A Generative Model for Parts-based Object Segmentation

S. M. Ali Eslami [1]   Christopher K. I. Williams [2]

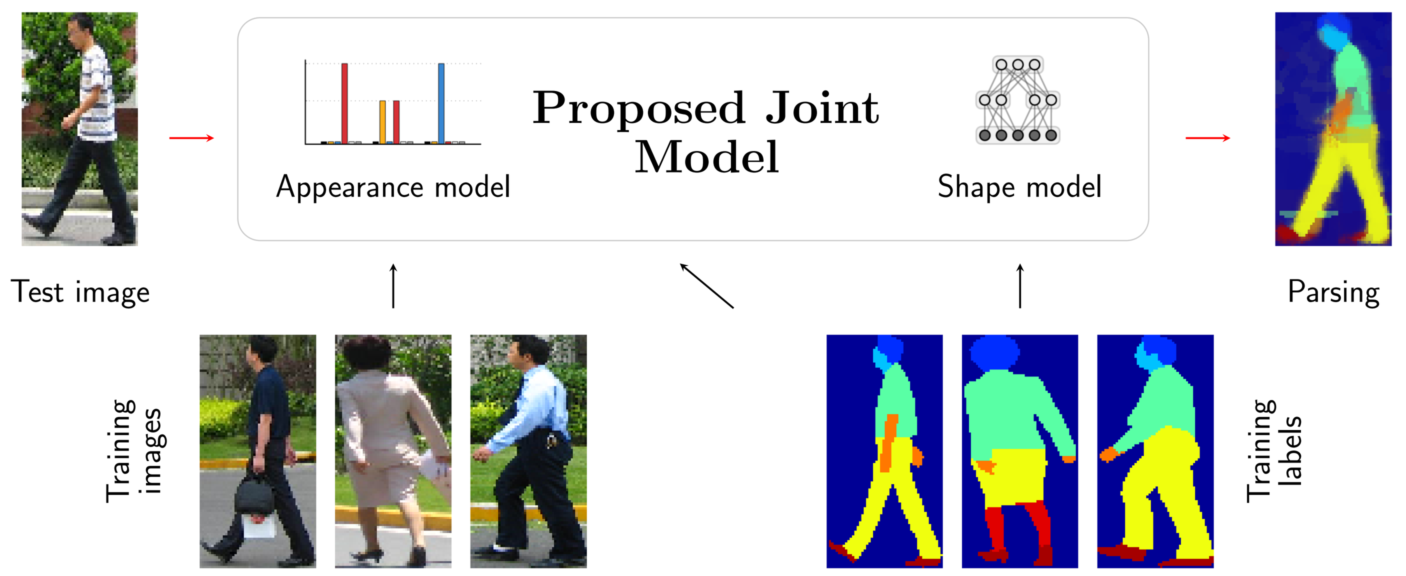
[1] s.m.eslami@sms.ed.ac.uk, [2] ckiw@inf.ed.ac.uk

THE UNIVERSITY of EDINBURGH informatics

ianc Institute for Adaptive and Neural Computation

1. Goal

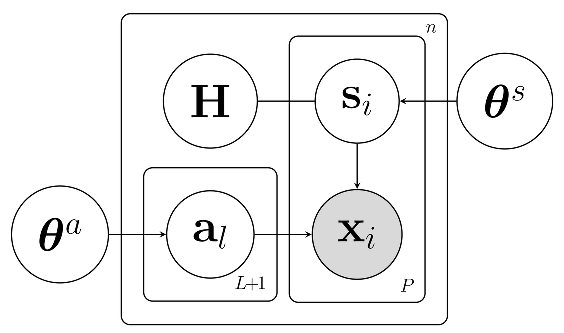
To obtain accurate **parts-based** segmentations of objects using **interpretable**, **generative**, Boltzmann-machine models.

