Autonomous Driving Lane Segmentation U-Net Variant Comparative Analysis

Rene Lisasi
CCSE
Kennesaw State University
Marietta, GA
renelisasi@hotmail.com

Yehong Huang
CCSE
Kennesaw State University
Marietta, GA
hyh159263@gmail.com

***Abstract*—The National Highway Traffic Safety Administration, part of the United States Department of Transportation issued a Standing General Order requiring identified manufacturers and operators to report to the agency certain crashes involving vehicles equipped with automated driving or advanced driver assistance systems. Since 2021, 4303 crashes have been reported with an average of 200 in May 2025 alone, resulting in loss of consumer trust. This study compares the current state of the art lane segmentation to reduce accidents further and increase customer trust by focusing on an important facet of autonomous driving, lane segmentation using advanced deep learning architectures.**

Keywords—Autonomous Driving, Consumer Trust, Lane Segmentation, Deep Learning

I. Introduction

Recently the automotive industry underwent rapid changes, from promises of fully autonomous driving to only mustering lane changing systems, companies like Tesla, Waymo, and Cruze are continually seeking improvement in the space, while legacy manufacturers are trying their best to keep up. One way for legacy manufacturers to overtake the more recent startups is to invest in autonomous research and development (R&D) but the question is where? Autonomous driving is made up of several key categories namely object detection, lane segmentation, reinforcement learning, path planning, risk evaluation, task planning, and local path generation to name a few.[1] Long term investment in any of these areas will see fruitful return in terms of advancement, and leadership. However, only a few will have influence on consumer trust, that is, the areas that have direct impact on the features of the decision making. One such area is lane segmentation.

To describe the impact of lane segmentation on consumer trust, the differences between the two types of autonomous driving systems must be understood. Automated driving systems (ADS) aims to perform the entire dynamic driving task on a sustained basis within a defined operational design domain without driver domain, while Level 2 ADS (ADAS) provides both speed and steering input when the driver assistance system is engaged but require the human driver to remain fully engaged in the driving task at all times. [2]

2089 of the 4303 NHTSA reported crashes were driver assisted (ADAS), while only 1600 were non-driver assisted (ADS). In the month of August they both dropped tremendously down to 15 crashes. Therefore the assistance of a human has proven to be detrimental to the number of reported crashes, so what can the reason be for lower trust? The answer is perceived safety and reliability while naturally, news and sights of or about crashes subdues trust. Enhancing safety, reliability, and privacy measures boost the adoption of autonomous vehicles. [3]

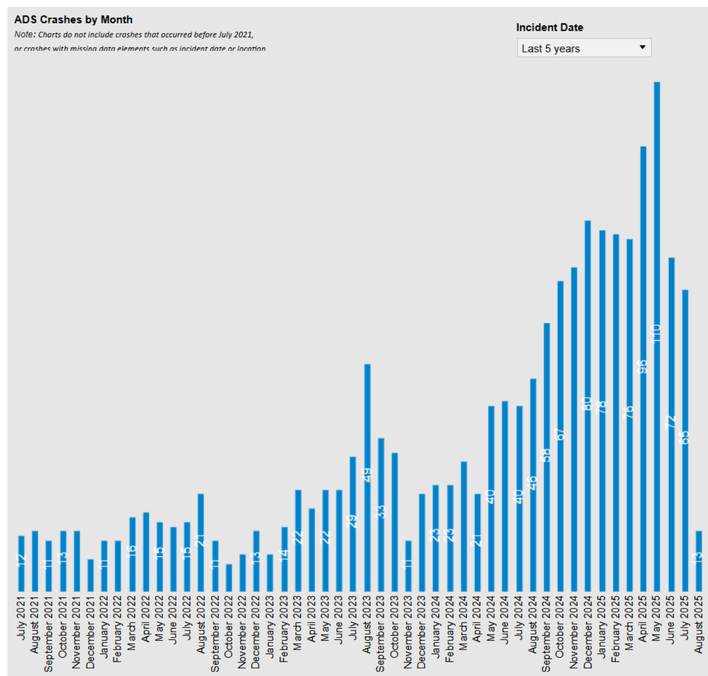

Figure 1 ADS Reported Crashes Since 2021

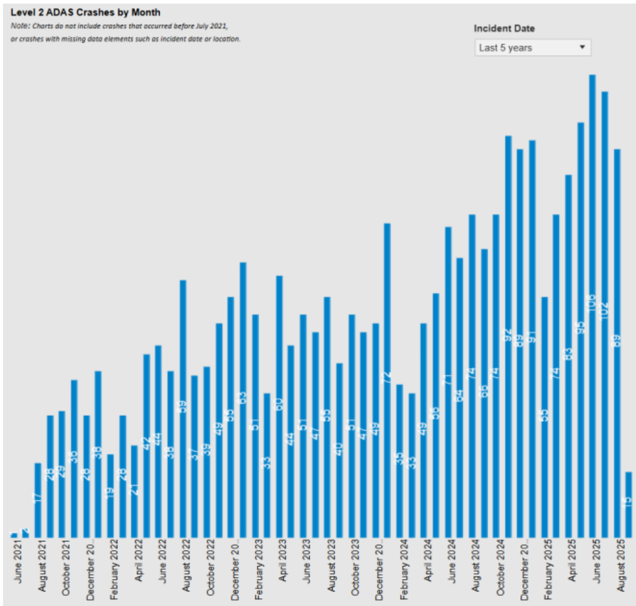

Figure 2 ADAS Reported Crashes Since 2021

A study was conducted into driver behavior and found that road markings influenced the behaviors of both the lateral lane position and driving speed of vehicles [4]. A faster and more precise lane segmentation of identification of longitudinal positioning would have a positive impact on autonomous vehicle reliability whether ADAS or ADS Evidence is shown in table 1:

TABLE I. ADS & ADAS CRASHES FREQUENCY TABLE

Roadway Type	Crash Frequency	
	ADAS	***ADS***
Highway	94	NA
Street	28	78
Unknown	71	NA
Intersection	21	50
Rural Road	5	NA
Parking Lot	1	14

Therefore, lane segmentation must happen faster and more precisely to increase customer trust, not simply creating arbitrary reliability and risk perception benchmarks.

