# The Performance Enhancement Model for Image Segmentation: U-Net++

Jae-Hyung Sim*    Seong-Ju Lee*    Yeon-Jun Jeong*

## Abstract

*In this paper, we propose an improved version of the U-Net architecture [1], called U-Net++ [2], to enhance the performance of image segmentation tasks. Image segmentation is a fundamental task in computer vision, and U-Net has shown remarkable success in this field. However, U-Net++ takes the U-Net's strengths further by introducing novel features to address its limitations.*

## 1. Introduction

Image segmentation plays a crucial role in various computer vision applications, such as medical image analysis, autonomous driving, and object recognition. The U-Net architecture, with its encoder-decoder structure and skip connections, has revolutionized image segmentation tasks by effectively capturing contextual information while preserving fine-grained details.

### 1.1. Background

U-Net has demonstrated remarkable accuracy in image segmentation tasks, effectively delineating objects. However, it suffers from difficulties in accurately localizing objects located far away in deep networks. Moreover, the model's relative abundance of parameters poses challenges for efficient training.

## 2. Method

Using the KITTI dataset for image classification, you implemented both the U-Net and U-Net++ models for image segmentation and extracted segmentation results. By comparing the Intersection over Union (IoU) values for each model, you verified the effectiveness of U-Net++.

### 2.1. Improvement

U-Net++ utilizes multiple paths to incorporate a broader context of information. Each path generates feature maps of different sizes through various down-sampling and up-sampling operations. These feature maps are then combined to capture a wider range of contextual information, allowing for better recognition of objects located at a distance and improving the accuracy of segmentation results.

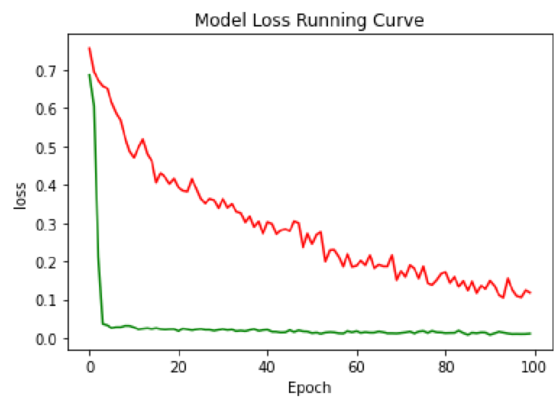


Figure 1. Learning Curve

Feature Map Reuse: U-Net++ reuses the feature maps generated from each path in different paths. This enables information sharing between paths, resulting in a more powerful representation with fewer parameters in the model. The reuse of feature maps enhances the efficiency of the model and allows for achieving high accuracy even with small datasets. Thus, these advancements have resulted in significant improvements.

### 2.2. Training

The U-Net++ model was trained with a specific configuration to achieve optimal performance. The training process consisted of 100 epochs, during which the Adam optimizer was utilized with a learning rate of 1e-4. To guide the model's training, the binary crossentropy loss function was employed, aiming to minimize the discrepancy between predicted and actual outputs.

Moreover, the model's performance was evaluated using the accuracy metric, which measures the proportion of correct predictions to the total number of predictions made. This allowed the researchers or developers to gauge how well the model performed on the given task.

To efficiently train the model, a batch size of 32 was chosen. This means that the dataset was divided into batches of 32 samples, and the model updated its parameters based on

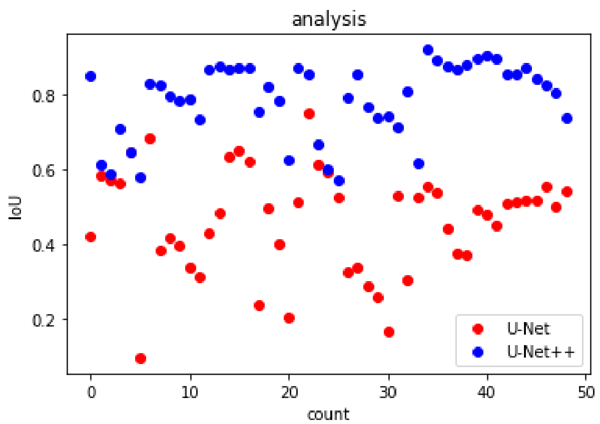


Figure 2. IoU Scatter Graph

| | AP50.0 | AP70.0 | AP90.0 |
|---|---|---|---|
| Unet | 0.251427 | 0.027829 | 0.000000 |
| Unet++ | 0.904309 | 0.623125 | 0.006494 |

Figure 3. AP score

the average loss across these batches, making the training process more memory-efficient.

## 3. Result

Based on Figure 2, which compares the IoU (Intersection over Union) of U-Net and U-Net++ on 50 KITTI road images each, the conclusive result is that U-Net++ achieved better segmentation results compared to U-Net. The IoU metric measures the overlap between predicted and ground truth segmentation masks, and a higher IoU indicates more accurate segmentation. Figure 3 shows that the average precision (AP) for different intersection over union (IoU) thresholds (AP50, AP70, and AP90) for U-Net and U-Net++ follow the same trend as described earlier. As can be seen from the results, U-Net++ outperforms U-Net at all three IoU thresholds and shows higher average precision at IoU levels 50, 70 and 90.

Therefore, U-Net++ demonstrated improved performance in accurately delineating road regions in the KITTI dataset, making it a more effective model for road image segmentation tasks.

## References

[1] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 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*, pages 234–241. Springer, 2015. 1

[2] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, pages 3–11. Springer, 2018. 1