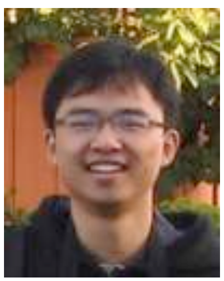
Layered Object Detection for Multi-Class Image Segmentation

Yi Yang



Sam Hallman

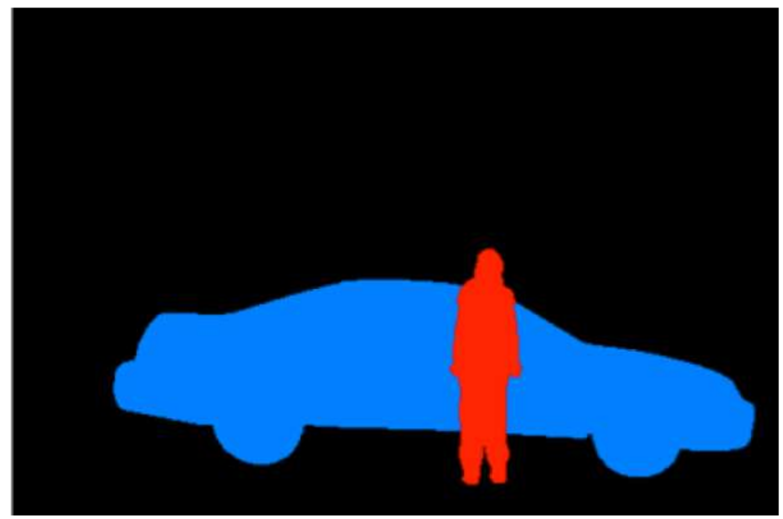
Deva Ramanan

Charless Fowlkes

UC Irvine

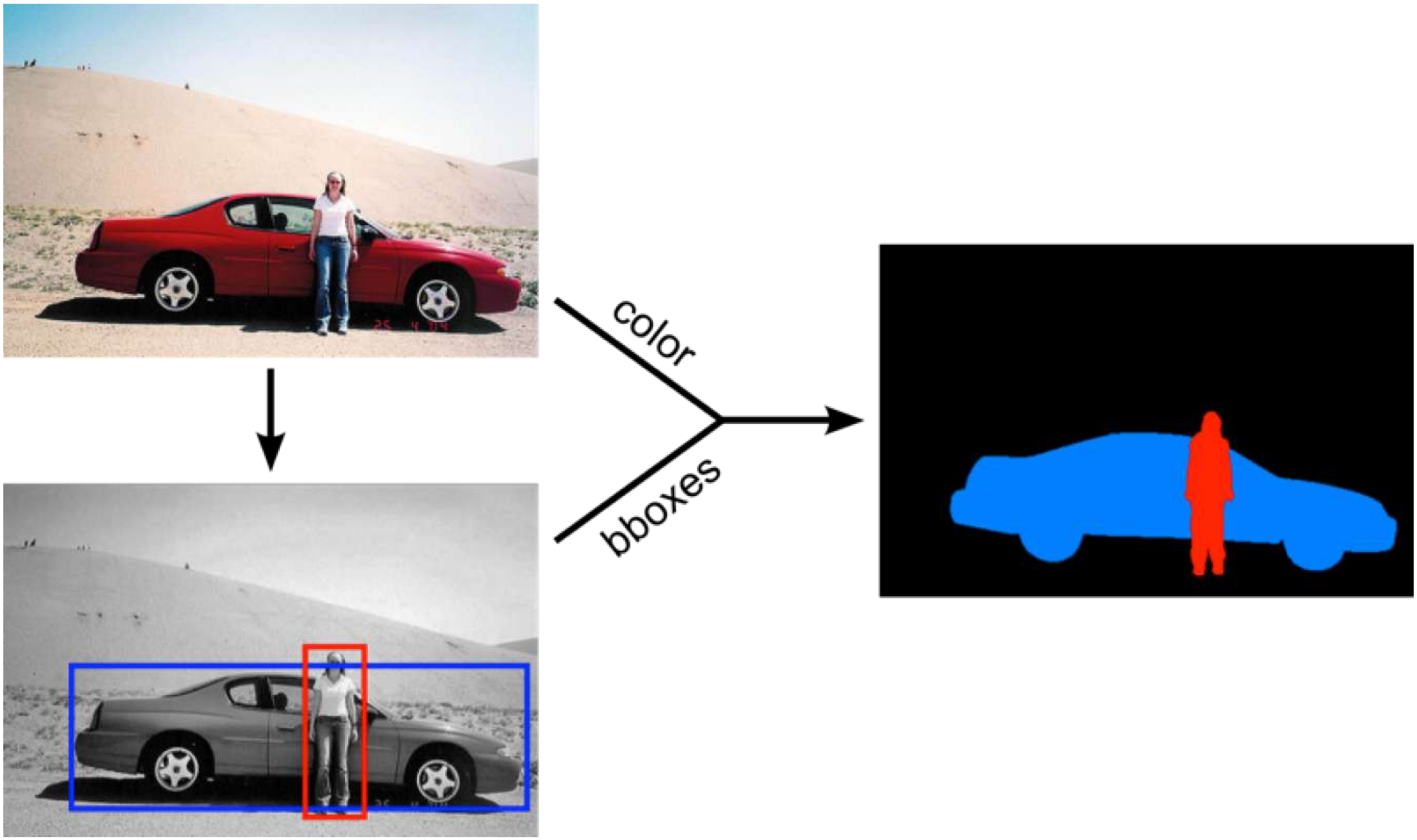
Introduction

- PASCAL competition
- 20 object categories + “background”

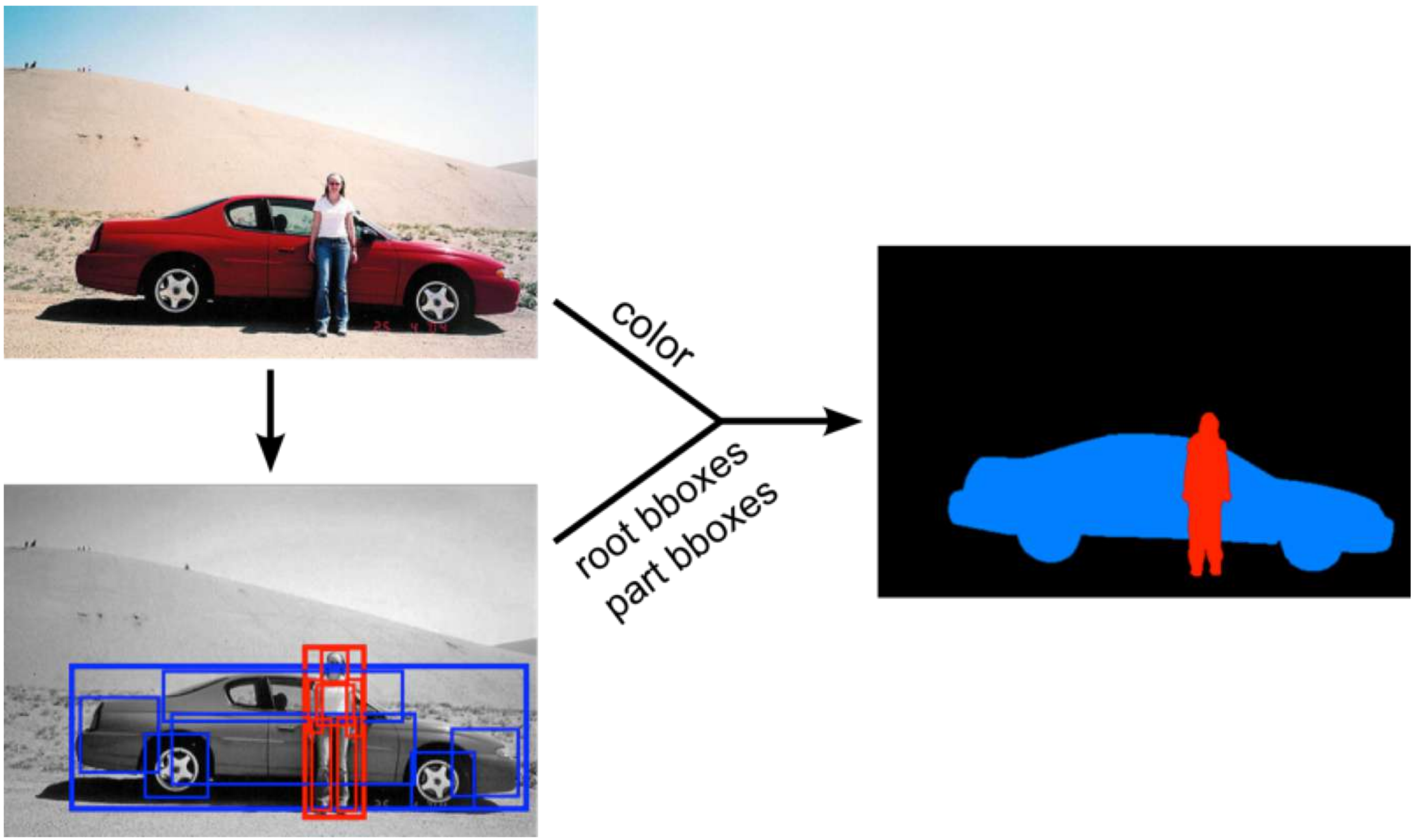


Desired result

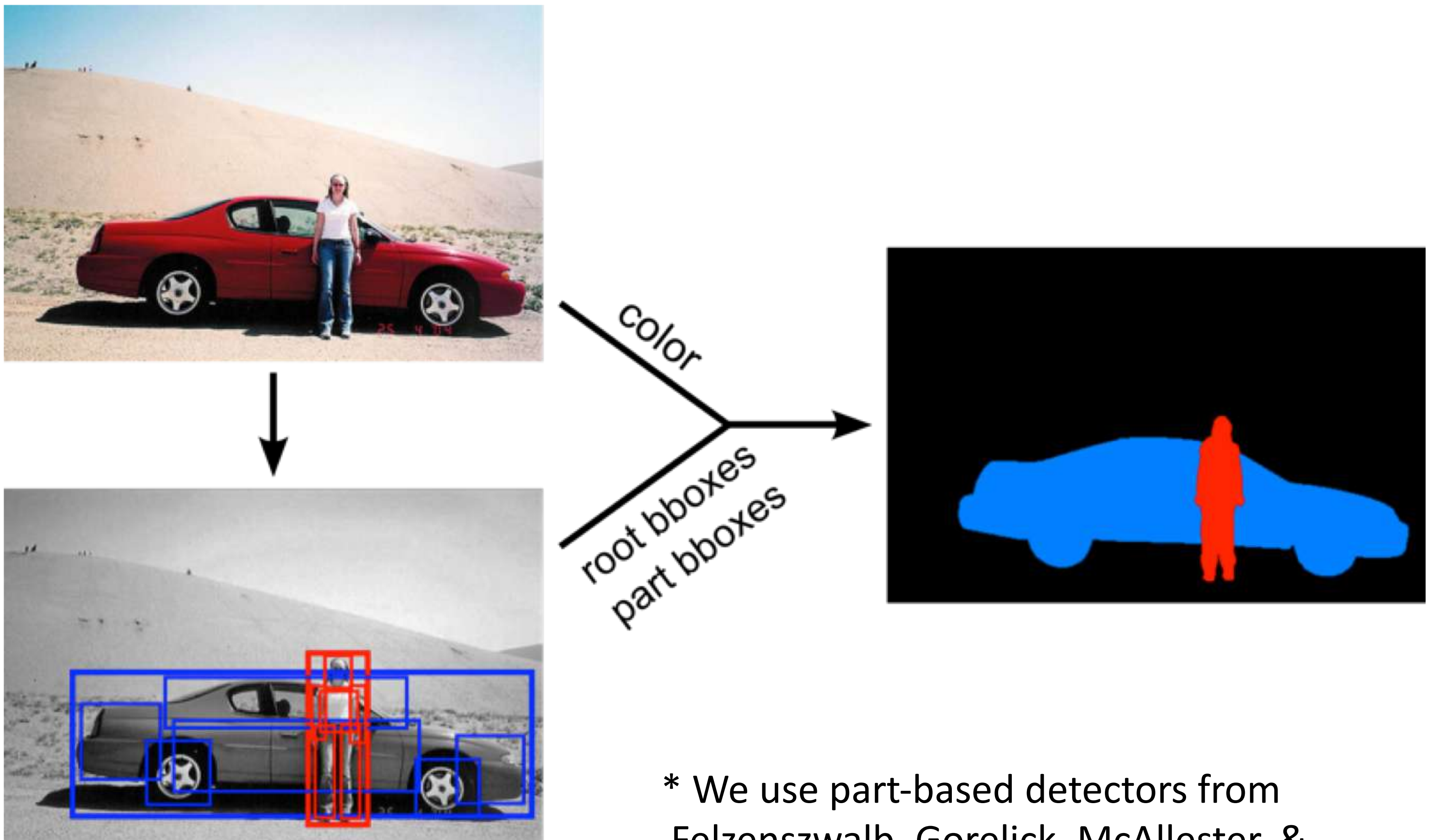
Introduction



Introduction



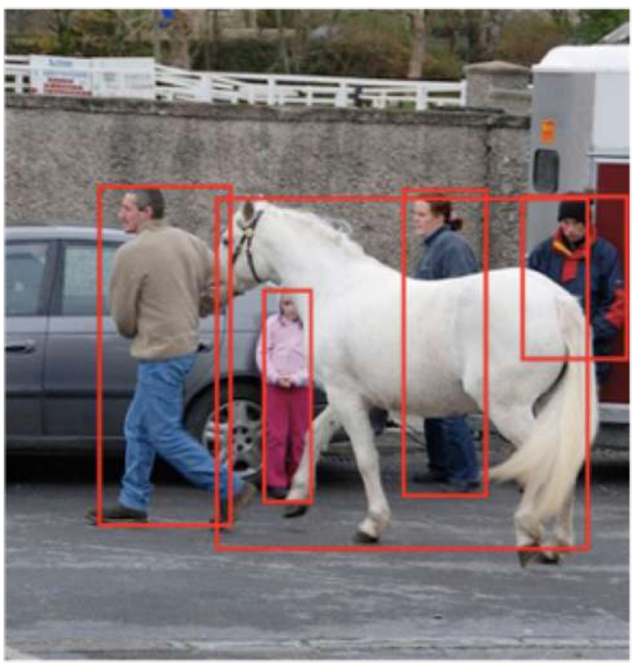
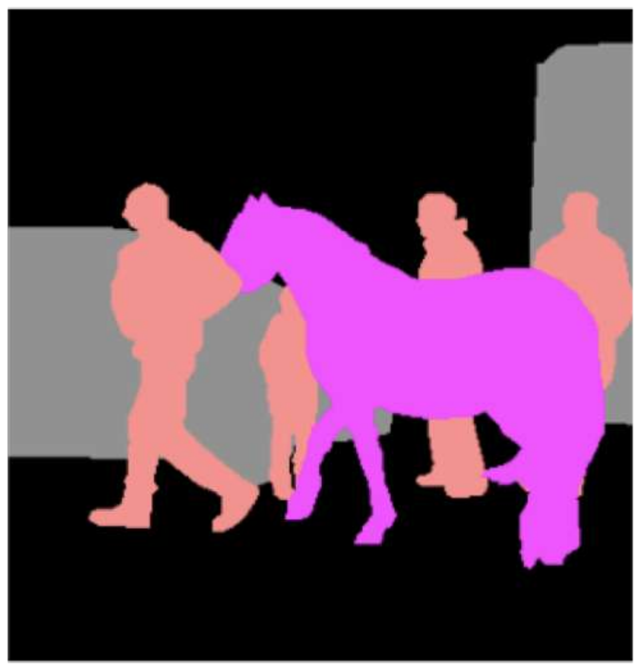
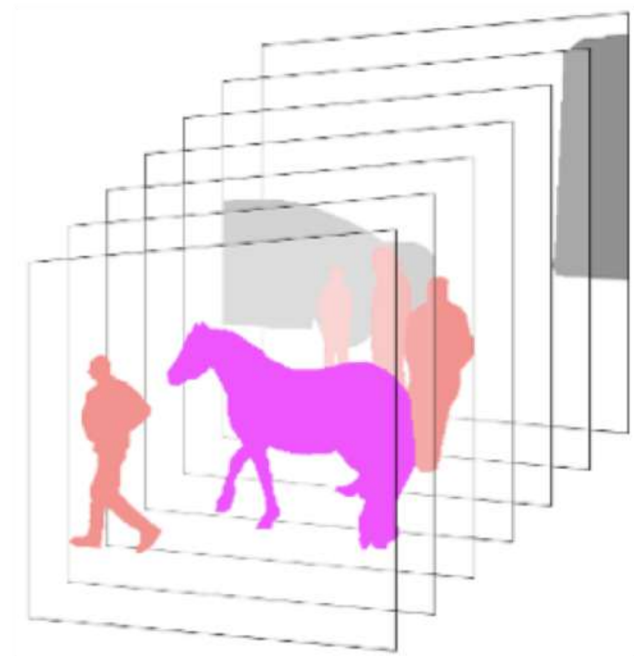
Introduction