II. RELATED WORKS

A precursor to lane segmentation is lane detection: feature extraction, feature segment grouping, and lane model fitting. Traditional lane detection require digital image processing hand-crafted features like Hough Line Detection or SIFT gradient detection, however, these are not robust enough to resist not just noise, weather, and light challenges but also curves as well as intersections with limited lane markings. [5] Researchers have developed intensity, color, and edge based methods [6]. Intersections specifically being the most challenging and needing an entirely new paradigm to capture the features. Lane detection is therefore bottlenecked by the need for perfect markings and as commonly known, road markings wear out within 1-2 years, therefore a different approach must be undertaken.

Identify applicable funding agency here. If none, delete this text box.

III. METHODOLOGY

A. Lane Detection, Traditional Image Processing & CNNs

That is where lane segmentation comes in, lane segmentation is highlighting regions of the road deemed to represent a drivable lane rather than markings depicting lanes, resulting in greater room for error. Some researchers have been able to harness convolutional networks (CNN) for lane segmentation but CNNs are better suited for feature extraction than segmentation. [CNN better at feature extraction than segmentation]. The conventional segmentation method is an encoder-decoder model and although there are many, the industry and research standard is a U-Net.[7]

B. Vanilla U-Net

The segmentation model is based on the U-Net architecture proposed by Ronneberger *et al.* [7]. To implement this architecture, we adopted the PyTorch-UNet library available at [8]

A Vanilla U-net is an encoder decoder model made up of several CNNs and the reason they are so good at segmentation is that the encoder extracts image features through multiple convolution and pooling operations, whereas the decoder gradually restores feature images through up-sampling. [9] In general, they can localize by training the encoder network predict the class label of each pixel by providing a local region (patch) around that pixel as input instead of an entire image with a single class label.

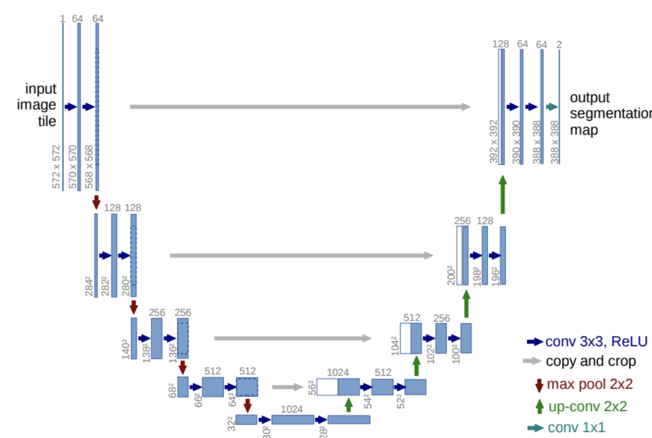

Figure 3 U-net architecture

C. Deep Residual U-Net (Res U-net)

In semantic segmentation, it is important to use low level details while retaining high level semantic information. Intuition would lead to believe that a deeper network would improve results, however, this leads to vanishing gradient. Residual U-Net consists of stacked residual units that facilitate information propagation without degradation by using skip connections. Skip connections feed the original input in later layers. The result is a U-Net with ease of training. [10]

