# Flower Segmenentation using Neural Networks

School of Computer Science, University of Nottingham, UK

Mohid Ali Gill
University of Nottingham
psxmg8@nottingham.ac.uk

***Abstract*—Semantic segmentation is a pivotal task in computer vision, classifying each pixel into a specific pre defined class. This research focuses on segmenting flower images from the Oxford Flower Dataset using the Convolutional Neural Networks (CNNs), specifically targeting binary classification of flower and background pixels. The study uses a well established U-Net architecture and a custom CNN model to perform the segmentation tasks. The dataset, comprising of 1360 high-resoulution images resized to 256x256 pixels is filtered to to include only images containing their respective labels. Data augmentation techniques are applied to make the models robust. The U-Net model provides better results with an overall mean IoU of 0.7679, a mean dice coefficient of 0.8574 and a global accuracy of 0.8937. In contrast the custom CNN model achieves the mean IoU of 0.7355, dice coefficient of 0.8344 and global accuracy of 0.8722. The results highlight the importance of advanced architectures like U-Net which have skip connections in acheiving high segmentation accuracy.**

## I. Introduction

One of the pivotal task in computer vision is semantic image segmentation which involves categorising each pixel in the image to a specific predefined class. This type of classification has numerous real-world applications such as autonomous driving, medical imaging and more. Segmenting accurately does not only help in object recognition but also helps in finding relationships between different objects in the image.

This research aims to develop and evaluate convolutional neural networks (CNNs) to segment flowers from the background in images, primarily classifying each pixel into one of the two categories: flower and background. A challenge this binary classification faces is to distinguish the flower from other visually similarly objects such as leaves and the sky.

A well known benchmark in the field of image segmentation and classification is the Oxford Flower Dataset which is also used in this study. A portion of dataset consisting of 1360 high resolution images resized to 256 x 256 pixels is used to perform the study. Only 846 out of these 1350 images contain there corresponding labels annotating the region of flowers. The segmentation labels are provided as color maps with pixel values indicating different classes: 0 for null and boundaries, 1 for flowers, and 3 for the background. These are the ones that are used in the study other than these labels for sky and leaves are also provided but will be ignored.

The study involves two main approaches to segment the images, firstly, an existing CNN model is used to segment the images and another custom model is developed from scratch. Both the models are then evaluated and the results are compared.

## II. Literature Review

Semantic image segmentation is a critical task in computer vision and it has been revolutionised by the Convolutions Neural Networks (CNNs) which delivered good results in capturing the spatial hierarchies and contextual information in the images.

One of the pioneering work in this field is by Long et al. (2015), who introduced the Fully Convolutual Networks (FCNs) for semantic segmentation which eliminated the need for the fully connected layers and produced consistent results for inputs of arbitrary sizes. FCNs resulted to a significant game changer for the segmentation tasks by enabling end to end learning.

Ronneberger et al. (2015) developed the U-Net architecture which played a major role in biomedical image segmentation. The U-Net architecture includes an encoder-decoder structure with symmetric skip connections. The U-net architecture is really good at preserving the spatial resolution during the upsampling process making it highly precise and effective for the segmentation tasks even with limited training data.

In the domain of flower segmentation, Nilsback and Zisserman (2006) developed a method using different computer vision techniques and traditional machine learning model to classify the flowers. Although, their research was before the deep learning era but still provided a strong foundation for the future research leveraging CNNs for the improved segmentation accuracy.

Milito et al. (2018) explained and demonstrated the practical applications of CNNs in the agriculture by using the U-Net model for real time segmentation of crops and plants. The work assisted in understanding how adaptable and robust CNN is in dynamic and challenging outdoor environments.

## III. Methodology

A detailed description of the process followed to develop the convolutional neural networks (CNN) for this task is in this section. The process involved several stages, starting from class selection and preparing the data to training both existing and a custom CNN model. Each stage played an important role in effectively segmenting the images.

