Improving Pathological Structure Segmentation Via Transfer Learning Across Diseases

Barleen Kaur, Paul Lemaitre, Raghav Mehta, Nazanin Mohammadi-Sepahvand, Doina Precup, Douglas L. Arnold and Tal Arbel

Workshop on Domain Adaptation and Representation Transfer, MICCAI 2019

Motivation

- Lack of access to large annotated datasets: major challenges in medical imaging analysis.

[1] Özgün Çiçek et al, MICCAI 2016

[2] Veronika and et al., MIA 2019

Motivation

- Lack of access to large annotated datasets: major challenges in medical imaging analysis.
- State-of-art models are based on deep learning methods, which perform well when trained on large datasets[1].

[1] Özgün Çiçek et al, MICCAI 2016

[2] Veronika and et al., MIA 2019

Motivation

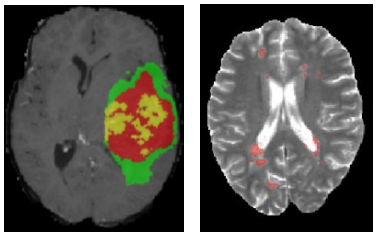
- Lack of access to large annotated datasets: major challenges in medical imaging analysis.
- State-of-art models are based on deep learning methods, which perform well when trained on large datasets[1].
- Transfer learning has been explored in various applications such as classification, detection and segmentation. See [2] for a survey.

[1] Özgün Çiçek et al, MICCAI 2016

[2] Veronika and et al., MIA 2019

Motivation

- Lack of access to large annotated datasets: major challenges in medical imaging analysis.
- State-of-art models are based on deep learning methods, which perform well when trained on large datasets[1].
- Transfer learning has been explored in various applications such as classification, detection and segmentation. See [2] for a survey.
- Pathology segmentation:
  - Public datasets are small. Most of the large datasets are inhouse.
  - Difficult to obtain ground truth.
  - Class imbalance and inter-subject variability.

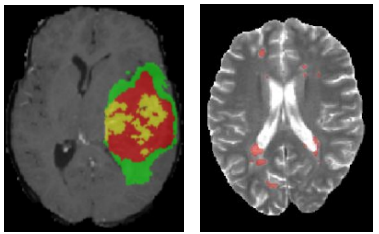


[1] Özgün Çiçek et al, MICCAI 2016

[2] Veronika and et al., MIA 2019

Motivation

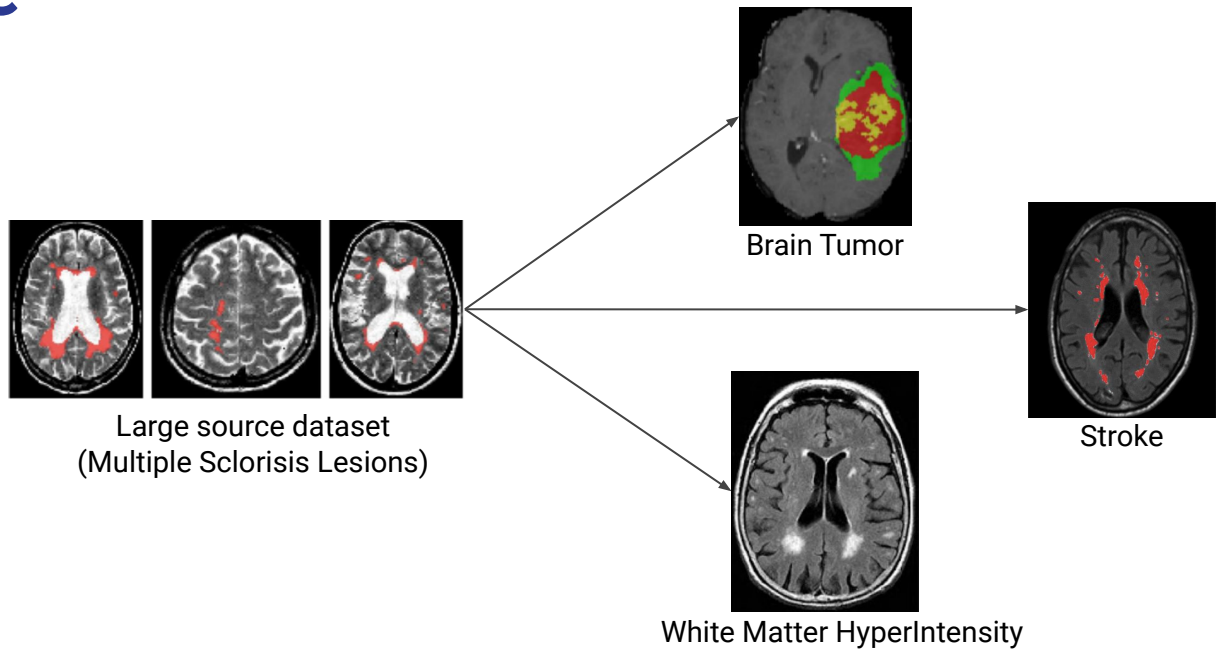
- Lack of access to large annotated datasets: major challenges in medical imaging analysis.
- State-of-art models are based on deep learning methods, which perform well when trained on large datasets[1].
- Transfer learning has been explored in various applications such as classification, detection and segmentation. See [2] for a survey.
- Pathology segmentation:
  - Public datasets are small. Most of the large datasets are inhouse.
  - Difficult to obtain ground truth.
  - Class imbalance and inter-subject variability.
  - **Leveraging models trained on large datasets** in order to improve **pathology segmentation** results on smaller dataset **across different diseases** could be impactful in medical image analysis.