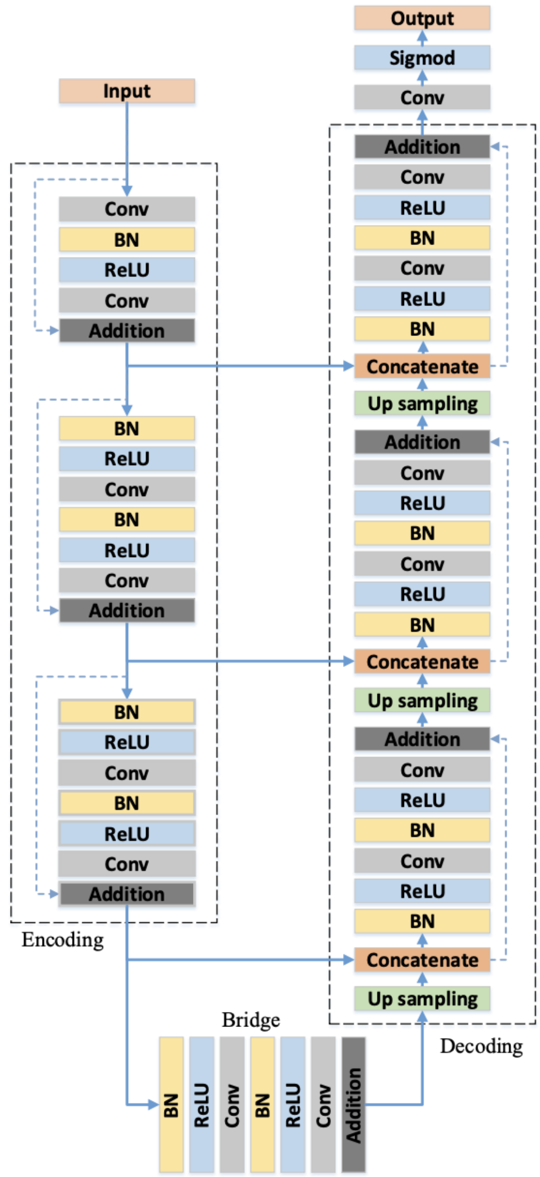

Figure 4 The architecture of the Residual U-net

D. Attention U-Net (Att U-Net)

To capture a wider receptive field (more local context), the features are gradually down-sampled in a CNN which can lead to degradation. The attention network however, suppresses irrelevant features and focuses on a subset of target structures. The gating vector prunes lower level features and is more computationally expensive but yields better performance. [11] For the specific implementation of these attention gates, we adopted the code framework provided by [12] and integrated it into our pipeline.

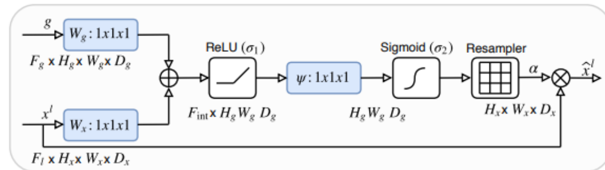

Figure 5 U-Net Attention Block

IV. EXPERIMENTAL SETUP

A. Implementation Details

Hardware/Software Environment: Python 3.11.9, PyTorch 2.1.2 + cu118, NVIDIA RTX 4090, 64 GB RAM. Hyperparameters: Learning rate of 1e-4, 100 epochs, batch size of 128 for lane segmentation and 8 for road segmentation, and finally AdamW optimizer. The models compared were residual U-net, U-net, and attention U-net. The implementation consisted of a custom data class, training, and inference. The data was collected from public sources.

B. Dataset Description

There are two datasets: SDCND Advanced Lane Lines dataset [13] and Semantic Segmentation Makassar (IDN) Road Dataset [14]. The specific version of the road dataset used had 299 training samples and 74 testing samples. The lane dataset is the collection of image frames of 720 pixels video clips, which contain around 12764 images of different environmental conditions. In this dataset, there are 17.4% clear night view, 16.4% rainy morning view, and 66.2% cloudy afternoon view. Besides, it contains around 26.5% straight roads, 30.2 curves roads, and 43.3% very curvy road. Any pixels with a 0 were background and any with 1 or positive integer were labels for both datasets. The road dataset was resized to 512 by 512.

C. Evaluation Metrics

The metrics used were Intersection over Union (IoU) [15], Dice Score (DSC) [16], and F1 Score (F1). IoU is a measurement based on the Jaccard Index, a coefficient of similarity for two sets of data. It measures the overlapping area between the predicted bounding box B_p and the ground truth bounding box B_{gt} divided by the area of union between them.

$$\boldsymbol{J(B_p, B_{gt}) = IoU = \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})}} \qquad (1)$$

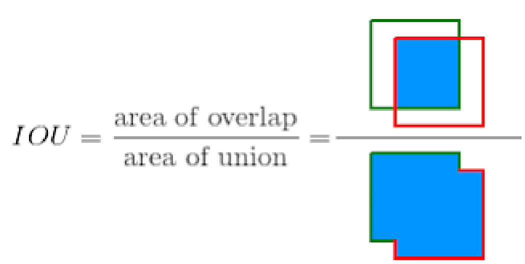

Figure 6 IoU Diagram

The DSC score assesses the similarity between ground truth, Y, and the prediction, $\hat{Y}$, binary masks by considering counts of true positives TP, false positives FP, and false negatives FN.

$$\boldsymbol{DSC} = \mathbf{2}\left(\mathbf{2} + \frac{\boldsymbol{FP}}{\boldsymbol{TP}} + \frac{\boldsymbol{FN}}{\boldsymbol{TP}}\right)^{-\mathbf{1}} \tag{2}$$

The F1 score is a score typically used in classification to indicate harmony. It is defined as the doubling of the product between precision and recall.

$$\boldsymbol{F1} = \frac{\mathbf{2} * \boldsymbol{TP}}{\mathbf{2} * \boldsymbol{TP} + \boldsymbol{FP} + \boldsymbol{FN}} \tag{3}$$

V. RESULTS

The following results show the input image, label mask, prediction, and absolute error between the label and prediction.

A. Lane Segmentation Scenario

The first set of results are lane segmentation. Every model failed to exclude the lane markings and predicted a whole surface. Nonetheless, there are some differences that can be visually distinguished. (120, 60-70) show how the attention U-net being slightly more aware of the lane limits. Also, (40,70) shows where the attention U-net is superior by correctly segmenting the lane markings. Besides that, very little can be observed visually. The loss converges much earlier for U-net than the attention U-net.

