FROM PETAL TO PIXELS: OPTIMIZING CONVOLUTIONAL NEURAL NETWORKS FOR FLOWER SEGMENTATION TASKS

Maithil Ishvarlal Tandel
MSc. Machine Learning in Sciences (Pursuing)
University Of Nottingham
Email Id: ppxmt4@nottingham.ac.uk

ABSTRACT

In the interdisciplinary fields of computer vision and botany, precise segmentation of floral images plays a crucial role in applications such as ecological monitoring and precision agriculture. This research delves into the application of convolutional neural networks, particularly DeepLabv3+ with a ResNet-18 backbone and a custom model, for precise floral image segmentation critical to ecological monitoring and precision agriculture. Leveraging the Oxford Flower Dataset, the study advances these models to markedly enhance segmentation accuracy, vital for digital herbarium management and agricultural innovation. The models were compared using various metrics including Pixel Accuracy and Mean IoU, with DeepLabv3+ featuring ResNet-18 showing superior performance with a Pixel Accuracy of 96.61%. Error indices like Weighted IoU, Mean BF Score are also potrayed so overall performance of the model can be evaluated. This study highlights the potential of specialized deep learning models to effectively handle intricate segmentation tasks, suggesting significant implications for sustainable farming practices through improved monitoring and decision-making. The findings propose transformative impacts for image-based botanical research, demonstrating the capabilities of tailored models in managing detailed, context-sensitive segmentation challenges.

***Index Terms*—** DeepLabV3+, ResNet-18, Segmentation,

1. INTRODUCTION

Accurate flower image segmentation is essential for progressing applications in the multiple disciplines of computer vision and botanical studies, like precision agriculture and ecological monitoring. These detailed natural scenes frequently have overlapping floral structures set against complex backgrounds, which pose significant challenges beyond the capabilities of conventional image processing techniques [1]. Image segmentation has advanced significantly as a result of machine learning advances, especially with regard to deep convolutional neural networks (CNNs) [2].

The DeepLabv3+ architecture, which uses a ResNet-18 backbone, is well-known for its ability to provide detailed contextual awareness and precision [3], which is critical for dealing with the unique challenges presented by botanical imagery, such as diverse flower morphology and dynamic changes in natural lighting [4]. This architecture excels in segmenting complicated images by using atrous convolution to collect multi-scale information without losing resolution, making it ideal for the deep analysis required in botanical studies [5]. This research provides a fresh custom model and strengths of the DeepLabv3+ with ResNet-18 as backbone architecture, specifically adapted to meet the peculiarities of the Oxford Flower Dataset.

Both models are evaluated to explore their synergistic potential and to understand the trade-offs involved in integrating these sophisticated approaches. This exploration is aimed at pushing the boundaries of what can be achieved with automated floral segmentation, providing valuable insights that could benefit both academic research and practical applications in fields reliant on precise and efficient image analysis.

Finally, the advances in segmentation techniques investigated in this study have far-reaching implications beyond academic research, influencing areas such as digital herbarium management and the development of sustainable agricultural practices, where improved image analysis capabilities can significantly contribute to better data-driven decisions.

2. LITERATURE REVIEW

The development of sophisticated image segmentation algorithms has substantially advanced the disciplines of botanical research and agricultural automation. These techniques, which are critical for discriminating between different plant sections in digital photos, have progressed from simple thresholding methods to complex deep learning models that provide subtle insights into plant anatomy. This paper examines the contributions of numerous scholars to the development of picture segmentation algorithms, with a special emphasis on their use in plant and flower detection.

A. Traditional Image Segmentation Techniques

Conventional image segmentation techniques, such as thresholding and edge detection, are fundamental ways for distinguishing foreground objects from their backgrounds in photographs. A study conducted in 2011 investigated the Preferential Image Segmentation (PIS) approach, which is known for its ability to segment leaf patterns from complicated backgrounds utilizing probabilistic curve evolution and particle filters [6]. This strategy improves the accuracy of plant recognition algorithms by highlighting stable elements such as leaf veins, which are critical for correct plant identification.