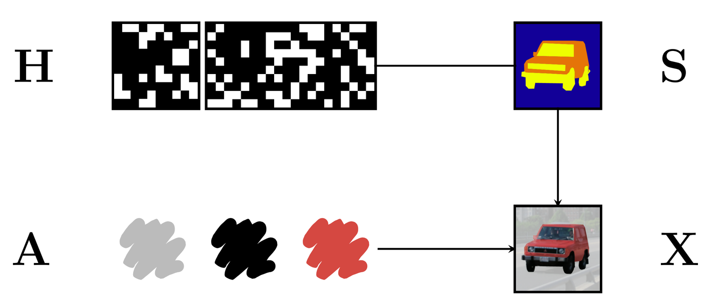


The parsing of an unseen image is simply obtained via **standard probabilistic inference** in the proposed joint model.

The model is **generic**, in that it makes few apriori assumptions about the shape and appearance properties of the modeled object.

2. The proposed joint model

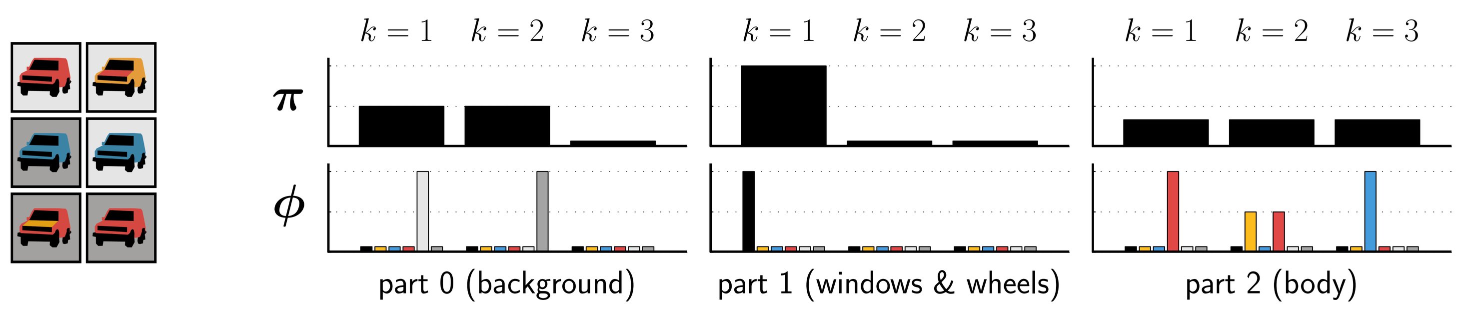




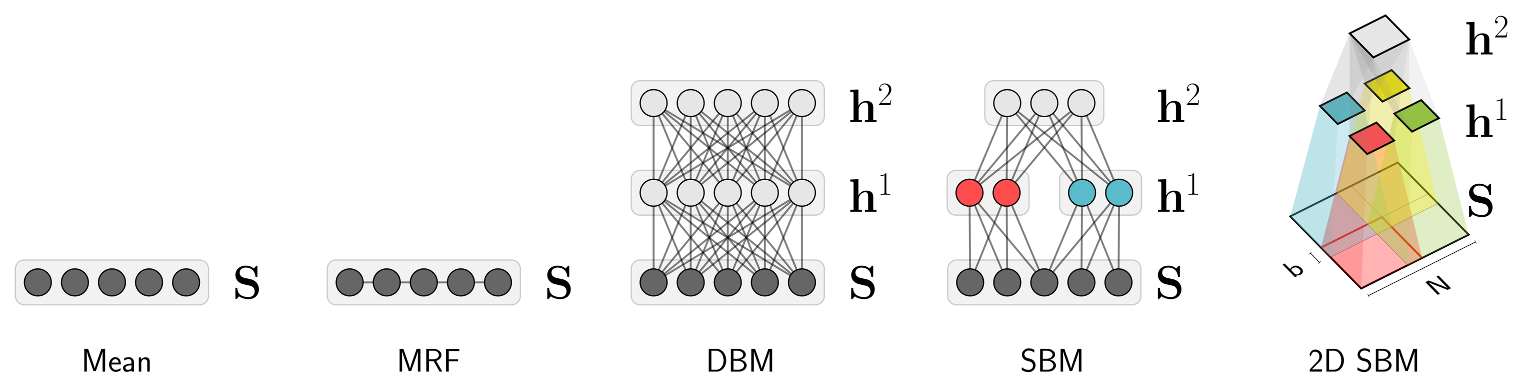
$$p(\mathbf{X},\mathbf{A},\mathbf{S},\mathbf{H}|\boldsymbol\theta) = \frac{1}{Z(\lambda)} p(\mathbf{A}|\boldsymbol\theta^a)p(\mathbf{S},\mathbf{H}|\boldsymbol\theta^s)\prod_i p(\mathbf{x}_i|\mathbf{A},\mathbf{s}_i,\boldsymbol\theta^a)^\lambda$$