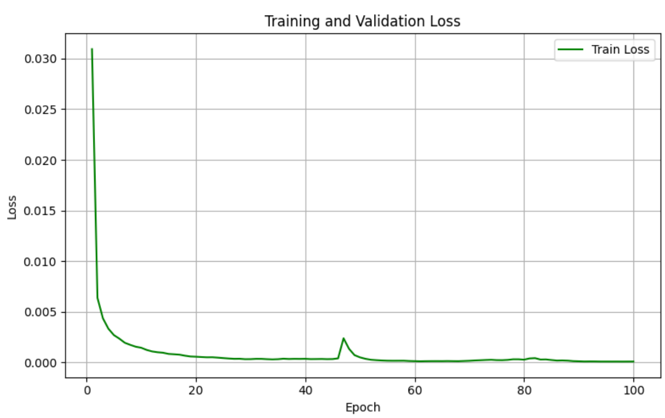

Figure 7 Lane U-net Loss vs. Epochs

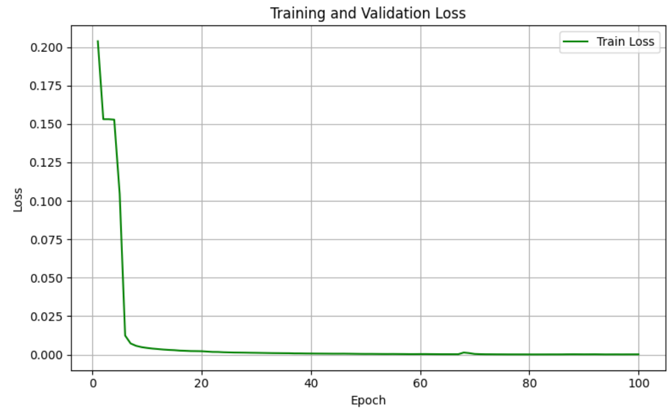

Figure 8 Lane Residual U-net Loss vs. Epochs

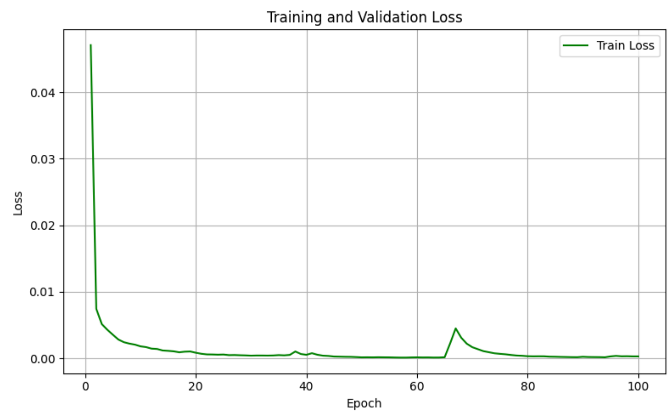

Figure 9 Lane Attention U-net Loss vs. Epochs

Table 1 Lane Model Performance

Metrics (%)	**Models**		
	U-net	***Res U-net***	***Att U-net***
IoU ↑	86.19	85.98	**86.5**
DSC ↑	92.58	92.46	**92.76**
F1 ↑	92.58	92.46	**92.76**

