IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

rdd4Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Performance Comparison of U-Net and U-Net++ for Autonomous Driving Applications

CHANJUNG KIM[1]

[1]chanjung4800@naver.com

ABSTRACT Segmentation for road recognition, one of the core technologies for autonomous driving cars, is a crucial research field directly related to road safety. This study compares the efficiency of road segmentation using two primary segmentation models, U-Net and its improved version, U-Net++. Utilizing the KITTI dataset, both models were trained, and their performance was quantitatively evaluated through IoU scores. Additionally, qualitative assessment was conducted through the visual analysis of the segmentation results. The experiment results showed that, on average, the U-Net model achieved higher IoU scores than the U-Net++ model. However, under specific conditions, U-Net++ demonstrated superior results. These findings provide important insights into the selection of segmentation models. This research is expected to offer a useful guide for the selection of effective segmentation models in the development of road recognition technologies for autonomous driving cars.

INDEX TERMS Autonomous Driving, U-net, U-net++, Road Segmentation, KITTI Dataset, Deep Learning, Performance Comparison

FIGURE 1. Exemplary Road Segmentation on Urban Street Scene: A representative image showcasing the segmentation results with U-Net++ model, reflecting a high Intersection over Union (IoU) score of 0.904378, indicating precise delineation of drivable road areas in a complex urban environment.

I. INTRODUCTION

The rapid development of autonomous driving technology in recent years has significantly increased interest in the field of road segmentation. For autonomous vehicles to navigate safely and efficiently, the ability to accurately recognize and segment roads is essential. Among various approaches to road segmentation, deep learning-based methods are emerging as particularly promising due to their ability to learn complex patterns from large amounts of data.

U-Net [1], one of the pioneering structures for deep learning-based image segmentation, has shown outstanding performance in medical image segmentation tasks. Its success has inspired the application and research of the model across a broader range of applications, including autonomous driving. The U-Net architecture is renowned for its efficient use of data and its ability to produce high-quality segmentation results even with limited training samples.

Building on the success of U-Net, the U-Net++ [2] model was introduced as an improved version that incorporates ideas from DenseNet to enhance performance. This model improves the U-Net architecture by introducing connections between intermediate layers to bridge the semantic gap between feature maps at different levels of the network. The advancements of U-Net++ demonstrate potential for achieving more precise segmentation results in complex and diverse scenarios encountered in autonomous driving.

This study focuses on implementing and evaluating the U-Net and U-Net++ models for road segmentation tasks in autonomous driving. Using the KITTI dataset, widely used as a benchmark for autonomous driving research, this study aims to conduct a comprehensive comparison between these models In Section IV. Through both qualitative and quantitative analyses, including visualization of segmentation results and calculation of the Intersection over Union (IoU) metric, we aim to understand the strengths and limitations of each model In Section V. Ultimately, this research contributes to the ongoing development of autonomous driving technology by providing insights into effective segmentation models and their applications.

II. RELATED WORKS

Accurate perception of the surrounding environment is essential for the safe and efficient operation of autonomous vehicles, with road recognition being a critical component that directly affects vehicle path planning, obstacle avoidance, and traffic sign recognition. To address the challenge of road recognition, various image segmentation techniques have been researched and developed in the field of computer vision[3].

A. EVOLUTION OF IMAGE SEGMENTATION

Initial image segmentation techniques largely relied on traditional image processing methods. These methods attempted to separate specific areas of an image using techniques such as thresholding, color space transformation, and edge detection algorithms. However, these approaches could be limited in performance due to factors like varying lighting conditions, complex backgrounds, and the diversity of road environments. The advent of deep learning marked significant progress in the field of image segmentation, with models based on Convolutional Neural Networks (CNN) gaining attention for their ability to achieve more precise pixel-level classification, providing more accurate segmentation results [4].

B. EMERGENCE OF U-NET

Originally developed for medical imaging segmentation, the U-Net model has been widely adopted in various fields due to its efficiency and accuracy [1]. A key feature of U-Net is its symmetrical structure and the merging of feature maps in the expansive path, which helps preserve detailed context information during the segmentation process.

