Graphical Models For Complex Health Data (P8124)

Daniel Malinsky

Columbia University
dsm2128@cumc.columbia.edu

Image analysis with VAEs

Unsupervised image segmentation

Goal: assign each voxel its most probable category label (e.g., one of L tissue types or brain ROIs)

Innovation: use VAEs to accomplish this with minimal supervision (most data has no labels)

Strategy: impose a MRF-structured prior, train VAE by maximizing ELBO, for new image use this VAE to determine most probable label

ELBO

Last week we saw:

$$\mathcal{L}_{\theta,\phi}(x) = -D_{KL}(q(s|x;\phi)||p(s;\theta)) + \mathbb{E}_q[\log p(x|s;\theta)]$$

Second term is evaluated by drawing samples $s \sim q(s|x)$

Choosing:

$$q(s|x) = \prod_{j=1}^{V} \text{Cat}(s_j|x;\phi)$$

and

$$p(x|s) = \prod_{j=1}^{V} N(x;\hat{x}_j(s;\theta),\sigma^2)$$

where $\hat{x}_j(s;\theta)$ is a "reconstruction" image computed by decoder NN

(using the notation from the paper)

ELBO

$$-\mathcal{L} = \sum_{i=1}^{N} D_{KL}(q(s|x;\phi)||p(s;\theta)) + \frac{V}{2}\log\sigma^2 + \frac{1}{2\sigma^2 K}\sum_{k=1}^{K}||x^{(i)} - \hat{x}(s_{ik};\theta)||_2^2$$

Consider two priors for $p(s)$:

$$p_{spatial}(s) = \prod_{j=1}^{V} p_j(s_j)$$

$$p_{MRF}(s) = \frac{1}{Z}\left[\sum_{j=1}^{V} V_j(s_j) + \sum_{j=1}^{V}\sum_{k\in N_j} V(s_k, s_j)\right]$$

ELBO

In both cases, $D_{KL}(q(s|x;\phi)||p(s;\theta))$ can be calculated exactly. The "extra" term to the ELBO introduced by the MRF prior is:

$$-\sum_{j=1}^{V}\left(\sum_{l_j=0}^{L-1} q_j(l_j|x^{(i)})\sum_{l_k=0}^{L-1}\sum_{k\in N_j} q_k(l_k|x^{(i)})V(s_k=l_k,s_j=l_j)\right)$$

What is this term doing? Introducing an affinity for a certiain label based on the label of neighbors...

ELBO

In both cases, $D_{KL}(q(s|x;\phi)||p(s;\theta))$ can be calculated exactly. The "extra" term to the ELBO introduced by the MRF prior is:

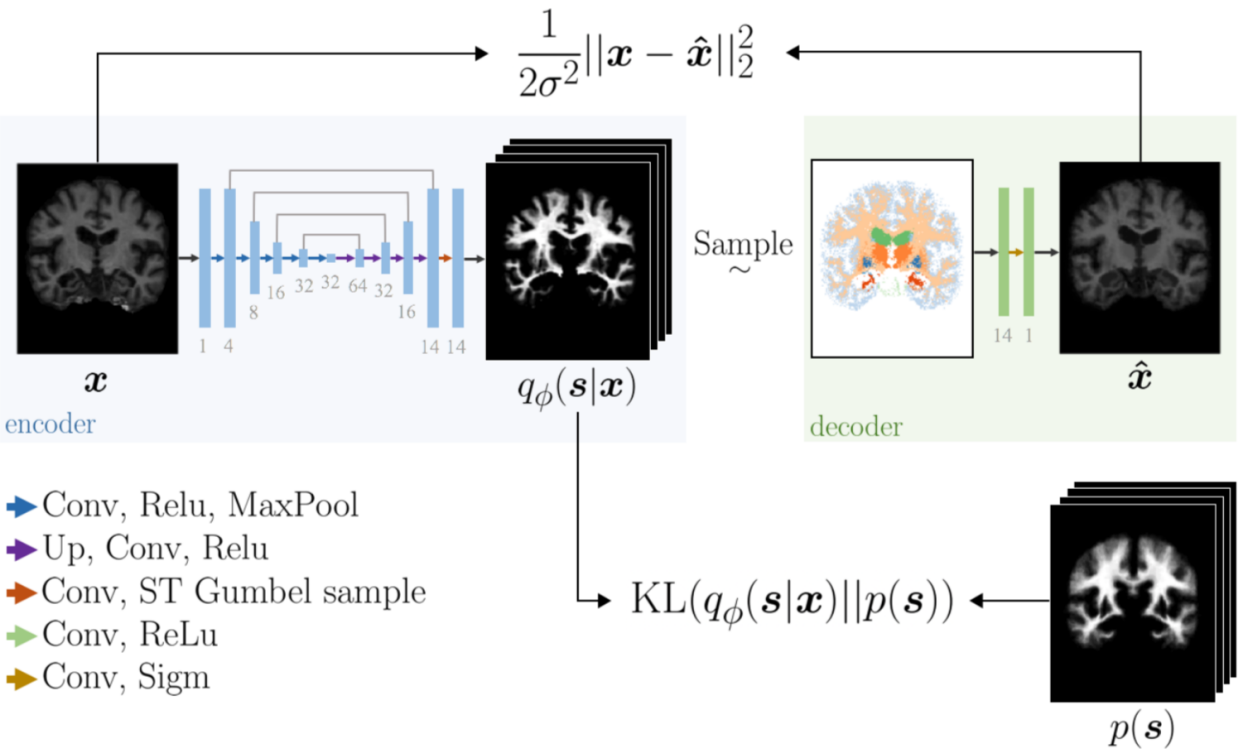
$$-\sum_{j=1}^{V}\left(\sum_{l_j=0}^{L-1} q_j(l_j|x^{(i)})\sum_{l_k=0}^{L-1}\sum_{k\in N_j} q_k(l_k|x^{(i)})V(s_k=l_k,s_j=l_j)\right)$$

