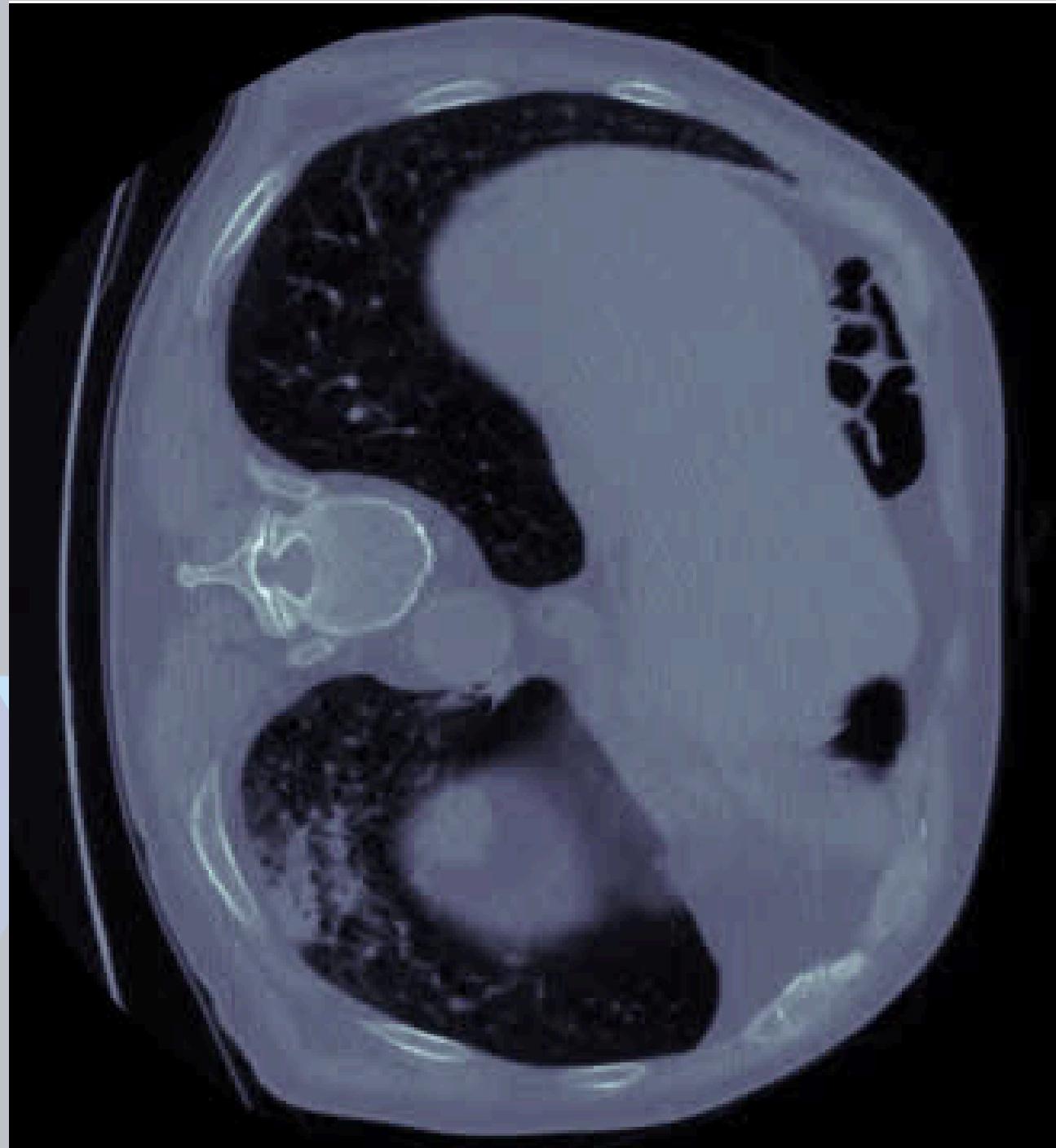


ENERGY EFFICIENT LUNG TUMOR SEGMENTATION

INTRODUCTION



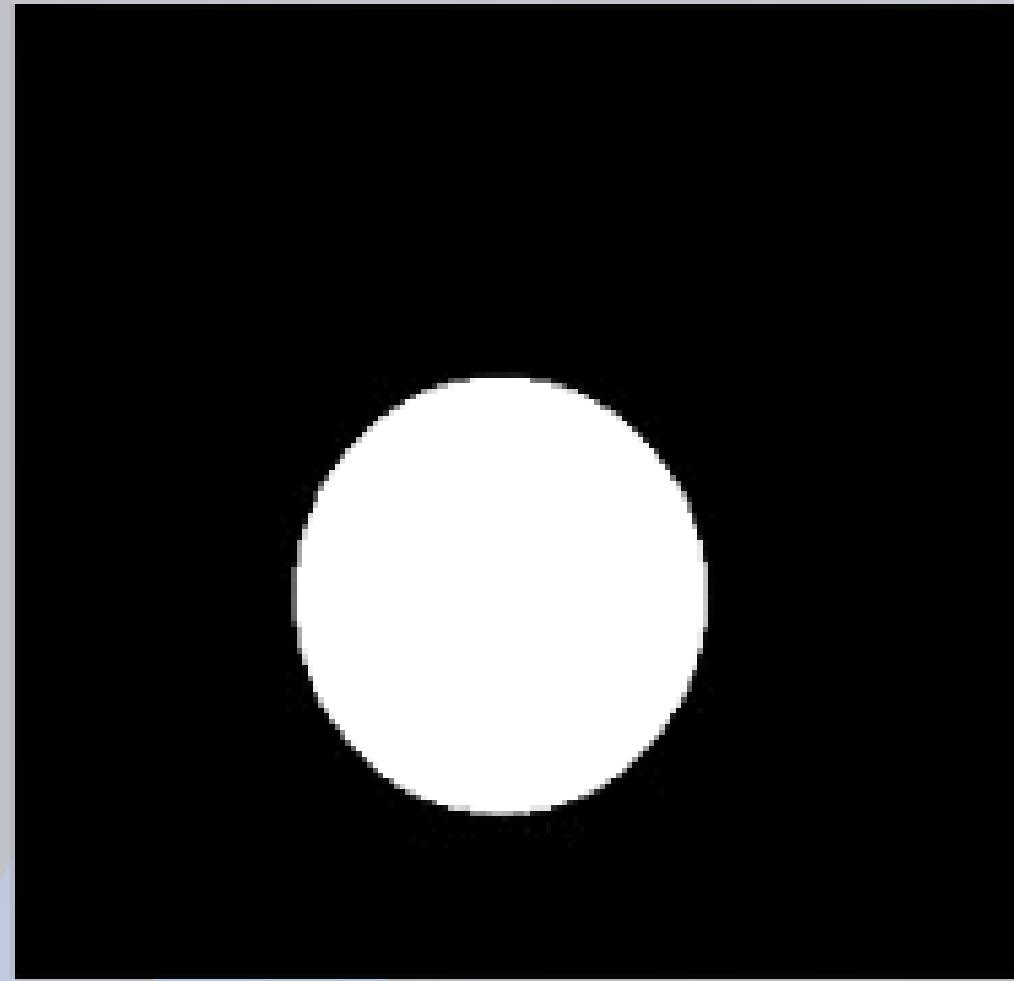
Fundamental Details of the problem

- Target: Lung and tumours
- Modality: CT
- Size: 96 3D volumes (64 Training + 32 Testing)
- Source: The Cancer Imaging Archive
- Challenge: Segmentation of a small target (cancer) in a large image

