**Lab 5 - Report**

***(Note: When we are talking about loss & accuracy, we are referring to the validation values)***

**Task 1**

**Answer)** No the performance is not consistent across all fold. We could observe a performance fluctuation between each fold, and this has to be accounted for when evaluating the performance. In order to deal with this one could perform averaging of the results from each fold which is standard practice when doing k-fold cross validation.

One thing that we also observed was that when increasing the number of folds to above 5 the GPU started producing memory issue errors. It ran out of memory. This occurred regardless if deleting the model and clearing the session in-between each fold.

We used data augmentation and loaded the data batch by batch with 150 epochs per fold.

**Task 2**

**Answer)** Yes, we could observe a discrepancy between the loss functions and the evaluation metrics. What we could observe was worse performance when adding weight maps than without weight maps. We had an 80/20 split, used data augmentation, and trained for 150 epochs. This was unexpected as we predicted that the weight map would improve the segmentation results by helping to add focus on the edges.

**Lorem ipsum …**

**Task 3**

**Answer) Lorem ipsum …**

**Final observations**

**Answer)** According to our results, ordinary data augmentation with a simple U-net, yields the best accuracy. We expected both the weighted dice loss and auto-context to perform better but this was not the case. Because of major memory issues, we could not perform data augmentation on auto-context with all training images. Which of course decreased the performance of the auto-context network since it overfitted to the training data. It was not possible to observe an increase in performance between auto-context steps which would be expected of the algorithm.