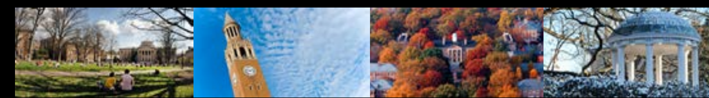
UNC SCHOOL OF MEDICINE



Deep Learning in Medical Imaging Analysis

Dinggang Shen, Ph.D.

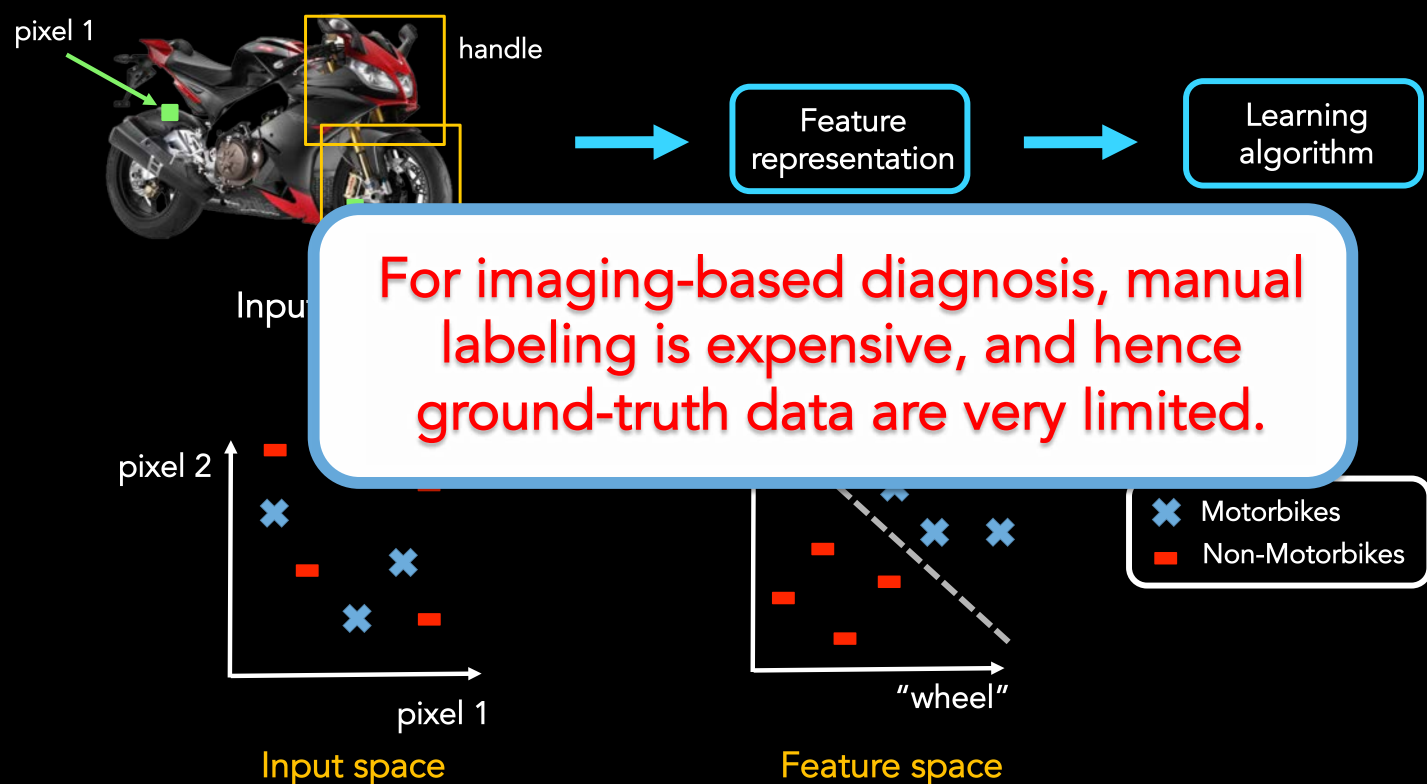
IDEA Lab, Department of Radiology, and Biomedical Research Imaging Center (BRIC)
University of North Carolina at Chapel Hill

BRIC BIOMEDICAL RESEARCH IMAGING CENTER

Research Lab IDEA

Deep Learning

Supervised Learning



For imaging-based diagnosis, manual labeling is expensive, and hence ground-truth data are very limited.

Unsupervised Learning

Method	Limitations
• PCA	✓ Linear ✓ Not optimal for non-Gaussian data
• Gaussian Mixture Models • K-Means	✓ Require knowledge for the number of clusters ✓ Challenging when applied to high-dimensional data
• ICA	✓ Linear model
• Sparse Coding • Non-Linear Embedding	✓ Shallow model (e.g., single-layer representation)

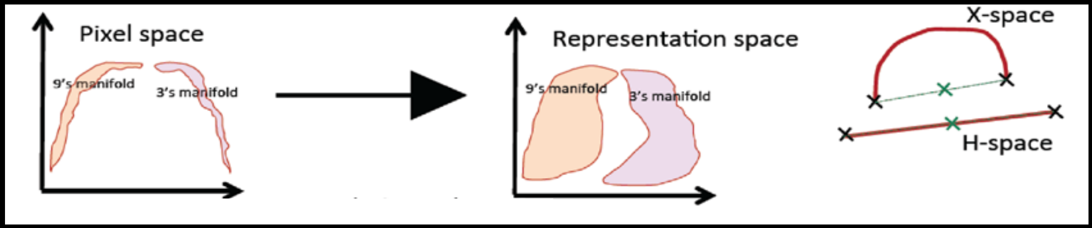
- All these methods involve just one step of mapping!

  → Mapping is shallow, not deep!