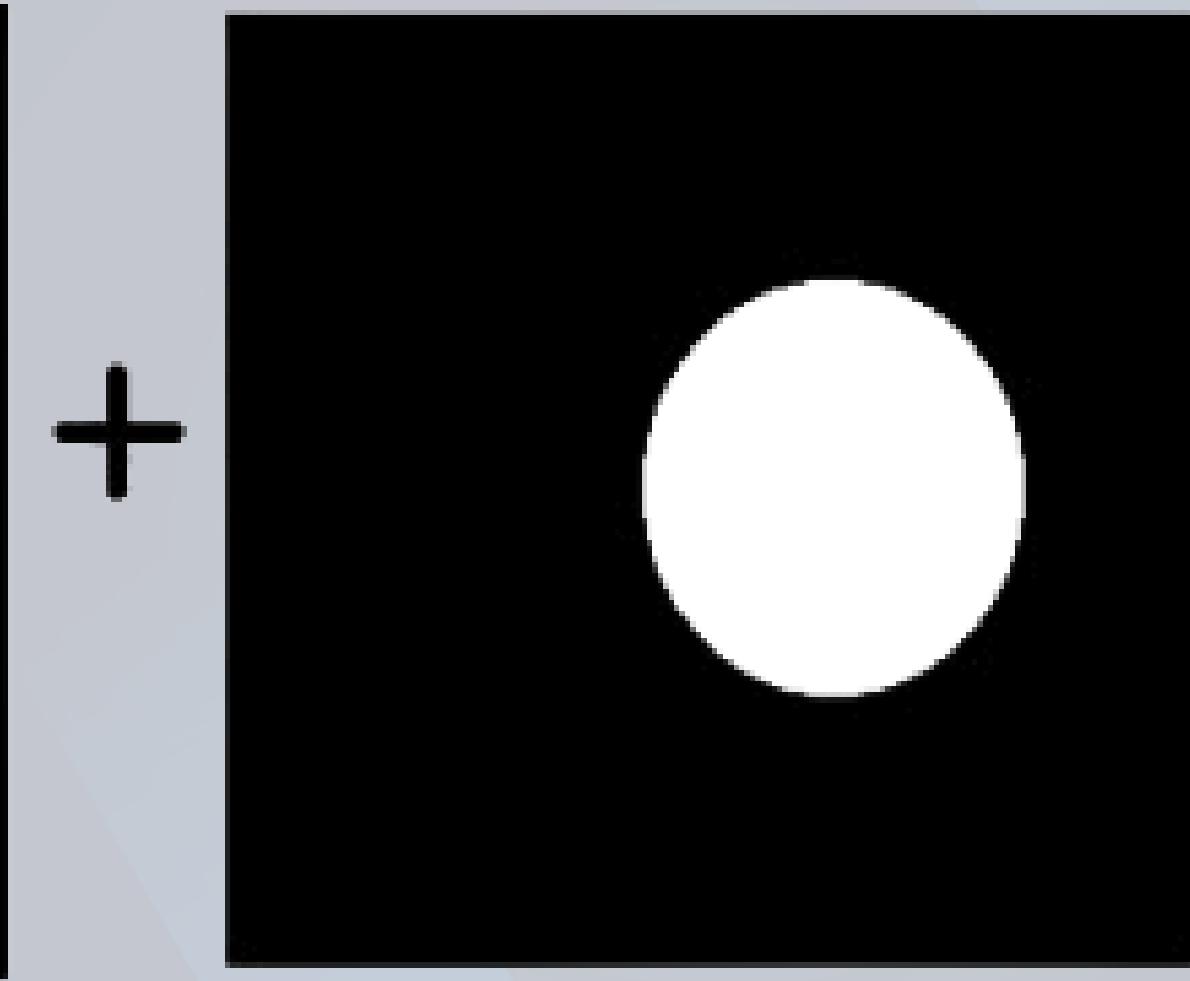
Issues to be addressed

- What is the differentiating factor between our work and existing research?
- How to develop a computationally efficient model while not sacrificing performance?