* We use part-based detectors from Felzenszwalb, Gorelick, McAllester, & Ramanan PAMI 09

Layered Representation

- Model each detection in “2.1D”: detections ordered by *depth*

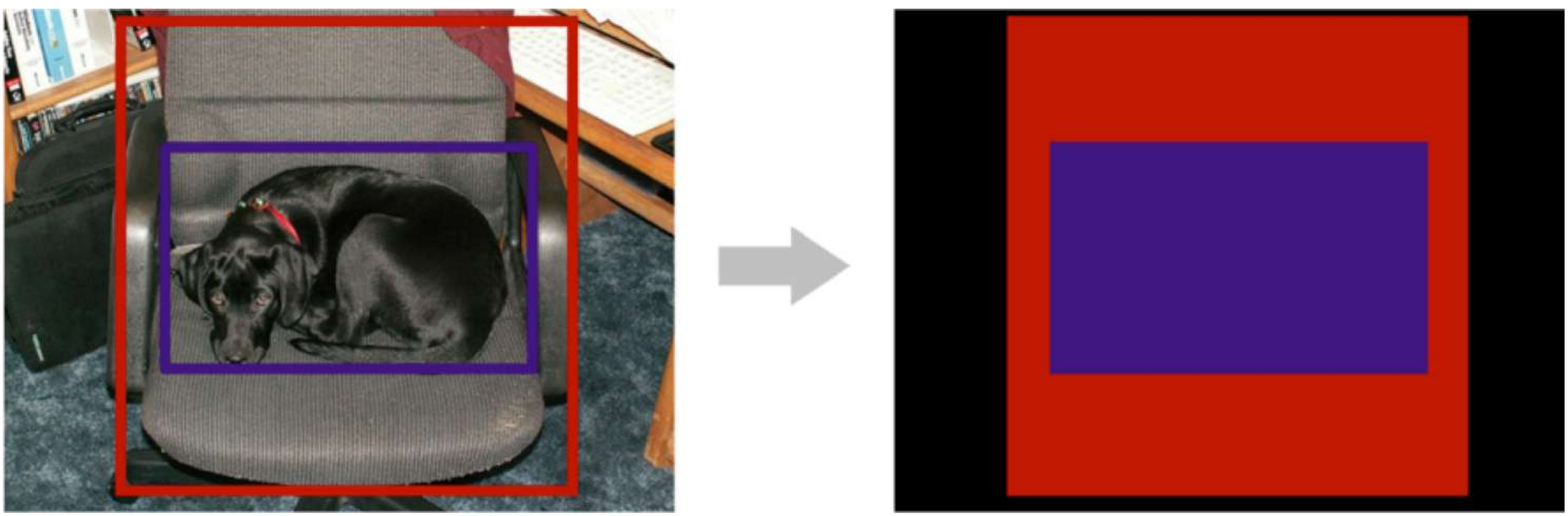
Desired result

Related work

- Related work combining segmentation and recognition
  - Biasing segmentation output based on object models [Yu et al. 03, Ramanan 06, Kumar et al. 05]
  - Iterating between bottom-up and top-down cues [Leibe et al. 04, Tu et al. 05]
- Related work on layers
  - Work in the video domain [Wang & Adelson 94, Jojic & Frey 01, Kumar et al. 04]
  - Some work in image domain [Gao et al. 07, Nitzberg et al. 93]

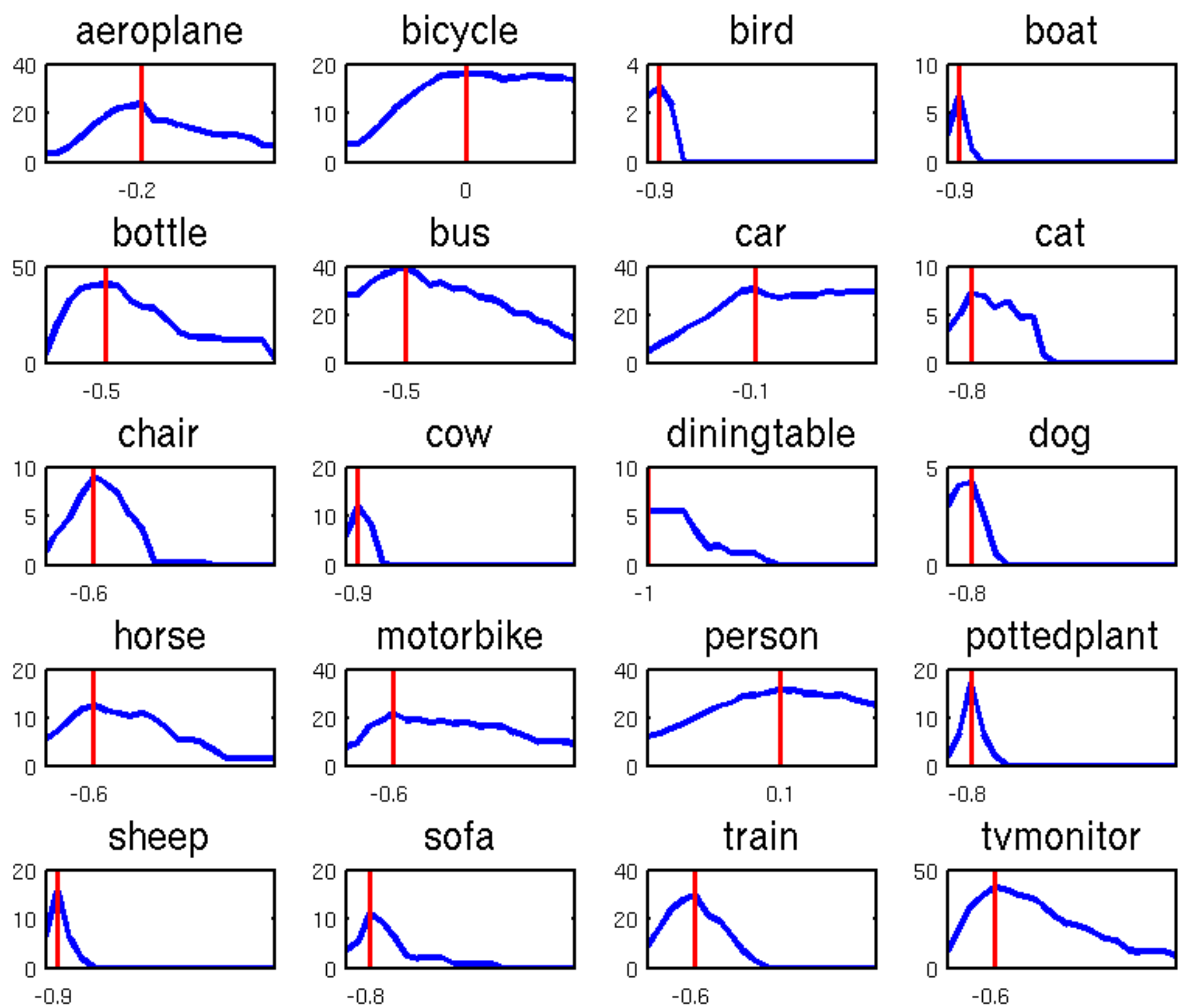
Issues

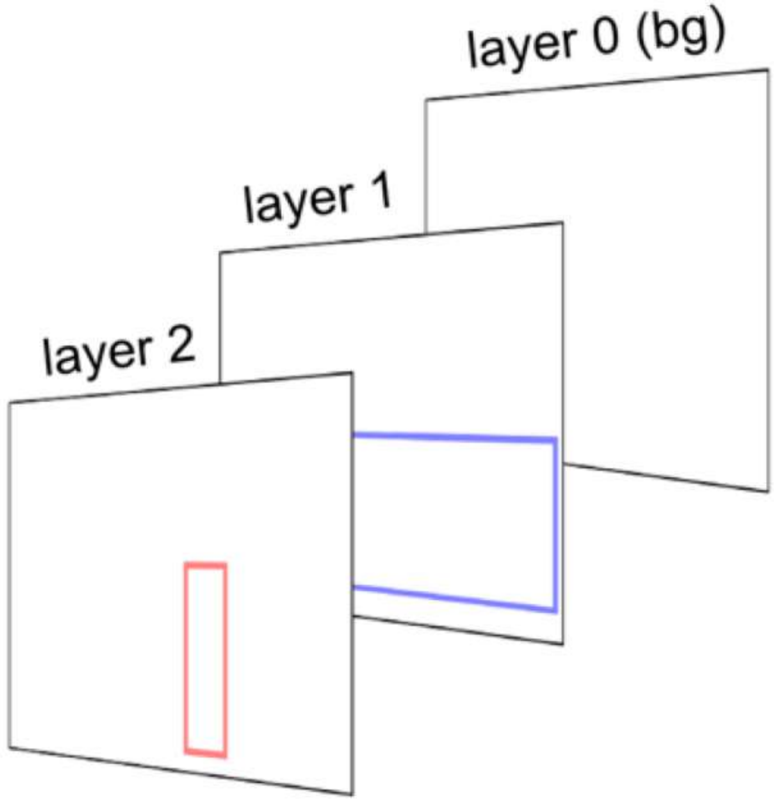
- Need detector thresholds
- Need comparable confidence scores

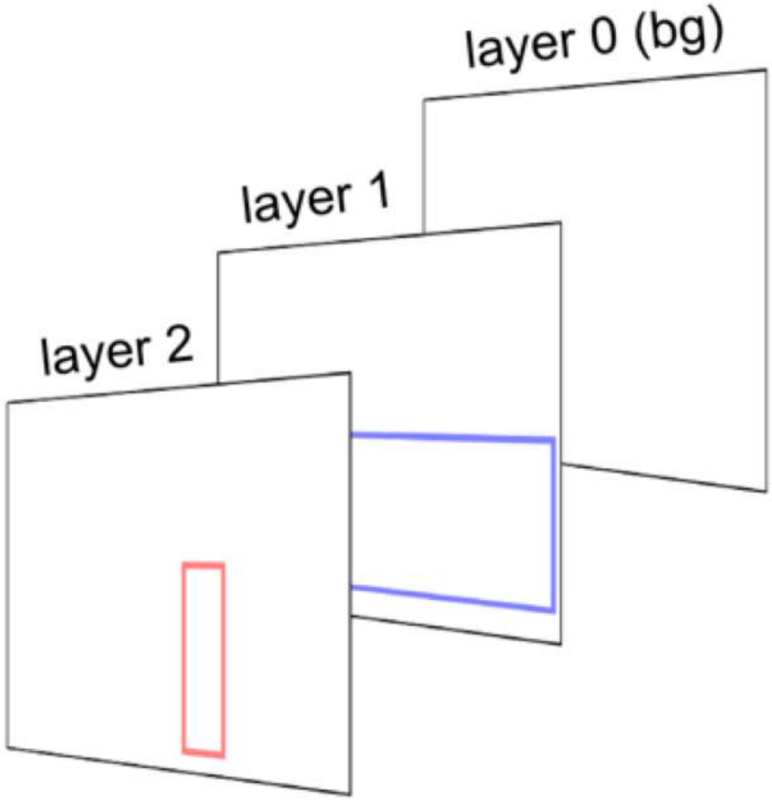


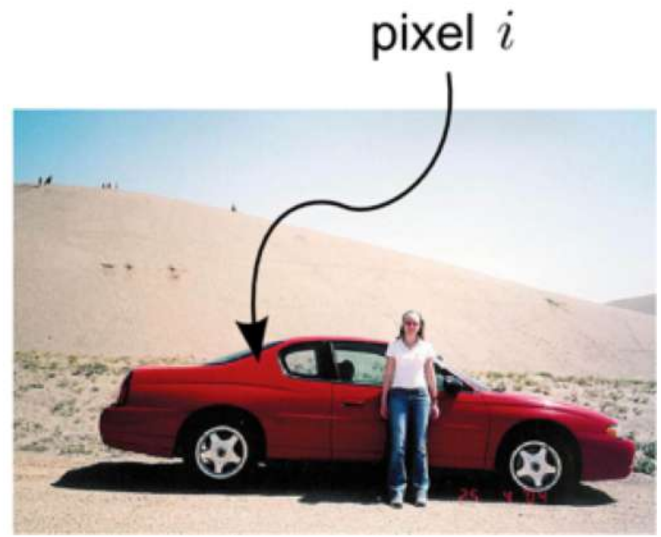
Detector calibration

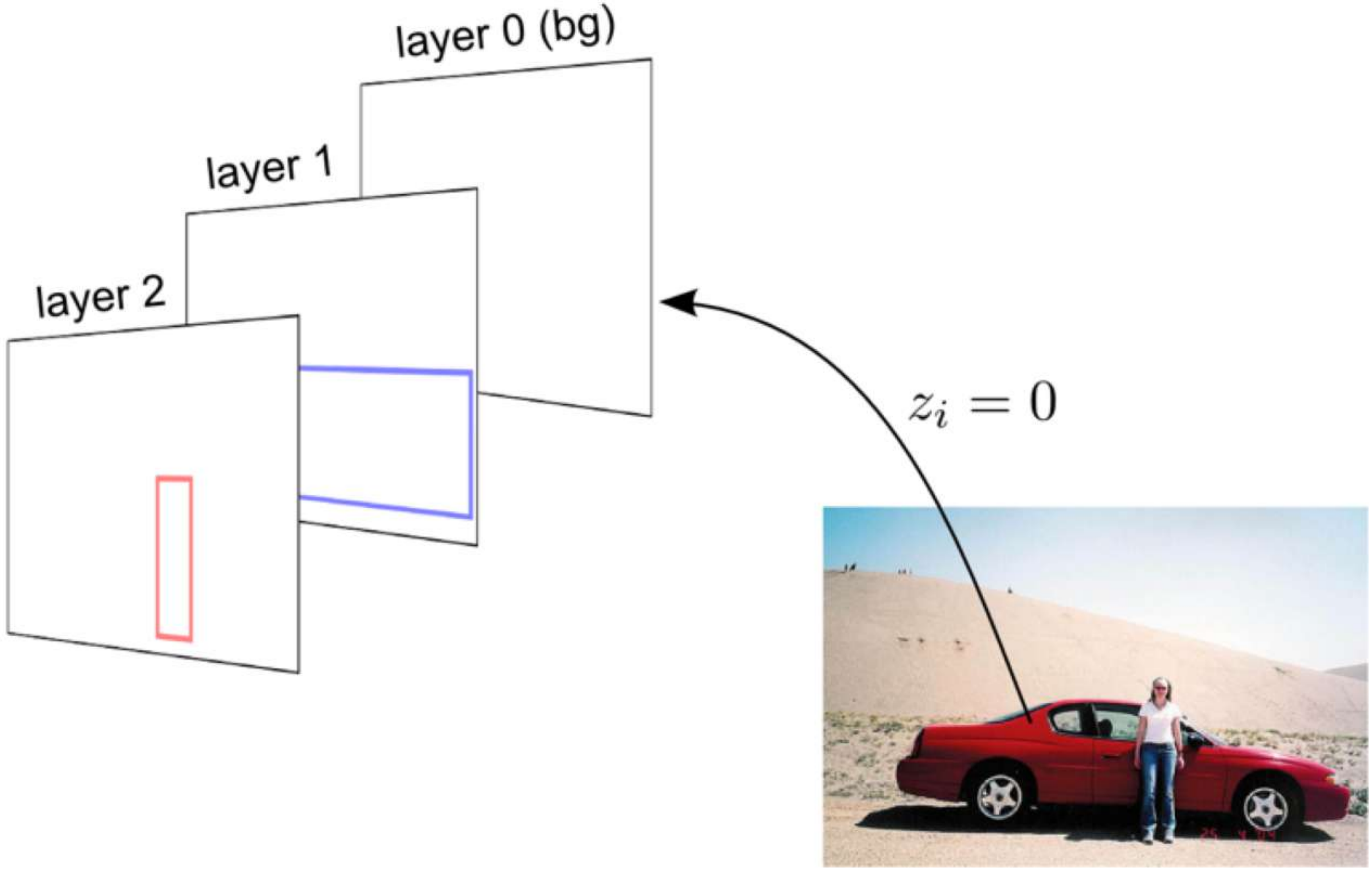
Find optimal threshold for pixel labeling

