UNIVERSITY OF CAPE TOWN

DEPARTMENT OF COMPUTER SCIENCE

CS/IT Honours Final Paper 2019

Title: Improving tree segmentation by combining U-Nets with engineered features

Author: Fergus Strangways-Dixon STRFER001

Project Abbreviation: TREESEG

Supervisor(s): Patrick Marais & Deshen Moodley

Category	*Min*	*Max*	Chosen
Requirement Analysis and Design	*0*	*20*	
Theoretical Analysis	*0*	*25*	
Experiment Design and Execution	*0*	*20*	
System Development and Implementation	*0*	*20*	
Results, Findings and Conclusion	*10*	*20*	
Aim Formulation and Background Work	*10*	*15*	
Quality of Paper Writing and Presentation	*10*		10
Quality of Deliverables	*10*		10
Overall General Project Evaluation (*this section allowed only with motivation letter from supervisor*)	*0*	*10*	
Total marks		**80**	

Improving tree segmentation by combining U-Nets with engineered features

UCT Honours Final Paper

Fergus Strangways-Dixon
Computer Science
University of Cape Town
South Africa
strfer001@myuct.ac.za

ABSTRACT

Convolution Neural Networks (CNN) are an industry-standard variant of neural networks for the classification and manipulation of two-dimensional image data. Although they are able to reach high levels of accuracy, they require large data sets to train on before becoming acceptably accurate. The U-Net variation of CNNs is a novel architecture designed to significantly reduce training times while preserving equivalent accuracy metrics. Little is understood about the effects of image pre-processing techniques on these models. An investigation into small to medium scale image transformation techniques and their effect on accuracy and loss of the U-Net model in the context of tree canopy segmentation is performed. This investigation establishes that the model reaches acceptable segmentation results with non-processed images, however, the addition of Morphological Closing and Histogram Equalization as additional layers to the base image results in higher accuracy, and lower cross-entropy and therefore higher confidence of the model when performing segmentation on tree canopy images. The most challenging areas for the model to segment around are ground vegetation at the base of the tree, and shadows caused by the tree. These processing techniques create a more contiguous and outlined tree canopy for the model to segment, therefore making it more confident when segmenting a canopy as well as reducing false negatives produced by shadows and ground vegetation.

CCS CONCEPTS

• **Computing Methodologies** → Machine Learning → Segmentation • **Applied Computing** → Agriculture

KEYWORDS

Artificial Intelligence, Image Processing, Feature Extraction, Colour Spaces, Segmentation, Vegetation

1 Introduction

There are many existing segmentation algorithms applied in the agricultural industry, these algorithms are designed to identify tree crops from the underlying landscape in an image. Due to the nature of the industry, image data is captured by remote-controlled drones capable of travelling the distances required. Once the images are captured, the images can be processed by a variation of a Convolutional Neural Network (CNN) called a U-Net to perform the segmentation. A CNN is a neural network designed specifically to exploit and analyse the 2-dimensional nature of image data. CNNs often have the drawback of requiring vast amounts of training data to become accurate, however the U-Net architecture was designed to reach industry-standard accuracy with training sets as small as 300 images. A robust solution will be able to reliably segment the images during any time of day, in any season. This may be achievable through vast amounts of training data captured at various times and seasons, however, this will greatly increase the cost of training the model and may not be guaranteed to work. This paper investigated alternative image representations that positively impact the accuracy of the model while keeping training costs to a minimum. While the architecture of the U-Net is important, the focus of this paper is the impact various techniques of pre-processing have on the accuracy of the model compared to the base RGB image. Specifically, image processing methods are separated into small scale transformations, such as colour space manipulation, Principal Component Analysis and Independent Component Analysis, as well as medium scale transformations, such as Mean Shift and Histogram Equalization transforms. The images provided for experimentation also included non-standard layers such as Near Infrared and height mapping layers. The purpose of this research is to determine which of these transformation methods increase accuracy by 1% or higher or reduce loss and cross-entropy by 0.05 or more when compared to the control models to justify them as a useful transformation for the purpose of improving the segmentation abilities of the U-Net model.

This research forms part of a wider evaluation of these image transformations on a variety of CNN architectures, and so the image transformations were implemented by the research group. Results from the group are yet to be published.

2 Background and Related Work

Since the development of digital image formats, researchers have manipulated image data for a variety of purposes. A large portion of this research was transforming the colour space of the existing image formats into new colour spaces for a wide variety of use cases. A detailed description of these colour spaces is defined

further in this paper, as well as their potential advantages towards more accurate tree segmentation. Furthermore, niche research groups have developed a wide variety of algorithms at various scales to extract useful features from an image, such as histogram equalization for contrast adjustment.

2.1 Per-Pixel Transformations

Per-pixel transformations are manipulations of the data in a single pixel and are often colour space transformations. These transforms convert image data from a source, usually RGB, into various other formats for different use cases.

The most common colour space is RGB, as this is the format computer screens require to display data. This format works well for describing light as components of 3 LEDs embedded in display pixels, however, is not a direct representation of how a human eye perceives light. This colour space can be useful for the task of image segmentation as the Green component can be weighted as the most important value of the pixel in the CNN, using the Red and Blue as auxiliary components to help isolate background data.

The L*a*b* colour space is closer to human vision as it includes a Lightness channel, as well as a* and b* chrominance channels. This allows us to separate the light factor from the colour values themselves, not possible in RGB. Hernández-Hernández et al. [6] tests a wide variety of colour spaces in the context of weed detection in agriculture and found their technique resulted in 99.2% accuracy in the L*a*b* colour space. Xiaosong et al. [13] found that the hues of tree canopies are predominantly found in the negative end of the a-channel, they then performed 2-dimensional OTSU segmentation of automatic threshold in the a-channel in order to reliably segment trees from an image. Xiao-Song Wang et al. [12] corroborate these findings, stating that differences in pixels in the L*a*b* colour space represents an equivalent difference in the human eyes' visual system and resulting in more accurate segmentation.