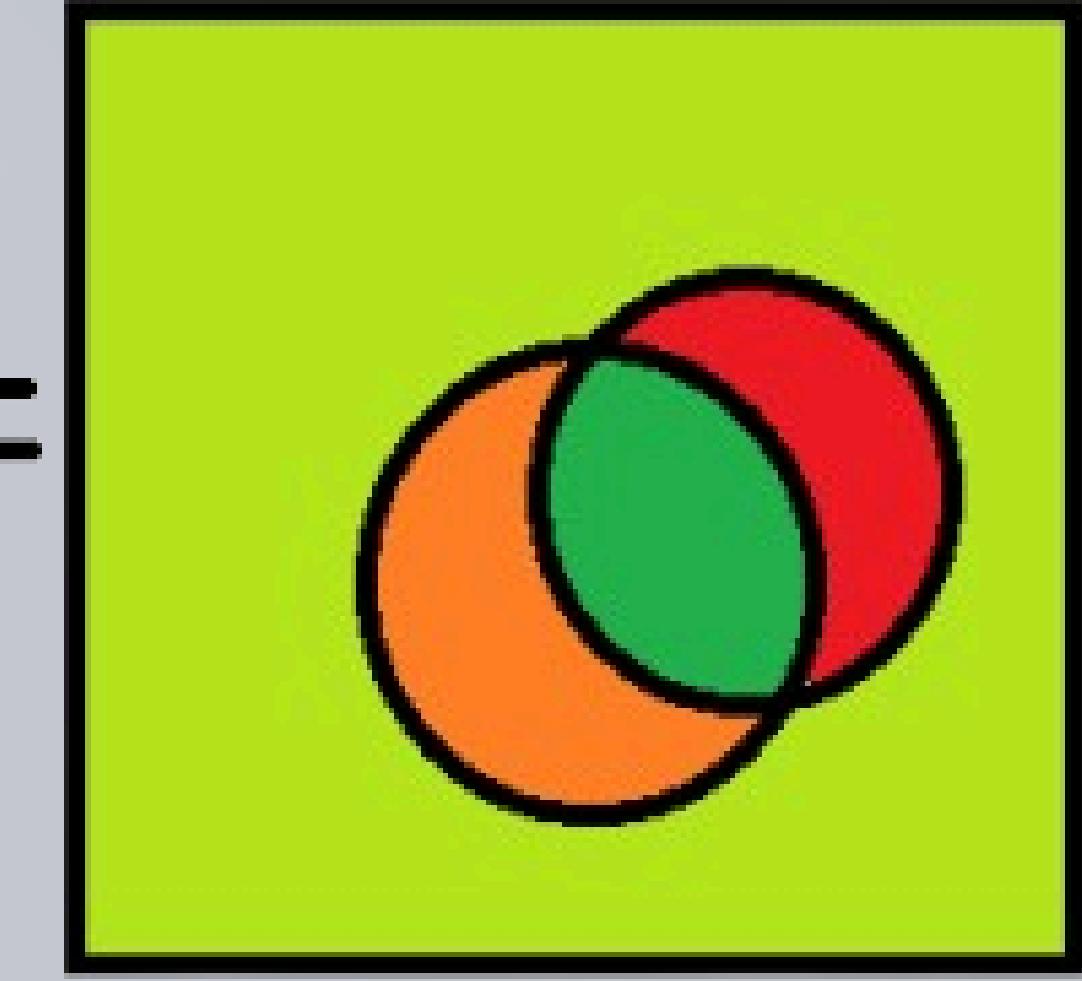
CALCULATION OF METRICS



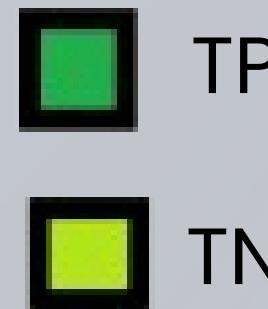
Prediction



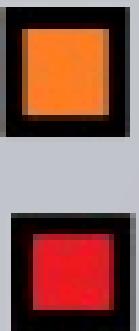
True Label



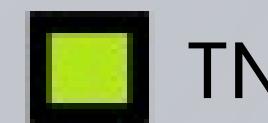
NOTE: The Dice Score is essentially just F1-Score for segmentation



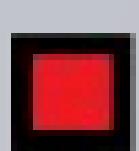
TP



FP



TN



FN

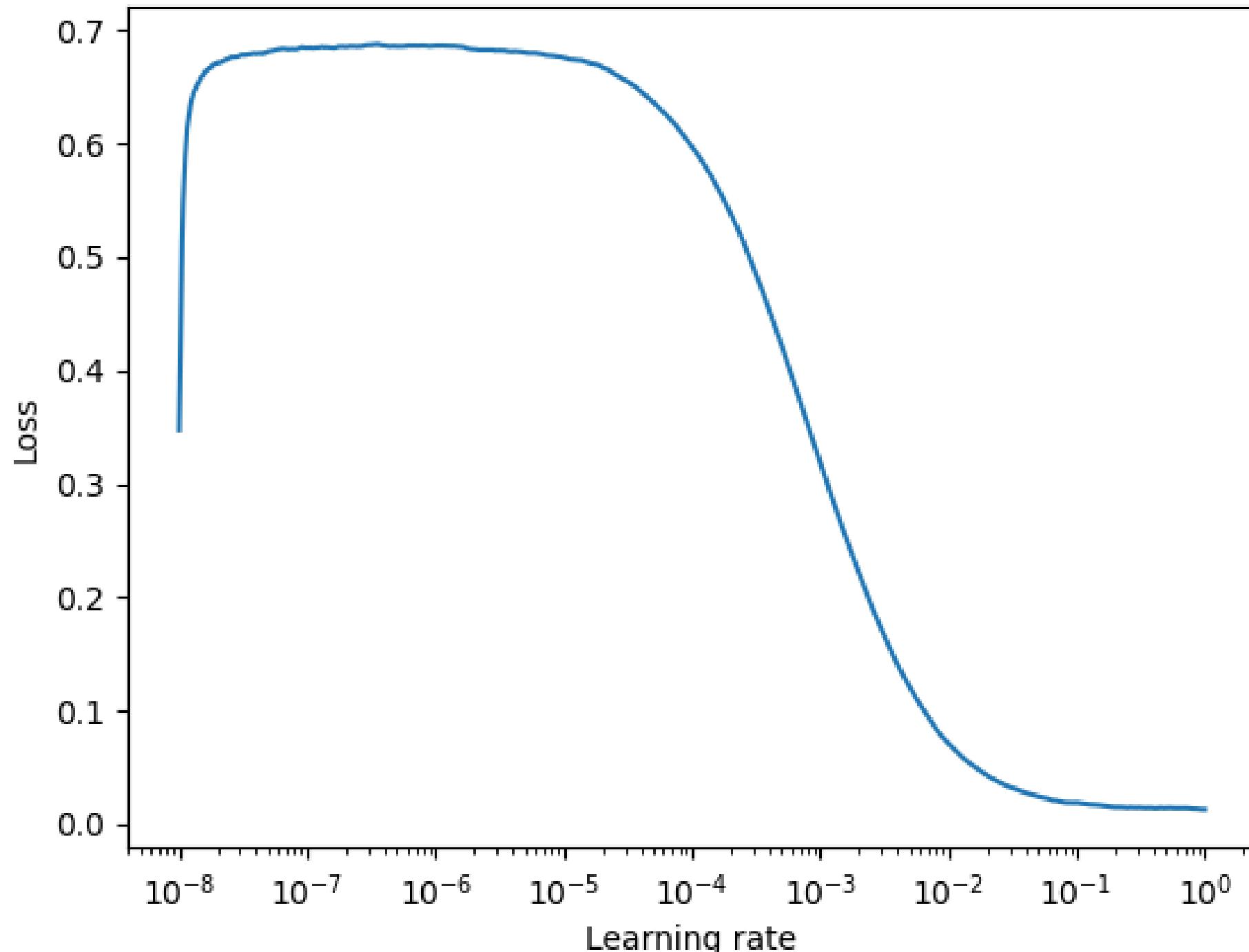
PROBLEM

- Data mismatch between training data and LIDC data, i.e. pixel values had differing ranges.
- Finding correct hyperparameters for focal loss even though current research literature recommends $\alpha=0.3$ and $\gamma=3.5$.
- Data augmentation techniques optimal for the problem.
- Model overfitting while using focal loss and oversampling.
- Exploding gradients with focal loss even with smaller learning rates.