Appearance

Represent the distribution over pixels inside each part using a mixture of colour histograms, capturing variability **within** and **across** images.



Shape



$$E(\mathbf{S},\mathbf{h}^1,\mathbf{h}^2|\boldsymbol\theta^s) = \sum_{i,l} b_{li}s_{li} + \sum_{i,j,l} w^1_{lij}s_{li}h^1_j + \sum_j c^1_j h^1_j + \sum_{j,k} w^2_{jk}h^1_j h^2_k + \sum_k c^2_k h^2_k$$

Inference

Block-Gibbs samplings of $\mathbf{A}$, $\mathbf{S}$ and $\mathbf{H}$:

procedure INFER($\mathbf{X},\boldsymbol\theta$)
  Initialize $\mathbf{S}^{(1)}$, $\mathbf{H}^{(1)}$
  for $t \leftarrow 2 : chain\_length$ **do**
    $\mathbf{A}^{(t)} \sim p(\mathbf{A}|\mathbf{S}^{(t-1)},\mathbf{H}^{(t-1)},\mathbf{X},\boldsymbol\theta)$
    $\mathbf{S}^{(t)} \sim p(\mathbf{S}|\mathbf{A}^{(t)},\mathbf{H}^{(t-1)},\mathbf{X},\boldsymbol\theta)$
    $\mathbf{H}^{(t)} \sim p(\mathbf{H}|\mathbf{S}^{(t)},\boldsymbol\theta)$
  end for
end procedure
return $\{\mathbf{S}^{(t)}\}_{t=burnin:chain\_length}$