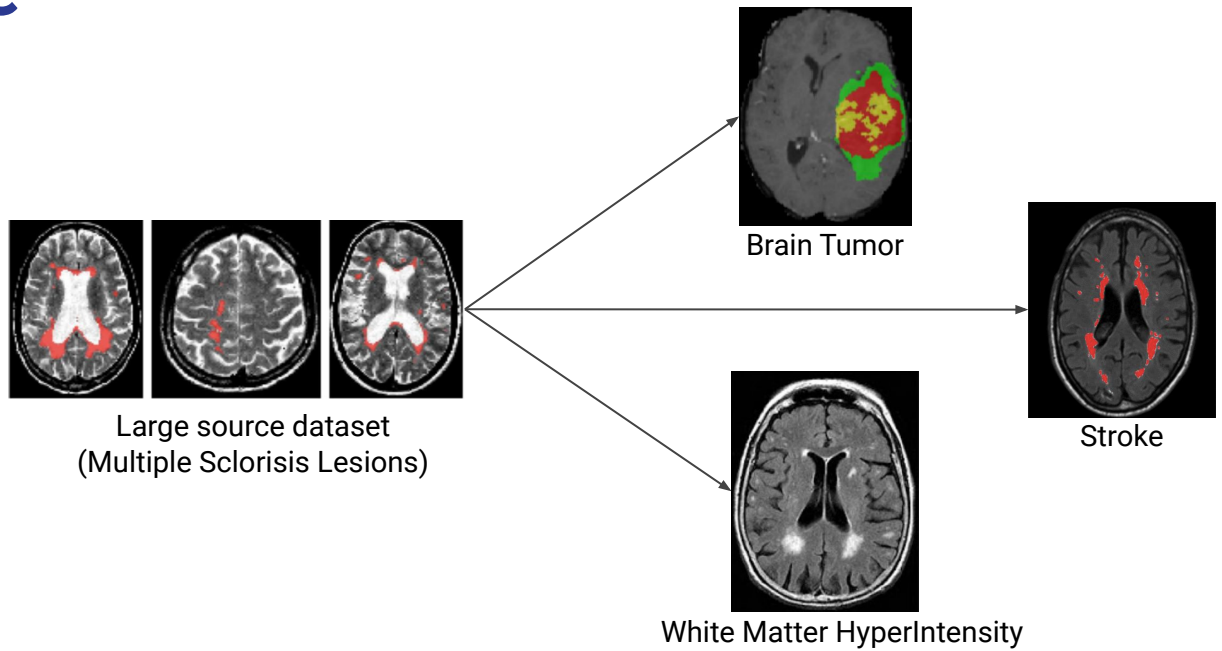
[1] Özgün Çiçek et al, MICCAI 2016

[2] Veronika and et al., MIA 2019

Objective


Brain Tumor
Stroke
Large source dataset
(Multiple Sclorisis Lesions)
White Matter HyperIntensity

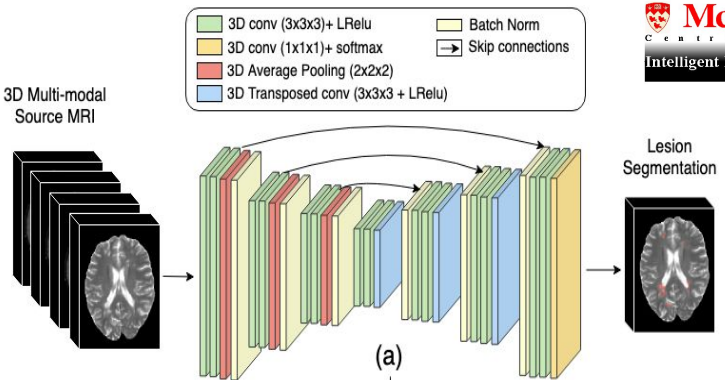
Objective



- Natural images: fine-tuning just last few layers helps. Is it same case in medical domain?
- We explore several fine-tuning strategies to see how to best leverage the source model and adapt it to the target dataset of varying sizes.

Methodology

First Phase: Pretraining the UNet[1] with source MS dataset.

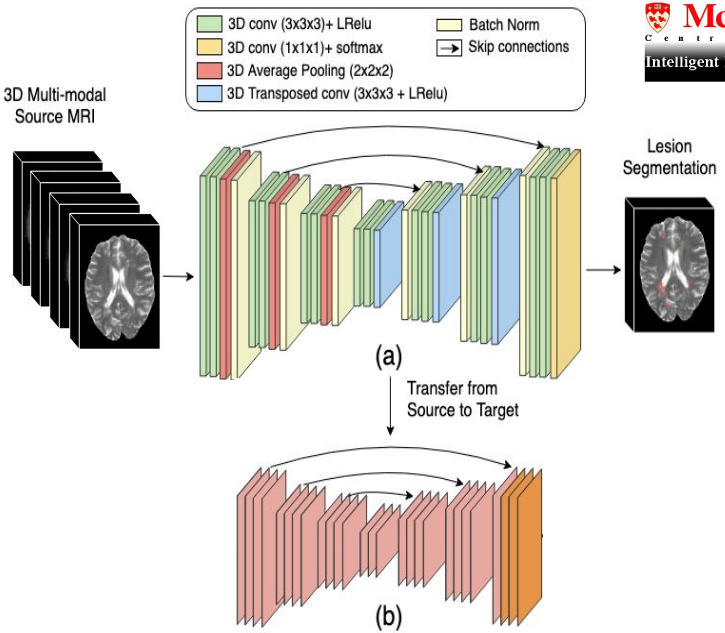


[1] Özgün Çiçek et al. MICCAI 2016

Methodology

First Phase: Pretraining the UNet[1] with source MS dataset.

Second Phase: Replacing the last three task-specific layers of the pre-trained MS network with **new layers** and then fine-tuning with target brain tumor in three different ways:



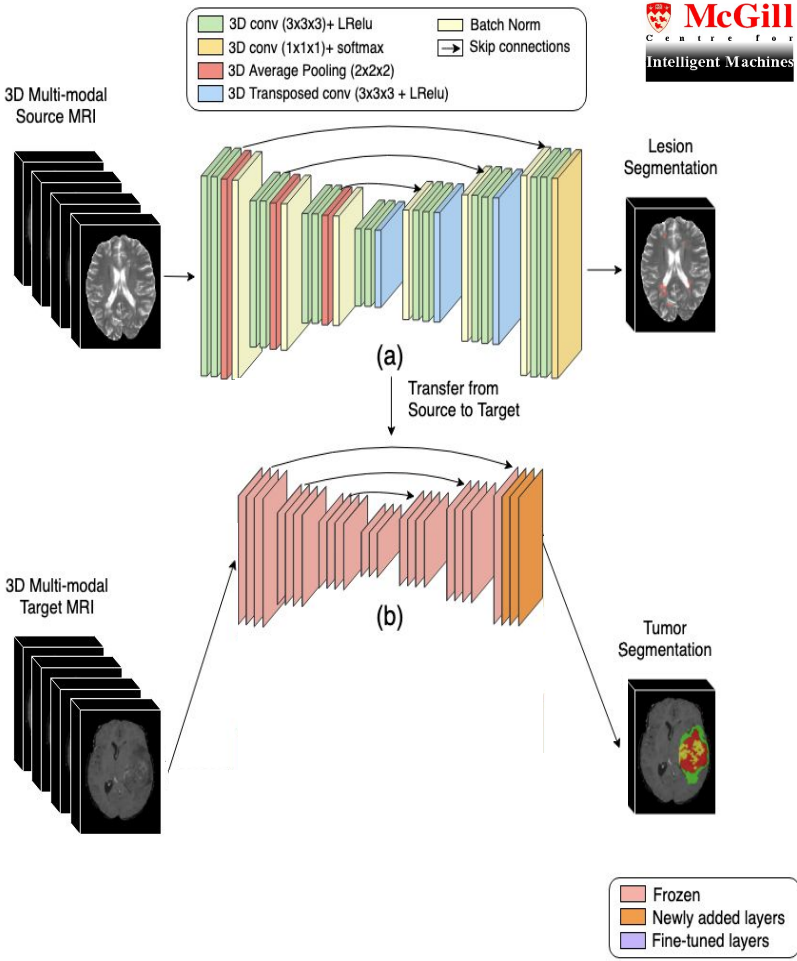
[1] Özgün Çiçek et al. MICCAI 2016

Methodology

First Phase: Pretraining the UNet[1] with source MS dataset.

Second Phase: Replacing the last three task-specific layers of the pre-trained MS network with **new layers** and then fine-tuning with target brain tumor in three different ways:

★ **FT_LastThree**: only the newly added layers are re-trained.