### A. Class Selection

The Oxford dataset used in this study has images with pixel-level annotations categorised into 5 classes: 0 for null/boundaries, 1 for flower, 2 for leaves, 3 for background and 4 for the sky. TO make it simpler in this study binary classification is used to segment flower from all the other labels, primarily focusing on 2 classes the flower (1) and the background (3), all the rest of classes are disregarded and treated as a background hence is is made sure that trained to segment flower from everything else.

### B. Filtering Data without labels

The dataset consists of 1360 images, each of them is resized to 256 x 256 pixels for simplicity but only 846 of these images have their corresponding segmentation labels. To train, test and validate data only the images containing the labels should be used hence a filtering process is implemented which matches the images with the corresponding label files and moving the files to another directory for which the labels are not provided. The resulting dataset are all the paired images and their labels, which are important for supervised learning. The filtering script can be found in **filtering_unlabelled.m**

### C. Splitting Data into Training, Testing and Validation Sets

Once the filtering is completed, the data is split into three subsets: training, testing and validation. To ensure that the data is distributed balanced between all three sets the data is assigned randomly to the sets. 70% of the data is used for training, the remaining 30% is divided equally among testing and validation sets. This 70/15/15 split resulted in 593 images for training, 127 images for testing and remaining 127 images for validation. Randomly distributing the images removed any chance of bias for selection.

### D. Data Augmentation

To make the segmentation model more robust data augmentation techniques are used on the training data. 3 different techniques are used which include random horizontal flips in which the images are flipped horizontally with a 50% probability. Random Translations are performed which translate the images up to 10 pixels in both x and y directions and lastly random rotations are performed rotating the images within a range of -10 to +10 degrees. These techniques increases the variability in the training data making the model more resilient to various types of input images.

### E. Training a U-Net model

To segment the flower images a U-Net model is chosen as it has proven effective in segmentation tasks, particularly with agriculture images. The U-Net model consists of contracting paths which are encoders and an expanding path which are decoders with skip connections that merge the feature maps from the encoders with the upsampled outputs from the decoder.

The U-Net model is initialised with an input size of 256x256x3 and output size corresponding to the two classes: flower and background. The training parameters used to train the model are selected carefully. The Adam (Adaptive Movement Estimation) optimiser is used for training. It is a very popular optimisation algorithm as it combines the best parts of both AdaGrad and RMSProp and lastly it adjusts the learning rate for every parameter dynamically which results in improving the convergence speed and performance. The initial learning rate is set to 0.0001. The learning rate decides how much to change the model by the resulted error each time the weights are adjusted. A lower learning rate is better in precise convergence but a higher learning rate would be much faster but can lead to overshooting the optimal solution. The model is trained for maximum 20 epochs with a mini batch size of 8. Mini batch size is the number of samples that will go through the network at once.

Data augmentation techniques are implemented and then passed to U-Net model with their respective labels. The model's performance is validated periodically by using the validation set and the resulting model is saved in a .mat file.

### F. Training a custom model

To compare with an existing U-Net model a custom CNN model is trained from scratch. The custom built model is designed to be simple yet effective for this segmentation task. The custom model contains series of downsampling and upsampling layers with ReLU activations and max-pooling layers which are designed to capture and reconstruct the spatial hierarchies.

The downsampling part of the network is made up of convolutional layer with filter size of 3x3 followed by a ReLU activation and a max-pooling layer off size 2x2 and stride 2. The downsampling was performed 4 times before the upsampling path was taken. Upsampling reconstructs the spatial dimensions of the feature maps using transposed convolutional layers, it reduces the depth of features and increase the spatial dimensions with skip connections to retain high-resolution features. The path consists of 4 blocks of transposed covolutional layes of size 4x4 and stride 2 followed by ReLU activation. A softmax layer is added to convert the logits to their probabilities and finally a classification layer to classify each pixel into one of the defined classes (flower or background).

