

ABSTRACT

Understanding road surface damage is crucial in autonomous driving systems to ensure passenger safety. Several semantic segmentation models have been implemented, but they typically rely on either convolutional neural network or transformer architectures. Some of these models even compromise segmentation accuracy in favor of improving inference speed. This opens up an opportunity to enhance model performance by applying a different approach to the model architecture. This study aims to explore the potential of combining both architectures by implementing and analyzing the performance of LETNet architecture on proposed model to produce predictions as quickly and accurately as possible. In this study, the model will be trained using two different datasets and compared to baseline models. The outcome shows that the combination of both architectures in LETNet model has a potential by outperforming baseline model's accuracy on majority of test configurations. On its best configuration, the model achieves 60.01% mIoU on PotholeMix and 54.83% on CQUPM dataset. The model still runs on average inference speed of 15 FPS which leaves a room for further improvements.

Key Words: *road damage, semantic segmentation, convolutional neural network, transformer*