SOLUTION

- Clip the values into the valid range of 3071 and -1024.
- Adaptive focal loss that calculates α per batch during run time while γ is kept constant.
- Research literature suggests CLAHE, noise addition, rotation, flipping, shearing, cropping - Addition of any more techniques hampers model performance.
- Without oversampling, the model learns nothing, use a different loss function instead of focal loss.

LEARNING RATE TUNER



- Used the in-built learning rate tuner of PyTorch Lightning.
- Num_steps=1000
- Low loss value for learning rate of $1e-8$ indicates its too small.
- The loss plateau until $5e-5$ indicates the learning rate is not big enough.
- The minima at the bottom is ideal is using SGD as an optimizer.
- Momentum-based optimizers like Adam require a learning rate on the middle or just above of the descent of the curve

EXPERIMENTATION RESULTS

Focal loss & it's combinations

- Increased data augmentation techniques, higher dropout probability and inclusion of weight decay with a higher learning rate - it didn't prevent overfitting and any harsher regularization caused underfitting
- Any combination of α and γ caused overfitting for any model tested

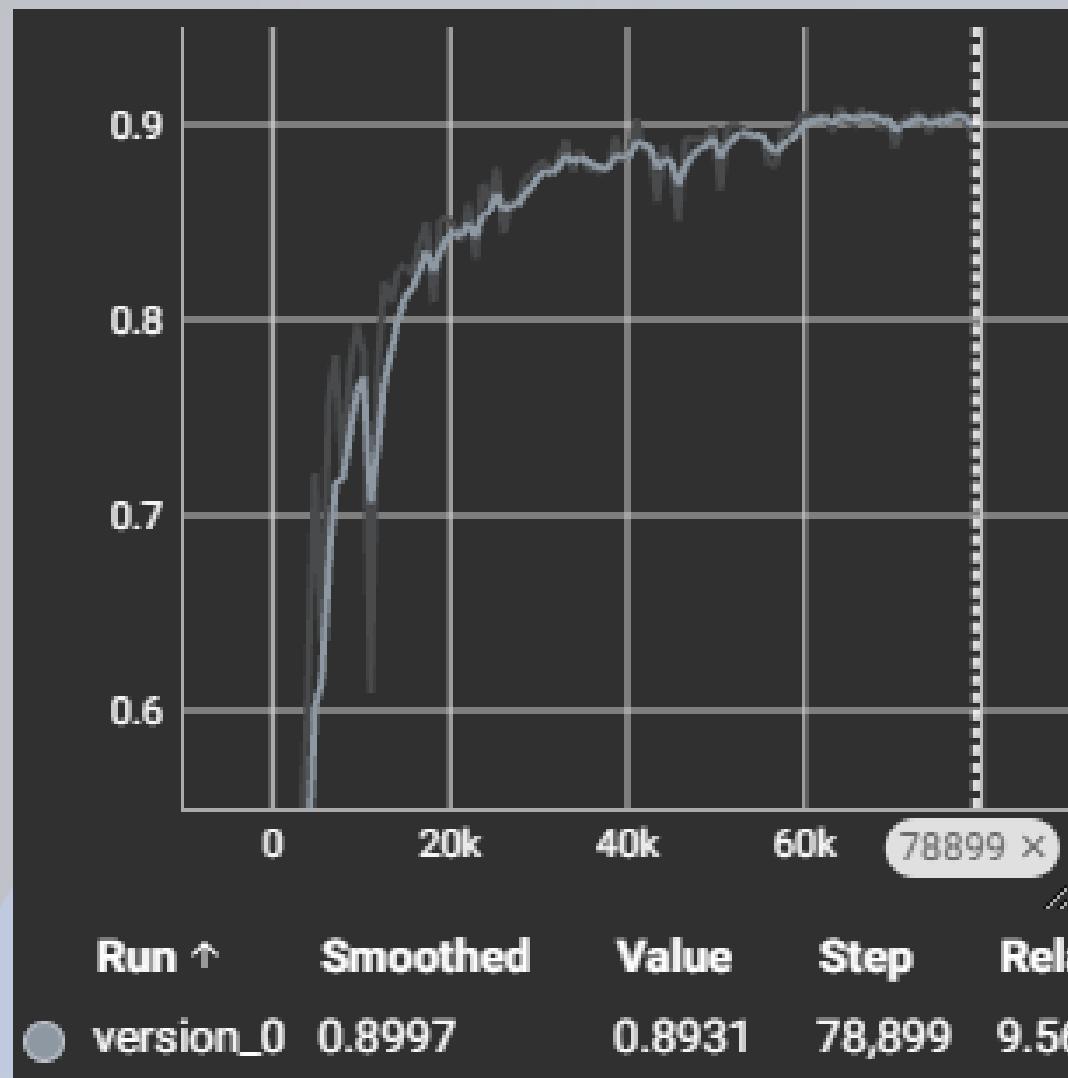
Binary Cross Entropy Loss

- Significantly improved model performance across all metrics for all models
- Doesn't cause overfitting even if oversampling is used
- Weighted Cross Entropy degrades performance with high recall and low precision

