# DUAL SUPER RESOLUTION LEARNING FOR SEMANTIC SEGMENTATION

Pulkit Garg (P21CS011), Misaal Khan (D20ID004), Satwik Srivastava (M21MA006)

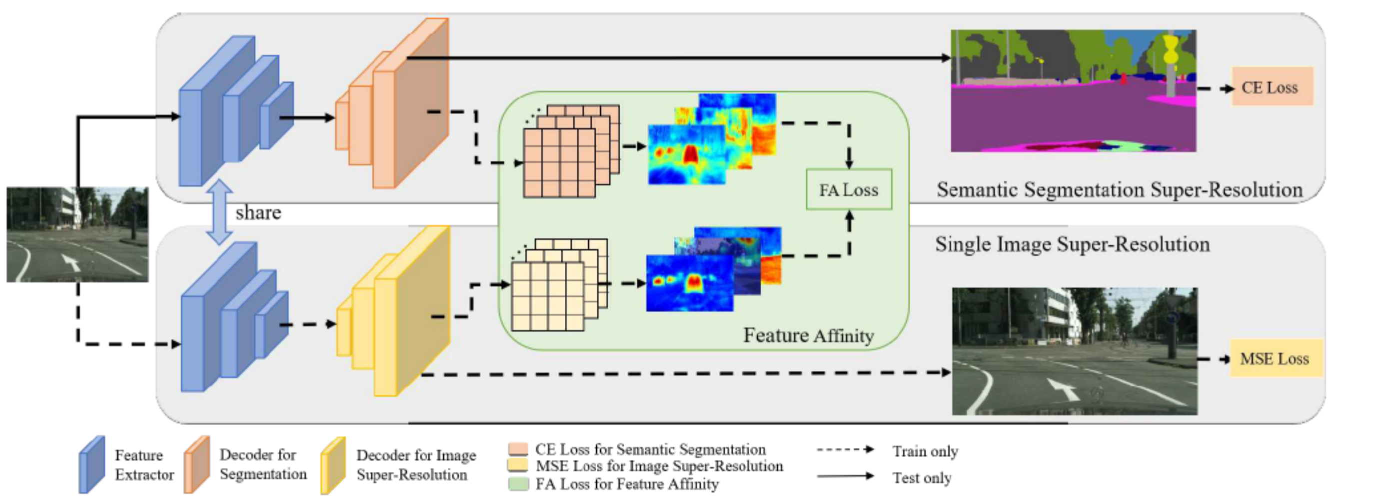
Indian Institute of Technology, Jodhpur

## INTRODUCTION

In today's world, many applications utilize semantic segmentation for various tasks, and with the increase in usage of smartphones and other devices, there is a need for fast and automated methods. However, the current state-of-the-art techniques, are computationally expensive and don't work in real-time. This is due to the fact that there is a tradeoff between accuracy and computation time/ complexity. Provided a high-resolution image, the semantic segmentation tasks can be performed with high accuracy but it results in more parameters and increased execution time.

To overcome this, the concept of super-resolution is utilized along with the existing semantic segmentation methodology. Since high-resolution deep features are crucial for achieving high accuracy, a high-resolution image is reconstructed from a low-resolution image. This step provides the architecture with a high-resolution image without explicitly passing it as an input. As a result, high performance can be attained using semantic segmentation without increasing the computational requirements.

## METHODOLOGY




A brief description of the used architecture is as follows:

- The technique utilizes an encoder-decoder based architecture for semantic segmentation, given an input image the model generates the semantic segmented output where each pixel is assigned a class label.

