COMPARISON OF U-NET AND IMPROVED IMPLEMENTATION IN SEGMENTATION TASKS: ANALYSING THE IMPACT OF ADAM AND RMSPROP OPTIMISERS ON TRAINING TIME, ACCURACY AND NOISE LEVELS

Id: 20370023

ABSTRACT

This report will compare the performance of two segmentation network architectures: a standard U-Net with the Adam optimiser, and a proposed implementation of the U-Net using the RMSprop optimiser. The objective is to evaluate training time, accuracy, loss, and noise levels in segmenting flowers from the background in images.
The standard U-Net with Adam, while widely used in deep learning, particularly in the medical field [1][2], has presented to have a lower performance when applied to the Oxford Dataset, with an average accuracy of 85% and high noise levels. Despite attempts to optimise batch sizes and epochs, it continued to produce lower-than-expected results.
Conversely, the proposed implementation with RMSprop, designed to have a more compact architecture, achieved significantly better results. This network maintained a high accuracy of 95% and displayed lower noise in the segmented images.
Based on these findings, the report will discuss the implications of these results for segmentation tasks, including the potential for reduced training times, lower computational costs, and improved segmentation quality. The report concludes with recommendations for future research and further exploration of optimisers and architectural adjustments.

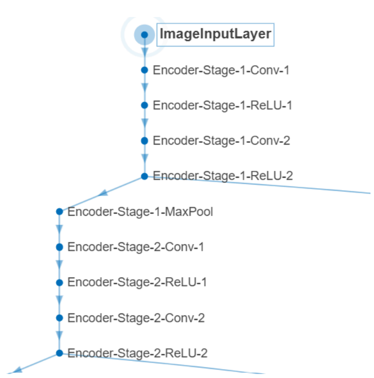
1. INTRODUCTION

Segmentation tasks in computer vision are essential for a wide variety of applications, such as medical imaging, object detection, and scene understanding in autonomous vehicles. These tasks involve dividing an image into several segments, where each segment corresponds to a category. Deep learning models have become a popular choice for segmentation due to their ability to learn complex features from raw image data.
When dealing with large amounts of data and hardware restrictions, optimisers play a crucial role in the training of these deep learning models. They control how the model updates the weights, which affects the speed, accuracy, and overall efficiency of the network. Different optimisers offer unique perspectives for adjusting these learning rates, some focusing on momentum, such as Adam, while others use adaptive learning strategies. The choice of optimiser can significantly affect the training time and computational resources required to achieve the desired level of accuracy.
This report will explore the performance of two segmentation models trained on the Oxford Flower Dataset, 17-class version. The task involves segmenting flowers from their backgrounds, a challenge rooted in the variability of shapes, colours, textures and density of the 2D images.
The standard U-Net architecture with the Adam optimiser is a well-known approach for segmentation. Unfortunately, its need for an extended amount of resources and long training times can imply a significant drawback. In addition to this, achieving high accuracy and low noise levels can be challenging in large, complex datasets like the Oxford Flower Dataset.
To address these issues, this paper proposes an alternative implementation - a smaller U-net-type architecture trained with the RMSprop optimiser. This approach aims to reduce training time and computational resources necessary while maintaining accuracy and clarity.
The scope of this paper includes a detailed comparison between the standard U-Net with Adam and the proposed implementation. The key metrics are training time, accuracy, loss, and noise levels, confusion matrix and noise levels to evaluate the effectiveness of these networks.

2. METHOD

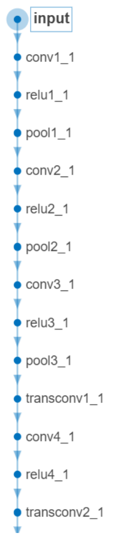
The Oxford Flower Dataset consists of 1360 256x256 images of various flowers from different angles, along with 846 image labels. Because not all images have been labelled, it was necessary to create a mapping between the images with labels and their respective labels.

2.1. UNet



The U-Net utilised here employs the Adam optimiser. This architecture consists of 58 layers, incorporating 3x3 convolutions throughout. With a total of 31 million learnable parameters, this segmentation method is more suitable to detailed tasks. The training setup involved a batch size of 4 and a learning rate of 0.001 for 10 epochs. In addition to this, ReLU activation functions are used to introduce nonlinearity.

2.2. Proposed Implementation



The proposed implementation is a custom CNN inspired by U-Net but with a simplified architecture. This design focuses on reducing the computational overhead while maintaining effective segmentation performance. It differs from the standard U-Net by having fewer layers, and significantly fewer learnable parameters, aiming to achieve a more resource-efficient model.

