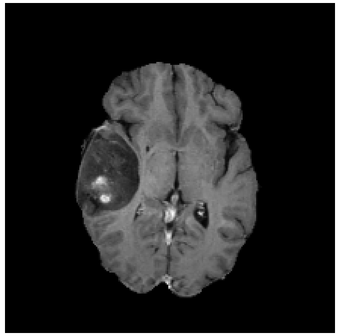
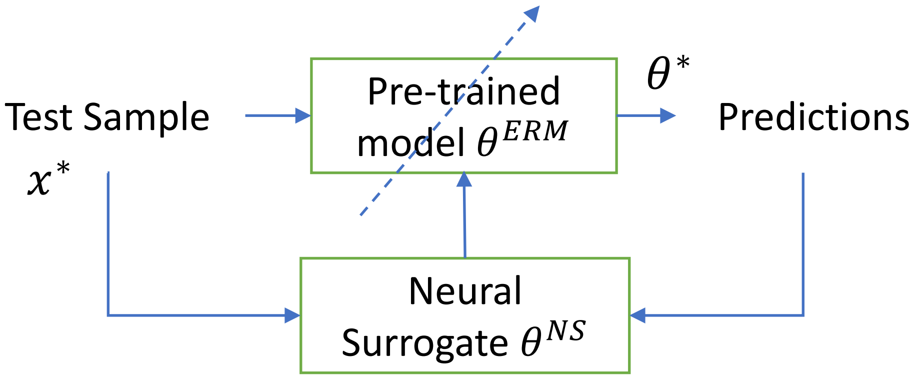
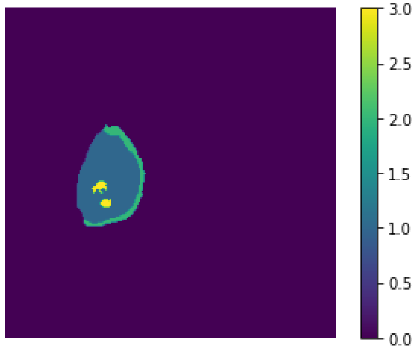
Test-time adaptation using neural surrogates for medical image segmentation – Hariharan Ravishankar

- Goal : Address generalization of semantic segmentation models by Test-time adaptation of pre-trained model's weights θ^{ERM} to θ^*, conditioned only on new input sample x^* in an unsupervised manner.
- Data-set : BRATS 2018 challenge High Grade Glioma (HGG) dataset.



Input 4 channels
[T1, T1Gd, T2, T2-Flair]





Output 4 channels
[0- Background, 1 – Core, 2 – Edema, 3 – EnhT]

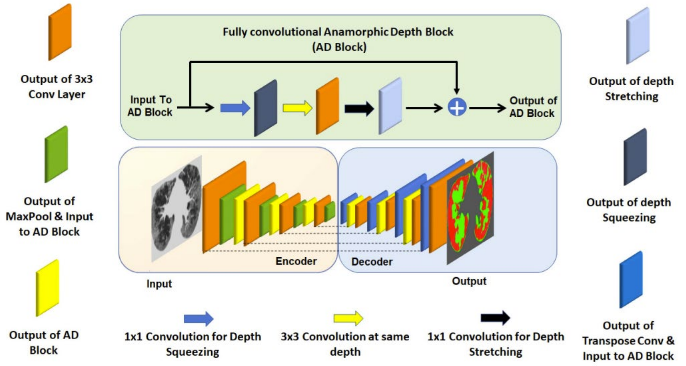
- Proposal : Learn θ^{NS}- Neural surrogate, which learns to predict quality of segmentation

$$\boxed{g_{\theta^{NS}}(x, y^{\wedge})\ \alpha\ \ l(f_\theta(x), y)}$$

- During test-time use learnt $g_{\theta^{NS}}$ for adapting to single sample

Base Model and Neural Surrogate(s) Details

Courtesy [1]