What is this term doing? Introducing an affinity for a certiain label based on the label of neighbors...

Note: the parameters of the priors, V_j and $V(l_k,l_j)$ are determined empirically from the atlas data. (Frequency of labels at each voxel, counts of co-occurances in neighboring voxels)

That's it for theory!

Having all the ingredients to evaluate the ELBO, they optimize over the parameters using SGD.



Proposed architecture. The encoder (blue) is a U-Net and decoder (green) is a simple CNN. (Conv) 3x3x3 convolution (Relu) rectified linear unit (Maxpool) 2x downsample (Up) 2x upsample (ST Gumbel) straight through Gumbel softmax (Sigm) sigmoid. The number of channels are displayed below each layer.

Data

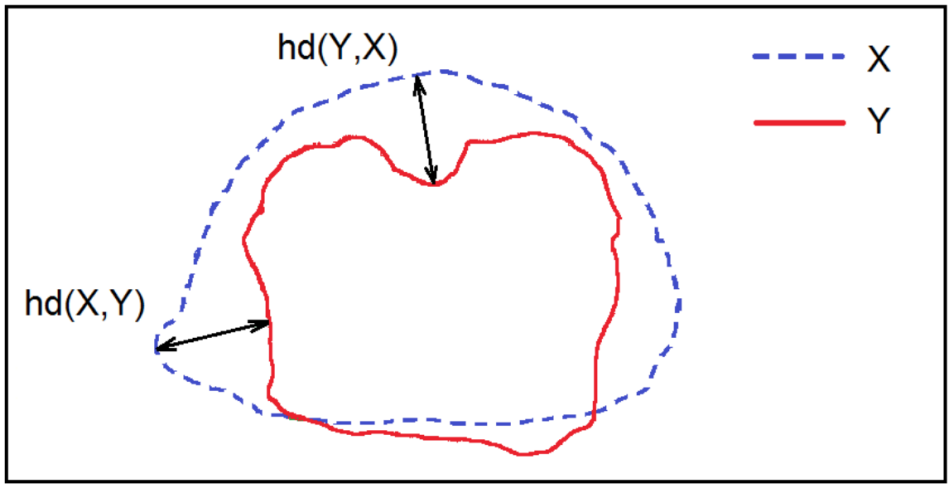
3D brain MRI scans on 38 subjects, preprocessed in various ways. Images were manually segmented into 12 ROIs but this segmentation information was not used in training the model (unsupervised), only for evaluation.

Priors were constructed based on two atlases. Atlas1 is a single segmented image with some Gaussian blurring. Atlas2 is a set of 20 manually labeled subjects.

Evaluation

DICE score: measures degree of overlap (% of voxels) between estimated segmentation and ground-truth (manual) segmentation. 100% is best.