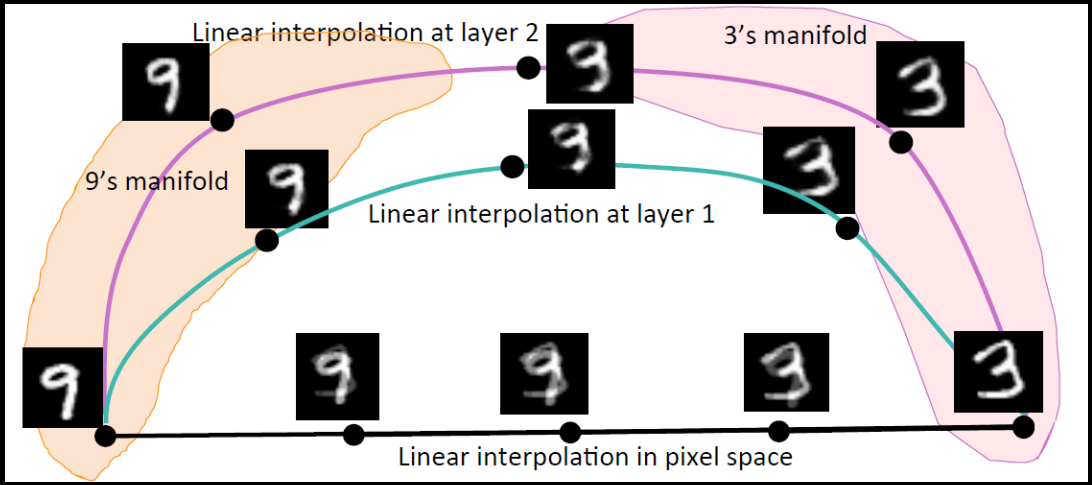
  → Thus, not able to represent the complex mapping!

Deep Learning – Why hot?

- Deep mapping and representation



Deeper representations
→ abstractions
→ disentangling

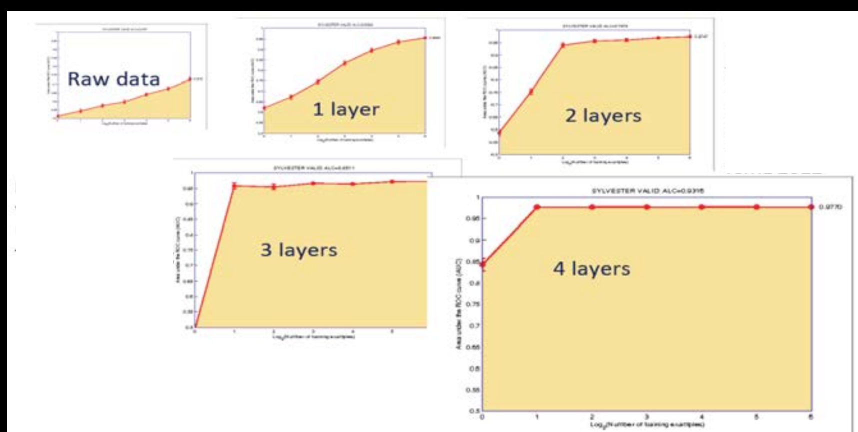


Manifolds are expanded and flattened

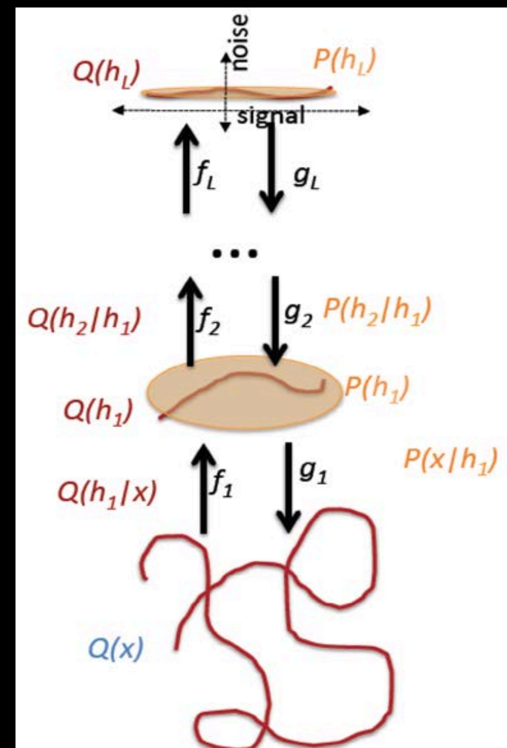
The following 5 slides edited from Dr. Yoshua Bengio's tutorial

Deep Learning – Why hot?

- Deep mapping and representation
  - Each level transforms the data into a representation, which can be easily modeled
    - → Unfolding it more will map the original data to a factorized (uniform-like) distribution

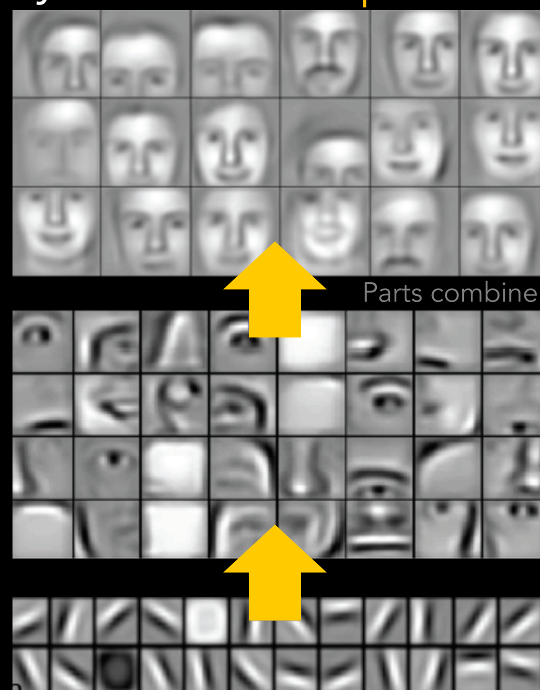


Performance increase with layers



Deep Learning – Why hot?

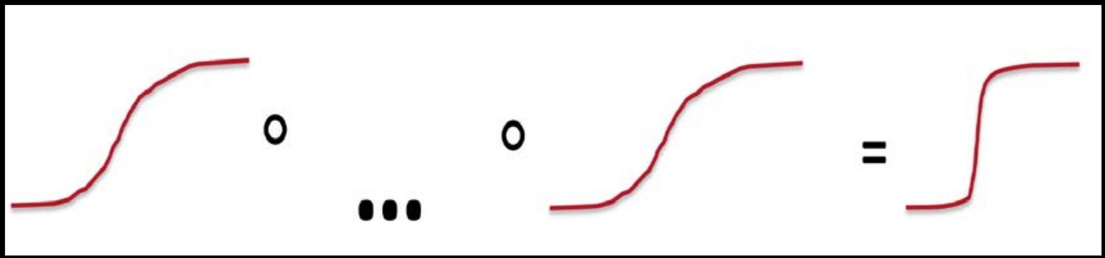
- Successive model layers learn deeper intermediate representations