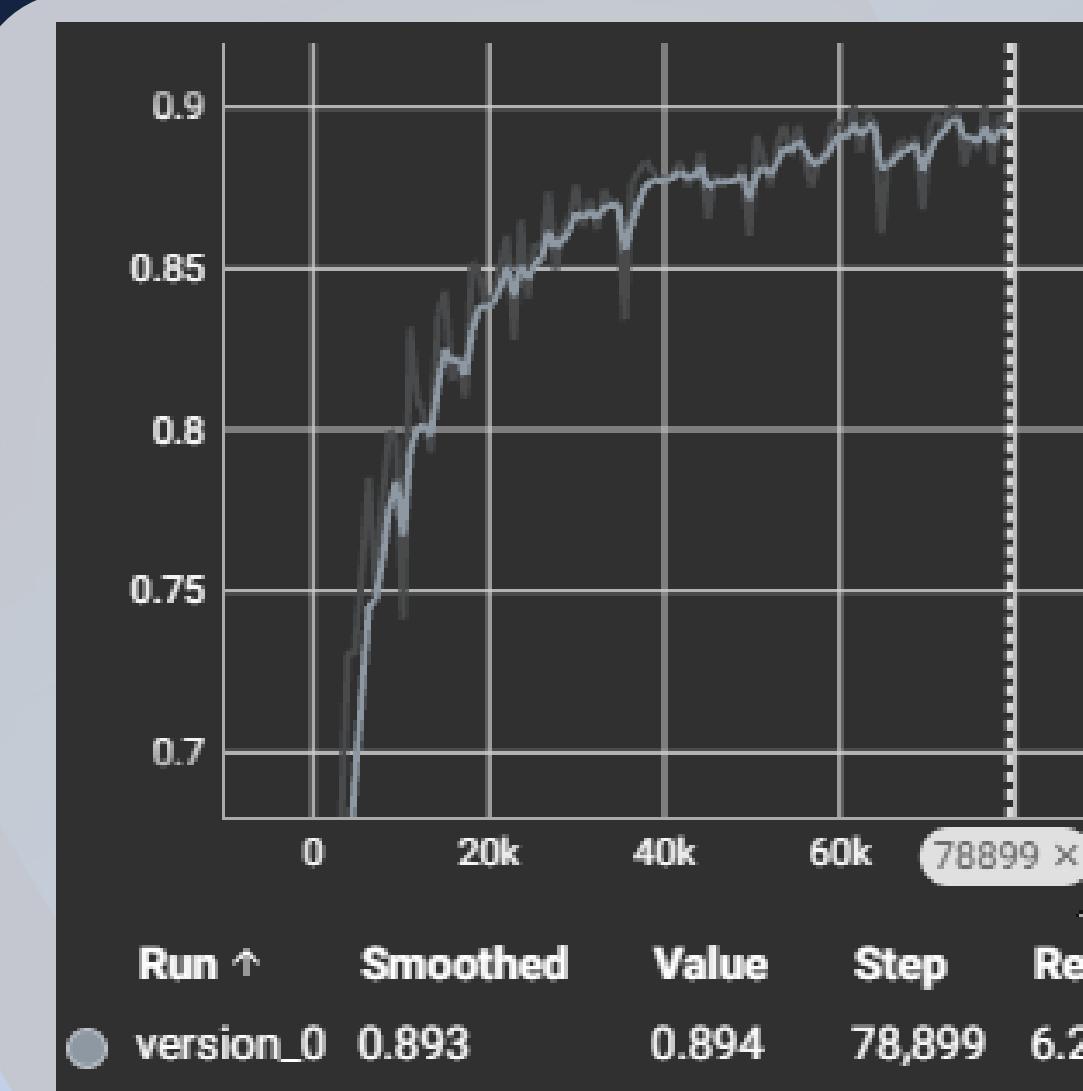
Miscellaneous results

- Dropout improves model robustness and makes ENet learn more properly
- UNet style skip connections and CBAM improved model performance

