

Multilevel Segmentation and Integrated Bayesian Model Classification

Jason J. Corso

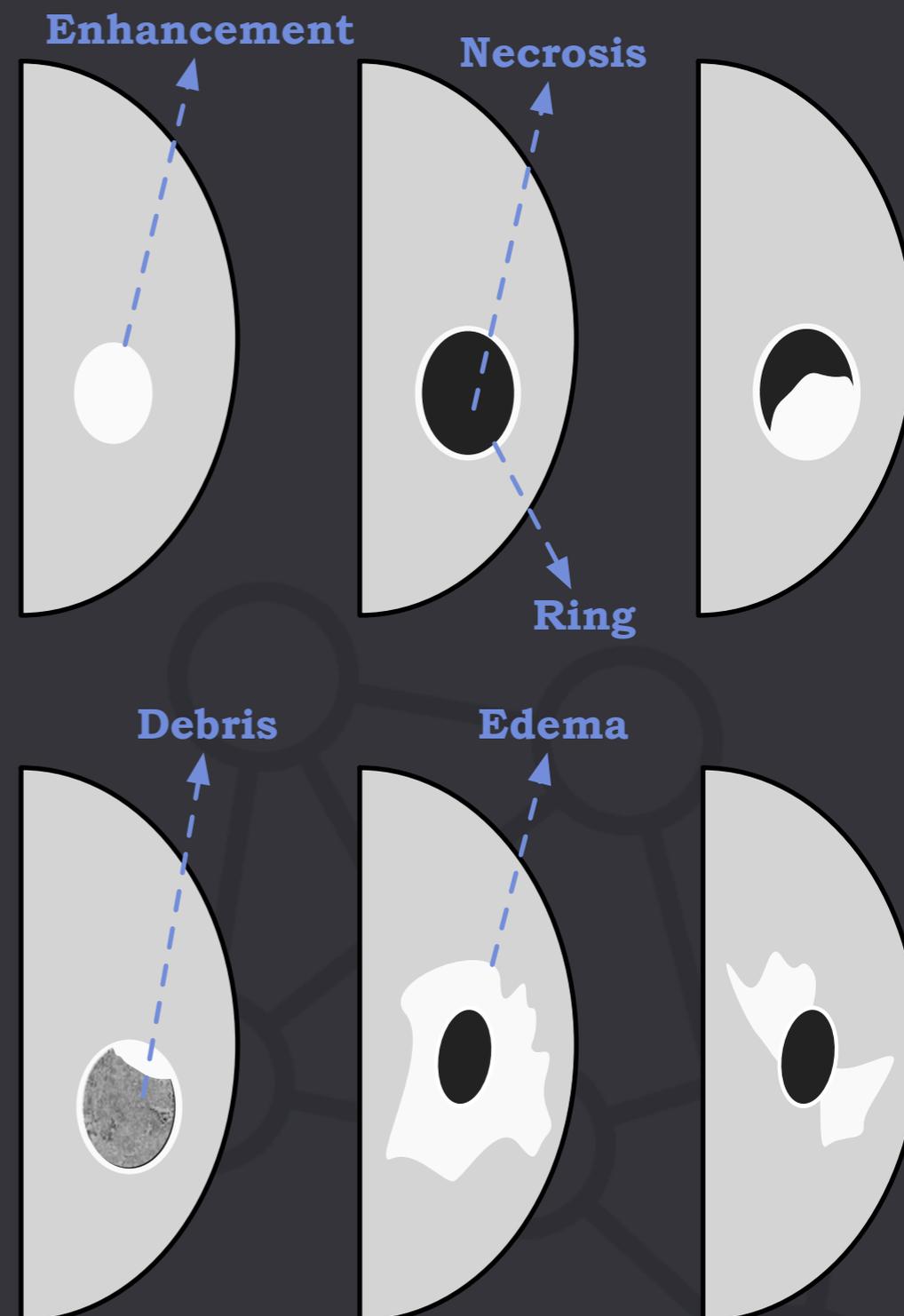
Postdoctoral Research Fellow

Laboratory of Neuro Imaging
Center for Computational Biology
Center for Imaging and Vision Science
Departments of Neuroscience and Statistics
University of California, Los Angeles
jcorso@ucla.edu

Motivation

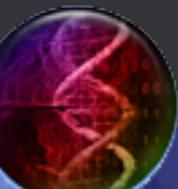
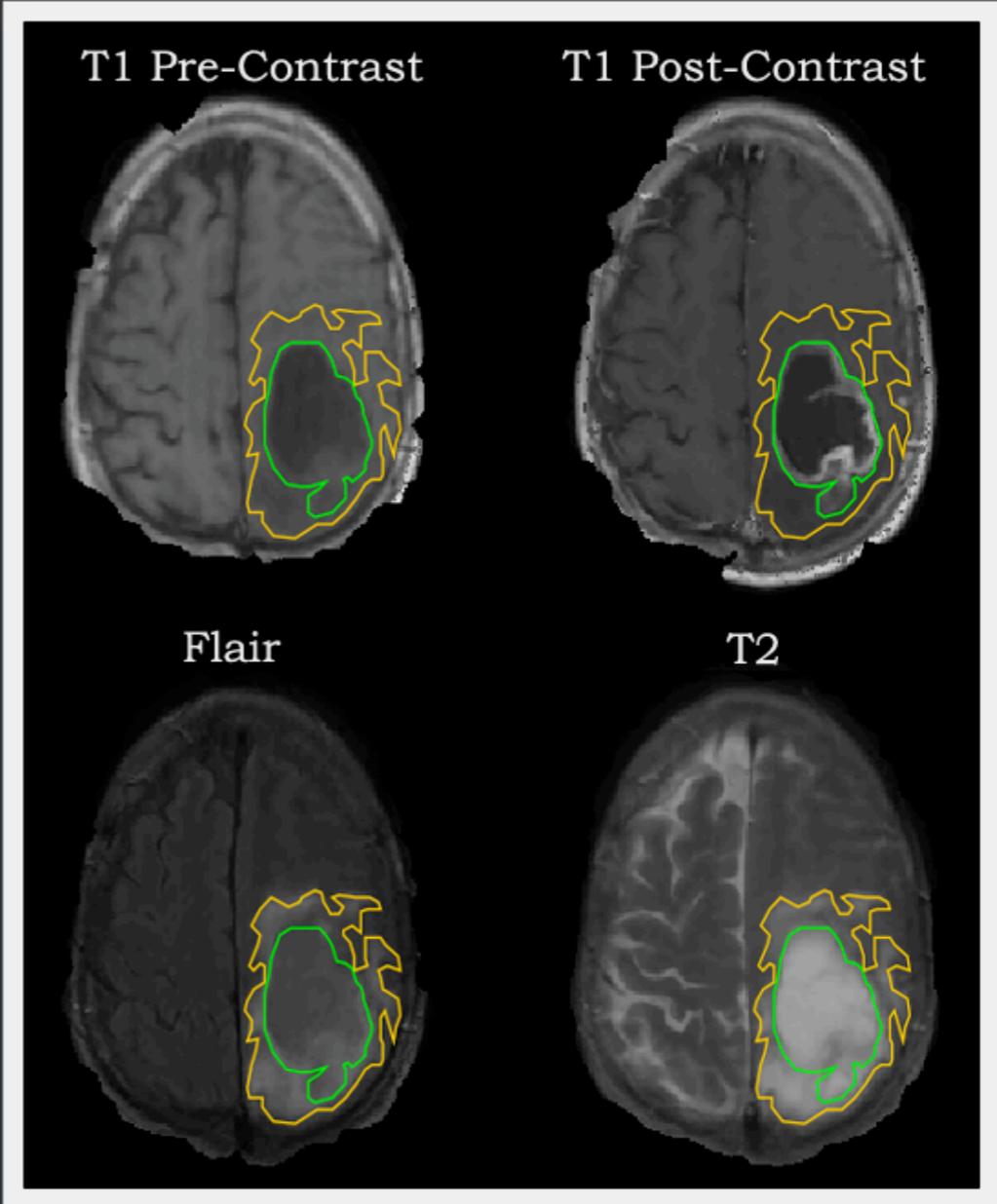
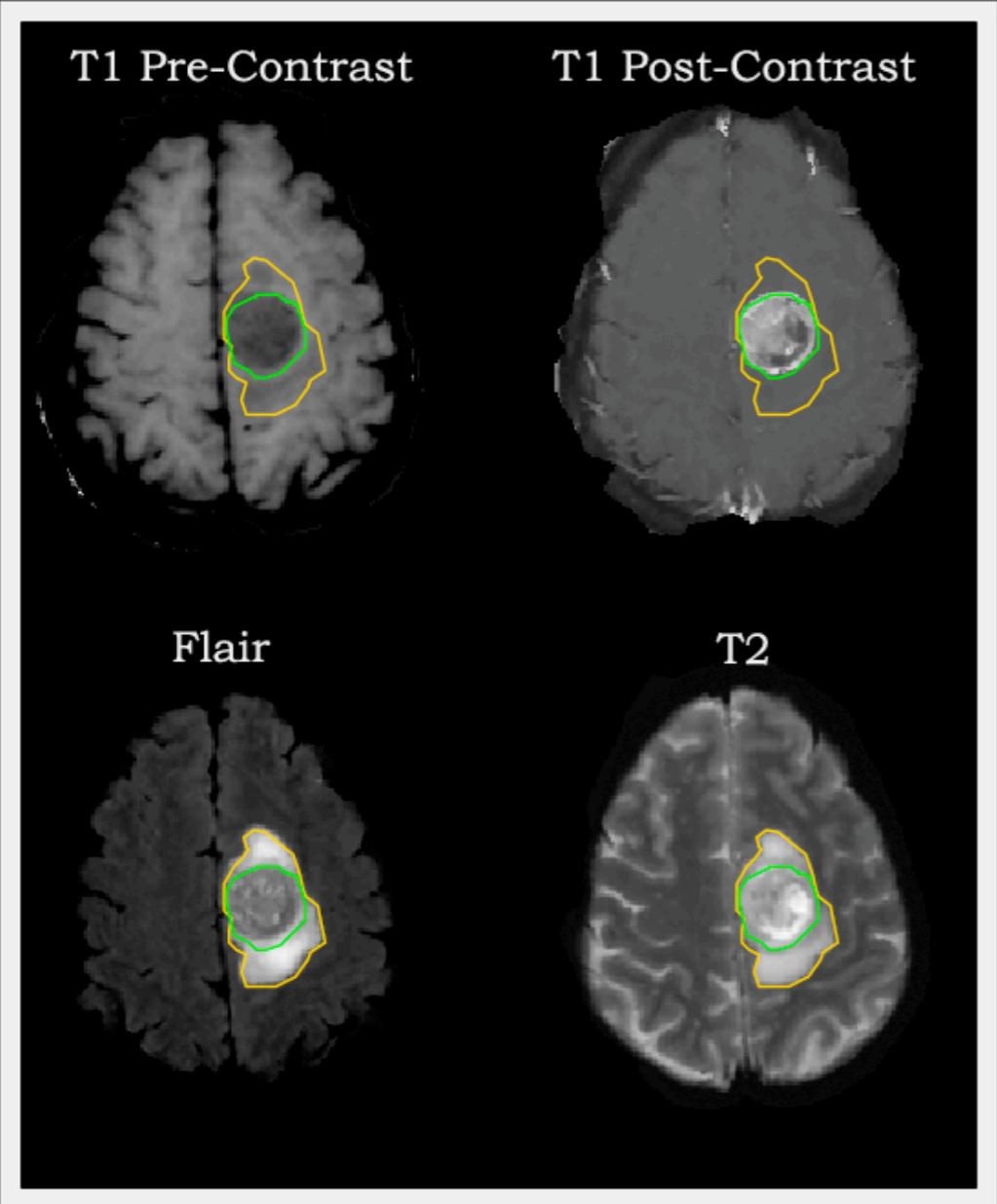
Segmentation of Brain Tumor is Important but Difficult.

- Important
 - Would quantify tumor growth.
 - Currently, only manual or interactive.
 - Aid in surgical planning.
 - Provide data for treatment analysis.
- Difficult
 - *Glioblastoma multiforme*.
 - Highly varying in appearance, size and shape.
 - Can deform nearby structures.
 - Impossible with simple thresholding.
 - Multiple modalities necessary.