Another industry standard for a digital format of human-like vision representation is the HSI colour space. The Hue component represents the dominant wavelength of the colour, the Saturation component is the relative purity of the Hue, and Intensity of zero describes pure white, and one describes pure black. L. Tang et al. [11] show that it is possible to decouple the intensity component to better represent how human vision perceives light in a digital manner, they also provide criticism of the HSI space by describing issues with sensor noise and minor reflectance variations leading to instabilities in the converted images, which can have serious negative effects on single dimension segmentation algorithms, however they find normalization of the pixel values in HSI colour space can theoretically lead to higher accuracy in higher-dimensional segmentation strategies.

G. Ruiz-Ruiz et al. [9] found that due to the separable nature of the HSI colour space, they were able to reduce the computational time of their clustering process and Bayesian classifier as they required only 2 and 1 components of the HSI space respectively. The removal of the Intensity component also increased accuracy for images with varying illumination levels present in real farm fields. This strategy resulted in a 25 times improvement in segmentation time compared to the original RGB segmentation strategy without a significant loss of accuracy. Dianyuan Han et al. [5] used an interesting approach to manipulate the HSI format for segmenting based on a dominant colour, green in the case of vegetation segmentation. The hue of tree canopies in an image was established, and the difference of each pixel to that hue was calculated and normalised, and used for segmentation. As with the L*a*b* space, Liying Zheng et al. [14] found that they could isolate green in the HSI colour space by the Hue component being roughly 120◦. They then used this in a mean-shift segmentation strategy discussed further in this paper to achieve better segmentation results for green vegetation.

A new approach to vegetation detection and segmentation is the use of infrared image overlays over original data, giving the segmentation an additional attribute to segment on. A. Colturato et al. [15] use drone-mounted FLIR cameras to detect potential diseases in vegetation, specifically tree trunks. This additional spectrum of light provided to the CNN can aid in both traditional segmentation, especially in detecting trees from background vegetation, as well as the additional feature of being able to assess the overall health of an area when compared to past data.

Principal Component Analysis (PCA), as well as Independent Component Analysis (ICA), are other image transformations tested in the following experiments. PCA is useful for compressing and collating data points. The transformation result is designed such that the first component explains the largest part of the variance in the data as possible, with further components following as long as they are orthogonal to preceding components [16]. This will highlight to the CNN where areas have the most variation and highlight edges of tree canopies. ICA is useful for separating sections of the data, specifically separating statistically independent portions of the light spectrum and eliminating noise from a mixed sensor source [17]. Adding these 2 layers to the base RGB channels should give more dimensions for the U-Net to make predictions on, and improve accuracy.

2.2 Medium-scale transformations

Medium-scale transforms are algorithms that take a local neighbourhood of pixels and perform operations on them, rather than the per-pixel operations. This is done with the goal of extracting features from the image, such as edge detection, or normalisation of outlying pixel values. This research examines the effects of Mean Shift, Histogram Equalization, Edge Detector, and Morphological Closing on the U-Net model's accuracy.

The Mean Shift algorithm estimates the average of surrounding pixels at a point. In practice, this results in homogenising colours in the image. This should result in the canopies of trees all becoming 1-pixel value representing green, and thus easier for the U-Net to identify. [18]

Histogram Equalization algorithms seek to enhance the contrast of a given image by spreading out the most frequent intensity values, theoretically allowing for a more distinguishable edge to tree canopies and thus allowing the U-Net to notice borders more effectively. [19]

Edge Detectors, specifically Canny edge detection performs a complex series of filters, gradient analysis and thresholding to detect the edges of objects in an image. Although the U-Net may eventually learn rules like this, providing the result as a channel in an image may significantly reduce training time required, as well as improving accuracy. [21]

Morphological closing uses mathematical dilation and erosion techniques on an image to fill in gaps of a grayscale image. This is expected to make the canopies of trees smoother, as gaps between branches and leaves should be removed, resulting in a more continuous surface for the U-Net to make predictions on, theoretically making segmentation simpler. [20]

2.3 U-Net architecture

Neural networks are machine learning models that consist of layers of neurons, with weighted connections between layers determining values at each layer. Data is entered at one side of the model and fed through these layers, with output being a segmentation mask in the context of this research. Convolutional Neural Networks (CNN) are variations of neural networks characterized by having at least one convolutional layer [7]. These convolutional layers aggregate neighbouring data points to draw relations between them.

U-Nets are a novel architecture that is based on a CNN proposed by Olaf Ronneberger et al. [8] initially for use in the medical field. The architecture connects all layers to each other, rather than the data following a linear path through each layer to the next. Their focus was segmenting cells from medical images, however, the task of segmenting cells in an image is similar to segmenting trees from a field.

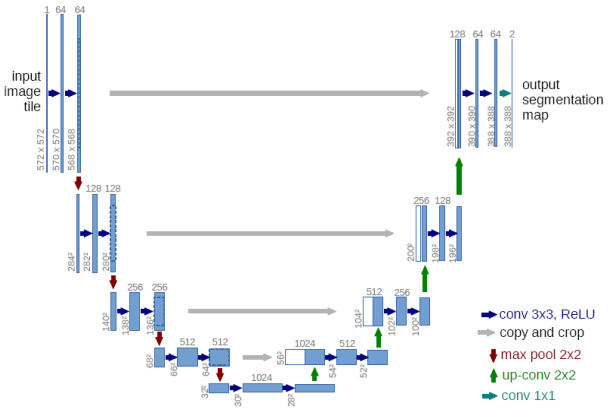


Figure 1: U-net architecture (example for 32x32 pixels in the lowest resolution). [8]

Figure 1 shows the architecture of a U-Net model. The differentiating factor of a U-Net is the down and up convolutional layers, that are also connected across the U. Due to this the model requires far fewer training images and results in higher accuracy on segmentation than existing industry-standard CNN models. [8]

3 Experimental Framework

3.1 Design

The experimental framework was divided into 3 layers: the image processing layer, the AI training harness, and the presentation layer as can be seen in figure 2. This research paper forms part of a larger investigation into the effects of image transformations on CNNs. The initial image processing layer was developed in cooperation with other student researchers investigating the effects of these same transformations on Fully Convolutional Neural Networks, and Atrous Neural Networks.