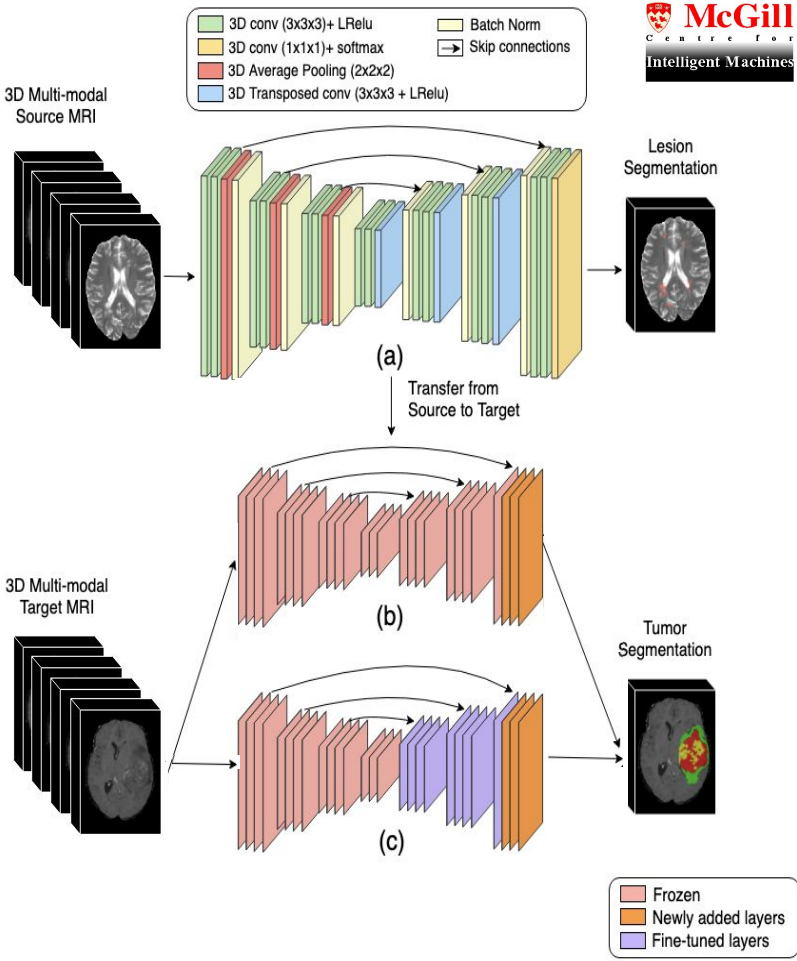
[1] Özgün Çiçek et al. MICCAI 2016

Methodology

First Phase: Pretraining the UNet[1] with source MS dataset.

Second Phase: Replacing the last three task-specific layers of the pre-trained MS network with **new layers** and then fine-tuning with target brain tumor in three different ways:

- ★ **FT_LastThree**: only the newly added layers are re-trained.
- ★ **FT_Decoder**: Encoder part is frozen and only the decoder is fine-tuned.



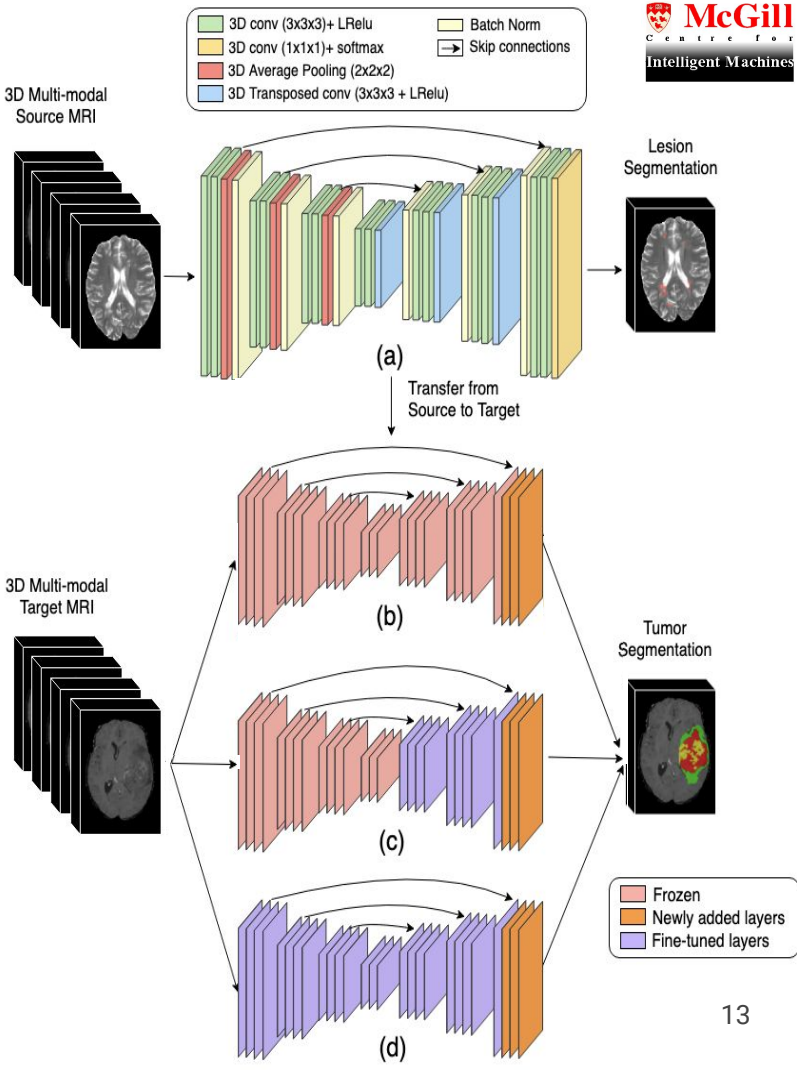
[1] Özgün Çiçek et al. MICCAI 2016

Methodology

First Phase: Pretraining the UNet[1] with source MS dataset.

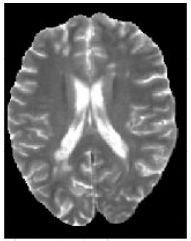
Second Phase: Replacing the last three task-specific layers of the pre-trained MS network with **new layers** and then fine-tuning with target brain tumor in three different ways:

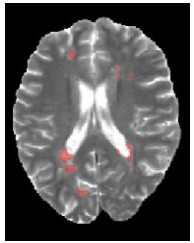
- ★ **FT_LastThree**: only the newly added layers are re-trained.
- ★ **FT_Decoder**: Encoder part is frozen and only the decoder is fine-tuned.
- ★ **FT_All**: The whole pretrained network is fine-tuned.



[1] Özgün Çiçek et al. MICCAI 2016

Data





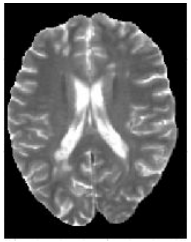
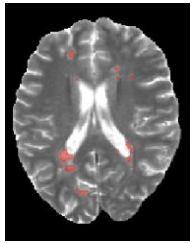
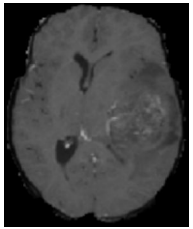
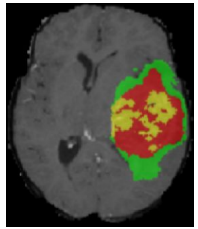
McGill Centre for Intelligent Machines

Source: Multiple Sclerosis Dataset

- Proprietary, multi-modal, multi-site, multi-scanner clinical trial dataset.
- 4 modalities (T1w, T2w, FLAIR, and T1 post-Gad)
- Resolution: 1 x 1 x 1 mm^3
- Dimensions: 229x193x193.
- Total patient scans: 3630 multimodal MRI
- T2 binary lesion segmentation mask provided.