Motivation

Two Typical Examples

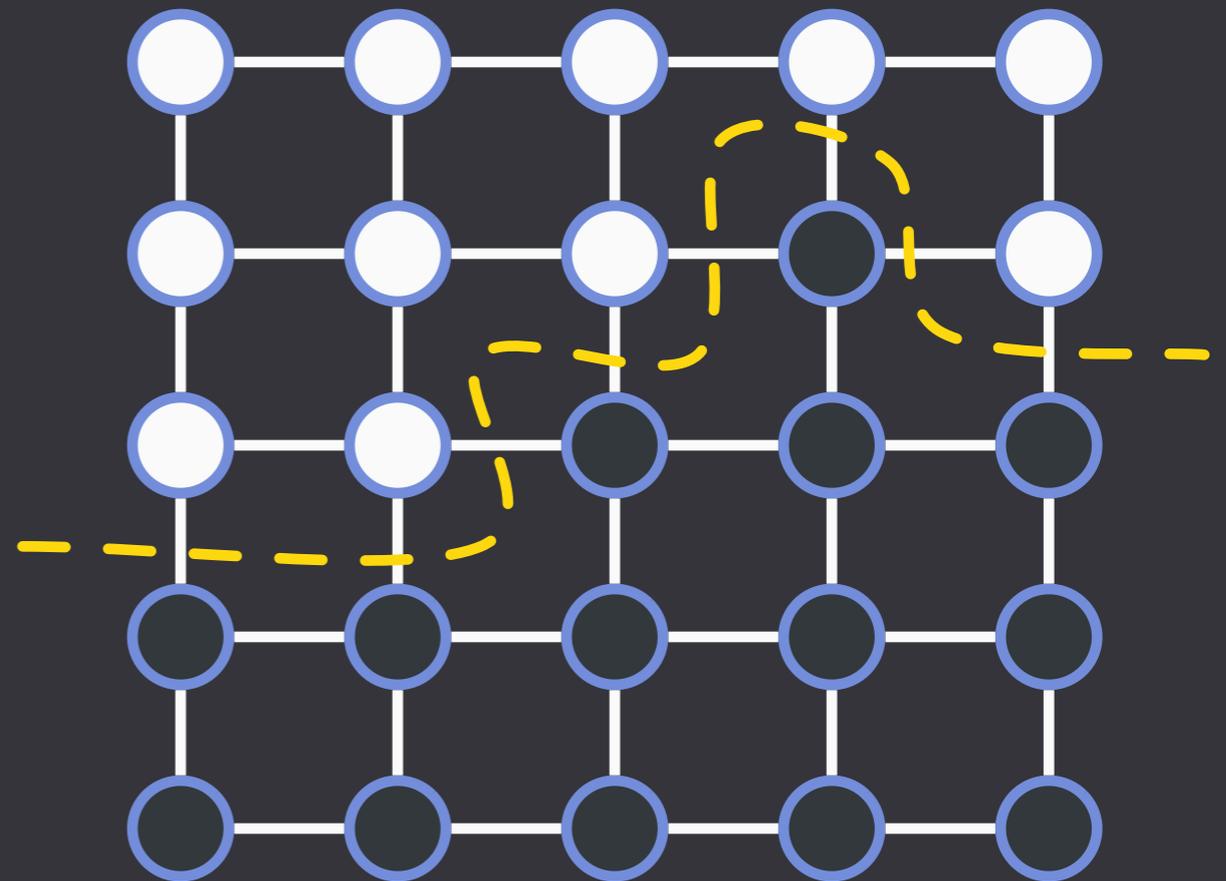
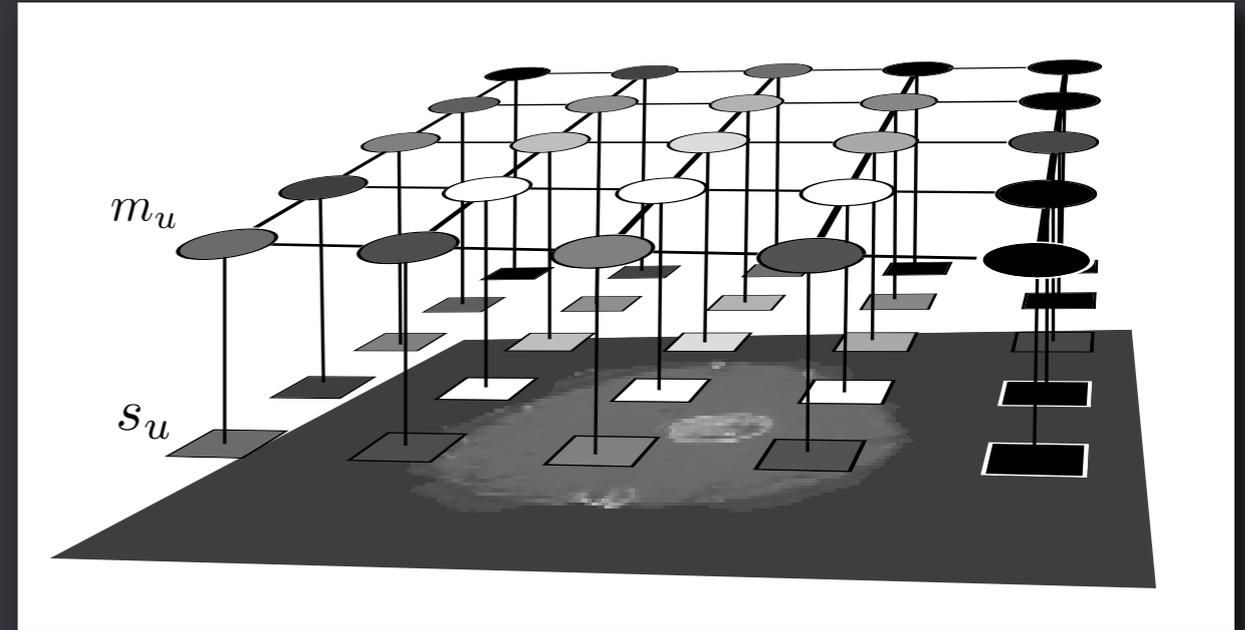


Graph Affinity-Based Segmentation

- Define the problem on a graph:

$$G = \{\mathcal{V}, \mathcal{E}\}$$

- Edges are sparse, to neighbors.
- Each pixel / voxel is a node.
- Augment nodes, for $v \in \mathcal{V}$
 - statistics: s_v .
 - class label: c_v .
- Define affinity between $u, v \in \mathcal{V}$
$$w_{uv} = \exp(-D(s_u, s_v; \theta))$$
,
 - where D is some non-negative distance function and θ are some predetermined values.
- Regions are defined by **cuts**.



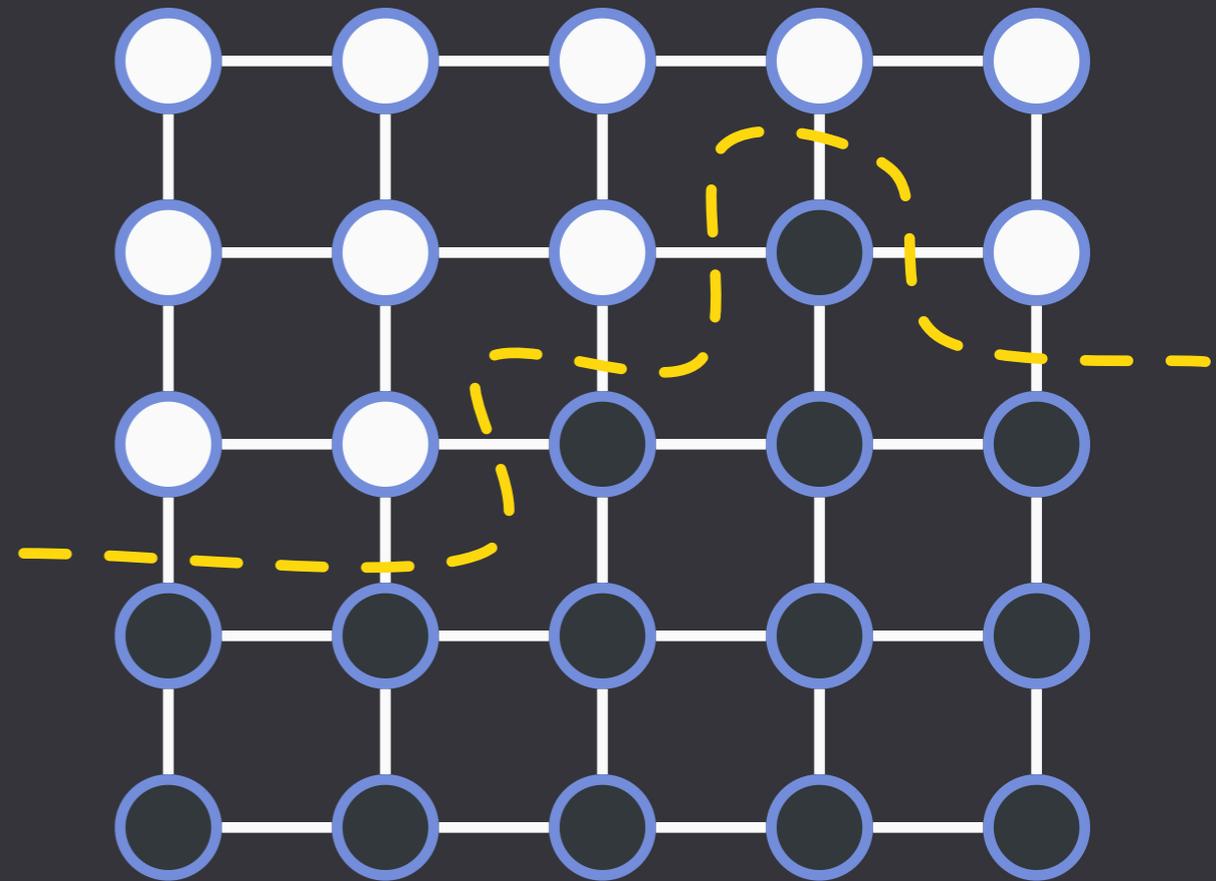
Graph Affinity-Based Segmentation

Segmentation by Affinities

- Define a region saliency measure

$$\Gamma(R) = \frac{\sum_{u \in R, v \notin R} w_{uv}}{\sum_{u, v \in R} w_{uv}} .$$

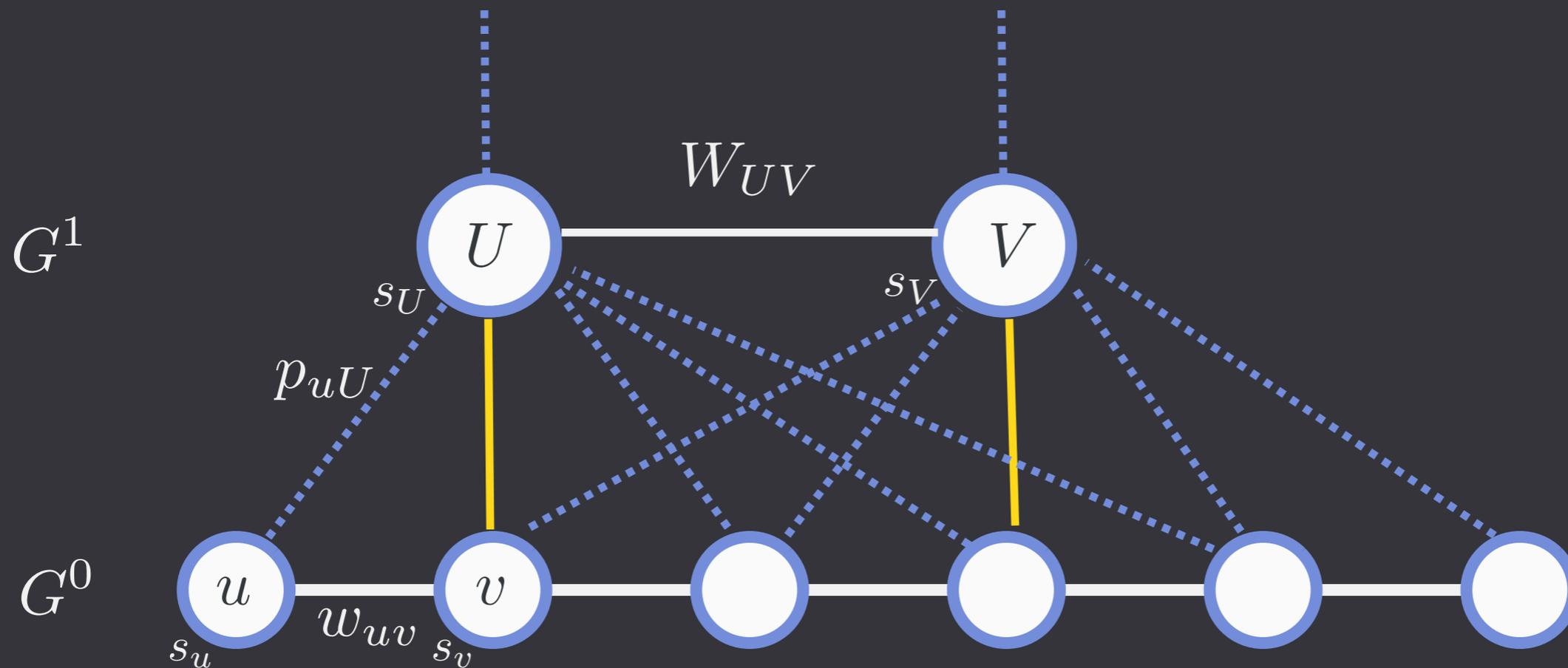
- Low $\Gamma(R)$ means good saliency:
 - low affinity on boundaries,
 - high affinity in interior.
- Normalized Cut (Shi & Malik)
 - Affinities at a pixel scale only.
 - Can take filters of varying size but erroneous at borders.



Graph Affinity-Based Segmentation

Segmentation by Weighted Aggregation

- Invented in natural image domain by Sharon et al. (CVPR 2000, 2001).
- First extended to medical imaging domain by Akselrod-Ballrin (CVPR 2006)
- **Efficient, multiscale process.**
- Results in a pyramid of recursively coarsened graphs that capture multiscale properties of the data.
- Affinities are calculated at each level of the graph.
- **Statistics in each graph node are agglomerated up the hierarchy.**



Segmentation with Generative Models

- Define a likelihood function $P(\{s_u\}|\{c_u\})$,
- And define priors on the class labels (models): $P(\{c_u\})$.
- Goal is to seek estimates of class labels that maximize the posterior:

$$P(\{c_u\}|\{s_u\}) \propto P(\{s_u\}|\{c_u\})P(\{c_u\}) .$$

- Segmentation and classification solved jointly.
- Exemplified by the DDMMCMC algorithm of Tu and Zhu.