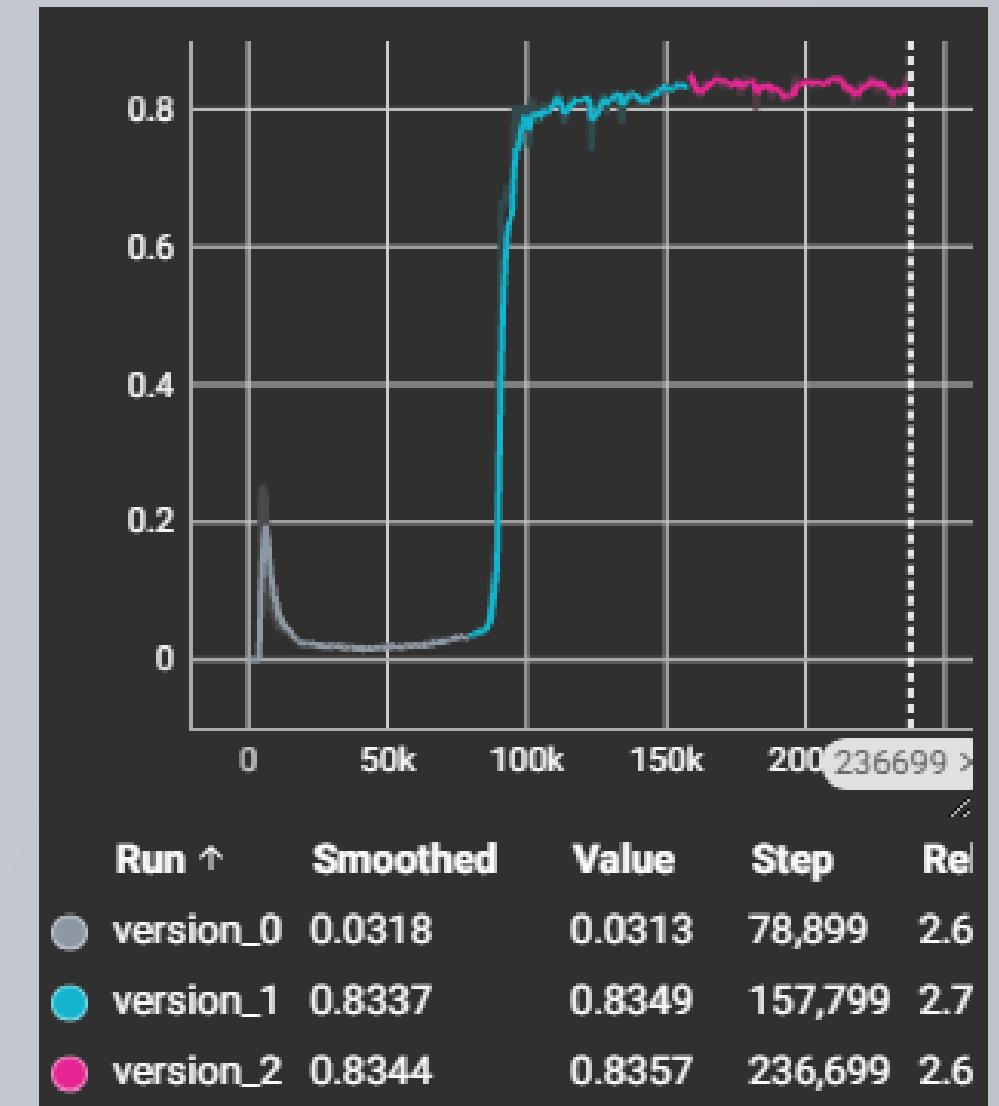
VALIDATION DICE SCORE



UNET



SEGNET

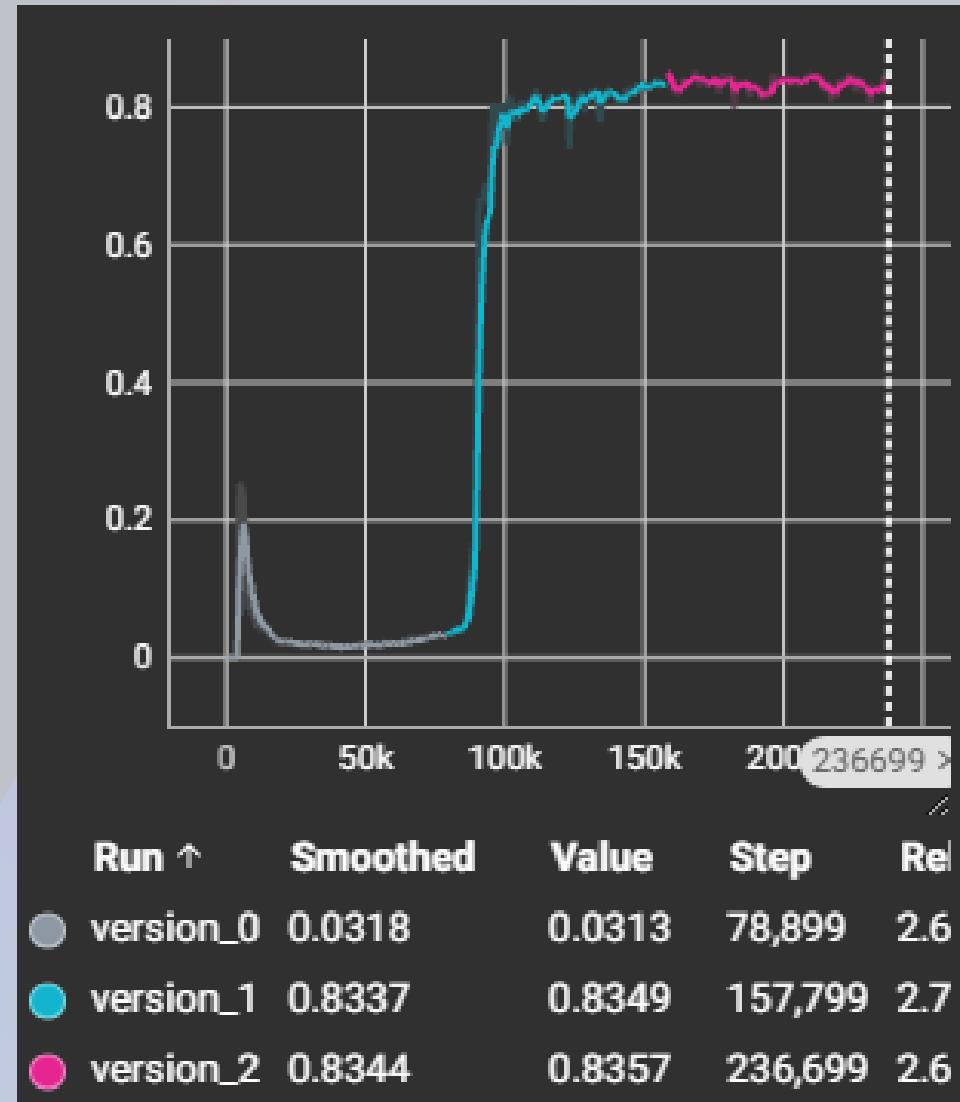


ENET

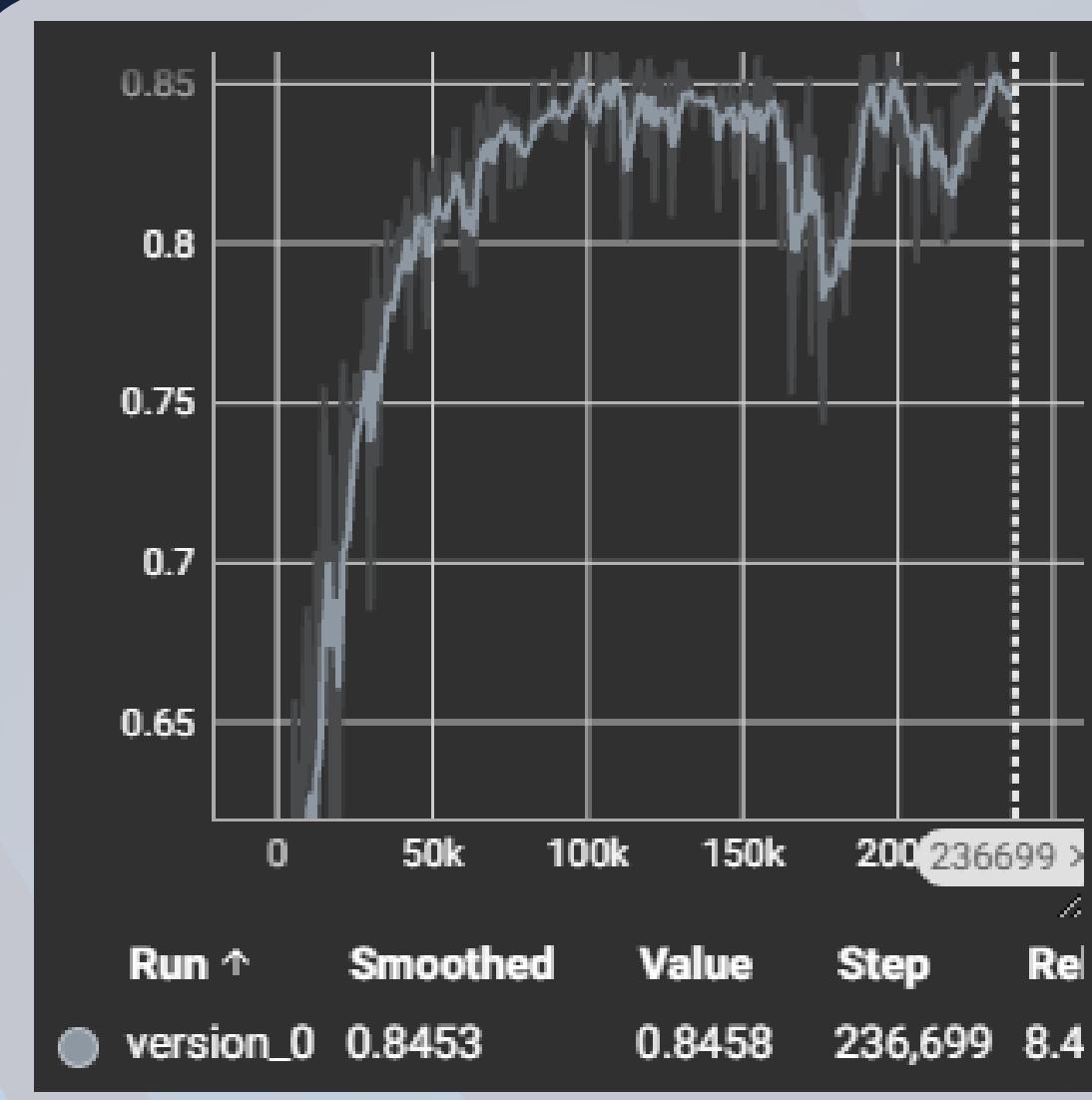
BASELINE RESULTS

Model	Total Params	GMAC	Dice Score	Precision	Recall	IoU
UNet	7,781,761	42.35 GB	0.874775	0.881955	0.867711	0.777422
SegNet	9,374,593	26.68 GB	0.860848	0.859121	0.862581	0.755691
ENet	348,852	490.92 MB	0.796514	0.811199	0.782351	0.661839