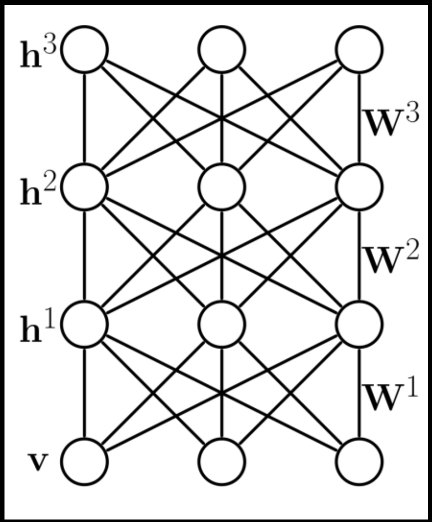


- Prior: Underlying factors and concepts compactly expressed without multiple levels of abstraction

Neural Network – Why not working

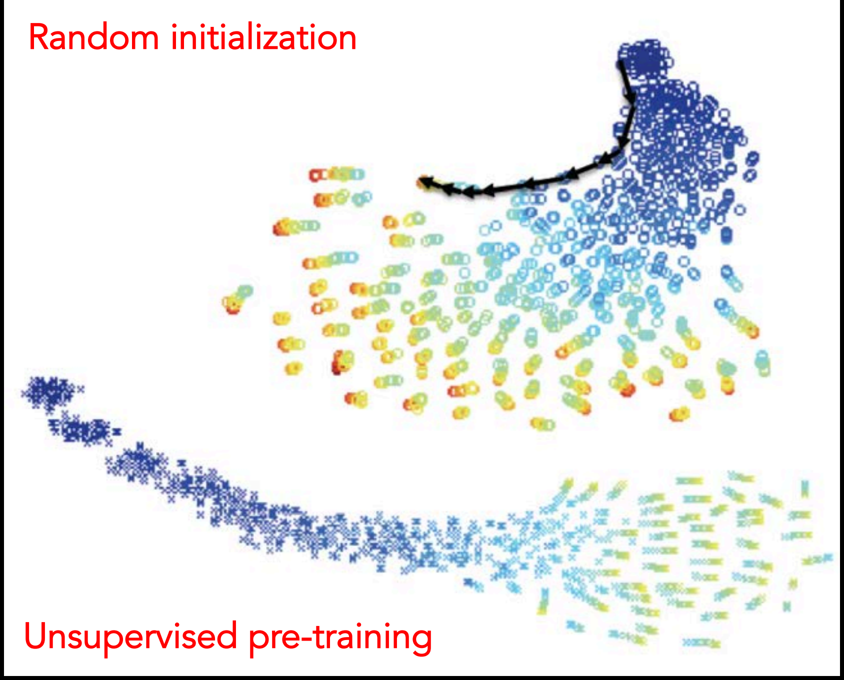
- Issues with previous neural network (NN)
  - Gradient-based method → propagate errors from the last layer to the previous layers
  - Last layer represents high nonlinear function (i.e., a jump function in binary classification) → unstable and large gradient in small range, but zero in most places





Neural Network – Why not working

- Effect of initial conditions in Deep Nets



No two training trajectories end up in the same place → huge number of effective local minima

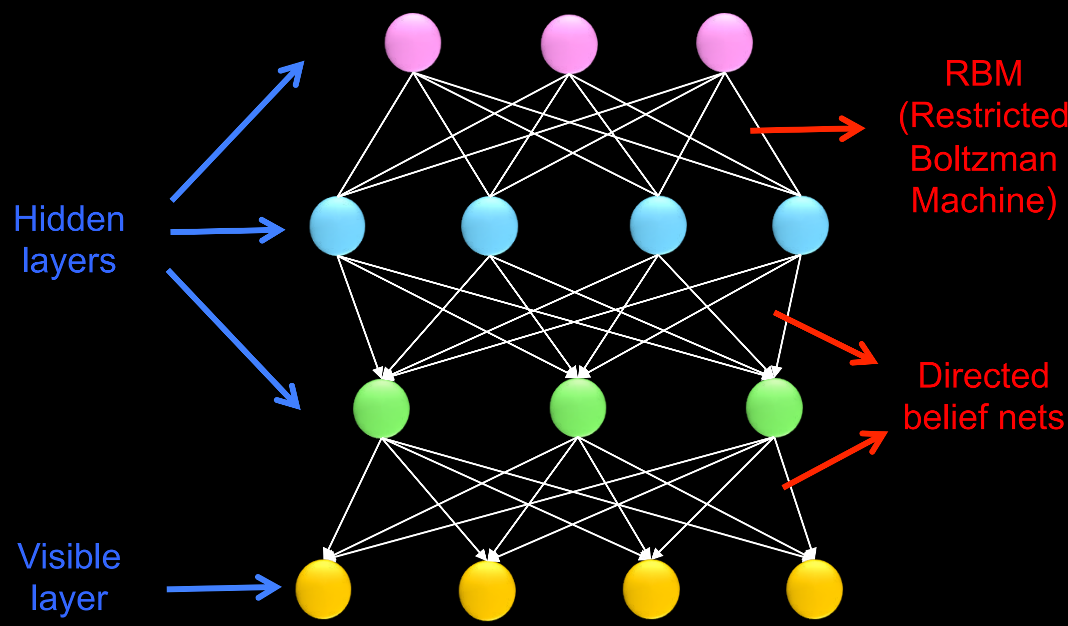
Pre-training: Transfer knowledge from previous learning (representation and explanatory factors) → cases with few examples → shared underlying explanatory factors, between P(X) and P(Y|X)

Deep Learning – Why working now

- Three main reasons
  - New layer-wise training algorithm [Science 2006]
    - Each time, train on simple task
  - Big data, compared to 20 years ago
  - Powerful computers
    - Previous algorithms may be theoretically working, but practically not converged to good local minima with the previous less-powerful computers

BRIC BIOMEDICAL RESEARCH IMAGING CENTER
Research Lab IDEA

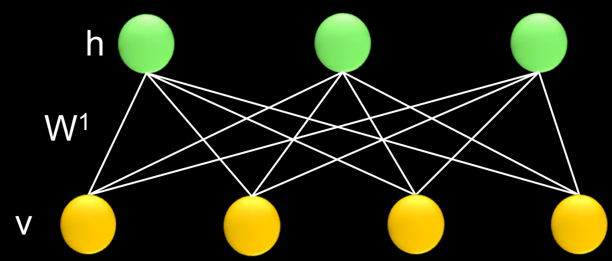
Deep Learning



$$P(v, h^1, h^2, \ldots, h^l) = P(v|h^1)\, P(h^1|h^2) \ldots P(h^{l-2}|h^{l-1})\, P(h^{l-1}, h^l)$$