[1] Menze et al The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS), TMI 2015

Data

McGill Centre for Intelligent Machines

Source: Multiple Sclerosis Dataset

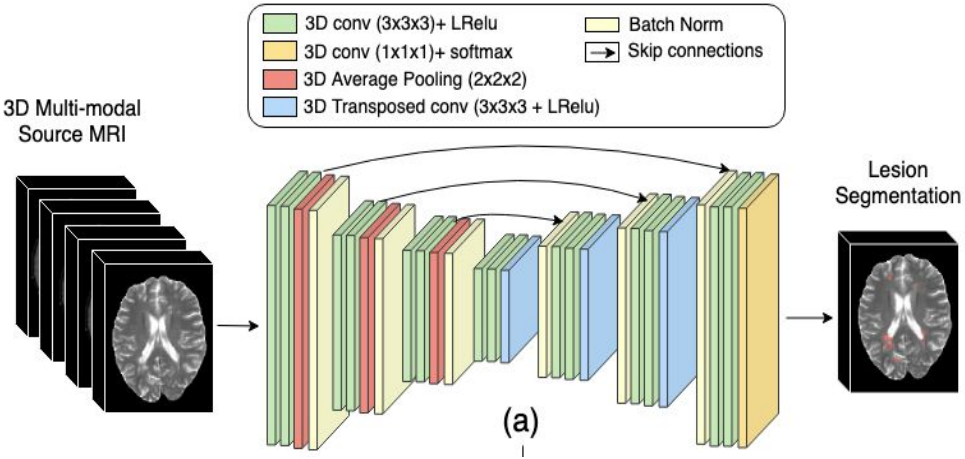
- Proprietary, multi-modal, multi-site, multi-scanner clinical trial dataset.
- 4 modalities (T1w, T2w, FLAIR, and T1 post-Gad)
- Resolution: 1 x 1 x 1 mm^3
- Dimensions: 229x193x193.
- Total patient scans: 3630 multimodal MRI
- T2 binary lesion segmentation mask provided.

Target: BraTS 2018 challenge Dataset[1]

- 4 modalities (T1, T2, FLAIR, T1c)
- Resolution: 1x1x1 mm^3
- Dimensions: 155 x 240 x 240
- Manual marking for 3 types of tumor (edema, necrotic core, and enhancing core)
- BraTS 2018 Training data (285 patients) for training (Ground Truth available)
- BraTS 2018 Validation data (66 patients) for testing (Ground truth not provided)

[1] Menze et al The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS), TMI 2015

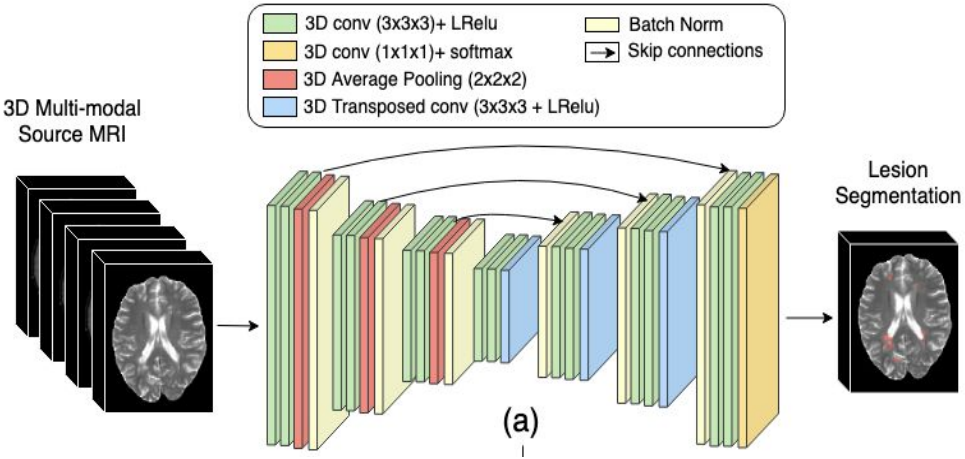
Experimentation (First Phase: Pre-training)



Pre-training the UNet with source MS data for T2 lesion segmentation.

- **MS Data** (3630 patient scans) → 80% **Training**, 20% **Validation**

Experimentation (First Phase: Pre-training)

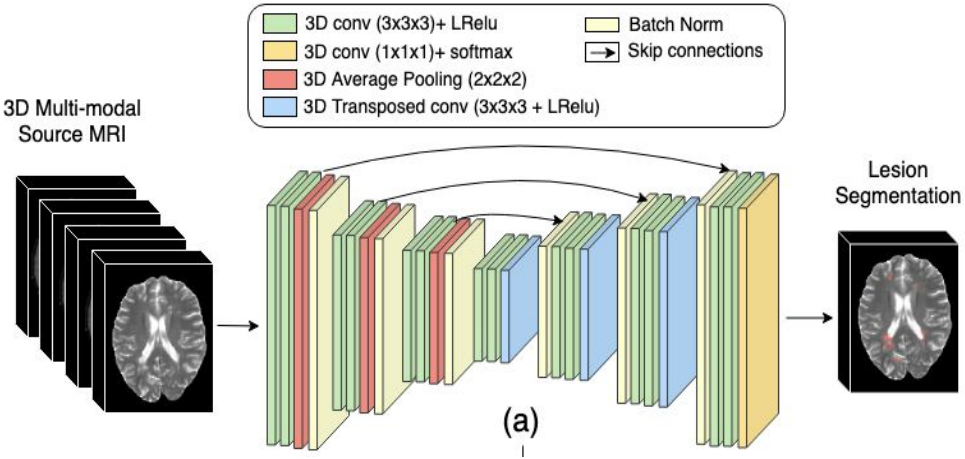


Pre-training the UNet with source MS data for T2 lesion segmentation.

- **MS Data** (3630 patient scans) → 80% **Training**, 20% **Validation**

- Weighted binary cross entropy was used as loss function.