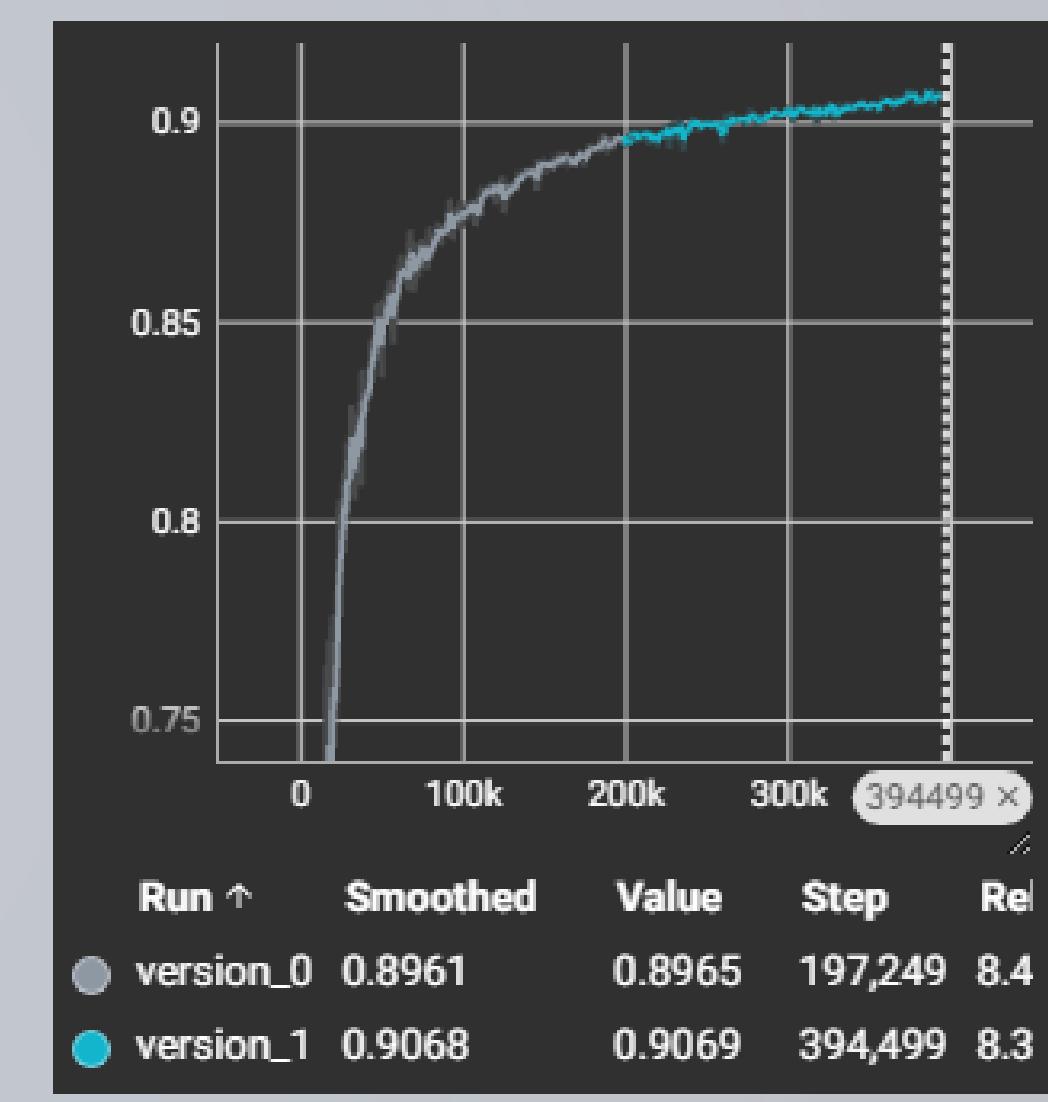
ENET VARIATIONS



BASELINE



SKIP



CBAM SKIP

Model	Total Params	GMAC (in MB)	Dice Score	Precision	Recall	IoU
Baseline	348,852	490.92	0.796514	0.811199	0.782351	0.661839
Skip	360,628	547.28	0.806953	0.806858	0.807049	0.67638
CBAM Skip (BCE)	430,052	549.04	0.861609	0.854944	0.86838	0.756866
CBAM Skip (Dice)	430,052	549.04	0.838374	0.846831	0.830083	0.721724

