Spatially Coherent Latent Topic Model for Concurrent Segmentation and Classification of Objects and Scenes

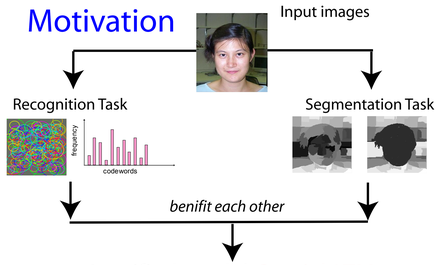
University of Illinois Urbana-Champaign, USA[1]

Liangliang Cao[1] cao4@ifp.uiuc.edu

Li Fei-Fei[2] feifeili@cs.princeton.edu

Princeton University, USA[2]

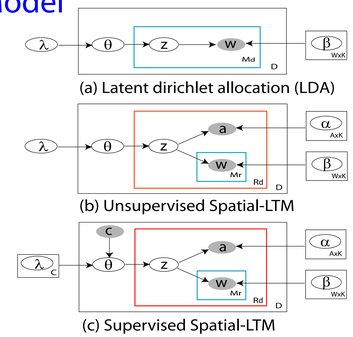
Motivation



A unified model: Spatial-LTM

key idea: assign **the same hidden topic** for the visual words within a neighboring region with similar appearance.

Model

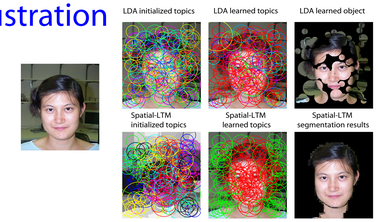


(a) Latent dirichlet allocation (LDA)

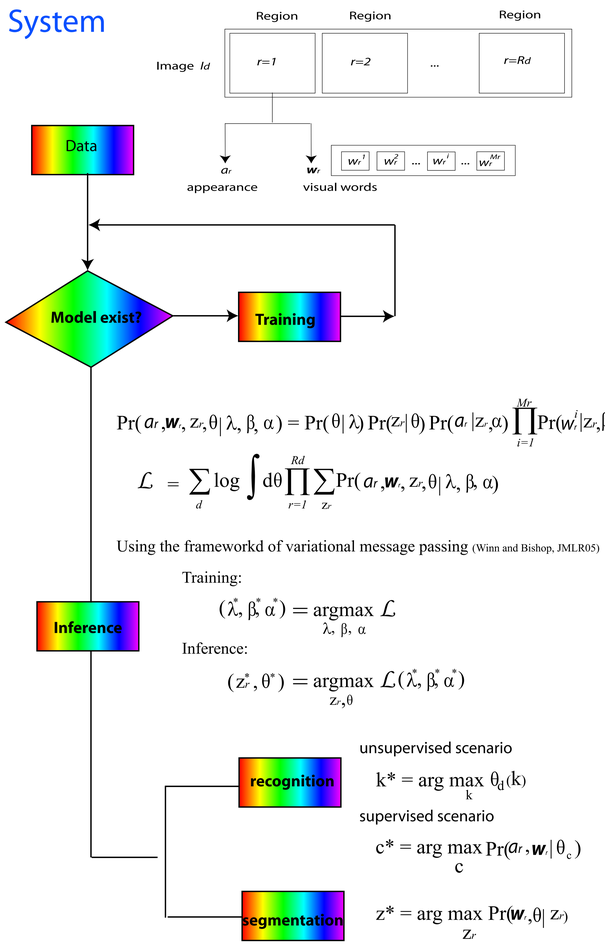
(b) Unsupervised Spatial-LTM

(c) Supervised Spatial-LTM

Illustration



System



$$\Pr(a_r, \mathbf{w}_r, z_r, \theta \mid \lambda, \beta, \alpha) = \Pr(\theta \mid \lambda)\Pr(z_r \mid \theta)\Pr(a_r \mid z_r, \alpha)\prod_{i=1}^{Mr}\Pr(w_r^i \mid z_r, \beta)$$

$$\mathcal{L} = \sum_d \log \int d\theta \prod_{r=1}^{Rd}\sum_{z_r}\Pr(a_r, \mathbf{w}_r, z_r, \theta \mid \lambda, \beta, \alpha)$$

Using the frameworkd of variational message passing (Winn and Bishop, JMLR05)

Training:

$$(\lambda^*, \beta^*, \alpha^*) = \underset{\lambda,\beta,\alpha}{\operatorname{argmax}}\ \mathcal{L}$$