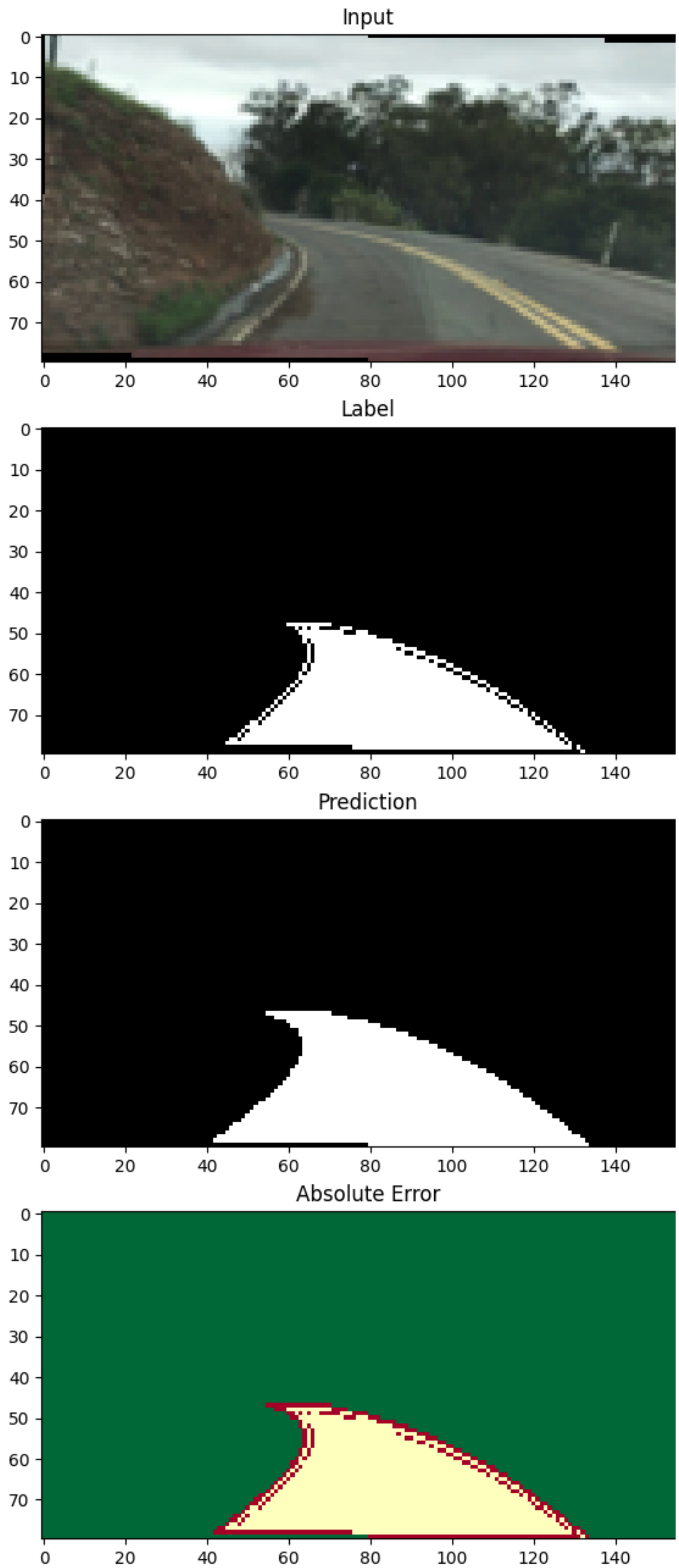

Figure 10 Lane U-net Results

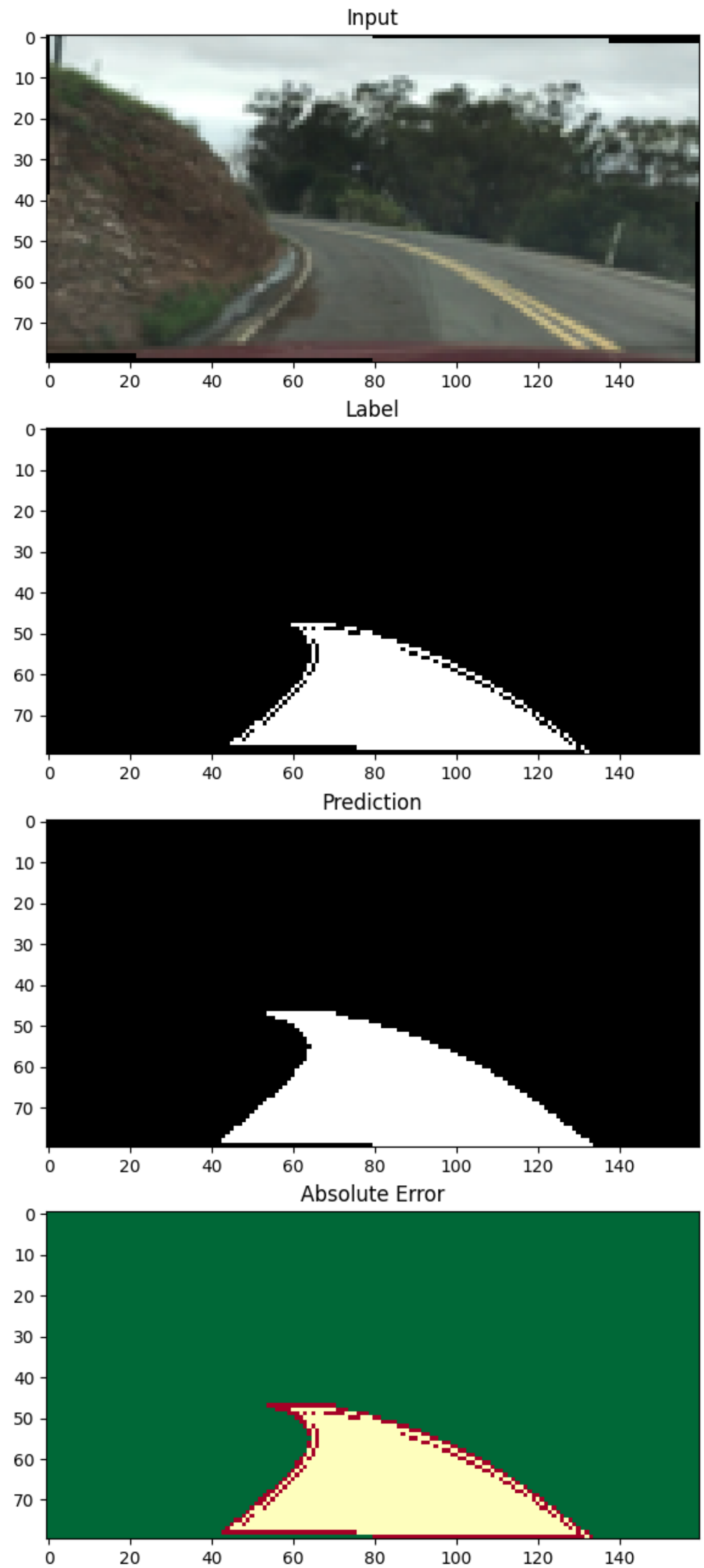

Figure 11 Lane Residual U-net Results

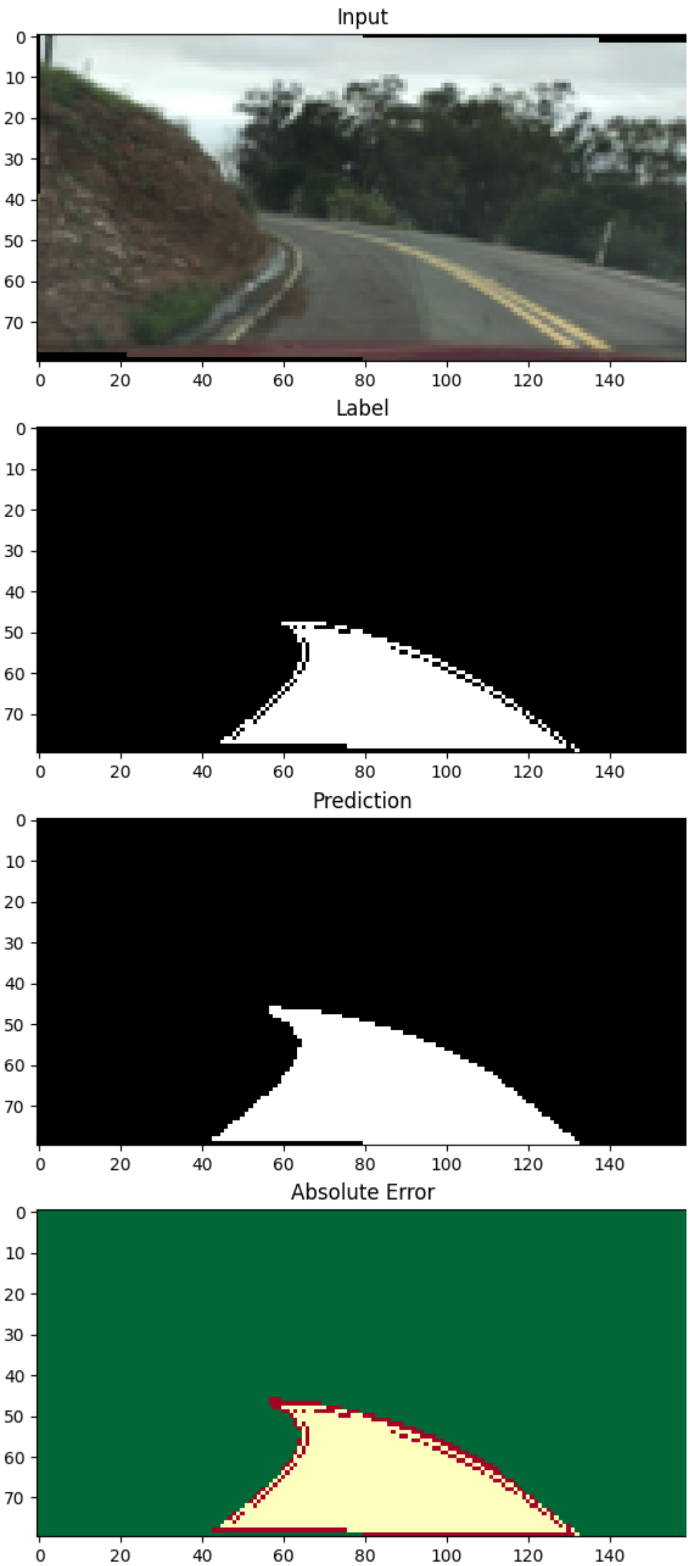

Figure 12 Lane Attention U-net Results

B. Road Segmentation Scenario

This scenario used roads as opposed to lanes. A wider surface area means visual differences are more pronounced. U-net had a few errors near edges, residual U-net had many large false positives, while the