Part-Based Layered Shape Models for Segmentation

Charless Fowlkes, Sam Hallman, Deva Ramanan, Yi Yang

UC Irvine

Overall approach

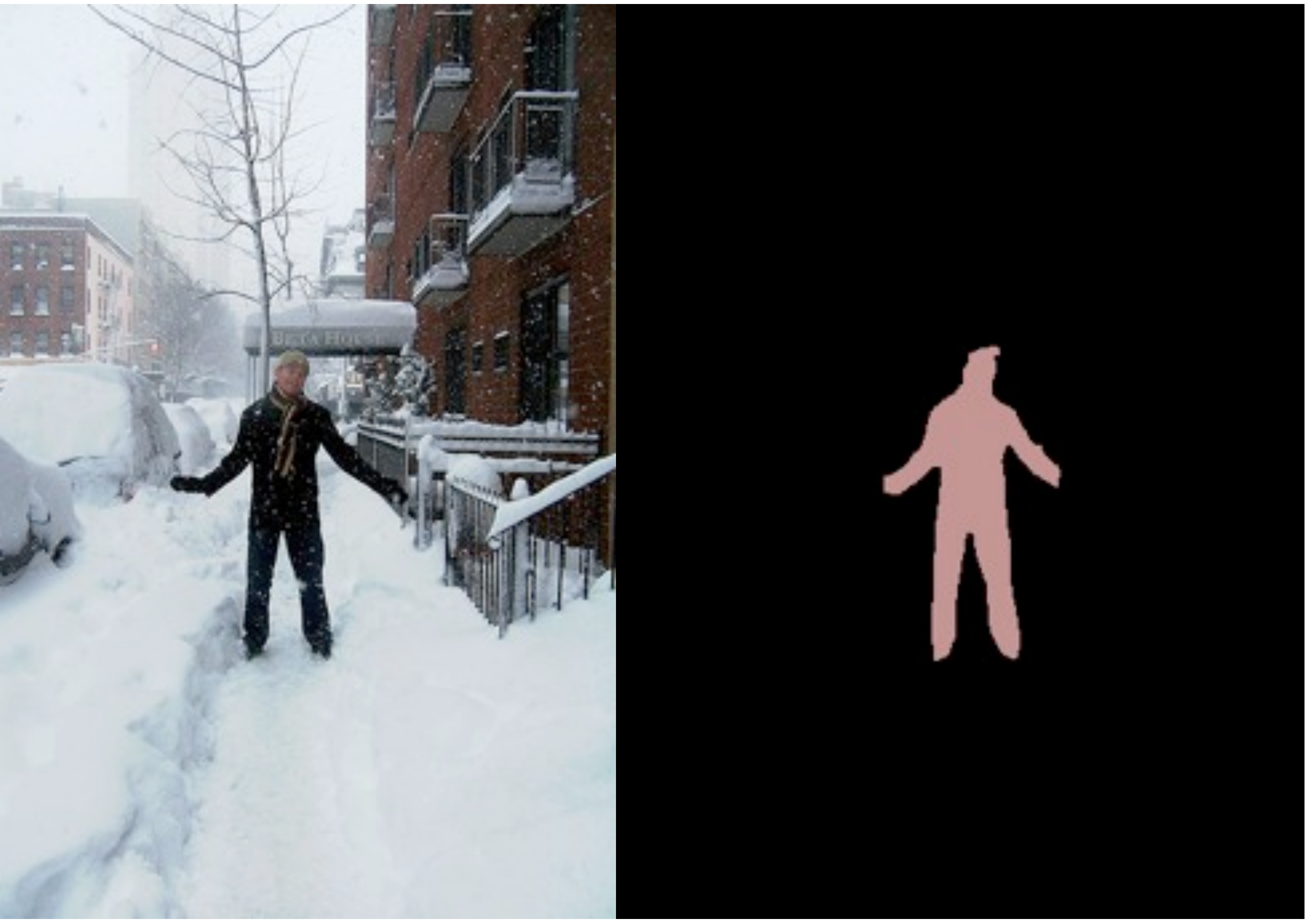
1) Detect putative bounding boxes
with scanning-window detector

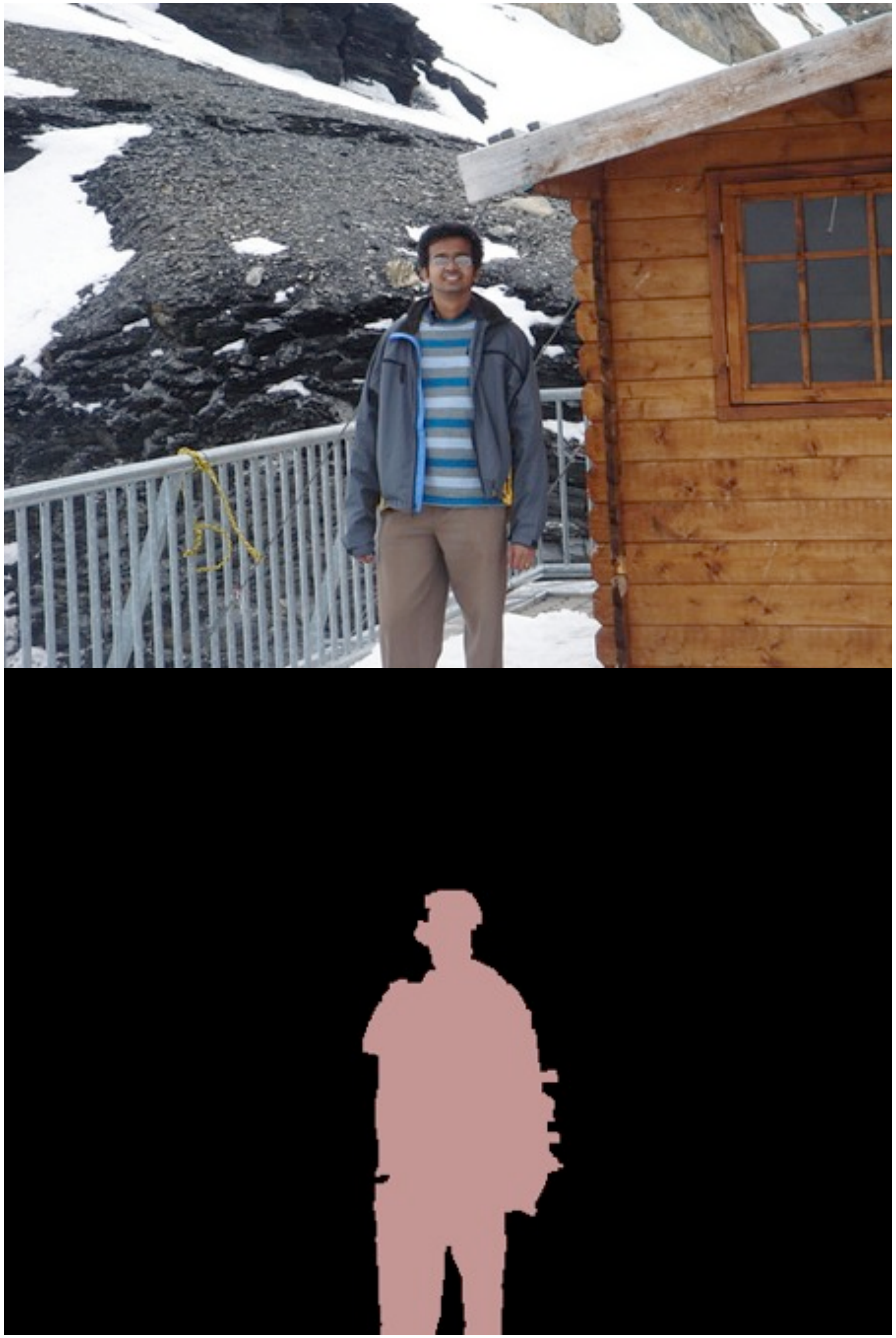
2) Integrate top-down shape priors
with bottom-up superpixels to
segment image

