Propagating Uncertainty Across Cascaded Medical Imaging Tasks for Improved Deep Learning Inference

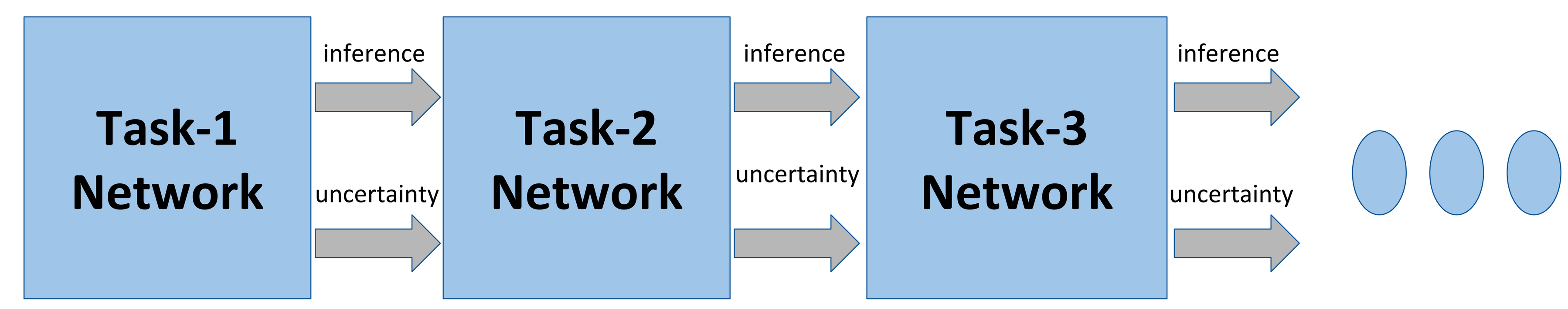
Raghav Mehta [1], Thomas Christinck [1], Tanya Nair [1], Paul Lemaitre [1], Douglas Arnold [2,3], and Tal Arbel [1]

[1] Centre for Intelligent Machines, McGill University, Montreal, Canada
[2] Montreal Neurological Institute, McGill University, Montreal, Canada
[3] NeuroRx Research, Montreal, Canada