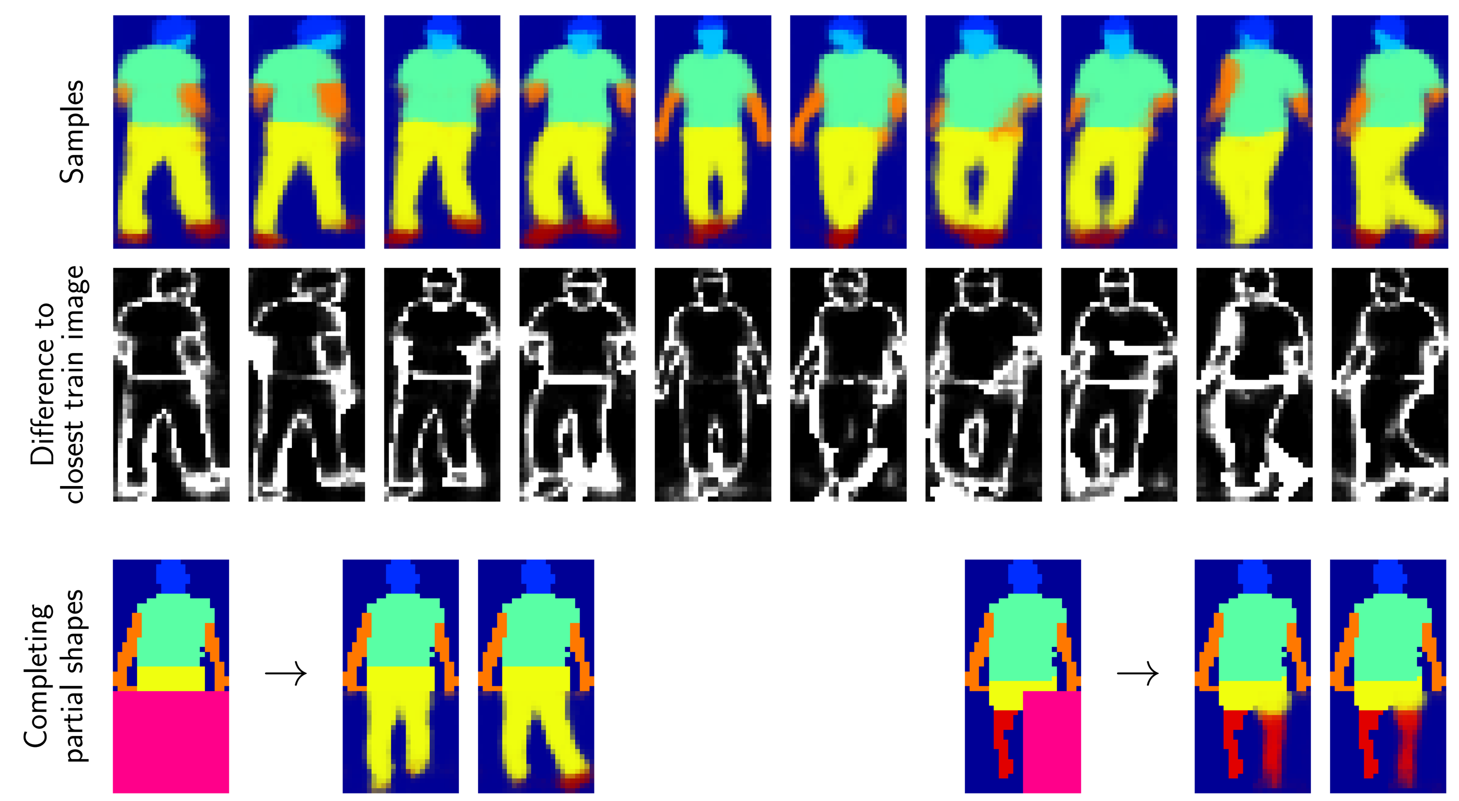
Seeding

Helpful to run several inference chains, each initializing $\mathbf{S}^{(1)}$ to a different value. The 'best' inference is retained and the others are discarded. The computation of the likelihood $p(\mathbf{X}|\boldsymbol\theta)$ of image $\mathbf{X}$ is intractable, so we approximate the quality of each inference using a scoring function:

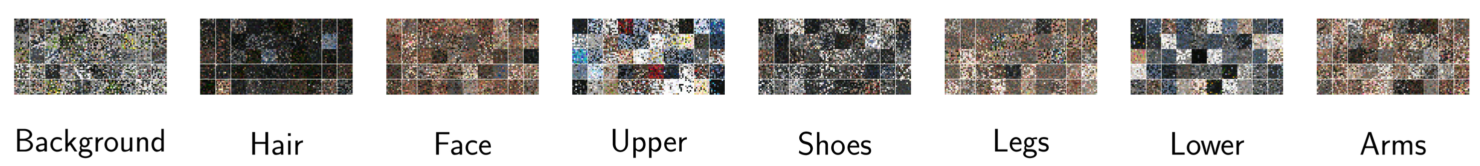
$$F(\mathbf{X}) = \frac{1}{T}\sum_t p(\mathbf{X},\mathbf{A}^{(t)},\mathbf{S}^{(t)},\mathbf{H}^{(t)}|\boldsymbol\theta)$$