Experimentation (First Phase: Pre-training)



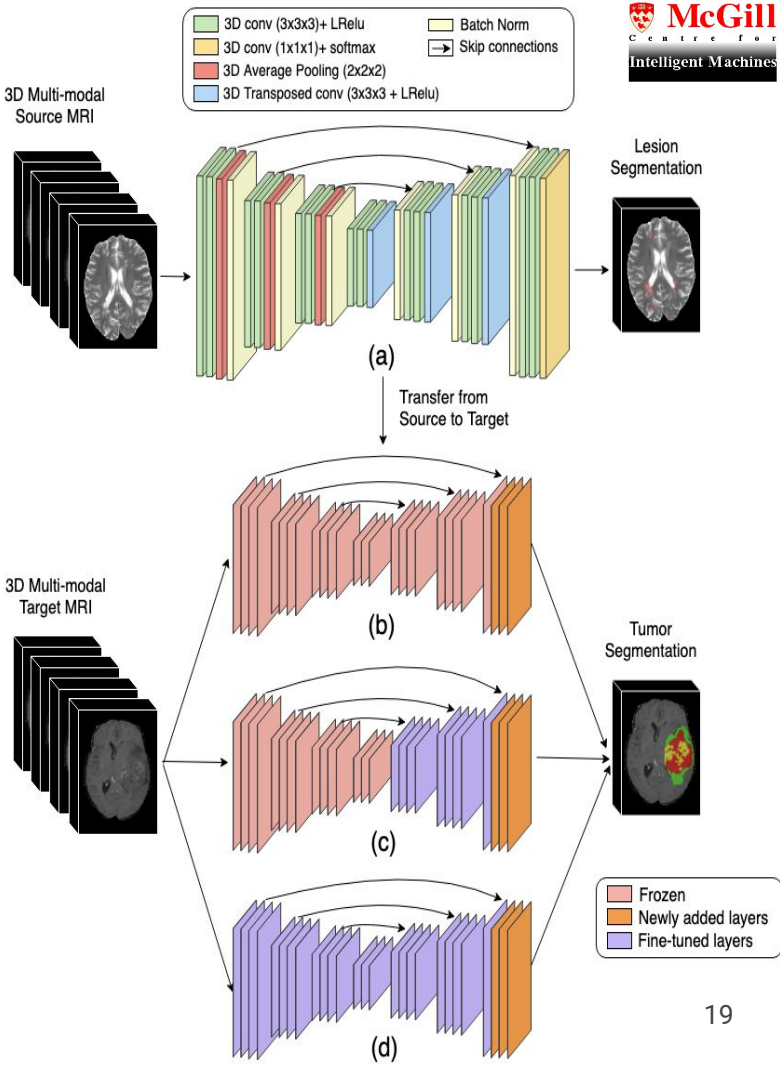
Pre-training the UNet with source MS data for T2 lesion segmentation.

- **MS Data** (3630 patient scans) → 80% **Training**, 20% **Validation**

- Weighted binary cross entropy was used as loss function.
- An **AUC of 0.77** was obtained on the validation (test) set.

Experimentation (Second Phase: Fine-tuning)

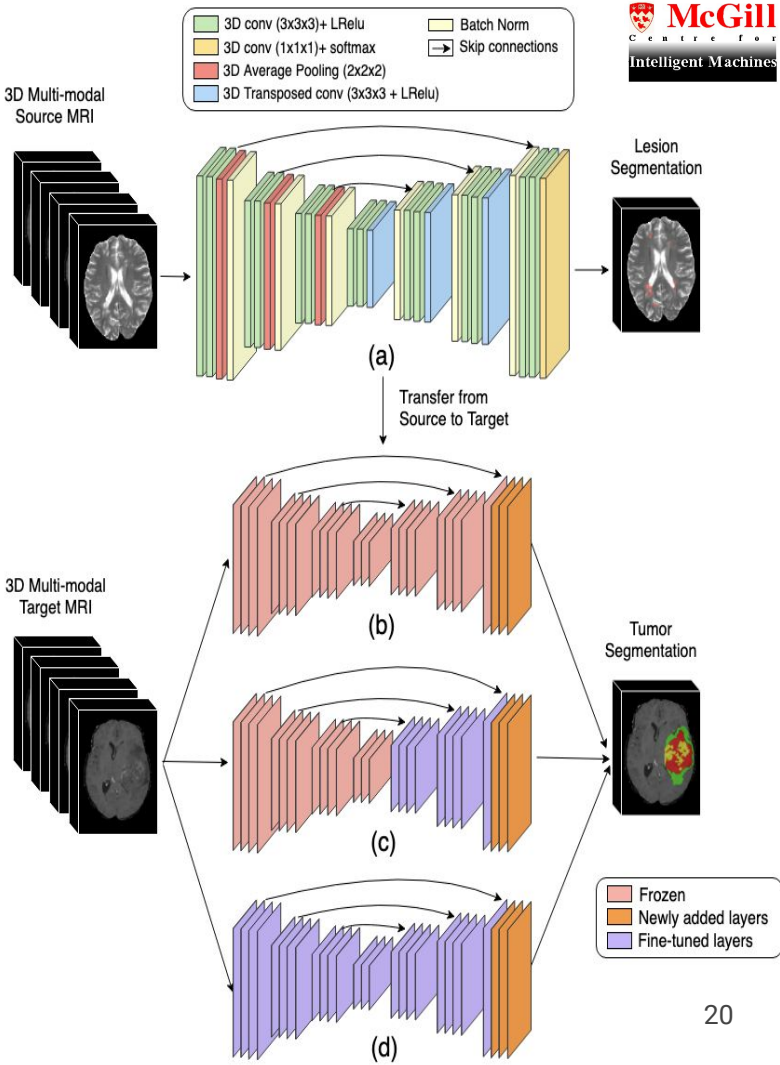
For **20, 50, 100, 150** brain tumor MRI scans:



Experimentation (Second Phase: Fine-tuning)

For **20, 50, 100, 150** brain tumor MRI scans:

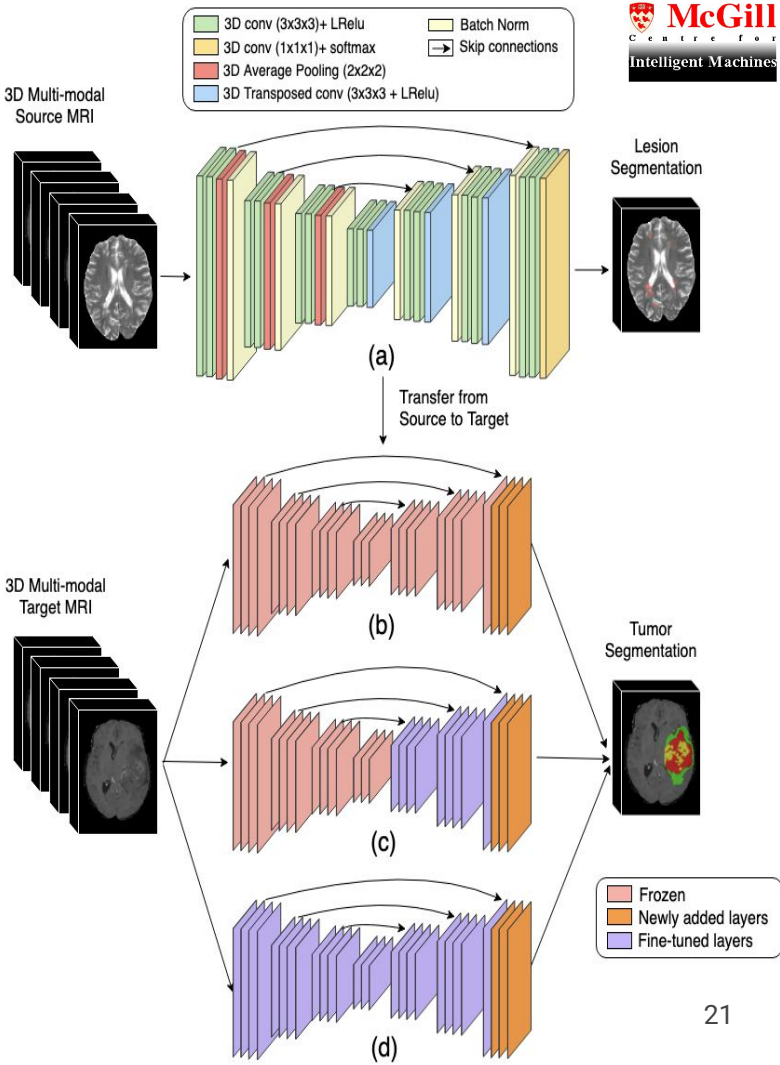
- ★ **Transfer Learning**
  - FT_Last Three
  - FT_Decoder
  - FT_All