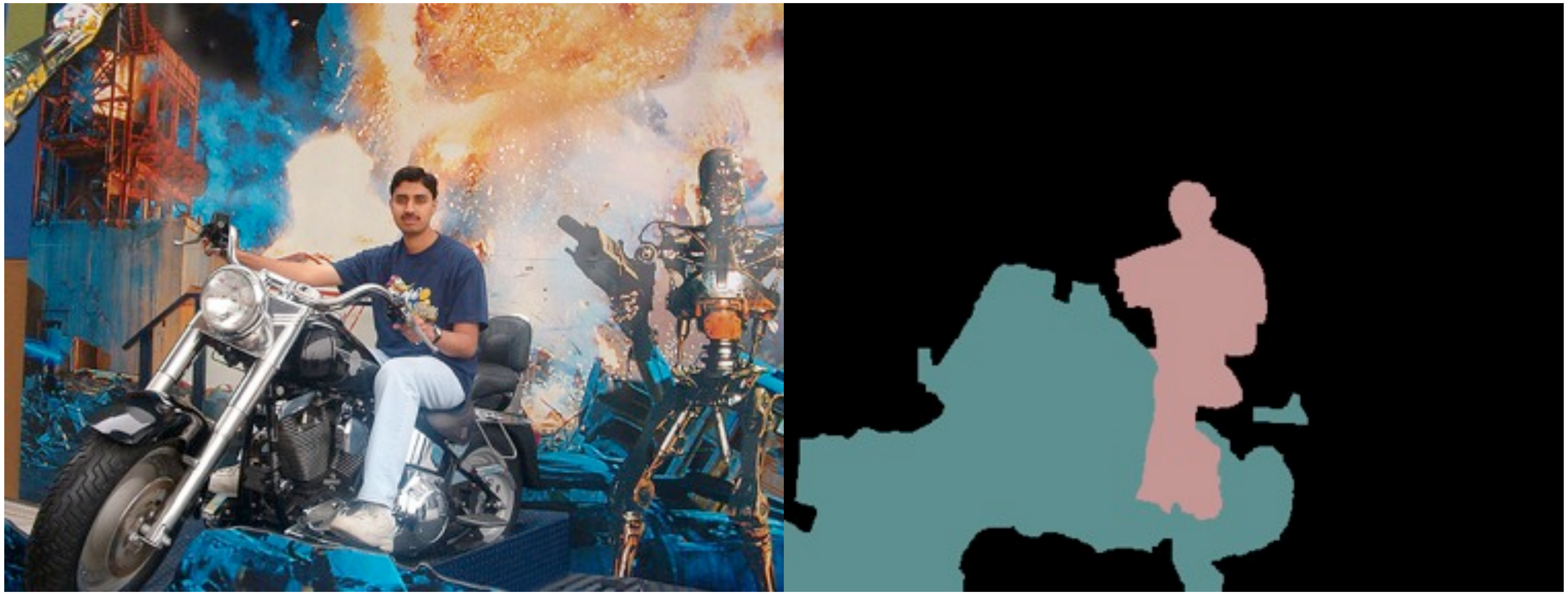
We use part-based detectors from
Felzenszwalb, Gorelick, McAllester, & Ramanan PAMI 09

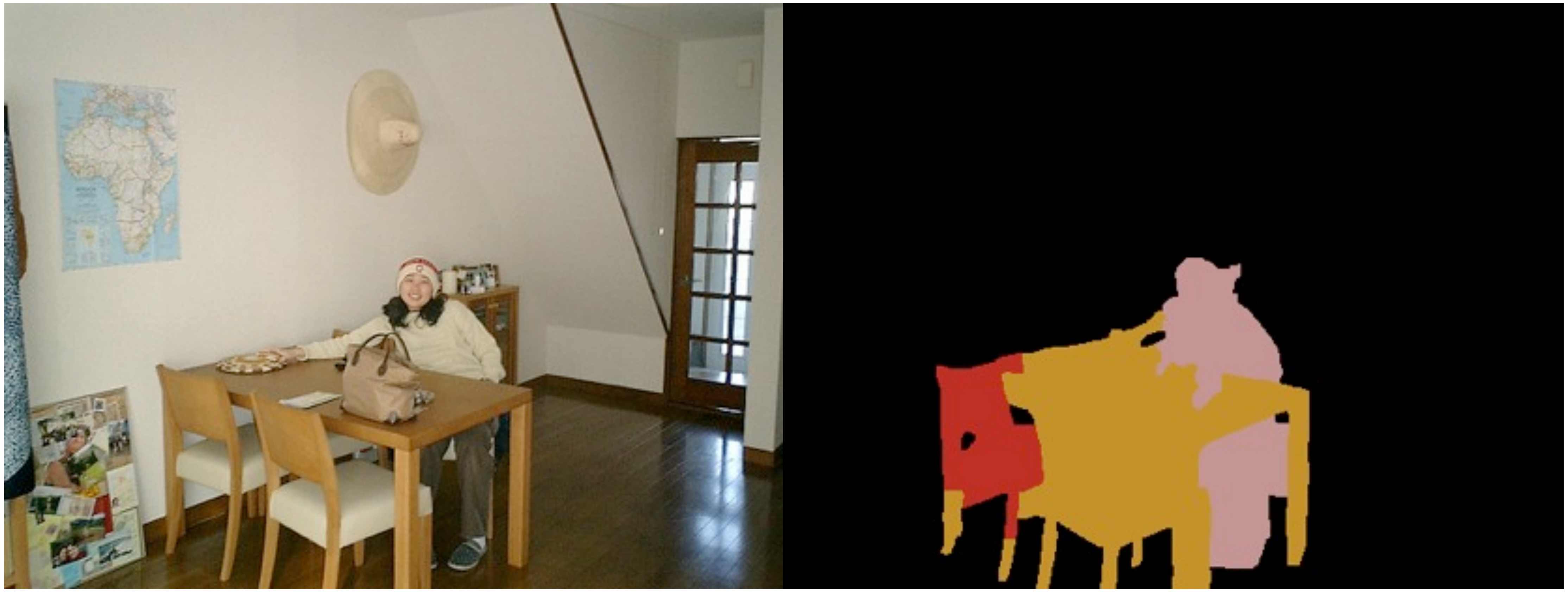
We use superpixels from
Arbelaez, Maire, Fowlkes, & Malik CVPR 09

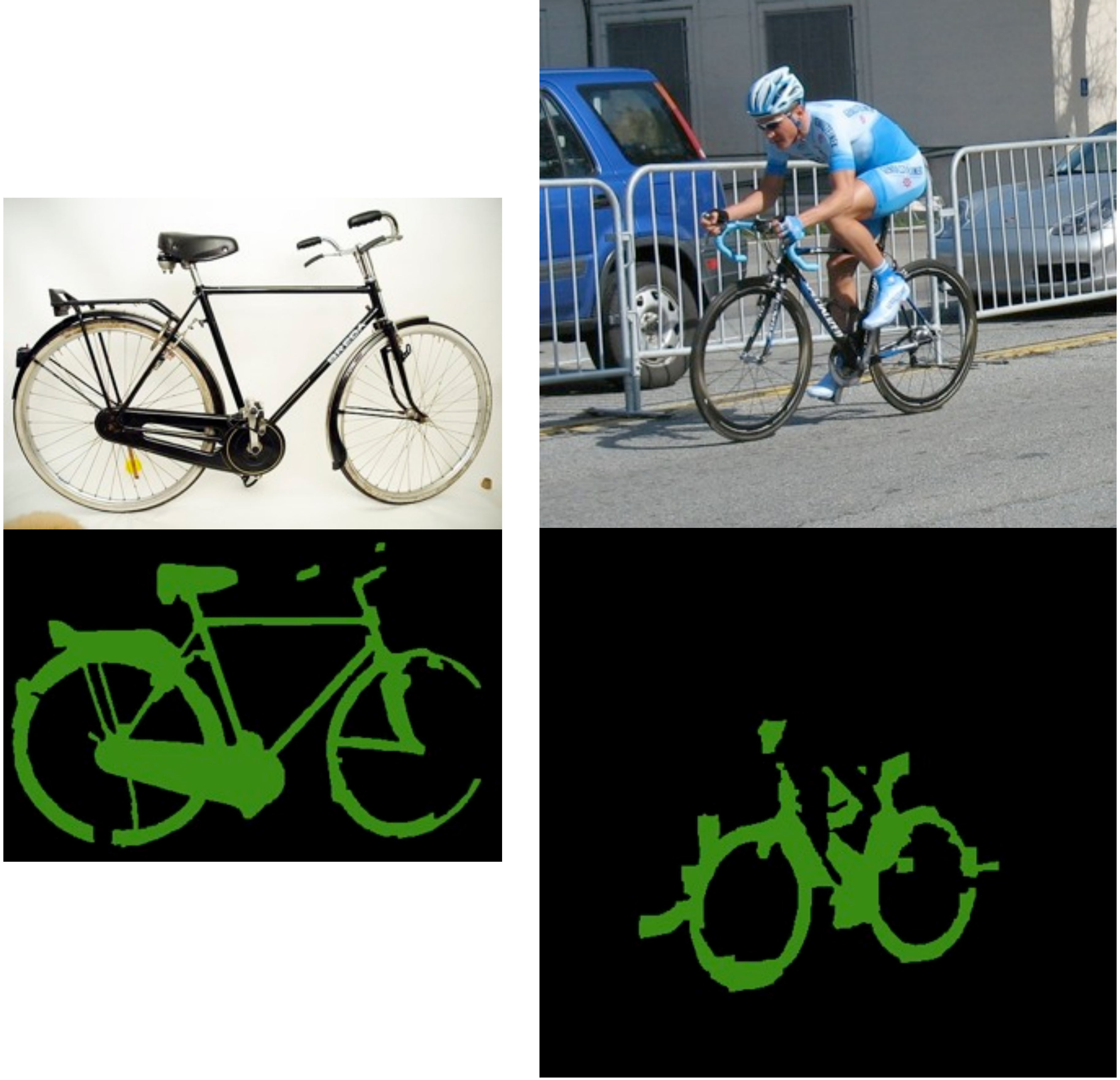
example results...