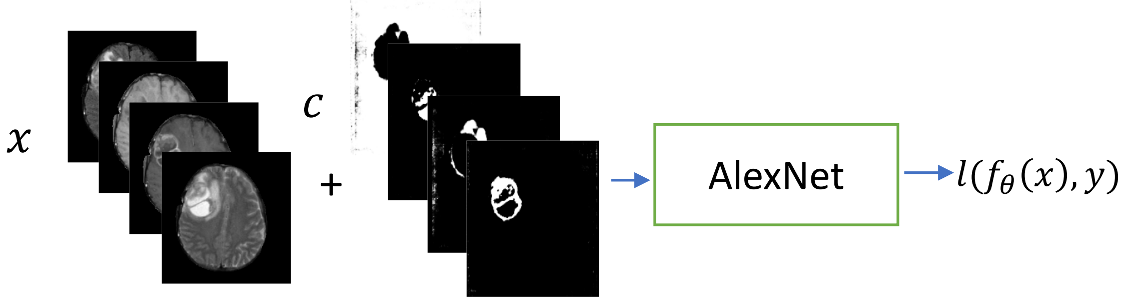


Base Model Details

- Multi-class segmentation network – AnamNet. Input and output channels = 4.
- Total # of subjects = 210. Training and Validation – 168 subjects, Testing – 42
- Number of slices per subject = 90. Optimizer = Adam, 100 epochs, batch size = 32.

Neural surrogate 1 – Regression to cumulative loss

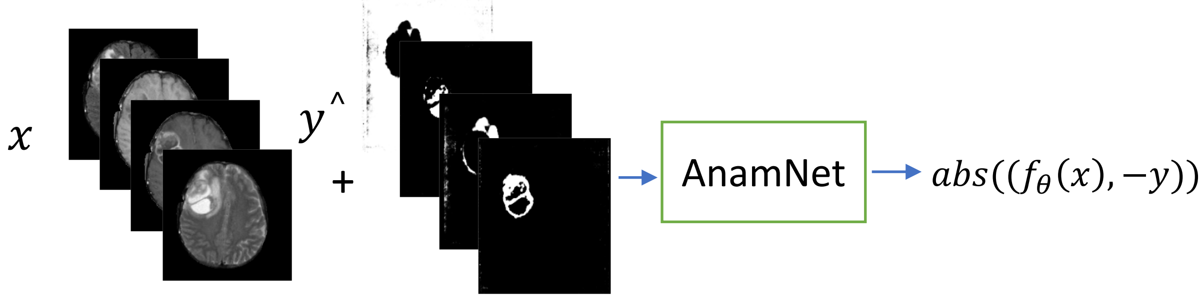
- Architecture used : AlexNet
- Input – 8 channels $\{x, y^{\wedge}\}$. Output – single loss value $l(f_\theta(x), y)$.
- Hyperparameters same as base model.



- Learning was inadequate. Single loss value could not capture pixel-wise mistakes made by base-model.