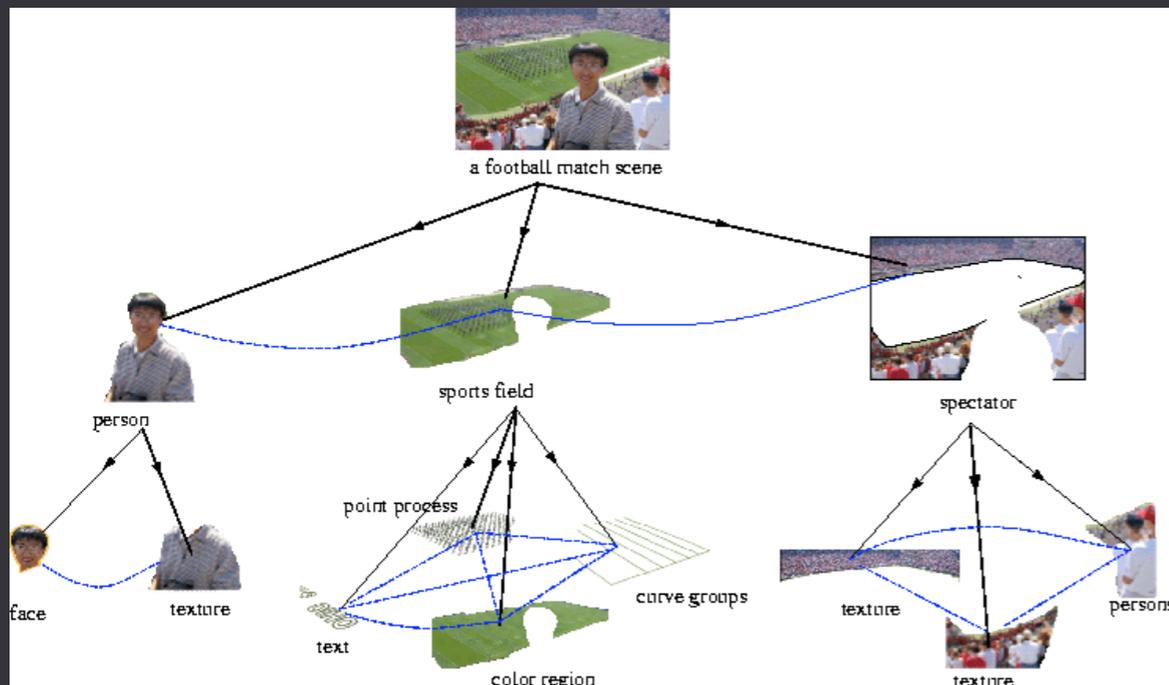
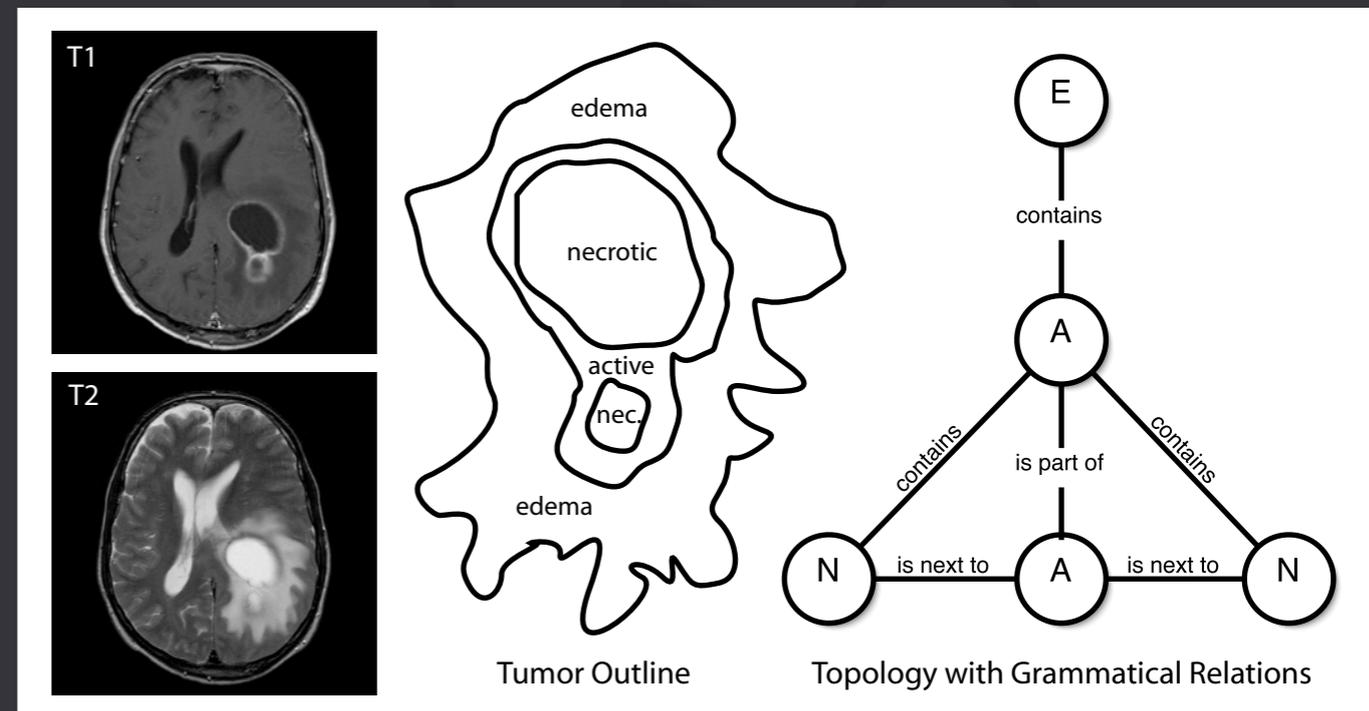


Image From Tu et al. IJCV 2005



Stochastic grammars are another example.

Comparison

- Graph Affinity-Based Methods
 - Rapid, feed-forward methods giving quick results.
 - Evidence of much power in these methods.
 - Not guaranteed to get meaningful answers.
 - Very memory intensive (especially with 3D medical data).
- Model-Based MAP Methods
 - Guaranteed to get answer from posterior.
 - Very computationally expensive.
 - The models are hard to design and train.
- Suggests a combination of the two.

Research Agenda: leverage the efficiency of graph-based bottom-up methods with the power and optimality of top-down generative models.



Combining Affinities and Models

- Define class dependent affinities in terms of probabilities:

$$\hat{P}(X_{uv} | s_u, s_v) = w_{uv} .$$

- X_{uv} is the binary event that both u and v are in the same region.



- X_{uv} is not deterministic on class labels (c_u, c_v) :
 - pixels in same region may have different labels,
 - pixels in different regions may have same labels.



Combining Affinities and Models

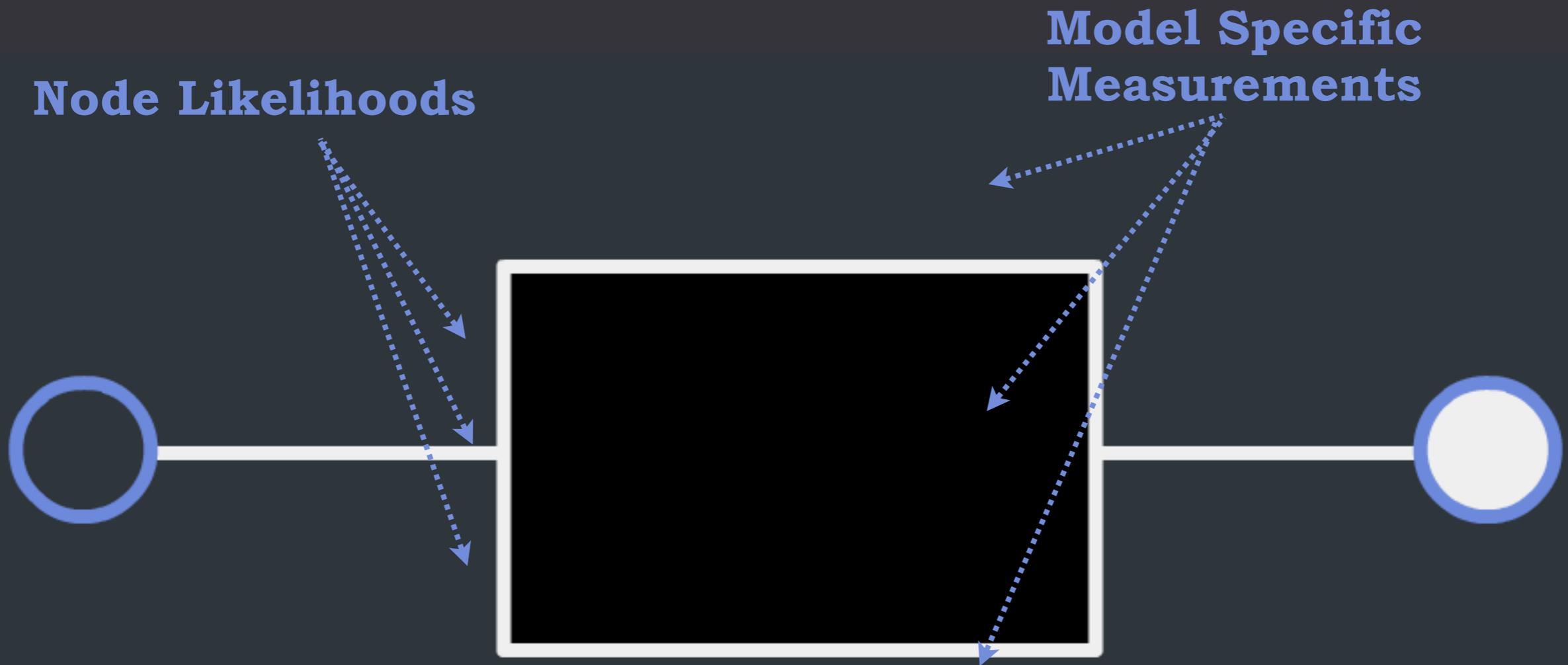
The Conventional Affinity Function

$$w_{uv} = \exp(-D(s_u, s_v; \theta))$$



Combining Affinities and Models

Model-Aware Affinity



The node likelihoods weigh the model specific measurements.



Combining Affinities and Models

Model-Aware Affinity

- Treat class labels as hidden variables and sum them out:

$$\begin{aligned} P(X_{uv} | s_u, s_v) &= \sum_{c_u} \sum_{c_v} P(X_{uv} | s_u, s_v, c_u, c_v) P(c_u, c_v | s_u, s_v) , \\ &\propto \sum_{c_u} \sum_{c_v} P(X_{uv} | s_u, s_v, c_u, c_v) P(s_u, s_v | c_u, c_v) P(c_u, c_v) , \\ &= \sum_{c_u} \sum_{c_v} \underbrace{P(X_{uv} | s_u, s_v, c_u, c_v)}_{\text{Model Specific Measurement}} \underbrace{P(s_u | c_u) P(s_v | c_v)}_{\text{Node Likelihoods}} \underbrace{P(c_u, c_v)}_{\text{Class Prior}} . \end{aligned}$$



Model-Aware Affinity

$$P(X_{uv} | s_u, s_v, c_u, c_v) = \exp\left(-D(s_u, s_v; \theta[c_u, c_v])\right)$$

- Defined as a function of class label pairs $\theta[c_u, c_v]$.
- Suitable for heterogeneous data.
 - Model-aware affinities modulated by evidence: $P(s_u | c_u)$.
 - At pixel scales, evidence for any class is low.
 - At coarser scales, evidence is high and models improve affinities.



Segmentation by Weighted Aggregation

- Overview
 - Multiscale graph-based algorithm.
 - Inspired by methods in Algebraic Multigrid.
- First extended to 3D medical domain by Akselrod-Ballrin (2006)
- Again, define a graph $G^t = (\mathcal{V}^t, \mathcal{E}^t)$
 - Superscript denotes level in a pyramid of graphs $\mathcal{G} = \{G^t : t = 0, \dots, T\}$
- Finest layer induced by voxel lattice
 - 6-neighbor connectivity
 - Node properties s_v set according to multimodal image intensities.
 - Affinities initialized by L1-distance: $w_{uv} = \exp(-\theta |s_u - s_v|_1)$.

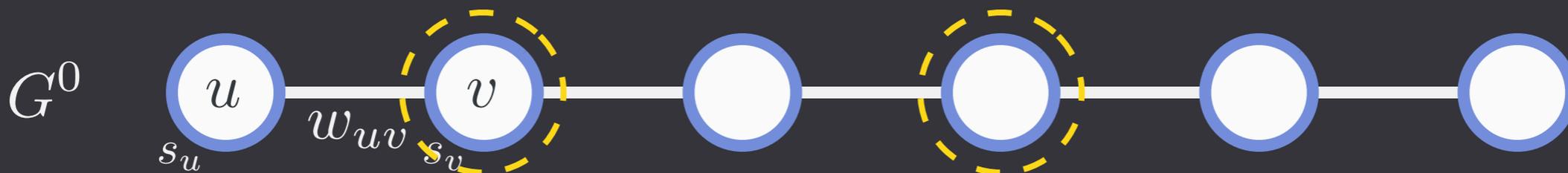


Segmentation by Weighted Aggregation

- Select a representative set of nodes satisfying

$$\sum_{v \in \mathcal{R}^t} w_{uv} \geq \beta \sum_{v \in \mathcal{V}^t} w_{uv}$$

- i.e., all nodes in finer level have strong affinity to nodes in coarser.

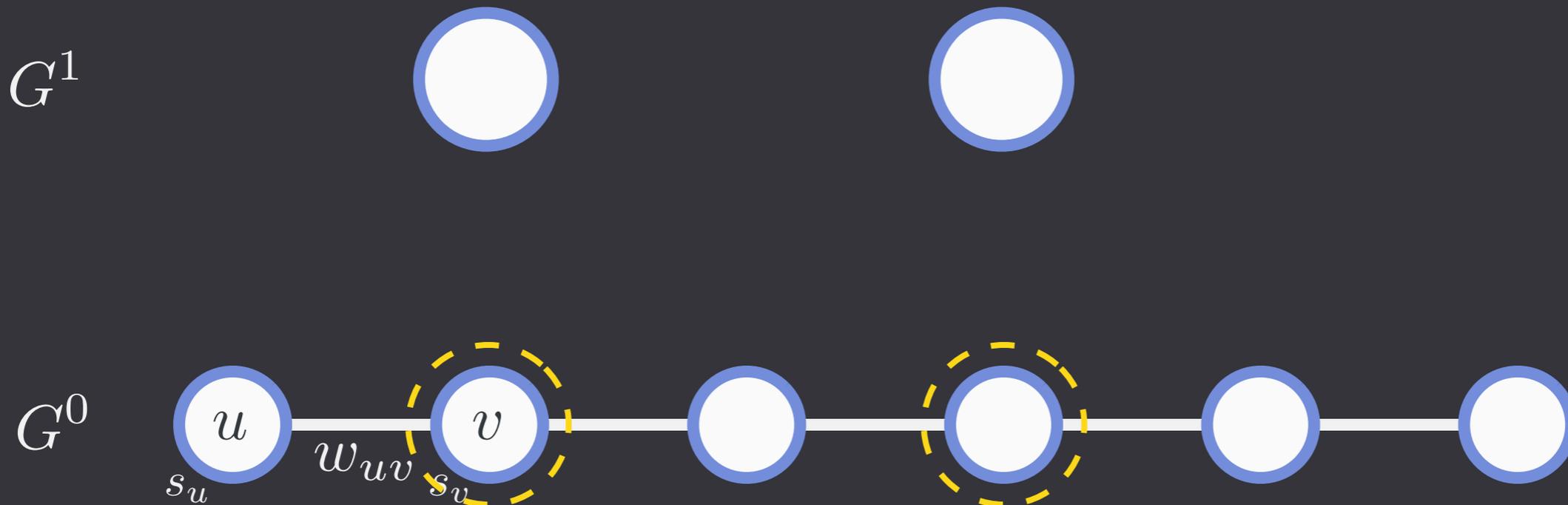


Segmentation by Weighted Aggregation

- Select a representative set of nodes satisfying

$$\sum_{v \in \mathcal{R}^t} w_{uv} \geq \beta \sum_{v \in \mathcal{V}^t} w_{uv}$$

