

Efficient Deep Learning of 3D Structural Brain MRIs for Manifold Learning and Lesion Segmentation with Application to MS

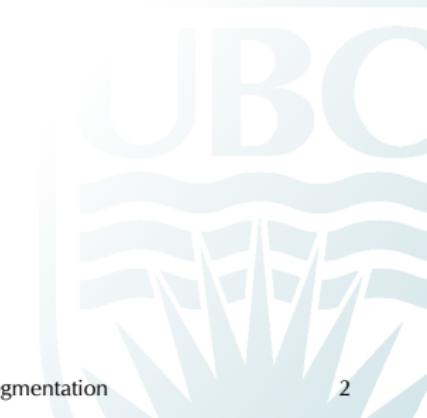
Tom Brosch

UBC MS/MRI Research Group
Electrical and Computer Engineering
University of British Columbia

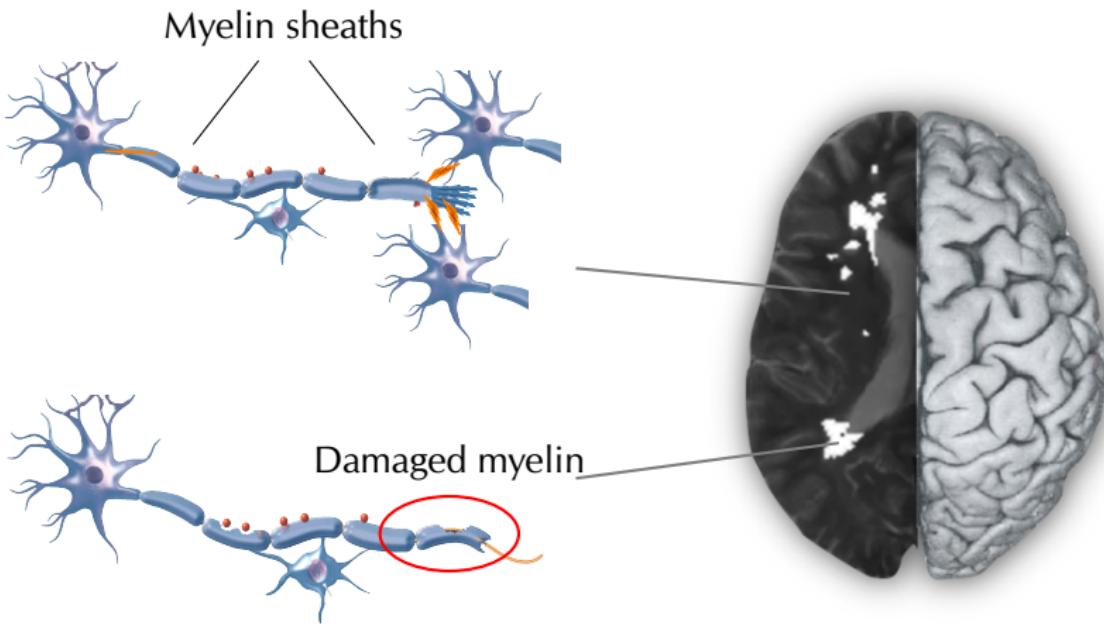
PhD Defence

Outline

- 1 Introduction
- 2 Efficient Deep Learning in the Frequency Domain
 - Training Algorithm
 - Evaluation
- 3 MS Lesion Segmentation
 - Convolutional Encoder Networks
 - Evaluation
- 4 Manifold Learning
 - Manifold Learning of Brain MRIs
 - Modelling the Variability of MS Brains
- 5 Conclusions

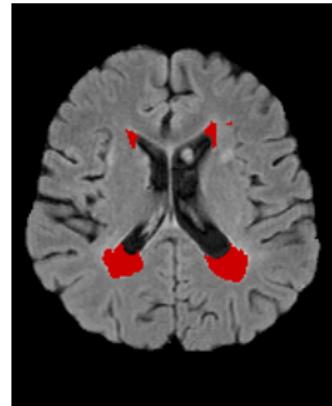
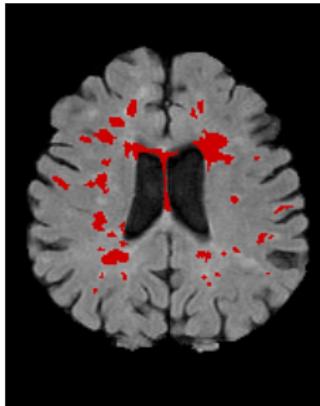
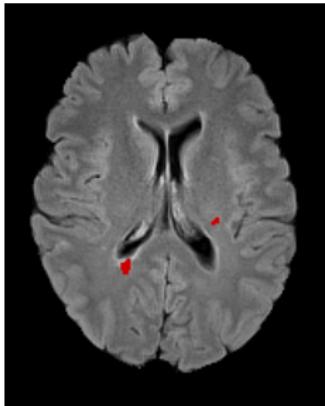


Multiple Sclerosis



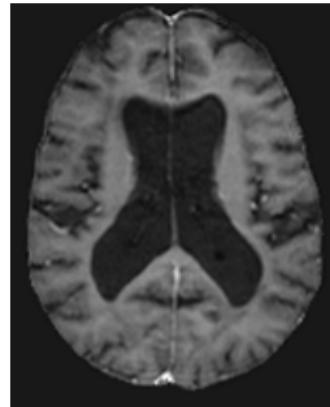
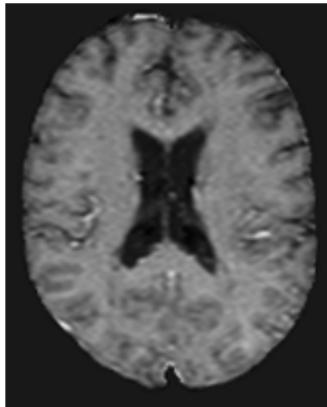
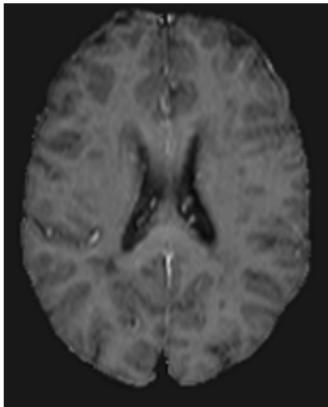
Multiple Sclerosis

- MS is a very heterogeneous disease
- MS lesions vary greatly in number, size, and location



Multiple Sclerosis

- MS is a very heterogeneous disease
- MS lesions vary greatly in number, size, and location
- MS also shows varying degrees of atrophy, most visible by an enlargement of the ventricles



Motivation and Challenges

Clinical motivation:

- **To measure disease state and progression**
- To automatically and accurately segmented MS lesions in order to derive lesion-based biomarkers such as lesion volume and lesion count
- To develop a method that can automatically discover patterns of variability in brain morphology and lesion distribution

Challenges:

- Large anatomical and pathological variability
- Large variability in contrasts produced by different scanners
- Patterns of morphological variability and lesion distribution are highly nonlinear

Objectives

Objectives

To develop deep learning methods for neuroimage analysis and to measure their performance.

Motivation:

- Can learn features that are robust to large variability
- Can discover highly nonlinear patterns of variability

Technical challenges:

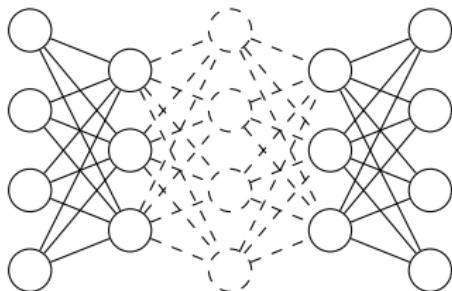
- Relatively small size of medical data sets
- High dimensionality of 3D medical images

Contributions

- Developed a computation and memory efficient training algorithm for convolutional deep learning models
- Developed a novel segmentation method that learns features at different scales that are tuned to a given combination of image types and segmentation task
- Proposed a novel objective function for neural networks that facilitates the training on vastly unbalanced training sets
- First work to demonstrate the use of deep learning for manifold learning of 3D medical images
- Developed a framework for modelling changes in brain morphology and lesion distribution

Introduction to Deep Learning

Input layer Hidden layers Output layer



- Multiple layers of nonlinear processing units for feature extraction
- Hierarchy of low-level to high-level features

Neural Networks

Supervised
Computational graph
Function approximation
Minimize the prediction error

Deep Belief Networks

Unsupervised
Probabilistic graphical model
Feature learning
Maximize the data likelihood

[Werbos, 1974; Hinton et al., 2006]

Introduction to Deep Learning

Fully connected models:

- Number of weights increases quadratically with the number of voxels
- Number of trainable weights too high for application to 3D volumes

Convolutional models:

- Number of weights greatly reduced due to local connectivity and weight sharing
- Strided convolutions further reduce the number of hidden units
- Scale much more efficiently with image resolution
- Calculation of convolutions computationally demanding

[LeCun et al., 1998]

Training of Convolutional Models

Observation

The computational bottleneck is the calculation of $O(NFC)$ convolutions.