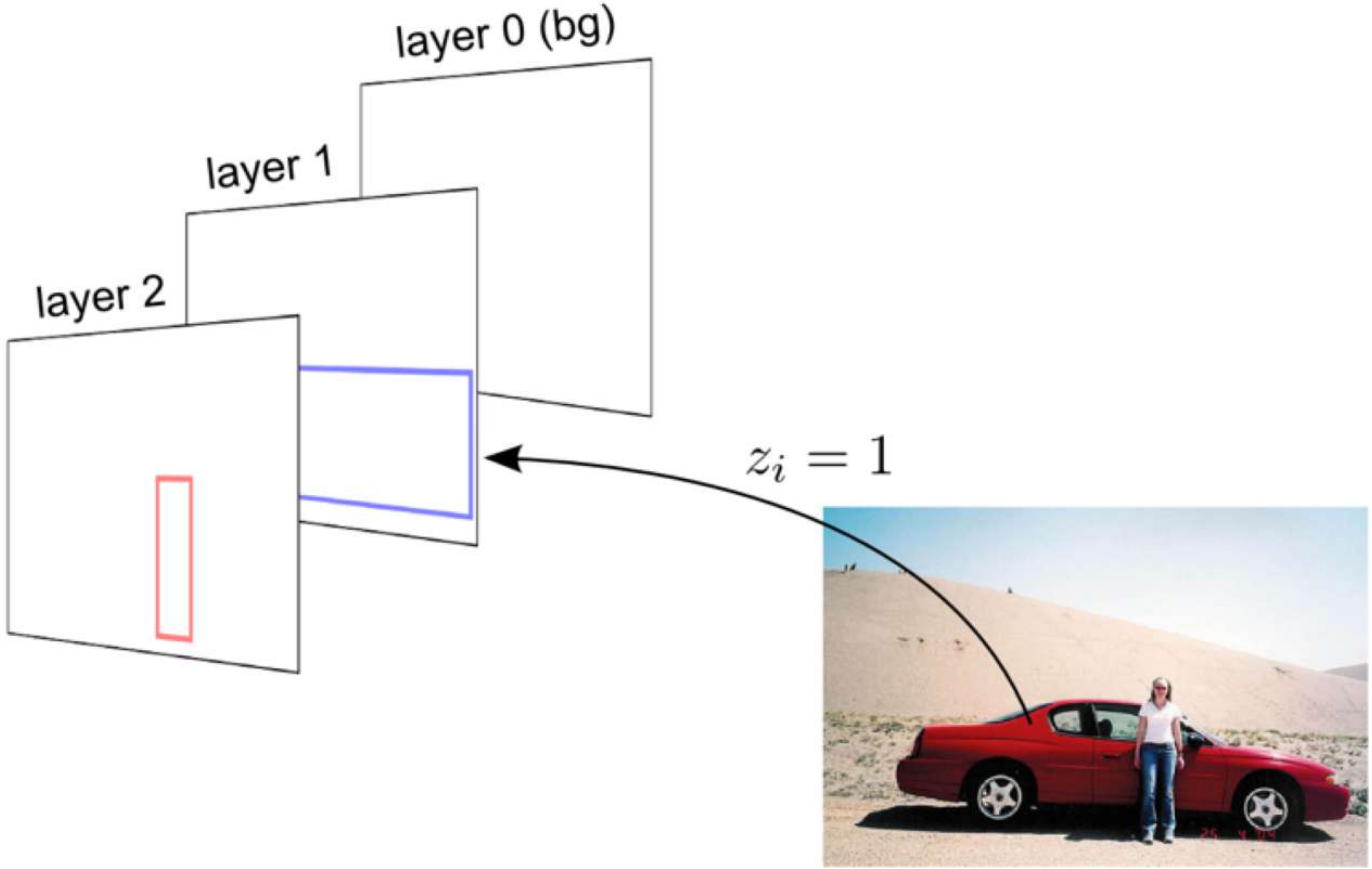


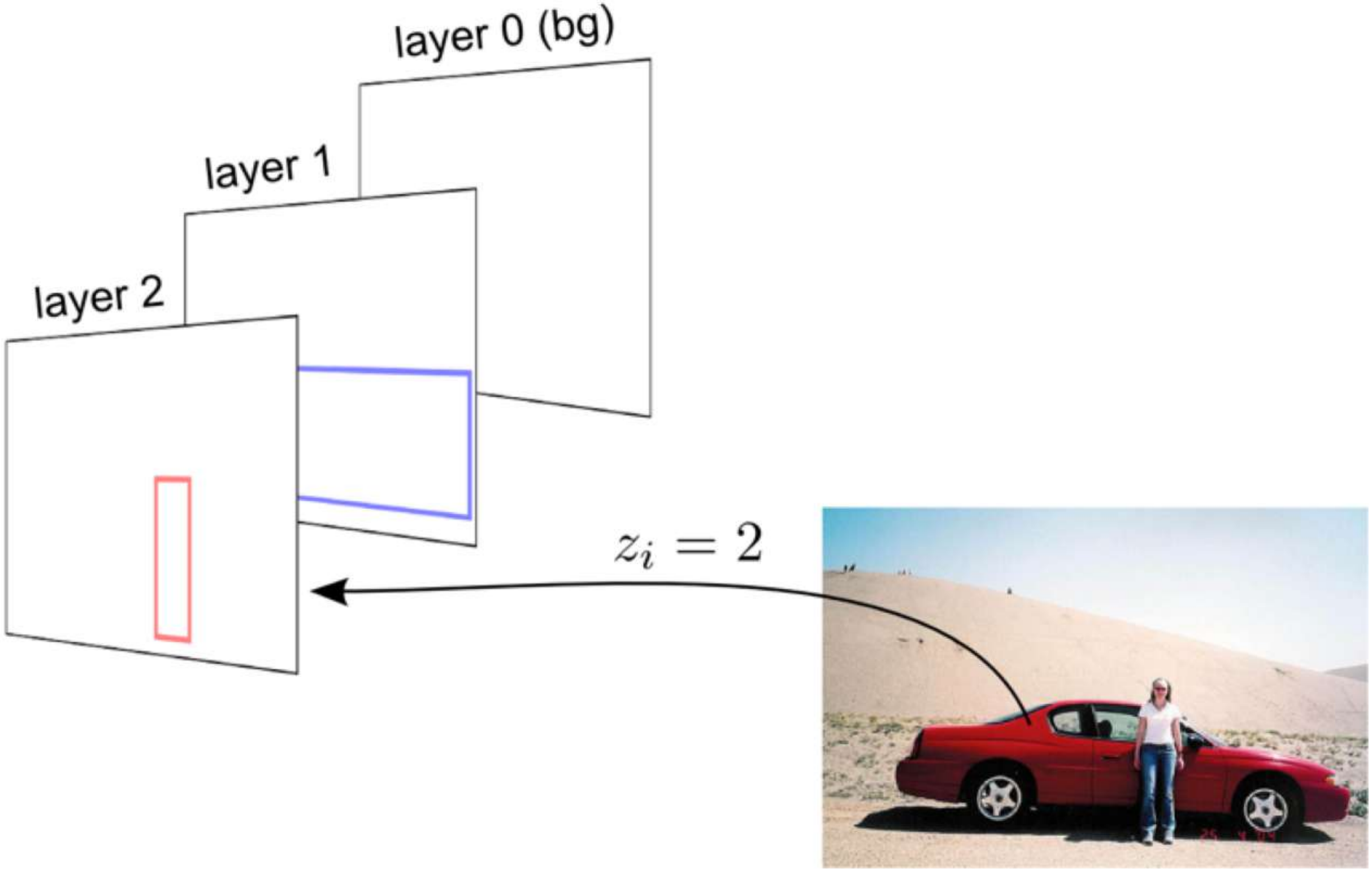

layer 0 (bg)
layer 1
layer 2


layer 0 (bg)
layer 1
layer 2

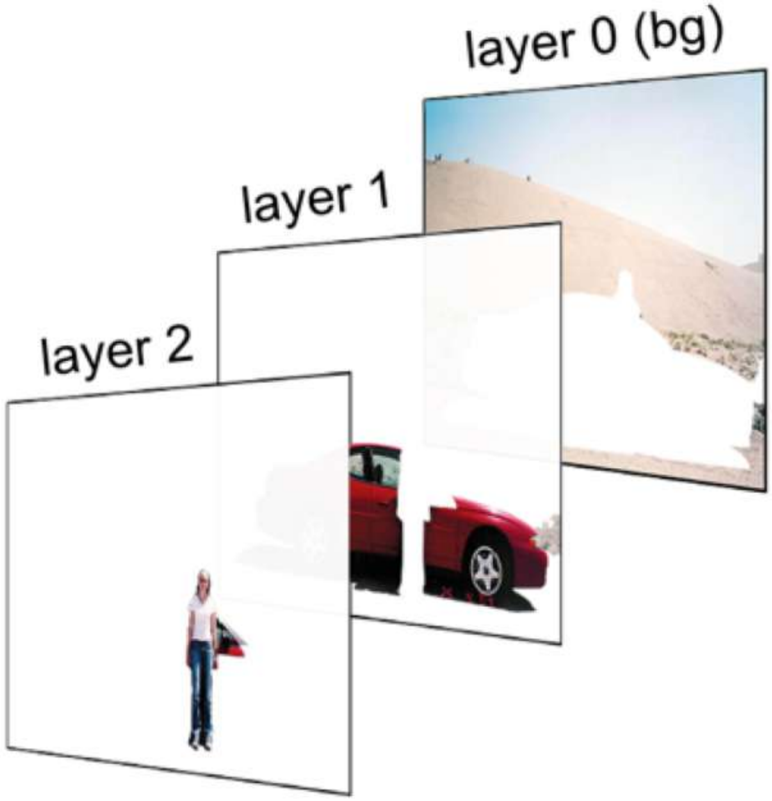

pixel i

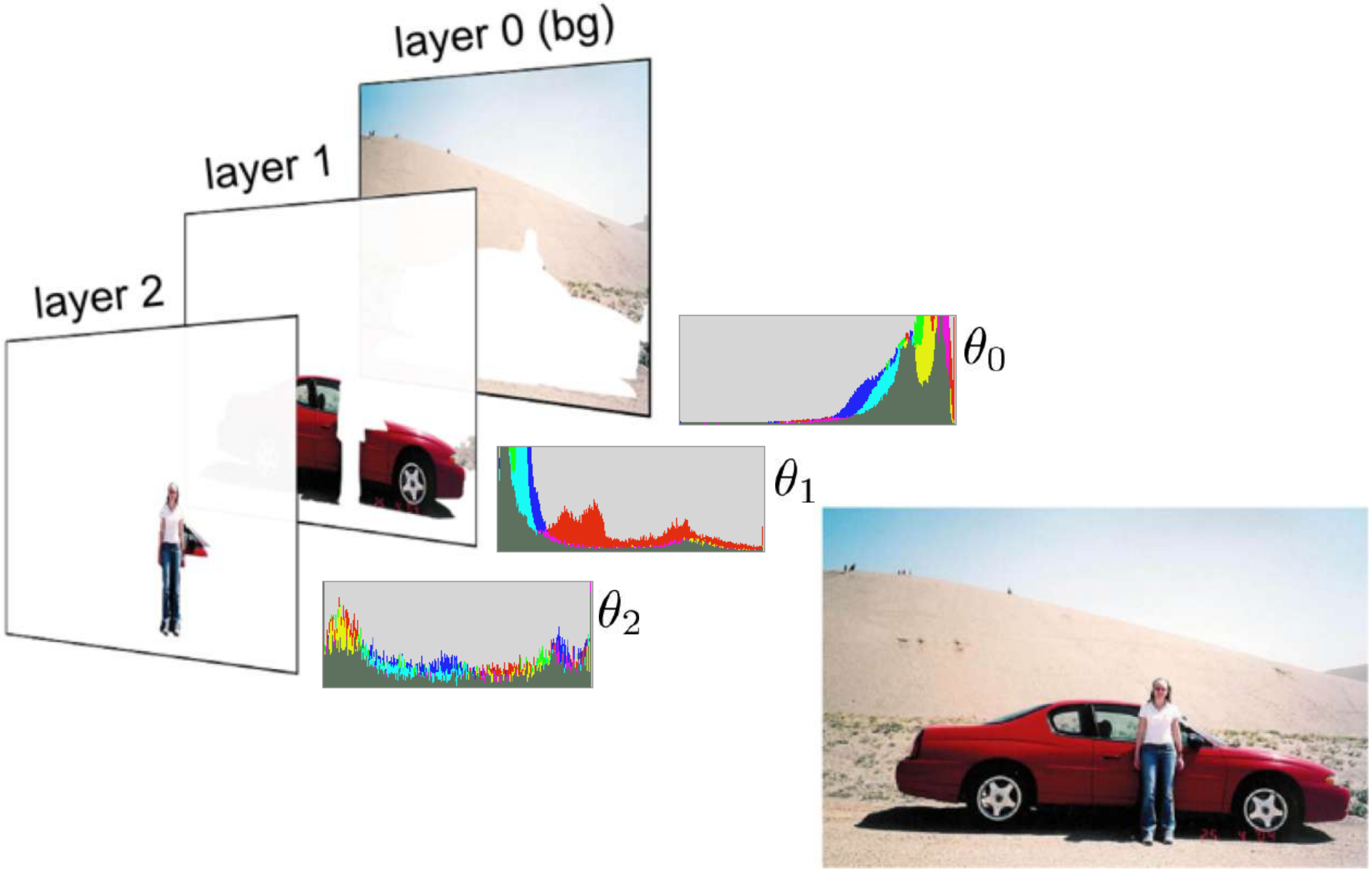

layer 0 (bg)
layer 1
layer 2
$z_i = 0$


layer 0 (bg)
layer 1
layer 2
$z_i = 1$



layer 0 (bg)
layer 1
layer 2
$z_i = 2$


layer 0 (bg)
layer 1
layer 2



layer 0 (bg)
layer 1
layer 2
θ_0
θ_1
θ_2

Model

N = # of detections

x_i = RGB values for pixel i

z_i = label for pixel i (1..N or 0 for background)

θ_n = color model for nth layer

d_π = set of detections with ordering π

$$P(z, x|\theta, d_\pi) = \underbrace{\prod_i}_{\text{product over pixels}} \underbrace{P(z_i|d_\pi)}_{\text{shape}} \underbrace{P(x_i|\theta_{z_i})}_{\text{color likelihood}}$$

Inference: coordinate descent

$$P(z, x|\theta, d_\pi) = \underbrace{\prod_i}_{\text{product over pixels}} \underbrace{P(z_i|d_\pi)}_{\text{shape}} \underbrace{P(x_i|\theta_{z_i})}_{\text{color likelihood}}$$

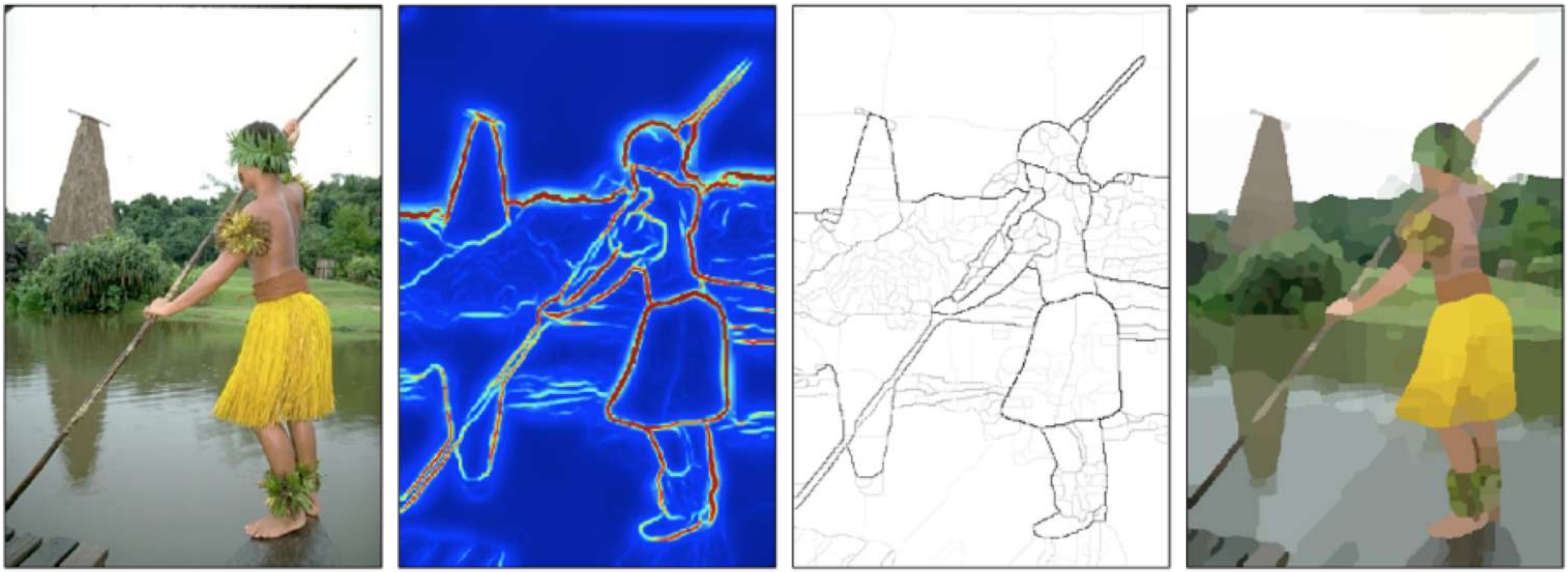
$$f(z, \theta) = -\log P(z, x|\theta, d_\pi)$$

Step 1: $\arg\min_z f(z, \theta)$ (per-pixel segmentation given shape prior & color likelihood)

Step 2: $\arg\min_\theta f(z, \theta)$ (fit color models to segments and background)

Color models are learned on-the-fly for each detection instance
(e.g., people may wear blue or red shirts)

Bottom-up segmentation



Use hierarchical segmentation of Arbelaez, Maire, Fowlkes, Malik (2009) at a threshold which generates ~200 segments per image.