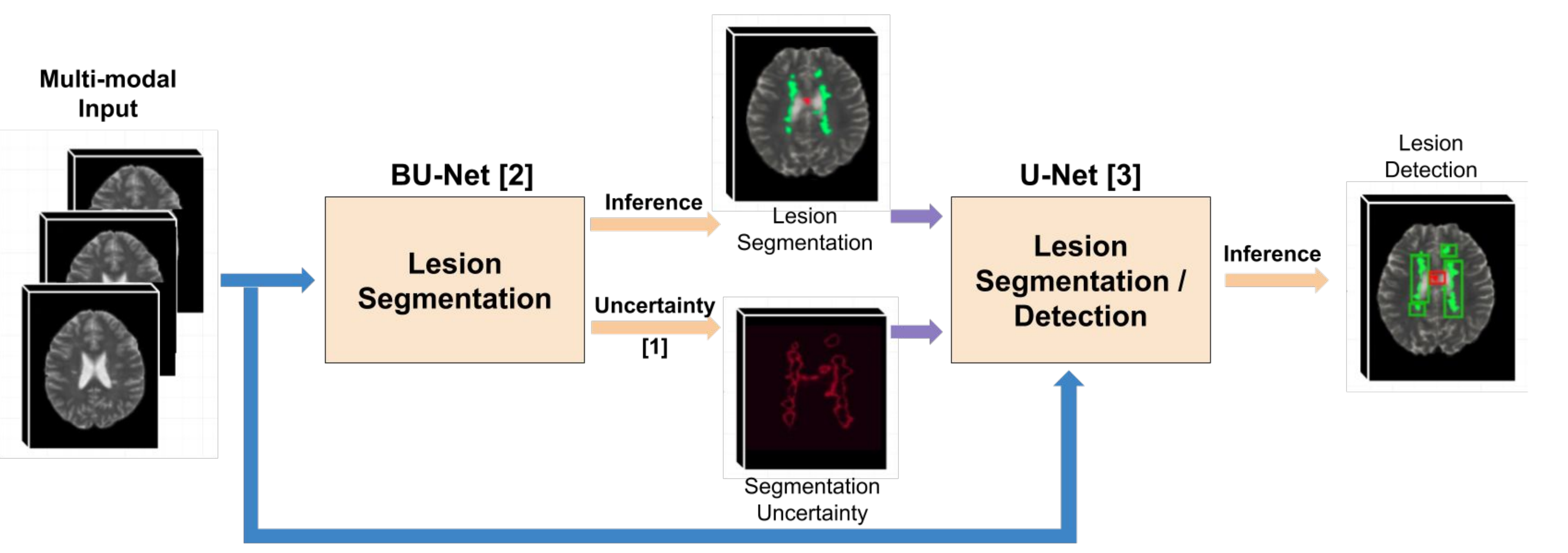
(1) Introduction

- Sequence of inference tasks in a medical image analysis pipeline
  - Registration
  - Skull-Stripping
  - Segmentation ...
- Errors in deterministic output can accumulate over sequential tasks
- **Hypothesis**: **Performance of the downstream task** can be **improved** by **propagating uncertainty** (e.g. MC-Dropout [1]) across sequential tasks