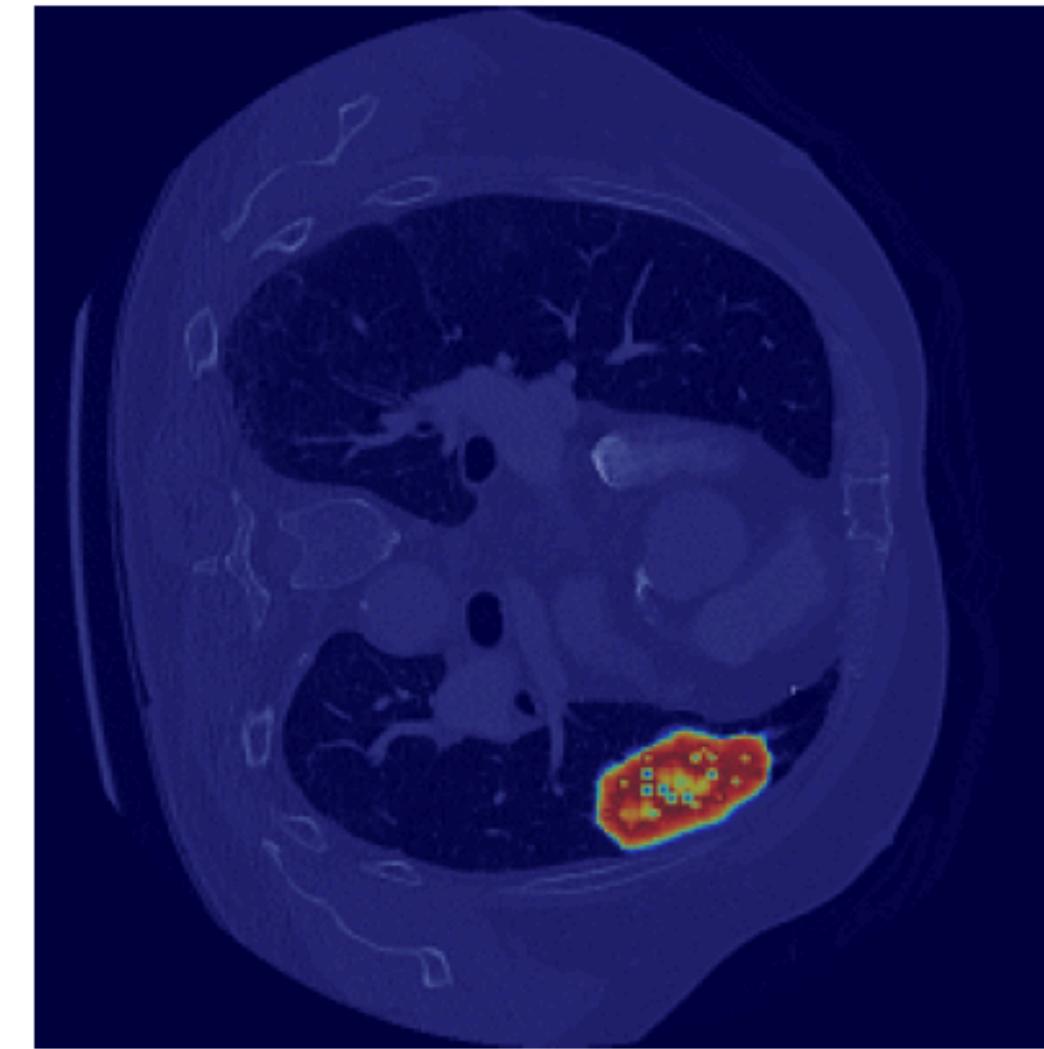
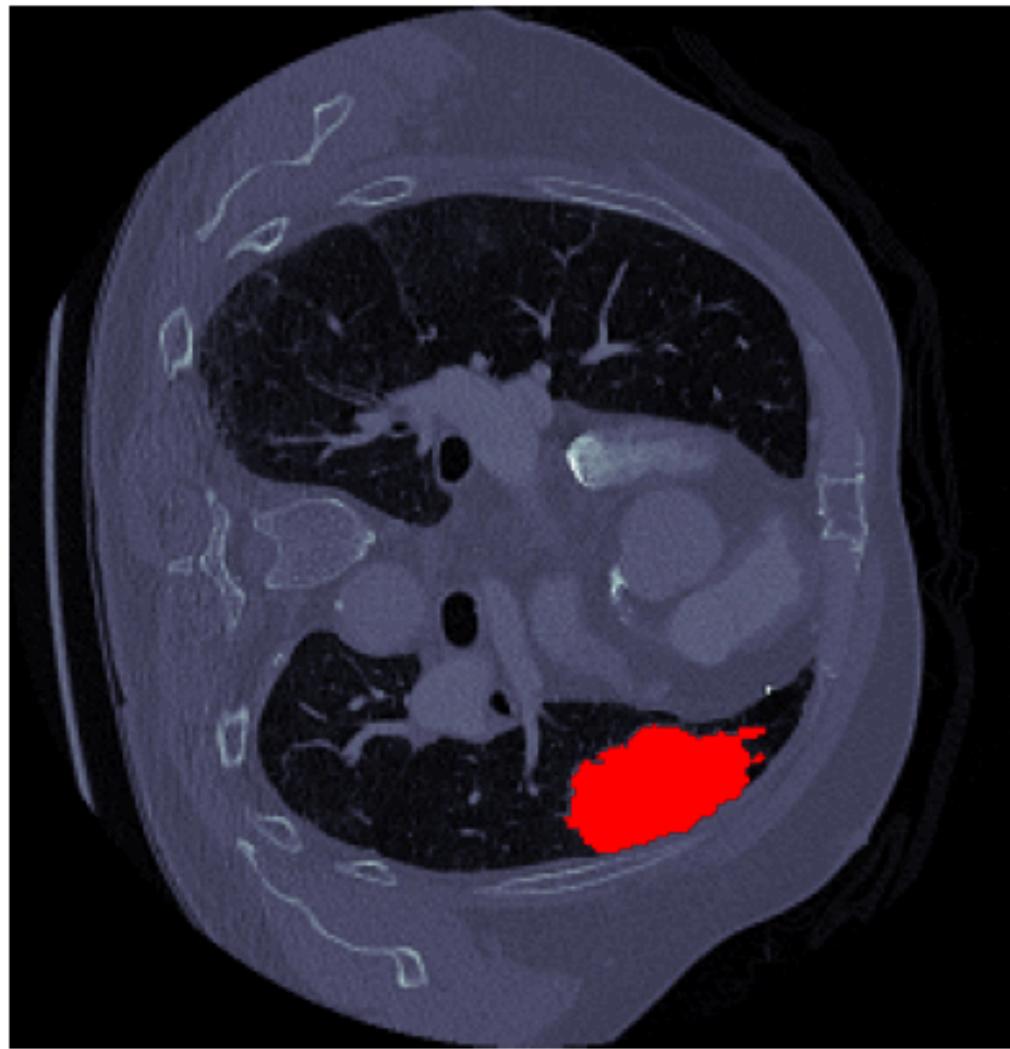
Dice Loss Metrics

Threshold	Dice Score	Precision	Recall	IoU
Pure Probs	0.838374	0.846831	0.830083	0.721724
0.5	0.838498	0.846927	0.830235	0.721908
0.8	0.839248	0.853021	0.825912	0.723021

BCE Loss Metrics

Pure Probs	0.861609	0.854944	0.86838	0.756866
0.5	0.90812	0.917452	0.898976	0.831703
0.8	0.871632	0.979533	0.785144	0.772471

GRAD-CAM VISUALIZATION



Tumor slice with the ground truth mask overlapped (left), heatmap overlay of Grad-CAM (middle) and model segmentation mask prediction (right)

WHAT FAILED?

- Focal loss and its varieties like focal tversky, focal jaccard index, etc did not work for any hyperparameter value and instead lead to overfitting. Similarly, weighted BCE lead to a skew in precision vs recall which harmed performance.
- Inclusion of data augmentation techniques like random brightness adjustment, pixel inversion, random image erasing, random sharpen adjustment and random cropping into the existing pipeline did not improve model's performance.
- Using Mish activation function in place of ReLU lead to model overfitting and harsher regularization hurt the model's performance.
- Not using dropout decreases model convergence speed by half with the trade-off being worse model performance.
- Integration of Conditional Random Fields in the form of either CRF-CNN or CRF-RNN as part of the post-processing did not improve the segmentation prediction.
- Using a Lung Window of 800 to -1024 Hounsfield Units didn't improve performance.
- Dice Loss doesn't train a better model in-spite of higher pure probability metrics.
- Experimentation with the Swin-UNetr architecture lead to overfitting due to the transformer architecture and less tumor slice to train on, with a best dice score of 0.64.

CONCLUSION

- No fancy loss function required, BCE loss is just fine
- Even with high threshold value of 0.8 the models are still displaying equivalent high validation metrics which shows that they are learning the cancer pixels properly with high confidence and robustness
- Experimentation results published have been found to be using a threshold value of 0.5 for final results
- The goal is to keep GMAC as low as possible with 700 MB being the maximum limit

THANK YOU