Inference: coordinate descent

product
over pixels

$$P(z, x|\theta, d_\pi) = \prod_i P(z_i|d_\pi) P(x_i|\theta_{z_i})$$

shape

color
likelihood

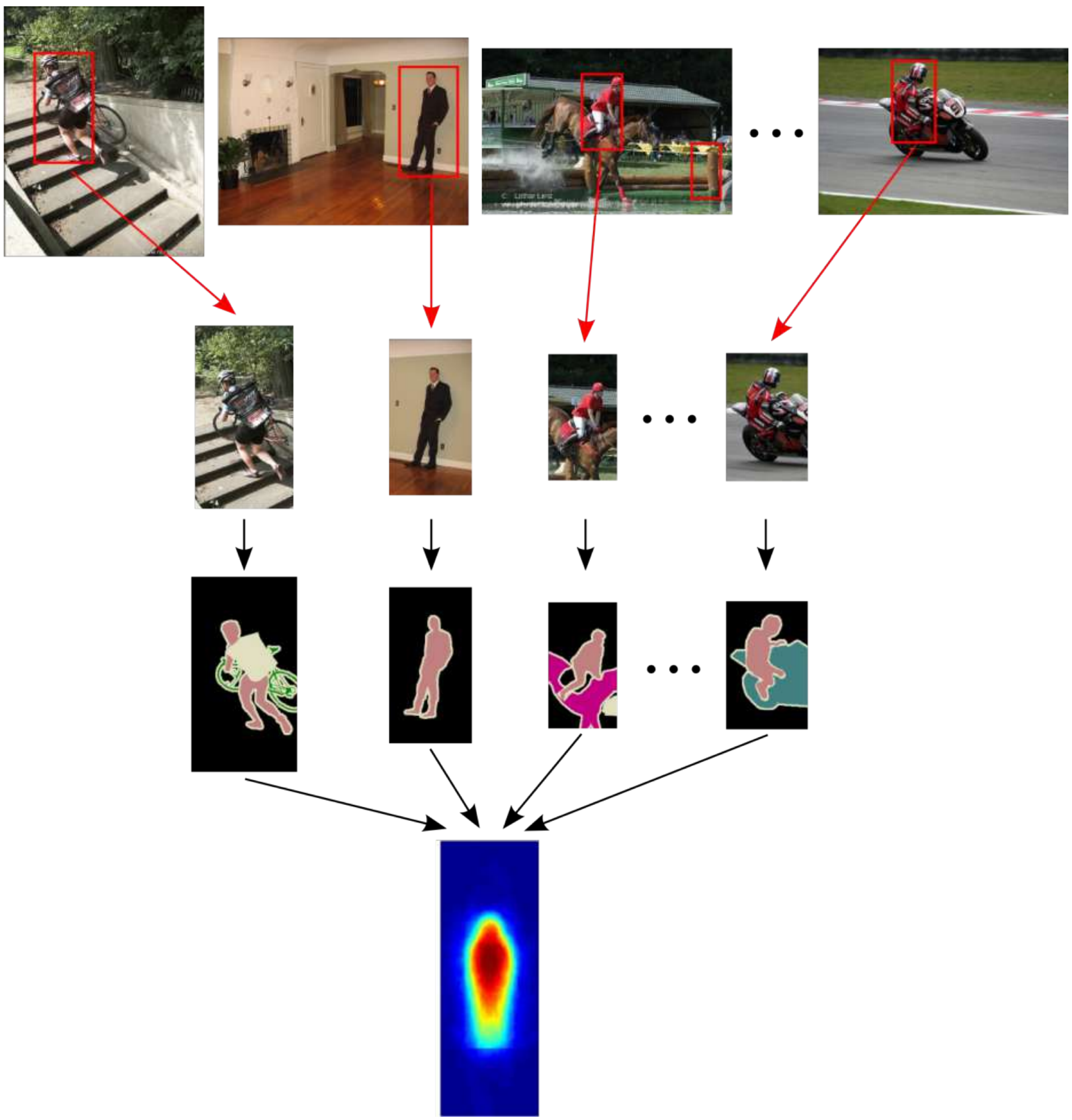
$$f(z, \theta) = -\log P(z, x|\theta, d_\pi)$$

Step 1: $\arg\min_{z \in \mathcal{Z}} f(z, \theta)$ (per-pixel segmentation given shape prior & color likelihood)

Step 2: $\arg\min_{\theta} f(z, \theta)$ (fit color models to segments and background)

$\mathcal{Z}$ = set of pixel labelings consistent with superpixel map

Building $P(z_i | d_\pi)$



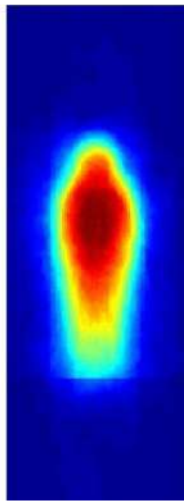
person detections

true positives

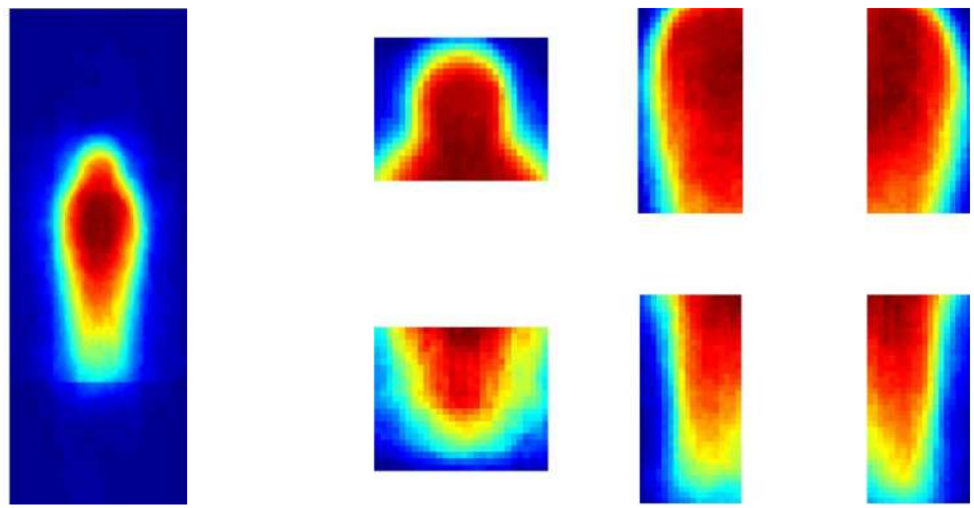
ground truth segmentations

shape prior for the *person* class

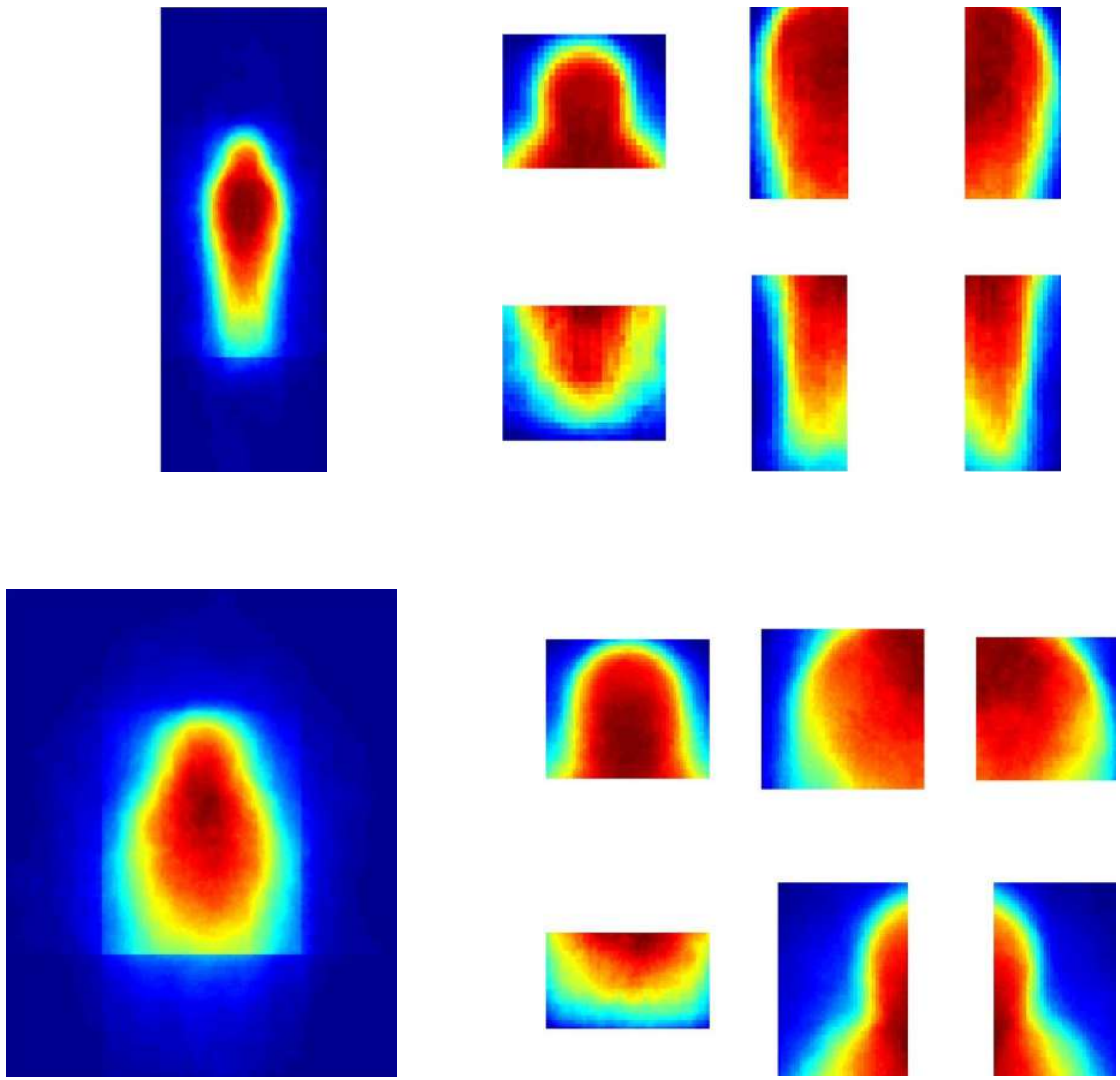
Shape priors (cont’d)



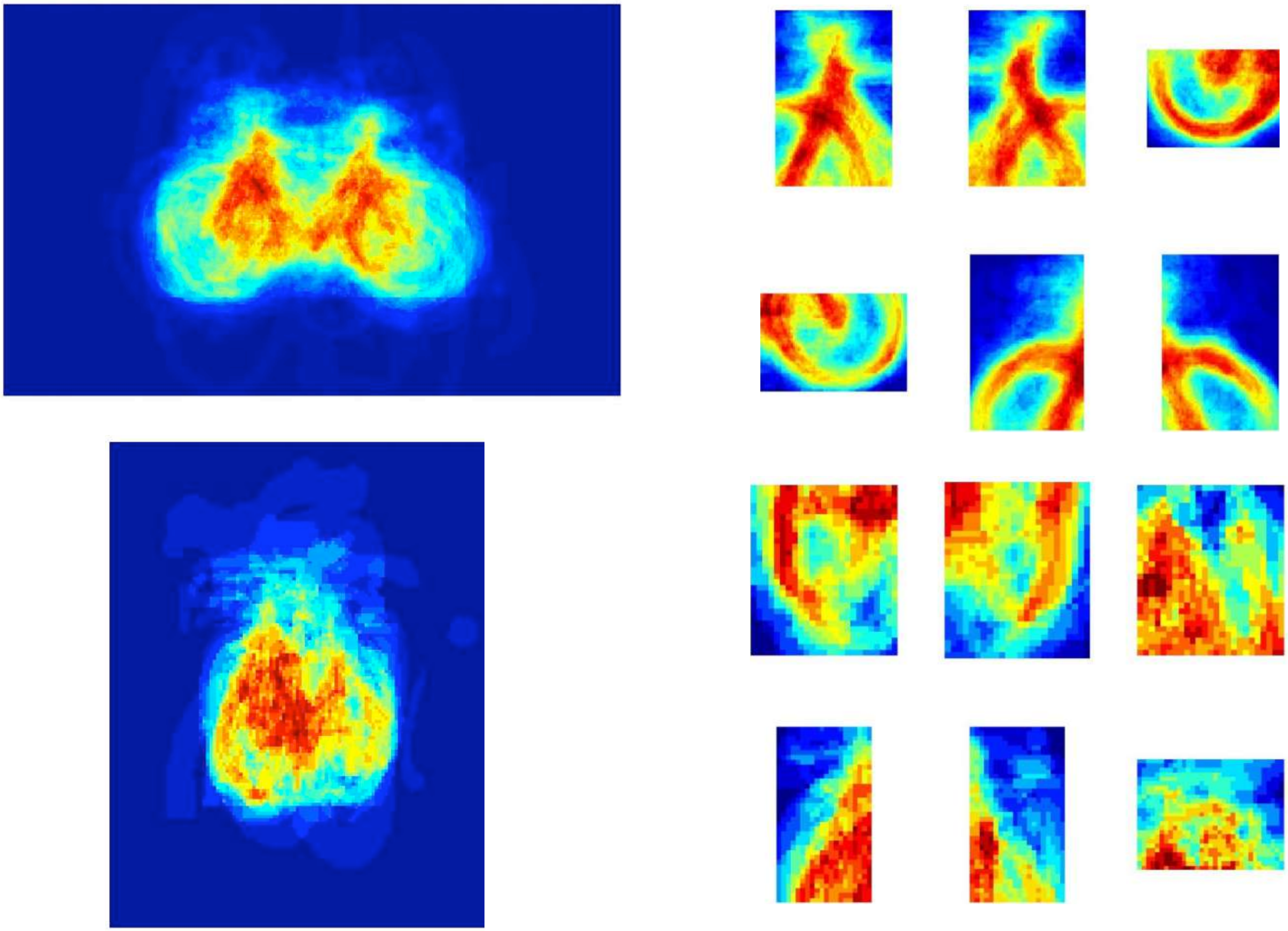
Shape priors (cont’d)



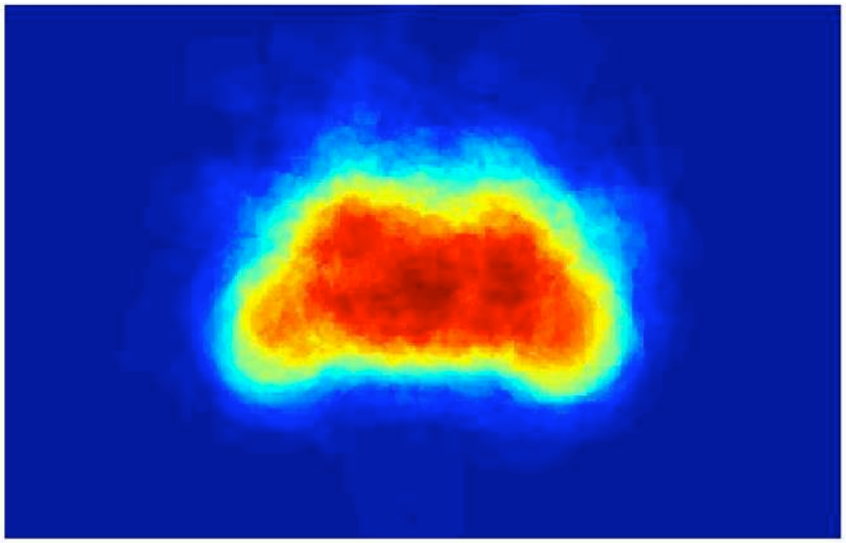
Shape priors (cont’d)