For training parameters the optimiser used is SGDM (Stochastic Gradient Descent with Momentum) optimiser is used. SGDM helps accelerate gradients vectors in the right directions, thus leading to a faster convergence. The learning rate is set to 0.01, the model is trained for 25 epochs with a mini batch size of 16. The resulted model is saved in a .mat file for evaluation process. Fig 1 displays the architecture of custom trained models.

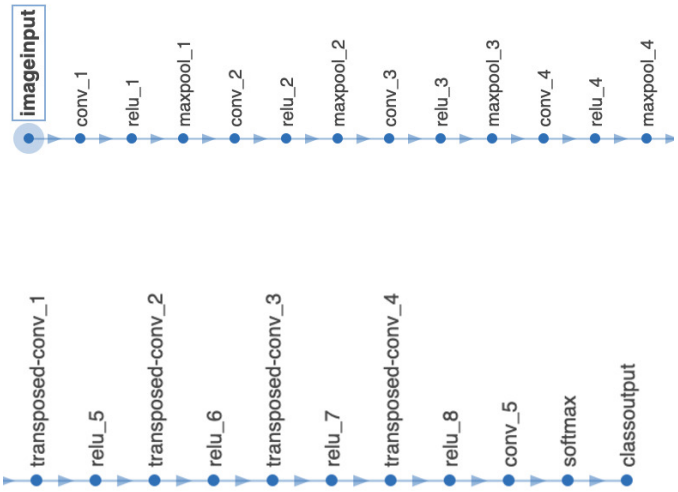


Fig. 1: Custom model architecture

### G. Evaluation Metrics

To evaluate the performance of the trained segmentation models, several evaluation metrics are used. These metrics provides both quantitative and qualitative insights for how good the model performs in segmenting the images. Global accuracy is measured the overall pixel-wise accuracy of the model. Mean Intersection over Union (IoU) also known as the Jaccard index, measures the overlap between the predicted segmentation and the ground truth. A number of true positives, true negative, false positive and falase negatives for each class are calculated.

## IV. Results and Evaluation

This section presents the results of the semantic segmentation models trained on the Oxford Flower Dataset and evaluates their performance using the metrics described in the methodology section. The performance of both the U-Net model and the custom CNN model is evaluated in this section.

### A. Quantitative Evaluation

The quantitative evaluation of the segmentation is based on the following metrics: Mean IoU, Mean Dice Coefficient, Precision, Recall, Global Accuracy and confusion matrix. These metrics provide a comprehensive assessment of the models' ability to accurately segment the flowers from the background.

*1) U-Net model performance:* The U-Net model achieved the following performance on the test dataset:

TABLE I: U-Net Model Performance Metrics

| Class | Mean IoU | Mean Dice | Precision | Recall |
|---|---|---|---|---|
| Flower | 0.68119 | 0.79711 | 0.70864 | 0.97422 |
| Background | 0.85459 | 0.91767 | 0.90429 | 0.94187 |

TABLE II: U-Net Model Confusion Matrix Elements

| Class | Total TP | Total FP | Total FN | Total TN |
|---|---|---|---|---|
| Flower | 2,292,500 | 942,590 | 60,662 | 5,092,800 |
| Background | 4,660,300 | 493,230 | 287,600 | 2,947,500 |

TABLE III: Overall Performance Metrics for U-Net Model

| Metric | Value |
|---|---|
| Overall Mean IoU | 0.7679 |
| Overall Mean Dice | 0.8574 |
| Global Accuracy | 0.8937 |

These results show that the U-Net model has high accuracy and segments both the flowers and background classes effectively. The high recall for the flower class suggests that the model is particularly good at identifying the flower pixels.