attention U-net correctly identified lane markings best. The loss starts much higher for worse models with fewer spikes.

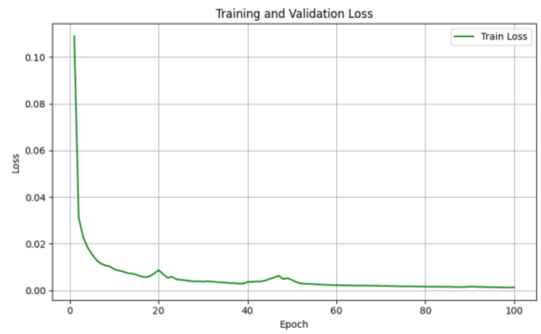

Figure 13 Road U-net Loss

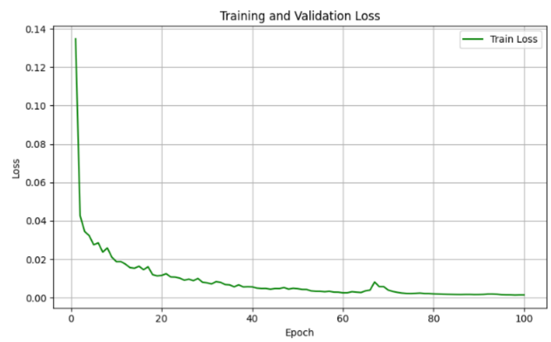

Figure 14 Road Residual U-net Loss

Figure 15 Road Attention U-net Loss

Table 2 Road Model Performance

Metrics (%)	**Models (Road)**		
	U-net	***Res U-net***	***Att U-net***
IoU	97.06	95.17	97.2
DSC	98.51	97.52	98.58
F1	98.51	97.52	98.58