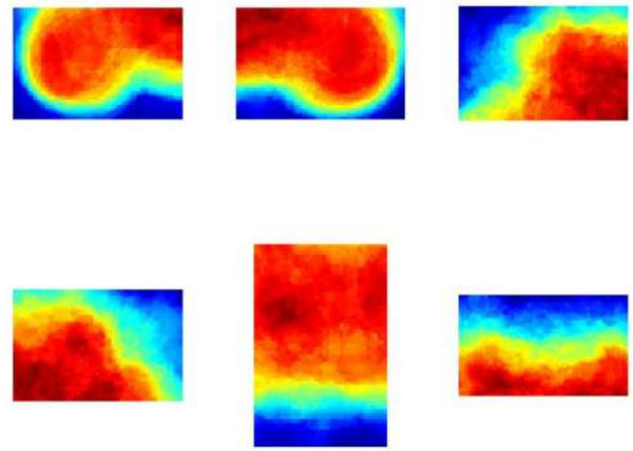


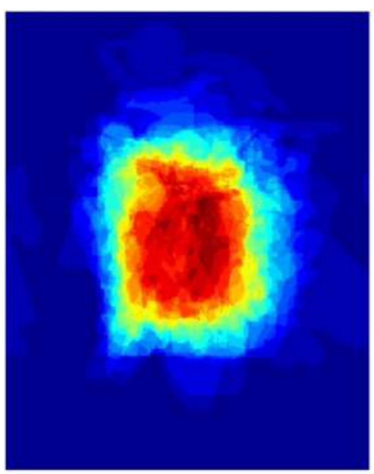
Bicycle part-based priors

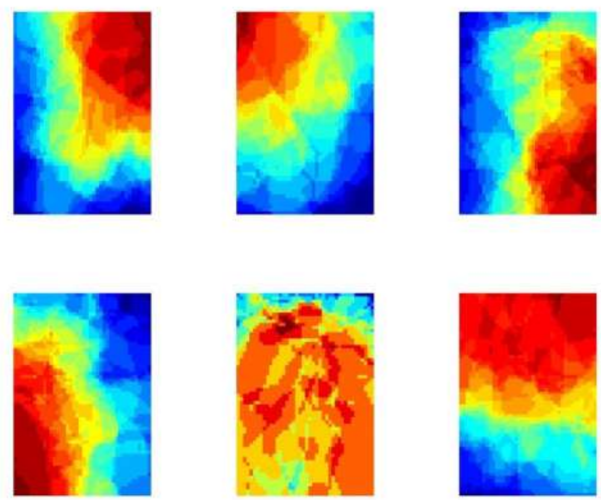


Motorcycle part-based priors

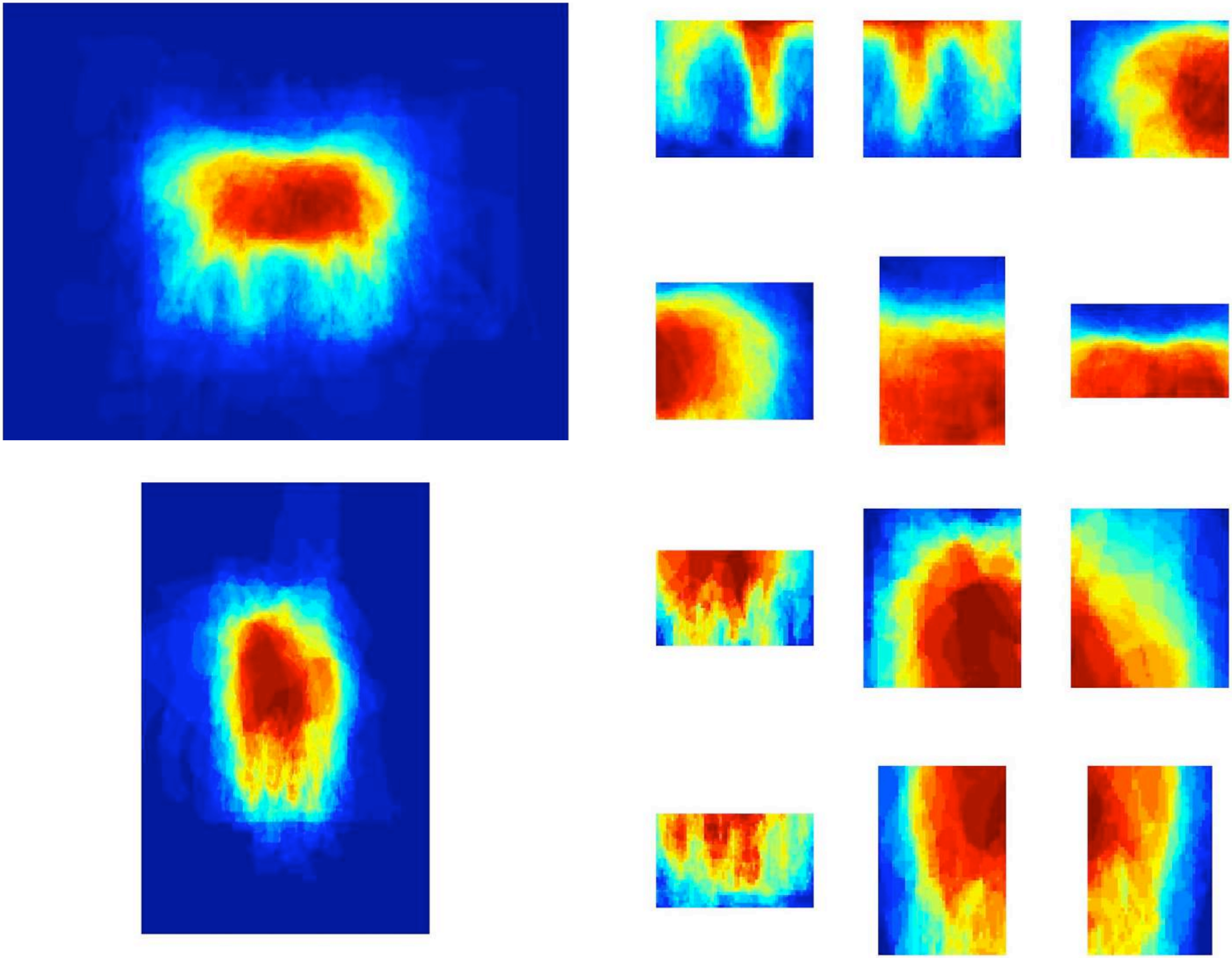




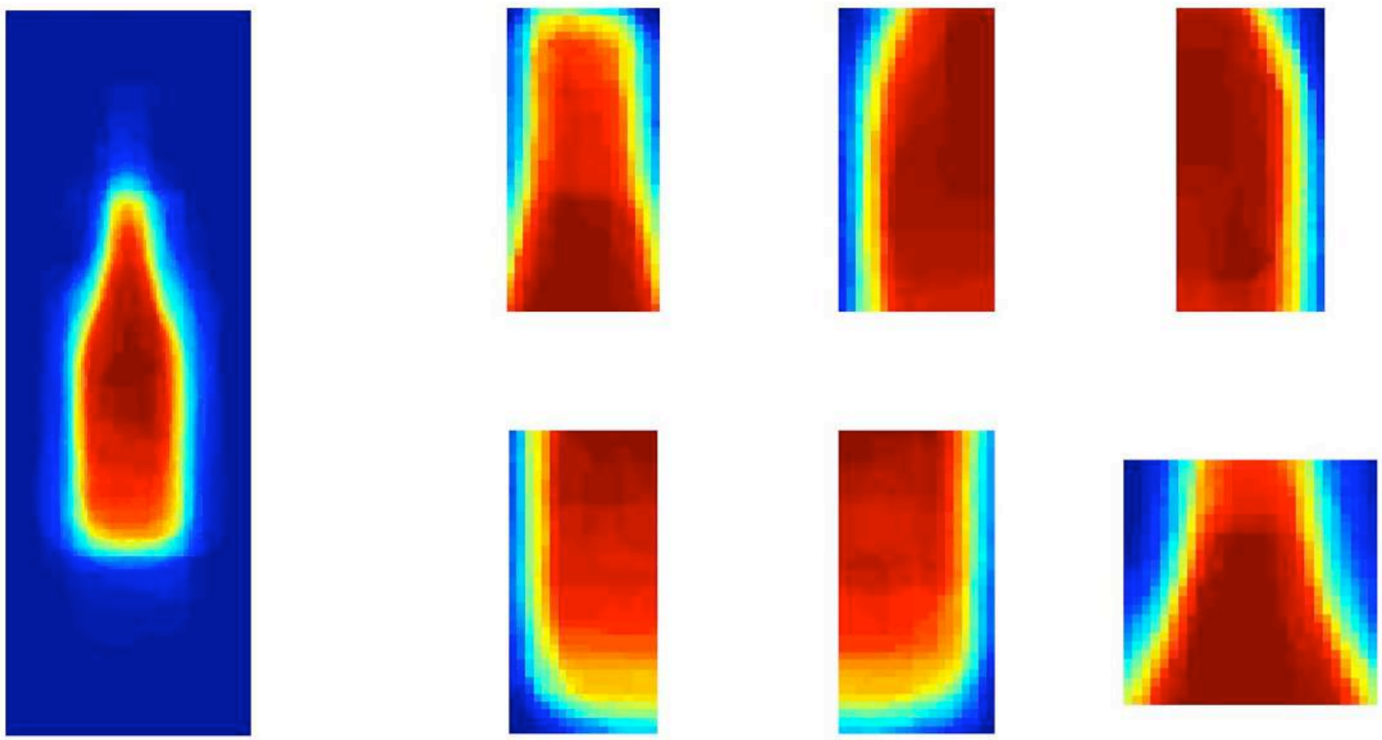




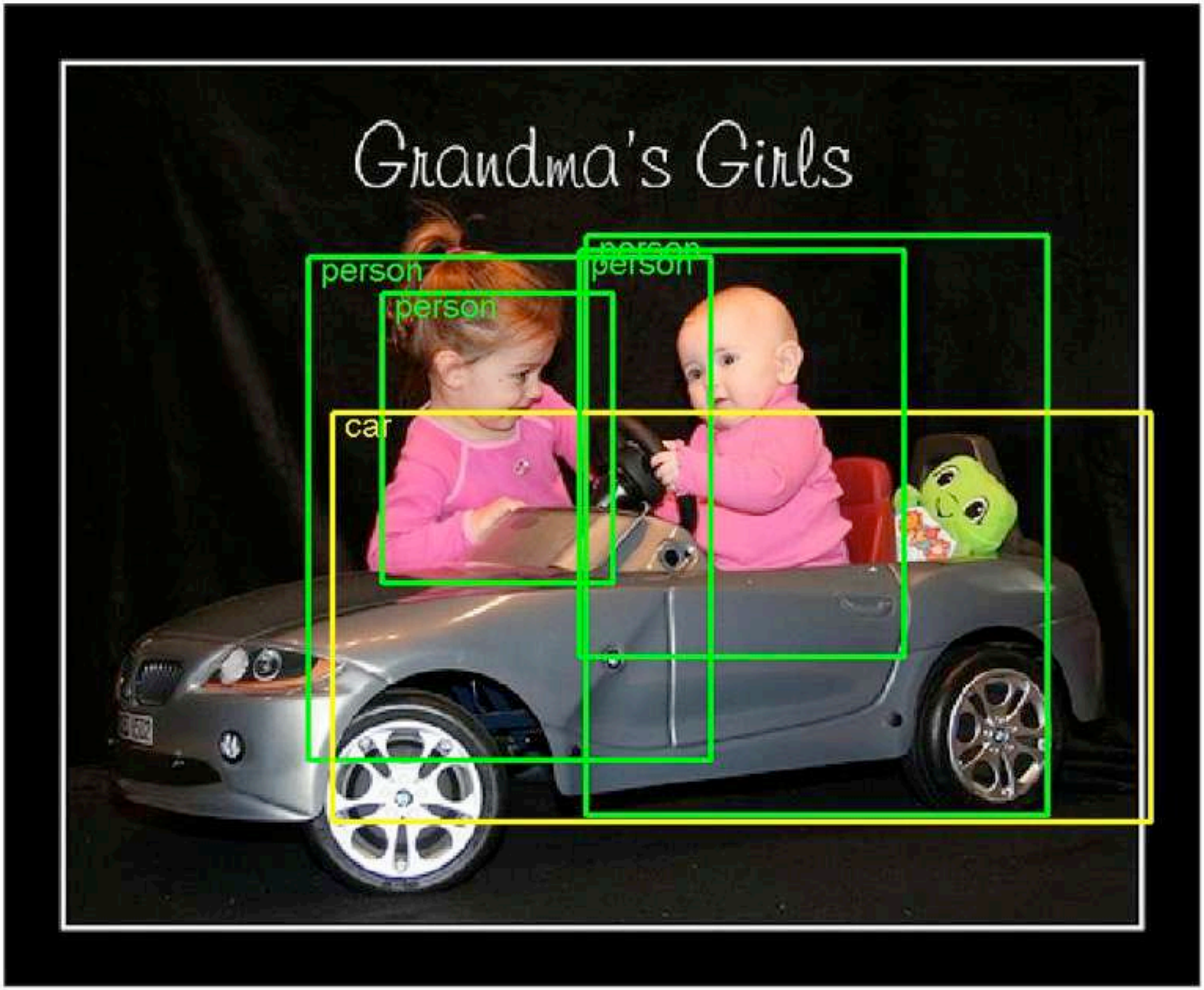
Horse part-based prior



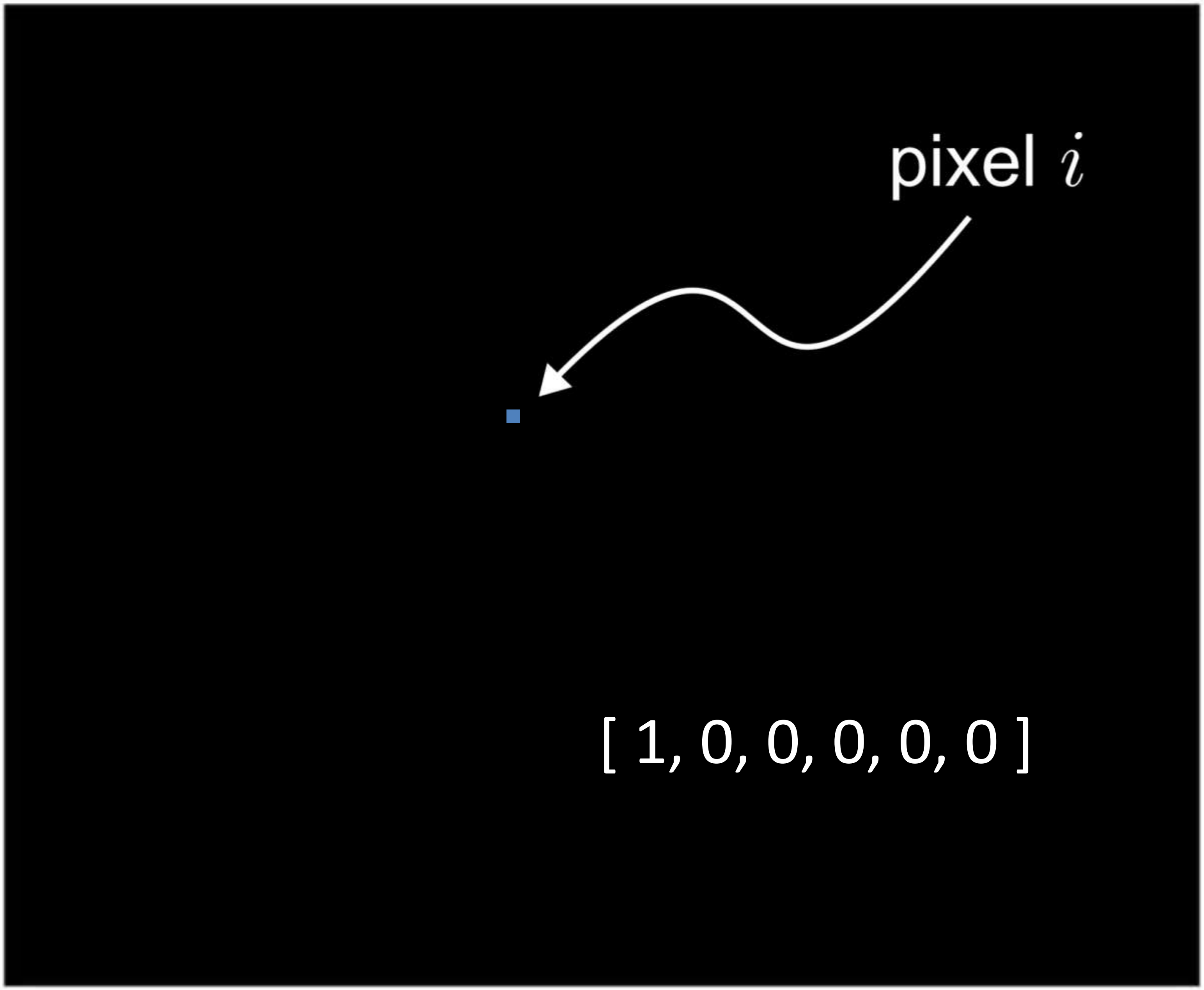
Bottle part-based priors



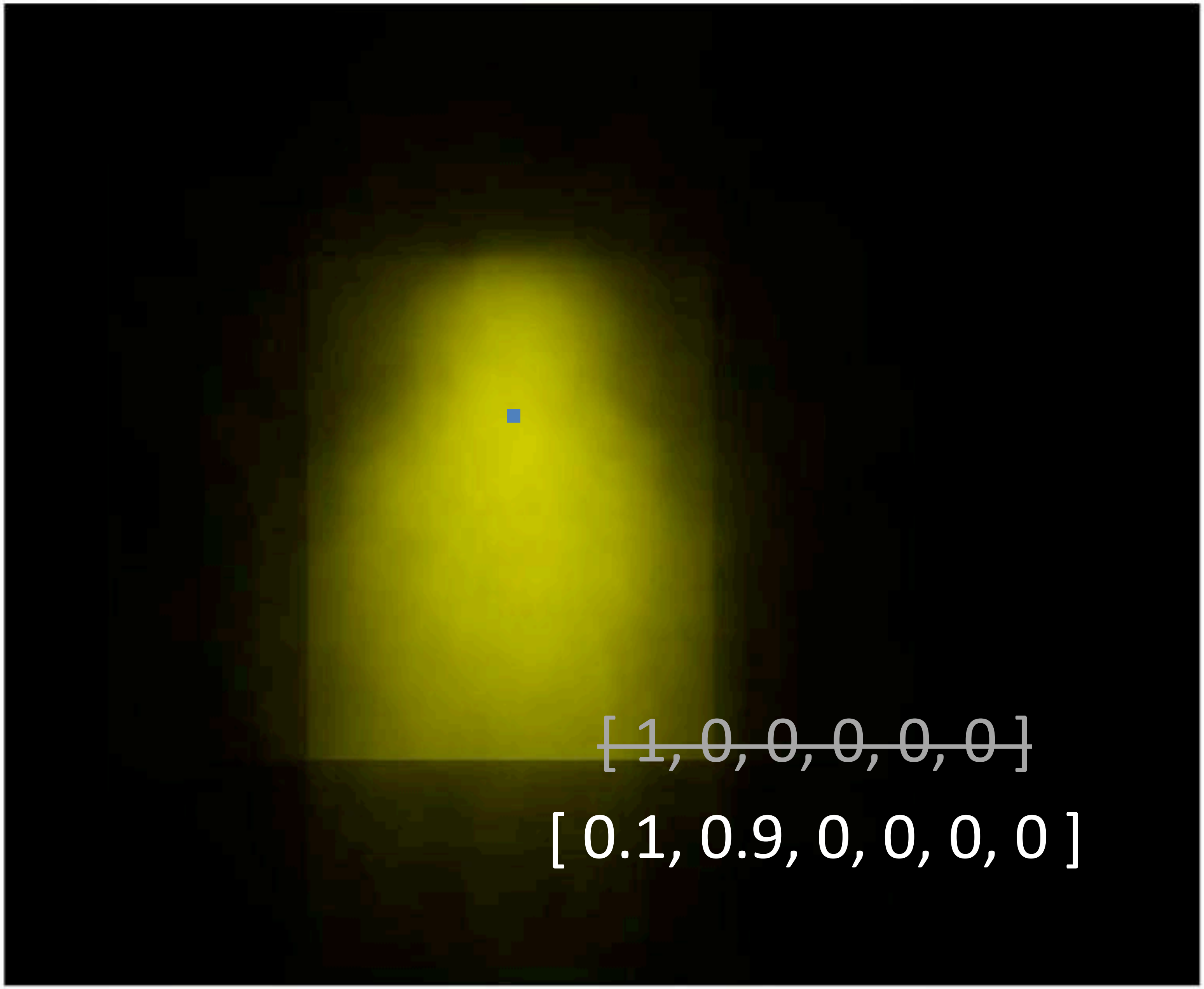
Building $P(z_i | d_\pi)$



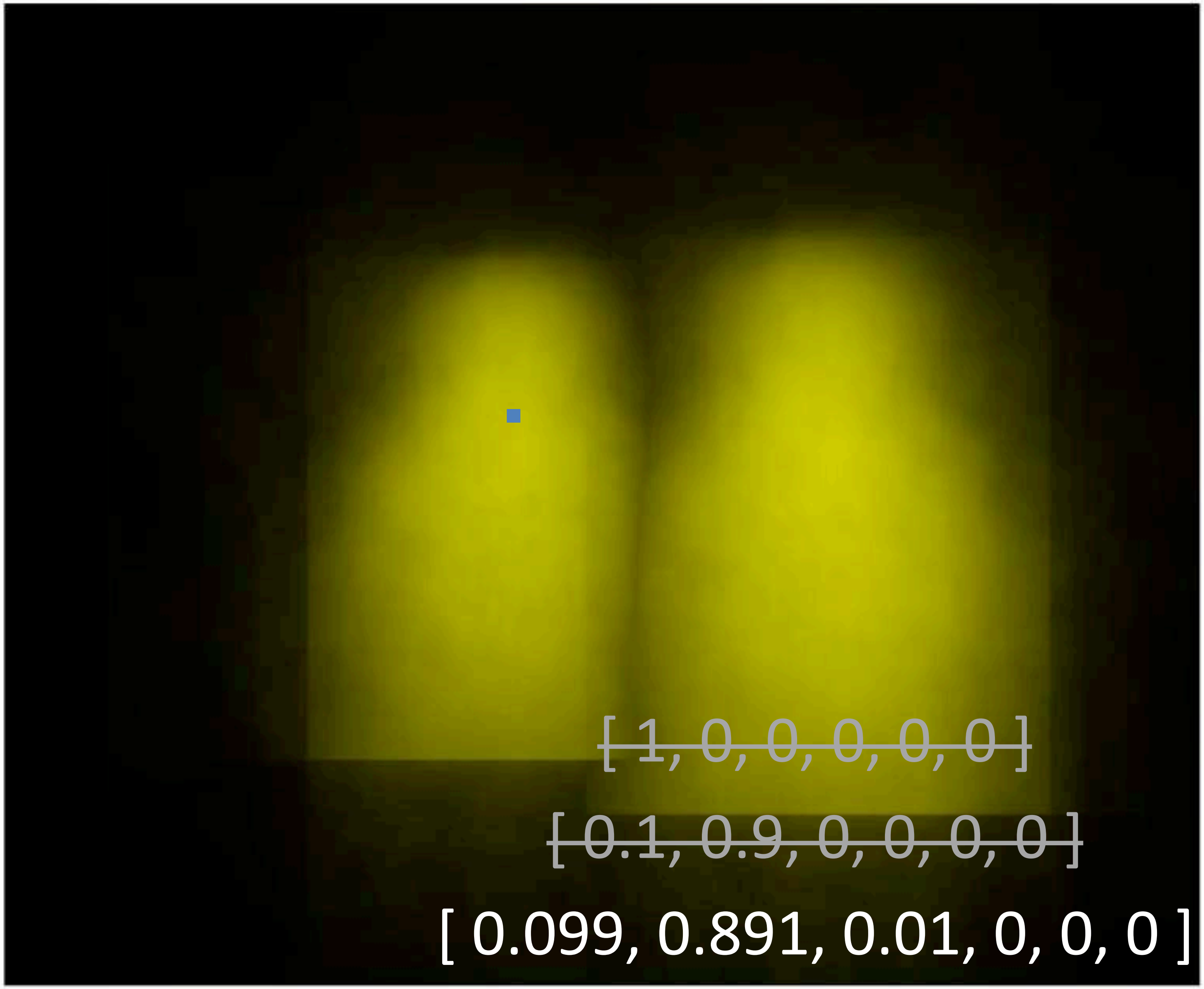
Building $P(z_i | d_\pi)$



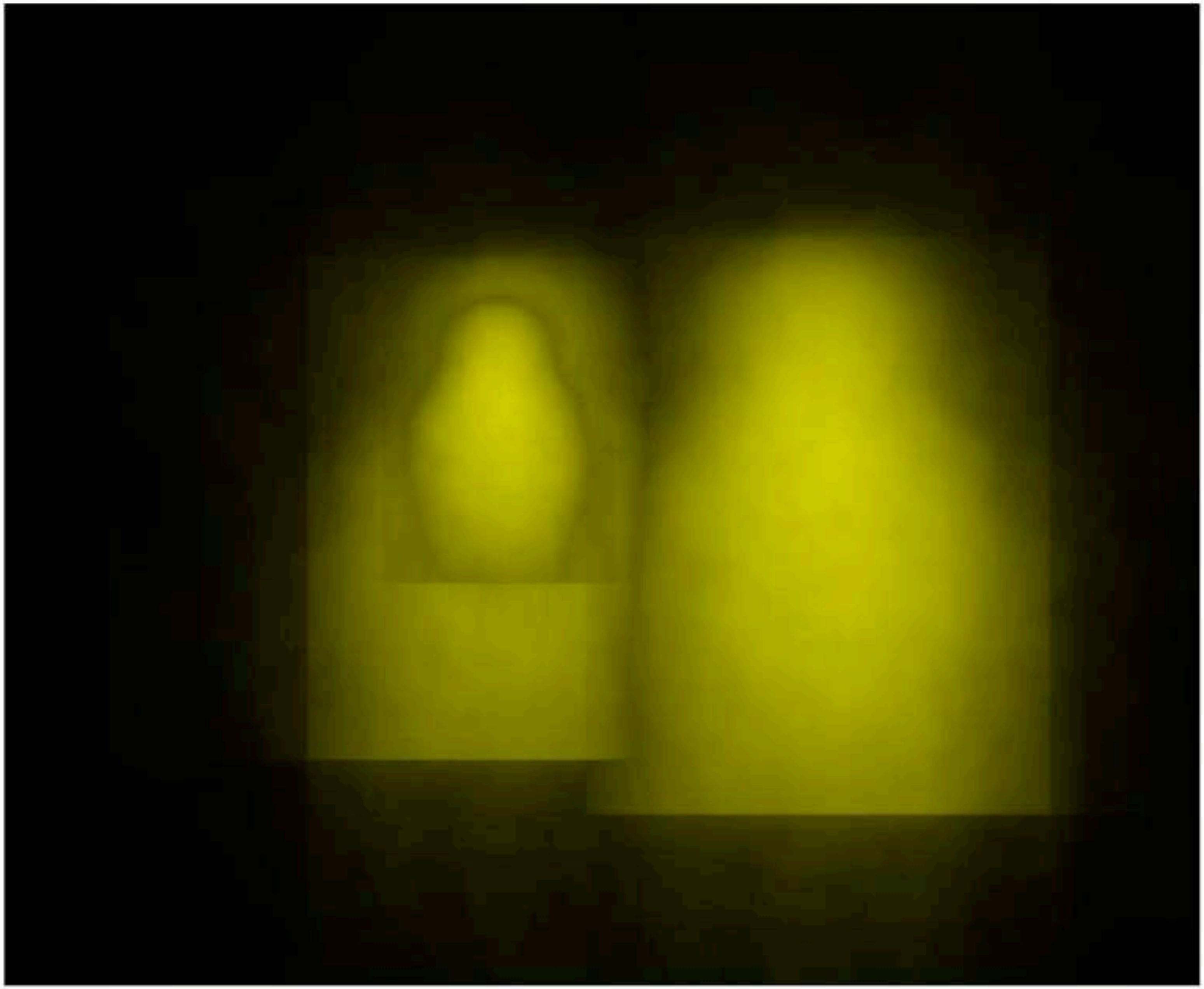
Building $P(z_i | d_\pi)$



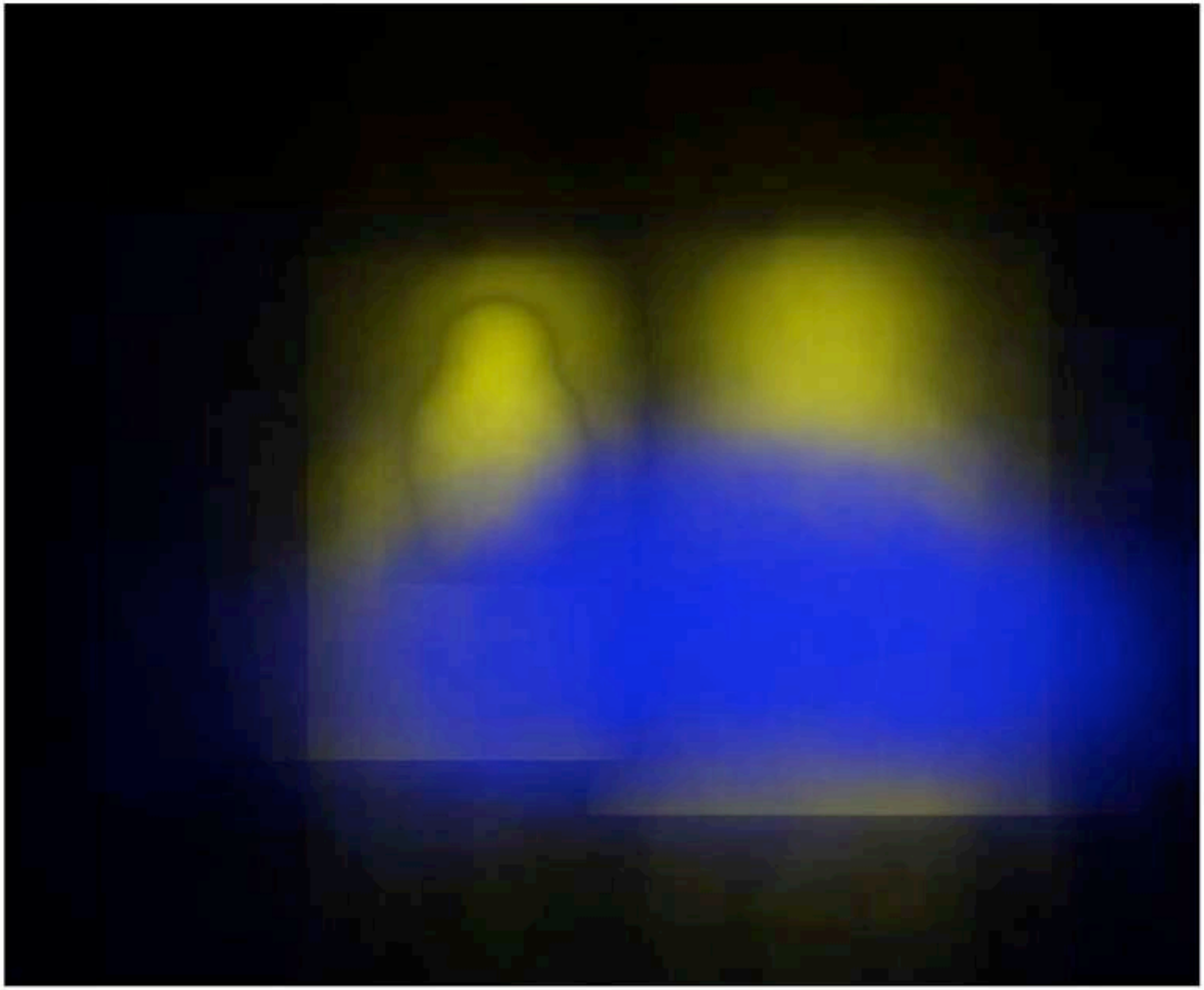
Building $P(z_i | d_\pi)$



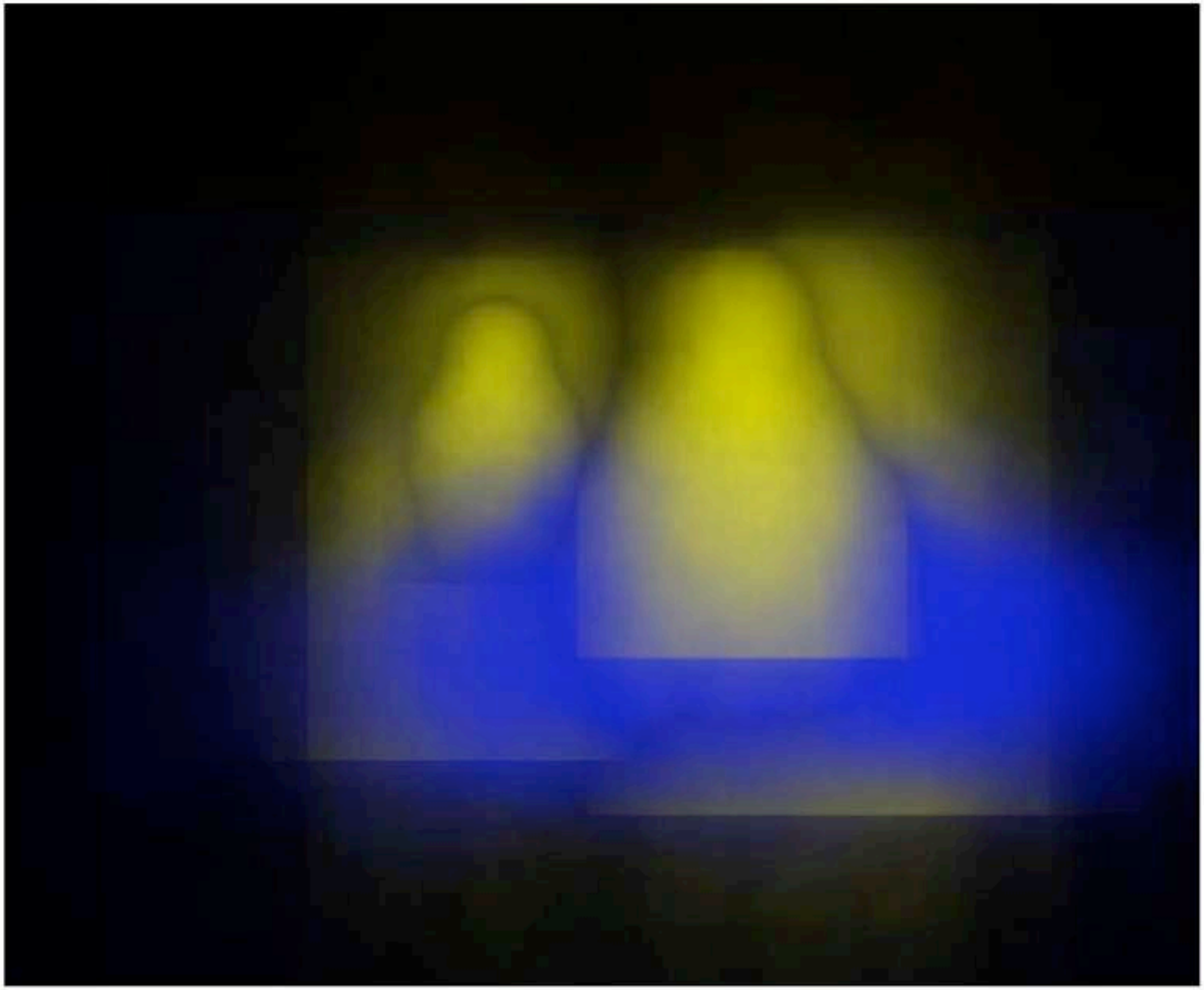
Building $P(z_i | d_\pi)$



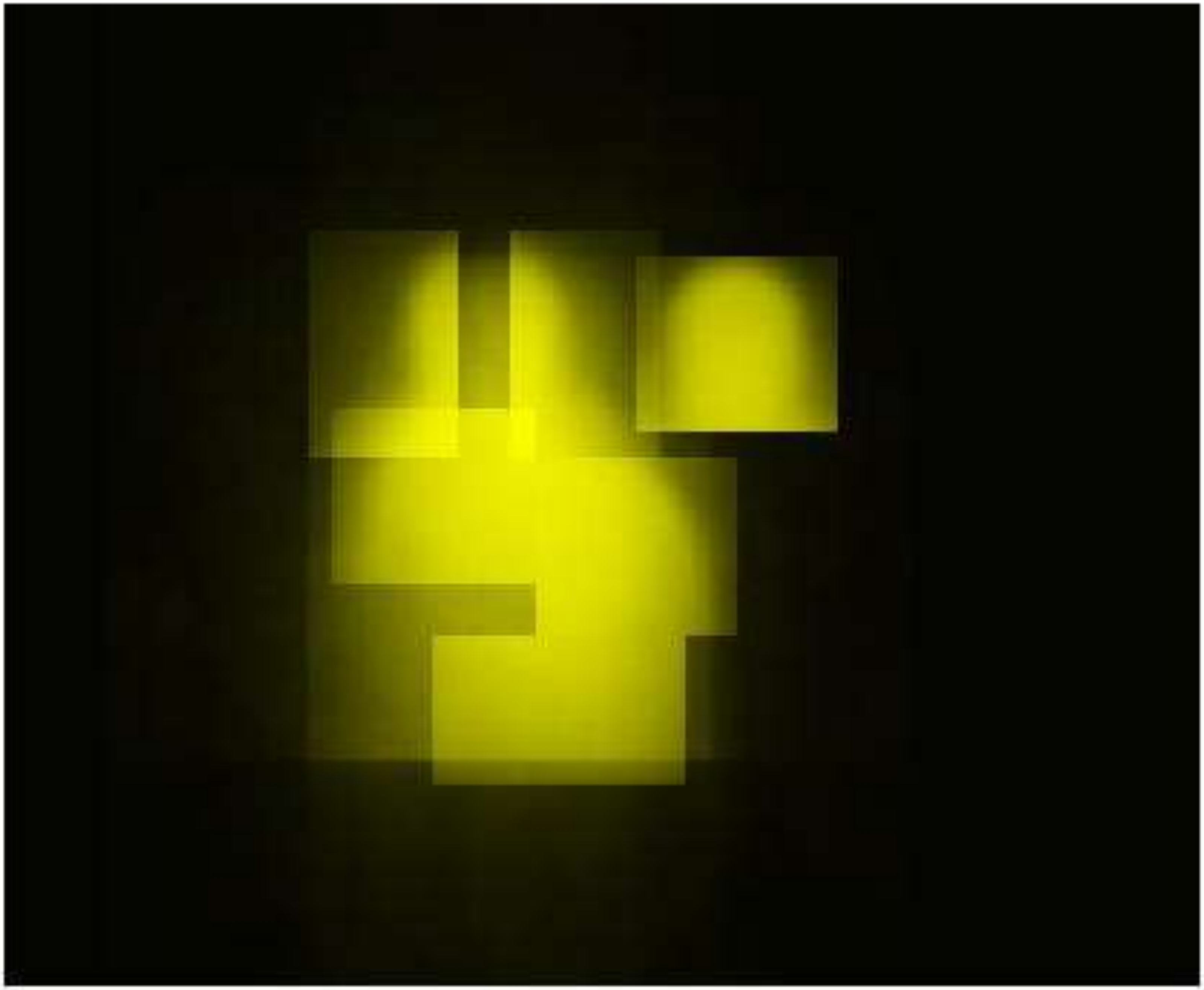
Building $P(z_i | d_\pi)$



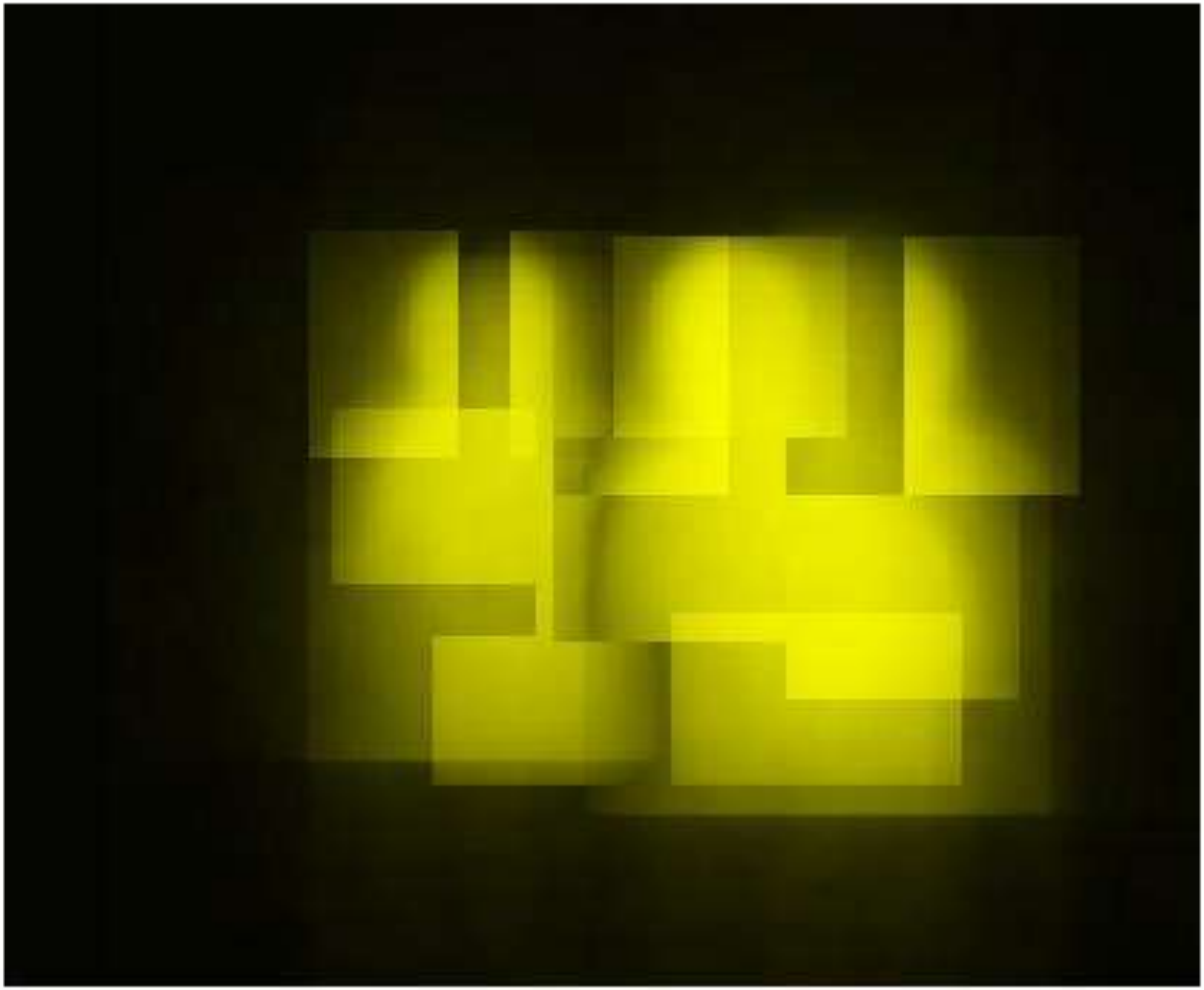
Building $P(z_i | d_\pi)$



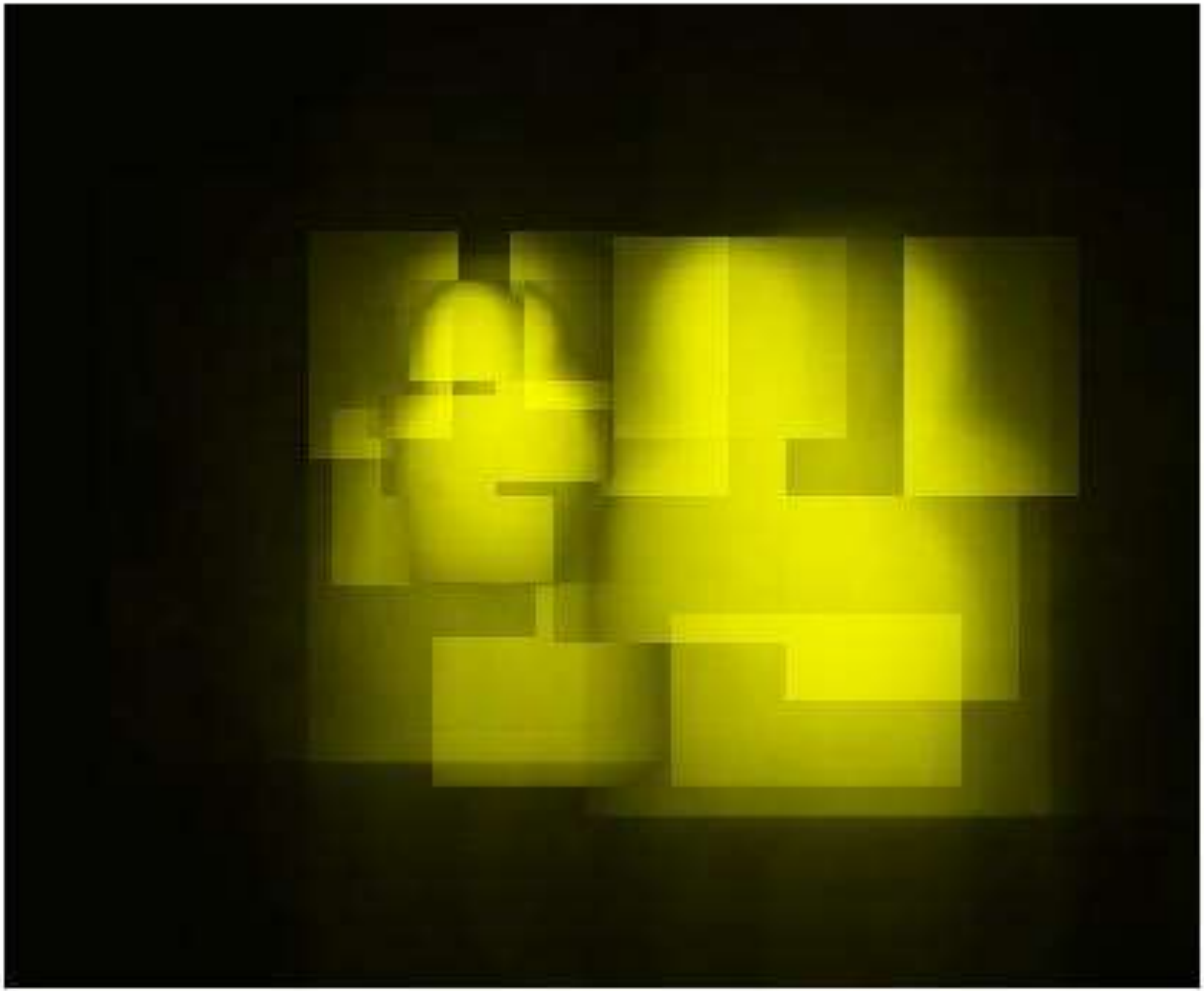
Building $P(z_i | d_\pi)$



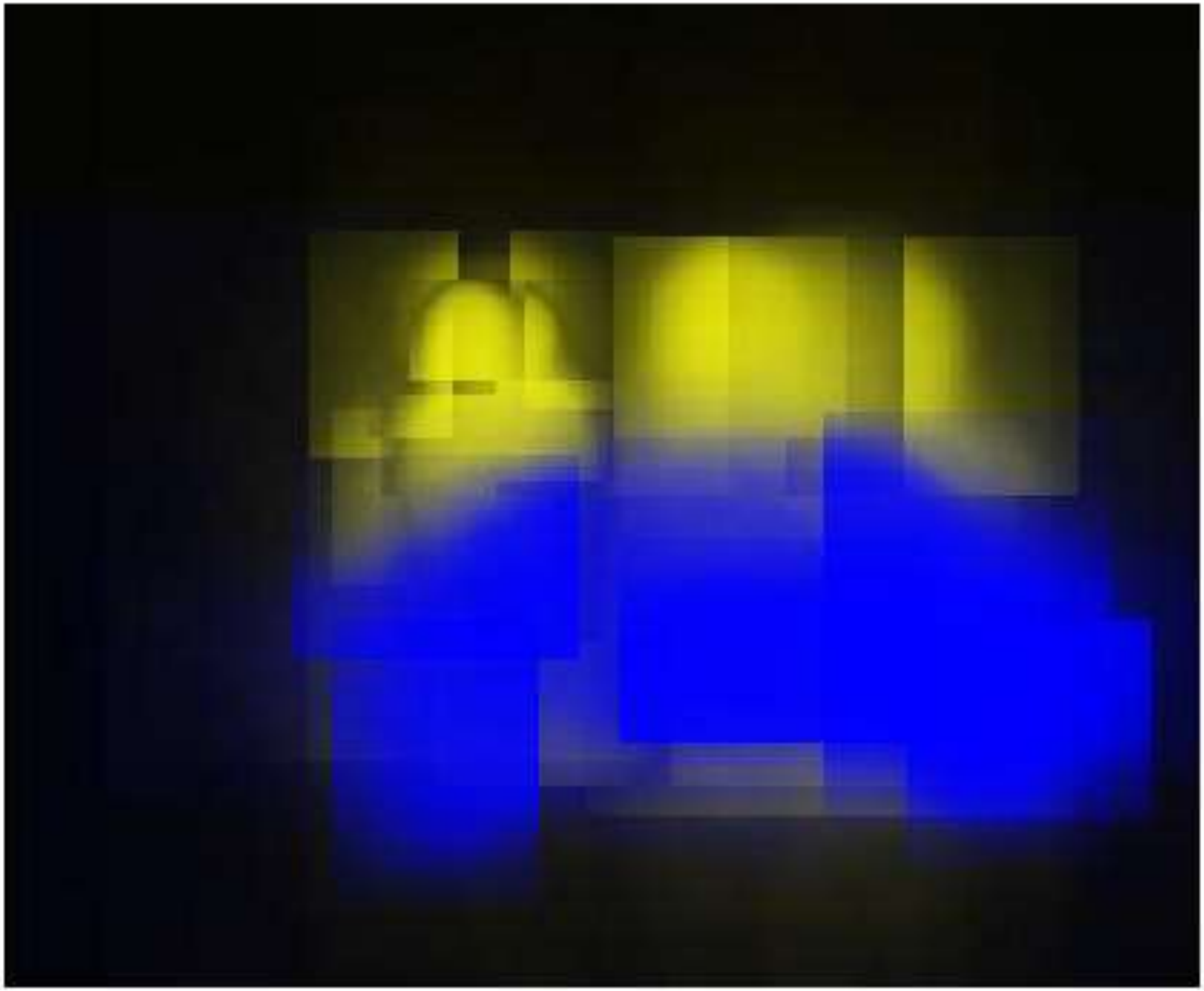
Building $P(z_i | d_\pi)$



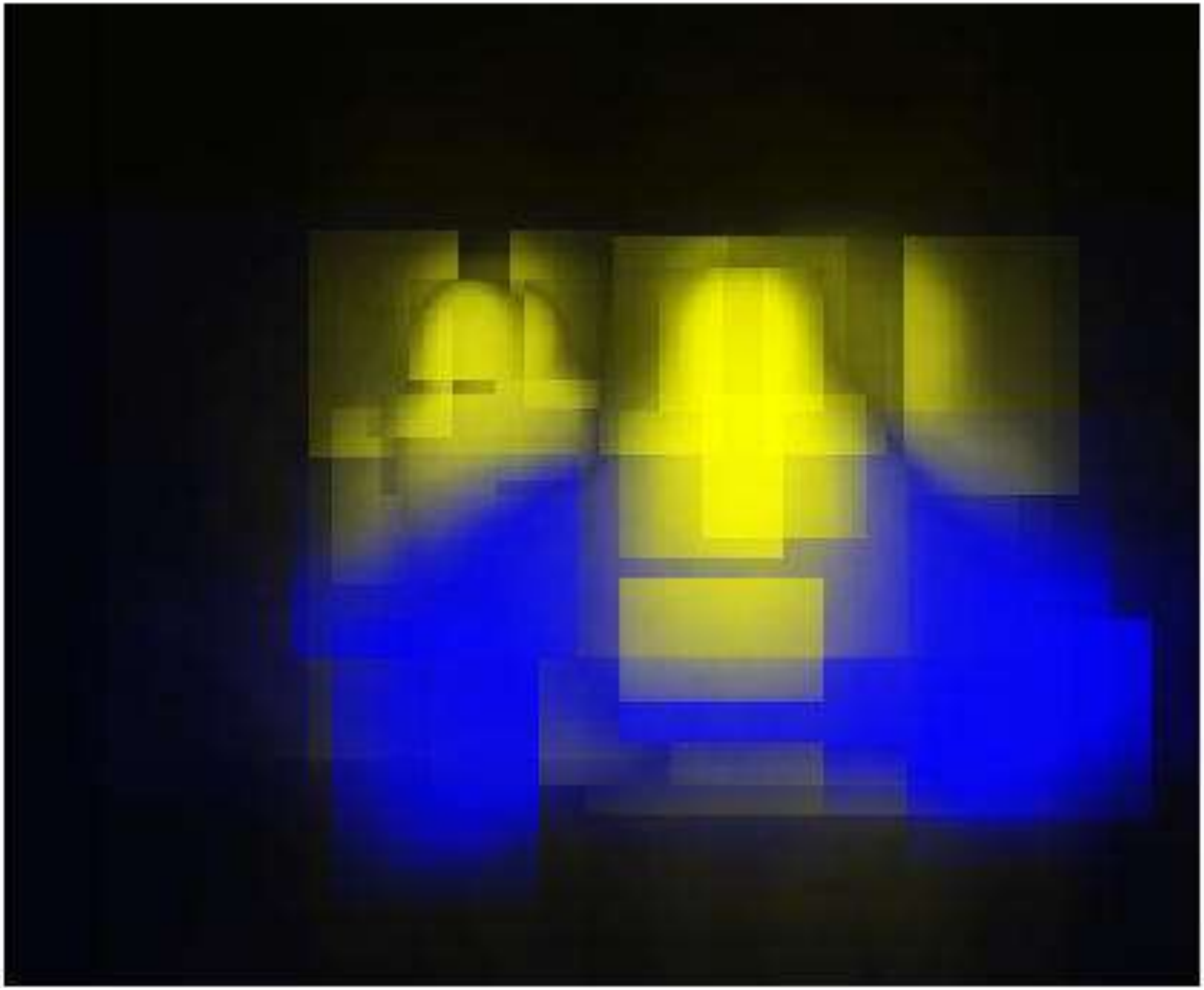