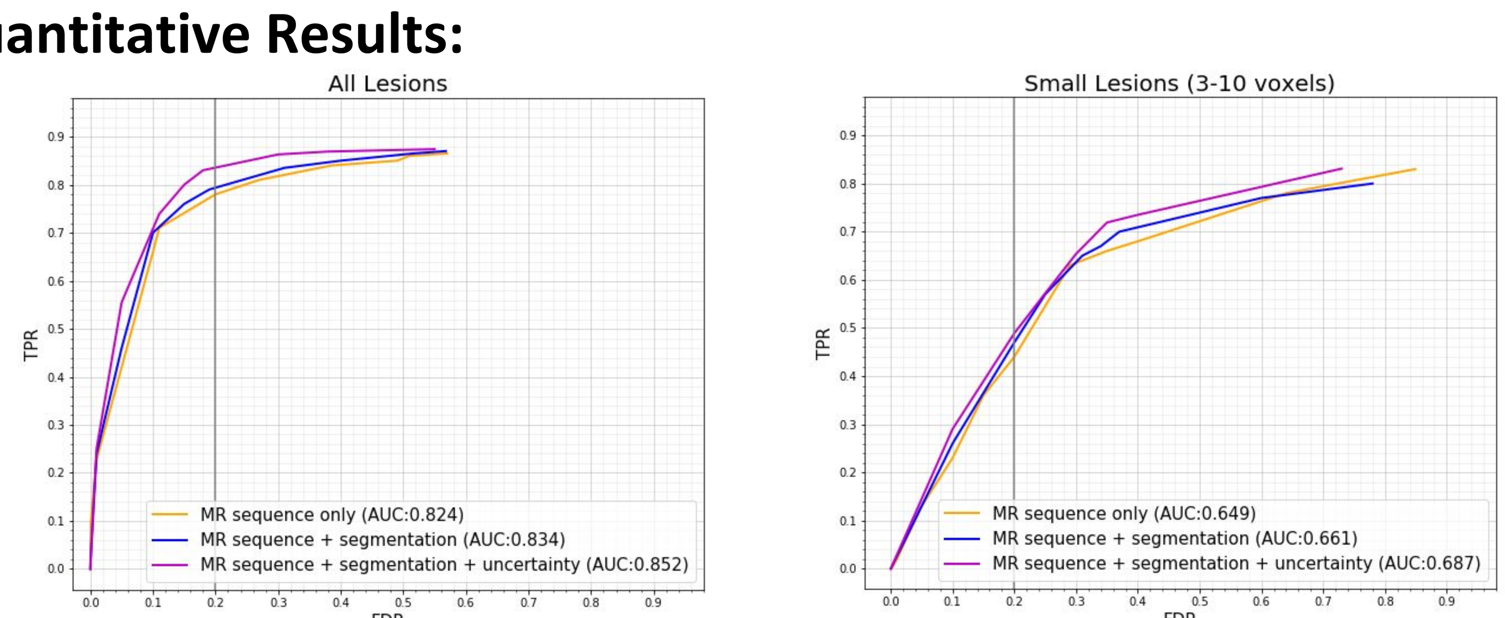
(2) Proposed Framework



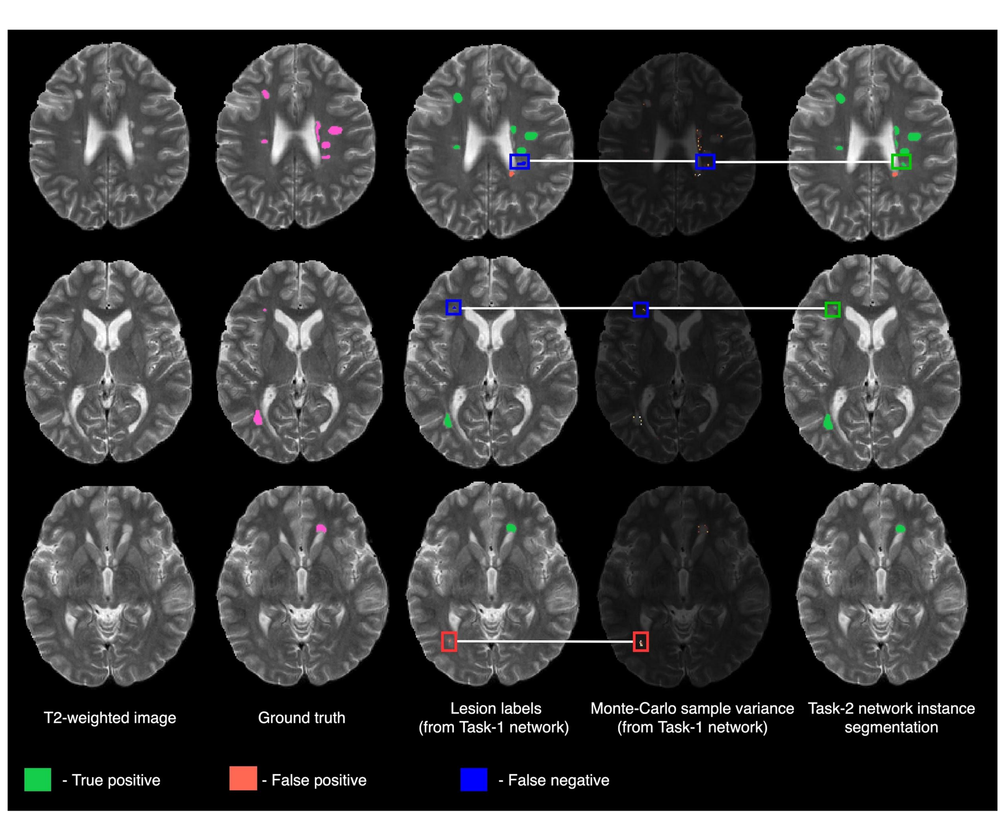
(3) MS T2 Lesion Segmentation/Detection Pipeline