Approach 1: Naïve FFT-based implementation

- Replaces convolutions with FFTs and element-wise multiplications
- Requires $O(NFC)$ FFT calculations

Approach 2: Training in the Frequency Domain

- Maps all operations to frequency domain, when possible
- Only calculates operations in spatial domain that cannot be calculated in the frequency domain
- Requires $O(NC + FC + NF)$ FFT calculations

[Brosch and Tam, Neural Computation, 2015]

Evaluation

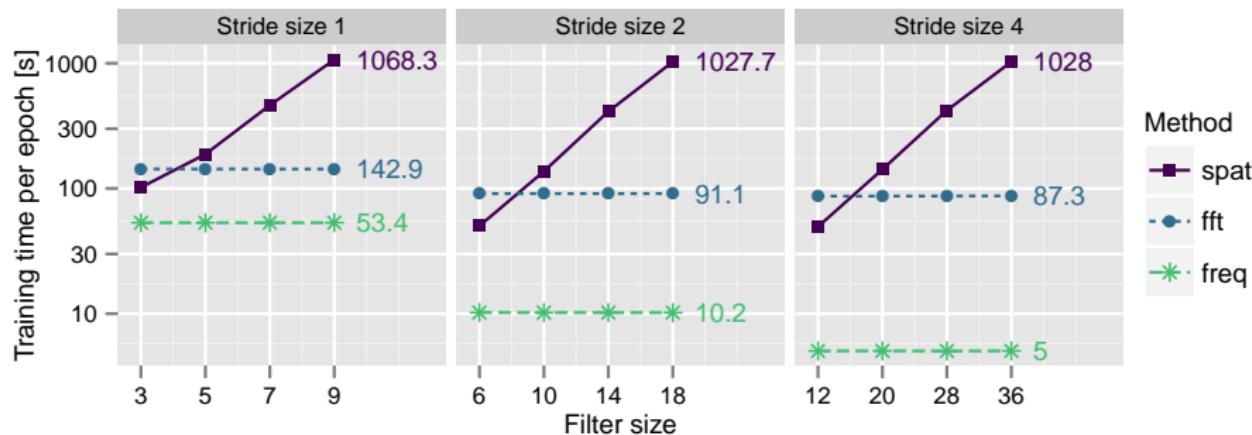
Data set:

- 100 T1-weighted MRIs of the brain from the OASIS data set
- Number of voxels: $128 \times 128 \times 128$
- Voxel size: $2 \times 2 \times 2 \text{ mm}$

Trained a strided convolutional RBM with varying parameters:

- Filter size
- Stride size

Evaluation



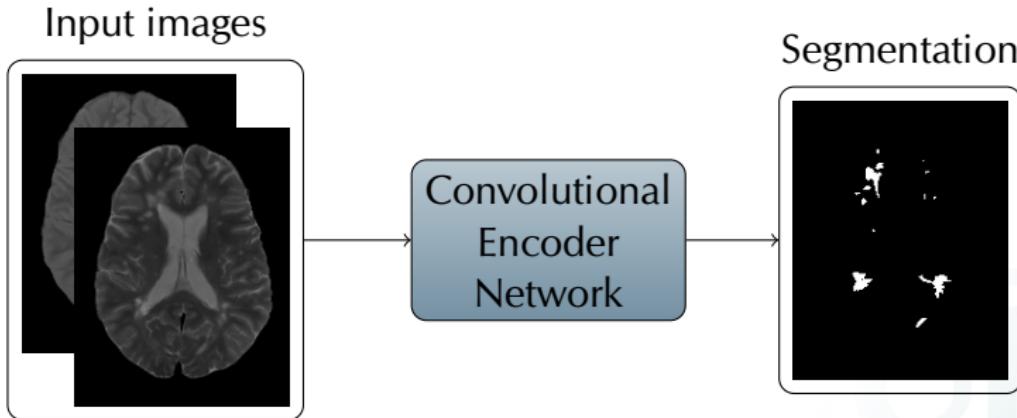
- Up to 200 times faster than training in the spatial domain
- Up to 17 times faster than a naïve FFT-based implementation
- More than **6 times faster** than cuDNN for the lesion segmentation network

[Brosch and Tam, Neural Computation, 2015]

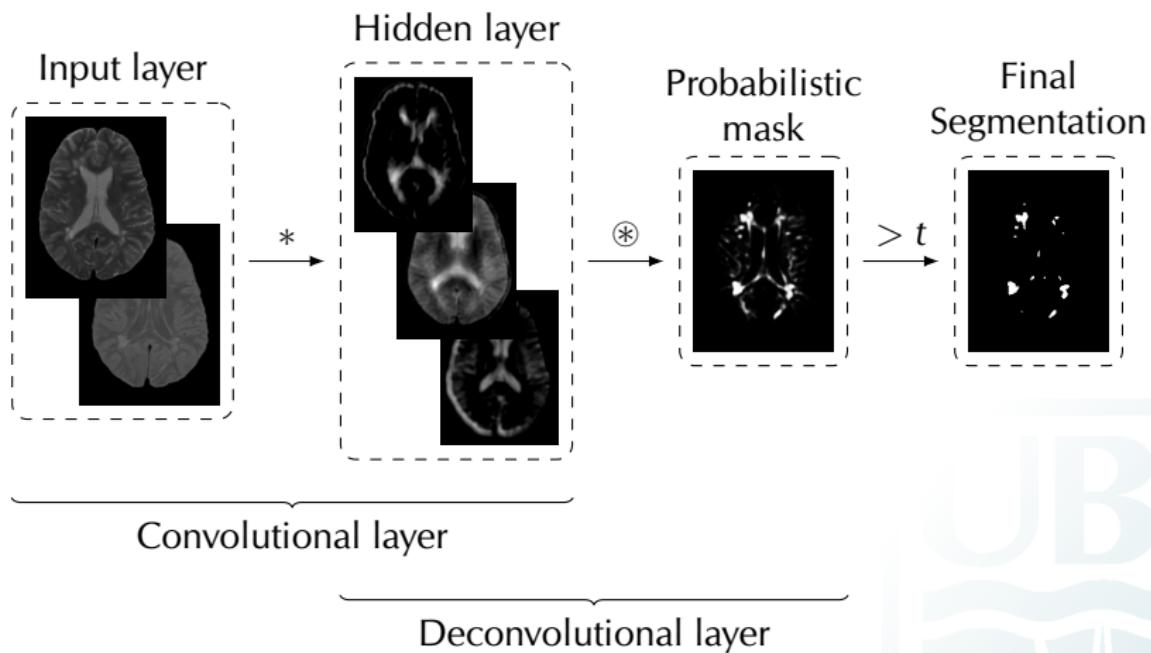
Lesion Segmentation

- Developed a computation and memory efficient training algorithm for convolutional deep learning models
- Developed a novel segmentation method that learns features at different scales that are tuned to a given combination of image types and segmentation task
- Proposed a novel objective function that facilitates the training using vastly unbalanced training sets, such as is the case for segmenting MS lesions
- First work to demonstrate the use of deep learning for manifold learning of 3D medical images
- Developed a framework for modelling changes in brain morphology and lesion distribution

Lesion Segmentation



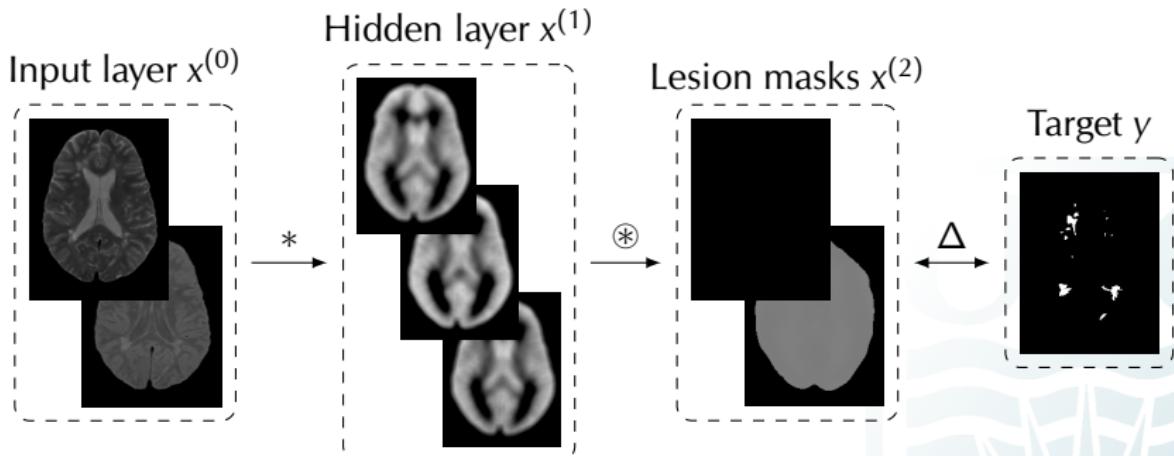
Convolutional Encoder Network: Architecture



[Brosch et al., MICCAI 2015]