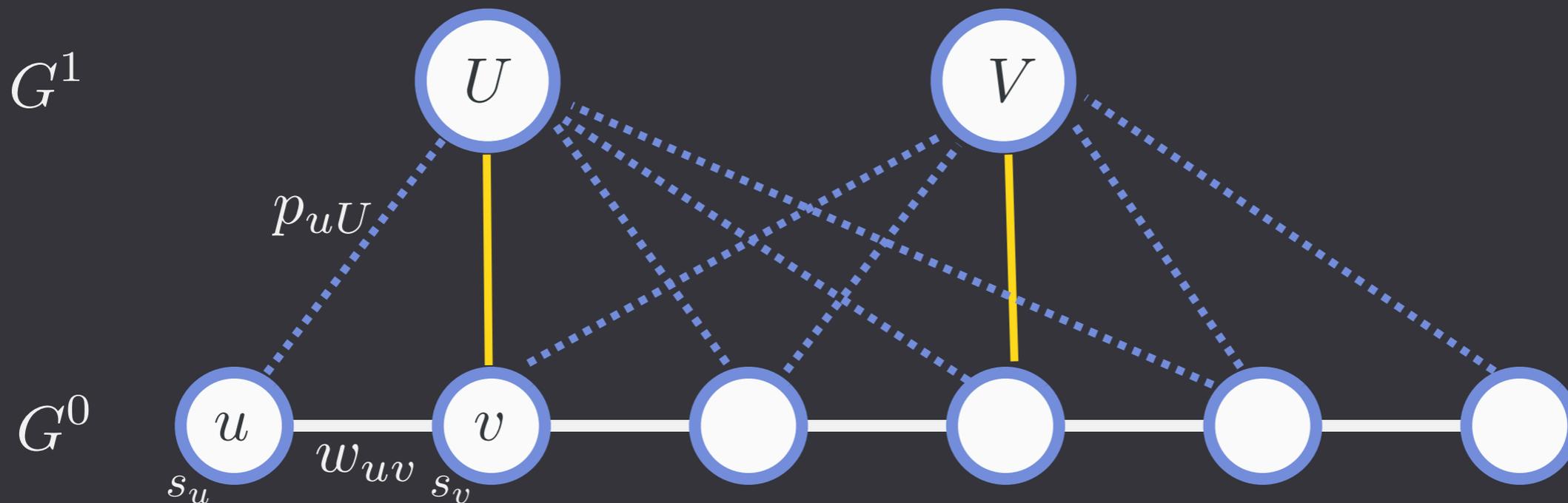
- i.e., all nodes in finer level have strong affinity to nodes in coarser.
- Begin to define graph $G^1 = (\mathcal{V}^1, \mathcal{E}^1)$



Segmentation by Weighted Aggregation

- Compute interpolation weights between coarse and fine levels

$$p_{uU} = \frac{w_{uU}}{\sum_{V \in \mathcal{V}^{t+1}} w_{uV}}$$



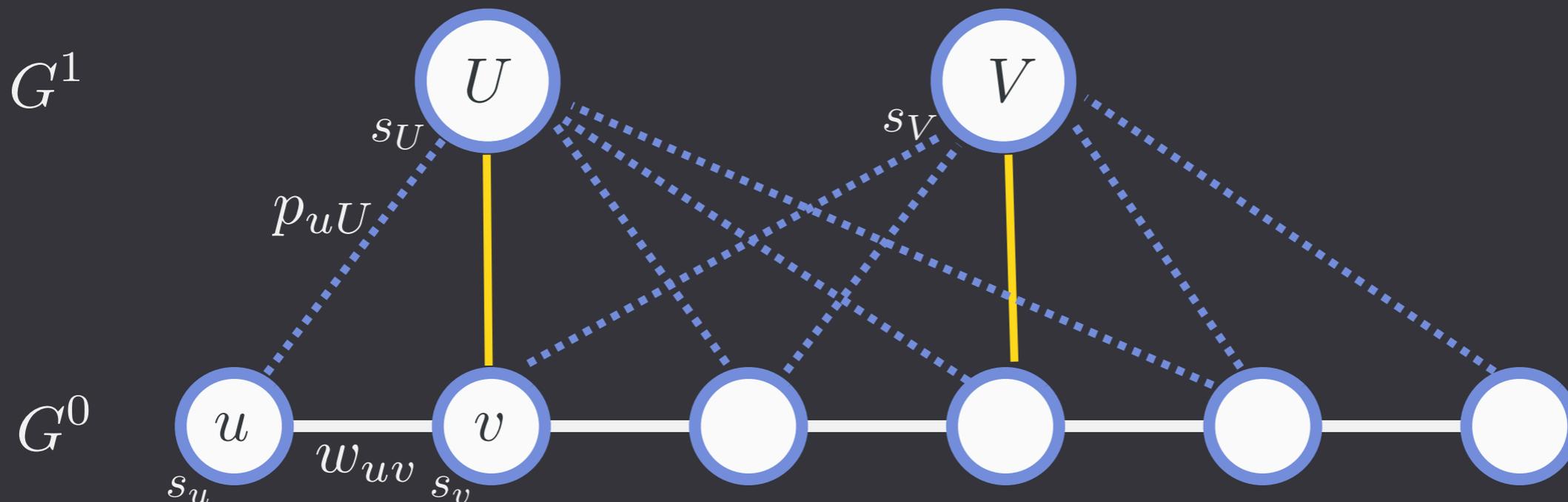
Segmentation by Weighted Aggregation

- Compute interpolation weights between coarse and fine levels

$$p_{uU} = \frac{w_{uU}}{\sum_{V \in \mathcal{V}^{t+1}} w_{uV}}$$

- Accumulate statistics at the coarse level

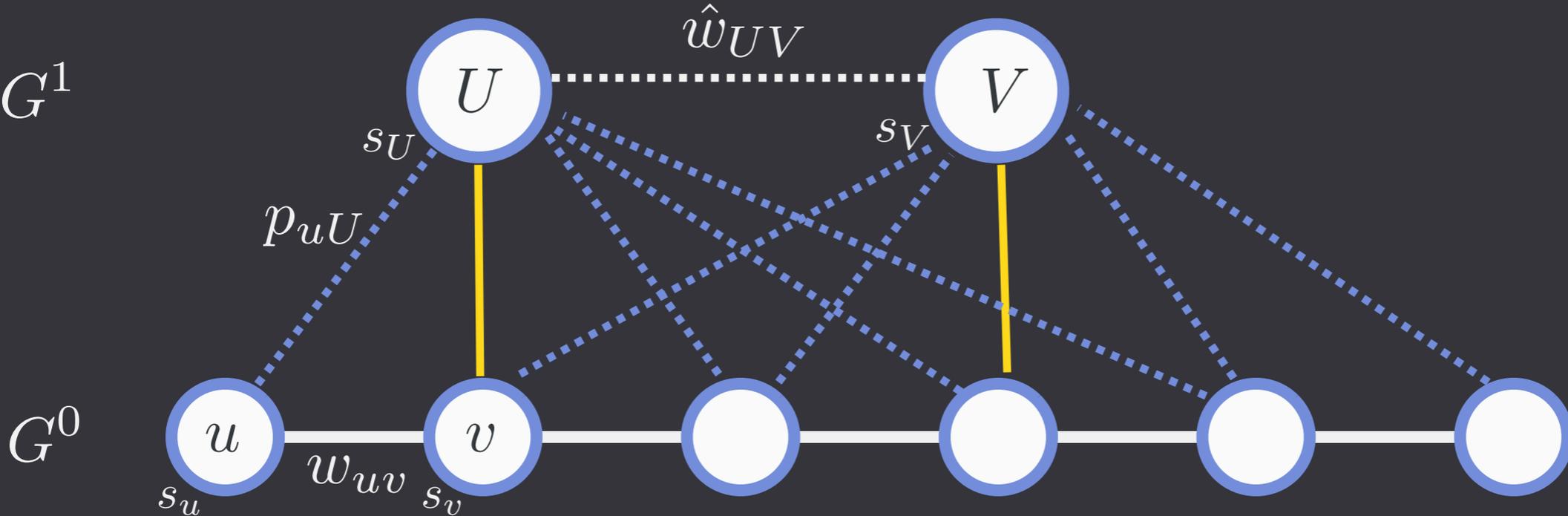
$$s_U = \sum_{u \in \mathcal{V}^t} \frac{p_{uU} s_u}{\sum_{v \in \mathcal{V}^t} p_{vU}}$$



Segmentation by Weighted Aggregation

- Interpolate affinity from finer levels

$$\hat{w}_{UV} = \sum_{(u \neq v) \in \mathcal{V}^t} p_{uU} w_{uv} p_{vV}$$



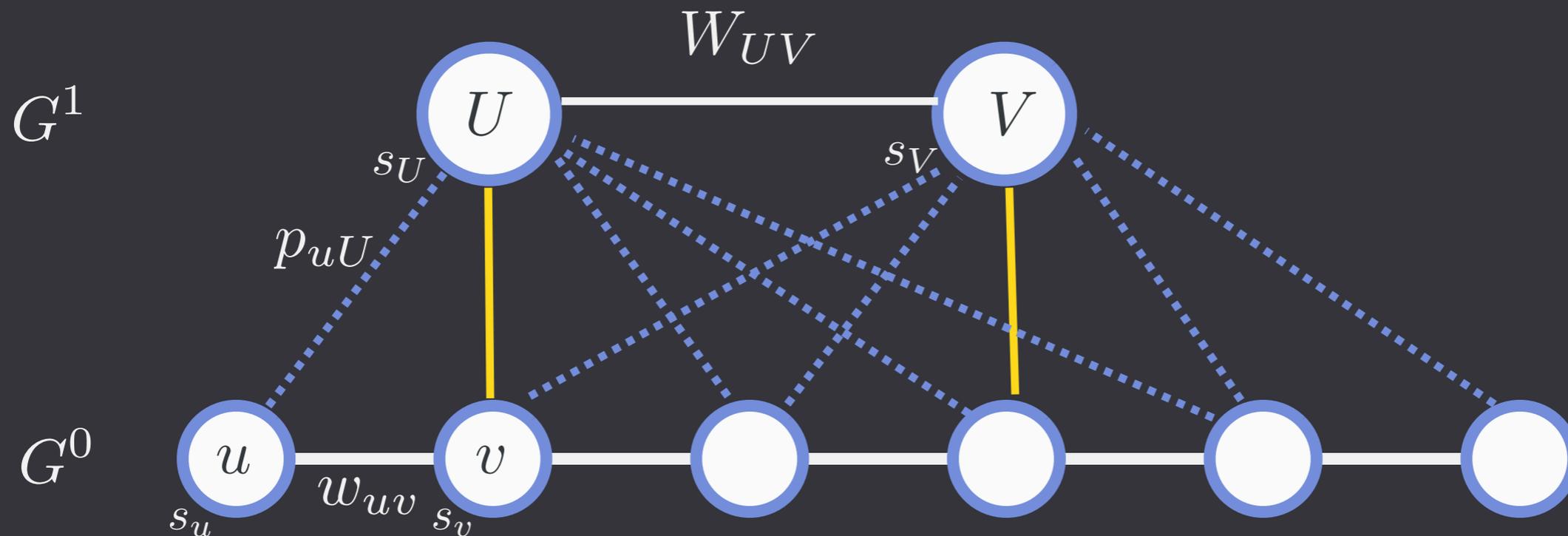
Segmentation by Weighted Aggregation

- Interpolate affinity from finer levels

$$\hat{w}_{UV} = \sum_{(u \neq v) \in \mathcal{V}^t} p_{uU} w_{uv} p_{vV}$$

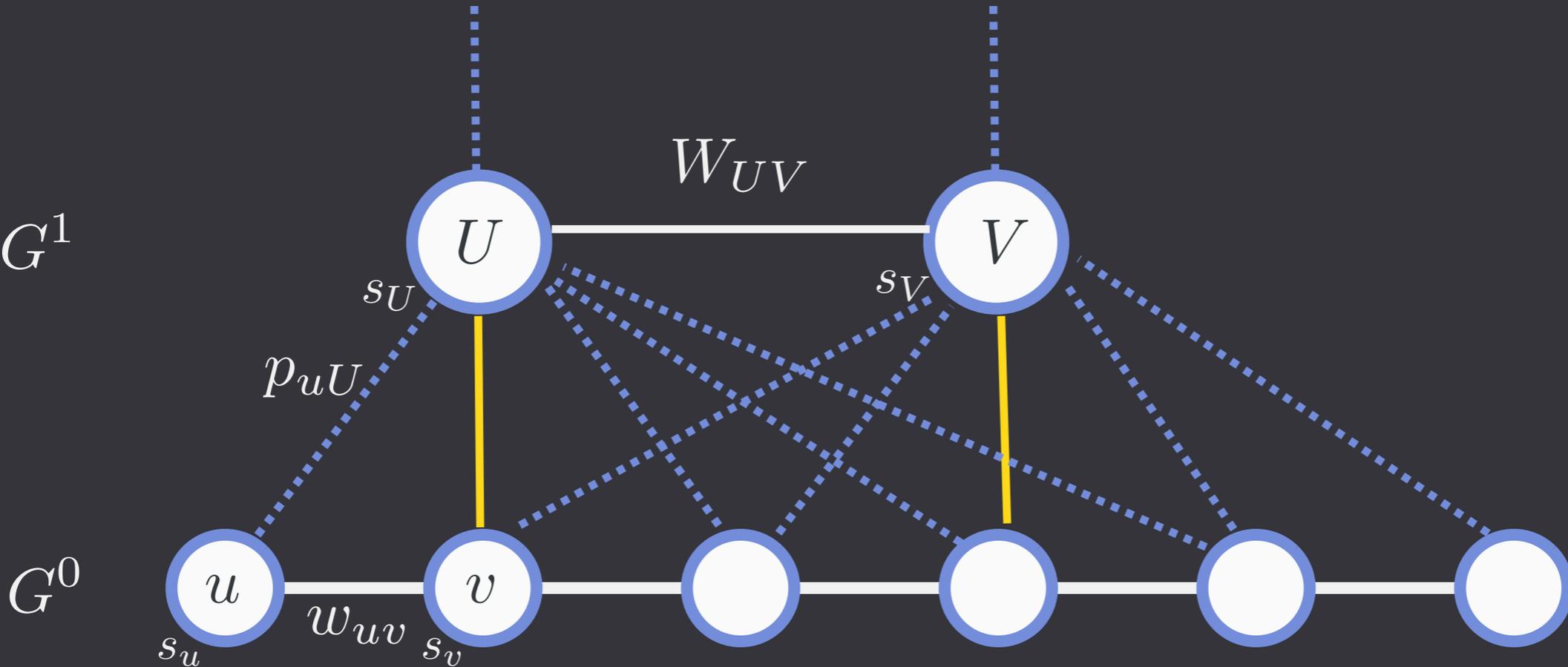
- Use coarse affinity to modulate the interpolated affinity

$$W_{UV} = \hat{w}_{UV} \exp(-D(s_U, s_V; \theta))$$



Segmentation by Weighted Aggregation

- Repeat ...