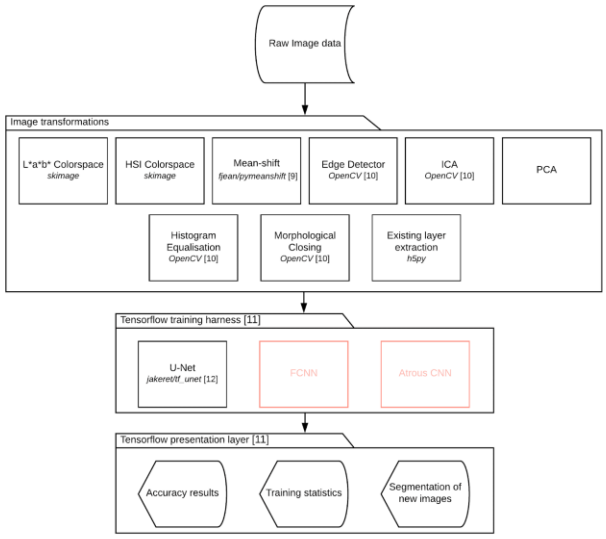


Figure 2: Framework architecture

The image processing layer takes HDF5 files as input, as this is the default format supplied by Aerobotics. These image files have many layers included, namely RGB, DEM, NDVI, NIR, and the ground truth segmentation mask. The image processing layer is responsible for reading in a batch of these images, extracting the specified existing layers, as well as performing small and medium scale transformations on these layers as additional channels to the image, and save it in a variety of formats. For example, a user may specify the RGB, NIR and DEM layers as well as histogram equalisation of the RGB layer as an additional layer to generate a deeper representation of the image to train a CNN model on.

The AI training harness layer is responsible for retrieving input images prepared by the initial processing layer and separating this batch into verification and training images for cross-validation. Additionally, this layer defines the training regimen the user wishes to perform in terms of epochs, training intervals and batch sizes, as well as configure the structure of the model, such as number of layers or pooling size. This layer is closely linked to the final presentation layer.

The presentation layer is responsible for displaying various statistics regarding the current training regime being conducted, as well as saving historical data for future analysis. This layer also

allows a user to view the structure of the model at various granularities in order to fine-tune its configuration.

3.2 Implementation

Each layer is implemented in Python, to generate easy to read and maintain code, as well as Python being a popular and mature data science language.

3.2.1 Image transformation layer

The image transformation layer uses a combination of the cv2, numpy, skimage, and sklearn Python libraries, as well as some other supporting libraries such as pandas and PIL. We used the h5py library to extract the various layers from the h5 images. The user is expected to specify the path to the source directory of the images, a list of layers and transformations required, as well as the output directory to save these images. The user may also specify the output format, such as a png, gif, or numpy array. When run, the image driver will give information as well as regular updates on the progress of the processing. Once complete, the resulting output image is an array of up to 20 channels per pixel, and ready to be processed by the training harness.

The driver can extract the existing layers as mentioned earlier, as well as any combination of the following small scale/pixel level transformations:

- 2 variations of RGB to HSI
- RGB to L*a*b*
- Independent Component Analysis (ICA)
- Principal Component Analysis (PCA)

The driver can produce the following medium-scale transformations:

- Mean Shift - Performed on RGB values to smooth green pixel areas
- Histogram Equalization - Performed on NDVI values to strengthen borders between shadow and tree canopy on NDVI spectrum
- Canny Edge Detection - Performed on RGB values to extract edges from RGB pixel values
- Morphological Closing - on NDVI values to fill gaps in the tree canopy

3.2.2 Image transformation layer.

The training harness was built on the Python TensorFlow framework, in particular using a U-Net implementation by Joel Akeret et al. [22] developed for use in radio frequency interference mitigation. TensorFlow requires access to the Nvidia cuDNN libraries which can be complex to install, and so to ensure easy reproducibility this layer also includes a Docker image based on the Nvidia TensorFlow Docker image to simplify setup.

3.2.3 Image transformation layer

The presentation layer was handled automatically by the TensorFlow implementation, leveraging the functionality of Tensorboard in order to track and graph the progress of each training run, as well as allowing the user to view historical data from past runs. Tensorboard also generates a graph representing the structure of the AI model at a point in time, allowing for further insight into the performance of the model.

4 Segmentation Tests

4.1 Test Dataset

The dataset used for segmentation testing was provided by Aerobotics, containing 636 drone-captured images of an apricot orchard from various times and days. The dataset was not randomized for each model's training run to keep the results as equivalent as possible. From this dataset, 6 images were kept aside as a verification batch, shown in Appendix 1. These images were selected to represent the different classes of challenges the model may face while performing segmentation.

- Images 1 and 4 were selected as tests for continuous rows of trees, under different lighting conditions.
- Image 2 was selected to test the response to the mask Aerobotics applies to neighbouring properties, visible in the top left corner, as well as severe shadows.
- Images 3 and 5 were considered the most straightforward and easy segmentation tasks.
- Image 6 was selected to test accuracy around the green vegetation at the base of each tree.

4.2 Test Strategy

In order to measure the relative change in accuracy of the model when transformations are applied, a base case is needed. This base case is a model trained on only the RGB channels, as this is considered the most standard image format. Aerobotics also provides additional layers such as NIR, and so a model trained on RGB as well as these additional layers will be a secondary base case for comparison.