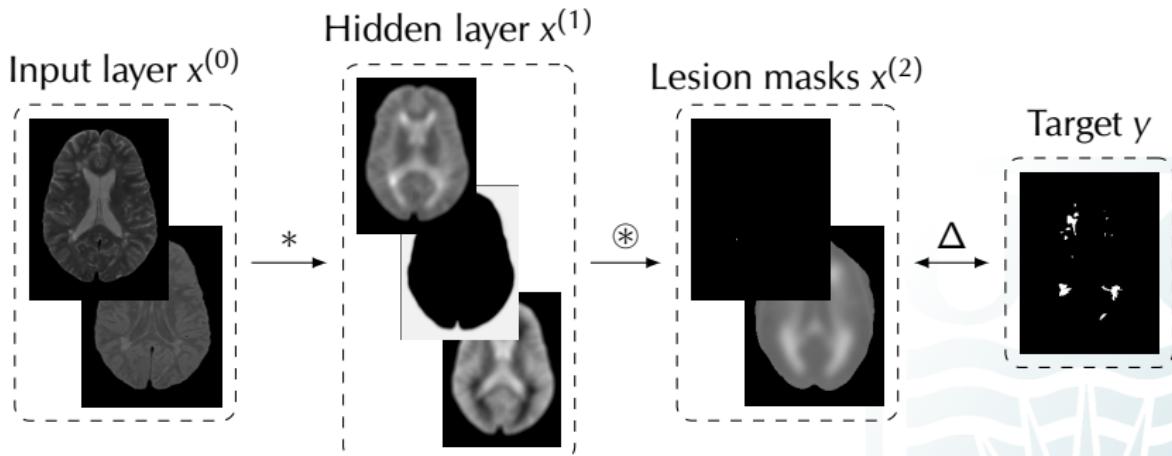
Convolutional Encoder Network: Training

- 1 Initialize filters with random values
- 2 Calculate initial guess of the segmentation
- 3 Optimize filters using stochastic gradient descent



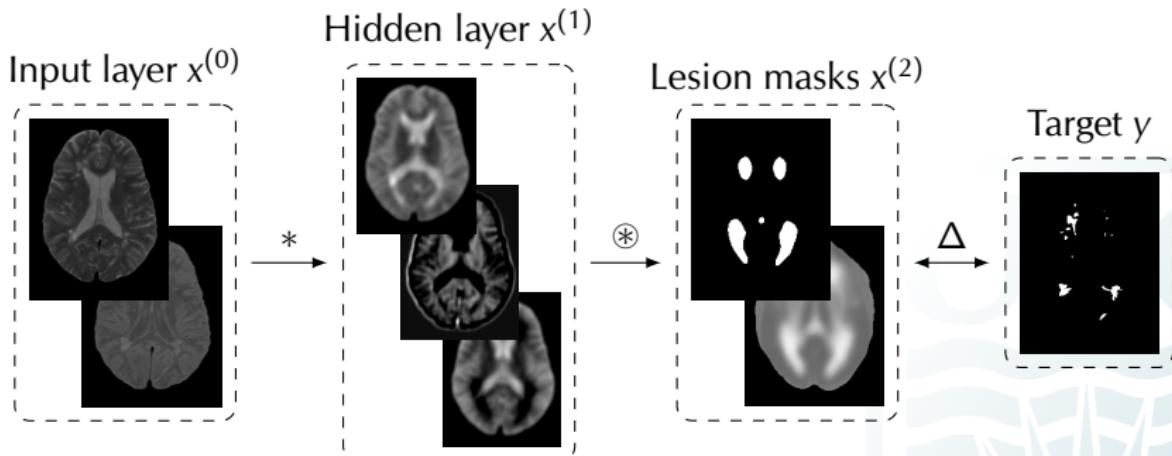
Convolutional Encoder Network: Training

- 1 Initialize filters with random values
- 2 Calculate initial guess of the segmentation
- 3 Optimize filters using stochastic gradient descent



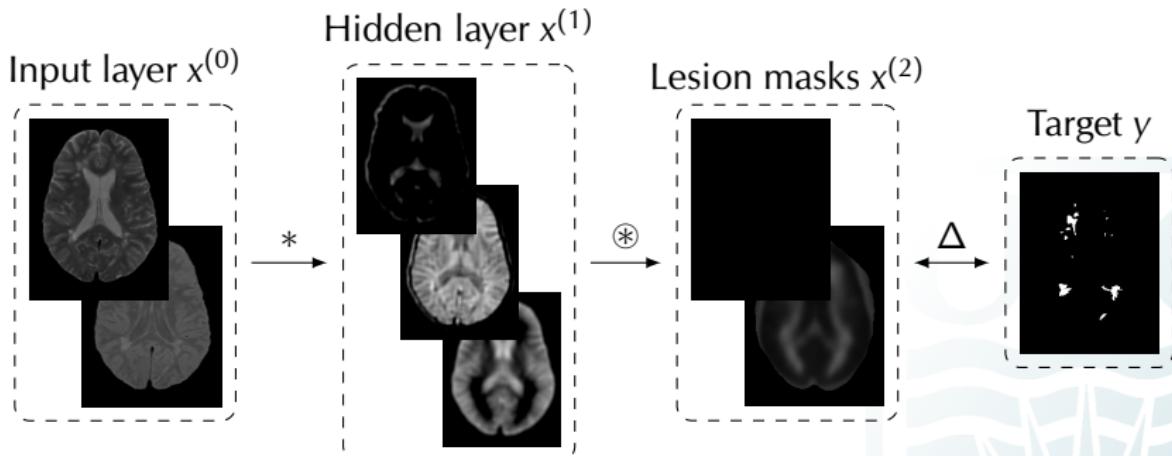
Convolutional Encoder Network: Training

- 1 Initialize filters with random values
- 2 Calculate initial guess of the segmentation
- 3 Optimize filters using stochastic gradient descent



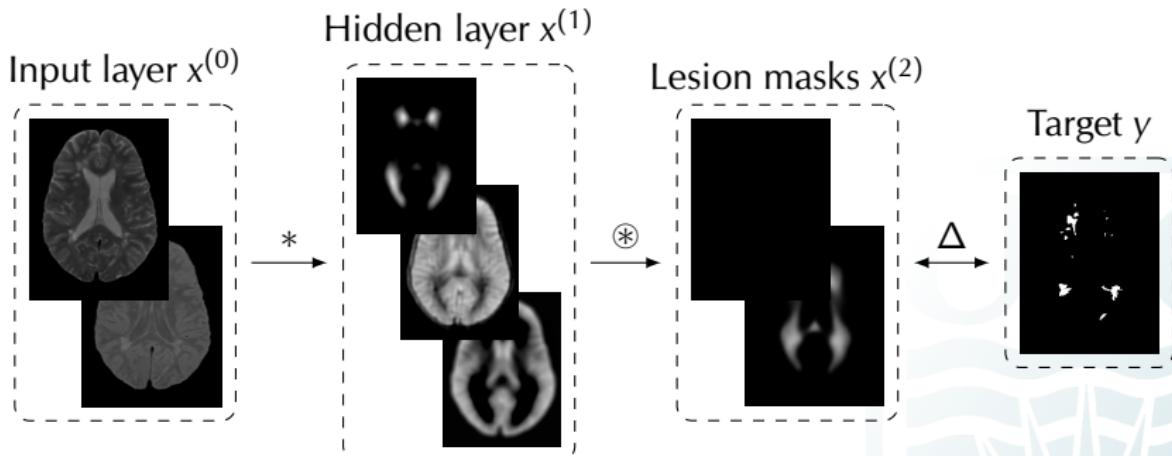
Convolutional Encoder Network: Training

- 1 Initialize filters with random values
- 2 Calculate initial guess of the segmentation
- 3 Optimize filters using stochastic gradient descent



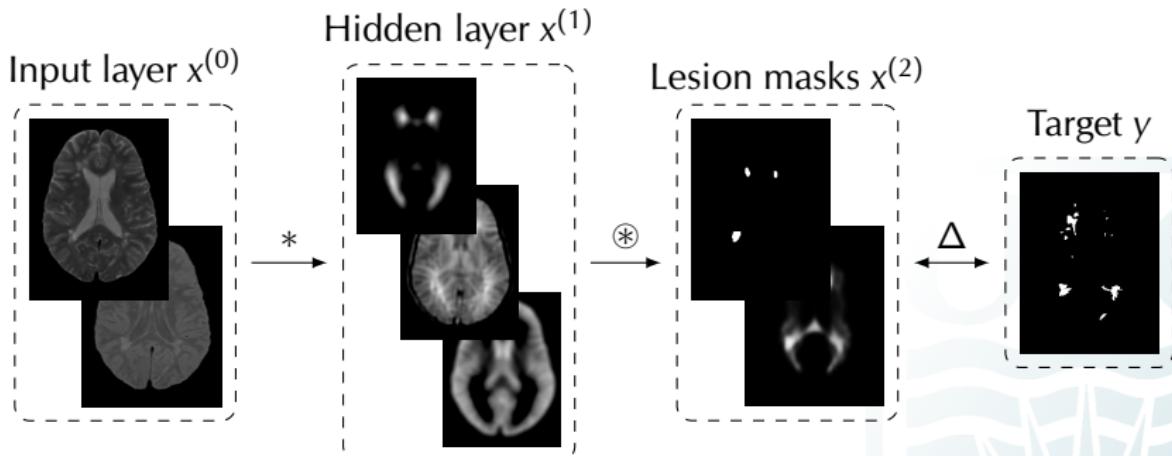
Convolutional Encoder Network: Training

- 1 Initialize filters with random values
- 2 Calculate initial guess of the segmentation
- 3 Optimize filters using stochastic gradient descent