Inference:

$$(z_r^*, \theta^*) = \underset{z_r,\theta}{\operatorname{argmax}}\ \mathcal{L}(\lambda^*, \beta^*, \alpha^*)$$

unsupervised scenario

$$k^* = \arg\max_k \theta_d(k)$$

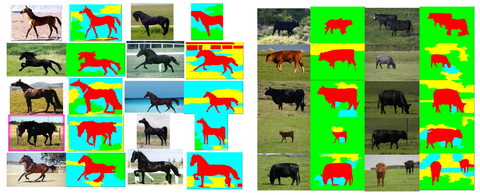
supervised scenario

$$c^* = \arg\max_c \Pr(a_r, \mathbf{w}_r \mid \theta_c)$$

$$z^* = \arg\max_{z_r} \Pr(\mathbf{w}_r, \theta \mid z_r)$$

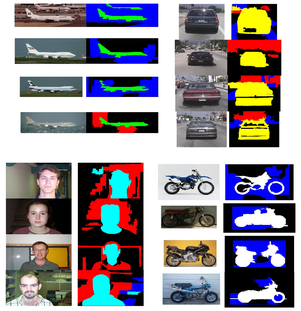
Experimental Results

Unsupervised recognition of animals (81.8% on Weizmann horse data set)

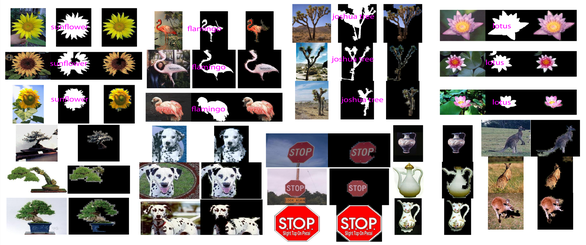


Handling occlusion and inversed directions

Unsupervised recognizing Caltecech 4 classes objects. (94.6%)

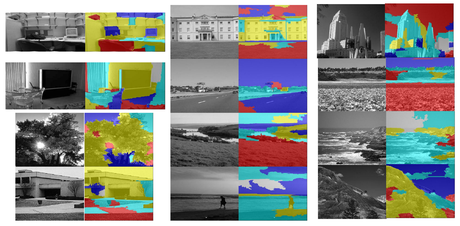


Supervised recognition of 28 objects classes from Caltech 101 (Overall precision: 69.4%)



	Faces easy	Leopards	Motorbikes	bonsai	brain	butterfly	cougar face	crab	cup	dalmatian	dolphin	elephant	euphonium	ewer	ferry	flamingo	grand piano	joshua tree	kangaroo	laptop	lotus	schooner	soccer ball	starfish	stop sign	sunflower	watch	yin yang
Faces easy	96																											
Leopards		73	18							1			2	1	2		1		1	1				1				
Moto/bikes			80	4			1			1				1				3										3
bonsai		2	2	72	1		1											1	5	1	2	1			1	2	1	1	5
brain					82	1			3	1					1	1	1											
butterfly	3	1	1	3	1	34	2	1			1		1		1			1	1	3		4		6			4	1	10
cougar face		2	2				82					2								2								2	
crab		6	3	4	3	2	3	44	5	2								2	4	4	2		4		6	2		4	2
cup		3	3	3			3		70	3	3			3					5										
dalmatian			2	2	3					91																			2
dolphin	2			2	2						84																		
elephant				2		2				2	56	4		2					5				2			5	2	2
euphonium			7	5	2								86																
ewer	3	2	8	6	2		5	2	3	2	6			26		2	2	2	2	2	3		3					6	15
ferry				3							2				2	91													
flamingo		2	9	11												2	49	2	2	2						2		2	9
grand piano																	191												
joshua tree		2	2	5	2					2		2						277				5							
kangaroo		3	2	9	6		2			2	3	2		2					361	2					2		2	2	2
laptop			2	2	3				2	2	2						13			62									10
lotus			2	2						4											80		2				4		
schooner			2	4																		41							5
soccer ball			2	3																	2		70			7	7	7	
starfish		2	5	3	5	2	6	5	2		2		2		2		3		3	3	2	3		45				2	9
stop sign											2		2					2							91				
sunflower				2														2	2	2				2		91		2	
watch			5		7		1	1	1	1	1		1					5		1				1			258	8	
yin yang				5													3								3		3		88

Unsupervised recognition of 13 scene classes (Overall precision: 66.4%)



	coast	forest	highway	mountain	opencountry	bedroom	kitchen	livingroom	PARoffice	CALsuburb	insidecity	street	tallbuilding
coast	85	2	2	1	10								
forest		98		2									
highway	16	2	66	6	6				1	2	1		
mountain	2	23	8	61	2				1		2	1	
opencountry	12	25	7	3	50					3			
bedroom	1	2	1			2	34	23	4	18	2	2	6	5
kitchen						2	49	3	29		14	1	2
livingroom		1		1		6	19	37	17	2	8	5	4
PARoffice							11		86		2		1
CALsuburb		4		1	3		1	1	5	83	1		1
insidecity	4	2			1		6	2	4	3	70	3	5
street	1	1	5	2	1		1			3	10	72	4
tallbuilding	2	6		6		1	3	1			6	5	70

Reference:

L. Cao and L. Fei-Fei. Spatially Coherent Latent Topic Model for Concurrent Segmentation and Classification of Objects and Scenes. IEEE International Conference on Computer Vision (ICCV), Rio de Janeiro, Brazil, 2007.