Hausdorff distance: quantifies the distance between boundaries of estimated vs ground-truth segementations. 0 means perfect boundaries.



from Karimi and Sulcudean (2019)

	Performance Measure	
Model	Haussdorff (mm)	Dice Overlap (%)
Baseline1	4.11±0.07	62.82±0.53
EM1 Baseline	4.25±0.09	71.24±0.71
Baseline2	3.50±0.06	71.45±0.65
SAE1 (w/o MRF)	3.88±0.05	74.64±0.30
SAE1 (w MRF)	3.81±0.05	75.36±0.32
EM2 Baseline	2.65±0.05	79.70±0.54
SAE2 (w/o MRF)	2.73±0.04	79.94±0.34
SAE2 (w MRF)	2.68±0.05	80.54±0.36
Supervised	2.23±0.07	84.60±0.26

Model	Test Time (s)
EM	61.07
SAE (CPU)	6.58
SAE (GPU)	1.58

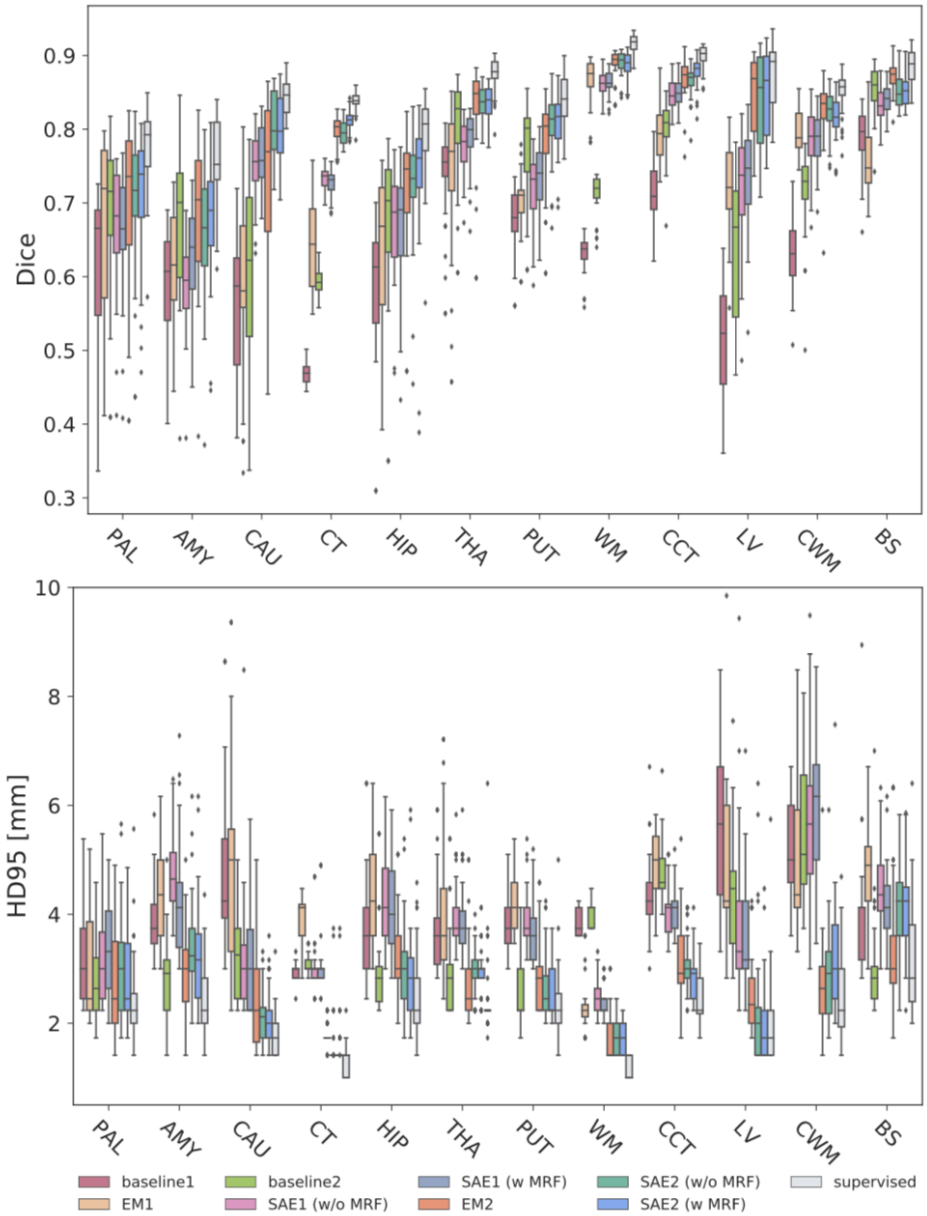