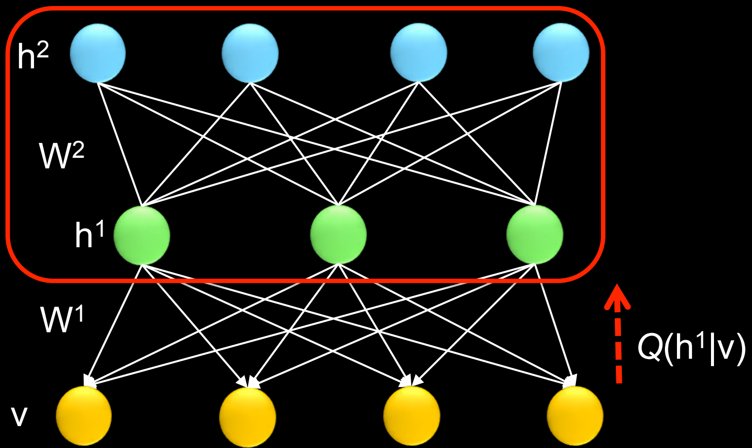
Deep Learning – Greedy Training

- First step
  - Construct an RBM with an input layer v and a hidden layer h
  - Train the RBM



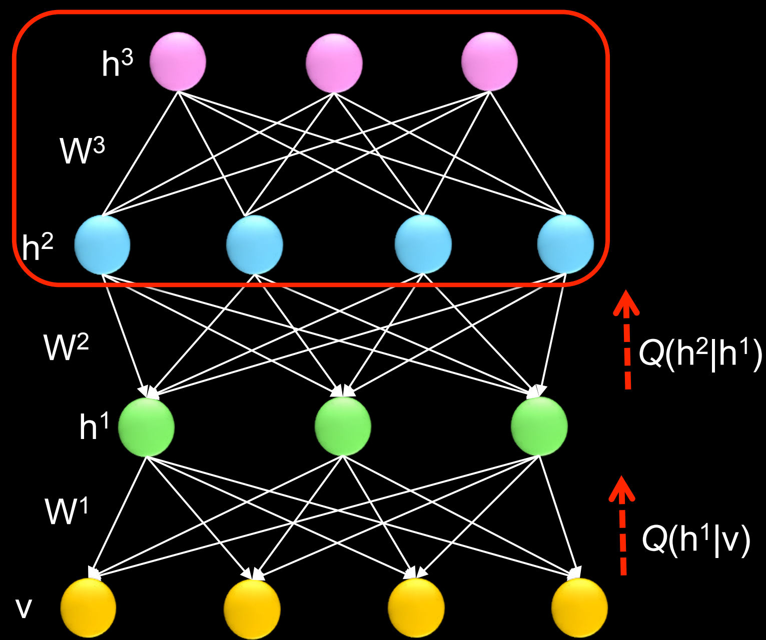
Deep Learning – Greedy Training

- Second step
  - Stack another hidden layer on top of the RBM to form a new RBM
  - Fix W^1, sample h^1 from $Q(h^1|v)$ as input
  - Train W^2 as RBM

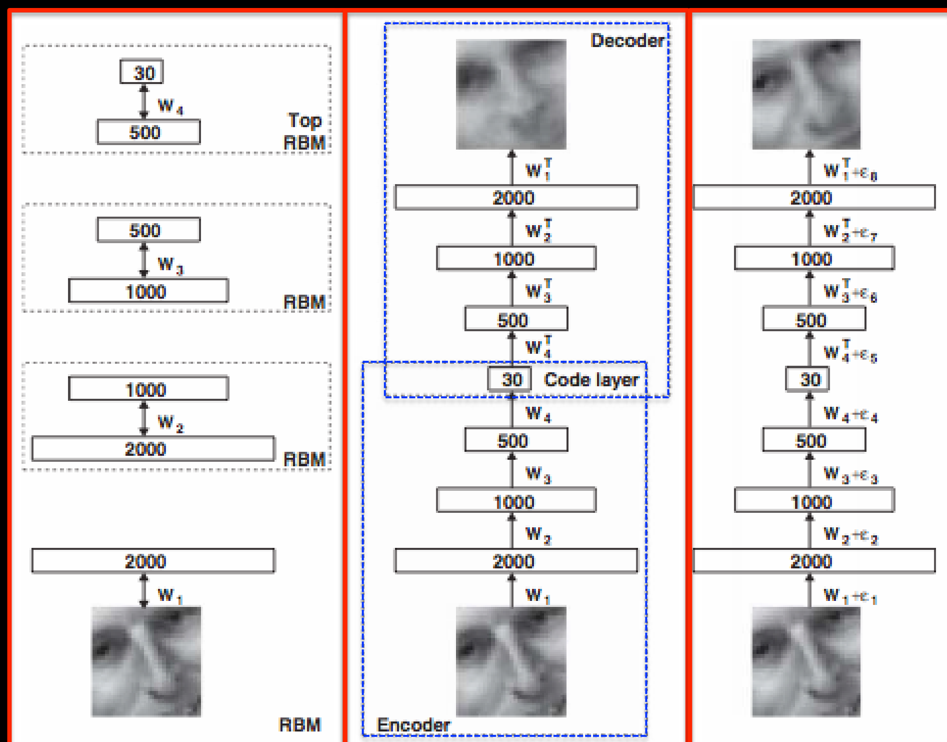


Deep Learning – Greedy Training

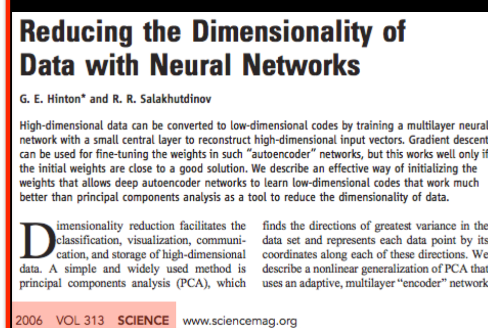
- Third step
  - Continue to stack layers on top of the network, and train it as previous step, with samples sampled from from $Q(h^2|h^1)$
- And so on…



Deep Learning – Stacked Auto-Encoder



Pretraining **Unrolling** **Fine-tuning**



Reducing the Dimensionality of Data with Neural Networks

G. E. Hinton* and R. R. Salakhutdinov

High-dimensional data can be converted to low-dimensional codes by training a multilayer neural network with a small central layer to reconstruct high-dimensional input vectors. Gradient descent can be used for fine-tuning the weights in such "autoencoder" networks, but this works well only if the initial weights are close to a good solution. We describe an effective way of initializing the weights that allows deep autoencoder networks to learn low-dimensional codes that work much better than principal components analysis as a tool to reduce the dimensionality of data.