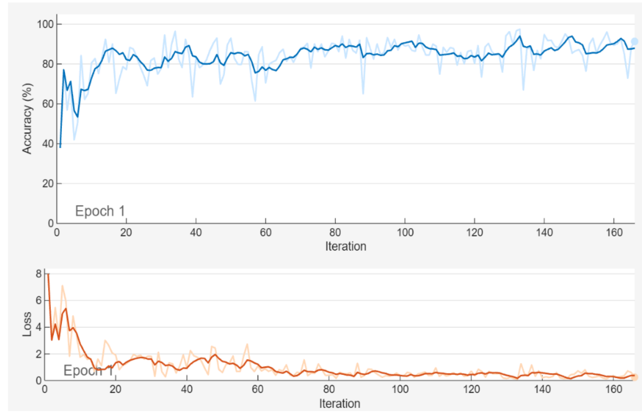
The proposed implementation has 22 layers and a total of 393.7k learnable parameters, a significant reduction compared to the standard U-Net. It uses 3x3 convolutional layers and 2x2 max pooling for downsampling. In addition to this, ReLU activation functions are used to introduce nonlinearity.

The primary metrics for evaluation are accuracy and loss. Additionally, the proposed implementation's segmentation results were evaluated for noise levels, with a focus on reducing unwanted artefacts.

The results from the paper "A Comparison of Adaptive Moment Estimation (Adam) and RMSProp Optimisation Techniques for Wildlife Animal Classification Using Convolutional Neural Networks" [3] highlight the importance of hyperparameter tuning, especially the learning rate, in achieving optimal performance. It was found that a learning rate of 0.001 often yielded the best results, suggesting that a similar approach would be beneficial for this segmentation study. Therefore, both methods have been assigned a consistent learning rate of 0.001.

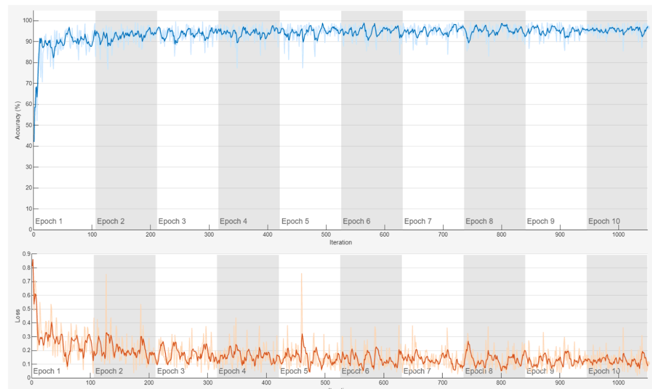
3. EVALUATION

The most noticeable difference between the 2 models is the training time. The U-Net architecture required around 87 minutes for the completion of 100 iterations, while the RMSprop required around 87 for the completion of the entire 10 epochs, 1050 iterations total. That is a huge difference in performance, with the proposed method being 10x faster.



Performance in U-Net using Adam

U-NET had an average of 1% loss with an average accuracy of 85%.



Overall Performance using RMSprop optimiser

On the other hand, the RMSprop optimiser allows a loss of <1% for the duration of the entire 10 epochs of the training process, and an accuracy of an average of 95%. In both cases, the accuracy does not improve after the first 100 iterations and both have displayed a low level of loss during the training process.

The segmentation accuracy has been mesured using global accuracy, mean accuracy, mean Intersection over Union (meanIoU), weighted Intersection over Union (weightedIoU), and mean Boundary F1-Score (meanBFScore). Global accuracy measures the overall correctness of the segmentation, accounting for all pixels in the dataset. Mean accuracy provides the average accuracy across all classes, offering insight into the model's performance on a per-class basis. MeanIoU calculates the average Intersection over Union for each class, giving a sense of the overlap between predicted and ground truth labels. WeightedIoU adjusts the meanIoU based on class frequency, providing a more balanced view when class distribution is imbalanced. Finally, meanBFScore assesses the segmentation's boundary precision and recall, indicating the model's ability to capture the correct edges and transitions between classes.

```
* Processed 846 images.

Evaluating semantic segmentation results
----------------------------------------
* Selected metrics: global accuracy, class accuracy, IoU, weighted IoU, BF score.
* Processed 846 images.
* Finalizing... Done.
* Data set metrics:

    GlobalAccuracy    MeanAccuracy    MeanIoU    WeightedIoU    MeanBFScore
    ______________    ____________    _______    ___________    ___________

       0.95503          0.95714       0.90388      0.91504        0.84434

Segmentation accuracy: 0.95503
>>
```

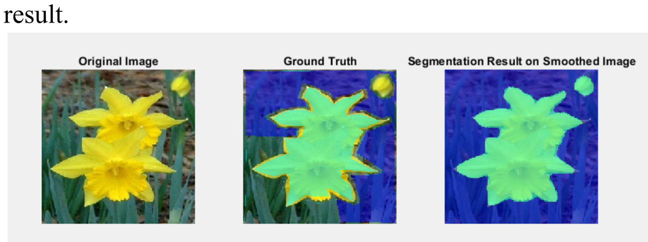
Accuracy stats for RMSprop

Based on these statistics, we can determine that, on average, the overlap between the predicted segmentation and ground truth was 90.4%, signifying a strong match between the model's predictions and the actual label. Furthermore, with a mean boundary of 84%, we can determine that the model captures the boundaries between different classes with a high degree of precision and recall.