Building $P(z_i | d_\pi)$



Building $P(z_i | d_\pi)$



Building $P(z_i | d_\pi)$



The algorithm

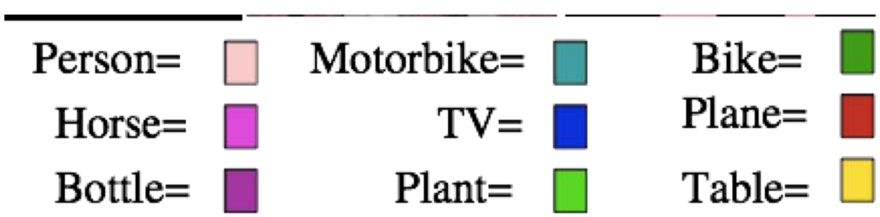
for each ordering π :

$\quad$ find $\arg\max_z P(z, x|\theta, d_\pi)$

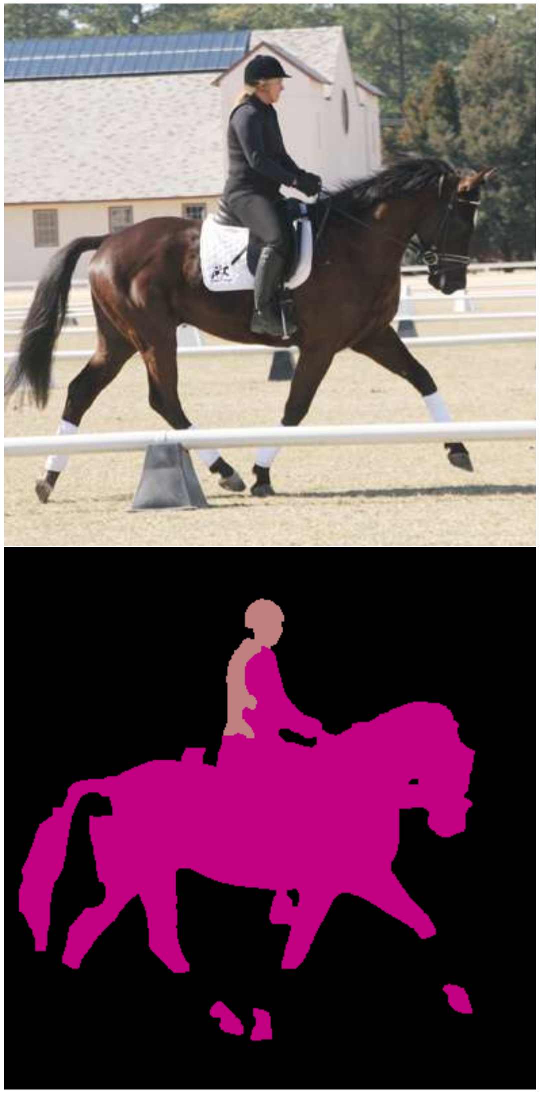
output superpixel labels with most probable π

- There is no secret method for finding π…
  Just try all of them!
- Luckily, the number of permutations to test is usually small.

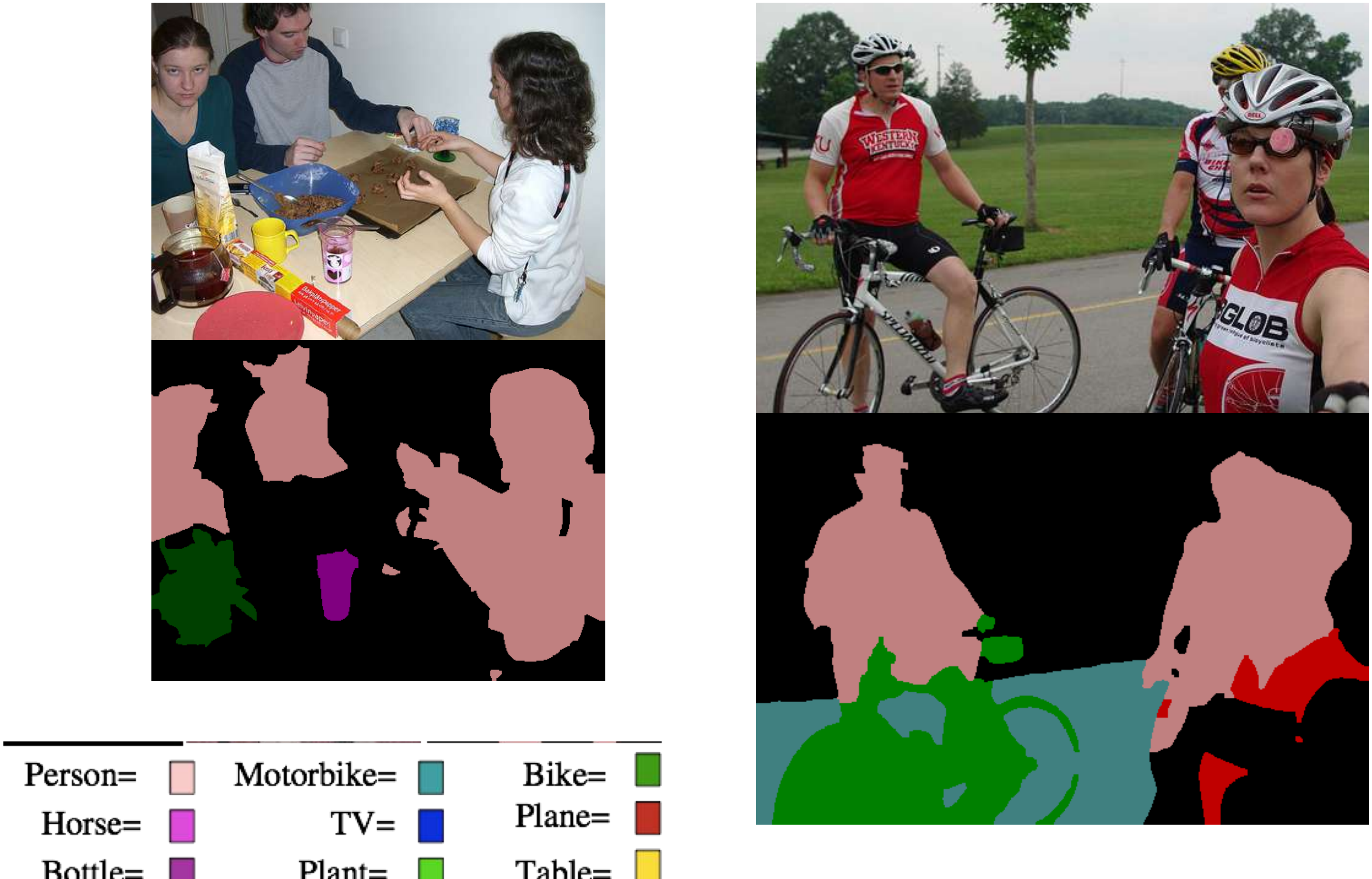
Results!


Person=
Motorbike=
Bike=
Horse=
TV=
Plane=
Bottle=
Plant=
Table=

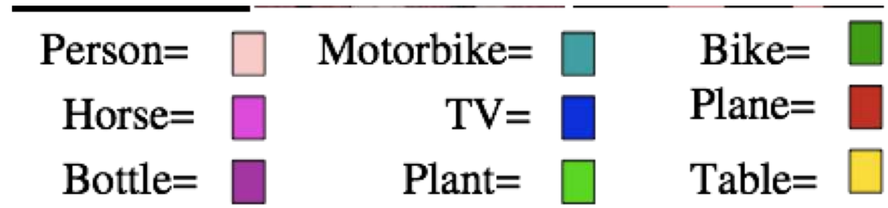

BIG BUCK
HUNTER
Pro



Person=	Motorbike=	Bike=
Horse=	TV=	Plane=
Bottle=	Plant=	Table=


Person=
Motorbike=
Bike=
Horse=
TV=
Plane=
Bottle=
Plant=
Table=


Schwinn Built
SCHWINN


Person=
Motorbike=
Bike=
Horse=
TV=
Plane=
Bottle=
Plant=
Table=

PASCAL 2009 Segmentation Challenge

	Mean	Max	Us	Our rank
background	41.2	83.5	78.0	8
aeroplane	18.8	56.3	32.8	7
bicycle	10.4	26.6	29.4	1
bird	11.0	40.6	3.2	17
boat	11.5	36.1	5.0	16
bottle	18.2	46.1	33.1	3
bus	25.5	50.5	43.4	3
car	20.6	42.3	43.8	1
cat	12.6	35.3	8.3	12