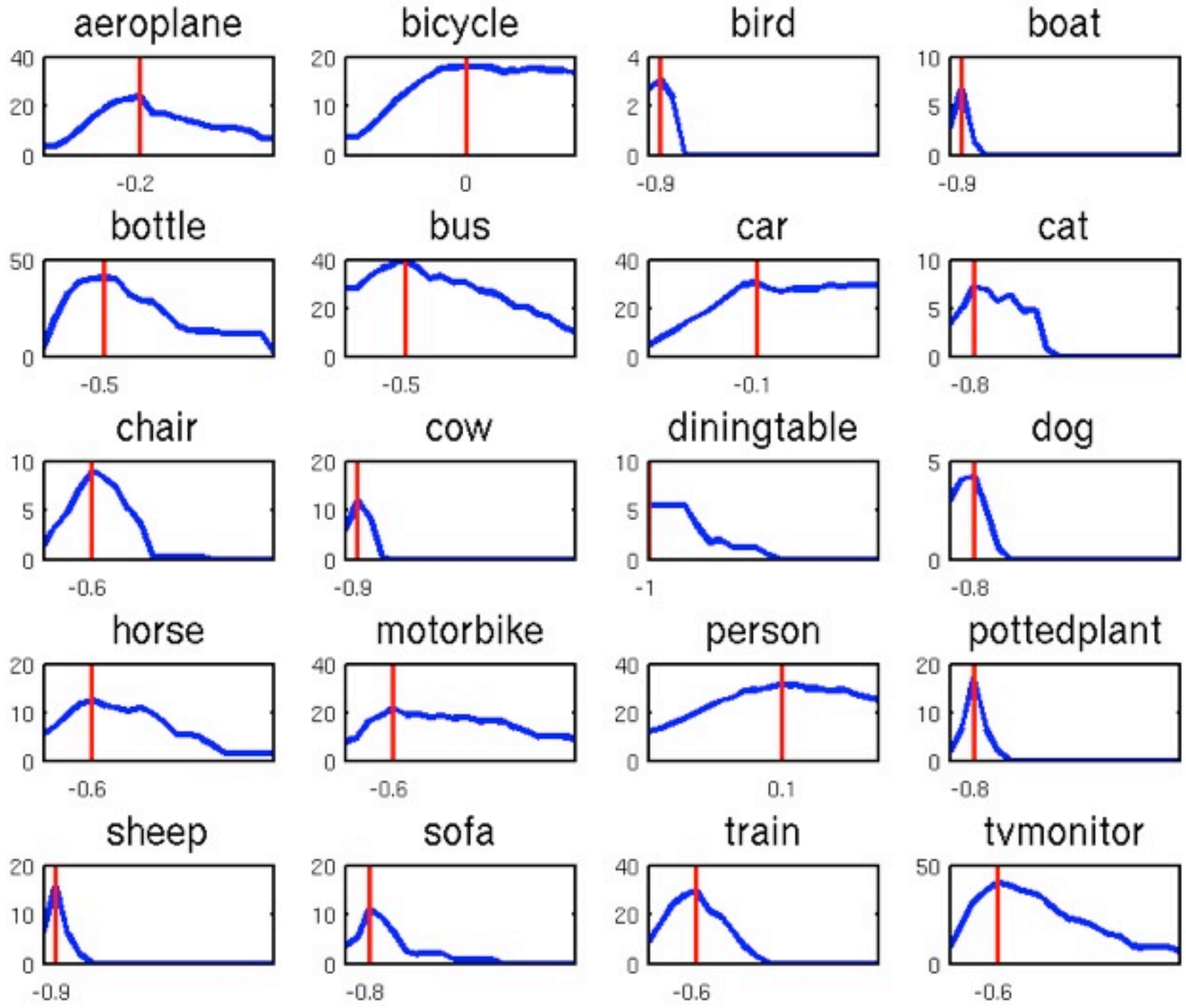






Detector calibration

Find optimal threshold for pruning detections



Model

x_i = RGB values for pixel i
z_i = label for pixel i (1..K or 0 for background)
c_n = color model of nth bounding box
D = set of bounding box detections

product
over pixels

$$P(Z, X|C, D) = \prod_{i=1}^{N} P(x_i|C_{z_i}) P(z_i|D)$$

color
likelihood

bbox
prior

Inference: coordinate descent

product over pixels

$$P(Z,X|C,D) = \underbrace{\prod_{i=1}^{N}}_{} \underbrace{P(x_i|C_{z_i})}_{\text{color likelihood}} \underbrace{P(z_i|D)}_{\text{bbox prior}}$$

color likelihood

bbox prior

$$f(C,Z) = -\log P(Z,X|C,D)$$

Step 1: $\underset{C}{\operatorname{argmin}} f(C,Z)$ (fit color models to segments and background)

Step 2: $\underset{Z}{\operatorname{argmin}} f(C,Z)$ (per-pixel segmentation given bbox prior & color likelihood)

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

Inference: coordinate descent

product
over pixels

$$P(Z, X|C, D) = \prod_{i=1}^{N} P(x_i|C_{z_i}) P(z_i|D)$$

color
likelihood

bbox
prior

$$f(C, Z) = -\log P(Z, X|C, D)$$

Step 1: $\underset{C}{\operatorname{argmin}} f(C, Z)$ (fit color models to segments)

Step 2: $\underset{Z \in \mathcal{Z}}{\operatorname{argmin}} f(C, Z)$ (per-pixel segmentation given bbox prior & color likelihood)

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

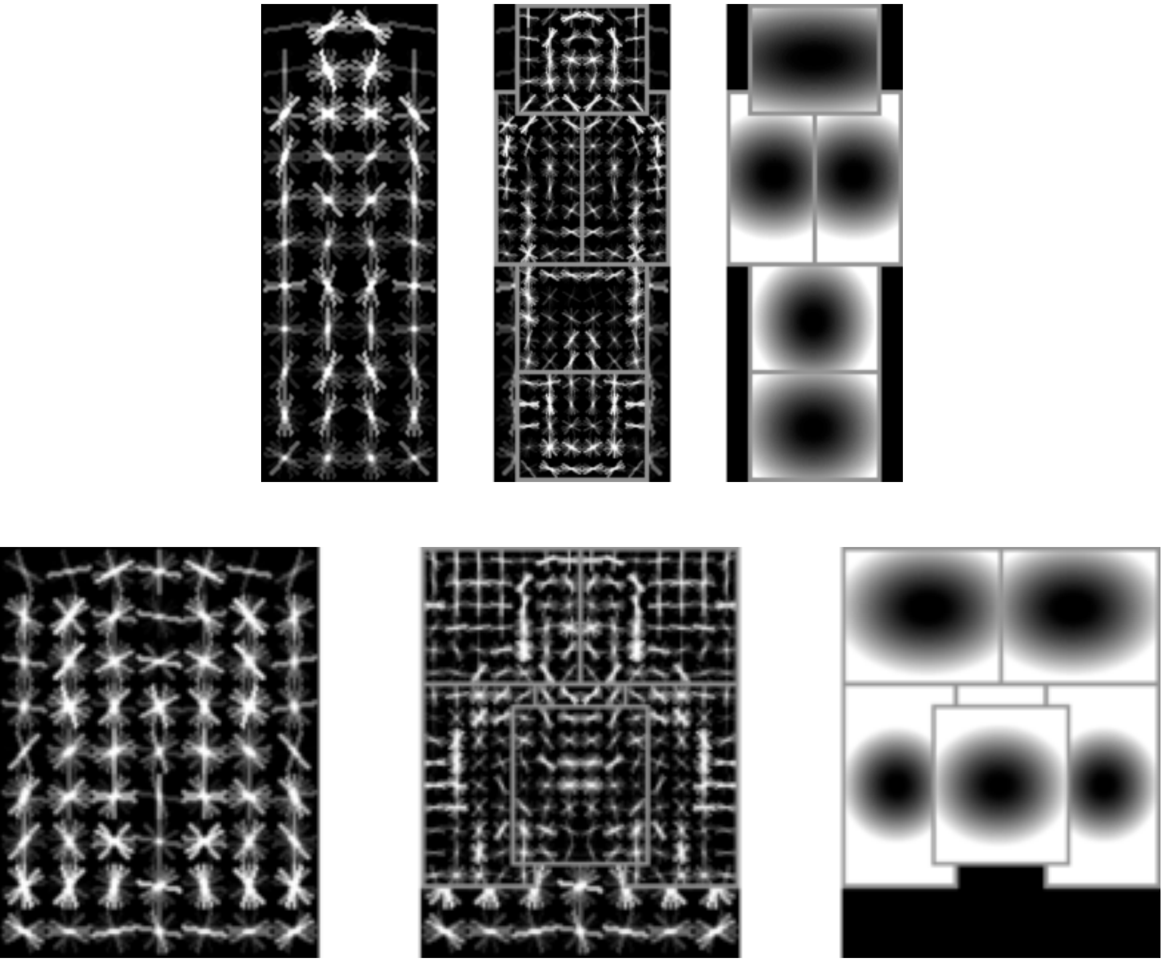
Building P(z$_i$|D)