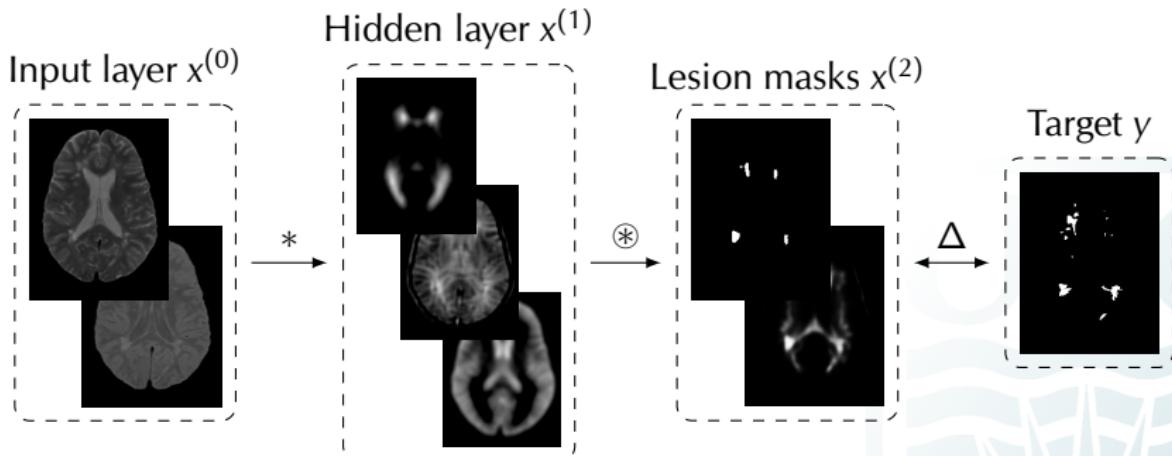
Convolutional Encoder Network: Training

- 1 Initialize filters with random values
- 2 Calculate initial guess of the segmentation
- 3 Optimize filters using stochastic gradient descent



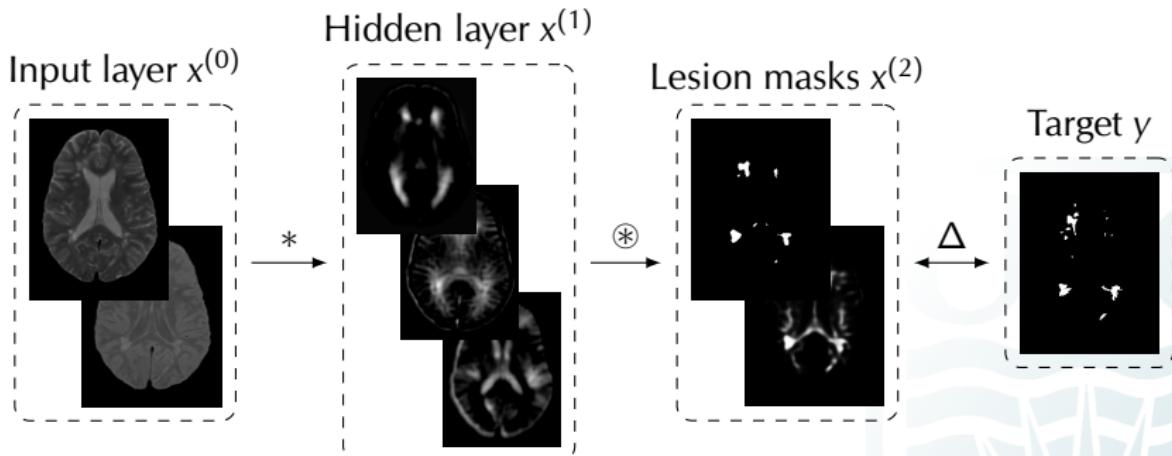
Convolutional Encoder Network: Training

- 1 Initialize filters with random values
- 2 Calculate initial guess of the segmentation
- 3 Optimize filters using stochastic gradient descent



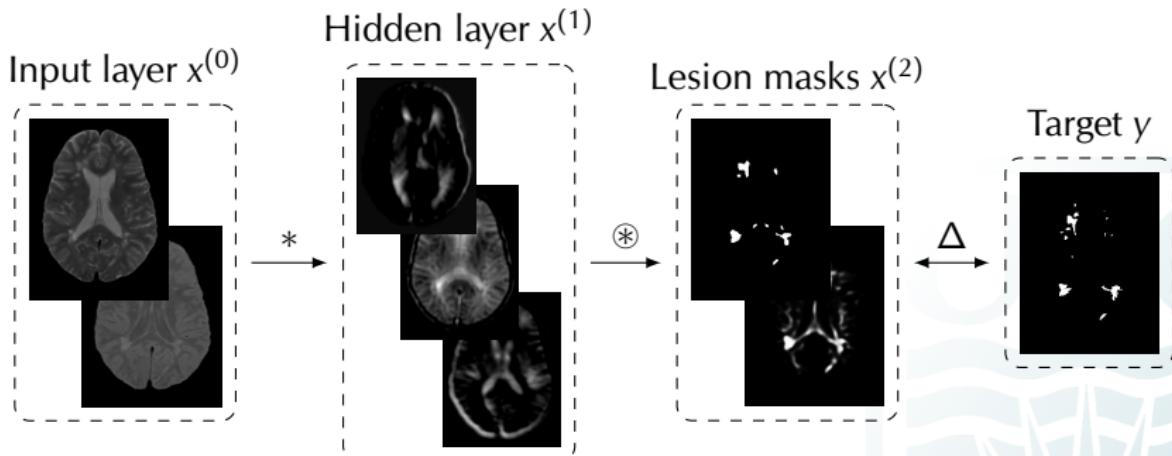
Convolutional Encoder Network: Training

- 1 Initialize filters with random values
- 2 Calculate initial guess of the segmentation
- 3 Optimize filters using stochastic gradient descent



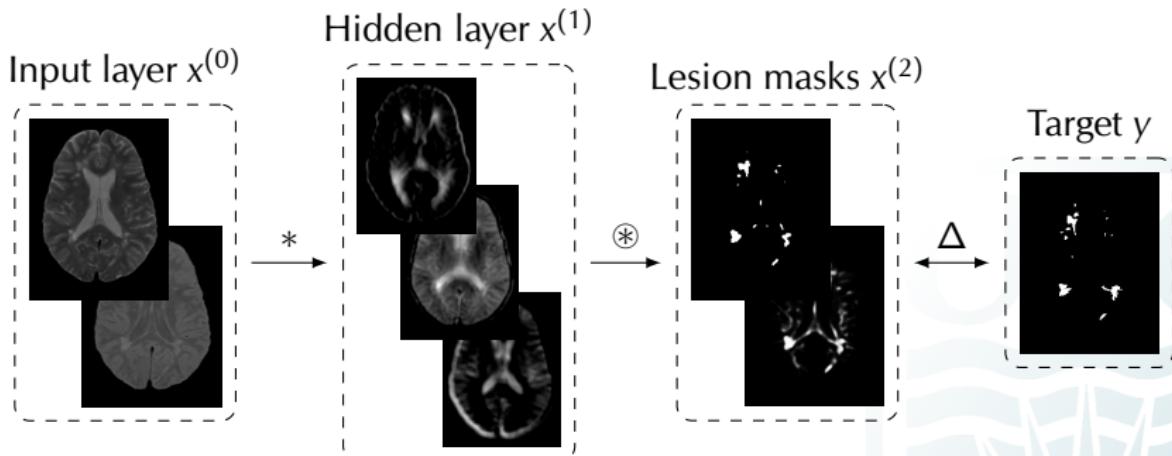
Convolutional Encoder Network: Training

- 1 Initialize filters with random values
- 2 Calculate initial guess of the segmentation
- 3 Optimize filters using stochastic gradient descent



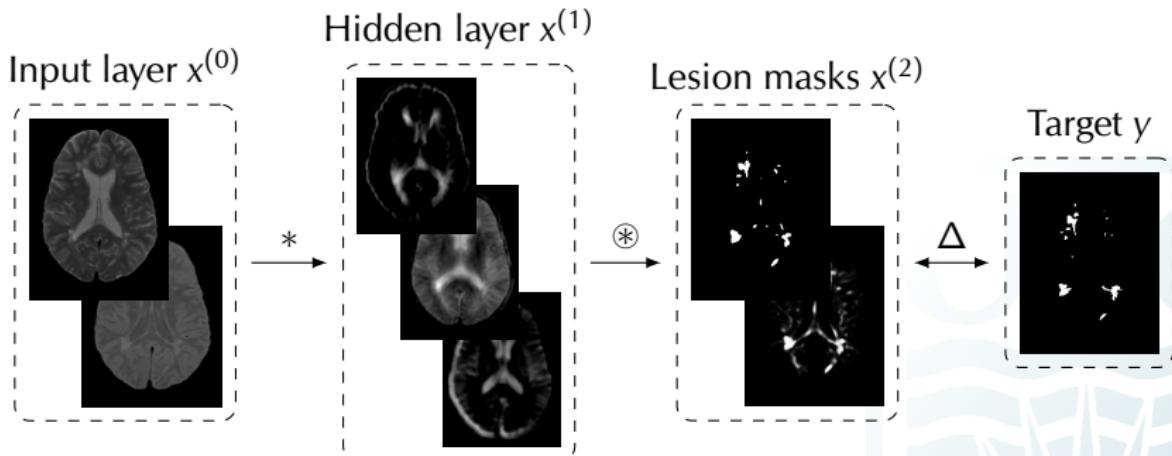
Convolutional Encoder Network: Training

- 1 Initialize filters with random values
- 2 Calculate initial guess of the segmentation
- 3 Optimize filters using stochastic gradient descent