Segmentation by Weighted Aggregation

Incorporating Model-Aware Affinities

- Alter the way coarse affinities are modulated.
- Currently

$$W_{UV} = \hat{w}_{UV} \exp(-D(s_U, s_V; \theta)) .$$

- Change to

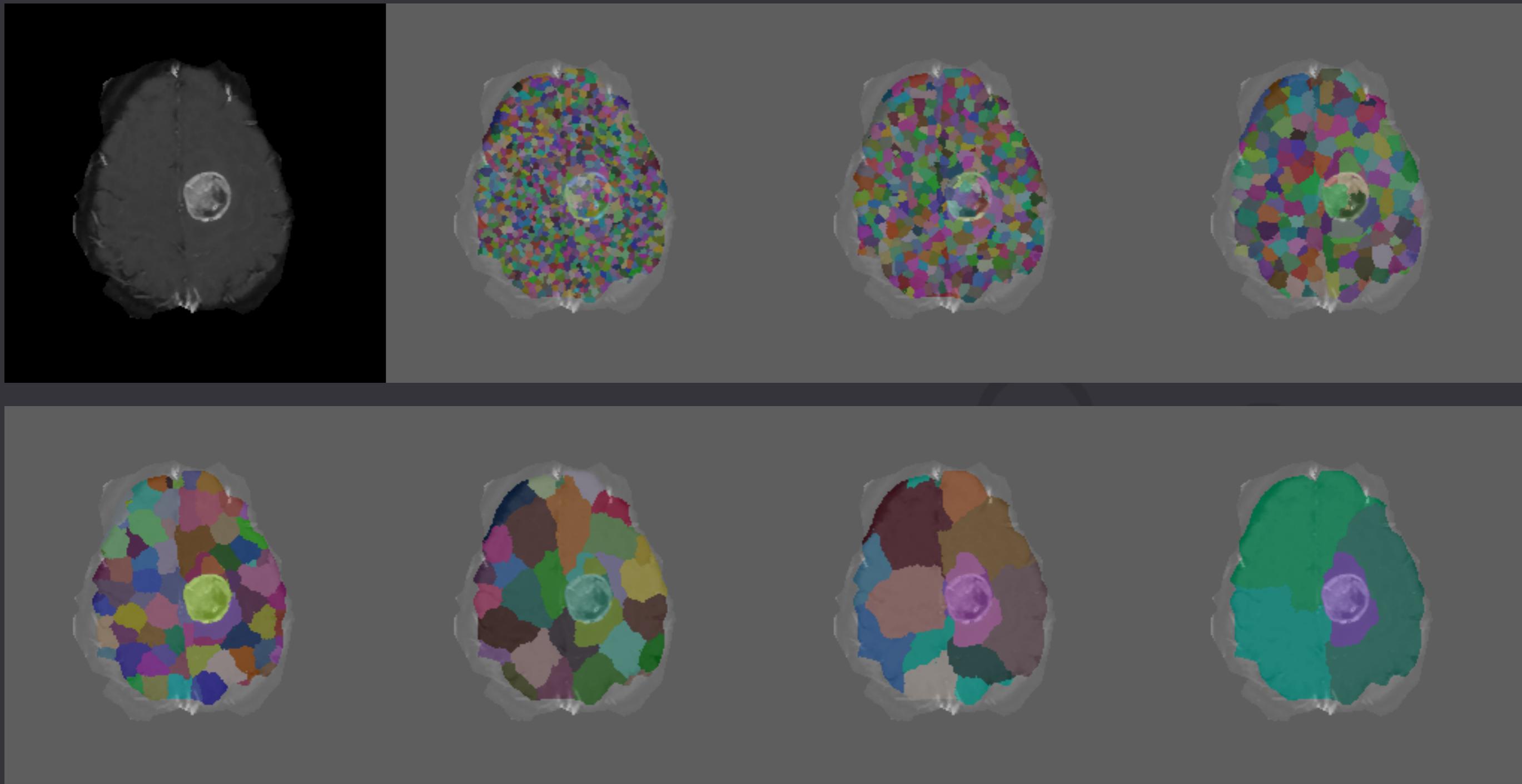
$$W_{UV} = \hat{w}_{UV} P(X_{UV} | s_U, s_V) .$$

- Associate the most-likely class with each node:

$$c_U^* = \arg \max_{c \in \mathcal{C}} P(s_U | c) .$$



Example of the Segmentation Pyramid



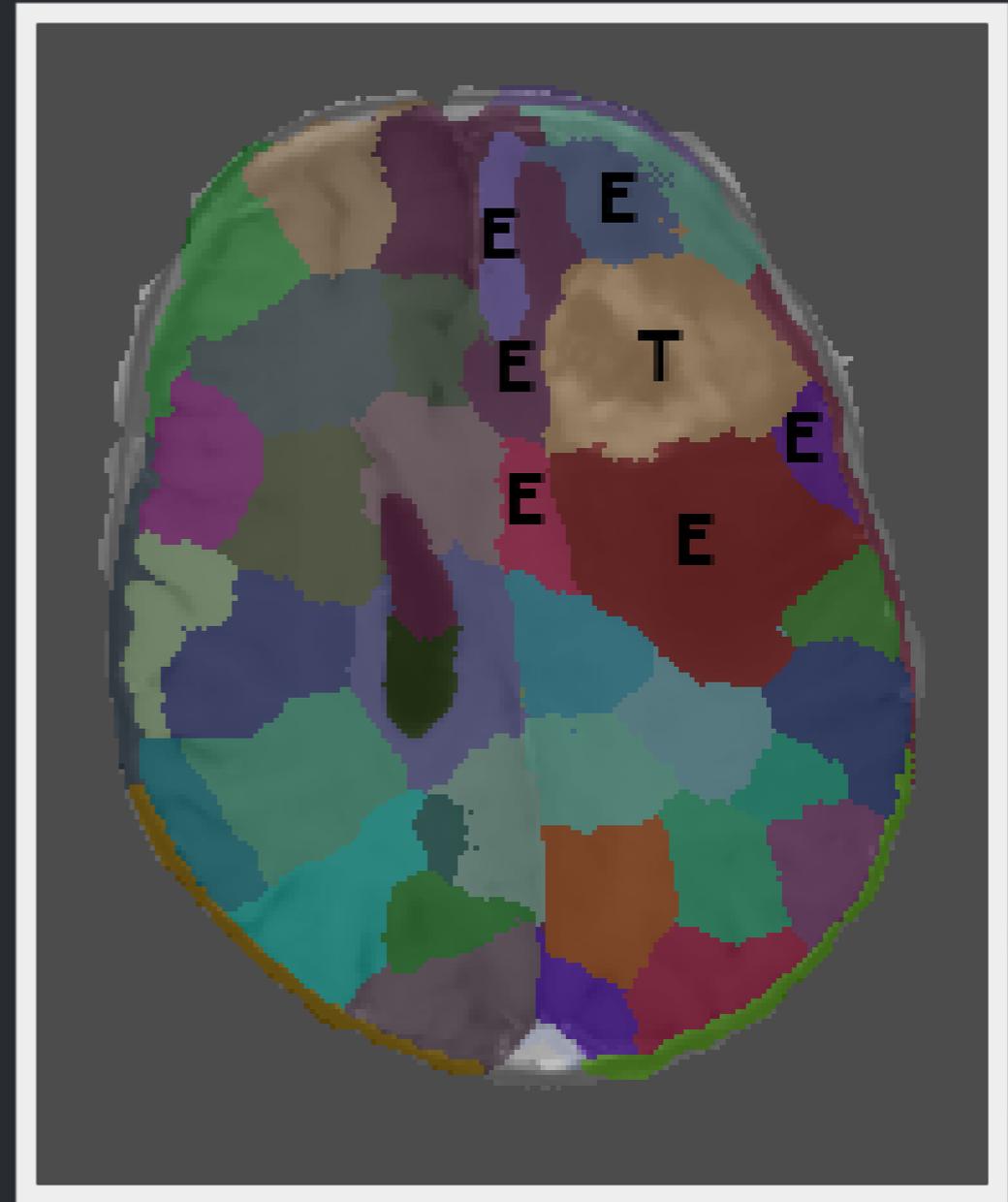
Extracting Information from the Hierarchy

The Problem:

- Gives hierarchy, need final answer.
- Original SWA suggests saliency.
- Model information provides more information for extraction.

Current Model-Based Extraction

- Compute class for each voxel for each level in the pyramid.
- Use interpolation weights.
- Accumulate class membership.
- Assign class the voxel was for most of the levels.



Extracting Information from the Hierarchy

Energy Based Methods -- Saliency Only

- Treat hierarchy as an input datum to a minimizer.
- Define a set of level variables $\{l_v\}, v \in \mathcal{V}$ over voxel lattice.
- A final segmentation is induced by an instance.

$$H_s(\{l\}) = \alpha_1 \sum_{\{i\}} \Gamma(v_i^{(l_i)}) - \alpha_2 \sum_{\langle i,j \rangle} \mathbf{1}(l_i = l_j)$$

“External” Potential

Pair Potential

- Can be minimized various techniques
 - Formulate as Gibbs Field and use simulated annealing.
 - Use popular min-cut/max-flow graph cuts method.



Extracting Information from the Hierarchy

Energy Based Methods -- Model-Based

- Models provide stronger information for extraction.
- Define a set of model variables $\{m_v\}$, $v \in \mathcal{V}$ over voxel lattice.
- Bayesian estimate of model likelihood over hierarchy:

$$\begin{aligned} P(m_i, \hat{s}_i) &= \sum_{t=\{0, \dots, T\}} P(\hat{s}_i, m_i, t) \\ &= \sum_{t=\{0, \dots, T\}} P(s_{v_i^{(t)}} | m_i) E(\hat{s}_i, t) \end{aligned}$$

- $E(\hat{s}_i, t)$ is level evidence, computed by entropy of likelihood dist:

$$E(\hat{s}_i, t) = \frac{\exp\left(-\kappa \sum_{m \in \mathcal{M}} P(s_{v_i^{(t)}} | m) \ln P(s_{v_i^{(t)}} | m)\right)}{\sum_{z=\{0, \dots, T\}} \exp\left(-\kappa \sum_{m \in \mathcal{M}} P(s_{v_i^{(z)}} | m) \ln P(s_{v_i^{(z)}} | m)\right)}$$



Extracting Information from the Hierarchy

Energy Based Methods -- Model-Based

- Define a similar Potts-type energy model:

$$H_m(\mathcal{M}) = -\alpha_1 \sum_{\{i\}} P(m_i, \hat{s}_i) - \alpha_2 \sum_{\langle i,j \rangle} \mathbf{1}(m_i = m_j)$$

- Use energy minimization or stochastic optimization techniques

- Min-flow/max-cut methods.

- Simulated annealing on Gibbs field: $\Pi_M(\mathcal{M}) = \frac{1}{Z} \exp\left(-\tau H(\{l\})\right)$

- Swendseng-Wang cuts similar to Barbu and Zhu.



Extracting Information from the Hierarchy

Activate Top-Down Generative Models

Treat the hierarchy as a set of model “proposals.”

Becomes plausible to use global object properties (e.g., shape).

Towards a comprehensive medical image parsing framework.

Keys to making these methods more robust and efficient.

Integrate Top-Down Generative Models

Drive hierarchical agglomeration by the top-down models.