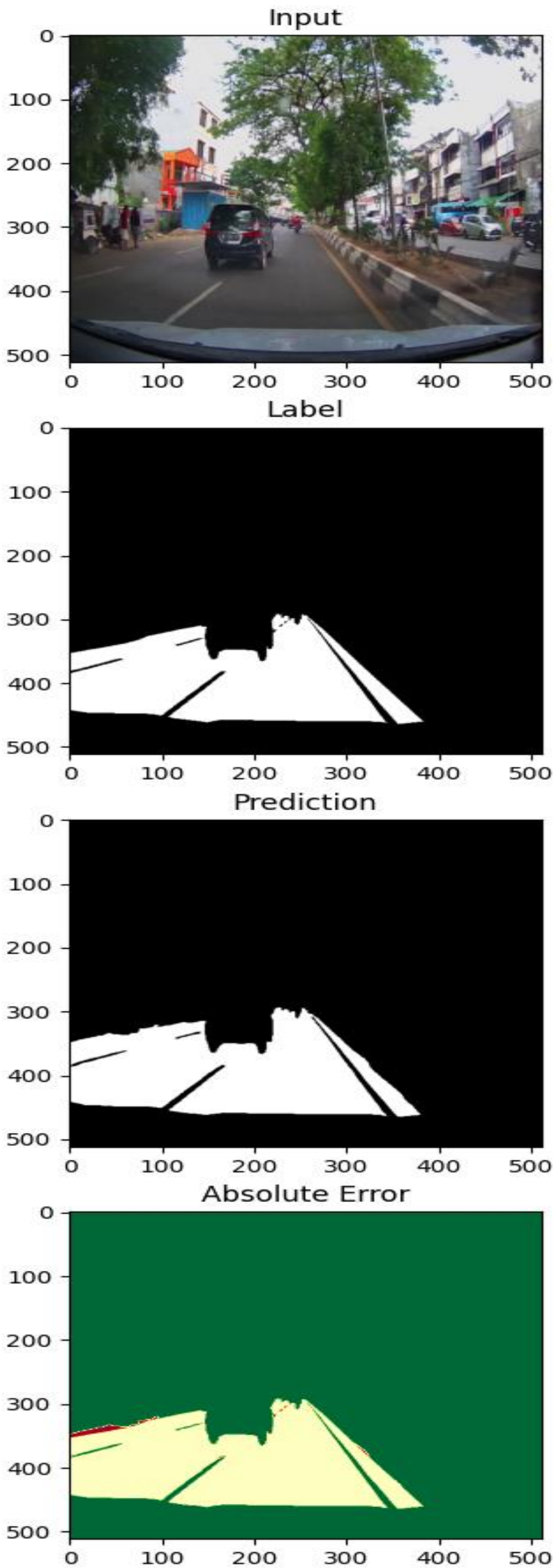

Figure 16 Road U-net Results

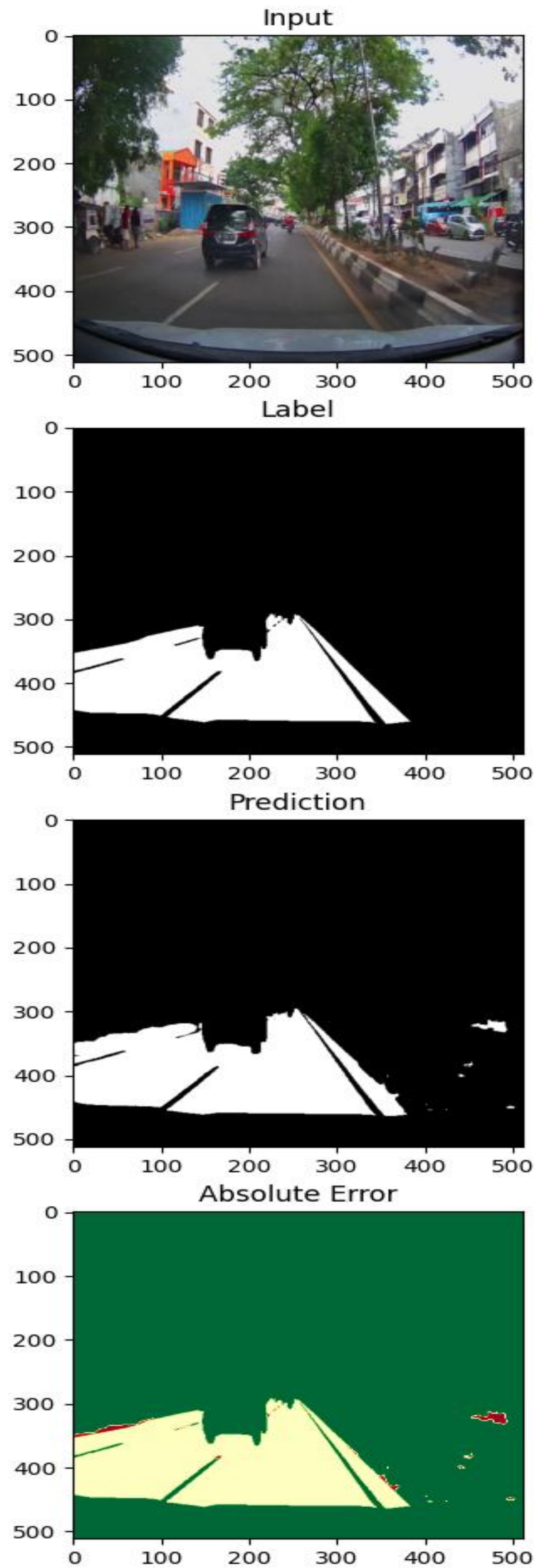

Figure 17 Road Residual U-net Results

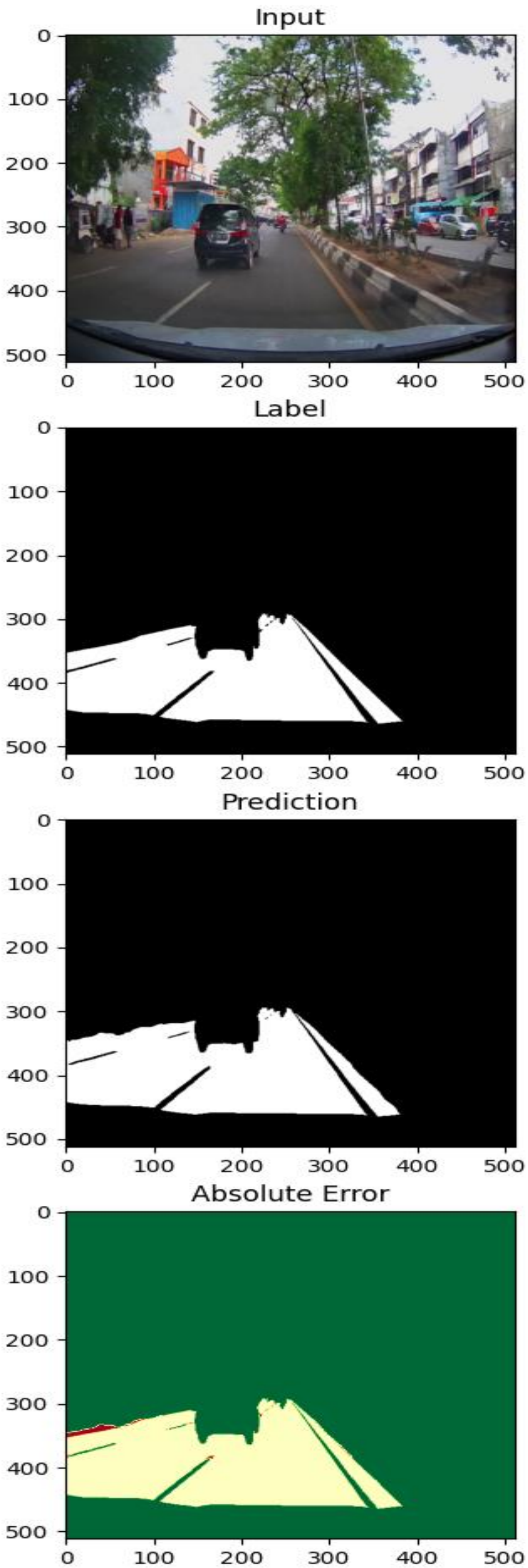

Figure 18 Road Attention U-net Results

VI. DISCUSSION

To summarize both scenarios, the residual U-net performed the worst on all fronts. It had the worst metrics, absolute visual error, and roughest loss. With the residual U-net out of the way, the main comparison takes place between the traditional U-net and the attention U-net. Although hard to distinguish, the lane error displays a few points where the attention U-net was better. While in the road scenario this is amplified. With the loss on the road scenario, it is hard to distinguish which model is better due to the scale being wider for the U-net than attention U-net. What remains is the metrics. The IoU, DSC, and F1 scores are all better on attention U-net.

VII. CONCLUSION

In conclusion, in an effort to improve consumer trust and reduce autonomous vehicle accidents, this study compared state of the art lane segmentation architectures. Based on the results, the attention U-net was determined to be the best model in both scenarios achieving the best IoU, DSC, and F1 scores in both lane segmentation and road segmentation. Despite autonomous driving still having a long way to go, this study served as a step in that direction. Future work could include training the attention U-net model further with larger datasets. In addition, the masks could be used as features for a fast-reasoning model to process while driving.

ACKNOWLEDGMENT

Sanghoon Lee is credited as the coordinating professor for this study, his insight was a great contribution.

References

[1] *SAE Level 3 Autonomous Driving Technology of the ETRI*. (2019, October 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/8939765

[2] Uknown Author, Uknown Date, https://www.nhtsa.gov/laws-regulations/standing-general-order-crash-reporting#83381

[3] Park, Ie & Kim, Seoyong & Moon, Jungwook. (2025). Why do people resist AI-based autonomous cars?: Analyzing the impact of the risk perception paradigm and conditional value on public acceptance of autonomous vehicles. PLOS ONE. 20. 10.1371/journal.pone.0313143.