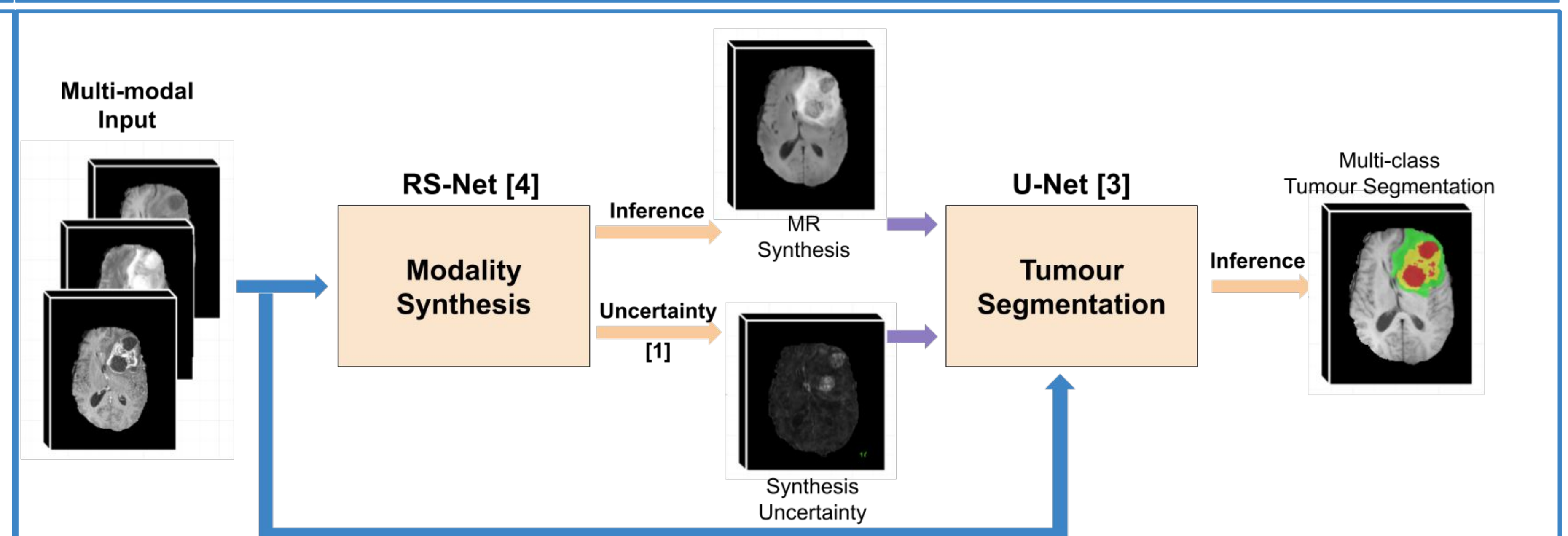
- **Dataset:**
  - Proprietary multisite, multi-scanner clinical relapsing-remitting MS (RRMS) trials
  - 5800 multi-modal MRI (T1, T2, FLAIR, and PD)
    - 40% of the available data to train BU-Net [2]
    - 50% of the remaining data to train 3D U-Net [3]
    - 10% held-out to test 3D U-Net [3]
- **Evaluation Metric:** ROC curves for lesion detection at various size
  - Segmentation converted into detection with connected component analysis
- **Quantitative Results:**