$$L_{ce} = \frac{1}{N} \Sigma_{i=1}^{N} - y_i \, log(p_i)$$

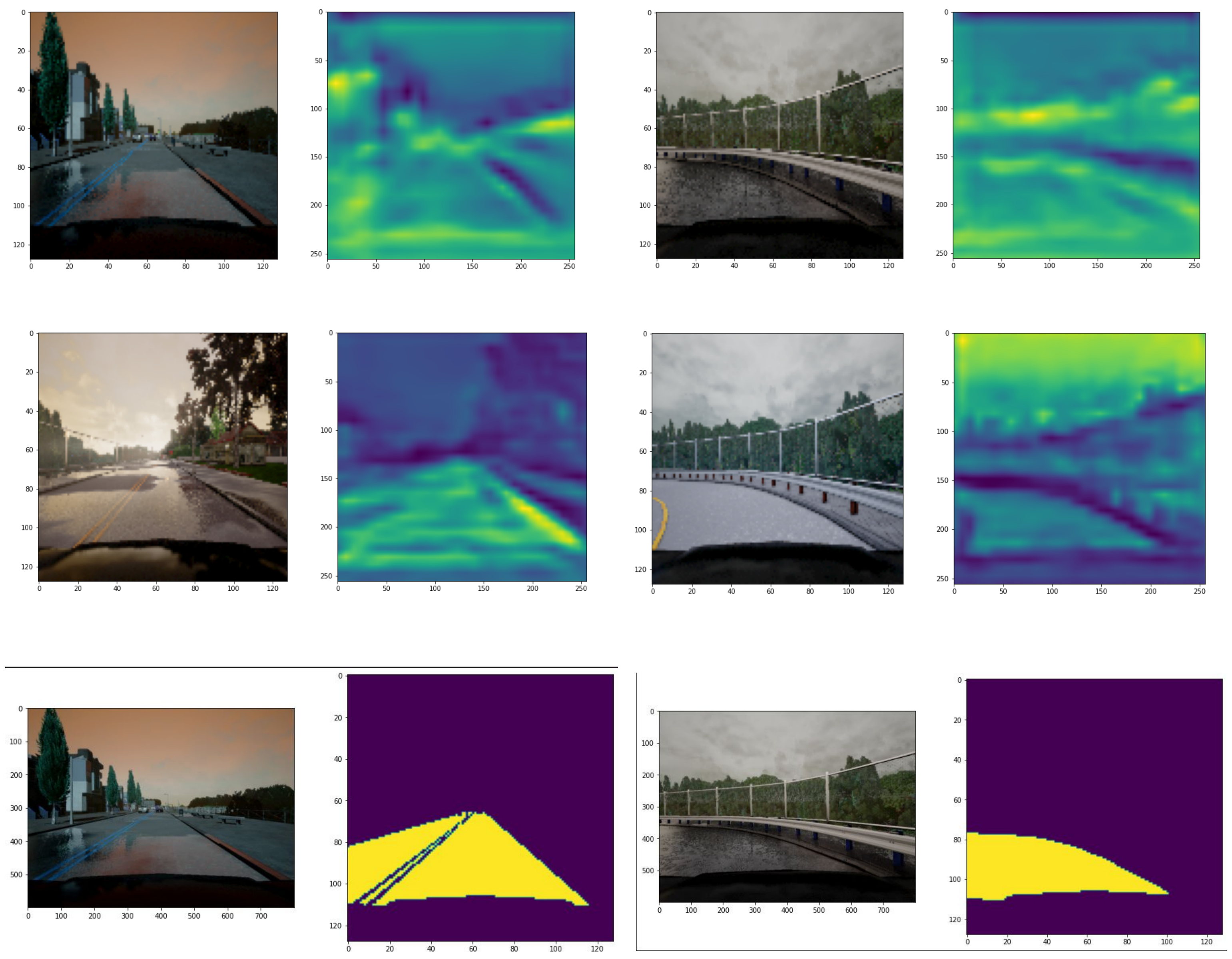
- The technique also develops a super-resolution architecture using encoder-decoder architecture. However, towards the end of the decoder, an extra up sampling layer is added to super resolve the decoded output.