Experimentation (Second Phase: Fine-tuning)

For **20, 50, 100, 150** brain tumor MRI scans:

- ★ **Transfer Learning**
  - FT_Last Three
  - FT_Decoder
  - FT_All
- ★ **Baseline** (Training from scratch)

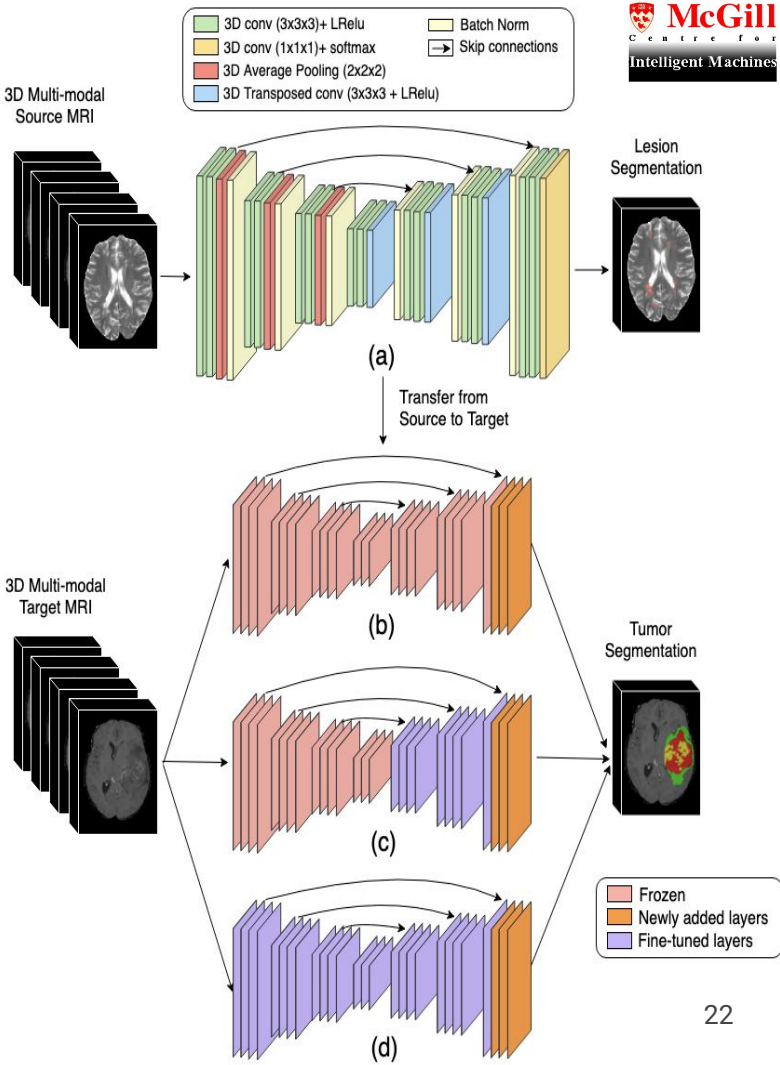


Experimentation (Second Phase: Fine-tuning)

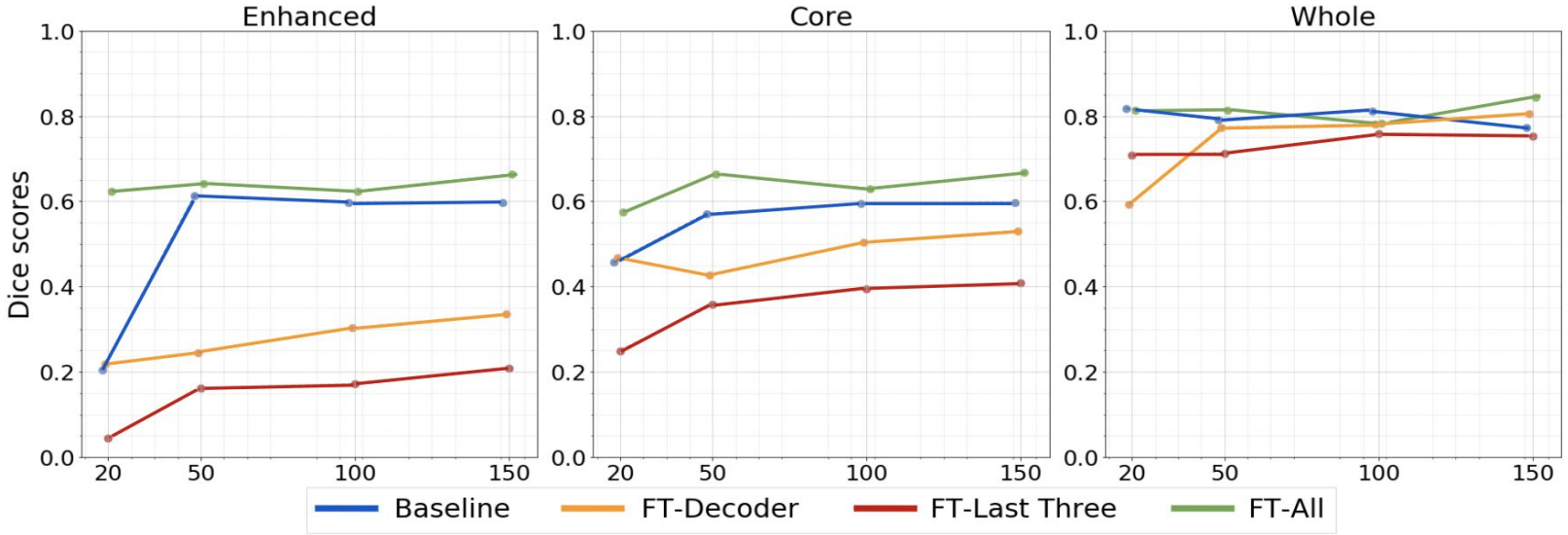
For **20, 50, 100, 150** brain tumor MRI scans:

- ★ **Transfer Learning**
  - FT_Last Three
  - FT_Decoder
  - FT_All
- ★ **Baseline** (Training from scratch)

- ➢ Weighted Cross entropy loss.
- ➢ Four-fold cross validation
- ➢ A **local validation set** of 50 samples is used to select the operating point.