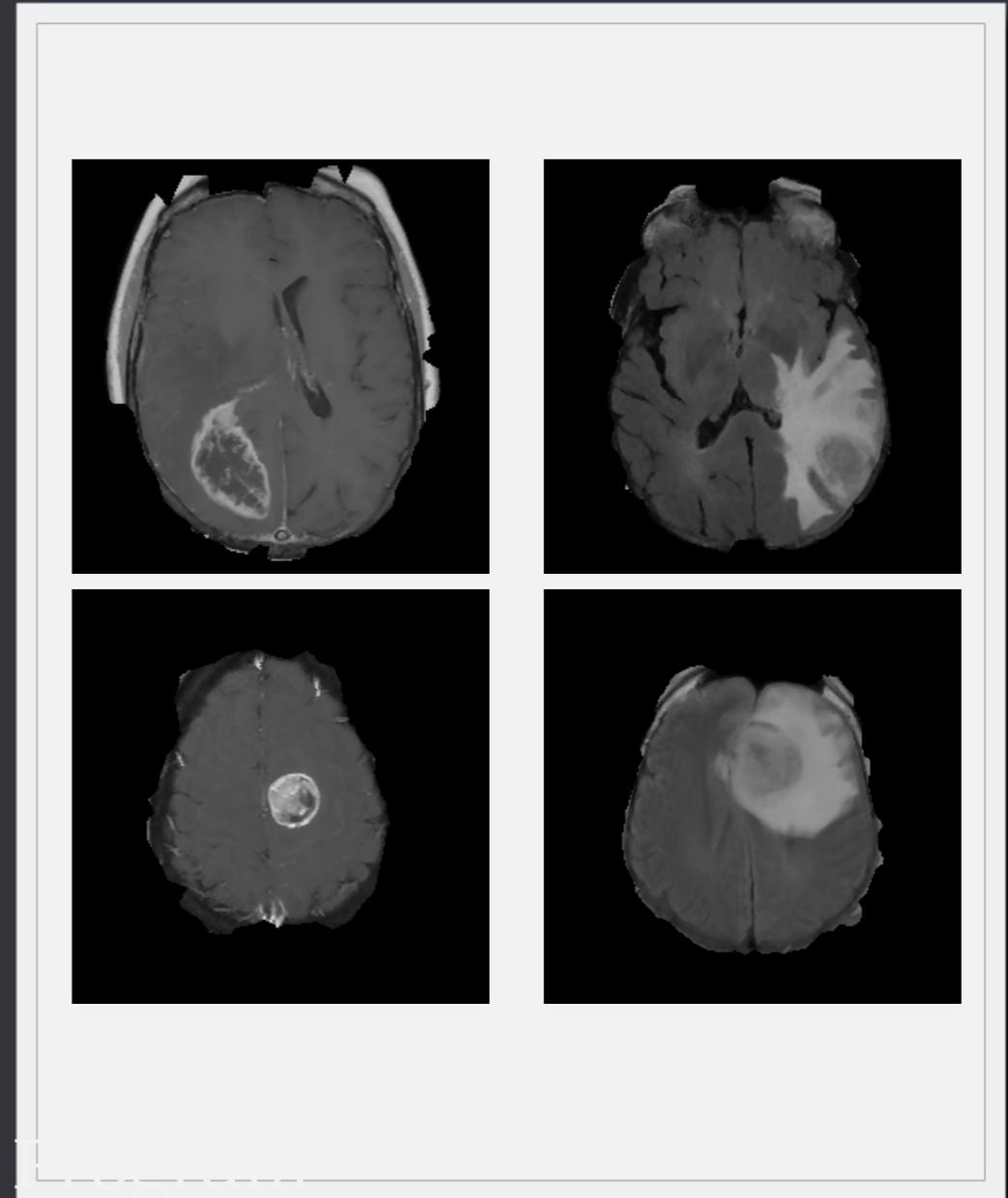
Need to define a new multiscale generative model incorporating appearance and shape properties.

New learning algorithms for the models and agglomeration.



Application to Brain Tumor Segmentation

- Dataset of 20 glioblastoma multiforme patients.
- 3D, 256x256x25.
- Images Pre-processed:
 - noise removal
 - skull removal
 - spatial registration
 - intensity standardization.
- Use 4 modalities
 - T1, T1 w/contrast, Flair, and T2.
- Expert annotated.
- Data graciously provided by Dr. Cloughesy of UCLA Henry E. Singleton Brain Cancer Research and preprocessed by Shishir Dube using FSL tools.



Model-Aware Affinity and Class Prior

- For model-aware affinity, use class-dependent weighted distance:

$$P(X_{uv} | s_u, s_v, c_u, c_v) = \exp \left(- \sum_{m=1}^M \theta_{c_u c_v}^m |s_u^m - s_v^m| \right).$$

- Coefficients $\theta_{c_u c_v}^m$ are set based on expert, domain knowledge.

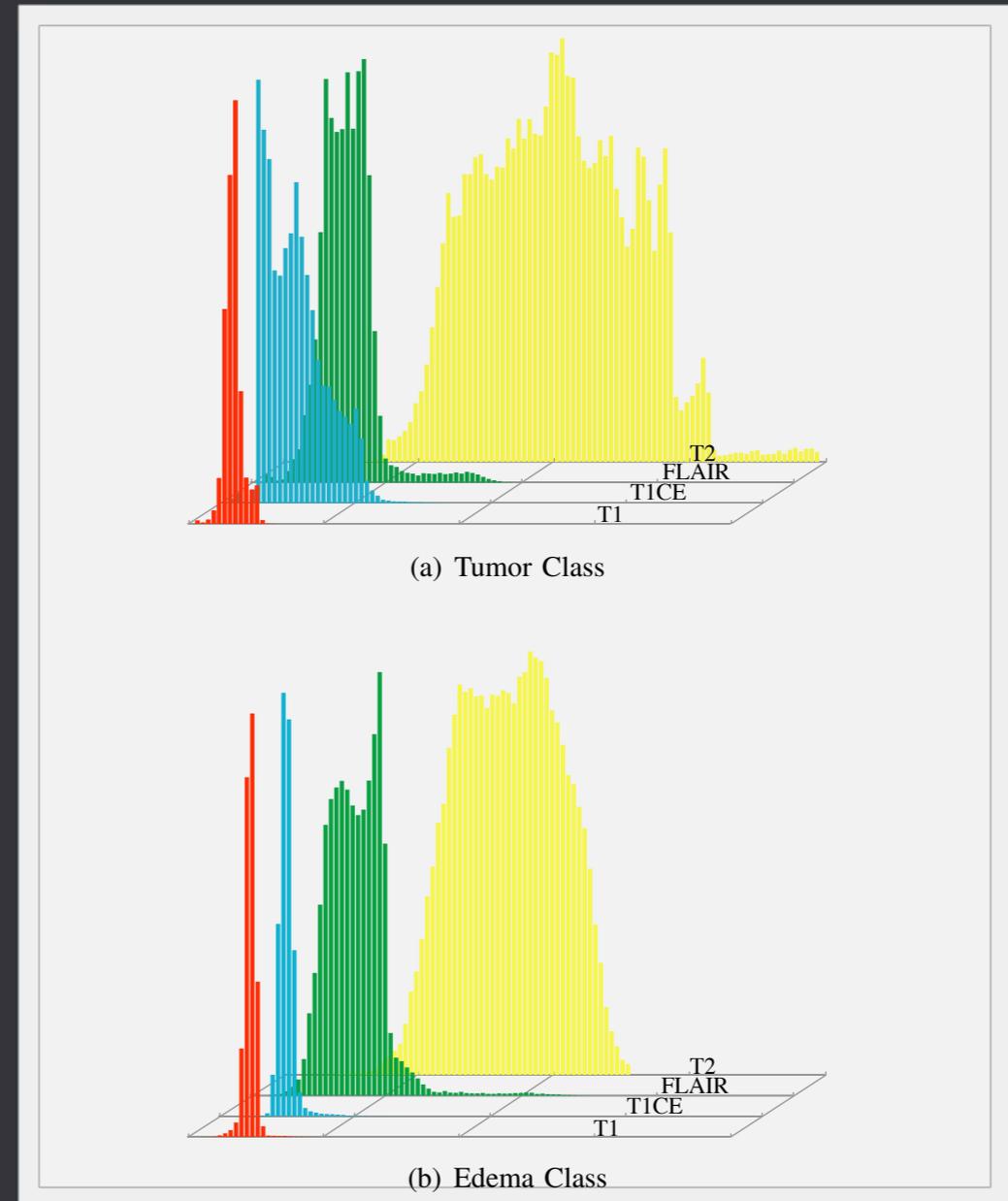
$c_u, c_v \backslash m$	T1	T1CE	FLAIR	T2
ND, Brain	1/4	1/4	1/4	1/4
Brain, Brain	1/4	1/4	1/4	1/4
Brain, Tumor	0	1	0	0
Tumor, Tumor	0	1	0	0
Brain, Edema	0	0	1/2	1/2
Tumor, Edema	0	1/2	1/4	1/4
Edema, Edema	0	0	1/2	1/2

- Feature statistic is simply average intensity.
- The class prior term
 - encodes obvious hard constraints (e.g. tumor cannot be next to non-data),
 - remaining set to uniform.

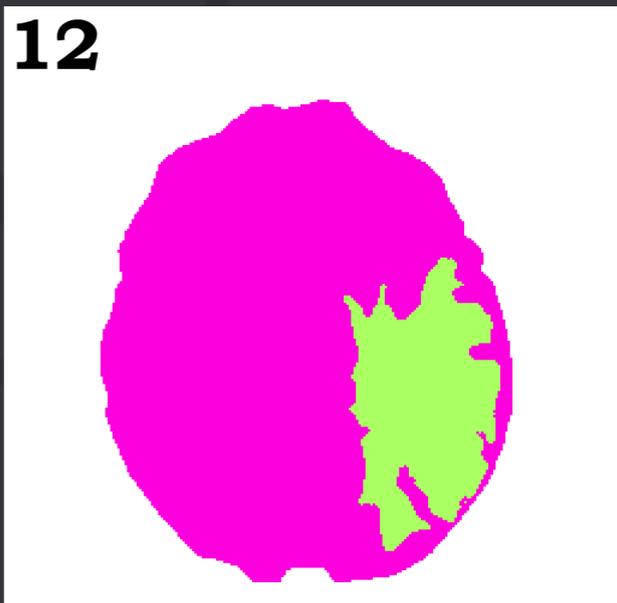
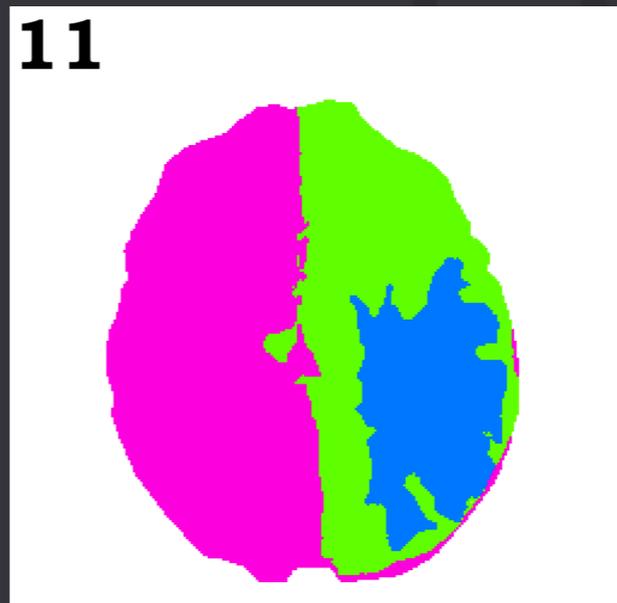
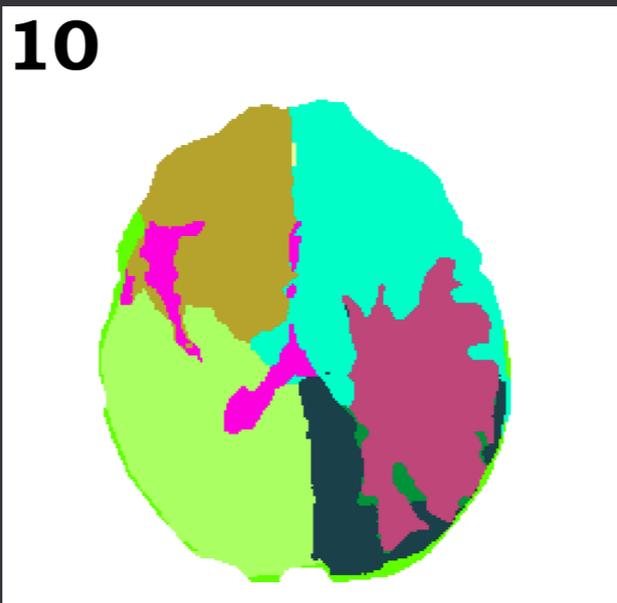
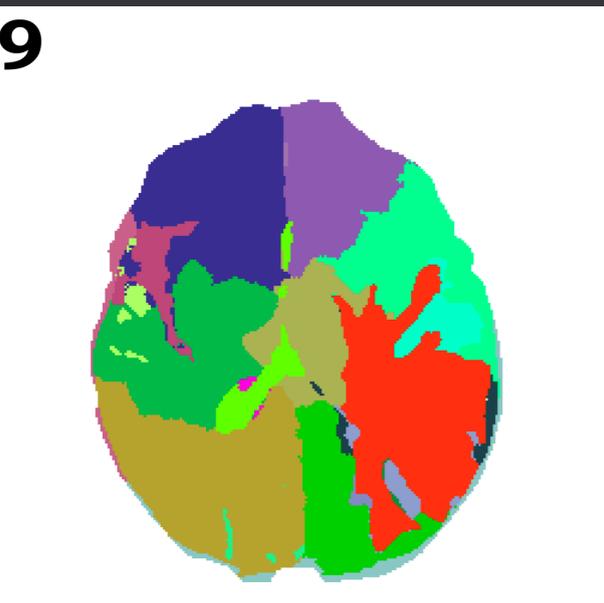
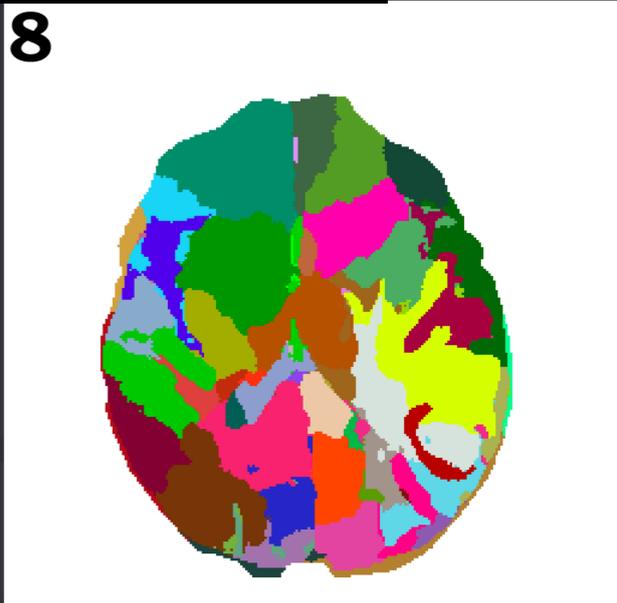
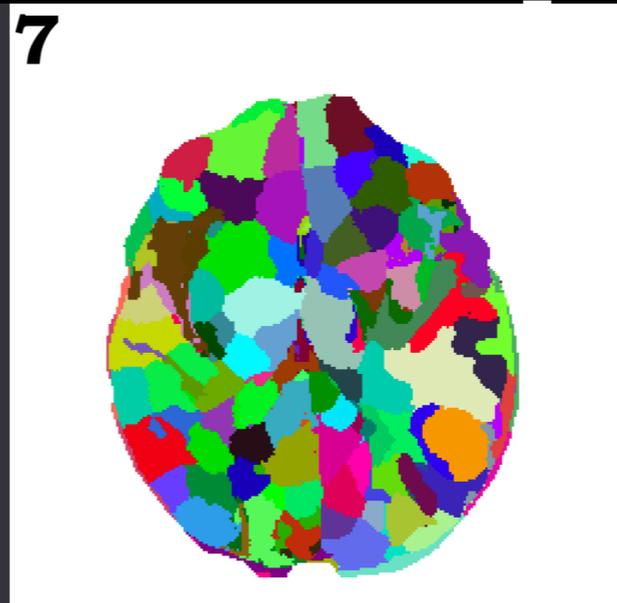
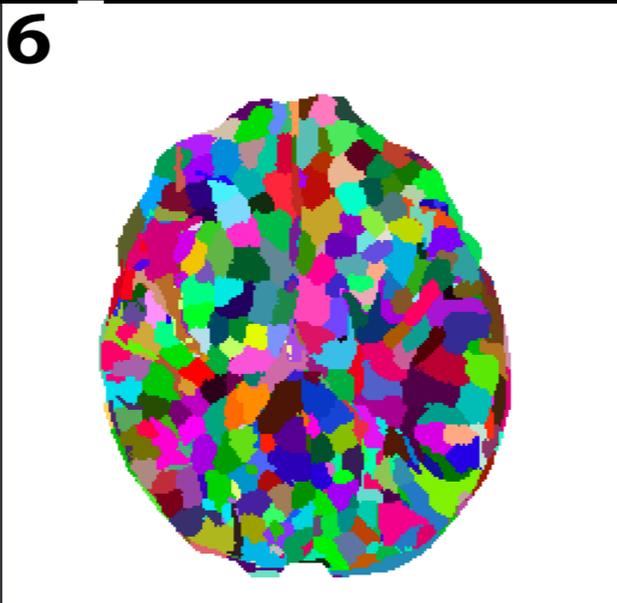
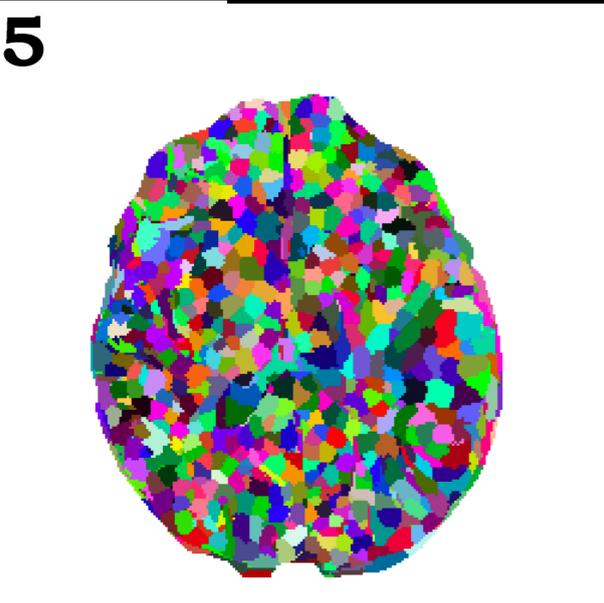
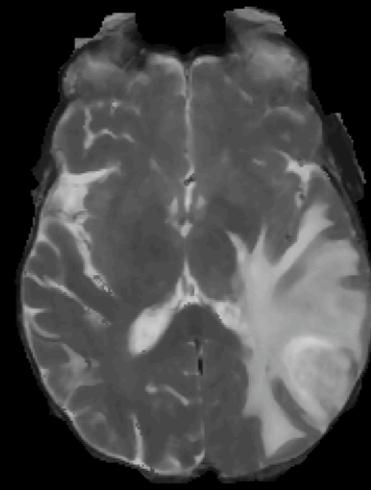
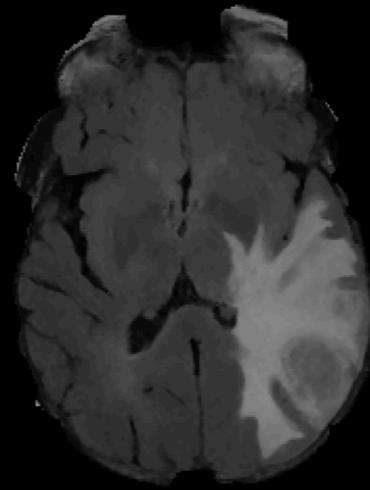
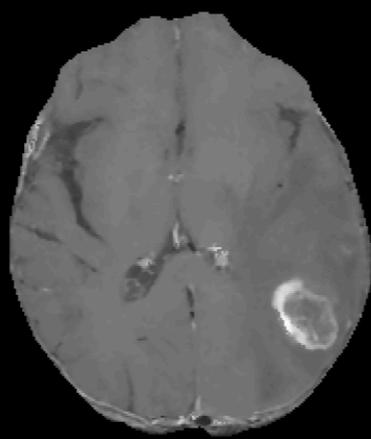
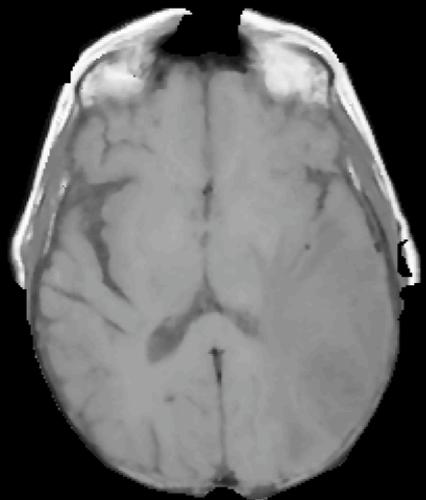


