Factored Shapes and Appearances for Parts-based Object Understanding

S. M. Ali Eslami s.m.eslami@sms.ed.ac.uk

Christopher K. I. Williams ckiw@inf.ed.ac.uk

THE UNIVERSITY *of* EDINBURGH informatics

ianc Institute for Adaptive and Neural Computation

1. Summary

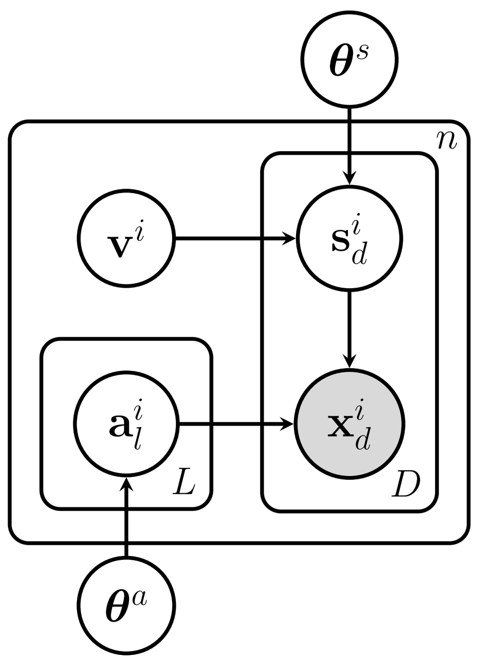
Modern computer vision algorithms rely on prior knowledge about images for performance. In this project we wish to learn **parts-based** models of object classes from **highly variable** data, in order to perform

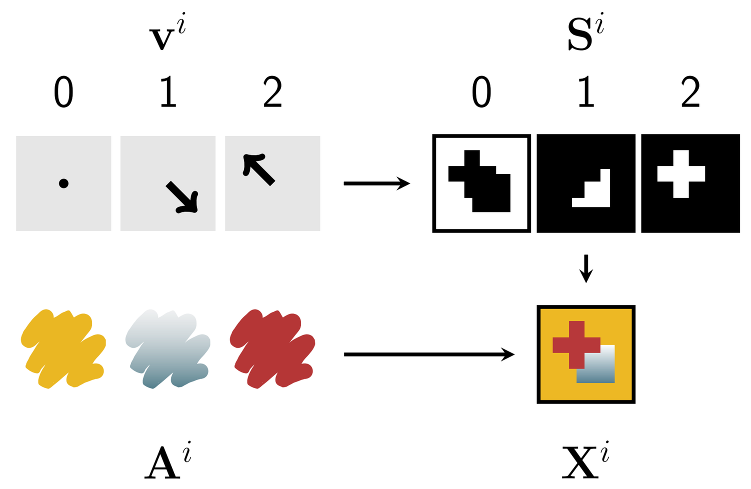
- object parsing,
- foreground/background segmentation, and
- fine-grained classification

on unseen images. Our probabilistic model employs a highly factored representation to learn and reason about both **appearance and shape variability** across datasets of images.

2. The FSA model

Given a dataset $\mathbf{D} = \{\mathbf{X}^i\}, i = 1...n$ of images $\mathbf{X}$, each consisting of D pixels $\{\mathbf{x}_d\}$ in some feature space, we wish to obtain an accurate understanding of the parts' extents by inferring a segmentation $\mathbf{S}$ for each image. Segmentations consist of labellings $\mathbf{s}_d$ for every pixel, where L is the fixed number of parts that combine to generate the foreground and $\mathbf{s}_d$ is a 1-of-$(L+1)$ encoded variable.