The foundation of an image is its colour channels, for the purpose of this research, this is limited to RGB, HSI and Lab formats. After investigating the effect these formats have on the accuracy of the model, a comparison will be made to discard the format with the worst accuracy statistics.

Taking these two foundations, various transformations will be added to them as additional channels to measure their interaction with the model. Due to time constraints, the initial two tests done per foundation as mentioned above are all small-scale transforms as additional layers, and all medium-scale transforms as additional layers.

A comparison will then be made between the performance of the groups of transforms and either small or medium scale transformations as a group may be discarded at this stage depending on their impact, or lack thereof, on accuracy. This is due to time constraints on the experimentation phase and justified

by the ability of a CNN to effectively discard channels that are not suitable. If there is an effective transform in the collection, the CNN will identify it and make use of it. If the collection as a whole does not impact accuracy positively, it can be safely discarded.

Finally, in order to isolate effective transformations, the transformation groups with an increase in accuracy will be split for test runs, for example running RGB with Edge Detection, as well as Lab with Edge Detection. At this stage, we will be able to make recommendations for the accuracy impact of each transformation.

The U-Net implementation allows for various configurations of both the structure of the U-Net model and its training regime. The model used to produce the results in this paper consisted of 5 active layers, 64 feature roots, a filter size of 3x3 and pool size of 2x2. The cost function used was cross-entropy, with no explicit class weights and a learning rate starting at 0.2. The data provided to it varied from 3 to 20 channels, and always 2 classes, tree or not-tree. The training regime for each run was done over 300 epochs with 25 training iterations, with a batch size of 3 images and a verification set of 6 images. All runs were conducted on an RTX2070 with 8Gb of VRAM. Runs take between 2 to 5 hours each, depending on the number of channels.

4.3 Performance Metrics

The key performance indicator for each run will be accuracy and will be supported by the loss and cross-entropy. Accuracy is measured as a percentage with 100% being perfectly correct segmentation and calculated as an average across the final training epoch. Cross entropy is a log loss function used for calculating loss on models that present a probability of classification, with 0 cross-entropy being a perfect classification, and increasing exponentially as the prediction tends to a complete incorrect classification. The aim of these transformations is to increase the accuracy, and reduce loss and cross-entropy when compared to the model trained on the original RGB image base case. Manual examination of the verification batch will also be made to analyze the effect the transforms have on each of the challenging test classes, such as shadow detection.

5 Results and Discussion

The purpose of these transformations is to improve the accuracy of the final trained model, as well as improve the confidence of the segmentation results produced. In the below tables, 'Aerobotics' layers are all the layers mentioned earlier provided with the base RGB image from Aerobotics, such as DEM and NDVI. Accuracy, cross-entropy and loss readings were taken from Tensorboard at the end of each training session.

5.1 Impacts on Metrics

5.1.1 Control Models

Image Base	Additional Layers	Accuracy	Cross-Entropy	Loss
RGB	-	88.64%	0.14	0.27
RGB	Aerobotics	91.63%	0.12	0.24

Table 1 – Control models

The additional layers provided by Aerobotics increased the accuracy when combined with the RGB image by 2.99%. These 2 models serve as the control models to compare all following experiments to. Each of the transformations applied is designed to increase accuracy, and reduce cross-entropy and loss when compared to these control results.

5.1.2 Small scale transforms

Image Base	Additional Layers	Accuracy	Cross-Entropy	Loss
HSI	-	73.90%	23	0.45
LAB	-	86.97%	0.15	0.3

Table 2.1 – Colour space transformation results

RGB and Lab performed at similar accuracy levels, with RGB being marginally more accurate by 1.67% on the verification batch as well as slightly lower cross-entropy and loss. HSI had a significant negative effect on accuracy of 14.74%, and so can be discarded for future test combinations. HSI's loss of accuracy may be attributed to the transformation losing accuracy on floating-point values of the Hue channel, or simply not having enough contrast between areas important for segmentation.

Image Base	Additional Layers	Accuracy	Cross-Entropy	Loss
RGB	Aerobotics	91.63%	0.12	0.24
LAB	Aerobotics	89.98%	0.14	0.29

Table 2.2 – Lab colour space with Aerobotics layers, compared to control case

The additional layers provided by Aerobotics increased segmentation accuracy when paired with the Lab base image by 3.01%. These layers give the model more context around a tree to make predictions, such as height, or NDVI values and so are expected to increase accuracy. The baseline accuracy of the RGB image and Aerobotics layers is still the most accurate, with 1.65% greater accuracy than the Lab image with the same additional layers, with lower cross-entropy and loss implying greater confidence in the segmentation effort.

Image Base	Additional Layers	Accuracy	Cross-Entropy	Loss
RGB	ICA and PCA	87.69%	0.15	0.29
LAB	ICA and PCA	70.51%	0.25	0.50

Table 2.3 – RGB and Lab image with ICA and PCA

The addition of the small-scale transformations, namely PCA and ICA, reduced accuracy of both the RGB and Lab base images. This was a small loss in the case of the RGB image, only reducing accuracy by 0.95%, however, the U-Net model failed on the Lab image with the small-scale transformation layers applied. Accuracy was reduced by 16.46%, and close to doubling loss and cross-entropy. This may be due to the PCA and ICA still being performed on the RGB values before being added to the Lab image, resulting in the model not understanding the relationship between the underlying Lab image, additional layers and segmentation mask. Due to this and loss of accuracy on the RGB image, small scale transformations other than the Lab base image will not be investigated further.

5.1.3 Medium-scale transformations

Image Base	Additional Layers	Accuracy	Cross-Entropy	Loss
RGB	All Medium Scale Transforms	89.22%	0.12	0.24
LAB	All Medium Scale Transforms	88.14%	0.13	0.27

Table 3.1 – RGB and Lab with all medium scale transformation layers

Both experiments with all of the medium-scale layers increased accuracy, with a 0.58% increase in RGB accuracy, and lower loss and cross-entropy. The medium-scale transformation layers increased the accuracy of predictions made on the Lab image by 1.17% and lowered cross-entropy loss. As the medium-scale transformations had a positive impact on accuracy, we will investigate which specific transformations had the most positive interaction on accuracy with the U-Net model.