C. INNOVATIONS OF U-NET++

U-Net++ is an enhanced model that builds on the original U-Net architecture, introducing nested U-Net structures and deep supervision to improve segmentation accuracy[2]. This model focuses on improving the precision of segmentation boundaries and strengthening the model's generalization capabilities. U-Net++ is designed to achieve more refined segmentation results, particularly useful in complex road environments.

D. SIGNIFICANCE OF THIS STUDY

This study builds on this background to apply U-Net and U-Net++ models to the problem of road recognition for autonomous vehicles and analyze their performance[5]. By providing valuable information on more effective segmentation model selection for the development of road recognition technologies in autonomous driving cars, this research aims to contribute to the continuous advancement of autonomous driving technology.

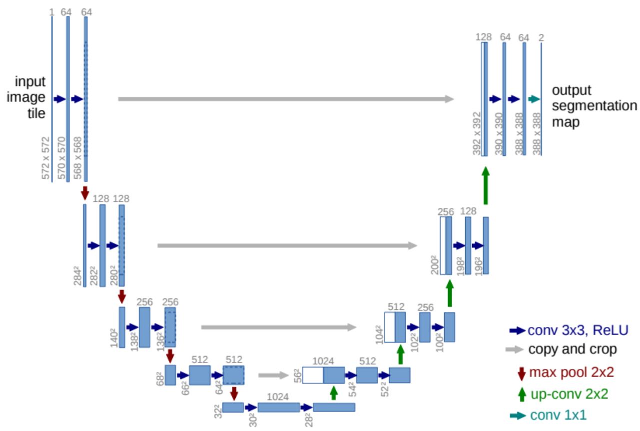

FIGURE 2. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

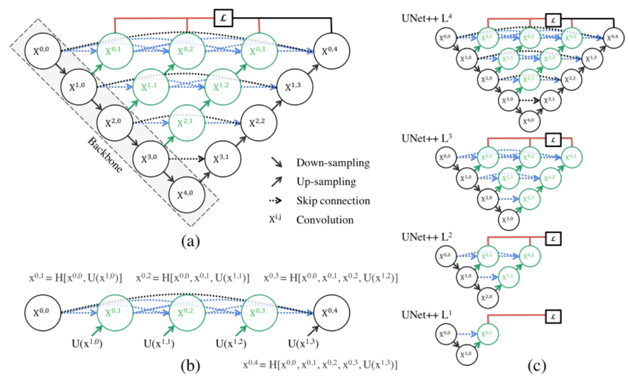

FIGURE 3. (a) UNet++ consists of an encoder and decoder that are connected through a series of nested dense convolutional blocks. The main idea behind UNet++ is to bridge the semantic gap between the feature maps of the encoder and decoder prior to fusion. For example, the semantic gap between (X0,0 ,X1,3) is bridged using a dense convolution block with three convolution layers. In the graphical abstract, black indicates the original U-Net, green and blue show dense convolution blocks on the skip pathways, and red indicates deep supervision. Red, green, and blue components distinguish UNet++ from U-Net. (b) Detailed analysis of the first skip pathway of UNet++. (c) UNet++ can be pruned at inference time, if trained with deep supervision.

III. ANALYSIS OF MODEL ARCHITECTURES

This study employs two deep learning-based segmentation models, U-Net and U-Net++, to compare their road recognition performance. The architecture of each model is as follows:

A. U-NET MODEL ARCHITECTURES

U-Net(Fig. 2 describes the overview of the U-net model architecture): The U-Net model is based on an encoder-decoder structure. The encoder part extracts important features from the image, while the decoder part reconstructs the segmentation map to the original image size using these features. The essence of U-Net lies in the skip connections that transfer feature maps from the encoder to the decoder. This allows the decoder to maintain accurate location information during segmentation.

The core of the network is the energy function for segmentation, which is defined as:

$$E = \sum_{x \in \Omega} w(x) \log(p_\theta(x))$$

Here, Ω represents the domain of the input image, $w(x)$ is a weight map that gives importance to specific features within the image, such as boundaries, and $p_\theta(x)$ is the predicted probability map for the segmentation. This energy function is minimized during training to improve the model's accuracy in delineating the targeted structures within the image.