B. Deep Learning Models

The rise of deep learning has resulted in strong models capable of managing the complex patterns inherent in plant life. In 2020, researchers investigated the use of self-supervised learning models for floral image segmentation, allowing deep networks to identify plant components without requiring considerable manual annotation [7]. These models, particularly those which use the U-Net architecture, excel at segmenting complicated floral imagery, which improves thorough plant analysis.

C. Advanced Techniques

The U-Net architecture is well-known for its effective use in medical image segmentation, utilizing a symmetrical design to ensure detailed capture and integration of contextual information crucial for accurate plant segmentation.

Enhancing these capabilities, the DeepLab v3+ model leverages the power of deep convolutional networks and builds upon the ResNet-18 backbone to further refine its performance. It incorporates advanced features such as atrous convolution and spatial pyramid pooling. These additions allow DeepLab v3+ to efficiently handle multi-scale contextual data without compromising the image resolution, making it exceptionally suitable for complex tasks like segmenting detailed plant images across various scales and growth patterns. This architecture's ability to manage such variability makes it a strong tool for botanical research and applications requiring precise image segmentation.

For evaluating the efficiency of these segmentation algorithms, metrics such as Intersection over Union (IoU) and mean Average Precision (mAP) are frequently utilized. These measures give a quantitative basis for comparing the performance of different segmentation models, guaranteeing that advances in the algorithms translate into real enhancements in plant recognition applications.

.

Table – 1: Literature Review

Ref No.	Technique	Key Contributions & Applications	Dataset Used	Metrics
[8]	U-net	Symmetric encoder-decoder for detailed context capture - Medical image segmentation	Medical Image Dataset	IoU, Dice Coefficient
[9]	Deep-Labv3+	Atrous convolutions for multi-scale context - Semantic segmentation	PASCAL VOC 2012	Mean IoU
[10]	DeepLab-V3+ with ResNet-18 as backbone	Atrous Spatial Pyramid Pooling learning - Deep feature extraction & Segmentation	PASCAL VOC 2012 and Cityscape datasets	Mean IoU

3. METHODOLOGY

A. Data Preparation

The Oxford Flower Dataset was utilized, consisting of images across 17 classes, each with a segmentation mask. Images and masks were resized to 256x256 pixels for uniformity and computational efficiency. The data was split into training (80%) and validation (20%) subsets. Some of the images had no labels which were discarded using matching function which extracts file names from both image and label datastores, compares them, and filters out non-matching entries. Mapped non-flower to background class

B. Network Architectures and Training

Three models were employed to tackle the task of image segmentation:

1. DeepLabV3+ with ResNet-18 Backbone: The DeepLabV3+ model was introduced for capitalizing on advances in image segmentation. It uses an atrous spatial pyramid pooling module to robustly segment objects at many sizes, as well as a ResNet-18 backbone for efficient feature extraction. The use of ResNet-18, which is noted for its balance of performance and computational load, enables for effective training on the available dataset while using reasonable resources. The models were trained using the 'rmsprop' and 'adam' optimizers with an initial learning rate of 0.001, which was dynamically modified based on the validation loss to avoid overfitting. Training was carried out for up to 15 epochs, with early stopping performed depending on validation performance to avoid overtraining.

2. Custom Model: Developed to address specific challenges encountered in preliminary tests, such as distinguishing overlapping flowers. This model features deeper convolutional layers and enhanced techniques aimed at improving segmentation precision and computational efficiency. The Fig-1 shows the network architecture of the custom model used in this paper for image segmentation of flowers. The image portrays information regarding the name of the layer, activations and the type of the layer.

