052>.
- Puliti, S., Dash, J.P., Watt, M.S., Breidenbach, J., Pearse, G.D., 2020. A comparison of uav laser scanning, photogrammetry and airborne laser scanning for precision inventory of small-forest properties. *For. Int. J. For. Res.* 93 (1), 150–162. <https://doi.org/10.1093/forestry/cpz057>.
- Raison, R.J., Khanna, P.K., Benson, M.L., Myers, B.J., McMurtrie, R.E., Lang, A.R., 1992. Dynamics of *Pinus radiata* foliage in relation to water and nitrogen stress: II. Needle loss and temporal changes in total foliage mass. *For. Ecol. Manage.* 52 (1–4), 159–178. [https://doi.org/10.1016/0378-1127\(92\)90500-9](https://doi.org/10.1016/0378-1127(92)90500-9).
- Redmon, J., 2013–2016. Darknet: Open source neural networks in c. <http://pjreddie.com/darknet/>.
- Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2015. You Only Look Once: Unified, Real-Time Object Detection. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2016-December*, pp. 779–788. URL: <https://arxiv.org/abs/1506.02640v5>.
- Ren, S., He, K., Girshick, R., Sun, J., 2016. Faster r-cnn: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (6), 1137–1149. <https://doi.org/10.1109/TPAMI.2016.2577031>.
- Roth, K.L., Roberts, D.A., Dennison, P.E., Alonzo, M., Peterson, S.H., Beland, M., 2015. Differentiating plant species within and across diverse ecosystems with imaging spectroscopy. *Remote Sens. Environ.* 167, 135–151. <https://doi.org/10.1016/j.rse.2015.05.007>.
- Safonova, A., Tabik, S., Alcaraz-Segura, D., Rubtsov, A., Baglinets, Y., Herrera, F., 2019. Detection of fir trees (*abies sibirica*) damaged by the bark beetle in unmanned aerial vehicle images with deep learning. *Remote Sens.* 11 (6), 643. <https://doi.org/10.3390/rs11060643>.
- Sankey, T., Donager, J., McVay, J., Sankey, J.B., 2017. Uav lidar and hyperspectral fusion for forest monitoring in the southwestern usa. *Remote Sens. Environ.* 195, 30–43. <https://doi.org/10.1016/j.rse.2017.04.007>.
- Shendryk, I., Broich, M., Tulbure, M.G., McGrath, A., Keith, D., Alexandrov, S.V., 2016. Mapping individual tree health using full-waveform airborne laser scans and imaging spectroscopy: A case study for a floodplain eucalypt forest. *Remote Sens. Environ.* 187, 202–217. <https://doi.org/10.1016/j.rse.2016.10.014>.
- Sohn, K., Zhang, Z., Li, C.-L., Zhang, H., Lee, C.-Y., Pfister, T., 2020. A simple semi-supervised learning framework for object detection. *arXiv preprint arXiv:2005.04757*.
- Solberg, S., Strand, L., 1999. Crown density assessments, control surveys and reproducibility. *Environ. Monit. Assess.* 56 (1), 75–86. <https://doi.org/10.1023/A:1005980326079>.
- Sustainable Forestry Social and environmental benefits of forestry, 2004. URL: www.forestry.gov.uk/sustainableforestry.
- Torresan, C., Berton, A., Carotenuto, F., Di Gennaro, S.F., Gioli, B., Matese, A., Miglietta, F., Vagnoli, C., Zaldei, A., Wallace, L., 2017. Forestry applications of UAVs in Europe: a review. *Int. J. Remote Sens.* 38 (8–10), 2427–2447. <https://doi.org/10.1080/01431161.2016.1252477>.

- Torres, P., Rodes-Blanco, M., Viana-Soto, A., Nieto, H., García, M., 2021. The role of remote sensing for the assessment and monitoring of forest health: A systematic evidence synthesis. *Forests* 12 (8), 1134. <https://doi.org/10.3390/f12081134>.
