1. Motivation and Objective

- Major challenges in pathology segmentation include:
  - Lack of access to large annotated datasets.
  - Existing small public datasets suffer from large class imbalance and inter-subject variability issues.
- State-of-the-art models are based on deep learning methods, which perform well when trained on large datasets.
- Leveraging models trained on large datasets in order to improve results on smaller dataset could be impactful.
- We explore several fine-tuning strategies to best leverage source model for target dataset of varying sizes.

2. Proposed Method

- First Phase: Pretraining the UNet[1] with source MS dataset.
- Second Phase:
  - Replacing last three layers of the pre-trained MS network with new layers.
  - Fine-tuning with target brain tumor in three different ways:
    - FT_LastThree: only the newly added layers are re-trained.
    - FT_Decoder: only the decoder is fine-tuned.
    - FT_All: the whole network is fine-tuned.

3. Data and Experimentation

- Source Data: Multiple Sclerosis (MS) dataset
  - Proprietary, multi-modal, multi-site, multi-scanner clinical trial dataset.
  - 3630 Multi-modal MRI (T1w, T2w, FLAIR, and T1 post-Gad).
- For our first phase, we use:
  - 80% of available data to train 3D UNet.
  - Remaining 20% of data to validate 3D UNet.
- Weighted Binary Cross Entropy loss.
- Evaluation metric: ROC curves for T2 lesion segmentation.
- An AUC of 0.77 is obtained on the validation set.

- Target Data: Brain Tumor dataset (BraTS 2018 challenge[2])
  - Multi-modal MRI (T1, T2, FLAIR, and T1ce).
  - Registration to same space as source data using ANTs tool[3].
  - For 20, 50, 100, 150 brain tumor samples (subset of BraTS 2018 training set):
    - Transfer learning: FT_LastThree, FT_Decoder, FT_All.
    - Baseline: Training from scratch with brain tumor MRI scans.
- Weighted Cross Entropy loss.
- Four-fold cross validation is performed.
- A local validation set of 50 samples is used to select operating point.

4. Quantitative Results

- FT-All outperforms the baseline in almost every case.
- High gain when number of tumor cases is extremely low, i.e., 20.
- Gain of FT-All over baseline diminishes with more samples.

5. Qualitative Results (fine-tuning with 20 brain tumor samples)

- FT-All is able to capture sub-structures of tumor better than other methods.
- Performance is better on HGG cases over LGG, as more HGG cases are present in training set.

6. Conclusion

- We explored different strategies for transfer learning across diseases for the task of focal pathology segmentation.
- We observed that fine-tuning the whole network outperforms baseline and other fine-tuning methods, especially when very small target datasets are available, unlike in case of natural images where fine-tuning just last few layers helps.
- We encourage public release of models trained on large datasets.

References:

Acknowledgement:
This work was supported by a Canadian Natural Science and Engineering Research Council (NSERC) Collaborative Research and Development Grant, Synaptive Medical and an award from the International Progressive MS Alliance.