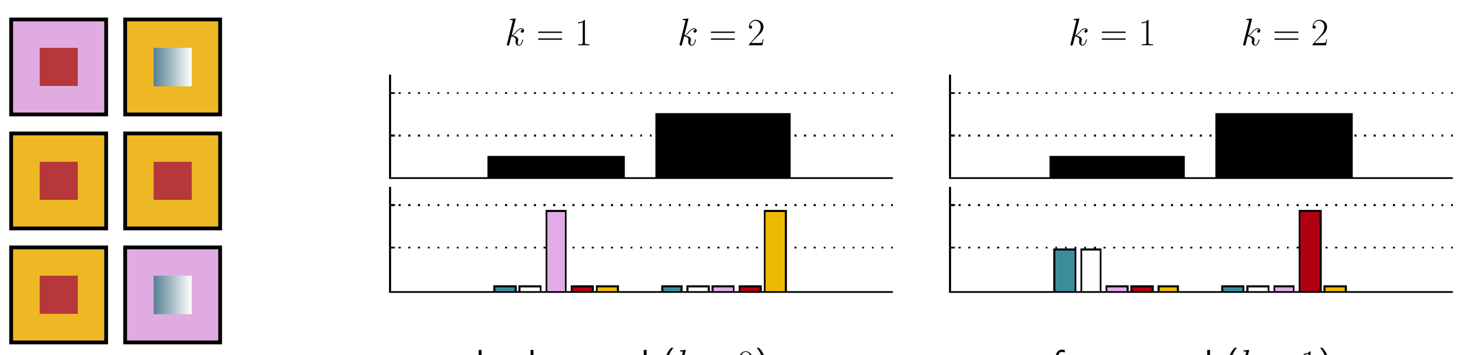


$$p(\mathbf{X}^i, \mathbf{A}^i, \mathbf{S}^i, \mathbf{v}^i | \boldsymbol{\theta}) = p(\mathbf{v}^i)\, p(\mathbf{A}^i | \boldsymbol{\theta}^a) \prod_{d=1}^{D} p(\mathbf{s}_d | \mathbf{v}^i, \boldsymbol{\theta}^s)\, p(\mathbf{x}_d^i | \mathbf{A}, \mathbf{s}_d^i, \boldsymbol{\theta}^a)$$

Shape

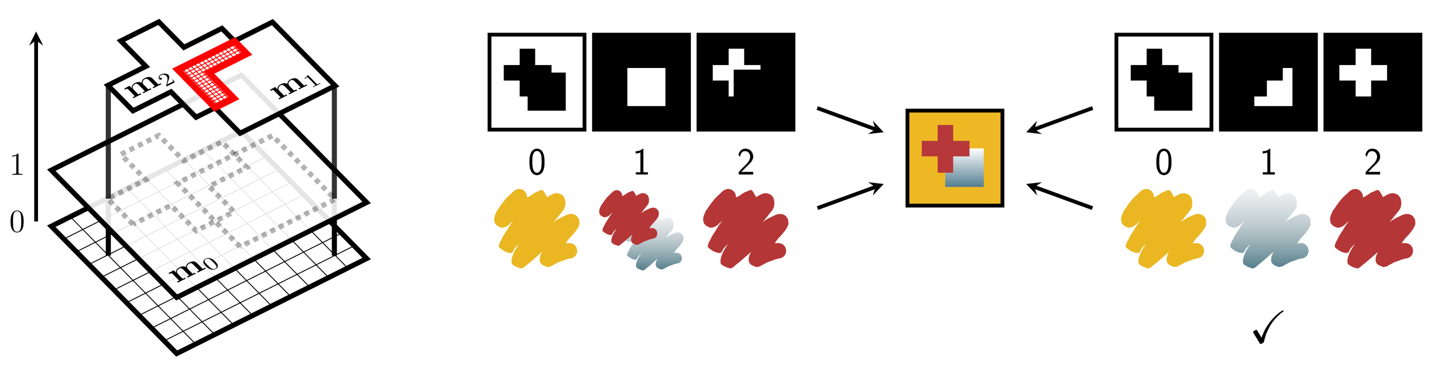
$$\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{H\times H}) \qquad \mathbf{m}_l = \mathbf{F}_l \mathbf{v} + \mathbf{c}_l \qquad p(s_{ld} = 1 | \boldsymbol{\theta}) = \frac{\exp\{\mathbf{m}_{ld}\}}{\sum_{k=0}^{L} \exp\{\mathbf{m}_{kd}\}}$$