*2) Custom CNN Model Performance:* The custom CNN model acheived the following performance on the test dataset:

TABLE IV: Custom CNN Model Performance Metrics

| Class | Mean IoU | Mean Dice | Precision | Recall |
|---|---|---|---|---|
| Flower | 0.68632 | 0.79444 | 0.79321 | 0.85823 |
| Background | 0.78475 | 0.87434 | 0.81358 | 0.96069 |

TABLE V: Custom CNN Model Confusion Matrix Elements

| Class | Total TP | Total FP | Total FN | Total TN |
|---|---|---|---|---|
| Flower | 2,019,600 | 526,520 | 333,620 | 5,508,900 |
| Background | 4,753,400 | 1,089,100 | 194,480 | 2,351,600 |

TABLE VI: Overall Performance Metrics for Custom CNN Model

| Metric | Value |
|---|---|
| Overall Mean IoU | 0.7355 |
| Overall Mean Dice | 0.8344 |
| Global Accuracy | 0.8722 |

These results suggest that the custom CNN model performs well but falls short of the U-Net model in terms of segmentation accuracy and boundary precision.

### B. Qualitative Evaluation

In addition to the quantitative metrics, the segmentation results are quantitatively evaluated through visual inspection of the predicted segmentation masks. Sample segmentation results for both the U-Net and custom model are shown below

*1) U-Net Model Segmentation Results:* The U-Net model shows a high accuracy in segmenting the flower regions, effectively capturing the fine details and the boundaries. The skip connections in the U-Net help preserve the high-resolution features leading to more precise segmentation. There were a few images that had a few errors Fig 2 and 3 displays a good and a poor segmentation results.

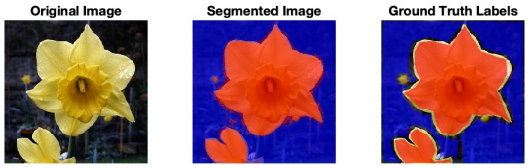


Fig. 2: U-Net result with good segmentation

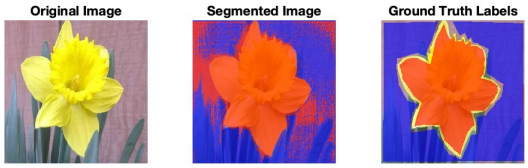


Fig. 3: U-Net result with poor segmentation

*2) Custom CNN Model Segmentation Results:* The custom CNN model also demonstrates good segmentation performance, though it falls short with finer details compared to the U-Net model. Sample results for a good and a poor segmentation for the custom model can be seen in Fig 4 and 5.

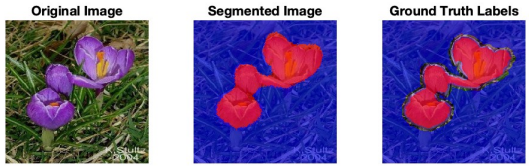


Fig. 4: Custom model result with good segmentation


Fig. 5: Custom model result with poor segmentation

## V. Conclusion and Future Work

The objective of this study was to segment the flower from the background. The U-Net model has better results across all the quantitative metrics achieving overall Mean IoU 0.7679, Dice coefficient 0.8574 and global accuracy 0.8937 whereas the custom model has mean IoU 0.7355, mean dice coefficient 0.8344 and global accuracy of 0.8722. The U-Net model effectively captured fine details and complex boundaries of the flower regions when the custom model struggled a bit with finer details. Advanced architectures like U-Net which incorporate skip connections and multi-scale feature integration are better suited for tasks that require precise boundary delineation and handling of fine details. In future the custom model can be enhanced such as skip connections can be added or layers can be added to capture more fine details. Additionally, the size and diversity of training data can be improved by adding more data which will help in both the models. Finally, the model can be tested for real time and can be improved.

## REFERENCES


[1] Long, J., Shelhamer, E. and Darrell, T., 2015. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431-3440.

[2] Ronneberger, O., Fischer, P. and Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, Cham, pp. 234-241.

[3] Nilsback, M.-E. and Zisserman, A., 2006. A Visual Vocabulary for Flower Classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1447-1454.

[4] Milioto, A., Lottes, P. and Stachniss, C., 2018. Real-time Semantic Segmentation of Crop and Weed for Precision Agriculture Robots Leveraging Background Knowledge in CNNs. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), pp. 2229-2235.