Dimensionality reduction facilitates the classification, visualization, communication, and storage of high-dimensional data. A simple and widely used method is principal components analysis (PCA), which finds the directions of greatest variance in the data set and represents each data point by its coordinates along each of these directions. We describe a nonlinear generalization of PCA that uses an adaptive, multilayer "encoder" network

2006 VOL 313 SCIENCE www.sciencemag.org

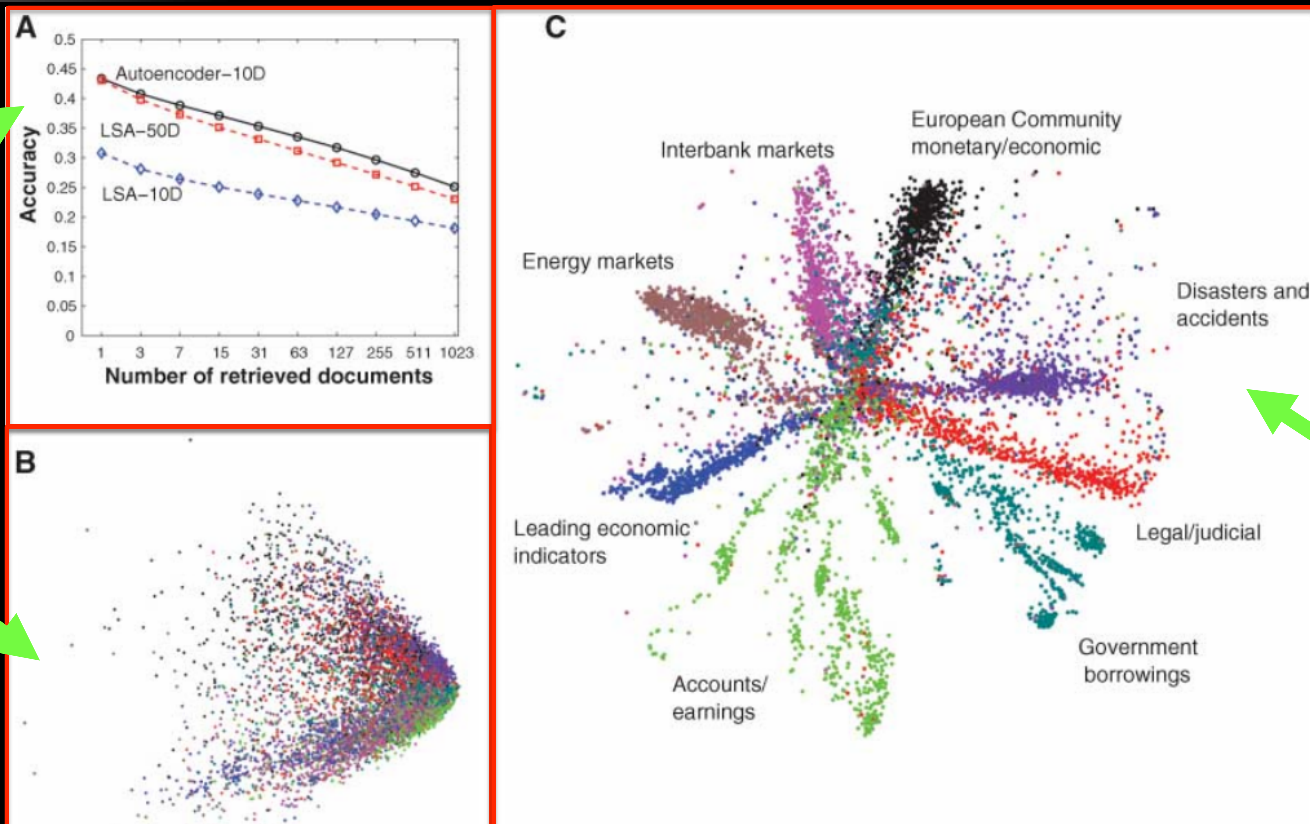
BRIC BIOMEDICAL RESEARCH IMAGING CENTER

Research Lab IDEA

Deep Learning – Stacked Auto-Encoder

The fraction of retrieved documents

The codes produced by 2D LSA



The codes produced by a 500-250-125-2 Auto-Encoder

Application 1

Segmentation

- Hippocampus Segmentation using 7T MRIs
- Infant Brain Segmentation

UNC SCHOOL OF MEDICINE

Hippocampus Segmentation

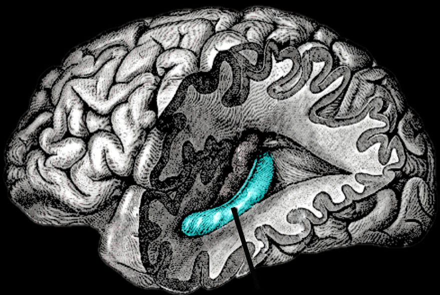
Hippocampus Segmentation Using 7T MR Images

BRIC BIOMEDICAL RESEARCH IMAGING CENTER

Research Lab IDEA

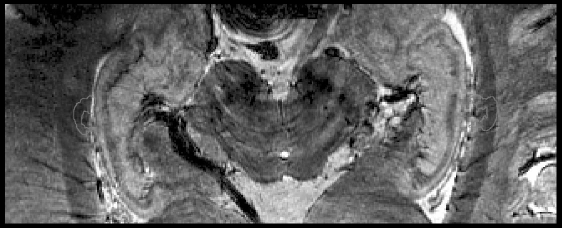
Hippocampus Segmentation

- Challenges in hippocampus segmentation using 1.5T/3T and 7T



Hippocampus
($\approx 35 \times 15 \times 7 mm^3$)

1.5T/3T ($1 \times 1 \times 1\ mm^3$)



7.0T ($0.35 \times 0.35 \times 0.35\ mm^3$)

- Low imaging resolution
- Low contrast

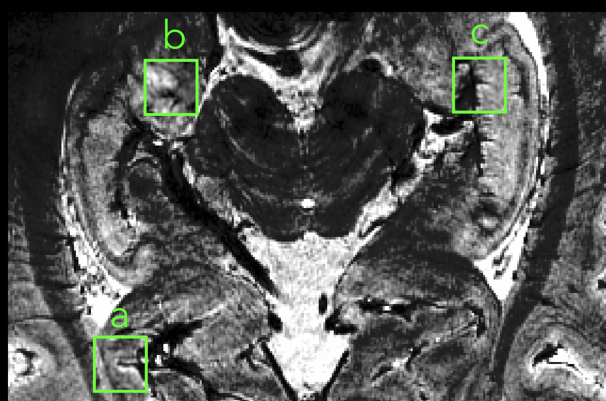
- Much richer structural information
- Less partial volume effect
- But, severe intensity inhomogeneity problem