Appearance

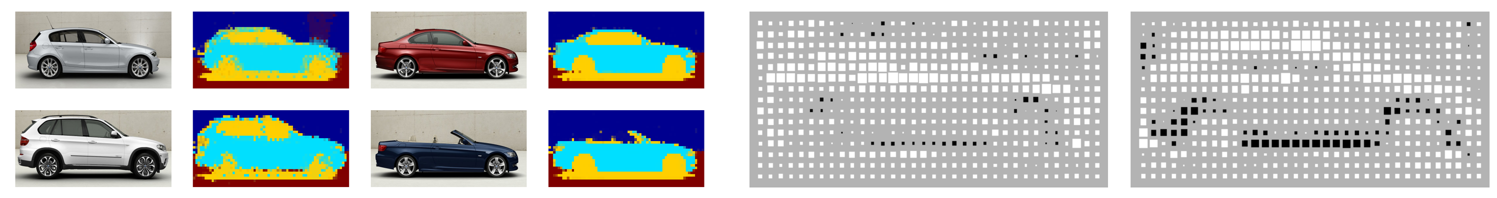


background ($l=0$) foreground ($l=1$)

Occlusion

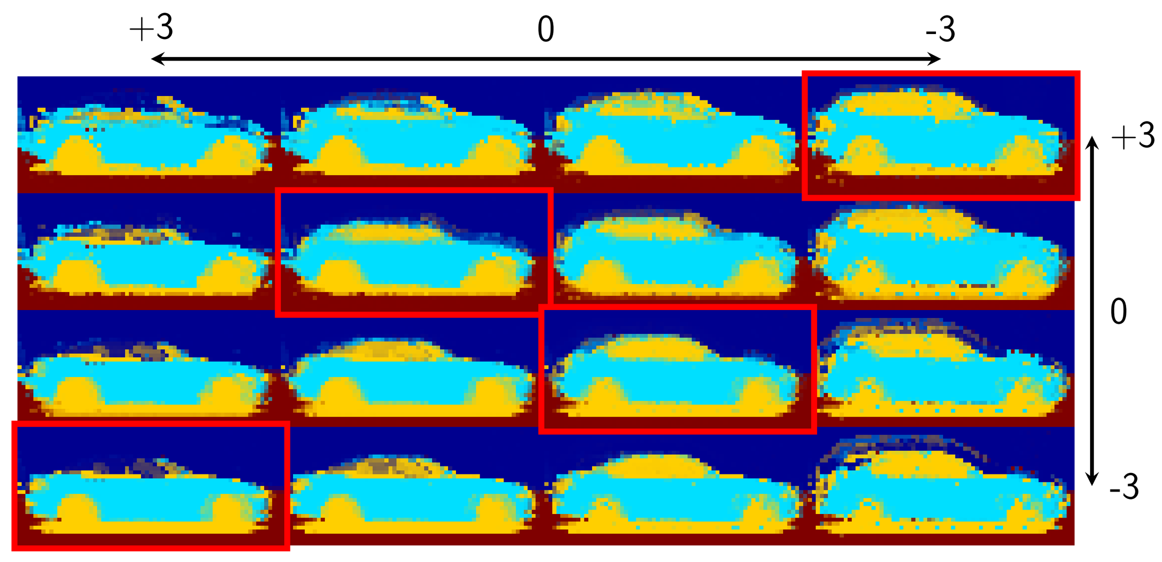


3. Results

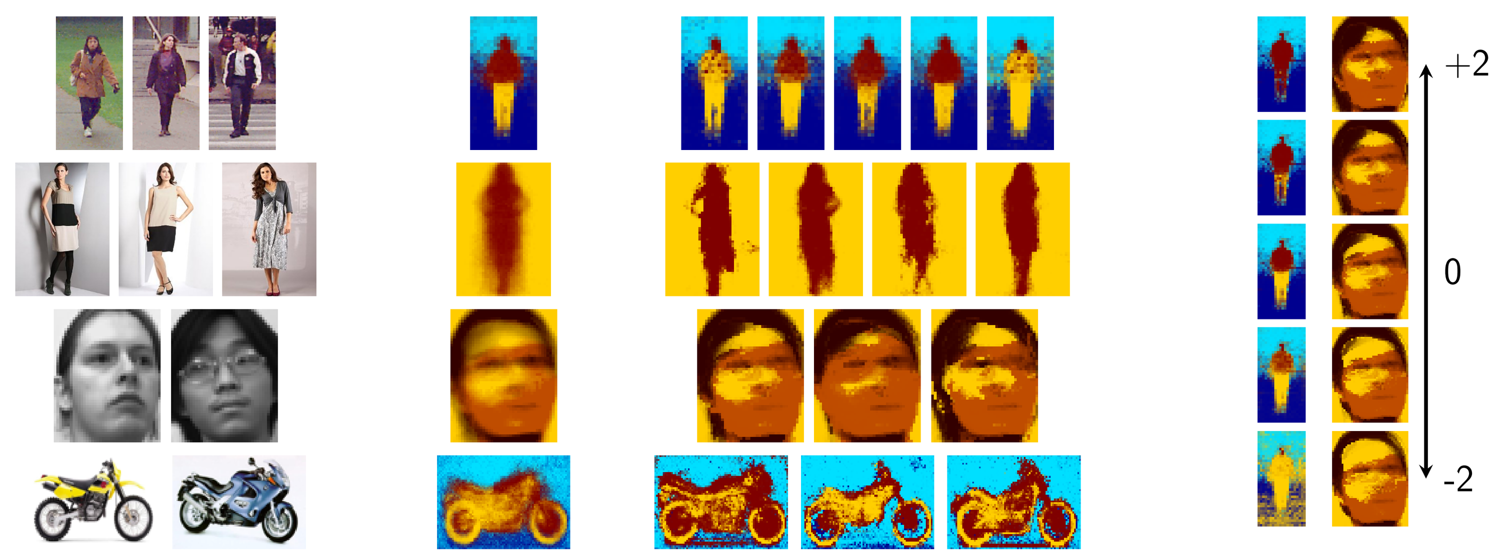


Segmentations Columns of $\mathbf{F}$ matrices

Using a leave-one-out SVM classifier on *only* the inferred $\mathbf{v}$s, we can classify the cars into the 5 distinct categories with **100% accuracy**:



Other datasets



Training images PIM Samples from FSA model Linear Variation

FSA uses shape information to increase segmentation accuracy, and its performance is comparable to that of state-of-the-art methods:

	Horses	Cars	Faces	Motorbikes	Airplanes
GrabCut [1]	83.9%	45.1%	83.7%	82.4%	84.5%
Borenstein [3]	93.6%	-	-	-	-
LOCUS [8]	93.1%	91.4%	-	-	-
Arora [2]	-	95.1%	92.4%	83.1%	93.1%
ClassCut [1]	86.2%	93.1%	89.0%	90.3%	89.8%
Unsupervised FSA	87.3%	82.9%	88.3%	85.7%	88.7%
Supervised FSA	88.0%	93.6%	93.3%	92.1%	90.9%

4. Related work

	FACTORED PARTS	FACTORED SHAPE & APP.	SHAPE VARIABILITY	APPEARANCE VARIABILITY
LSM [4]	✓(layers)	-	✓(FA)	✓(FA)
Sprites [7]	✓(layers)	-	-	-