- Wagner, F.H., Ferreira, M.P., Sanchez, A., Hirye, M.C., Zortea, M., Gloor, E., Phillips, O. L., de Souza Filho, C.R., Shimabukuro, Y.E., Arag Ao, L.E., 2018. Individual tree crown delineation in a highly diverse tropical forest using very high resolution satellite images. *ISPRS J. Photogramm. Remote Sens.* 145, 362–377. <https://doi.org/10.1016/j.isprsjprs.2018.09.013>.
- Waite, C.E., van der Heijden, G.M., Field, R., Boyd, D.S., 2019. A view from above: Unmanned aerial vehicles (UAVs) provide a new tool for assessing liana infestation in tropical forest canopies. *J. Appl. Ecol.* 56 (4), 902–912. <https://doi.org/10.1111/1365-2664.13318>.
- Waser, L.T., Ginzler, C., Kuechler, M., Baltsavias, E., Hurni, L., 2011. Semi-automatic classification of tree species in different forest ecosystems by spectral and geometric variables derived from airborne digital sensor (ads40) and rc30 data. *Remote Sens. Environ.* 115 (1), 76–85. <https://doi.org/10.1016/j.rse.2010.08.006>.
- Webster, C., Westoby, M., Rutter, N., Jonas, T., 2018. Three-dimensional thermal characterization of forest canopies using uav photogrammetry. *Remote Sens. Environ.* 209, 835–847. <https://doi.org/10.1016/j.rse.2017.09.033>.
- Wu, X., Shen, X., Cao, L., Wang, G., Cao, F., 2019. Assessment of individual tree detection and canopy cover estimation using unmanned aerial vehicle based light detection and ranging (uav-lidar) data in planted forests. *Remote Sens.* 11 (8), 908. <https://doi.org/10.3390/rs11080908>.
- Xiao, F., Zheng, P., di Tria, J., Kocer, B.B., Kovac, M., 2021. Optic flow-based reactive collision prevention for mavs using the fictitious obstacle hypothesis. *IEEE Robot. Autom. Lett.* 6 (2), 3144–3151. <https://doi.org/10.1109/LRA.2021.3062317>.
- Yilmaz, V., Güngör, O., 2019. Estimating crown diameters in urban forests with unmanned aerial system-based photogrammetric point clouds. *Int. J. Remote Sens.* 40 (2), 468–505. <https://doi.org/10.1080/01431161.2018.1562255>.
- Yin, D., Wang, L., 2019. Individual mangrove tree measurement using uav-based lidar data: Possibilities and challenges. *Remote Sens. Environ.* 223, 34–49. <https://doi.org/10.1016/j.rse.2018.12.034>.
- Yurtseven, H., Akgul, M., Coban, S., Gulci, S., 2019. Determination and accuracy analysis of individual tree crown parameters using uav based imagery and obia techniques. *Measurement* 145, 651–664. <https://doi.org/10.1016/j.measurement.2019.05.092>.
- Zarco-Tejada, P., Hornero, A., Hernández-Clemente, R., Beck, P., 2018. Understanding the temporal dimension of the red-edge spectral region for forest decline detection using high-resolution hyperspectral and sentinel-2a imagery. *ISPRS J. Photogramm. Remote Sens.* 137, 134–148. <https://doi.org/10.1016/j.isprsjprs.2018.01.017>.
- Zhang, C., Xia, K., Feng, H., Yang, Y., Du, X., 2020. Tree species classification using deep learning and RGB optical images obtained by an unmanned aerial vehicle. *J. For. Res.* 1, 3. <https://doi.org/10.1007/s11676-020-01245-0>.
- Zheng, P., Tan, X., Kocer, B.B., Yang, E., Kovac, M., 2020. Tilt drone: A fully-actuated tilting quadrotor platform. *IEEE Robot. Autom. Lett.* 5 (4), 6845–6852. <https://doi.org/10.1109/LRA.2020.3010460>.
- Zheng, J., Fu, H., Li, W., Wu, W., Yu, L., Yuan, S., Tao, W.Y.W., Pang, T.K., Kanniah, K.D., 2021. Growing status observation for oil palm trees using unmanned aerial vehicle (uav) images. *ISPRS J. Photogramm. Remote Sens.* 173, 95–121. <https://doi.org/10.1016/j.isprsjprs.2021.01.008>.