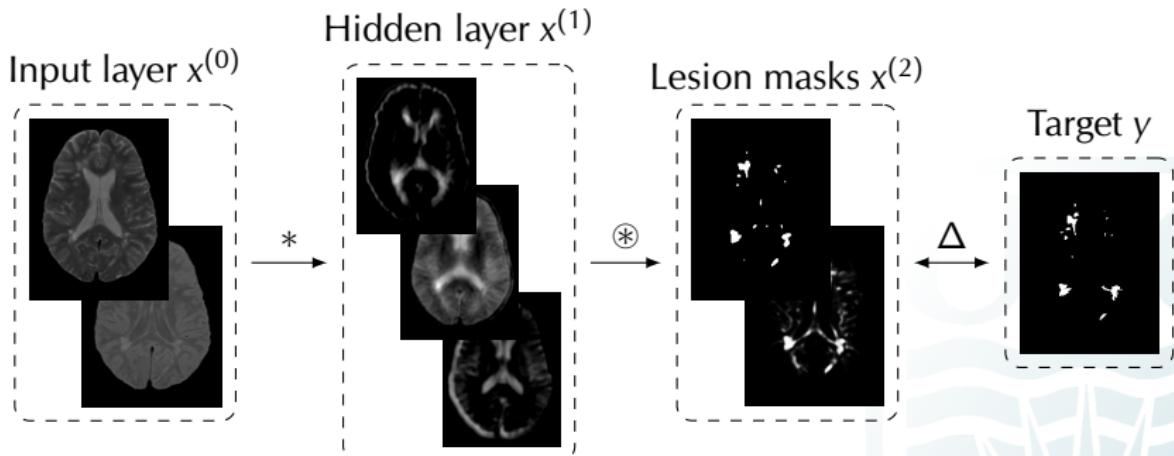
Convolutional Encoder Network: Training

- 1 Initialize filters with random values
- 2 Calculate initial guess of the segmentation
- 3 Optimize filters using stochastic gradient descent



Convolutional Encoder Network: Training

- 1 Initialize filters with random values
- 2 Calculate initial guess of the segmentation
- 3 Optimize filters using stochastic gradient descent



Objective Functions

Sum of squared difference:

- $\Delta = \frac{1}{2} \sum_{\mathbf{p}} (y(\mathbf{p}) - x^{(2)}(\mathbf{p}))^2$
- Assigns same importance to every voxel
- Can learn to ignore the minority class

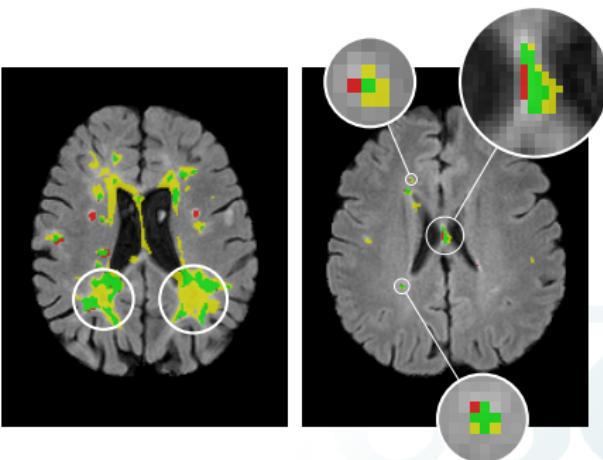
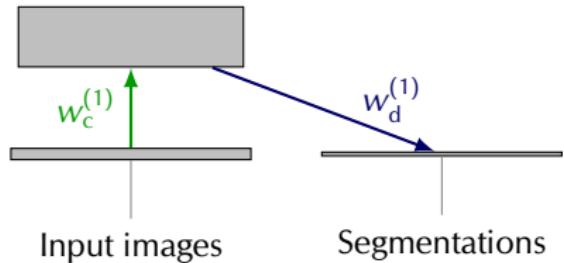
Combination of **sensitivity** and **specificity**:

$$\bullet \Delta = r \frac{\sum_{\mathbf{p}} (y(\mathbf{p}) - x^{(2)}(\mathbf{p}))^2 y(\mathbf{p})}{\sum_{\mathbf{p}} y(\mathbf{p})} + (1 - r) \frac{\sum_{\mathbf{p}} (y(\mathbf{p}) - x^{(2)}(\mathbf{p}))^2 (1 - y(\mathbf{p}))}{\sum_{\mathbf{p}} (1 - y(\mathbf{p}))}$$

- Accuracies for both classes are calculated separately and then combined to yield a single score
- Formulated as squared error terms to yield smooth gradients

[Brosch et al., MICCAI 2015]

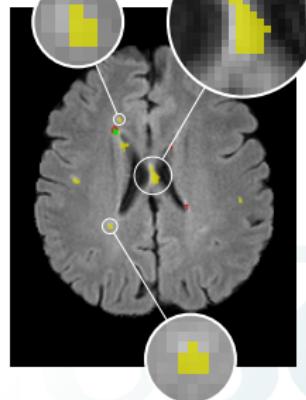
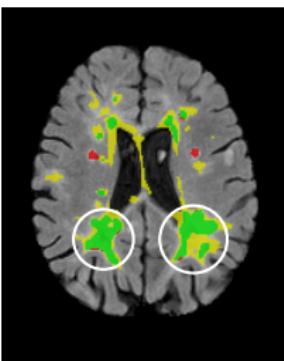
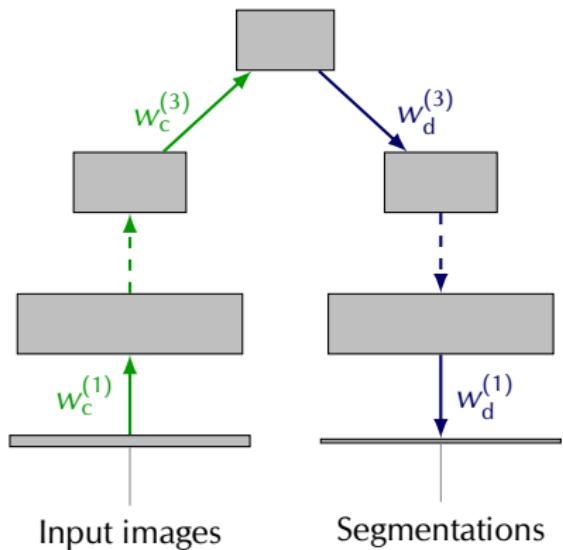
Network Architecture: 3-layer CEN



Detected lesions
Missed lesions
False positives

[Brosch et al., TMI 2016]

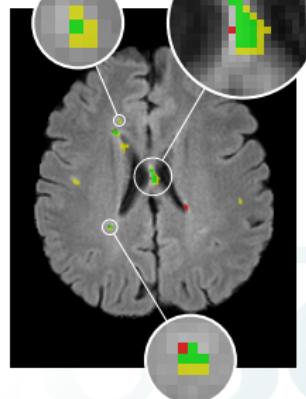
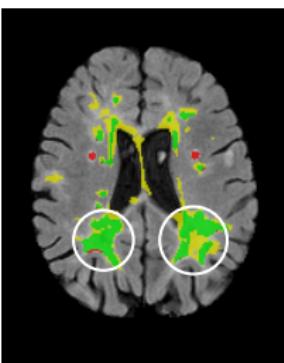
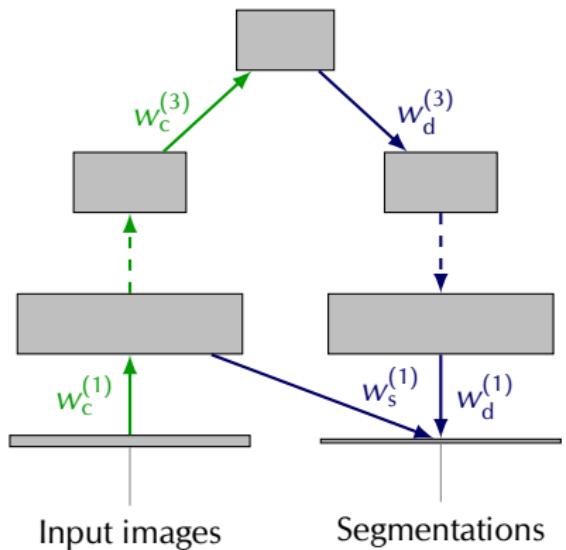
Network Architecture: 7-layer CEN



Detected lesions
Missed lesions
False positives

[Brosch et al., TMI 2016]

Network Architecture: 7-layer CEN with shortcut



Detected lesions
Missed lesions
False positives

[Brosch et al., TMI 2016]

Evaluation

Evaluated on two data sets:

- MICCAI 2008 MS Lesion Segmentation Challenge
 - 20 training, 23 testing images
 - Modalities: T1w, T2w, FLAIR
- Clinical trial data set
 - 250 training, 77 testing images
 - Modalities: T1w, T2w, PDw, FLAIR
 - Compared to 5 freely available methods
 - Stratified by average lesion size

Evaluation measures:

- Dice similarity coefficient (DSC)
- Lesion true positive rate (LTPR)
- Lesion false positive rate (LFPR)
- Volume difference (VD)

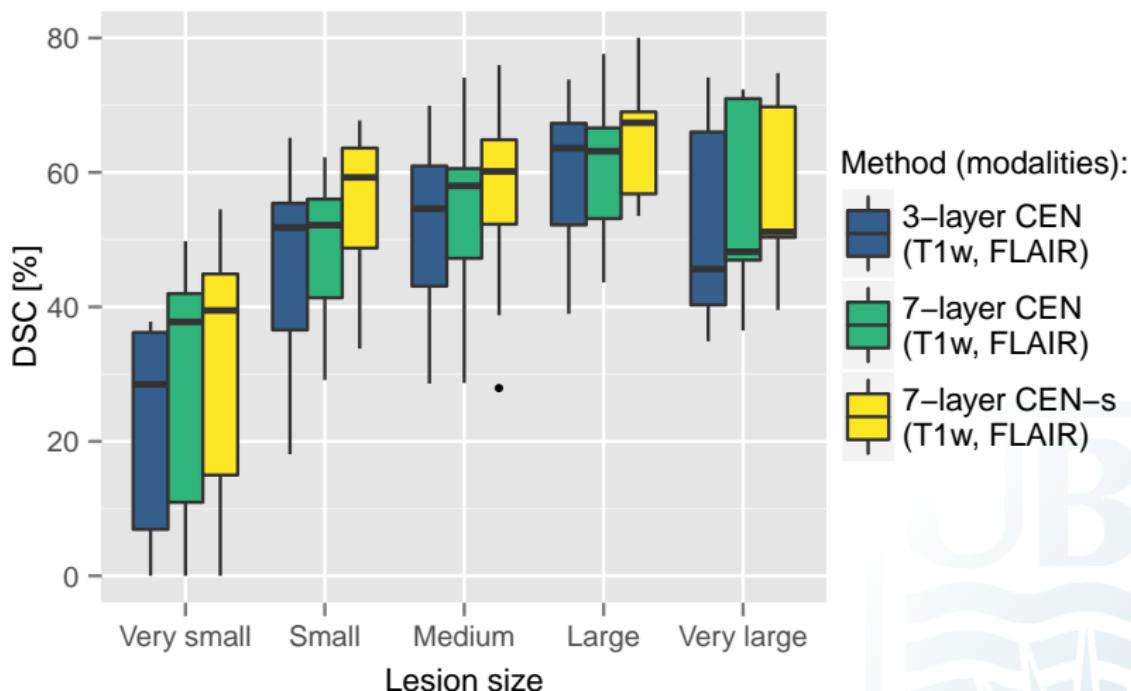
MICCAI 2008 Lesion Segmentation Challenge

Selected methods out of the 52 entries submitted for evaluation to the MICCAI 2008 MS lesion segmentation challenge.

Rank	Method	Score	LTPR	LFPR	VD
1, 3, 9	Jesson et al. (2015)	86.94	48.7	28.3	80.2
2	Guizard et al. (2015)	86.11	49.9	42.8	48.8
4, 20, 26	Tomas-Fernandez et al. (2012)	84.46	46.9	44.6	45.6
5, 7	Jerman et al. (2015)	84.16	65.2	63.8	77.5
6	Our method	84.07	51.6	51.3	57.8
11	Roura et al. (2015)	82.34	50.2	41.9	111.6
13	Geremia et al. (2010)	82.07	55.1	74.1	48.9
24	Shiee et al. (2010)	79.90	52.4	72.7	74.5
33	Sudre et al. (2014)	77.96	22.3	18.1	285.6

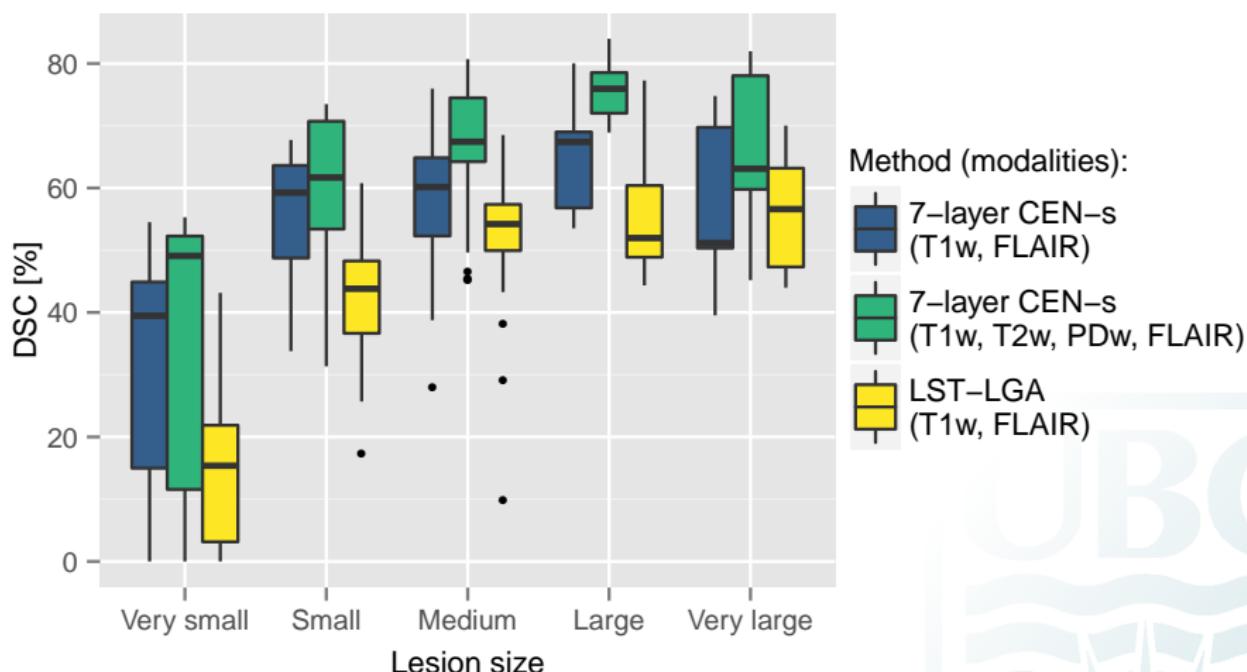
[Brosch et al., TMI 2016]

Clinical Trial Data Set



[Brosch et al., TMI 2016]

Clinical Trial Data Set



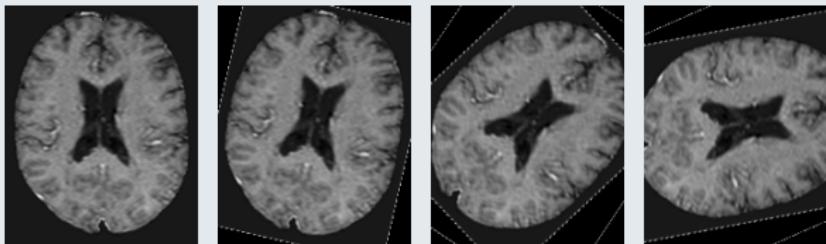
[Brosch et al., TMI 2016]

Manifold Learning

- Developed a computation and memory efficient training algorithm for convolutional deep learning models
- Developed a novel segmentation method that learns features at different scales that are tuned to a given combination of image types and segmentation task
- Proposed a novel objective function that facilitates the training using vastly unbalanced training sets, such as is the case for segmenting MS lesions
- First work to demonstrate the use of deep learning for manifold learning of 3D medical images
- Developed a framework for modelling changes in brain morphology and lesion distribution

Modelling Variability by Manifold Learning

Toy Example: Rotated Brain MRIs

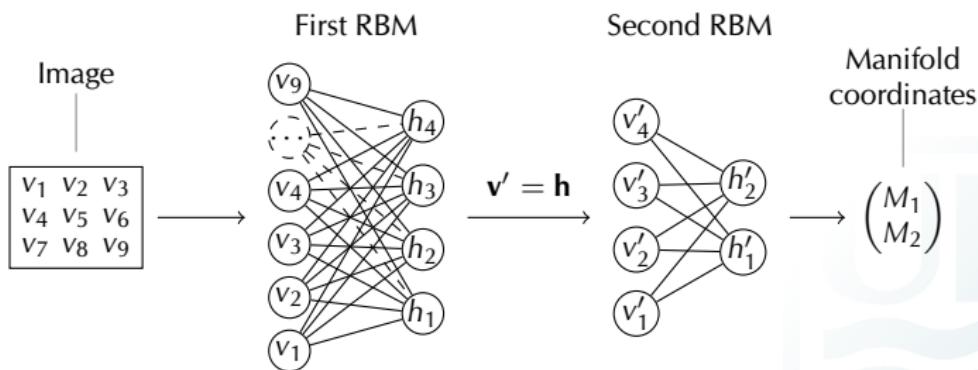


- **Manifold assumption:** Brain images lie on a low-dimensional manifold, where the manifold coordinates represent the modes of variation.
- ? What are the modes of variations of brain images affected by Alzheimer's disease and multiple sclerosis?
- ? Can the manifold coordinates be used to measure disease state and progression?