Neural surrogate 2 – Predicting pixel-wise error maps

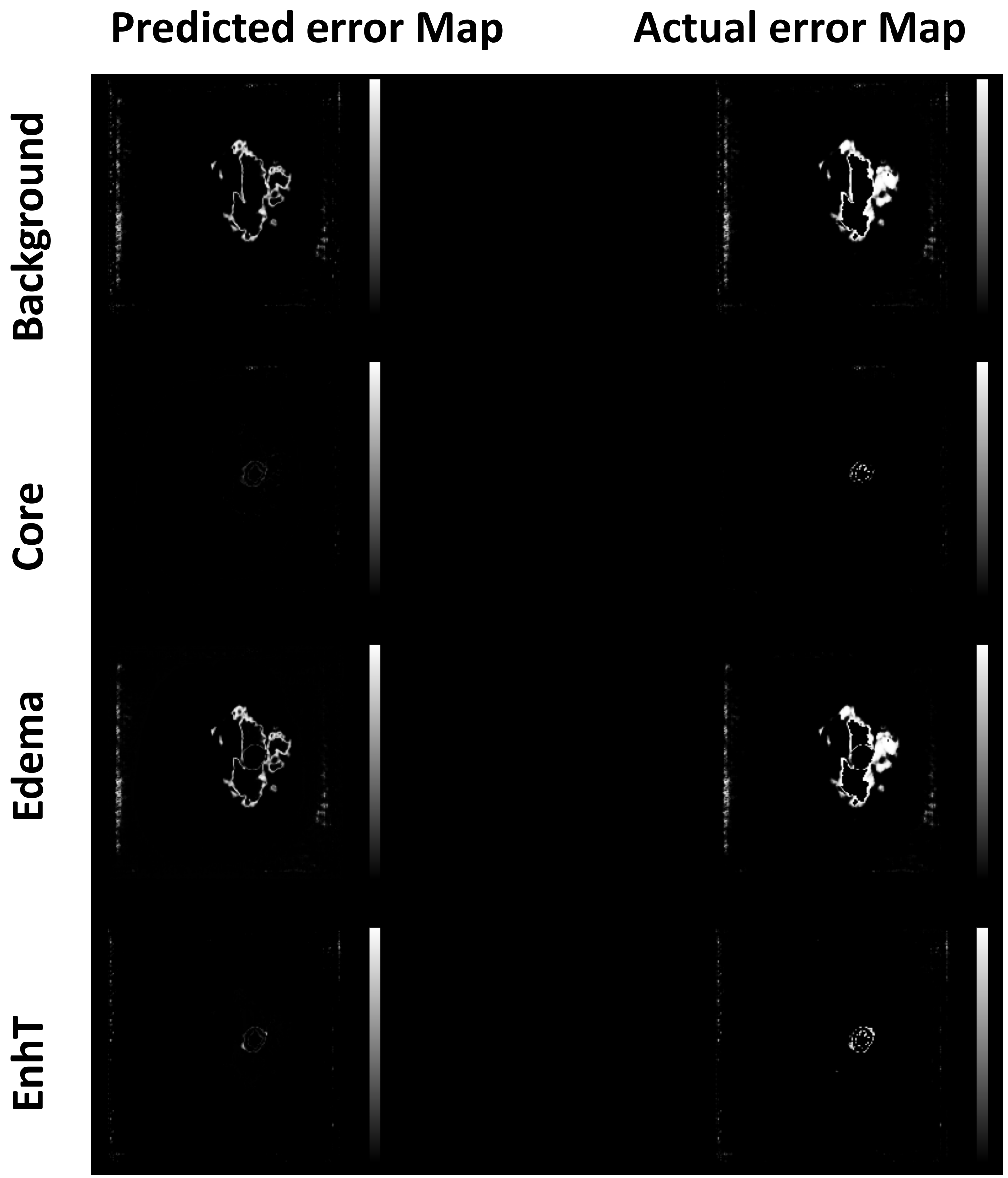
- Architecture used : AlexNet
- Input – 8 channels $\{x, y^{\wedge}\}$. Output – 4 channels - $abs((f_\theta(x), -y))$.
- Hyperparameters same as base model.



- **NS2 was able to approximate error maps adequately.**

1. Paluru, Naveen, et al. "Anam-Net: Anamorphic depth embedding-based lightweight CNN for segmentation of anomalies in COVID-19 chest CT images." IEEE Transactions on Neural Networks and Learning Systems 32.3 (2021): 932-946.