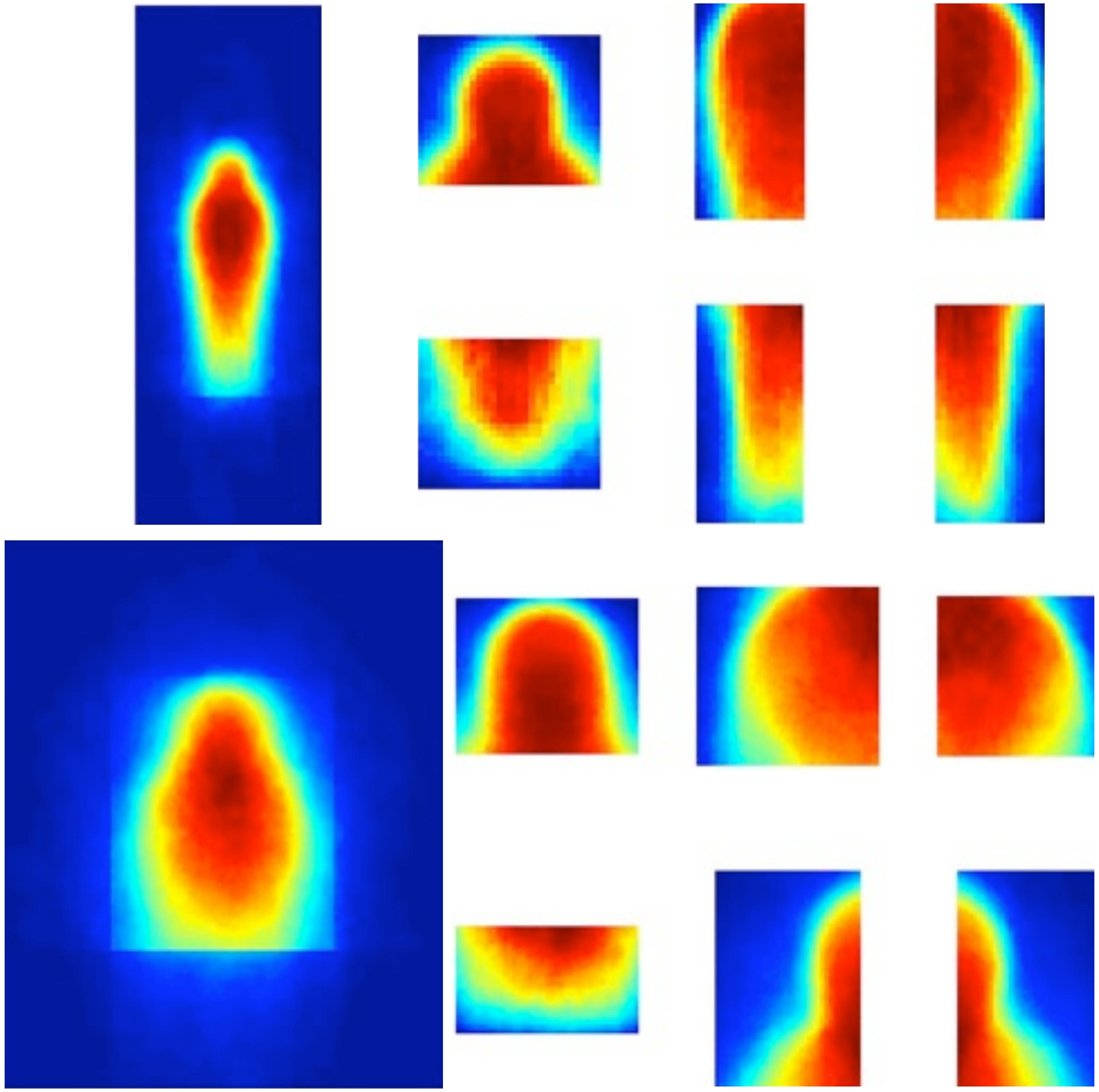
$$P(Z, X|C, D) = \overset{\text{product over pixels}}{\prod_{i=1}^{N}} \underset{\text{color likelihood}}{P(x_i|C_{z_i})} \underset{\text{bbox prior}}{P(z_i|D)}$$

Build a soft segmentation mask for each detection category

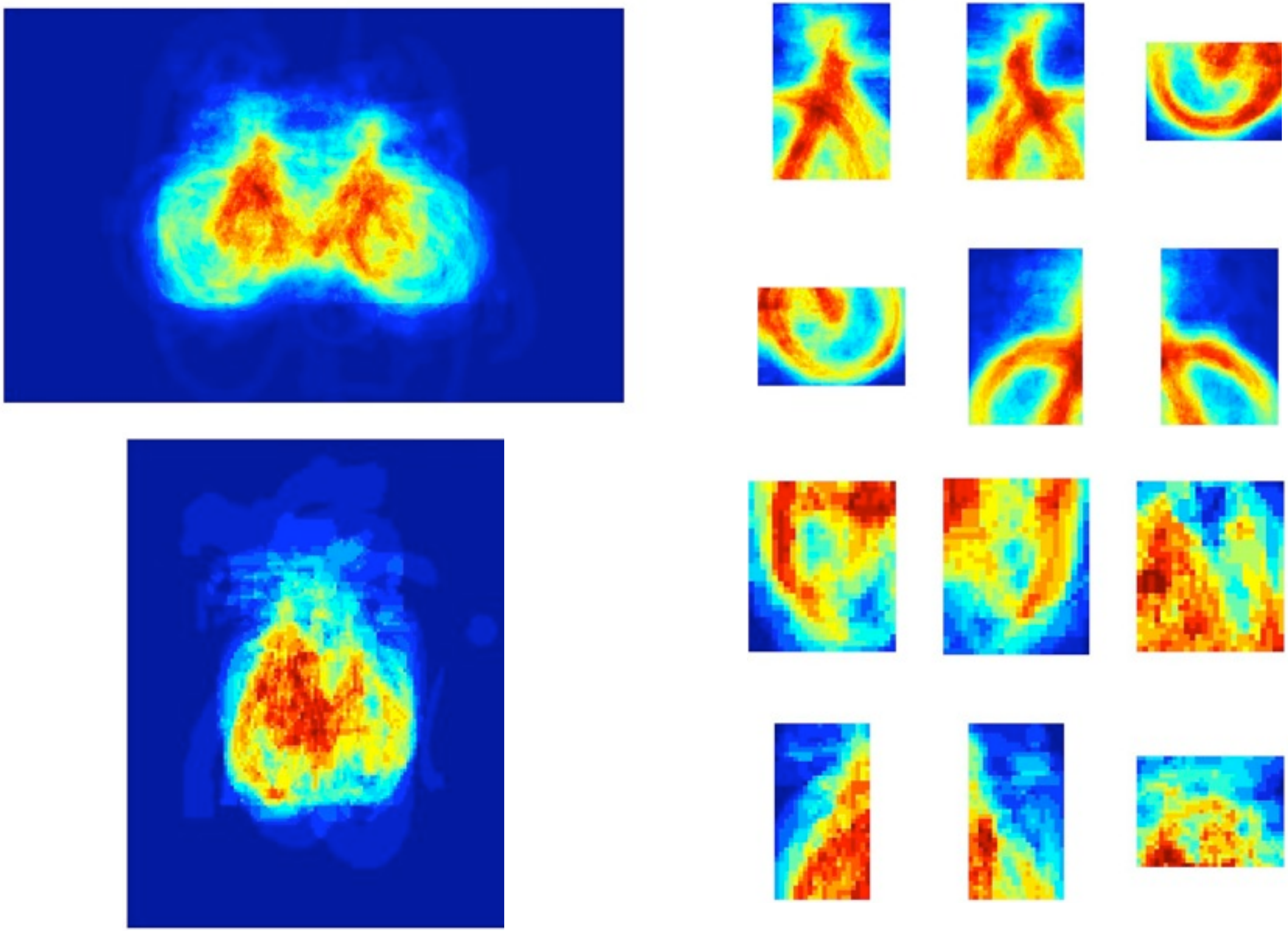
(By counting fg/bg pixels in ground truth segmentation masks)

