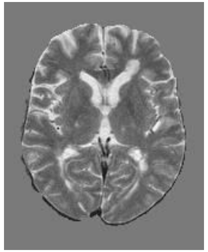
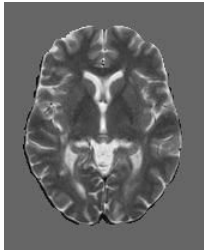
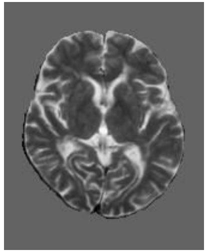
Cohort Bias Adaptation in Aggregated Datasets for Lesion Segmentation

Brennan Nichyporuk[1,2], Jillian Cardinell[1,2], Justin Szeto[1,2], Raghav Mehta[1,2], Sotirios A. Tsaftaris[3,4], Douglas L. Arnold[5,6], and Tal Arbel[1,2]

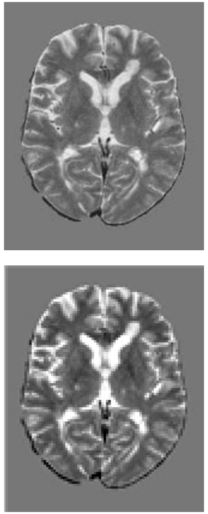
[1]Centre for Intelligent Machines, McGill University, Montreal, Canada
[2]MILA (Quebec Artificial Intelligence Institute), Montreal, Canada
[3]School of Engineering, University of Edinburgh, Edinburgh, UK
[4]The Alan Turing Institute, London, UK
[5]Department of Neurology and Neurosurgery, McGill University, Montreal, Canada
[6]NeuroRx Research, Montreal, Canada

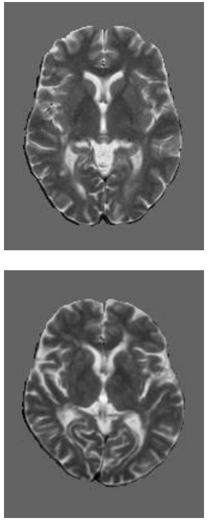
Motivation

- Deep learning methods for focal pathology segmentation and detection require large annotated datasets, which are not generally available.
- Common strategy to build a large dataset: aggregate multiple datasets together (‘naive pooling’)
  - May *decrease* performance due to cohort biases across datasets
- Goal: Train on multi-cohort dataset accounting for individual cohort biases
  - Improved inference results over naive pooling
  - Adaptation to new cohort biases with few samples

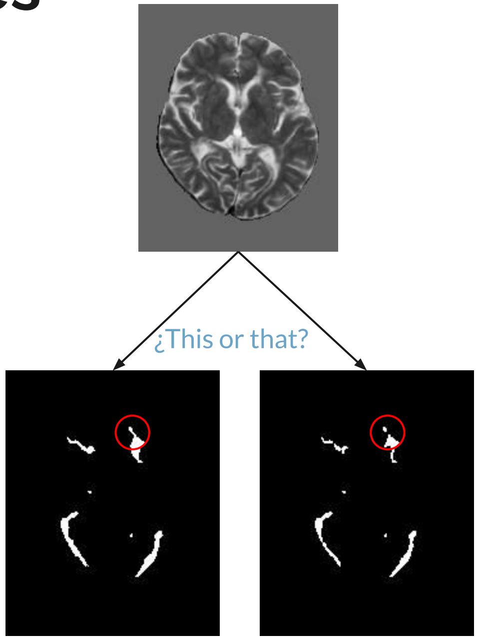
Sources of Cohort Biases



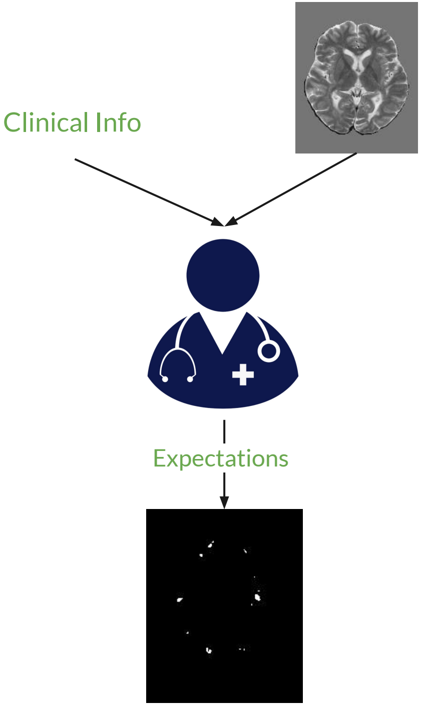
Different Acquisitions



Population Variability



Labelling Style



Observer Bias

Aggregating Datasets - Proposed Solution

Source-Conditioned Instance Normalization (SCIN):