$$L_{mse} = \frac{1}{N} \sum_{i=1}^{N} ||SISR(X_i) - Y_i||^2$$

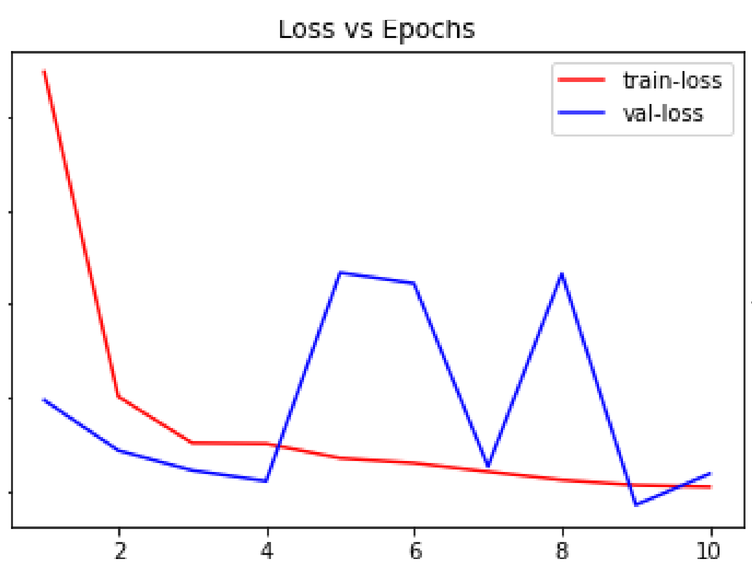
- FA aims to learn the distance of similarity matrix between SISR and SSSR branch. Specifically, for a W' × H' × C, feature map F, we formulate the relationship between every two pixels and Sij is relationship between the ith and the jth pixels

$$L = L_{ce} + w_1 L_{mse} + w_2 L_{fa} \qquad L_{fa} = \frac{1}{W'^2H'^2} \sum_{i=1}^{W'H'} \sum_{j=1}^{W'H'} ||S_{ij}^{seg} - S_{ij}^{sr}||_q$$

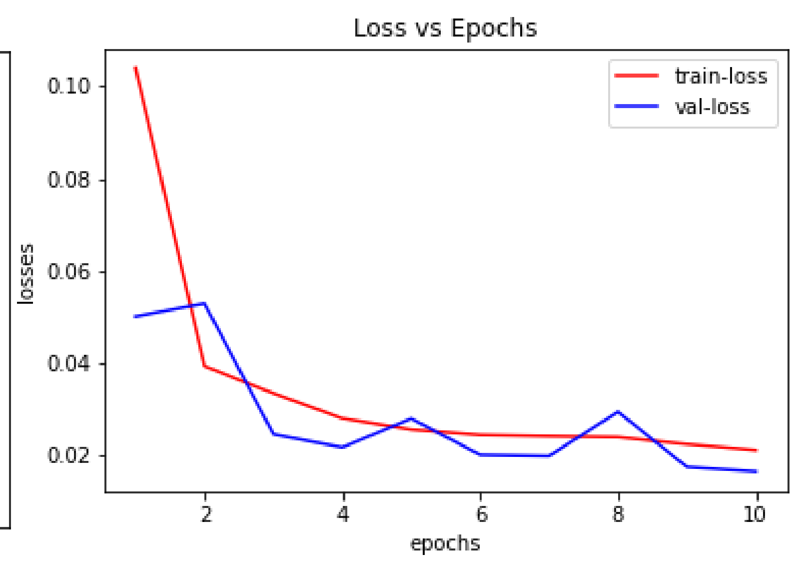
## RESULTS



Model1




Model 2




## CONCLUSIONS & OBSERVATIONS

A dual super-resolution learning framework is implemented for semantic segmentation. The semantic segmentation super-resolution branch helps learn higher-resolution representations for dense label prediction, the single image super-resolution branch can recover detailed structure information, and the feature affinity module is introduced to enhance the high-resolution representations of semantic segmentation through the detailed structural information

## REFERENCES

Wang, L., Li, D., Zhu, Y., Tian, L. and Shan, Y., 2020. Dual super-resolution learning for semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 3774-3783).