LOCUS [8]	-	✓	✓(deformation)	✓(colours)
MCVQ [6]	-	✓	-	✓(templates)
SCA [5]	-	✓	✓(convex)	✓(histograms)
FSA	✓(softmax)	✓	✓(FA)	✓(histograms)

[1] B. Alexe, T. Deselaers, and V. Ferrari. ClassCut for unsupervised class segmentation. In *ECCV*, 2010.

[2] H. Arora, N. Loeff, D. Forsyth, and N. Ahuja. Unsupervised Segmentation of Objects using Efficient Learning. *CVPR*, 2007.

[3] E. Borenstein, E. Sharon, and S. Ullman. Combining Top-Down and Bottom-Up Segmentation. In *CVPR*, 2004.

[4] B. Frey, N. Jojic, and A. Kannan. Learning appearance and transparency manifolds of occluded objects in layers. In *CVPR*, 2003.

[5] N. Jojic, A. Perina, M. Cristani, V. Murino, and B. Frey. Stel component analysis. In *CVPR*, 2009.

[6] D. Ross and R. Zemel. Learning Parts-Based Representations of Data. *JMLR*, 7, 2006.

[7] C. K. I. Williams and M. Titsias. Greedy learning of multiple objects in images using robust statistics and factorial learning. *NC*, 2004.

[8] J. Winn and N. Jojic. LOCUS: Learning object classes with unsupervised segmentation. In *ICCV*, 2005.