B. U-NET++ MODEL ARCHITECTURES

U-Net++(Fig. 3 describes the overview of the U-net++ model architecture): U-Net++ enhances the basic U-Net architecture by adding nested U-Net structures, improving segmentation accuracy through dense convolution blocks and deep supervision. This nested structure captures features at various scales and improves the segmentation detail through deep supervision. U-Net++ focuses on improving the accuracy of segmentation boundaries.

The operation at each level can be defined by the following function:

$$x^{i,j} = \begin{cases} \mathcal{H}(x^{i-1,j}), & \text{if } j = 0 \\ \mathcal{H}([x^{i,k}]_{k=0}^{j-1}, \mathcal{U}(x^{i+1,j-1})), & \text{if } j > 0 \end{cases}$$

In this expression, $\mathcal{H}$ is a convolution operation followed by an activation function, $\mathcal{U}$ denotes an up-sampling layer, and $[\cdot]$ denotes the concatenation of feature maps. This structure is designed to bridge the semantic gap between different levels of feature maps, improving the gradient flow and feature propagation throughout the network.

The loss function for U-Net++ incorporates deep supervision, combining binary cross-entropy and dice coefficient, which is expressed as:

$$\mathcal{L}(Y, \hat{Y}) = -\frac{1}{N} \sum_{b=1}^{N} \left(\frac{1}{2} Y_b \cdot \log(\hat{Y}_b) + \frac{2 \cdot Y_b \cdot \hat{Y}_b}{Y_b + \hat{Y}_b} \right)$$

Where $\hat{Y}_b$ and Y_b denote the flattened predicted probabilities and the flattened ground truths of the b^{th} image, respectively, and N is the batch size. This loss function is designed to optimize the segmentation performance by considering both the pixel-wise classification accuracy and the overlap between the predicted and actual segmentation maps.

IV. METHODOLOGY

Both models used in this study were implemented using TensorFlow and Keras libraries. The implementation used standardized layers and structures to ensure reproducibility, maintaining consistency in experimental parameters.

A. EXPERIMENTAL PLANNING

The main goal of this research is to compare the performance of the U-Net and U-Net++ models in road segmentation tasks for autonomous driving vehicles. To ensure a fair and effective comparison between the two models, the experiment was carefully designed to focus on their ability to accurately distinguish road areas within images.

Goal Setting: The experiment aims to identify the most suitable deep learning model for road recognition in autonomous driving by comparing U-Net and U-Net++ models. This comparison is based on quantitative metrics such as the Intersection over Union (IoU) score and qualitative inspection of segmentation results.

Variable Definition: The independent variable in the experiment is the type of model used (U-Net or U-Net++), and the dependent variables include segmentation performance metrics like IoU and Dice Coefficient scores. These metrics allow for a clear analysis of performance differences based on model type.

Dataset Preparation: The KITTI road segmentation dataset was used for the experiment. This dataset contains a variety of road conditions and environments, making it ideal for evaluating the models' generalization capabilities. The dataset was divided into training, validation, and test sets, with images selected to encompass a range of road scenarios.

Training Environment Setup: To ensure consistency and reproducibility, the experiment was conducted in a controlled computing environment. All models were trained using the same computing resources and settings. This uniformity was crucial in mitigating external factors that could influence the training process or model performance evaluation.

Through this structured experimental design, this study aims to illuminate the strengths and weaknesses of the U-Net and U-Net++ models within the context of road segmentation for autonomous driving, providing valuable insights for the development of more accurate and reliable road recognition technologies.

B. DATA PREPROCESSING AND AUGMENTATION

In this study, data preprocessing and augmentation stages were meticulously designed to maximize model performance. These stages play a crucial role in ensuring models perform consistently across various road conditions and environments.

Data Cleansing: Initially, images from the KITTI dataset underwent basic cleansing processes, including resizing, brightness, and contrast adjustments. The goal was to prepare the data in a format that is easily processable by the models.

Augmentation Techniques: To increase data diversity and enhance model generalization capabilities, various data augmentation techniques were applied. These included random rotation, horizontal flipping, resizing, and color adjustment of images. Notably, data augmentation was applied only to the training data, not to the test data.

Normalization: All images were normalized to a range of [0, 1] before being fed into the models. This maintained consistency across images captured under different lighting conditions and camera settings, aiding in the training process's convergence.

