Experimentation with Vision Transformers

Q2)

https://colab.research.google.com/drive/18dXcYQMrnwZg6IKzyQ8f4Mite4KjKfTU?usp=sharing In this google collab notebook i used MADS dataset which is human segementation dataset to finetune the pre-trained ViT based image segmentation on 'nvidia/segformer-b0-finetuned-ade-512-512' dataset and the metrics are shown in notebook itself and here i used IoU(to check alignment) and Pixel accuracy metrics(simple one) and also showed(visualized) how predictions are wrt original images and their true masks

I trained for 3 epochs(instead of 10) as it was taking time and by the time i realised i didn't use precision, recall, F1 for metrics ,my notebook runtime was up(used gpu)