Image Base	Additional Layers	Accuracy	Cross-Entropy	Loss
RGB	All Medium scale + Aerobotics	90.80%	0.10	0.21
LAB	All Medium scale + Aerobotics	90.72%	0.11	0.22

Table 3.2 – RGB and Lab images with all medium scale and Aerobotics layers

Adding the medium-scale transformations to the base image and layers provided by Aerobotics, there was a minor negative impact on accuracy of less than 1% for the RGB image, and a less than 1% increase in accuracy on the Lab image, however it resulted in a fairly significant reduction in cross-entropy and loss when compared to the control case metrics, implying the model was more confident about the predictions it made, however, they were not more accurate. This may be a limitation of the ground truth masks supplied by Aerobotics, as they are not always completely accurate. These results show that although the additional layers did not improve accuracy, having additional channels to segment on made the model more confident in these segmentations.

Image Base	Additional Layers	Accuracy	Cross-Entropy	Loss
RGB	Morphological Closing	89.37%	0.12	0.24
LAB	Morphological Closing	88.10%	0.13	0.26

Table 3.3 – RGB and Lab with Morphological Closing

Morphological closing had a positive impact on accuracy, increasing it by 0.73% on the RGB base image, and 1.13% on the Lab base image. It also slightly decreased cross-entropy and loss, implying the transform did not improve confidence, only accuracy. This can be expected as the purpose of morphological closing is to remove small blemishes in tree canopies, presenting a more contiguous surface to the U-Net for segmentation.

Image Base	Additional Layers	Accuracy	Cross-Entropy	Loss
RGB	Mean Shift	86.50%	0.16	0.33
LAB	Mean Shift	74.12%	0.24	0.49

Table 3.4 – RGB and Lab with Mean shift layer

The addition of the mean shift layer to the RGB and Lab base images resulted in a reduction in accuracy, and a gain in cross-

entropy and loss. The RGB base image with mean shift was not impacted as heavily as the Lab image, which had an accuracy reduction of 12.85%, as well as significant gains in cross-entropy and loss.

Image Base	Additional Layers	Accuracy	Cross-Entropy	Loss
RGB	Histogram Equalization	89.78%	0.14	0.27
LAB	Histogram Equalization	82.04%	0.23	0.45

Table 3.5 – RGB and Lab with histogram equalization layer

Adding the histogram equalization result to each of the base images marginally increased the accuracy of the RGB image, and significantly lowered the accuracy on the Lab image. The combination of the histogram equalization layer and the Lab image hindered the ability of the U-Net model to perform reliable segmentation.

Image Base	Additional Layers	Accuracy	Cross-Entropy	Loss
RGB	Edge Detector	78.34%	0.20	0.41
LAB	Edge Detector	79.10%	0.22	0.43

Table 3.6 – RGB and Lab with edge detector layer

Combining the result of the edge detector with each of the base images resulted in significantly lower segmentation accuracy. The features extracted by the edge detector confused the U-Net model and did not allow for it to construct its own more accurate understanding of edges.

Image Base	Additional Layers	Accuracy	Cross-Entropy	Loss
RGB	-	88.64%	0.14	0.27
RGB	Morphological Closing and Histogram Equalization	90.31%	0.13	0.27

Table 3.7 – RGB control result, and RGB with Histogram Equalization and Morphological Closing layers

The addition of Morphological Closing and Histogram Equalization layers to the base RGB image increased segmentation accuracy by 1.67%, with negligible changes in loss and cross-entropy. This combination of transformations therefore successfully improved the segmentation performance of the U-Net model

Image Base	Additional Layers	Accuracy	Cross-Entropy	Loss
RGB	Aerobotics	91.63%	0.12	0.24
RGB	Aerobotics, Morphological Closing and Histogram Equalization	92.91%	0.09	0.19

Table 3.8 – RGB with Aerobotics layer control result, and RGB, Aerobotics, Histogram Equalization and Morphological Closing result

The addition of Morphological Closing and Histogram Equalization layers to the base RGB image with Aerobotics layers increased segmentation accuracy by 1.28% from the control image with only RGB and Aerobotics layers, as well as lowering cross-entropy by 0.03 and loss by 0.05. This is the highest accuracy model, with the lowest cross-entropy and loss. The final verification batch for this model can be found in the Supplementary Information appendix 2.

5.1.4 Accuracy Findings

These results provide evidence that manipulating the colour space of an image does not improve segmentation accuracy of the U-Net model in the case of drone-captured tree images. PCA and ICA do not improve performance metrics either. The U-Net model can learn the most effective small scale or pixel-level transformations without our assistance, and so adding these extra channels only added noise to the model's input data, and so reduced accuracy and increased cross-entropy and loss.

Medium-scale transformations were able to extract features that marginally aided the U-Net in segmentation. Specifically, the results show that morphological closing increased accuracy and decreased cross-entropy and loss on both variants of the base image. The histogram equalization transformation also increased the accuracy of the model when paired with the RGB base image, however, decreased accuracy significantly when paired with the Lab base image.

The model with the best performance metric results was the RGB base image along with Aerobotics layers, Morphological closing, and histogram equalization. This resulted in 92.91% final epoch accuracy, successfully improving upon both control models.