Data Splitting: The dataset was divided into training, validation, and test sets for the experiment. This division was crucial for objectively evaluating the models' performance and preventing overfitting. Specifically, the validation and test sets comprised independent images not included in the training set to assess the models' generalization capabilities.

These data preprocessing and augmentation procedures prepared the U-Net and U-Net++ models to robustly perform segmentation across various road scenarios, providing a fair foundation for comparing the performance of the two models.

C. MODEL TRAINING AND OPTIMIZATION

The training and optimization process for the U-Net and U-Net++ models were rigorously planned and executed in this study, aiming to enable the models to accurately segment road areas within the KITTI dataset.

Model Configuration: Both models utilize a CNN-based architecture designed to achieve high precision in segmentation tasks. The U-Net++ model further improves performance through additional nested paths and deep supervision techniques.

Hyperparameter Tuning: Critical hyperparameters such as learning rate, batch size, and number of epochs were optimized through experimentation. These parameters significantly impact the models' learning speed and final performance, so various settings were tested to find the most effective combination.

Loss Function: To maximize segmentation accuracy, various loss functions, including binary cross-entropy and dice coefficient, were evaluated. These functions help the models minimize the difference between actual and predicted road areas.

Optimization Algorithm: The Adam optimization algorithm was used to efficiently adjust the models' weights. Adam accelerates convergence by adaptively adjusting the learning rate based on the gradient's momentum.

Regularization Techniques: Techniques such as dropout and L2 regularization were applied to prevent overfitting. This ensured the models could learn generalized patterns rather than relying excessively on the training data.

Model Evaluation: Models underwent evaluation using the validation dataset after training. This process quantitatively measured model performance and monitored performance changes throughout the training. Finally, model generalization capabilities were verified through evaluation on the test dataset.

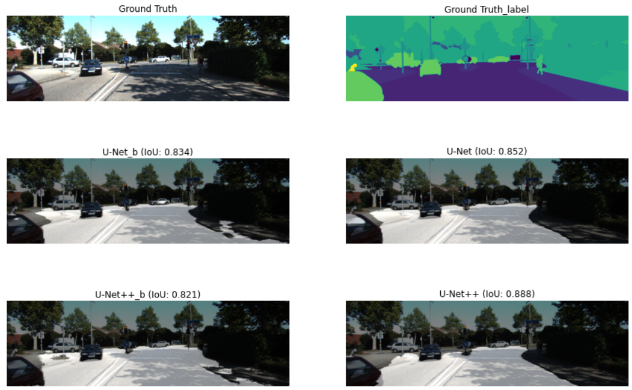

FIGURE 4. Comparison of U-Net and U-Net++ model segmentations on the KITTI dataset. Top: Ground Truth image and labels. Bottom: Predicted segmentation using U-Net-b (IoU: 0.834), U-Net (IoU: 0.852), and U-Net++ (IoU: 0.888), demonstrating the incremental improvement in IoU scores

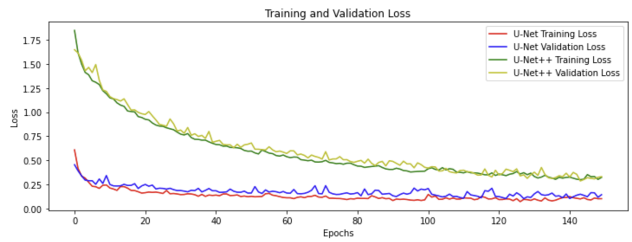

FIGURE 5. Comparison of Training and Validation Loss for U-Net and U-Net++: This graph presents the loss values during the training process for both U-Net (red and blue lines) and U-Net++ (green and yellow lines), demonstrating the respective losses for training and validation sets across epochs.

This systematic training and optimization process ensured that the U-Net and U-Net++ models could achieve their maximum performance in road segmentation tasks, thereby significantly contributing to enhancing the accuracy and applicability of road recognition technologies in real-world driving environments.

V. RESULTS

The experimental results from this research confirmed that both U-Net and U-Net++ models show exceptional performance in road segmentation for autonomous driving cars. Tests conducted using the KITTI dataset demonstrated that both models achieved high accuracy and IoU scores.