- Condition network using cohort or source-specific instance normalization parameters

Overview

Case Study:

- Utilize SCIN for Multiple Sclerosis (MS) lesion segmentation and detection
- Cohorts: Several MS clinical trials datasets

Experiments show SCIN can:

- Strategically pool diverse datasets by learning a cohort-specific bias
- Adapt to new cohort bias by fine-tuning SCIN parameters on a few samples
- Model complex cohort biases: e.g. rater ignores small lesions

Methodology

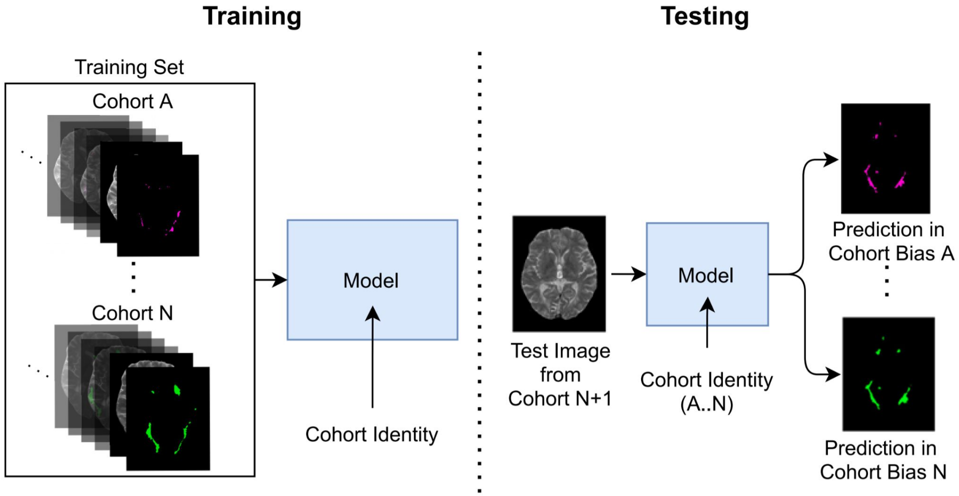
Overview

Training

- Train on multi-cohort dataset
- Use SCIN to condition on cohort identity

Testing

- Provide cohort identity
- Get prediction output with bias corresponding to cohort



Methodology

Conditioning

- Conditional Instance Normalization (CIN) [1]
- Source-specific instance normalization parameters γ_s and β_s model cohort biases

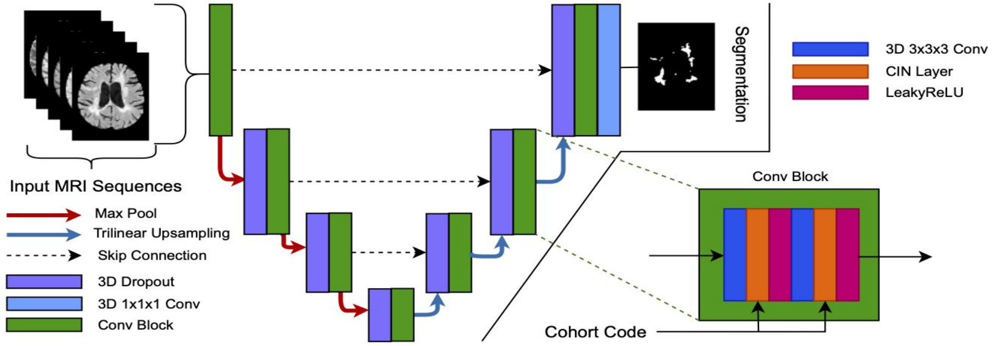
$$\mathrm{CIN}(z) = \gamma_s \left(\frac{z - \mu(z)}{\sigma(z)} \right) + \beta_s$$