3. Samples from the generative components

Shape

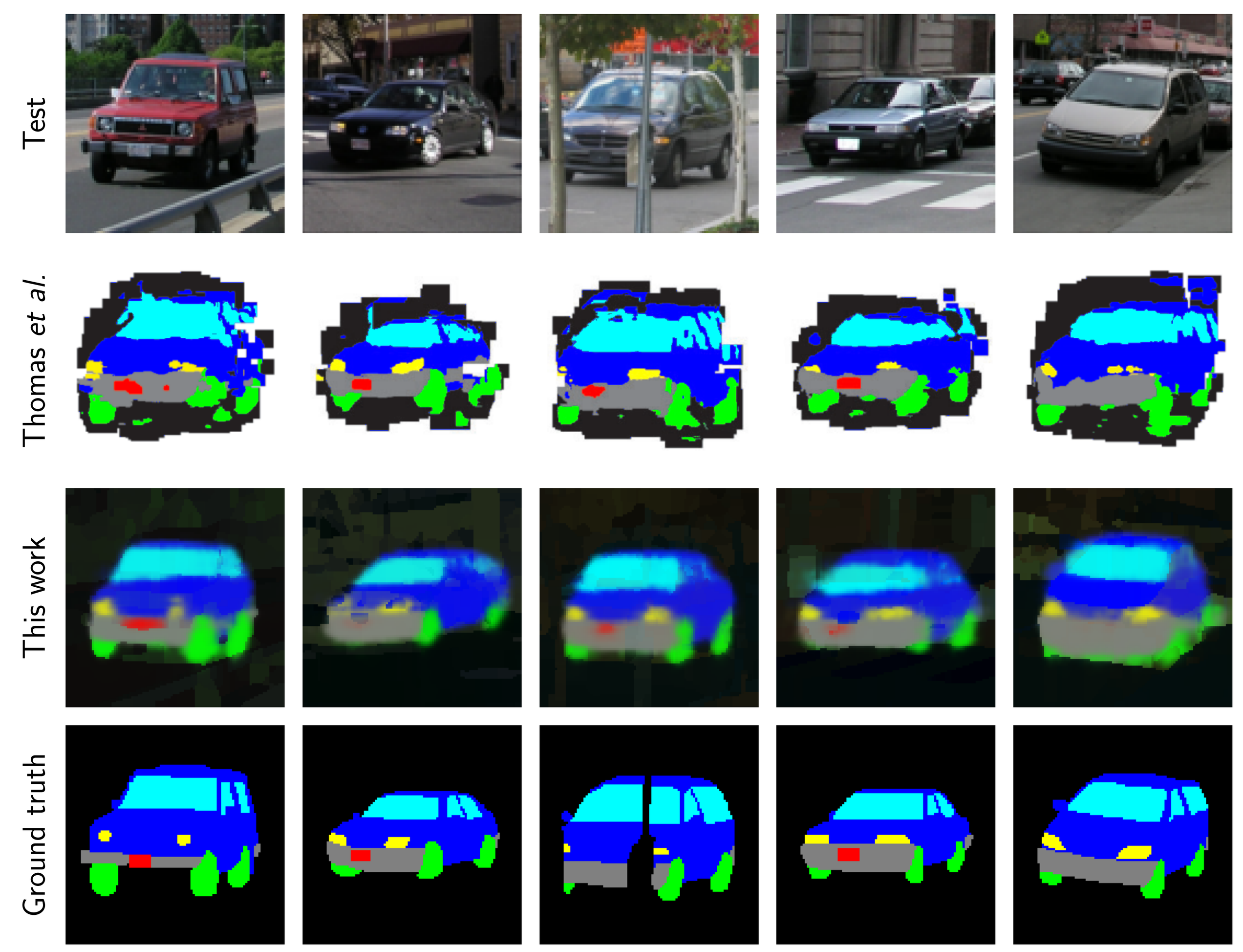


Appearance



4. Experimental results

	FG	BG	Upper Body	Lower Body	Head	Average
Bo and Fowlkes	73.3%	81.1%	73.6%	71.6%	51.8%	69.5%
This work	70.7%	72.8%	68.6%	66.7%	53.0%	65.3%
This work + superpixels	71.6%	73.8%	69.9%	68.5%	54.1%	66.6%
Top seed	59.0%	61.8%	56.8%	49.8%	45.5%	53.5%
Top seed + superpixels	61.6%	67.3%	60.8%	54.1%	43.5%	56.4%



	BG	Body	Wheel	Window	Bumper	License	Light	Average
ISM	93.2%	72.2%	63.6%	80.5%	73.8%	56.2%	34.8%	86.8%
This work	94.6%	72.7%	36.8%	74.4%	64.9%	17.9%	19.9%	86.0%
Top seed	92.2%	68.4%	28.3%	63.8%	45.4%	11.2%	15.1%	81.8%