5.2 Manual analysis of verification batch

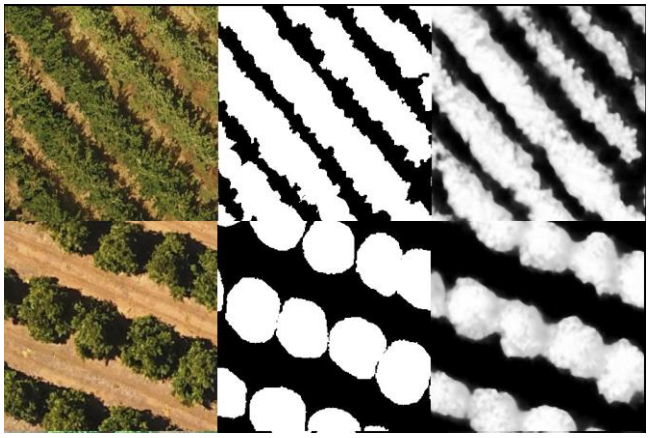
The verification batch was selected to test a wide variety of segmentation challenges the model will encounter. Manual verification is focused on how the model segments rows of trees as well as individual trees, and exclusion of ground vegetation and shadows from tree segmentation. Verification image 2 also has a farm border that is masked in the image, the model should not identify any trees in that area. In the figures below, the leftmost image is the base RGB or Lab image, the middle is the segmentation mask provided by Aerobotics, and the rightmost image shows the segmentation performed by the U-Net model on

the 300th training epoch. The closer the colour is to white, the more confident the model is segmenting that pixel as "tree".

5.2.1 Control model results

Figure 3. RGB verification results

With only the RGB channels, the model was able to reliably segment most well-defined tree canopy borders for both rows of trees and individual trees. However, it did not reliably exclude shadows and ground vegetation shown in figure 3. The model was able to exclude most of the shadow along the border of the final image shown in figure 3, however still identified portions of it as a tree which is incorrect.



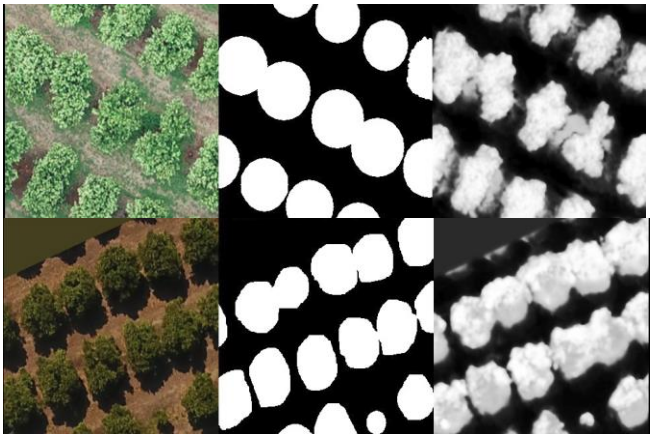


Figure 4. RGB with Aerobotics layers (Left) with ground truth (middle) and model segmentation (right)

The addition of layers provided by Aerobotics such as NIR and DEM gave the model far more context to use in its segmentation task, shown in figure 4. These layers allowed the model to define clearer borders between the tree and the ground for both rows of trees and individual trees, as well as reduce the amount of shadow and ground vegetation incorrectly segmented as tree. The addition of these layers also reduced the amount of incorrect segmentation at the border of the final image.

5.2.2 Highest accuracy transformations

Although the two control cases had high accuracy, some medium scale transforms had positive effects on accuracy and served to reduce cross-entropy and loss. No medium-scale transform is perfect on its own, but rather serve to provide different aids reflected in figures 5 and 6.

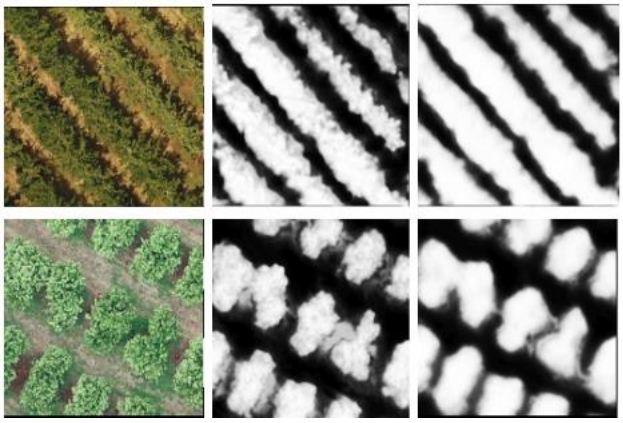


Figure 5. RGB with Aerobotics (middle) and Medium-scale transformations & Aerobotics layers (right)

Figure 5 shows the results of combining all the medium-scale transformations with the RGB image and Aerobotics layers (right), compared to the original RGB with Aerobotics results (middle). This model displayed equivalent accuracy results, but lower cross-entropy and loss. This can be confirmed visually by analysing the solid white colouring of these segmentation

attempts, showing the model was far more confident when segmenting the centre of each tree canopy. The results also show more confidence when discarding shadow and vegetation, however, some reaction to it is still present.

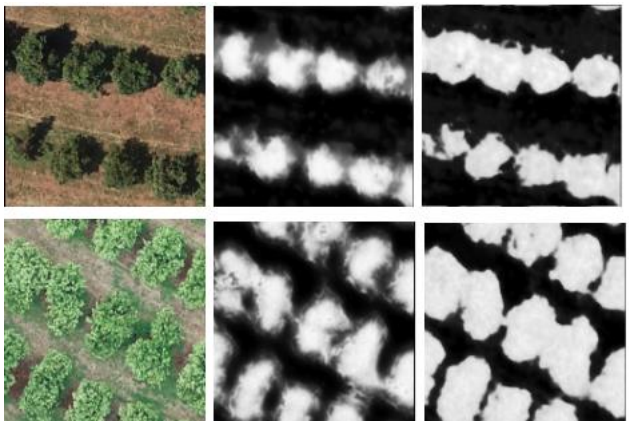


Figure 6. RGB with Morphological Closing (middle) and Histogram Equalization (right) results

Morphological closing and Histogram equalization resulted in the highest accuracy when paired with the RGB base image. These transformations are based on the NDVI channel, and so provide the U-Net model with a modified spectrum to perform segmentation with. Morphological closing is intended to fill gaps in the canopy resulting in the high confidence seen in figure 6 and the centre of each canopy. It is also an effective transformation for reducing the false positives around shadows and ground vegetation. Histogram equalization served to create clean borders around the edges of each tree, at the cost of often including ground vegetation within these borders.