M. Kim, G. Wu, D. Shen, "Unsupervised Deep Learning for Hippocampus Segmentation in 7.0 Tesla MR Images," *MLMI*, 2013.

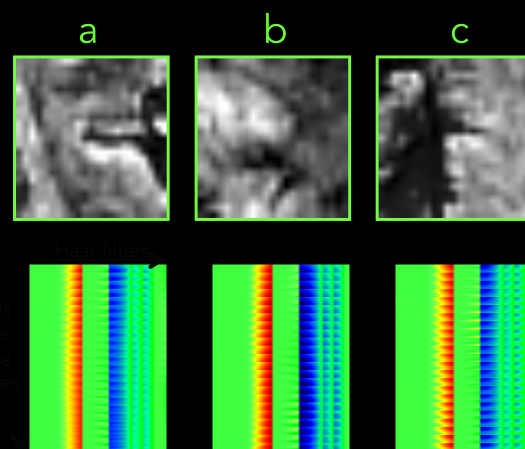
BRIC BIOMEDICAL RESEARCH IMAGING CENTER
Research Lab IDEA

Hand-Crafted Features

- Limited discriminative power of hand-crafted features



Extracting patches from a 7T MR image



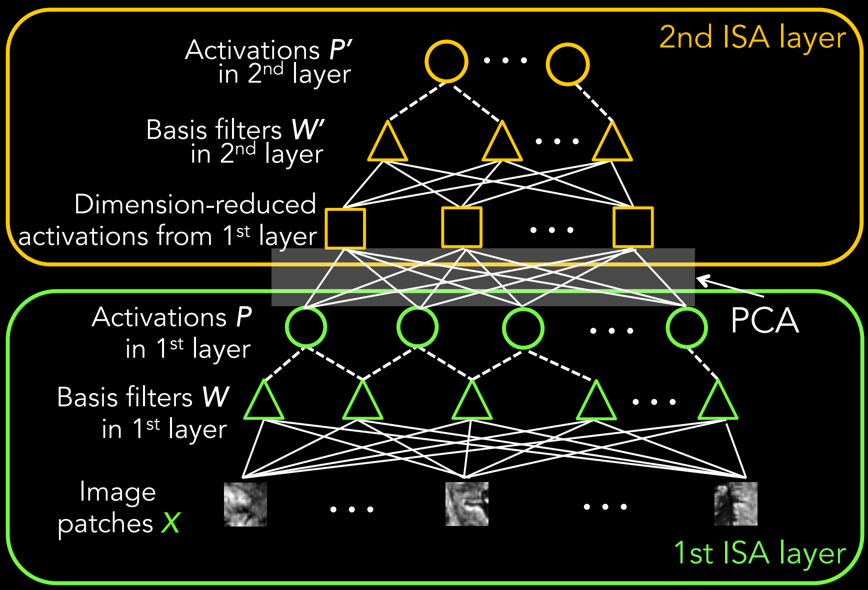
Responses of Haar filters for the image patches

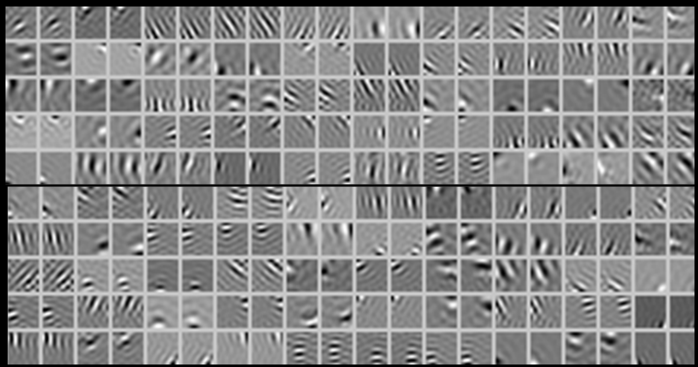
BRIC BIOMEDICAL RESEARCH IMAGING CENTER

Research Lab IDEA

Hierarchical Feature Extraction via Unsupervised Deep Learning

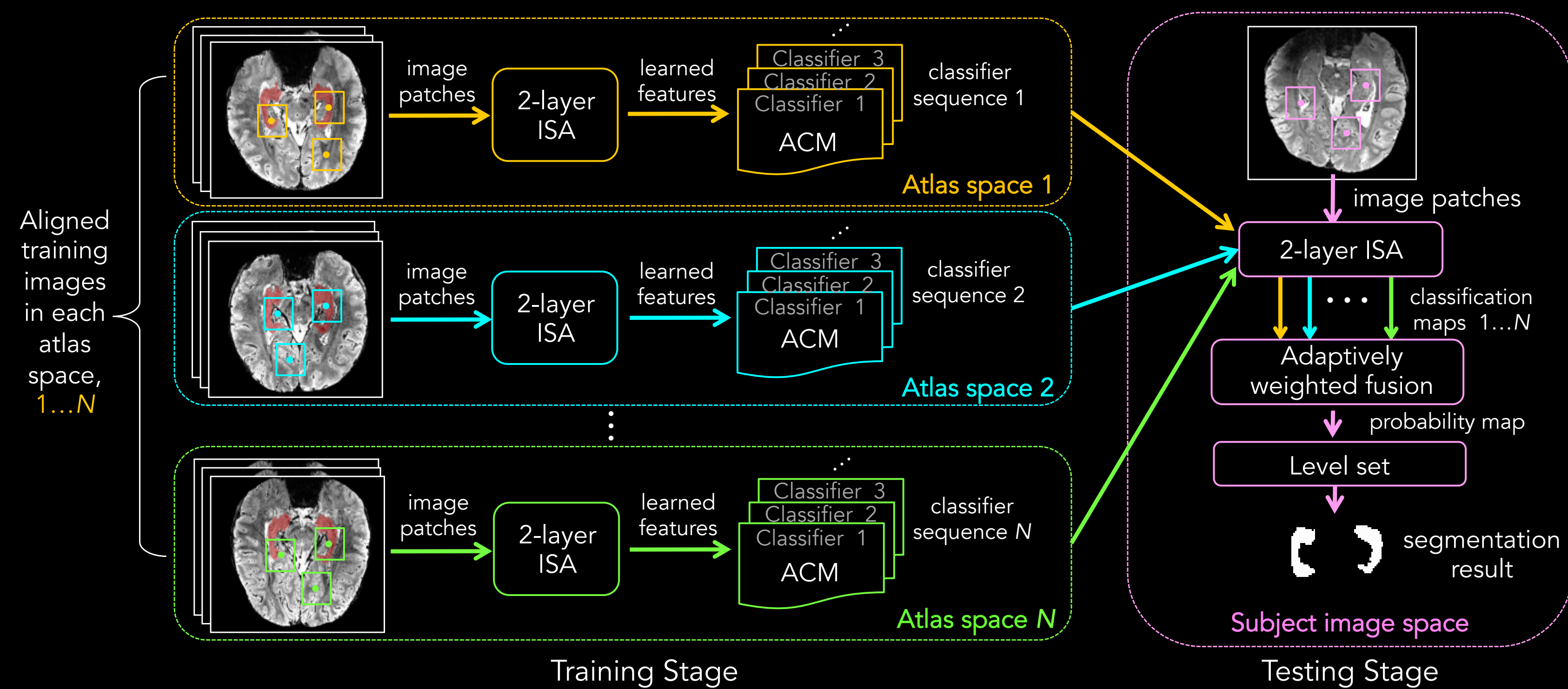
- Stacked two-layer convolutional ISA (Independent Subspace Analysis)