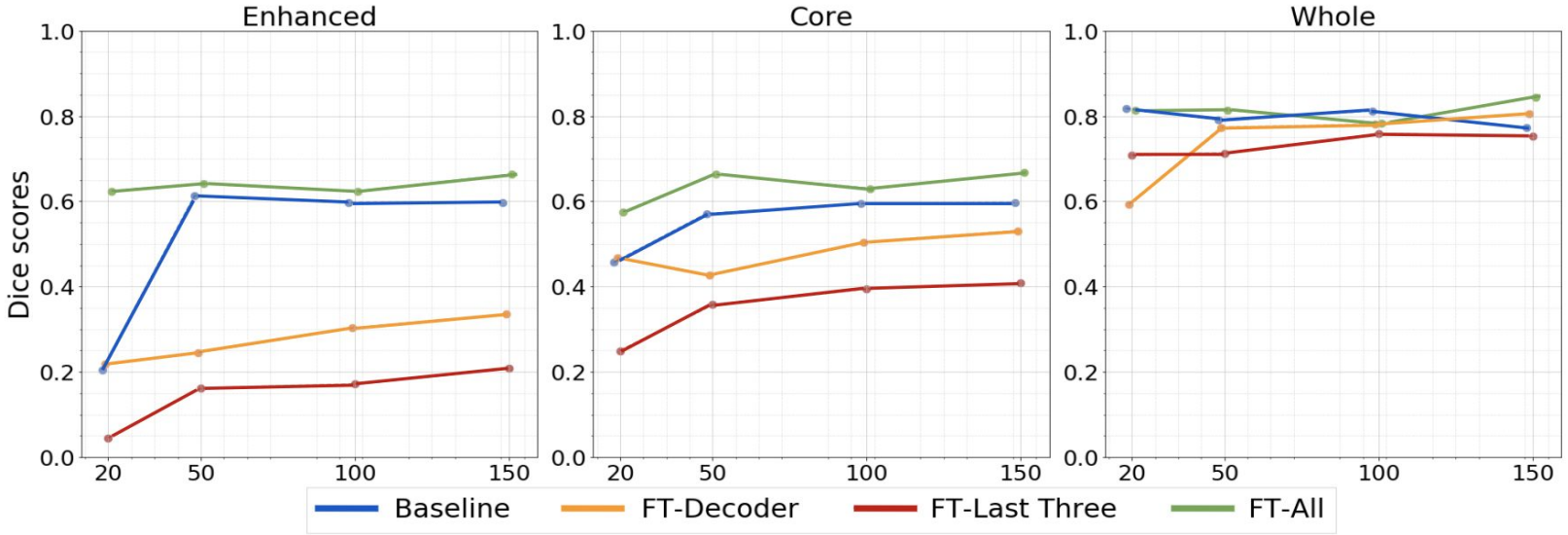


Quantitative Results (on BraTS 2018 Validation set)



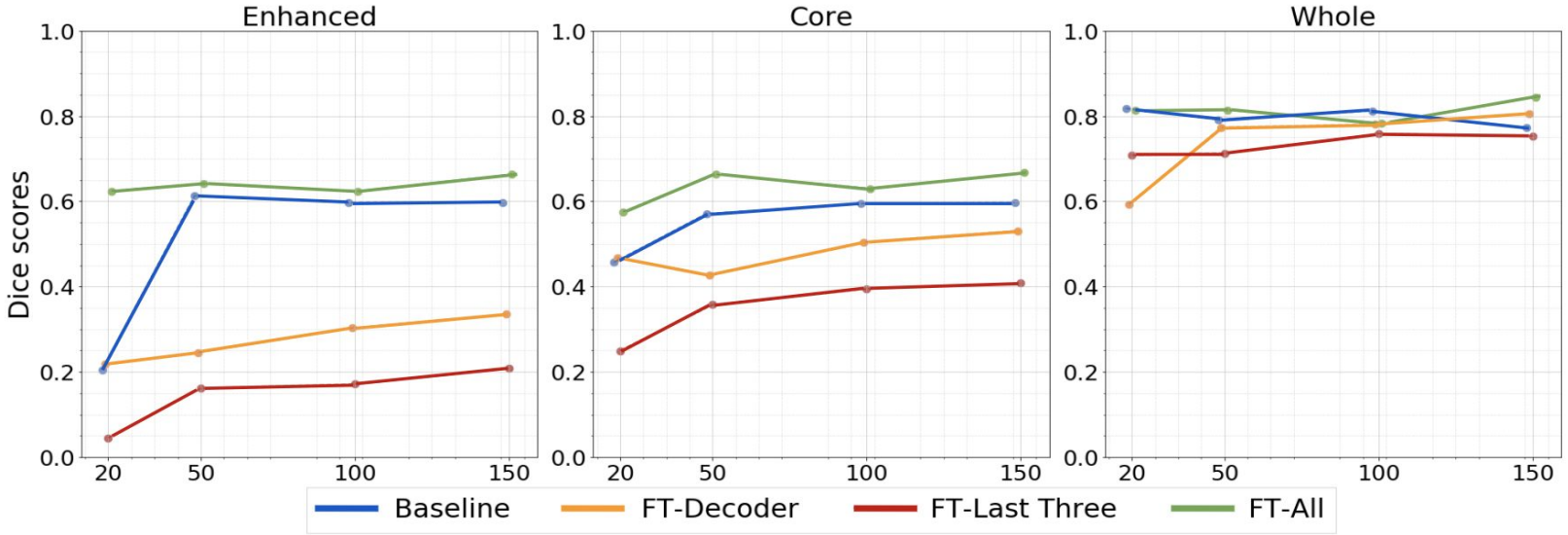
➢ **FT-All outperforms the baseline** in almost every case.

Quantitative Results (on BraTS 2018 Validation set)



- **FT-All outperforms the baseline** in almost every case.
- Best when the **number of tumor cases is extremely low**, i.e. 20.

Quantitative Results (on BraTS 2018 Validation set)