[4] Babić, Dario, Fiolić, Mario, Babić, Darko, Gates, Timothy, Road Markings and Their Impact on Driver Behaviour and Road Safety: A Systematic Review of Current Findings, *Journal of Advanced Transportation*, 2020, 7843743, 19 pages, 2020. **https://doi.org/10.1155/2020/7843743**

[5] *Multi-Class Lane Semantic Segmentation using Efficient Convolutional Networks*. (2019, September 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/8901686

[6] 17. Z. Teng, J.-H. Kim, and D.-J. Kang, “Real-time lane detection by using multiple cues,” in International Conference on Control, Automation and Systems, 2010

[7] Ronneberger, Olaf, Philipp Fischer, Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation." arXiv:1505.04597, 2015.

[8] LeeJunHyun (2018) Image_Segmentation [Source code]. https://github.com/LeeJunHyun/Image_Segmentation

[9] Jiangtao, W., Ruhaiyem, N.I.R., Panpan, F. (2025). A Comprehensive Review of U-Net and Its Variants: Advances and Applications in Medical Image Segmentation. arXiv preprint arXiv:2502.06895.

[10] Zhang, Z., Liu, Q., Wang, Y. (2017). Road Extraction by Deep Residual U-Net. arXiv preprint arXiv:1711.10684.

[11] Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D. (2018). Attention U-Net: Learning Where to Look for the Pancreas. arXiv preprint arXiv:1804.03999.

[12] Milesial (2017) Pytorch-UNet [Scource code]. https://github.com/milesial/Pytorch-UNet

[13] Mamun, Abdullah & Em, Poh Ping & Hossen,. (2021). Lane marking detection using simple encode decode deep learning technique: SegNet. International Journal of Electrical and Computer Engineering. 11. 3032-3039. 10.11591/ijece.v11i4.pp3032-3039.

[14] Azkalani, Nublan. (2024) Semantic Segmentation Makassar (IDN) Dataset. Kaggle. https://www.kaggle.com/datasets/nublanazqalani/semantic-segmentation-makassaridn-road-dataset

[15] Padilla, Rafael & Netto, Sergio & da Silva, Eduardo. (2020). A Survey on Performance Metrics for Object-Detection Algorithms. 10.1109/IWSSIP48289.2020.

[16] V. Raina *et al*., "Tackling Bias in the Dice Similarity Coefficient: Introducing NDSC for White Matter Lesion Segmentation," *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, Cartagena, Colombia, 2023, pp. 1-5, doi: 10.1109/ISBI53787.2023.10230755

EOD.