[1] Vincent D., et al. 2017

Methodology

Architecture

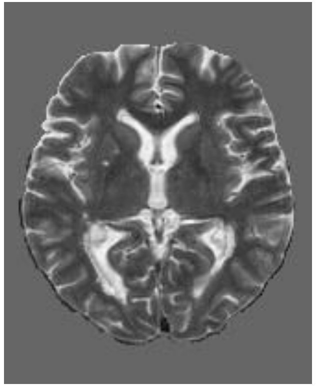
- Learn cohort-specific biases by conditioning on cohort identity
- Learn cohort-specific instance normalization parameters

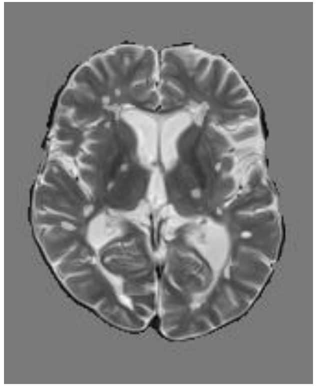


Experiments

Data - Cohorts

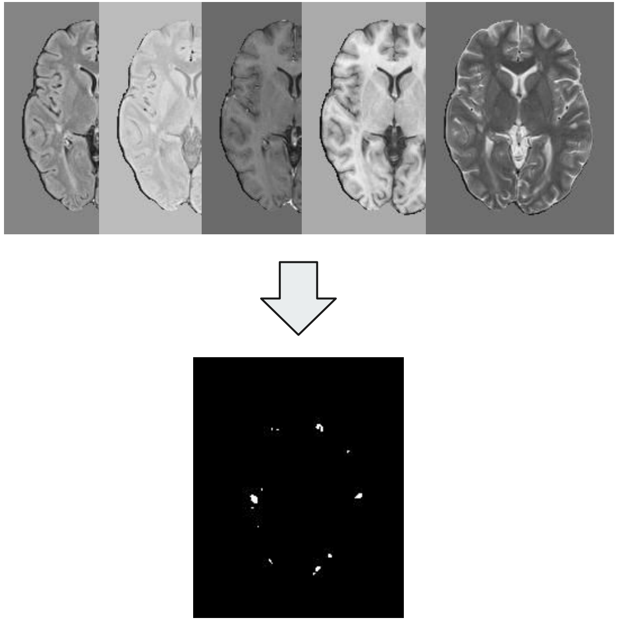
- Trial-A (2011-2015): Late-stage Secondary-Progressive (SPMS), 1000 Samples
- Trial-B (2008-2011): Relapsing Remitting (RRMS), 1000 Samples
- Trial-C (2004-2009): Early-stage SPMS, 500 Samples





Experiments:
Data

- MRI Sequences
  - FLAIR, PDW, T2, T1, and Gadolinium Enhanced T1
- T2 lesion segmentation and detection