- **Qualitative Results:**



(4) Brain Tumour Segmentation Pipeline

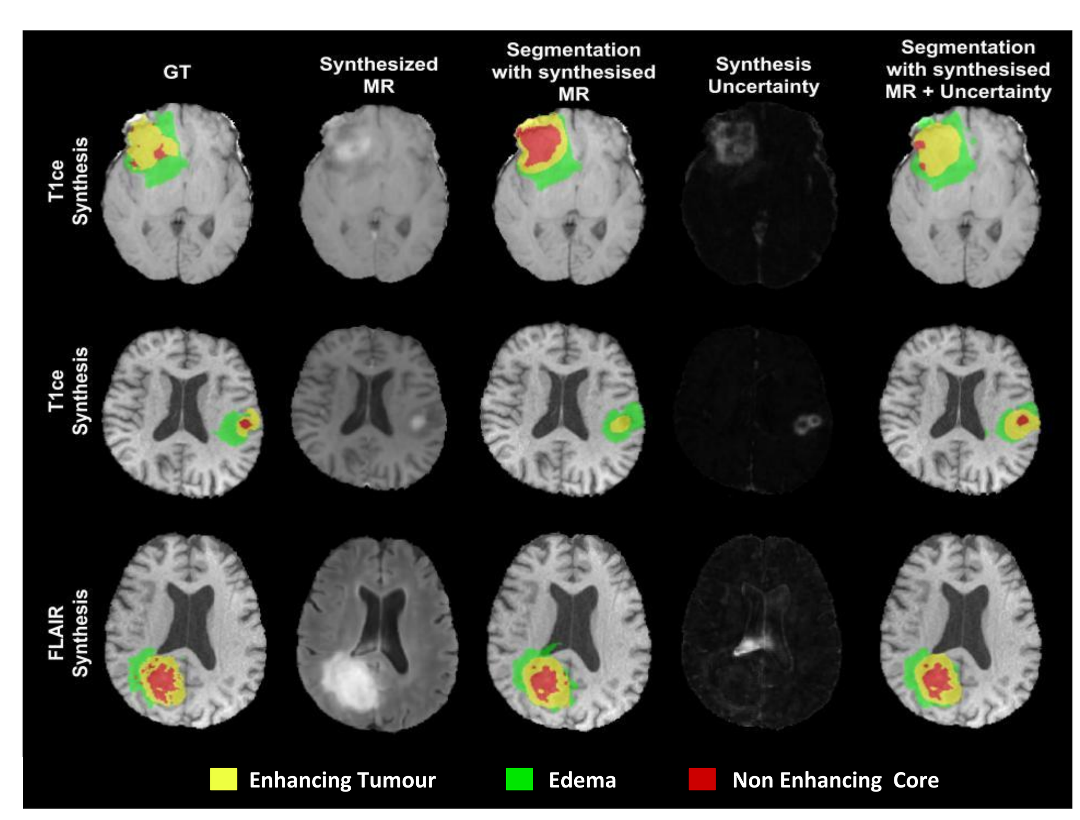


- **Dataset**:
  - Brain Tumour Segmentation (BraTS) 2018 [5] challenge dataset
  - Multi-modal MRI (T1, T2, FLAIR, and T1ce)
    - BraTS 2018 Training set to train and validate RS-Net [4] and 3D U-Net [3] (285 patients)
    - BraTS 2018 Validation set (held-out) to test 3D U-Net (66 patients)
- **Evaluation Metric**: Dice scores for three different tumour subtypes: enhancing tumour (DE ■), whole tumour (DT ■■■), and tumour core (DC ■■) [5]
- **Quantitative Results:**

	T1ce synthesis			FLAIR synthesis		
	DT	DC	DE	DT	DC	DE
real(3) sequences	87.17	50.25	26.89	83.27	73.91	71.07
real(3)+synthesized sequences	86.72	52.80	27.35	84.56	76.72	72.89
real(3)+synthesized+uncertainty	**88.20**	**57.29***	**32.86***	**85.84***	**79.25***	**74.51***

(*) indicates statistically significant (p ≤ 0.05) differences between second and third row.

- **Qualitative Results:**



(5) Conclusion

- Proposed **a general deep learning framework** for the propagation of uncertainty across a sequence of inference tasks within a medical image analysis pipeline for improved inference
- **Evaluation on two different contexts** of MS T2 lesion segmentation/detection and Brain Tumour segmentation
- **2-10% improvement** for both tasks on their respected **quantitative** measures
- **Clearly visible qualitative improvement**
- **Future work** will explore how to properly develop **a complete end-to-end system** that includes uncertainty propagation across the inference modules

Reference:
1) Y. Gal, Z. Ghahramani.: "Dropout as a Bayesian approximation: representing model uncertainty in deep learning", ICML 2016
2) T. Nair, et al.: "Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation", MICCAI 2018
3) O. Cicek, et al.: "3D U-Net: learning dense volumetric segmentation from sparse annotation", MICCAI 2016
4) R. Mehta, and T. Arbel: "RS-Net: regression-segmentation 3D CNN for synthesis of full resolution missing brain MRI in the presence of tumours", SASHIMI 2018
5) S. Bakas, et al.: "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge", arXiv preprint arXiv:1811.02629 (2018)

Acknowledgment: This work was supported by a Canadian Natural Science and Engineering Research Council (NSERC) Collaborative Research and Development Grant (CRDPJ 505357 - 16), Synaptive Medical, the Canadian NSERC Discovery and CREATE grants, and an award from the International Progressive MS Alliance (PA-1603-08175).