Class Likelihood Model

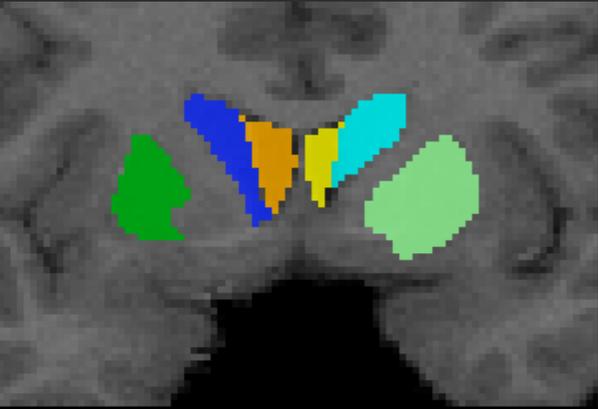
- Each class is modeled with a Gaussian mixture model.
- Likelihood is computed directly against this model.
- For some structures more node statistics must be used:
 - Shape moments
 - Surface curvature
 - Location
 - Relational



Example of the Segmentation Pyramid



Example of the Segmentation Pyramid



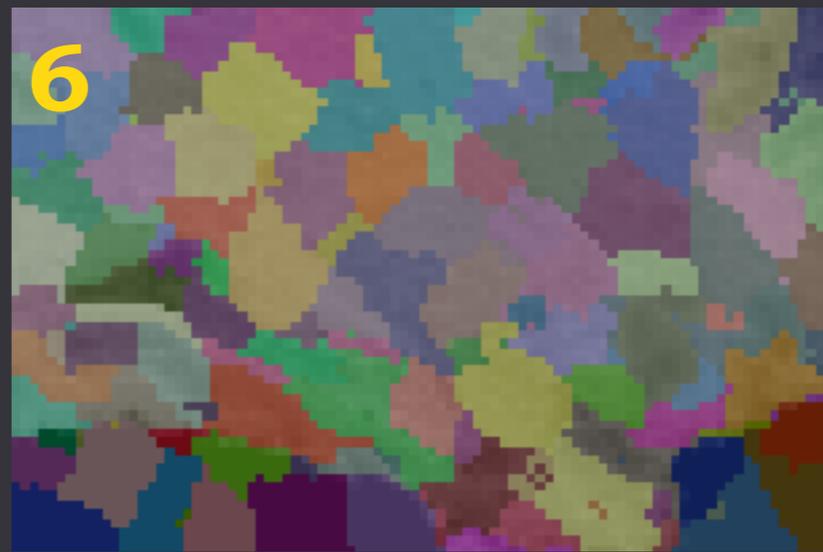
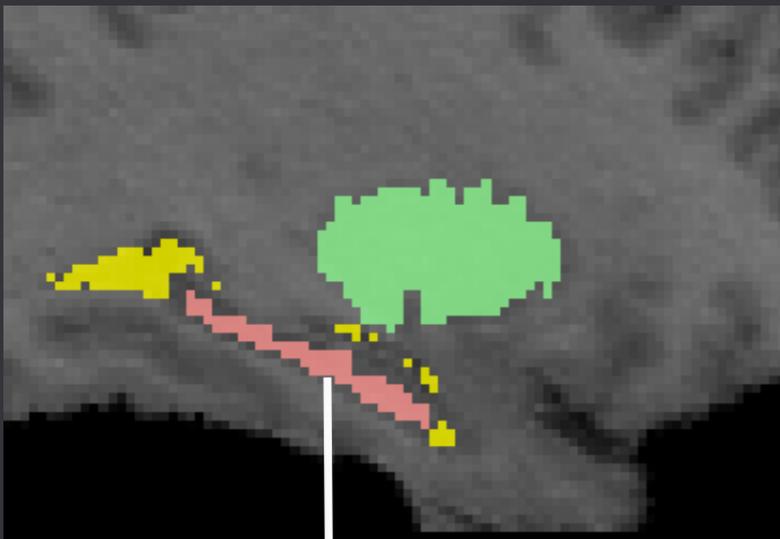
Caudate
Ventricle
Putamen

Can see the relevant structures emerging.

Preliminary results from sub-cortical project (began Sept. 2006).

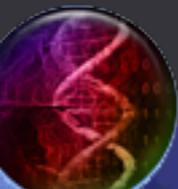


Example of the Segmentation Pyramid



Hippocampus

Hard to segment structures emerging too.



Quantitative Results

Volume Overlap Comparison

Results on Training Set						
Algorithm	Tumor			Edema		
	Jac	Prec	Rec	Jac	Prec	Rec
Single Voxel Classifier	42%	48%	85%	43%	49%	78%
Saliency-Based Extractor	44%	51%	64%	47%	55%	76%
Model-Based Extractor	62%	75%	81%	54%	66%	72%

Results on Testing Set						
Algorithm	Tumor			Edema		
	Jac	Prec	Rec	Jac	Prec	Rec
Single Voxel Classifier	49%	55%	81%	56%	66%	76%
Saliency-Based Extractor	48%	61%	63%	56%	66%	71%
Model-Based Extractor	66%	80%	79%	61%	78%	71%



Quantitative Results

Volume Measurement

Study ID	Tumor		Edema	
	Truth	Auto	Truth	Auto
TR01	93335	79894	138972	151524
TR02	17344	12516	20709	38145
TR03	71211	30769	122168	188732
TR04	798	1552	10143	11730
TR05	25411	34776	122829	118247
TR06	7643	13826	10768	16470
TR07	31829	34699	34733	38670
TR08	24594	26570	134487	162264
TR09	64486	79718	253682	268981
TR10	220243	227251	90120	36970
TE01	24851	38817	262421	213046
TE02	54352	51054	171869	153000
TE03	46651	41242	65919	70784
TE04	10499	8177	81740	74495
TE05	46783	57454	101198	84437
TE06	41395	39328	37355	29709
TE07	36826	24079	72625	36251
TE08	52256	39348	140799	146390
TE09	16503	19989	220270	241509
TE10	95422	92311	183315	186080

mm³

Quantitative Results

Surface Measurement

Study ID	Tumor		Edema	
	Mean	Median	Mean	Median
TR01	1.23	0	1.01	0
TR02	7.73	0	76.40	97.68
TR03	3.71	1.8	3.03	0
TR04	5.11	0.86	1.95	0
TR05	1.56	0	0.56	0
TR06	7.63	1.22	13.60	6.5
TR07	10.28	0	7.02	0
TR08	1.17	0	1.12	0
TR09	4.11	0	0.71	0
TR10	0.68	0	5.57	1.72
TE01	29.97	0.78	2.52	0
TE02	1.56	0	2.49	0
TE03	1.12	0	0.54	0
TE04	0.39	0	1.24	0
TE05	1.02	0	0.77	0
TE06	0.74	0	19.69	26.78
TE07	20.23	2.32	10.25	1.83
TE08	2.23	0	1.63	0
TE09	2.14	0	2.04	0
TE10	4.54	0	3.36	0