chair	4.2	9.1	5.1	9
cow	11.7	33.1	11.9	9
diningtable	9.1	27.0	8.2	11
dog	9.1	24.5	5.6	14
horse	17.5	42.7	21.0	7
motorbike	23.4	56.4	24.4	9
person	20.9	37.5	38.6	1
pottedplant	9.7	37.1	14.6	6
sheep	19.7	43.6	14.8	13
sofa	8.5	21.9	3.5	17
train	19.2	41.0	27.5	7
tv/monitor	22.3	47.8	45.7	2
average	16.4	36.2	23.7	7

Performs well where detector works well (objects with well defined, fairly rigid shapes)

What aspects of the model are useful?

	¬ordering	¬color	¬superpixel	¬parts	all
background	79.37	78.93	78.65	**79.62**	79.36
aeroplane	35.26	32.39	30.61	**37.22**	35.26
bicycle	**25.46**	23.12	20.7	24.58	25.45
bird	**2.81**	2.78	2.68	2.79	**2.81**
boat	**9.87**	9.16	9.64	9.14	**9.87**
bottle	41.44	39.73	**41.76**	40.19	41.29
bus	49.83	48.52	48.54	48.72	**49.87**
car	46.88	45.66	44.25	46.14	**47.03**
cat	**18.4**	17.68	16.81	15.06	**18.4**
chair	**10.05**	9.06	9.57	8.37	10
cow	17.74	16.83	**18.1**	15.91	17.77
diningtable	6.94	6.8	6.79	6.85	**7.27**
dog	**11.53**	10.55	11.18	10.91	**11.53**
horse	16.07	14.6	15.33	15.19	**16.21**
motorbike	**25.72**	24.38	24.46	24.88	25.62
person	**36.88**	34.98	35.3	32.4	36.81
pottedplant	**15.55**	14.92	15.17	14.32	**15.55**
sheep	21.09	18.77	20.33	17.77	**21.1**
sofa	**12.63**	12.2	12.14	12.05	**12.63**
train	**28.6**	27.43	27.88	27.86	**28.6**
tvmonitor	**46.41**	46.01	46.36	43.67	46.28
average	**26.6**	25.45	25.53	25.41	**26.6**

Superpixels really help bikes

Color helps person segmentation

Parts provide small but noticeable improvement

Does ordering help?

- Default ordering based on detector score works fairly well (after calibration)
- PASCAL segmentation benchmark limitations:
  - images often only have a single object
  - scoring is per-class rather than per-instance
    - Ordering between same class detections doesn't affect benchmark score
- We found ordering helped on a subset of PASCAL with overlapping objects

Conclusions

- A simple layered model for compiling multi-object detections into segmentations
  - Deformable shape prior built on part-based detector
  - Per-instance color model
- Provides a globally consistent 2.1D interpretation
- Good performance on PASCAL segmentation benchmark

Thanks!