Results

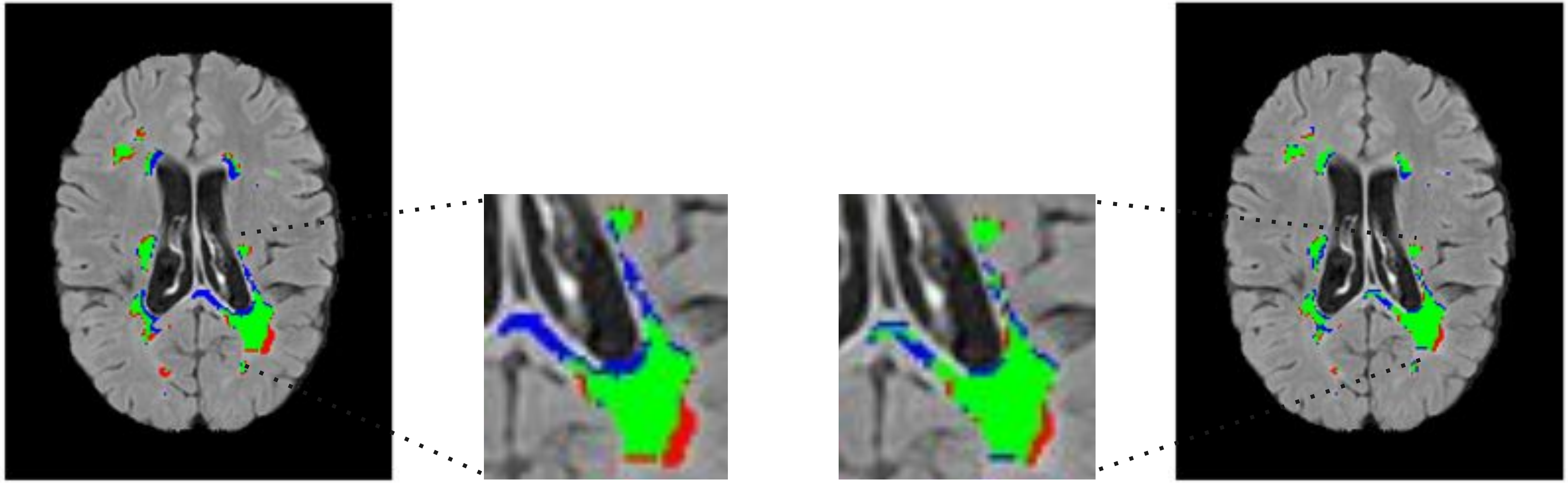
Experiment 1

- Naive pooling decreases performance on individual trials
- SCIN-Pooling achieved a higher DICE compared to Naive-Pooling

#	Model	Train Set		Conditioned On		Test Performance	
		Trial-A	Trial-B	Trial-A	Trial-B	Trial-A	Trial-B
1	Single-Trial	✔		-		0.793	0.689
2	Single-Trial		✔	-		0.715	0.803
3	Naive-Pooling	✔	✔	-		0.789	0.748
4	SCIN-Pooling	✔	✔	✔		0.794	0.700
5					✔	0.725	0.797

Results

Experiment 1



Cond.: Trial-A

Green: True Positive, Red: False Positive, Blue: False Negative

Cond.: Trial-B

Results

Experiment 2

- Naive Pooling Experiment: Train on A&B, Fine-tune only IN parameters on 10 samples
- SCIN Pooling Experiment: Fine-tuned SCIN-pooling model achieves best DICE on new Trial C

#	Model	Fine-Tuned On Trial-C	Conditioned On			Test Performance (Trial-C)
			Trial-A	Trial-B	Trial-C	
1	Naive-Pooling		-			0.774
2		✔	-			0.819
3	SCIN-Pooling		✔			0.763
4				✔		0.806
5		✔			✔	0.834

Results
Experiment 3

- Simulate cohort bias where small lesions are missed (i.e. not labeled)
  - Create Missed Small Lesion (MSL) Trial
- DICE Relative to Single Trial Model: Naive-pooling suffers, **SCIN-pooling improved**
- Detection F1 Relative to Single Trial Model: SCIN successfully learns the cohort bias of MSL

<table>
<tr><th rowspan="2">#</th><th rowspan="2">Model</th><th colspan="2">Train Set</th><th colspan="2">Conditioned On</th><th colspan="2">Test Performance (Trial-Orig)</th></tr>
<tr><th>Trial-Orig</th><th>Trial-MSL</th><th>Trial-Orig</th><th>Trial-MSL</th><th>Sm Lesion F1</th><th>Voxel DICE</th></tr>
<tr><td>1</td><td>Single-Trial</td><td>✔</td><td></td><td colspan="2">-</td><td>0.795</td><td>0.844</td></tr>
<tr><td>2</td><td>Single-Trial</td><td></td><td>✔</td><td colspan="2">-</td><td>0.419</td><td>0.837</td></tr>
<tr><td>3</td><td>Naive-Pooling</td><td>✔</td><td>✔</td><td colspan="2">-</td><td>0.790</td><td>0.797</td></tr>
<tr><td>4</td><td rowspan="2">SCIN-Pooling</td><td rowspan="2">✔</td><td rowspan="2">✔</td><td>✔</td><td></td><td>0.784</td><td>0.854</td></tr>
<tr><td>5</td><td></td><td>✔</td><td>0.496</td><td>0.850</td></tr>
</table>

Conclusions

- SCIN enables training on aggregated datasets by accounting for individual cohort biases
- SCIN can be used to adapt to a new cohort using few samples

CIFAR | ICRA
CANADIAN INSTITUTE FOR ADVANCED RESEARCH | INSTITUT CANADIEN DE RECHERCHES AVANCÉES

Mila

CIM CENTRE FOR INTELLIGENT MACHINES

THE UNIVERSITY OF EDINBURGH

The Alan Turing Institute

INTERNATIONAL PROGRESSIVE MS ALLIANCE
CONNECT TO END PROGRESSIVE MS

NSERC CRSNG

neuro
Montreal Neurological Institute-Hospital

Thank you and please join us for our discussion and poster session!