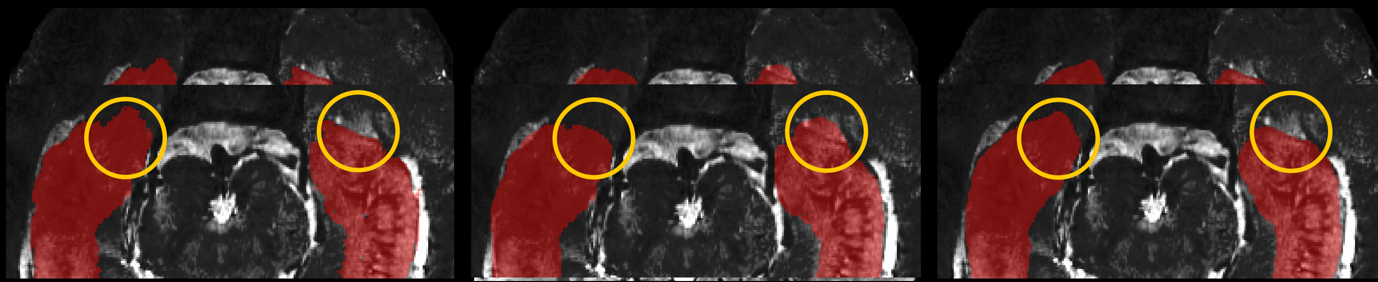
Learned basis filters by the 1st ISA

Multi-Atlas-based Segmentation using Deep Learning Features



BRIC
BIOMEDICAL RESEARCH IMAGING CENTER
Research Lab IDEA

Results



Ground Truth | Haar + Texture Features | Hierarchical Features

Comparison Results Using 20 Leave-One-Out Cases

	P	R	RO	SI
Hand-Crafted Haar + Texture Features	0.843	0.847	0.772	0.865
Hierarchical Patch Representations	0.883	0.881	0.819	0.894