mm

Classification Comparison

1

2

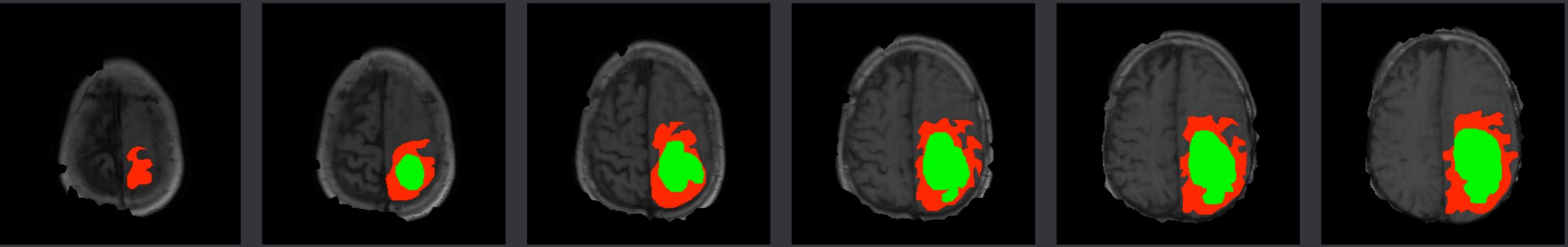
3

4

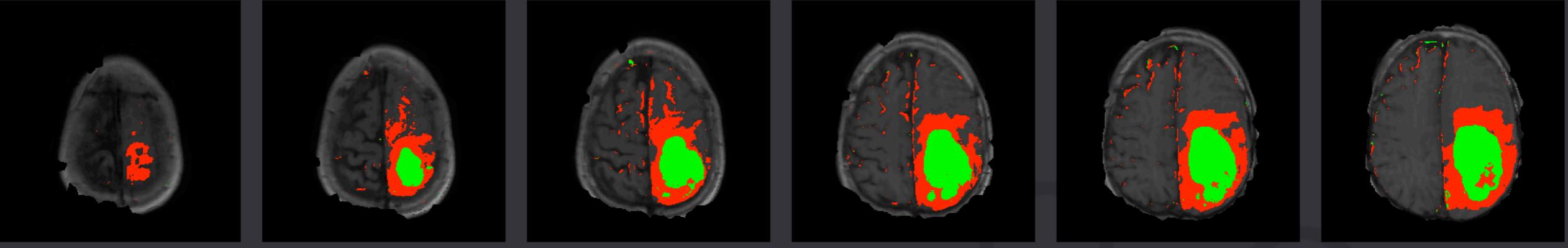
5

6

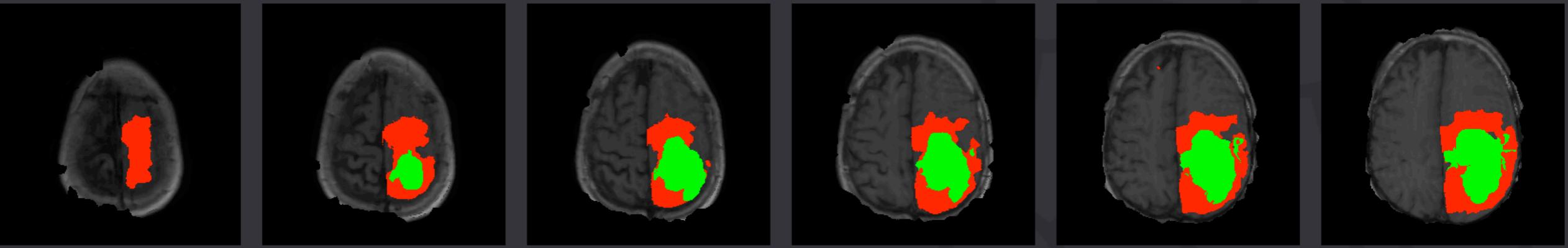
Manual



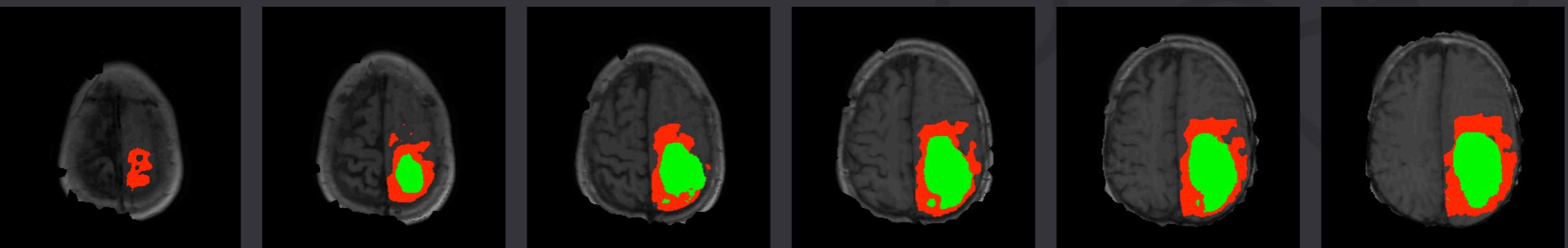
Single Voxel



Saliency



Models



Classification Comparison

7

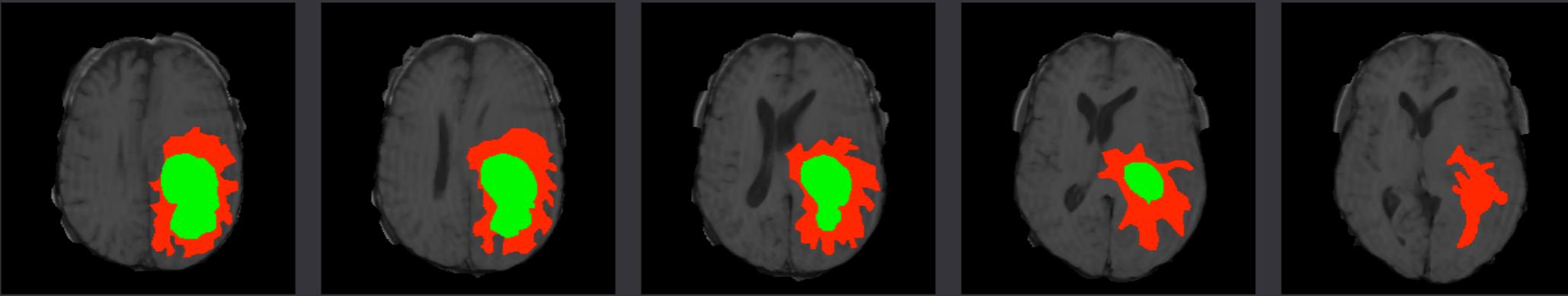
8

9

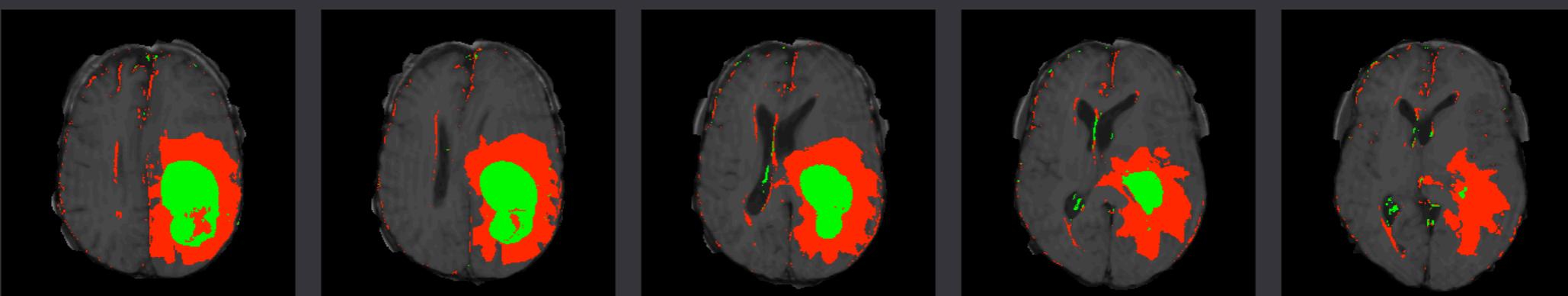
10

11

Manual



Single Voxel



Saliency



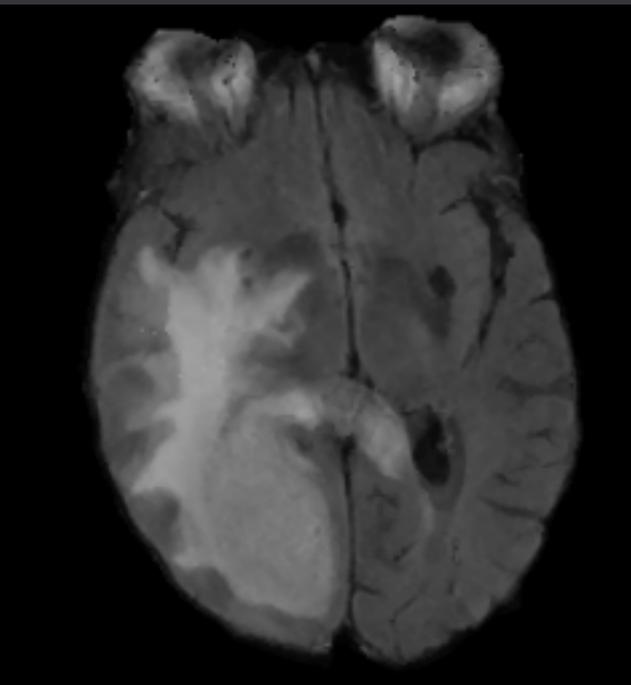
Models



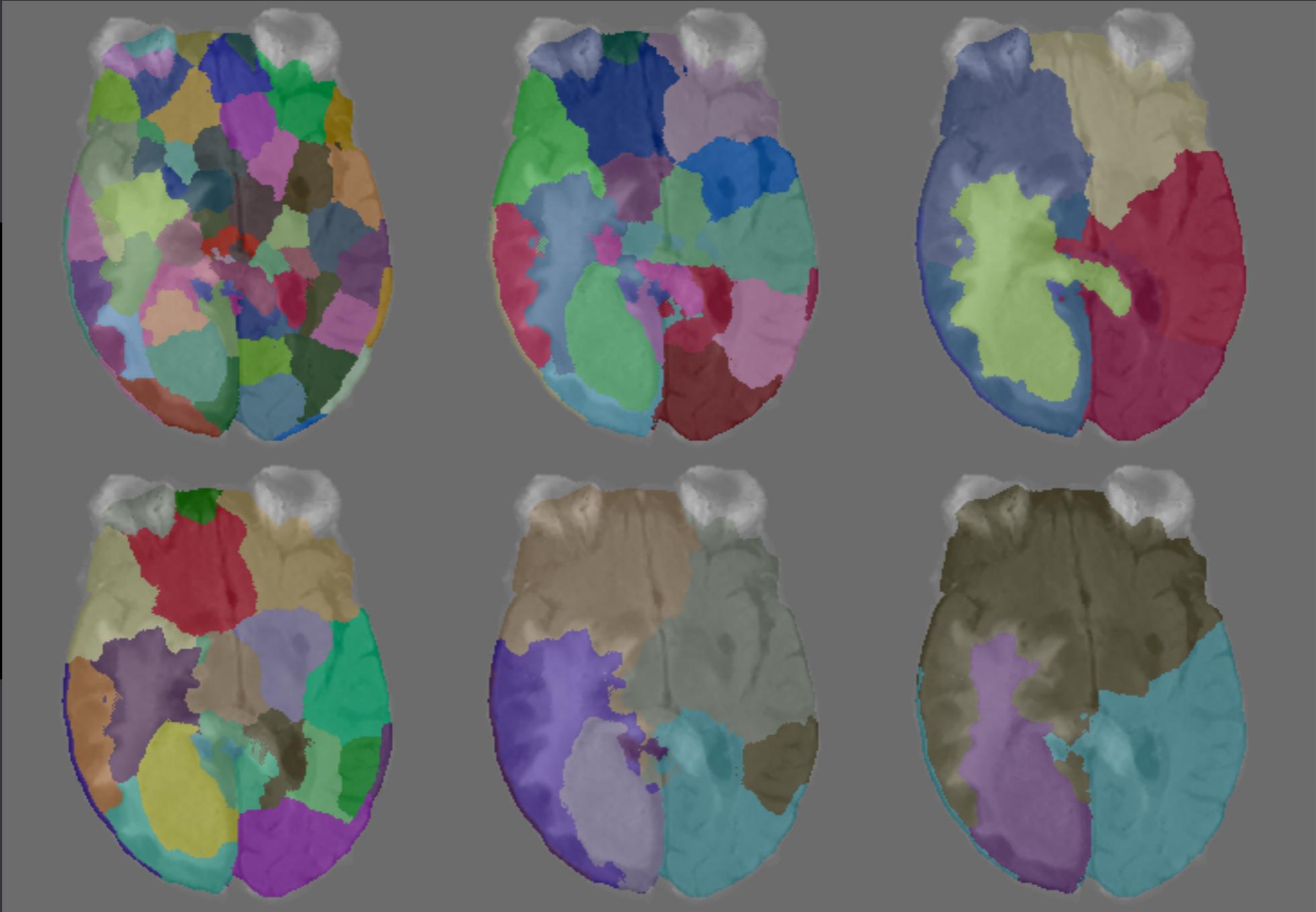
Results

Comparing SWA with Model-Aware Affinities to Original SWA

Model-Aware



Original



Implementation Details

- Built with Java and Swing.
- Uses ImageIO: supports many image formats using the LONI medical image plugins.
- Developed **new core data structures to support transparent out-of-core data manipulation**, which is often necessary when working with medical imaging (especially hierarchical).
- Developed with principled OO techniques:
 - **Polymorphic Model-Class design for “plug-and-play” affinity function and likelihood measurements based on specific application.**
 - Design patterns to make it easy to generate an application specific segmentation algorithm built with atop same core software.
- Computational time: 90x60x90 volume segmented in about 2 minutes on a 2Ghz PowerPC G5.
- Extraction time depends on method.



Results - Tool for Interactive Analysis Using Models

Select Slices to View (3 axes)

Select Modality

Select Hierarchy Level

Select Transparency

Overlay Manual Contours

JavaSeg Gui

File

Segmentation Visualization

XY (17/21) YZ (139/255) XZ (126/255)

Axial

Sagittal

Coronal

3D Rendering

T1CE

Pyramid Level

4 8 12

Overlay Blend

0 64 128 192

Hilite 5 Most Salient

Overlay Ground Truth Contours

Run Tumor Analysis (ClassAware)

Grab XY Slice

Stats & Info

Node ID: 6

Saliency: 5.6916702E-12

Conclusions

- Main contribution:
 - Classes as hidden variables for bridging graph-based affinities and model-based techniques.
- Incorporated into the SWA algorithm.
- Computational time: 90x60x90 volume segmented & classified in about 2 minutes on a PowerPC G5 (in Java).
- Software contribution:
 - A complete system for experimenting with the Bayesian computational methods, including learning, segmentation, visualization, and analysis.
- Future:
 - Better models with more feature statistics,
 - Learning the model-aware affinity class dependent parameters,
 - Leverage the efficiency of graph-based bottom-up methods with the comprehensive and optimality of top-down generative models.
- Thanks for listening.



Acknowledgements

- Funding by NLM Medical Informatics Training Grant #LM07356 and NIH NCBC Center for Computational Biology (LONI).
- Dr. Cloughesy of the Henry E. Singleton Brain Cancer Research Program, University of California, Los Angeles, for providing the data.
- Shishir Dube for helping with and performing the data preprocessing.
- Drs. Yuille, Sharon, Taira and Sinha for guidance.



Research Overview

- Concepts
 - Automatic Knowledge discovery
 - Domain expert knowledge
 - Models
 - Bayesian methodology (generative modeling)
 - Manual creation
 - Model discovery (e.g. learning grammars, manifold learning)
 - Inference
 - Estimate the models (find instances of the concepts) in data
 - Efficient and robust
 - Integrate bottom-up, feed-forward with top-down models
 - Applications
 - Medical Informatics
 - Surgical planning
 - Various related biomedical application (e.g. tissue microarrays)
- Statistical Learning
- 