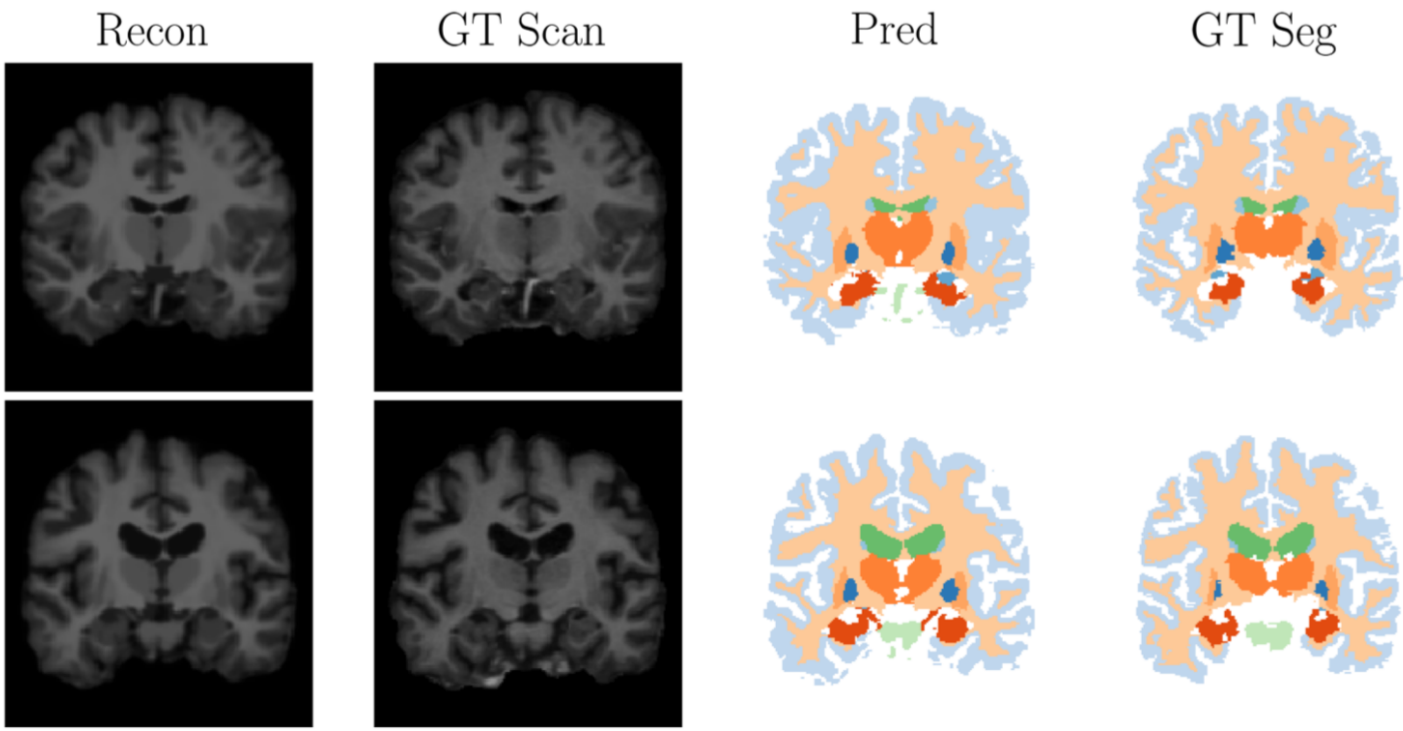


Figure 2: Boxplot of dice and Hausdorff distance. Legend: (PAL) pallidum (AMY) amygdala (CAU) caudate (CT) cerebral cortex (HIP) hippocampus (THA) thalamus (PUT) putamen (WM) white matter (CCT) cerebellar cortex (LV) left ventricle (CMW) cerebral white matter (BS) brainstem.



Representative segmentation results obtained with SAE2 (w/ MRF) on two subjects. Recon is the output of the decoder. GT scan and segmentation are the input MRI and manual segmentation, respectively. Pred is the segmentation obtained through `argmax` of the one-hot encoding $q_\phi(\boldsymbol{s}|\boldsymbol{x}^{(i)})$.