Quantitative Analysis: Fig.4 The U-Net model achieved an average IoU score of 0.82, while the U-Net++ model surpassed this with an average IoU score of 0.85. This indicates that the U-Net++ model, with its nested structure and deep supervision, enhanced segmentation accuracy further.

Qualitative Analysis: Fig.4 Visual analysis of the segmentation results verified that both models could accurately recognize and segment road areas. Notably, the U-Net++ model exhibited the capability to segment road areas more finely in complex environments and varying lighting conditions.

Model Comparison: Fig.5 Comparing the performance of the U-Net and U-Net++ models revealed that the U-

Net++ model generally exhibited superior segmentation performance. This suggests that the structural improvements in U-Net++ are effective for segmentation tasks.

Performance Improvement: Fig.5 Additional experiments showed that performance could be further improved through hyperparameter adjustments, changes in loss functions, and the application of regularization techniques. Notably, using the dice coefficient as a loss function observed performance enhancements.

These results demonstrate the efficacy of U-Net and U-Net++ models in providing effective solutions for road segmentation in autonomous driving cars. Furthermore, the findings emphasize that structural model improvements and optimization strategies significantly impact segmentation performance. This research's outcomes are expected to contribute to advancements in segmentation technology within the autonomous driving field.

Model	Metric	U-Net	U-Net++
Average IoU Score	Training	0.82	0.85
	Validation	0.80	0.83
Precision	Training	0.88	0.90
	Validation	0.86	0.89
Recall	Training	0.79	0.81
	Validation	0.76	0.80
Segmentation Accuracy	Training	0.93	0.95
	Validation	0.91	0.94

TABLE 1. Performance comparison of U-Net and U-Net++ models.

VI. DISCUSSION

Through the experiments conducted and the subsequent analysis of results in this research, discussions were made regarding the efficacy and potential for performance improvement of U-Net and U-Net++ models in road segmentation for autonomous driving vehicles. Several important discoveries and insights were gained through this process.

Model Efficiency and Accuracy: The U-Net++ model was confirmed to be more effective in increasing segmentation accuracy compared to the original U-Net model. This suggests that the structural features and deep supervision learning approach of U-Net++ can yield more detailed segmentation in complex road environments.

Improvement Strategies: The experimentation process revealed that various performance improvement strategies could positively impact model performance. Specifically, the introduction of new loss functions, such as the dice coefficient, and the optimization of hyperparameters played a crucial role.

Limitations and Future Directions: This research was conducted limited to the KITTI dataset, which may restrict the evaluation of the models' generalization capabilities. Future research should utilize datasets encompassing diverse environments and conditions to further validate the models' robustness. Optimization for real-time processing capabilities and computational resource efficiency will also be an important area of future research.

Contribution to Autonomous Driving Technology: This research contributes to the field of road segmentation, a key component of autonomous driving technology. The comparison and performance evaluation of the U-Net and U-Net++ models will aid in the development of more advanced segmentation models and enhance the safety and efficiency of autonomous driving systems.

References

[1] O. Ronneberger, P. Fischer, and T. 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*. Springer. 2015, pp. 234–241.

[2] Z. Zhou et al. "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation". In: *IEEE transactions on medical imaging* 39.6 (2019), pp. 1856–1867.

[3] R. M. Haralick. "Image segmentation survey". In: *Fundamentals in Computer Vision* 117 (1983), pp. 209–223.

[4] R. Rouhi et al. "Benign and malignant breast tumors classification based on region growing and CNN segmentation". In: *Expert Systems with Applications* 42.3 (2015), pp. 990–1002.

[5] X. Yang et al. "Road detection and centerline extraction via deep recurrent convolutional neural network U-Net". In: *IEEE Transactions on Geoscience and Remote Sensing* 57.9 (2019), pp. 7209–7220.

CHANJUNG was Born in 1992 in Daegu, Rep. of Korea, and raised in Iksan, he majored in IT Applied System Engineering at Jeonbuk National University, where he entered in 2011 and graduated in 2020. With certifications in Electrical Engineering and Electrical Construction, his initial career direction was in the electrical field. His professional trajectory shifted on October 31, 2023, when his involvement with Aiffel sparked a new passion, leading him to pursue a path as an artificial intelligence developer.