Part-based bbox priors

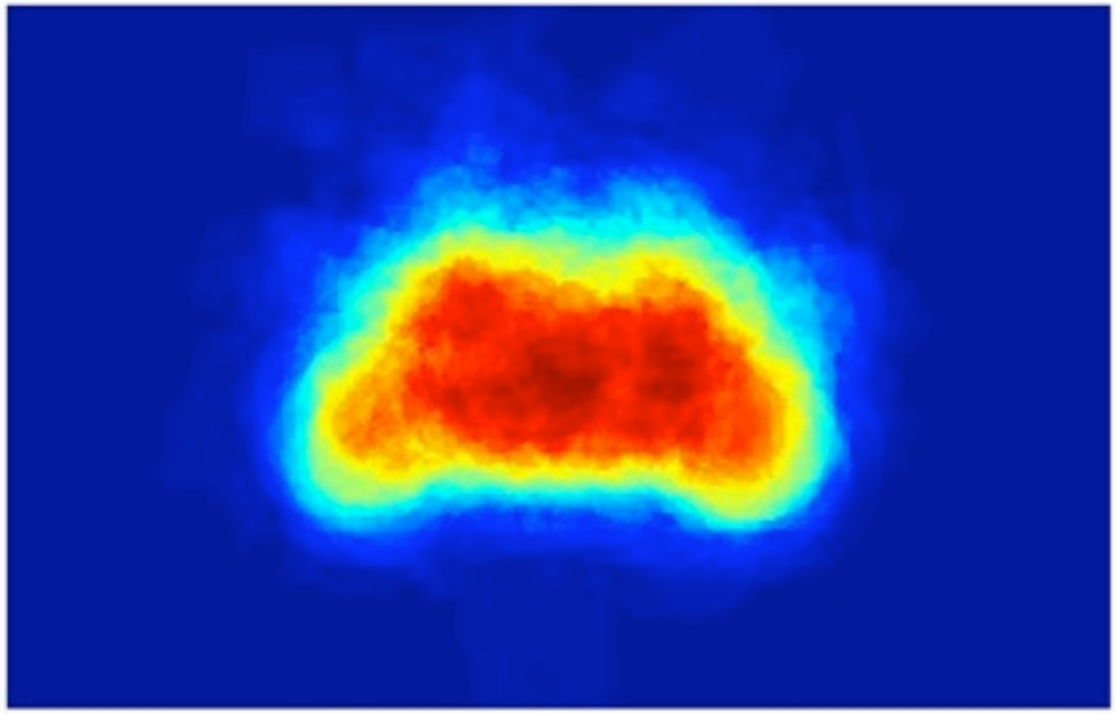


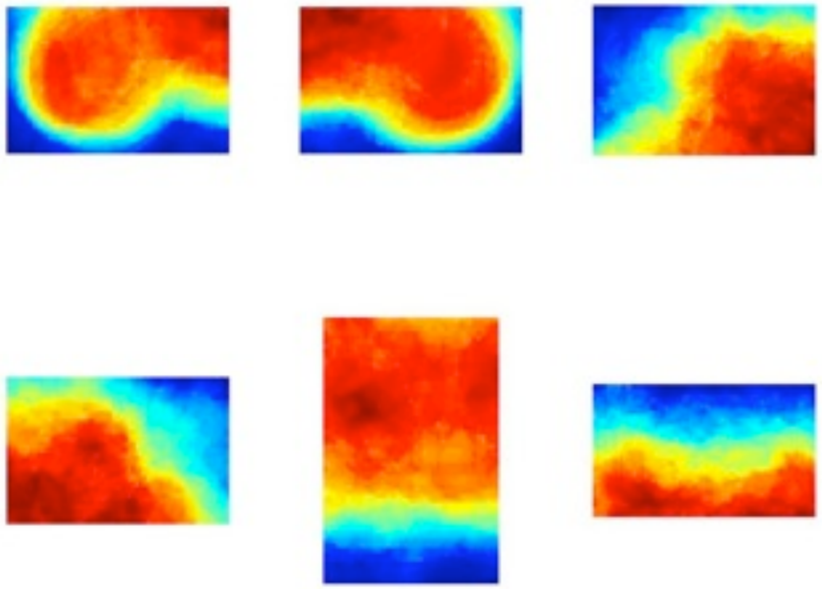


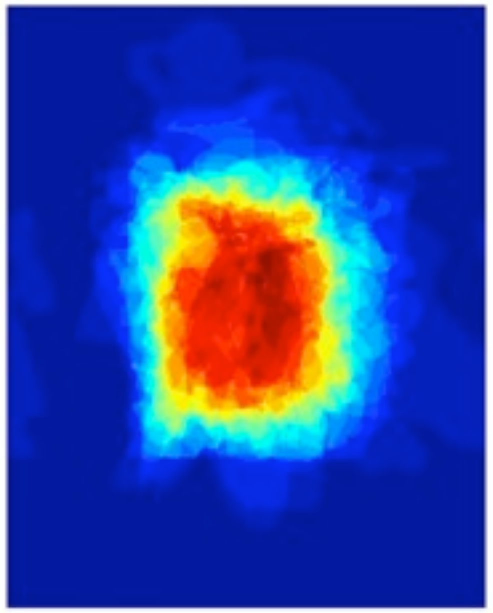
Bicycle part-based priors

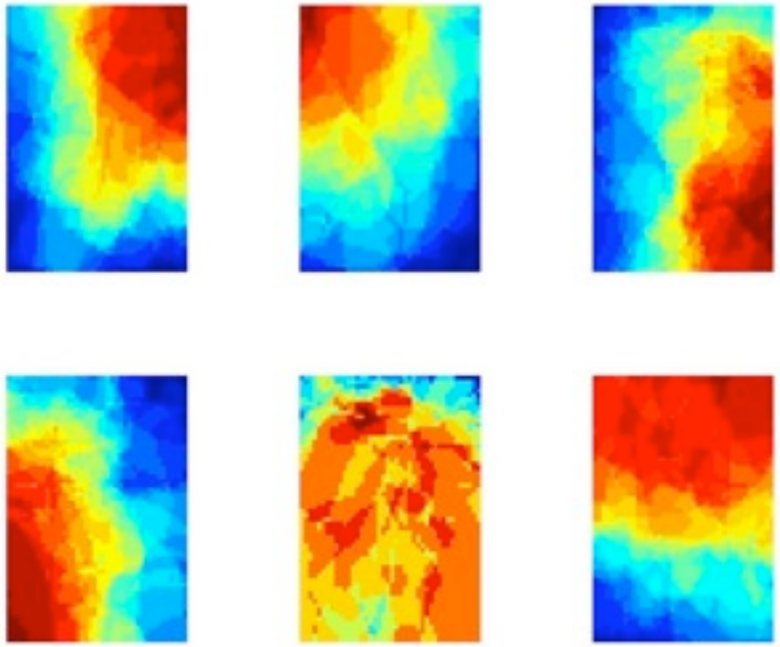


Motorcycle part-based priors

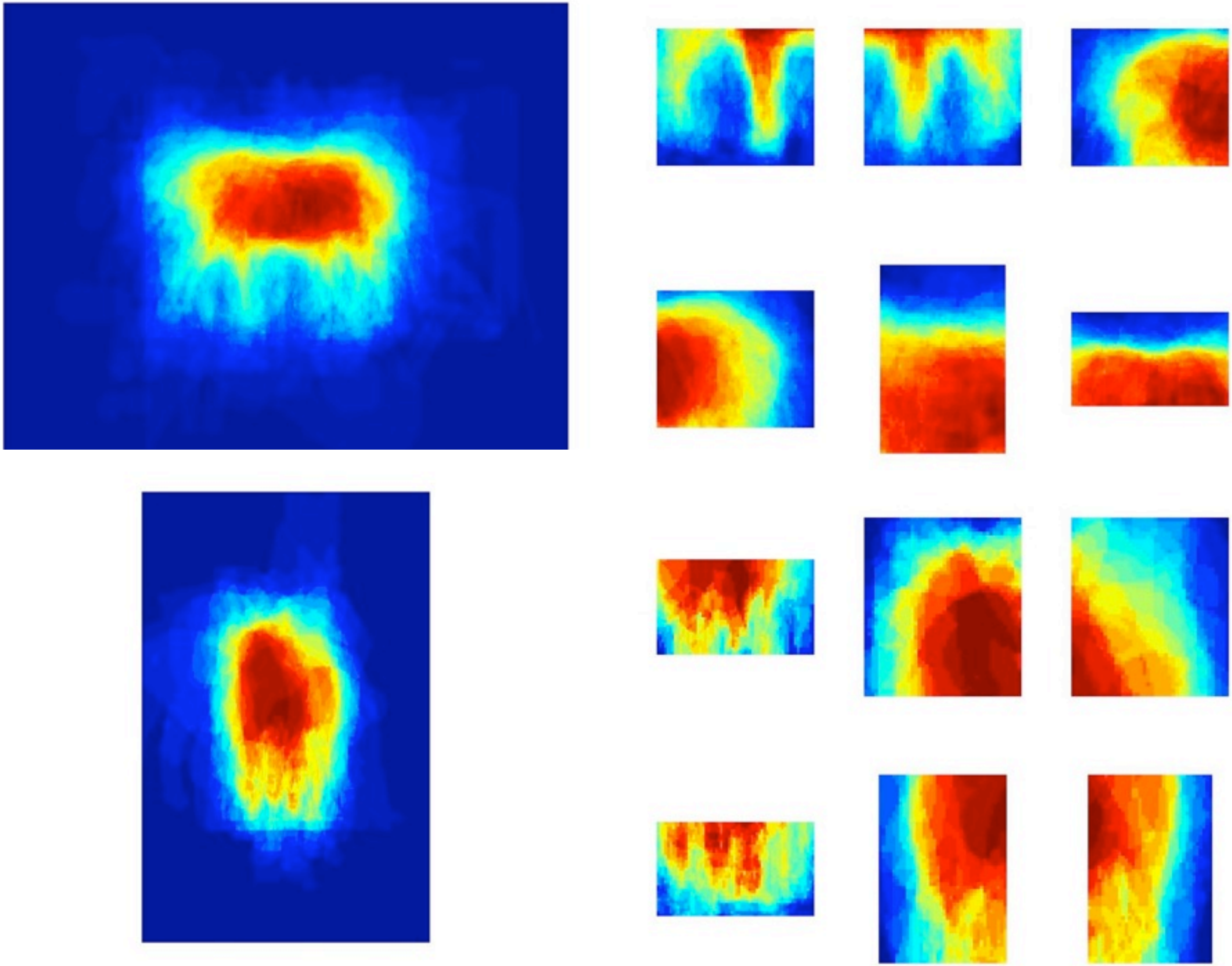




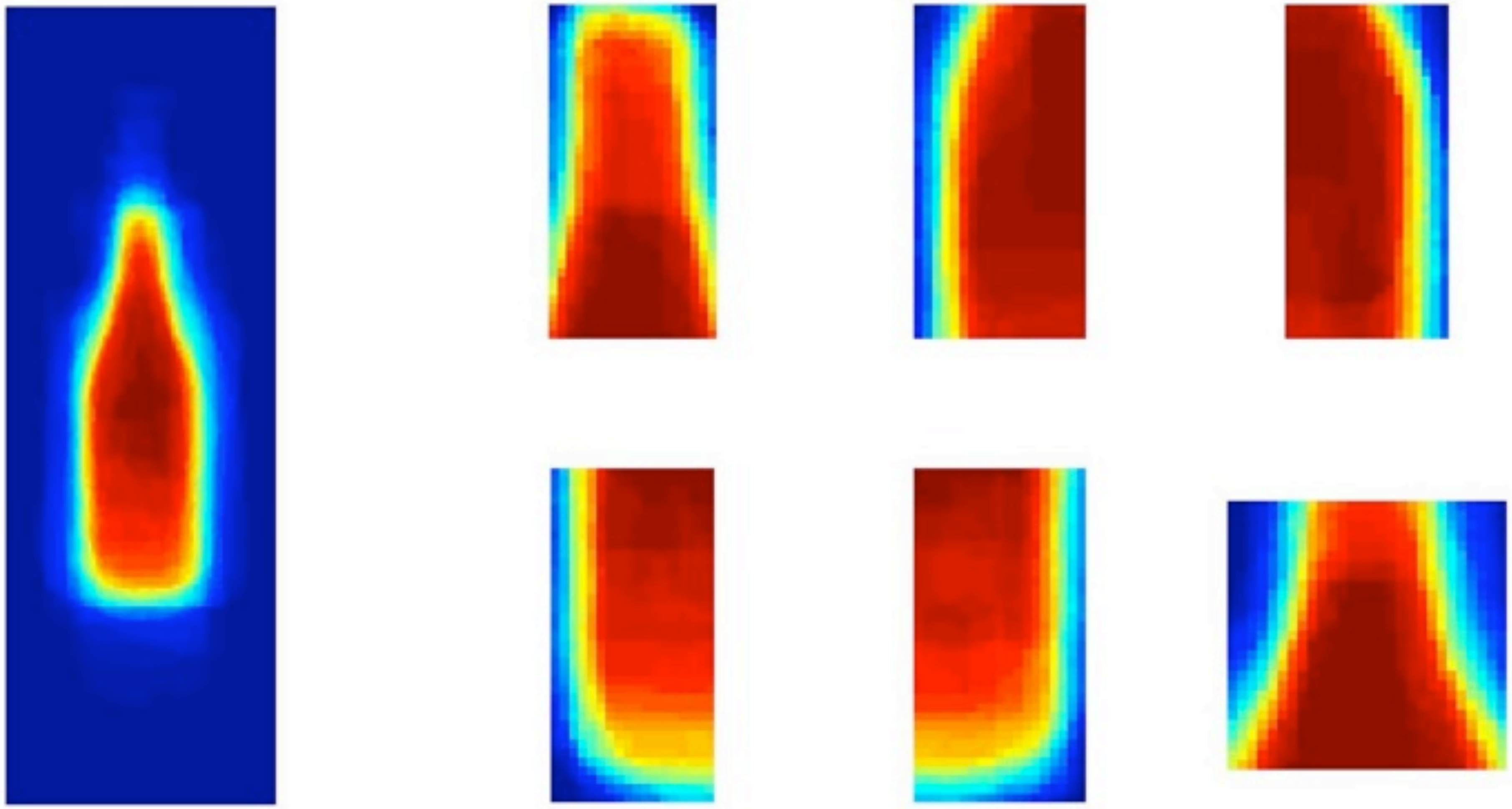




Horse part-based prior