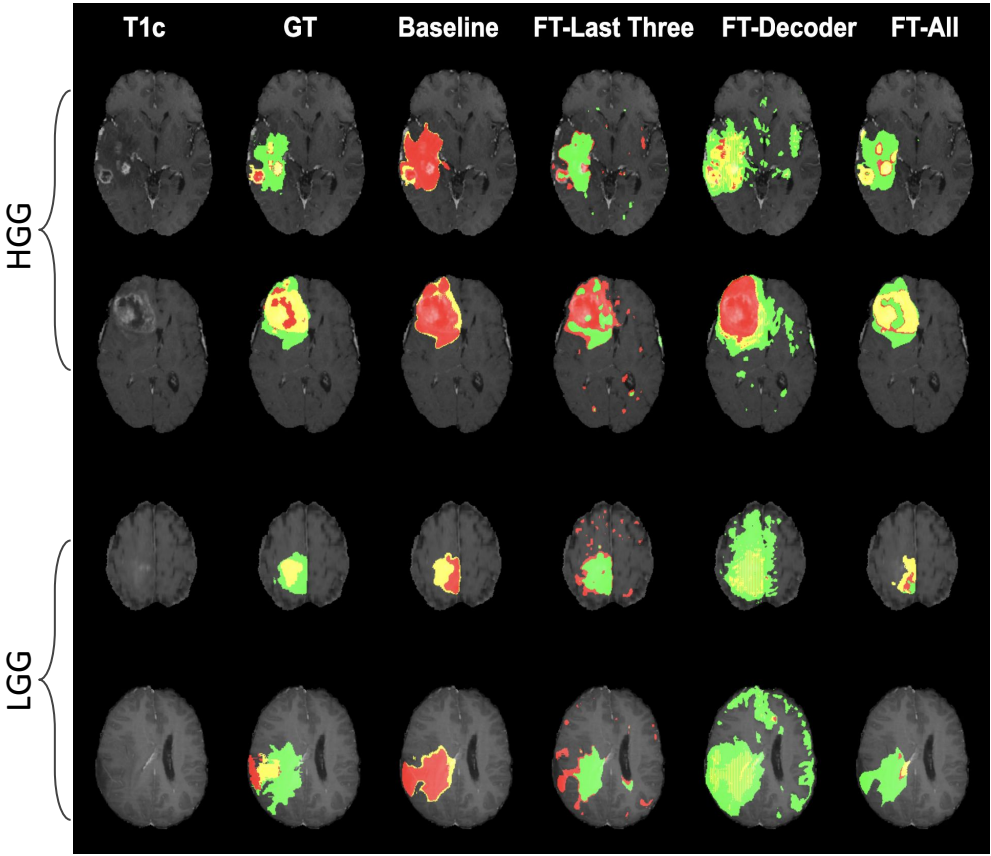


- **FT-All outperforms the baseline** in almost every case,
- Best when the **number of tumor cases is extremely low**, i.e. 20.
- As the number of brain tumor samples increase, **the gain of FT-All over baseline diminishes**.

Qualitative Results

- **FT-All is able to capture sub-structures of tumor** better than the other methods.

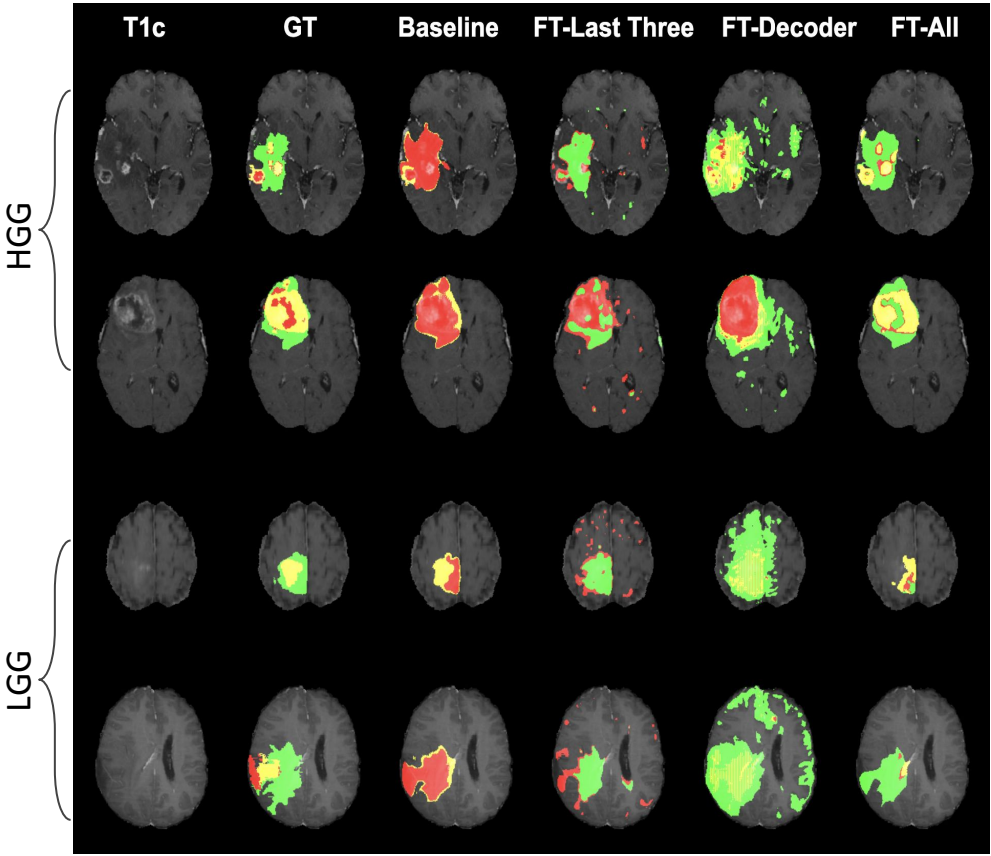
Fine-tuned with 20 brain tumor cases



Qualitative Results

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

Fine-tuned with 20 brain tumor cases



Conclusions and Discussions

McGill Centre for Intelligent Machines

- We explored different strategies for transfer learning across diseases for the task of focal pathology segmentation.

Conclusions and Discussions

- We explored different strategies for transfer learning across diseases for the task of focal pathology segmentation.
- We observed that **fine-tuning the whole network** works best, especially when **very small target datasets are available**.

Conclusions and Discussions

- We explored different strategies for transfer learning across diseases for the task of focal pathology segmentation.
- We observed that **fine-tuning the whole network** works best, especially when **very small target datasets are available**.
- We also observed that as in case of natural images, where fine-tuning just the last few layers works, **it's not the same case in medical domain**.

Conclusions and Discussions

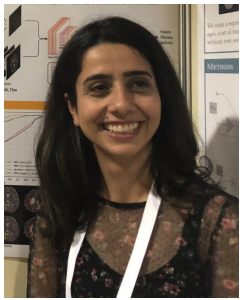
- We explored different strategies for transfer learning across diseases for the task of focal pathology segmentation.
- We observed that **fine-tuning the whole network** works best, especially when **very small target datasets are available**.
- We also observed that as in case of natural images, where fine-tuning just the last few layers works, **it's not the same case in medical domain**.
- We motivate **public release of models** trained on large datasets.

Acknowledgement

- Lab mates and supervisors: Prof Tal Arbel and Prof Doina Precup.








- Sponsors

NSERC CRSNG

INTERNATIONAL PROGRESSIVE MS ALLIANCE

CONNECT TO END PROGRESSIVE MS

Thank you for your patient listening!

Questions??

Data Preprocessing

McGill Centre for Intelligent Machines

Source: Multiple Sclerosis Dataset

- Brain extraction[2]
- N3 bias field inhomogeneity correction[3]
- Nyul image intensity normalization[4]
- Registration to the MNI-space.
- Intensity Normalization (mean subtraction, divide by standard deviation, re-mapping to 0-1)
- Cropped and zero-padded to 240x192x192.

Target: BraTS 2018 challenge Dataset [1]

- Skull stripping
- Co-registration
- Registration to same space as source data using ANTs tool[5]
- Intensity Normalization (mean subtraction, divide by standard deviation, re-mapping to 0-1)
- Cropped and zero-padded to 240x192x192.

[1] Menze et al, TMI 2015
[2] Smith et al, HBM 2002
[3] Sled et al TMI 1998
[4] Nyul et al TMI 2000
[5] Avants et al Neuroimage 2011