Neural Surrogate 2 performance and test-time adaptation

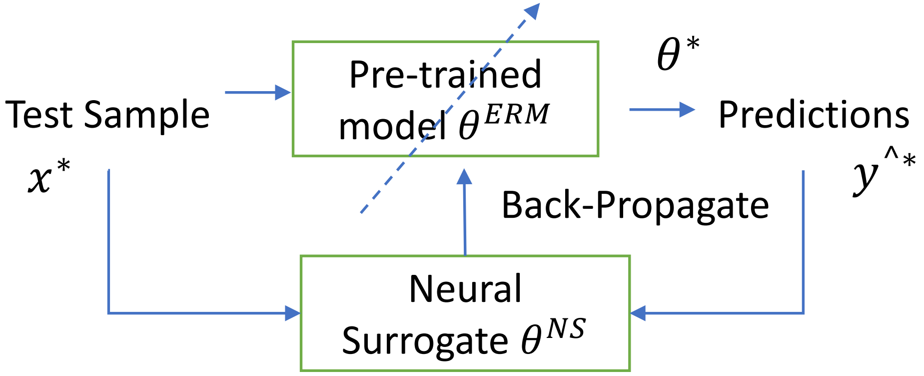


Neural Surrogate Performance

- NS2 – $g_{\theta^{NS}}(x, y^{\wedge})$: was able to capture mistakes made by base Model.
- Note : Overlapping information in background and one of the classes.
- Criterion used : MSE. Training error ~ 0.0010

Iterative procedure for test-time adaptation of using NS2

- For batch of samples from test-data, initial predictions were made using pre-trained model - θ^{ERM}
- The output predictions $y^{\wedge *}$ and inputs x^* were passed to $g_{\theta^{NS}}$.
- Output of $g_{\theta^{NS}}$ was used to backpropagate and update weights.
- Only batch norm-parameters were updated using this procedure.

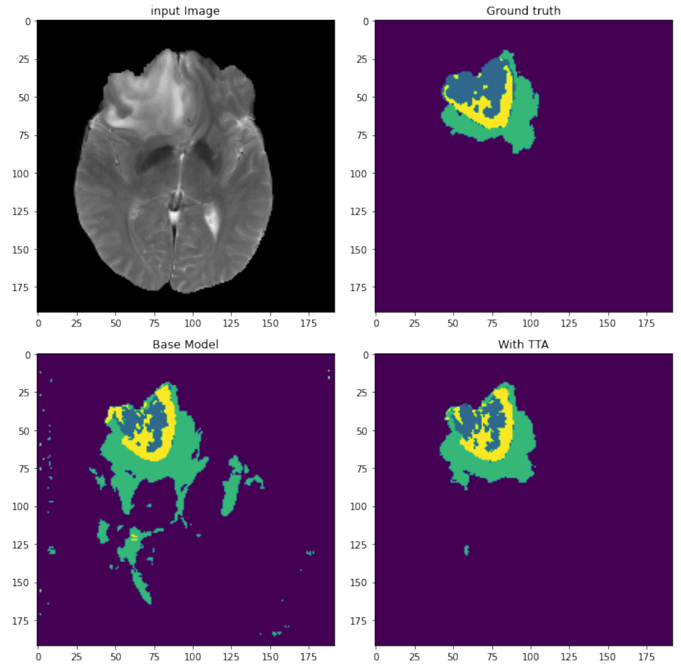


Quantitative results with Test-time adaptation using NS2

Dice overlap scores on test data-set for different components of tumor and for varying training data sizes

Class / Training Data Fraction	0.25		0.5		1.0	
	Orig Model	With TTA	Orig Model	With TTA	Orig Model	With TTA
Core	0.442	**0.527**	0.535	**0.655**	0.69	**0.692**
Edema	0.650	**0.660**	0.714	**0.724**	0.718	**0.765**
EnhT	0.429	**0.692**	0.800	**0.800**	0.780	**0.82**
Whole Tumor	0.664	**0.762**	0.799	**0.834**	0.818	**0.864**

Qualitative visualization of improvements in segmentation results with TTA using neural surrogates



Key Take-aways

- TTA with NS2 consistently improves performance on all types of tumor tissues with roughly 5% increase on full training dataset.
- TTA with 50% data beats orig model's performance on full data.
- Increase in performance due to TTA is possible for all training sizes, including 25% of training data with close to 10% increase in dice overlap.

Other Details

- Resources : MIG Lab server and workstation with RTX A600 gpu cards.
- Orig Model inference – 10ms, Adapt procedure – 1s for batch size of 24.
- Experiments with more architecture and additional problems in progress.