Bottle part-based priors



Compositing detections

Order detections in depth from back to front

d_1 = furthest in depth (lowest score)
d_N = closest in depth (highest score)
D_n = set of n furthest (lowest scoring)

Initialize labels to background

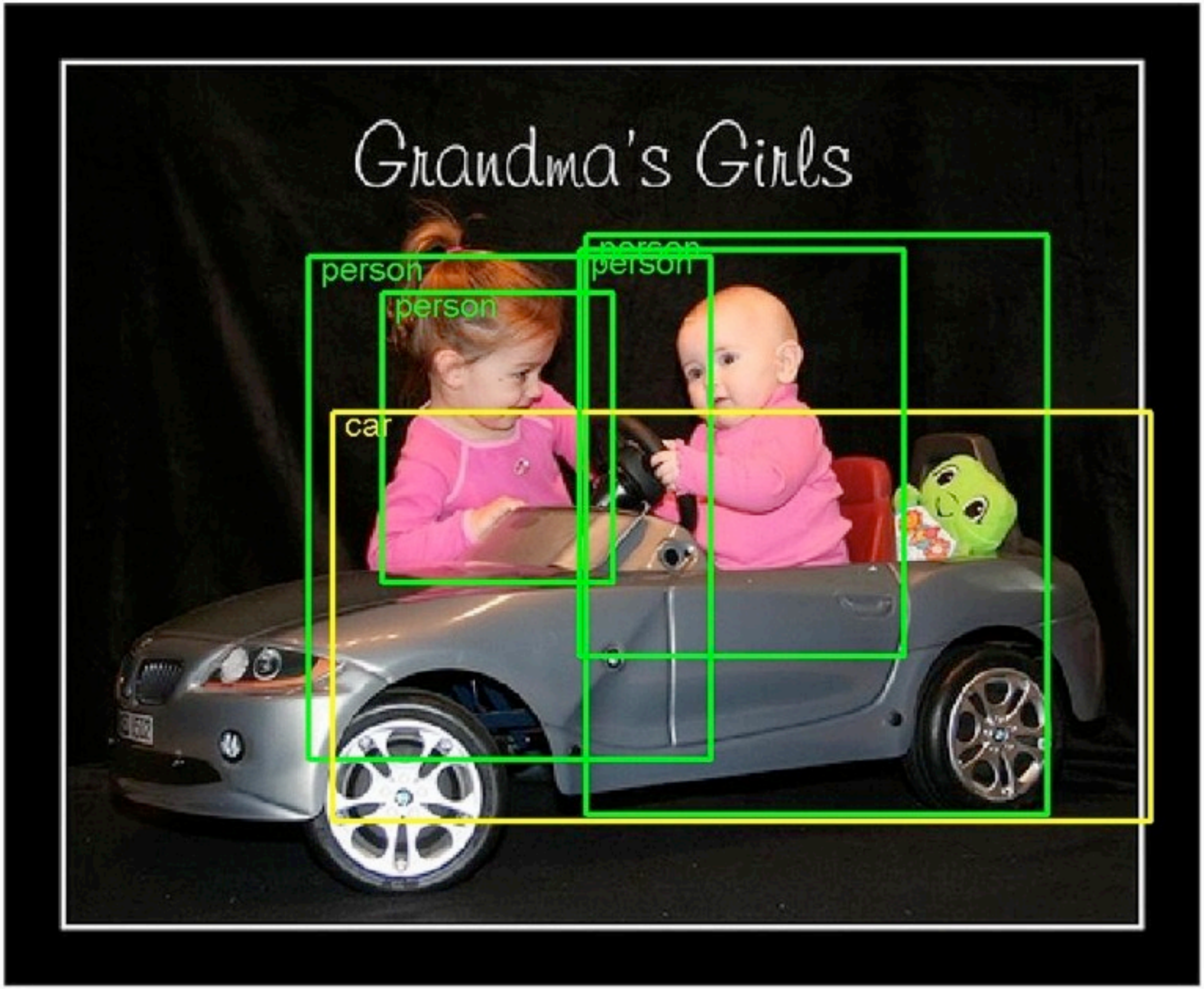
$$P(z_i|D_0) = \begin{bmatrix}1 & 0 & \dots & 0\end{bmatrix}$$