	Name	Type	Activations
1	input	Image Input	256(S) × 256(S) × 3(C) × 1(B)
2	conv1	2-D Convolution	256(S) × 256(S) × 64(C) × 1(B)
3	relu1	ReLU	256(S) × 256(S) × 64(C) × 1(B)
4	maxpool1	2-D Max Pooling	128(S) × 128(S) × 64(C) × 1(B)
5	conv2	2-D Convolution	128(S) × 128(S) × 128(C) × 1(B)
6	relu2	ReLU	128(S) × 128(S) × 128(C) × 1(B)
7	maxpool2	2-D Max Pooling	64(S) × 64(S) × 128(C) × 1(B)
8	conv3	2-D Convolution	64(S) × 64(S) × 256(C) × 1(B)
9	relu3	ReLU	64(S) × 64(S) × 256(C) × 1(B)
10	maxpool3	2-D Max Pooling	32(S) × 32(S) × 256(C) × 1(B)
11	transConv1	2-D Transposed Convolution	64(S) × 64(S) × 256(C) × 1(B)
12	relu4	ReLU	64(S) × 64(S) × 256(C) × 1(B)
13	transConv2	2-D Transposed Convolution	128(S) × 128(S) × 128(C) × 1(B)
14	relu5	ReLU	128(S) × 128(S) × 128(C) × 1(B)
15	transConv3	2-D Transposed Convolution	256(S) × 256(S) × 64(C) × 1(B)
16	relu6	ReLU	256(S) × 256(S) × 64(C) × 1(B)
17	conv4	2-D Convolution	256(S) × 256(S) × 2(C) × 1(B)
18	softmax	Softmax	256(S) × 256(S) × 2(C) × 1(B)
19	pixelLabels	Pixel Classification Layer	256(S) × 256(S) × 2(C) × 1(B)

Fig-1: Custom model architecture

4. MODEL EVALUATIONS

Mean Intersection over Union (Mean IoU): This metric computes the average overlap between predicted segmentation and ground truth, normalized by the sum of these two areas. It gives a thorough measure of segmentation accuracy for all classes in the sample. Mean IoU represents the average IoU score of all classes in a given image.

IoU score = TP / (TP + FP + FN) ---------------------------(i)

Where TP – True Positive
FP – False Positive
FN – False Negative

Pixel Accuracy: The percentage of successfully categorized pixels divided by the total number of pixels. This measure is simple; however, it may be biased towards classes with a greater area in the image.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad \text{----------------(ii)}$$

Where TP – True Positive
TN – True Negative
FP – False Positive
FN – False Negative

5. RESULTS

A set of best 4 models are studied for the image segmentation task, 2 pre-existing models and 2 custom models, but with a small change in training with optimizers.

Table-2: Models with best metrics scores

Models	Optimizers	Accuracy	Mean IoU Score
DeeplabV3+ With ResNet -18 – **M1**	Adam	96.61%	0.91657
DeeplabV3+ With ResNet -18 – **M2**	RMSProp	95.33%	0.88859
Custom Model-**M3**	Adam	92.84%	0.83103
Custom Model-**M4**	RMSProp	93.80%	0.85535

	flower	background
flower	0.92909	0.070915
background	0.020706	0.97929

Fig-2: DeeplabV3+ With ResNet -18 (Adam optimizer) Confusion matrix

GlobalAccuracy	MeanAccuracy	MeanIoU	WeightedIoU	MeanBFScore
0.96614	0.95419	0.91657	0.93497	0.75473

Fig-3: DeeplabV3+ With ResNet -18 (Adam optimizer) Metrics Scores

	flower	background
flower	0.93684	0.063161
background	0.04095	0.95905

Fig-4: DeeplabV3+ With ResNet -18 (RMSProp optimizer) Confusion matrix

GlobalAccuracy	MeanAccuracy	MeanIoU	WeightedIoU	MeanBFScore
0.95328	0.94794	0.88859	0.91246	0.74857

Fig-5: DeeplabV3+ With ResNet -18 (RMSProp optimizer) Metrics Scores

	flower	background
flower	0.81728	0.18272
background	0.030882	0.96912

Fig-6: Custom (Adam optimizer) Model Confusion Matrix

GlobalAccuracy	MeanAccuracy	MeanIoU	WeightedIoU	MeanBFScore
0.92839	0.8932	0.83103	0.86683	0.65577

Fig-7: Custom Model (Adam optimizer) Metrics Scores

	flower	background
flower	0.85287	0.14713
background	0.029464	0.97054

Fig 8: Custom model (RMSProp optimizer) confusion Matrix

GlobalAccuracy	MeanAccuracy	MeanIoU	WeightedIoU	MeanBFScore
0.93796	0.9117	0.85535	0.88367	0.67943

