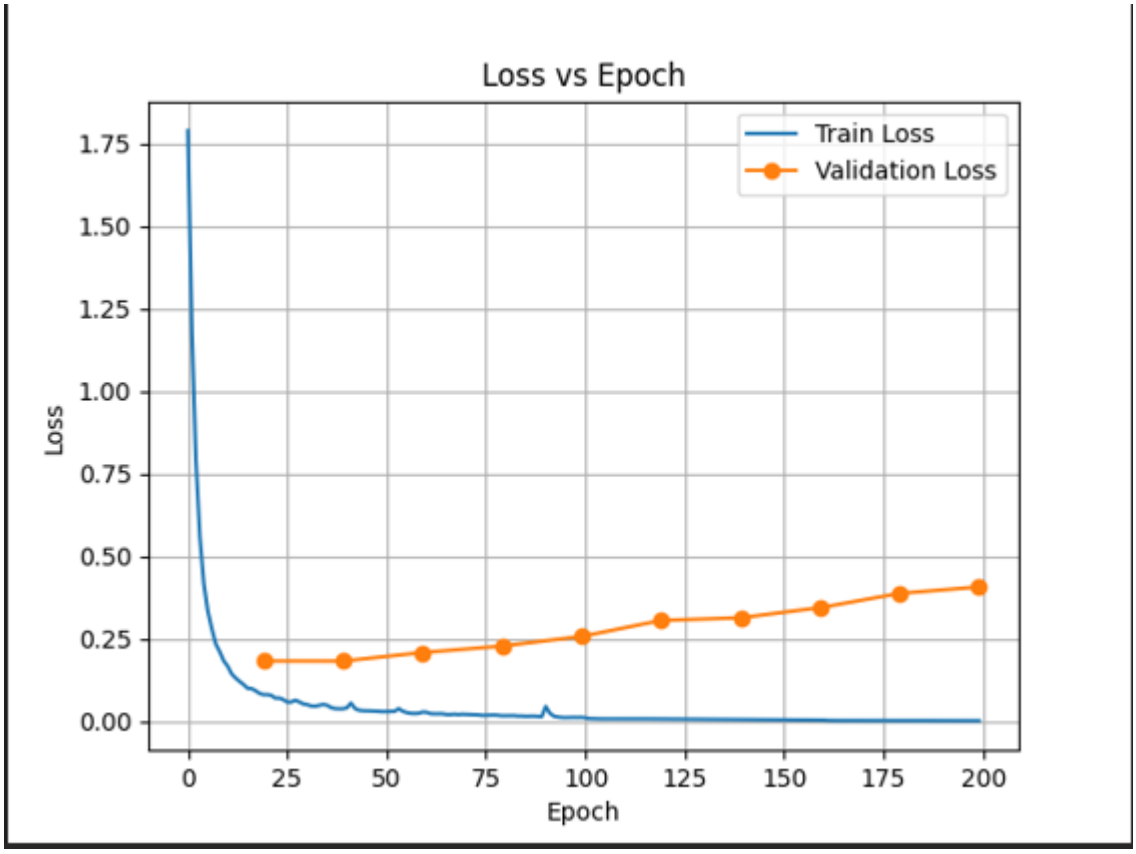
DINOv2 with U-Net Segmentation Head

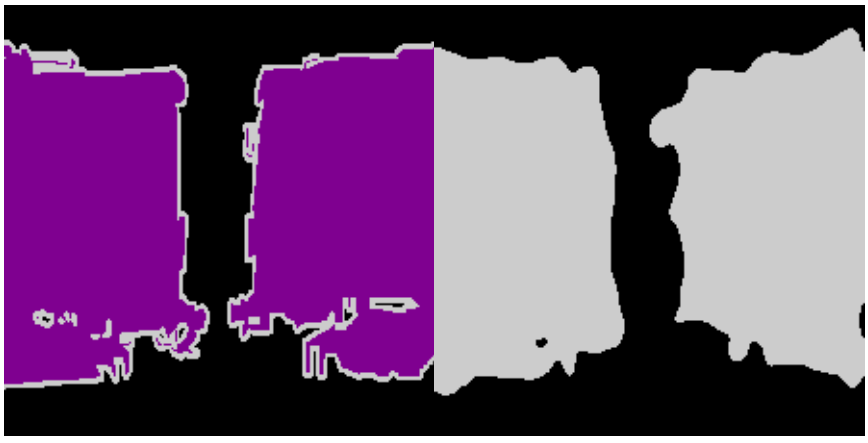
Observation:

1.) Tried first without fine tuning the DINOv2 weights. We see some kind of Overfitting (could be due to more numbers of epoch). With 100 epoch the validation loss was ok. So, tried with Reduced epoch with other architecture.

With 200 epochs. Case of OVERFITTING



2.) Few results on the validation set. Left is GT and Right is Pred.

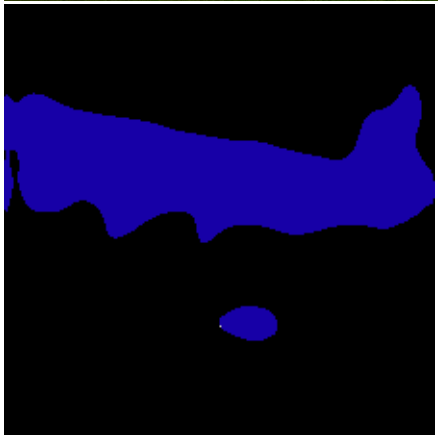


3.) Found fine-tuning DINO weights was even more OVERFITTING. Guessing that the data is to less for this model to get fine-tuned. We can use this setting here to turn fine-tuning on and off.

```python
class DinoSegModel(nn.Module):
    def __init__(self, freeze_dino=True, num_classes=21):
        super().__init__()
        self.dino = Dinov2Model.from_pretrained("facebook/dinov2-large")
        self.decoder = UNetDecoder(in_channels=1024, num_classes=num_classes)

        if freeze_dino:
            for param in self.dino.parameters():
                param.requires_grad = False
```

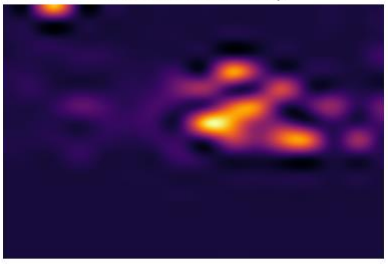
3.) Result on Test Data (Not from the train or validation set).



4.) Attention Visualization.

Original Image

DINOv2 Attention Map



Segmentation Prediction