Composite detections from back to front

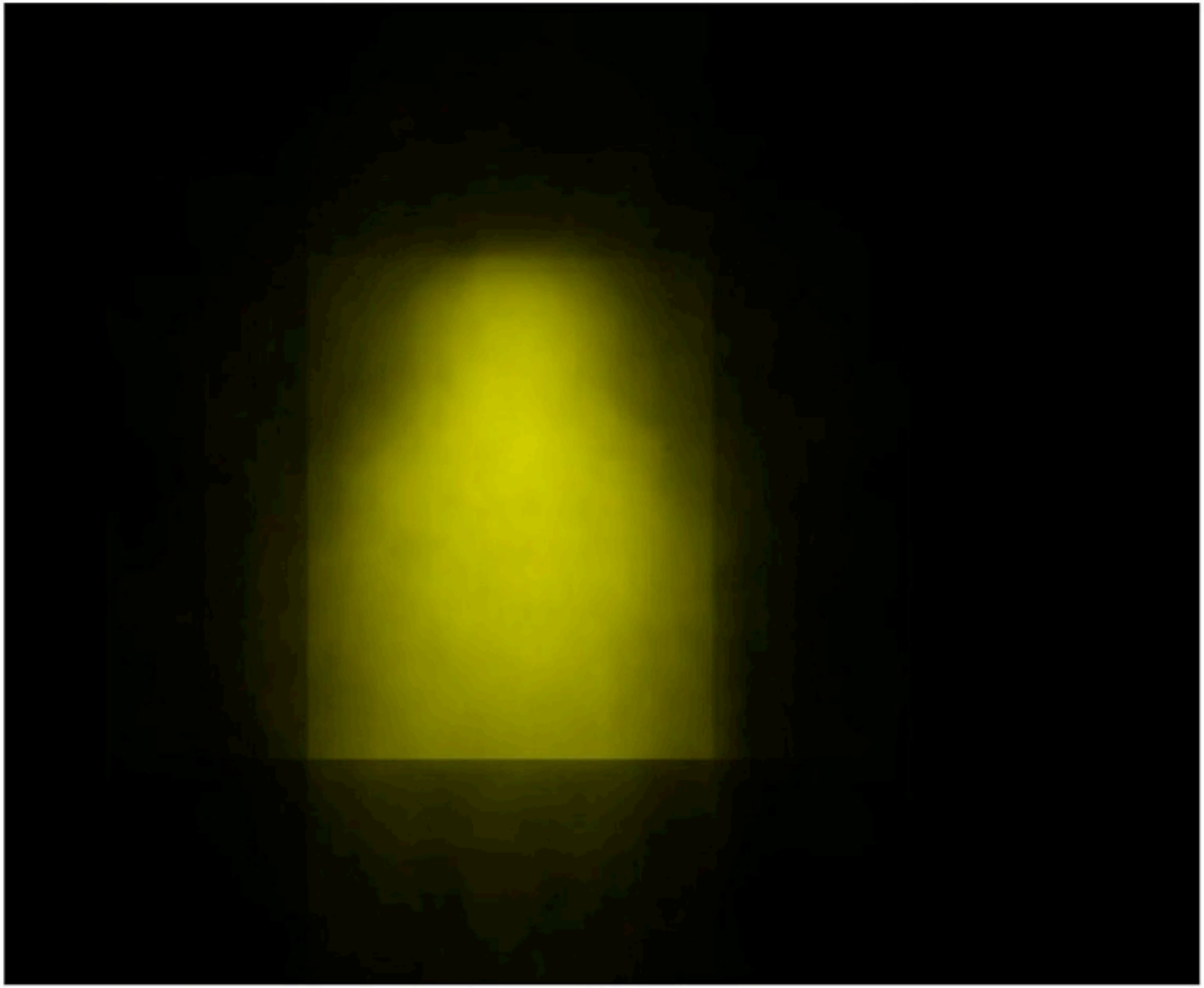
$$P(z_i|D_n) = P(z_i|d_n) + (1 - P(z_i|d_n))P(z_i|D_{n-1})$$