Fig-9: Custom Model Metrics (RMSProp optimizer) Scores

6. DISCUSSION

The training progression and evaluation metrics for the M1, M2, M3, and M4 models demonstrate their strong performance in image segmentation tasks, particularly in distinguishing between flower and background categories. The M1 model, which used DeepLabV3+ with a ResNet-18 architecture, trained with 'Adam' optimizer showed good generalization with a validation accuracy of 96.61% and high-class separation accuracy, as proven by a Mean IoU of 88.59%. It successfully defined object boundaries, with a Boundary F1 Score of 74.87%.

The M2 model demonstrated significant stability following initial fluctuations, with a validation accuracy of 95.33%, indicating good learning without overfitting. The mean IoU score for M2 was 83.103%, showing accurate object delineation.

M3's highest validation accuracy reached 92.84%, demonstrating its ability to handle complicated segmentation tasks. The Mean IoU and Boundary F1 scores demonstrated the model's ability to retain accurate object boundaries.

Finally, the M4 model's validation accuracy, with a peak of 93.8%, confirming its generalizability. The Mean Boundary F1 Score of 67.943%, however lower than other models, showed reasonable boundary detection effectiveness.

In the Fig-10, we can see that Deeplabv3+ with Resnet-18 as backbone (Model M1) has almost correctly segmented the flower and background while the custom model had a little bit of deficiency while producing the results as accurate as model M1.

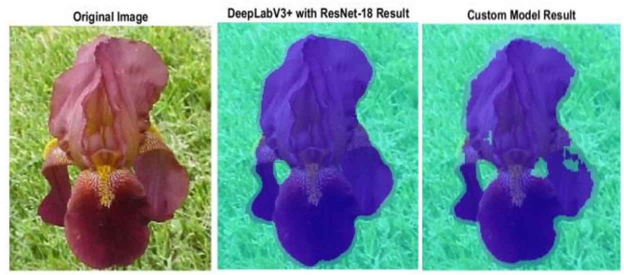

Fig-10: Original image and images produced with best different models of pre-existing and custom model.

7. CONCLUSION

Among the evaluated models, the M1 model, which utilizes the deeplabv3+ architecture enhanced with a resnet-18 backbone, trained with ‘adam’ optimizer, stands out as the best performer. This model achieved the highest validation accuracy at 96.61% and exhibited exceptional class separation capabilities with a mean intersection over union (iou) of 88.59%. Additionally, its boundary F1 score of 74.87% indicates superior ability in accurately delineating object boundaries, crucial for detailed segmentation tasks. The combination of high accuracy, precise boundary delineation, and robust generalization to unseen data makes the M1 model particularly effective for complex segmentation tasks such as differentiating flowers from their backgrounds in varied imaging conditions.

8. REFERENCES

[1] Nilsback, M.E. and Zisserman, A., 2008, December. Automated flower classification over a large number of classes. In *2008 Sixth Indian conference on computer vision, graphics & image processing* (pp. 722-729). IEEE.

[2] He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[3] Abd Almisreb, A., Aliyev, S.Z., Hamzah, R. and Jamil, N., 2022. Performance Evaluations of Convolutional Neural Network (CNN)-Based Models for Semantic Segmentation of Plant Leaf Diseases. *Journal of Positive School Psychology*, *6*(3), pp.8623-8635.

[4] Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A.L., 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, *40*(4), pp.834-848

[5] Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 801-818).

[6] Valliammal, N. and Geethalakshmi, S.N., 2011. Automatic recognition system using preferential image segmentation for leaf and flower images. *Computer Science & Engineering*, *1*(4), p.13.

[7] Saha, S., Sheikh, N., Banerjee, B. and Pendurkar, S., 2020, November. Self-supervised Deep Learning for Flower Image Segmentation. In *2020 14th International Conference on Innovations in Information Technology (IIT)* (pp. 126-130). IEEE.

[8] Ronneberger, O., Fischer, P. and Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18* (pp. 234-241). Springer International Publishing.

[9] Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 801-818).

[10] Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 801-818).