Manifold Learning of Brain MRIs by Deep Learning

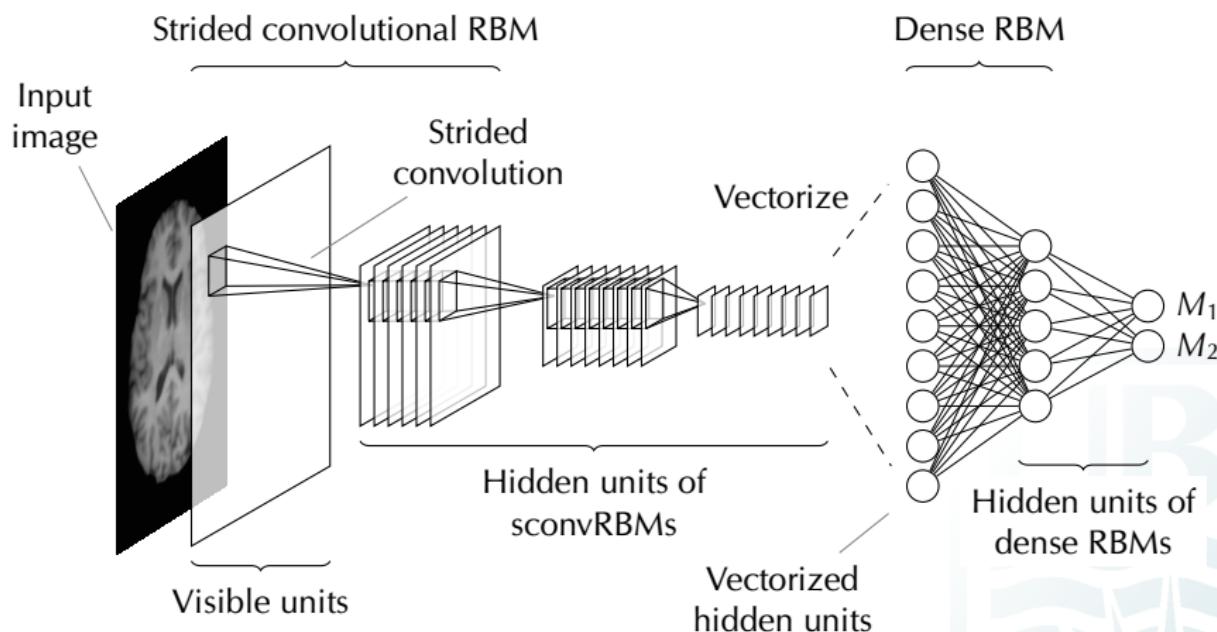
Proposed Framework:

- Based on deep belief networks
- Fewer assumptions than alternative manifold learning methods



[Brosch and Tam, MICCAI 2013]

Manifold Learning of Brain MRIs by Deep Learning



[Brosch and Tam, MICCAI 2013]

Evaluation

Data set:

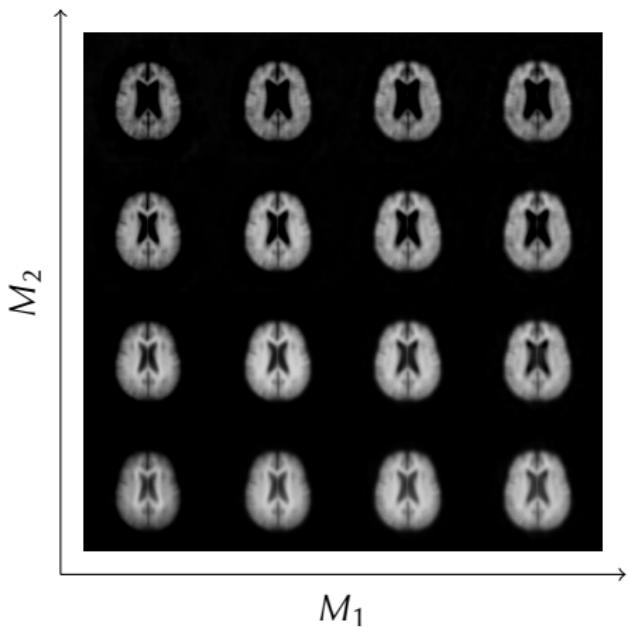
- 300 T1-weighted MRIs of AD (150) and normal (150) subjects from the ADNI data set
- Number of voxels: $128 \times 128 \times 128$
- Voxel size: $2 \times 2 \times 2$ mm

Experiments:

- Visualization of patterns of variability
- Visualization of population differences in manifold space

[Brosch and Tam, MICCAI 2013]

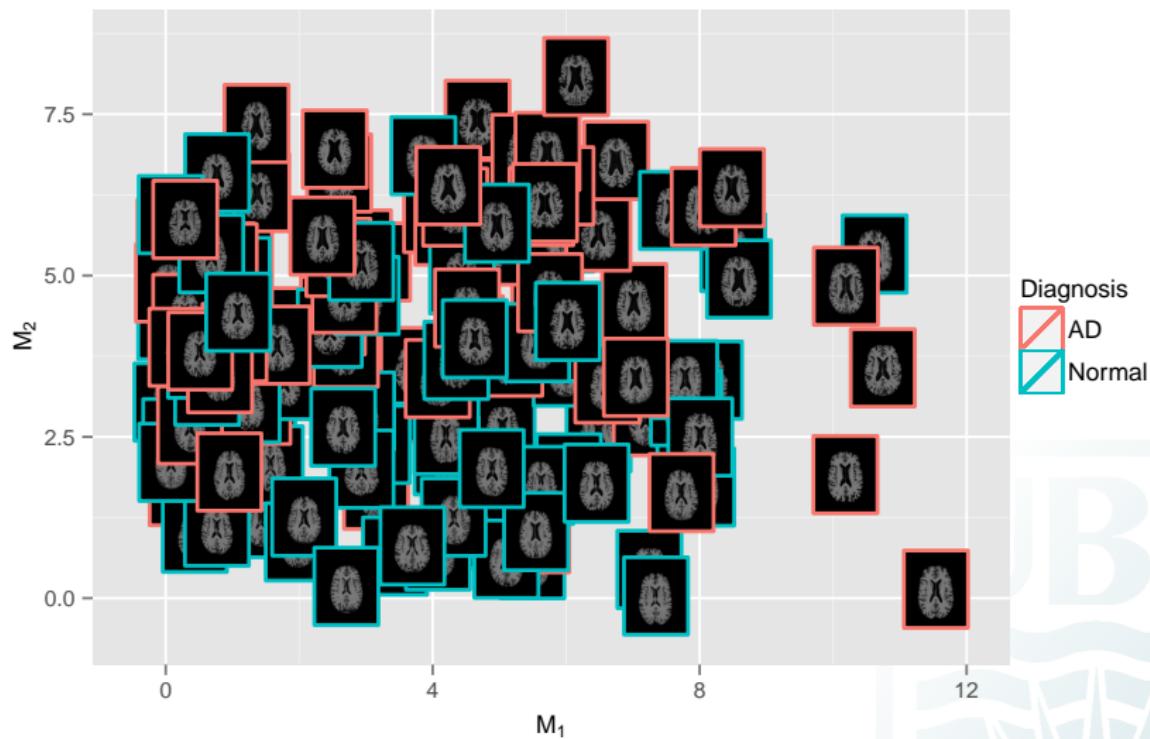
Visualizing the Modes of Variation



- M_1 visually correlates with an increase in brain size
- M_2 visually correlates with an increase in ventricle size

[Brosch and Tam, MICCAI 2013]

Visualization of the Data Set in Manifold Space



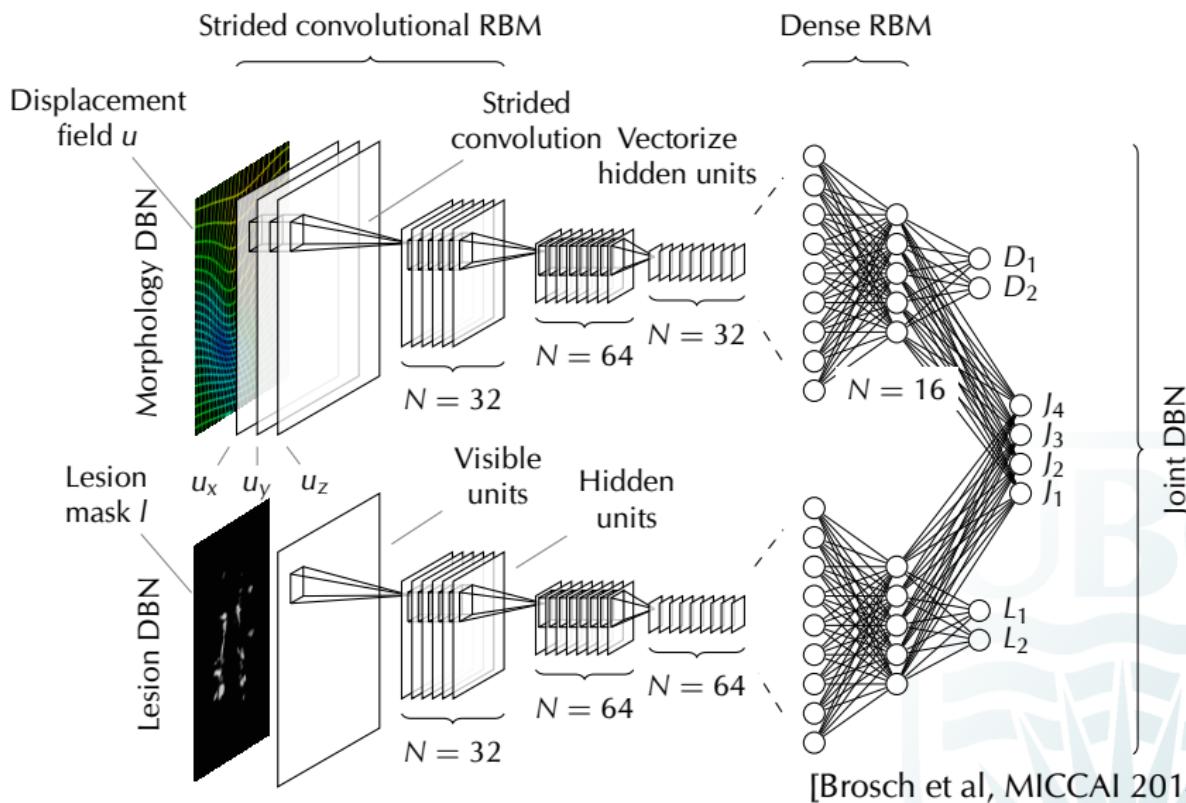
[Brosch and Tam, MICCAI 2013]