5.2.2 Lowest accuracy transformations

Some transformations significantly reduced the ability of the U-Net model to perform segmentation tasks, shown in figures 7 and 8

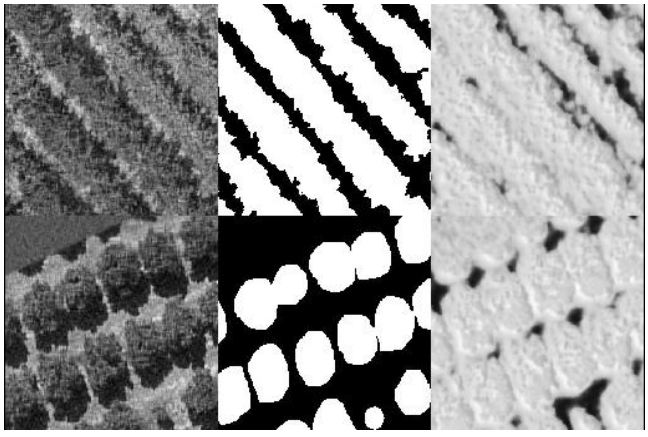


Figure 7. HSI with ground truth (middle) and verification results (right)

The U-Net model was hampered by the HSI format, classifying the entire border region as a tree in the second image of figure 7, as well as failing to identify the gaps between rows of trees in both images. Shadows and ground vegetation were segmented with the same confidence as the tree canopy, leading to this transformation being discarded early in the experimental process.

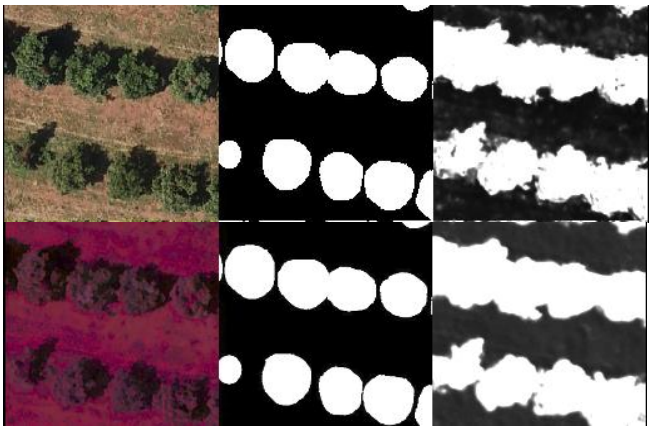


Figure 8. RGB and Lab image with ground truth (middle) and Canny Edge Detector (right)

The addition of the edge detector to each set of images shown in figure 8 resulted in higher confidence when segmenting the tree canopy, as the edges were already highlighted for the U-Net model. However, due to this, the U-Net was unable to discard shaded areas as they were contained within the edges, leading to a nearly 10% loss in accuracy when compared to the base images. The model was also unable to distinguish individual trees when the edge detector layer was added, further reducing accuracy. From these results it is recommended to allow the U-Net model to create its own smaller-scale rules and transformations, rather than adding unneeded noise and focusing efforts on extracting larger-scale features the model finds difficult to understand.

5.3 Conclusion

The U-Net CNN model is well suited to the task of segmenting tree canopies from the underlying landscape, reaching industry-standard accuracy results on a training set of only 600 images. Although the control models performed with acceptable accuracy, the addition of Morphological Closing and Histogram Equalization layers served to reduce cross-entropy and therefore improve the confidence of the model when performing segmentation. The ideal model from these experiments was that trained on an RGB base, with layers supplied by Aerobotics and Morphological Closing, and Histogram Equalization layers. The transformations found to have the most significant positive impacts on performance metrics when compared to both control models were Histogram Equalization and Morphological closing. Small scale transformations such as colour space manipulation, as well as ICA and PCA, had negative effects on the key success metrics of the model and should not be used to improve segmentation.

6 Future Work

This research was performed as an initial investigation into the effects of various image transformation techniques on the U-Net model, using a limited dataset and training hardware. Future research may investigate the effects these transformations have on a larger production scale dataset, as well as testing on live segmentation tasks in various contexts. This research also identified transformations that were not successful in improving the segmentation ability of the model, and so future research may investigate a new set of transformations. As this was an initial investigation, hardware resources were limited and so investigation into the effects these transformations have on larger models with more layers and feature neurons should be performed.

Acknowledgements

This research paper was part of an extended research group evaluating the effects of these image transformations on various CNN architectures, postgraduate student researchers Michael Scott and Charl Ritter assisted in the development of the image transformation layer and analysed the effects on Atrous Neural Networks and Fully Convolutional Neural Networks respectively in their own papers yet to be published. This research would not have been possible without the assistance and guidance of Aerobotics South Africa, providing both the sample data set as well as potential research areas. Professor Patrick Marais and Professor Deshen Moodley supervised the extended research group, providing valuable insight and recommendations.

REFERENCES

[1] Marc Antonini, Pierre Mathieu, and Ingrid Daubechies. 1992. Image Coding Using Wavelet Transform. *IEEE Transactions on Image Processing* 1, 2 (April 1992), 205–220.

[2] Min Bai and Raquel Urtasun. 2017. Deep Watershed Transform for Instance Segmentation. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), 5221–5229. DOI:http://dx.doi.org/10.1109/cvpr.2017.305

[3] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. 2017. Rethinking Atrous Convolution for Semantic Image Segmentation. *CoRR* abs/1706.05587 (2017).