P = Precision; R = Recall; RO = Relative overlap; SI = Similarity index

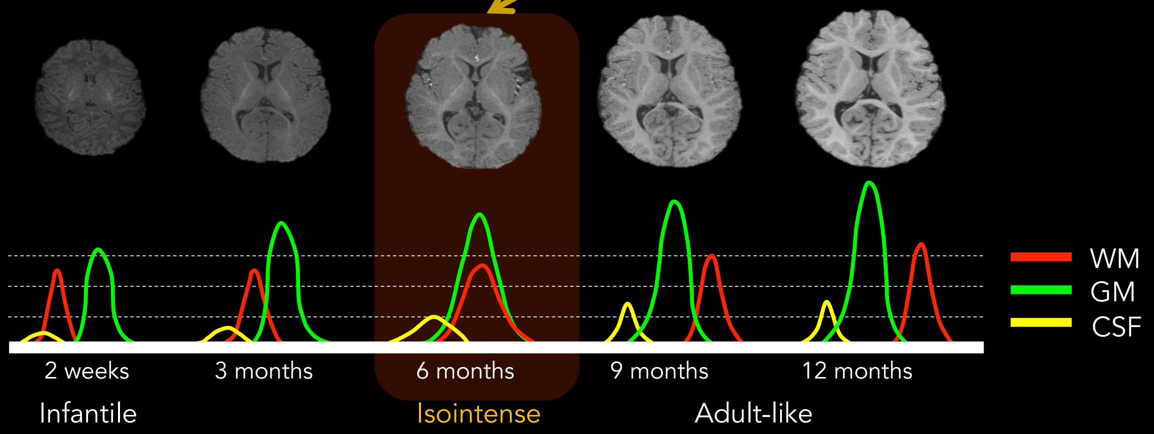
Infant Brain Segmentation

Multi-modality Isointense Infant Brain Image Segmentation

Infant Brain Segmentation

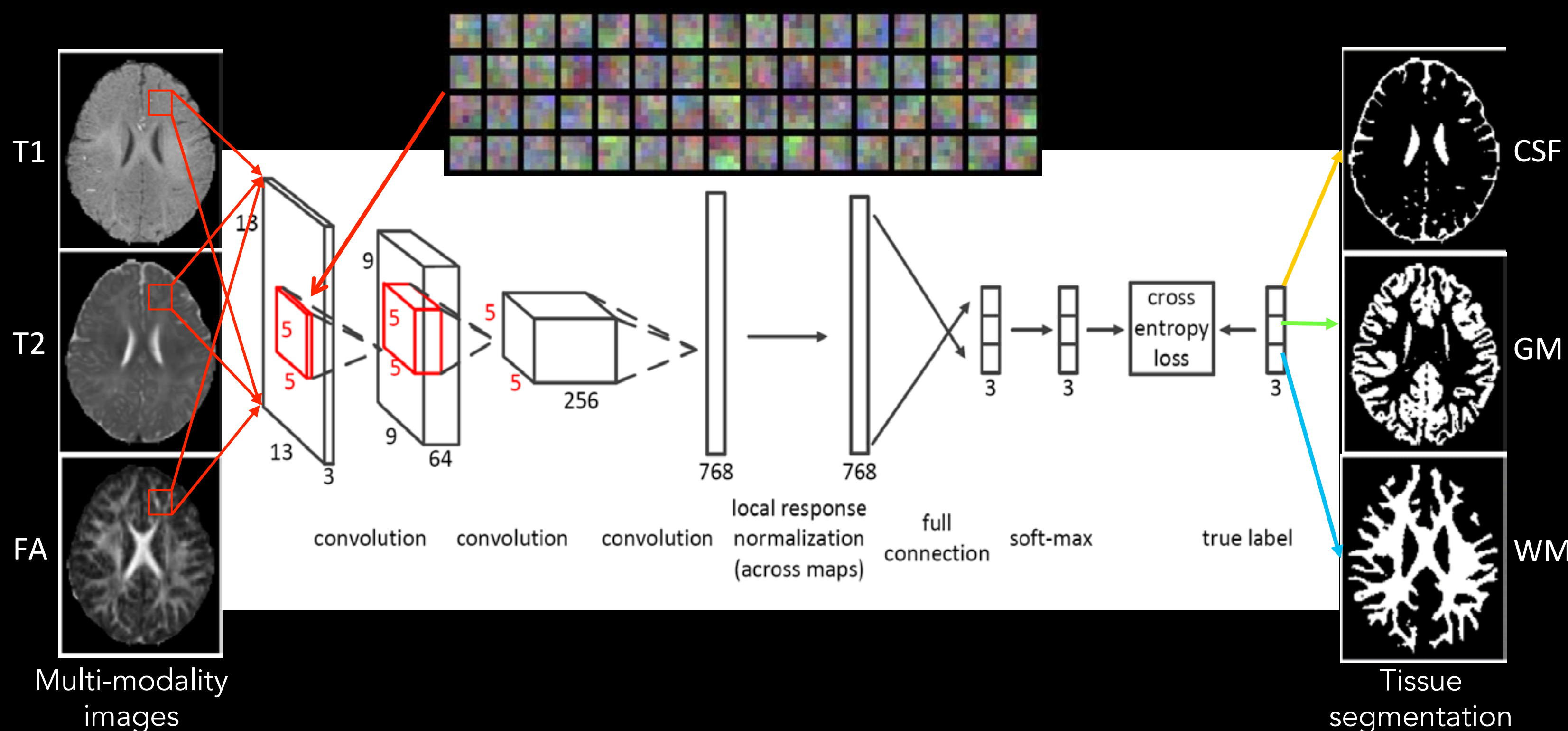
- Challenges in infant brain segmentation
  - Low tissue contrast
  - Low spatial resolution

WM and GM exhibit almost the same level of intensity in both T1 and T2 MR images



W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, D. Shen, "Deep Convolutional neural networks for multi-modality isointense infant brain image segmentation," *Neuroimage*, 2015.

Deep Convolutional Neural Network (CNN)



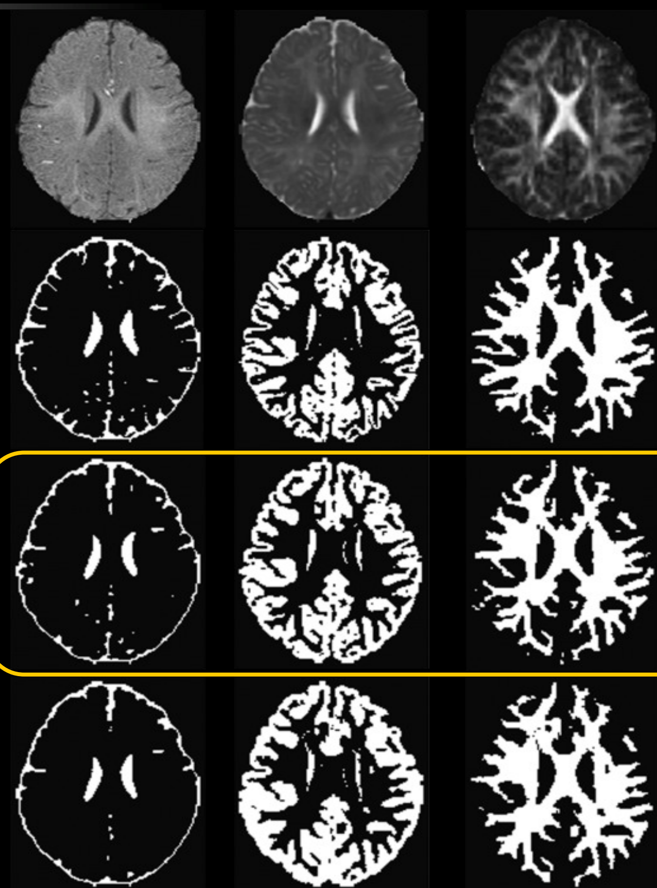
Results

Segmentation performance in terms of Dice ratio achieved by the CNN, RF, SVM, CLS, MV

		Subj. 1	Subj. 2	Subj. 3	Subj. 4	Subj. 5	Subj. 6	Subj. 7	Subj. 8
CSF	CNN	0.83	0.83	0.83	0.84	0.85	0.85	0.82	0.83
	RF	0.82	0.81	0.83	0.81	0.83	0.85	0.79	0.80
	SVM	0.74	0.77	0.77	0.74	0.70	0.78	0.72	0.73
	CLS	0.81	0.82	0.73	0.86	0.84	0.82	0.81	0.83
	MV	0.71	0.69	0.68	0.63	0.63	0.61	0.69	0.69
GM	CNN	0.85	0.86	0.88	0.82	0.81	0.87	0.86	0.86
	RF	0.83	0.85	0.88	0.81	0.80	0.85	0.85	0.84
	SVM	0.79	0.80	0.83	0.75	0.74	0.80	0.80	0.80
	CLS	0.83	0.84	0.85	0.83	0.81	0.87	0.86	0.84
	MV	0.85	0.84	0.85	0.80	0.78	0.80	0.84	0.83
WM	CNN	0.88	0.81	0.88	0.85	0.87	0.87	0.87	0.88
	RF	0.86	0.78	0.87	0.84	0.85	0.86	0.84	0.84
	SVM	0.82	0.74	0.76	0.80	0.80	0.79	0.71	0.76
	CLS	0.84	0.81	0.80	0.82	0.84	0.82	0.83	0.81
	MV	0.86	0.80	0.85	0.82	0.84	0.84	0.84	0.84

CNN: Convolutional Neural Network
RF: Random Forest
SVM: Support Vector Machine
CLS: Coupled Level Sets
MV: Majority Voting

Results



Original multi-modality data (T1, T2 and FA)

Manual segmentations (CSF, GM, and WM)

Segmentation results by CNN

Segmentation results by RF

BRIC BIOMEDICAL RESEARCH IMAGING CENTER

Research Lab IDEA