U-Net Results from Segmentation

Although the results of the classic U-NET implementation have more defined edges, there is a significant amount of artifacting in the background, picking up unnecessary data. Even after pre-processing the images by applying a Gaussian filter to mitigate the issue, it was to no avail as the issue continued to persist.

One of the possible reasons for this artefact is the much deeper structure and numerous skip connections. This might lead to overfitting, leading to noise in the segmentation. This would be a reasonable assumption as the ground truth does not contain the same level of noise as the segmentation result.



Proposed Implementation Results

As you can see in the image above, the improved network accurately segments the image in the flower and the background. The edges are not as sharp as in the previous method, but there is a significantly less amount of noise. Conversely, one of the reasons this method shows less noise might be because RMSprop uses a single running average of squared gradients to adjust learning rates, which may lead to more stable and consistent updates.

3.1. Challenges

Due to the complexity of the pre-existing U-Net layer configurations, the network requires a significantly larger amount of memory to support the training process. A batch size of 8 has been assigned to both models, however, U-Net quickly ran out of memory when attempting to train on the Oxford set. The batch size had to be decreased to 4 to support the training. The smaller size of the improved network architecture greatly increased the efficiency of resource allocation, leading to less computational overhead and allowing to continue the process by decreasing the need for high-end hardware.

3.2. Implications

The accuracy results show a strong performance across a range of metrics. This level of accuracy and reliability has significant implications for practical use, suggesting a better applicability for this type of method on plant segmentation. This implementation highlights how modern top performing algorithms in certain scenarios are not a panacea to every segmentation problem. Here, a much simpler implementation outperformed a pre-existing network on a new set of data by focusing on the actual output of these networks.

4. CONCLUSION

In this comparative paper, the performance of 2 architectures has been evaluated, that being a standard U-Net architecture using the Adam optimiser against a proposed implementation with RMSprop in the context of segmentation tasks for the Oxford Flower Dataset. The analysis consisted of key metrics such as training time, accuracy, loss, and noise levels.

The results demonstrated that the proposed implementation with RMSprop significantly outperformed the standard U-Net with Adam in terms of training time, accuracy and noise reduction. The proposed implementation achieved a global accuracy of 95%, featuring a smaller architecture, with a 10x reduction in training time.

In contrast, the standard U-Net with Adam displayed a lower performance, with an average accuracy of 85%, higher loss, and increased noise levels in the segmentation results. Even after attempts to reduce the noise, U-Net continued to produce artefacts and inconsistent results.

These findings suggest that a simpler architecture with RMSprop can be more effective for certain segmentation tasks, especially when resource efficiency and reduced training time are critical. This can reduce costs long-term Given the results, further research into optimised architectures and alternative optimisers would be recommended. It would be beneficial to take into consideration other datasets with a different composition to the Oxford Flower Dataset to evaluate much simpler alternatives to improving efficiency.

5. REFERENCES

[1] Ronneberger, O., Fischer, P. and Brox, T. (2015) 'U-Net: Convolutional Networks for Biomedical Image Segmentation', in N. Navab et al. (eds) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer International Publishing, pp. 234–241. Available at: https://doi.org/10.1007/978-3-319-24574-4_28.

[2] Astono, I.P. *et al.* (2020) 'Optimisation of 2D U-Net Model Components for Automatic Prostate Segmentation on MRI', *Applied Sciences*, 10(7), p. 2601. Available at: https://doi.org/10.3390/app10072601.

[3] Kartowisastro, I.H. and Latupapua, J., 2023. A Comparison of Adaptive Moment Estimation (Adam) and RMSProp Optimisation Techniques for Wildlife Animal Classification Using Convolutional Neural Networks. Revue d'Intelligence Artificielle, 37(4).

[4] Poojary, R. and Pai, A. (2019) 'Comparative Study of Model Optimization Techniques in Fine-Tuned CNN Models', in *2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*. *2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*, pp. 1–4. Available at: https://doi.org/10.1109/ICECTA48151.2019.8959681.