[4] V. Grau, A.u.j. Mewes, M. Alcaniz, R. Kikinis, and S.k. Warfield. 2004. Improved Watershed Transform for Medical Image Segmentation Using Prior Information. *IEEE Transactions on Medical Imaging* 23, 4 (April 2004), 447–458. DOI:http://dx.doi.org/10.1109/tmi.2004.824224

[5] Dianyuan Han and Xinyuan Huang. 2010. A Tree Image Segmentation Method Based on 2-D OTSU in HSI Colour Space. *2010 International Conference on Computational Intelligence and Software Engineering* (2010). DOI:http://dx.doi.org/10.1109/wicom.2010.5600669

[6] J.L. Hernández-Hernández, G. García-Mateos, J.M. González-Esquiva, D. Escarabajal-Henarejos, A. Ruiz-Canales, and J.M. Molina-Martínez. 2016. Optimal colour space selection method for plant/soil segmentation in agriculture. *Computers and Electronics in Agriculture* 122 (2016), 124–132. DOI:http://dx.doi.org/10.1016/j.compag.2016.01.020

[7] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015), 3431–3440. DOI:http://dx.doi.org/10.1109/cvpr.2015.7298965

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* 9351 (2015), 234–241. DOI:http://dx.doi.org/10.1007/978-3-319-24574-4_28

[9] G. Ruiz-Ruiz, J. Gómez-Gil, and L.M. Navas-Gracia. 2009. Testing different colour spaces based on hue for the environmentally adaptive segmentation algorithm (EASA). *Computers and Electronics in Agriculture* 68, 1 (2009), 88–96. DOI:http://dx.doi.org/10.1016/j.compag.2009.04.009

[10] R.N. Strickland and Hee Il Hahn. 1996. Wavelet transforms for detecting microcalcifications in mammograms. *IEEE Transactions on Medical Imaging* 15, 2 (April 1996), 218–229. DOI:http://dx.doi.org/10.1109/42.491423

[11] L. Tang, L. Tian, and B.L. Steward. 2000. Colour Image Segmentation With Genetic Algorithm For In-Field Weed Sensing. Transactions of the ASAE 43, 4 (2000), 1019–1027. DOI:http://dx.doi.org/10.13031/2013.2970

[12] Xiao-Song Wang, Xin-Yuan Huang, and Hui Fu. 2009. The Study of Colour Tree Image Segmentation. *2009 Second International Workshop on Computer Science and Engineering* (2009), 303–307. DOI:http://dx.doi.org/10.1109/wcse.2009.818

[13] Xiaosong Wang, Xinyuan Huang, and Hui Fu. 2010. A Colour-Texture Segmentation Method to Extract Tree Image in Complex Scene. *2010 International Conference on Machine Vision and Human-machine Interface* (2010). DOI:http://dx.doi.org/10.1109/mvhi.2010.138

[14] Liying Zheng, Jingtao Zhang, and Qianyu Wang. 2009. Mean-shift-based colour segmentation of images containing green vegetation. *Computers and Electronics in Agriculture* 65, 1 (2009), 93–98. DOI:http://dx.doi.org/10.1016/j.compag.2008.08.002

[15] Adimara Bentivoglio Colturato et al. 2013. Pattern Recognition in Thermal Images of Plants Pine Using Artificial Neural Networks. *Engineering Applications of Neural Networks Communications in Computer and Information Science* (September 2013), 406–413. DOI:http://dx.doi.org/10.1007/978-3-642-41013-0_42

[16] M. Mirzaie, R. Darvishzadeh, A. Shakiba, A.a. Matkan, C. Atzberger, and A. Skidmore. 2014. Comparative analysis of different uni- and multi-variate methods for estimation of vegetation water content using hyper-spectral measurements. International Journal of Applied Earth Observation and Geoinformation 26 (February 2014), 1–11. DOI:http://dx.doi.org/10.1016/j.jag.2013.04.004

[17] M. Lennon, G. Mercier, M.c. Mouchot, and L. Hubert-Moy. 2001. Spectral unmixing of hyperspectral images with the independent component analysis and wavelet packets. IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217) 6 (July 2001), 2896–2898. DOI:http://dx.doi.org/10.1109/igarss.2001.978198

[18] Liying Zheng, Jingtao Zhang, and Qianyu Wang. 2009. Mean-shift-based colour segmentation of images containing green vegetation. Computers and Electronics in Agriculture 65, 1 (2009), 93–98. DOI:http://dx.doi.org/10.1016/j.compag.2008.08.002

[19] Le Yu, Alok Porwal, Eun-Jung Holden, and Michael Charles Dentith. 2011. Suppression of vegetation in multispectral remote sensing images. International Journal of Remote Sensing 32, 22 (September 2011), 7343–7357. DOI:http://dx.doi.org/10.1080/01431161.2010.523726

[20] J.a. Benediktsson, M. Pesaresi, and K. Arnason. 2003. Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. IEEE Transactions on Geoscience and Remote Sensing 41, 9 (2003), 1940–1949. DOI:http://dx.doi.org/10.1109/tgrs.2003.814625

[21] Qingsheng Liu, Jinfa Dong, Gaohuan Liu, Chong Huang, and Chuanjie Xie. 2011. Using the Canny edge detector and mathematical morphology operators to detect vegetation patches. Third International Conference on Digital Image Processing (ICDIP 2011) (July 2011). DOI:http://dx.doi.org/10.1117/12.896163

[22] J. Akeret, C. Chang, A. Lucchi, and A. Refregier. 2017. Radio frequency interference mitigation using deep convolutional neural networks. Astronomy and Computing 18 (2017), 35–39. DOI:http://dx.doi.org/10.1016/j.ascom.2017.01.002

Supplementary Information

Appendix 1: Verification Batch

Appendix 2: Highest Accuracy model results (RGB, Aerobotics, Morphological Closing, Histogram Equalization)