Modelling the Variability of MS Brains

- ➊ What are the predominant morphological variations in MS?
- ➋ How are MS lesions distributed?
- ➌ How do these key pathological features interact?
- ➍ Enforce the learning of morphological and lesion distribution changes
- ➎ DBN trained on **deformation fields** and binary **lesion masks**

[Brosch et al, MICCAI 2014]

Model



Evaluation

Data set:

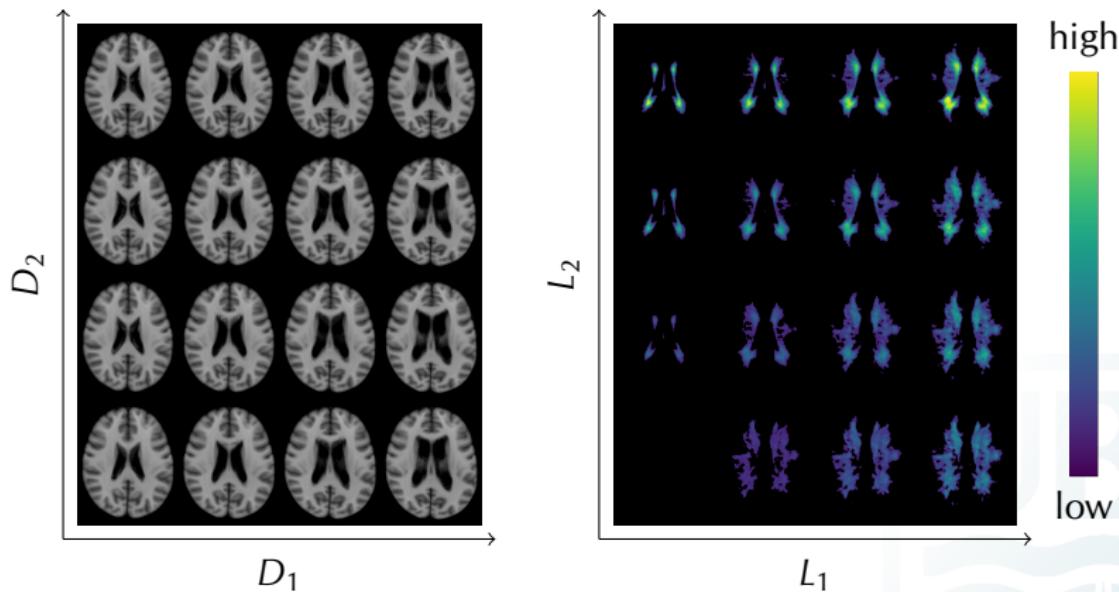
- 474 T1-weighted, T2-weighted, and PD-weighted MRIs of secondary progressive MS patients
- Number of voxels: $256 \times 256 \times 50$
- Voxel size: $0.937 \times 0.937 \times 3.00 \text{ mm}$

Experiments:

- Visualization of patterns of morphological and lesion distribution changes
- Correlation of manifold coordinates with MS clinical scores

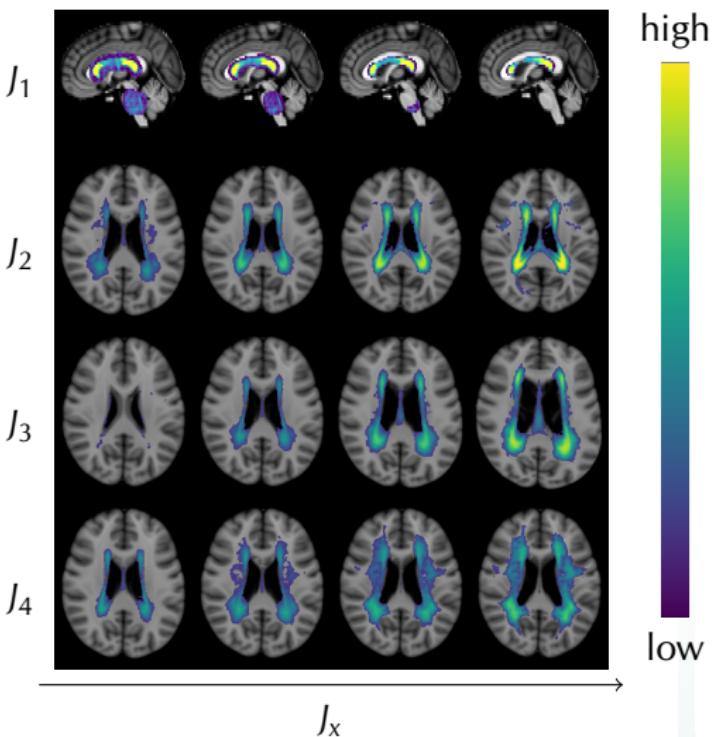
[Brosch et al, MICCAI 2014]

Morphology and Lesion Distribution Patterns



[Brosch et al, MICCAI 2014]

Joint Patterns of Variability



[Brosch et al, MICCAI 2014]

Pearson Correlations with Clinical Scores

		T25W	9-HPT	PASAT	MSFC
Individual models	D_1	-0.129**	-0.215***	-0.282***	-0.315***
	D_2	0.087	0.116*	0.089	0.139**
	L_1	-0.058	-0.231***	-0.392***	-0.367***
	L_2	-0.091	-0.354***	-0.427***	-0.464***
Joint model	J_1	0.107*	0.286***	0.336***	0.379***
	J_2	-0.038	-0.210***	-0.227***	-0.256***
	J_3	-0.118*	-0.369***	-0.453***	-0.494***
	J_4	-0.049	-0.206***	-0.383***	-0.346***
Imaging biomarkers	nBV	0.053	0.144**	0.247***	0.235***
	LL	-0.074	-0.286***	-0.400***	-0.406***

Abbreviations: Timed 25-Foot Walk (T25W), 9-Hole Peg Test (9-HPT), Paced Auditory Serial Addition Test (PASAT), MS Functional Composite (MSFC), normalized brain volume (nBV), lesion load (LL)

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Summary and Conclusions

- Presented a novel training algorithm for convolutional models that performs training in the frequency domain
- Minimizes the number of FFTs through the mapping of operations to the frequency domain
- Significantly faster than the state-of-the-art spatial domain implementation
- Facilitates the application of convolutional deep learning models for 3D neurological images analysis

Summary and Conclusions

- Presented a fully automatic MS lesion segmentation method based on convolutional encoder networks
- Joint training of feature extraction and classification layers allows for the learning of features that are tuned to any given combination of image types and segmentation task
- Our framework is very flexible and can be readily applied to various segmentation problems
- Presented a novel method for the automatic discovery of patterns of variability in multiple sclerosis
- Framework allows for the visualization and quantification of morphology and lesion distribution patterns
- Parameters of our model correlate stronger with MS clinical scores than volumetric imaging biomarkers

Publications Arising from this Thesis

- Brosch, Tom and R. Tam, for the ADNI. Manifold learning of brain MRIs by deep learning. In: K. Mori et als. (Eds.): MICCAI 2013, Part II, LNCS 8150, pages 633–640, 2013.
- Brosch, Tom, Y. Yoo, A. Traboulsee, D.K.B. Li, and R. Tam. Modeling the variability in brain morphology and lesion distribution in multiple sclerosis by deep learning. In: P. Golland et al. (Eds.): MICCAI 2014, Part II, LNCS 8674, pages 462–469, 2014.
- Brosch, Tom and R. Tam. Efficient training of convolutional deep belief networks in the frequency domain for application to high-resolution 2D and 3D images. *Neural Computation*, 27(1), pages 211–227, 2015.

Publications Arising from this Thesis

- Brosch, Tom, Y. Yoo, L.Y.W. Tang, D.K.B. Li, A. Traboulsee, and R. Tam. Deep convolutional encoder networks for multiple sclerosis lesion segmentation. In: N. Navab et al. (Eds.): MICCAI 2015, Part III, LNCS 9351, pages 3–11, 2015.
- Brosch, Tom, L.Y.W. Tang, Y. Yoo, D.K.B. Li, A. Traboulsee, and R. Tam. Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation. IEEE Transactions on Medical Imaging, 2016. (accepted)

Thank you for your attention.