Multi-channel alpha matte compositing

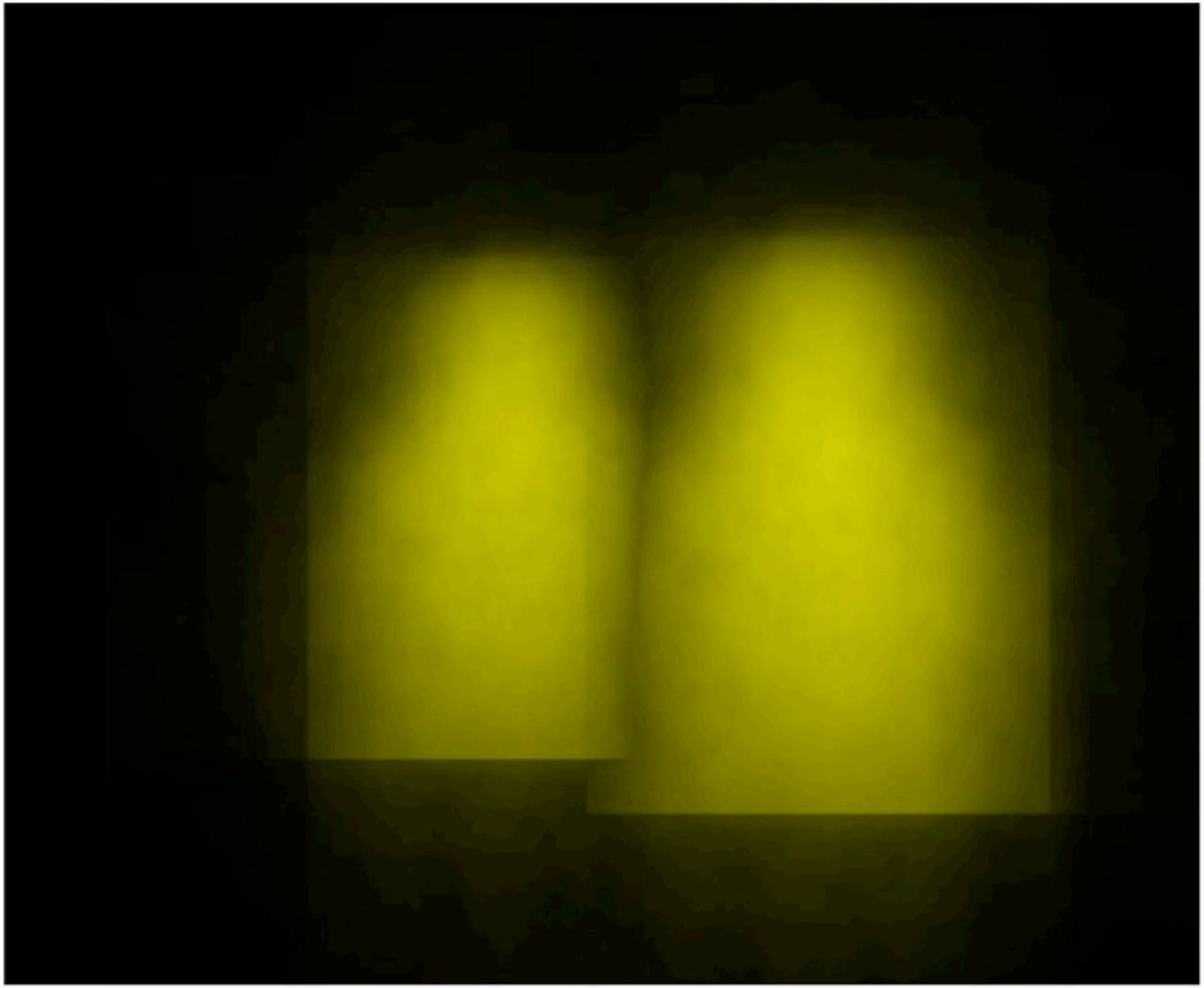
Compositing detections



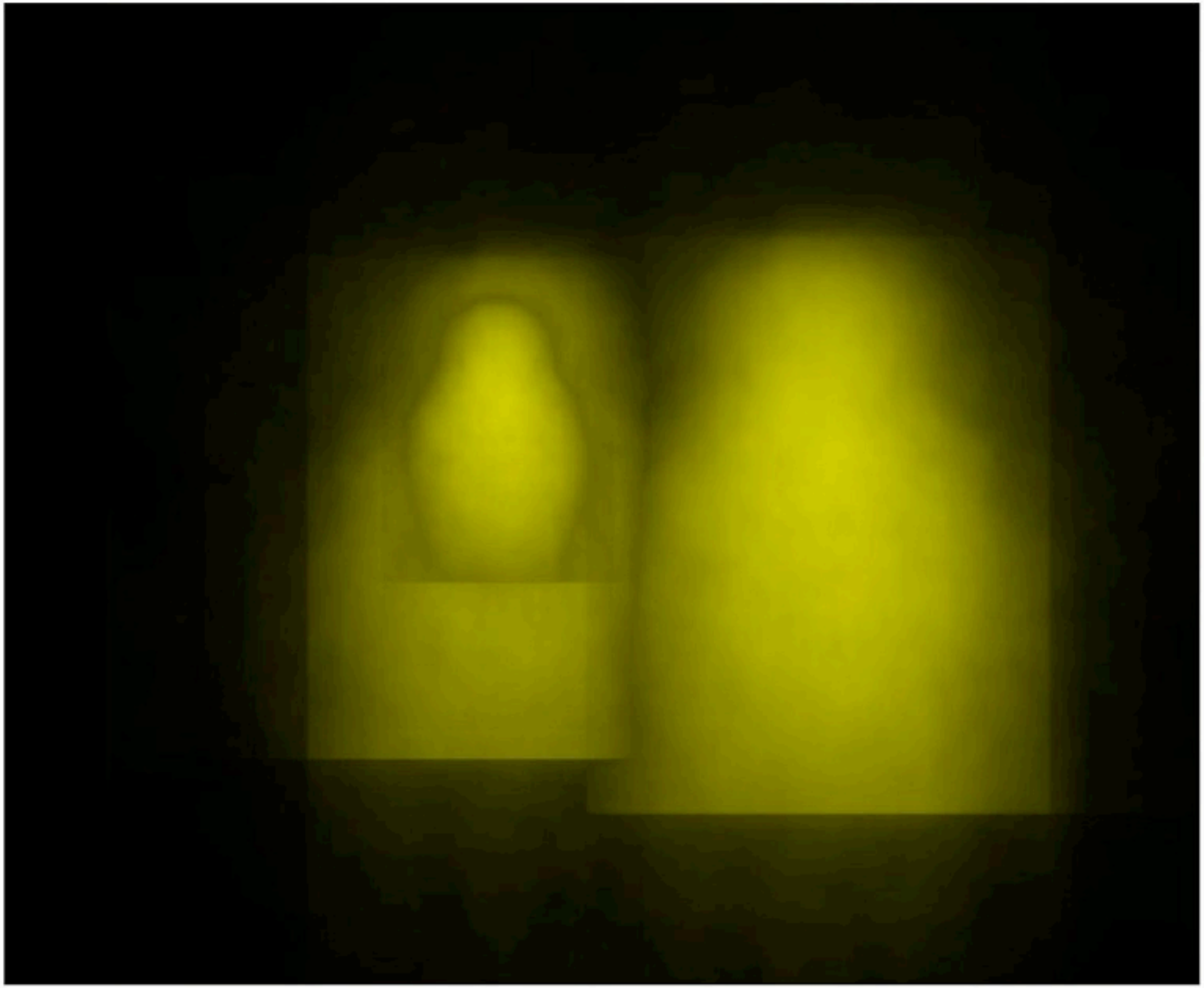
Compositing detections



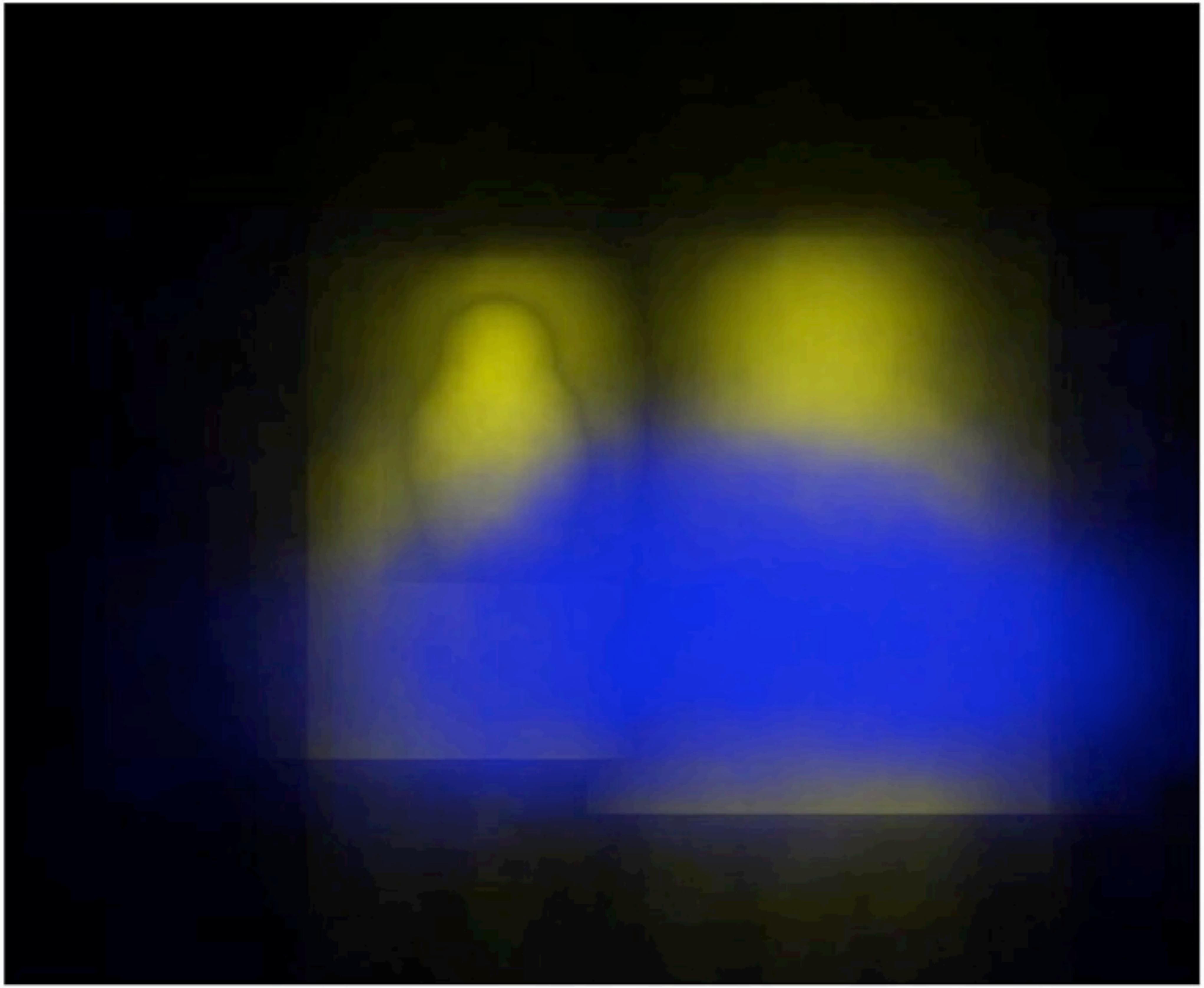
Compositing detections



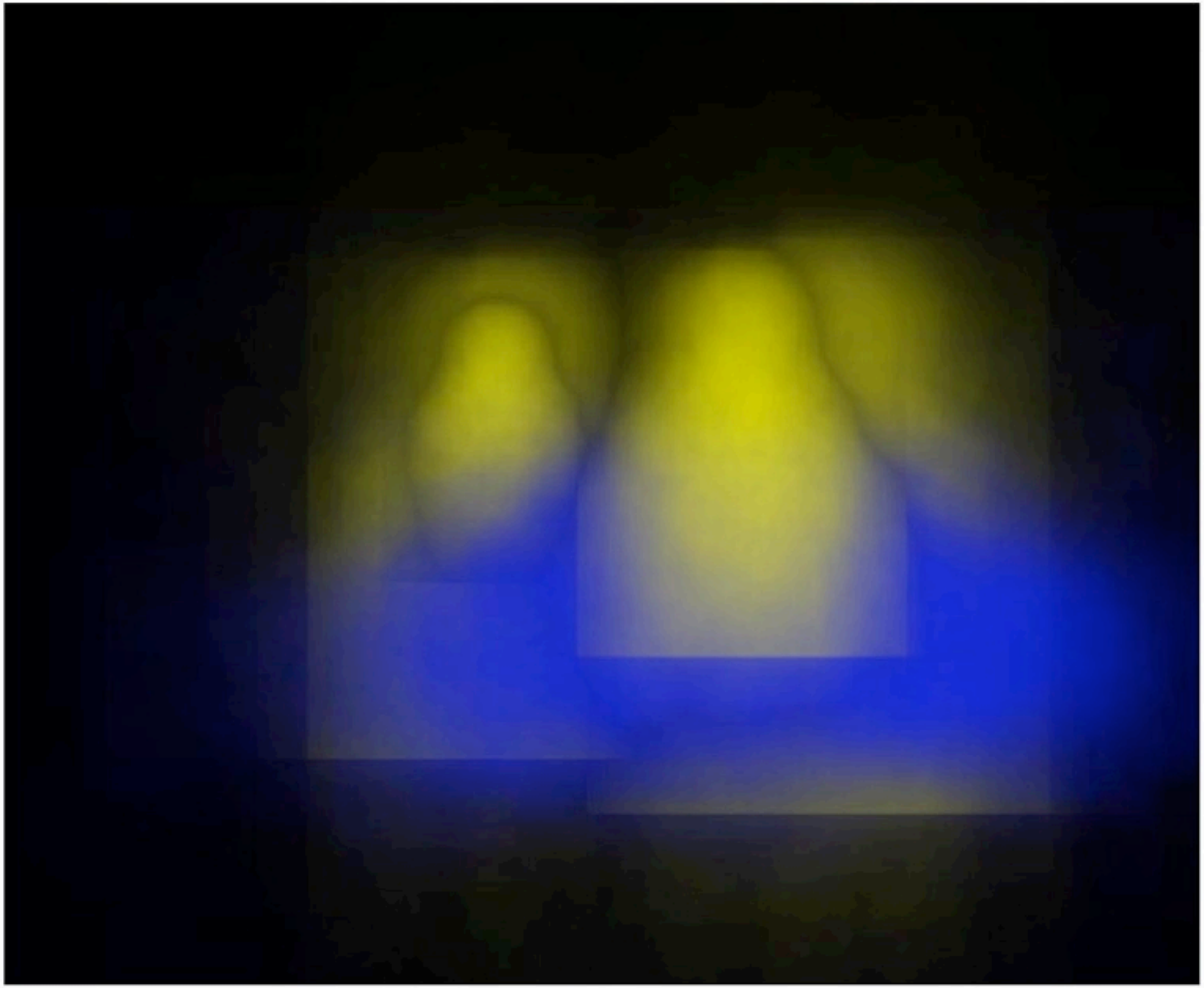
Compositing detections



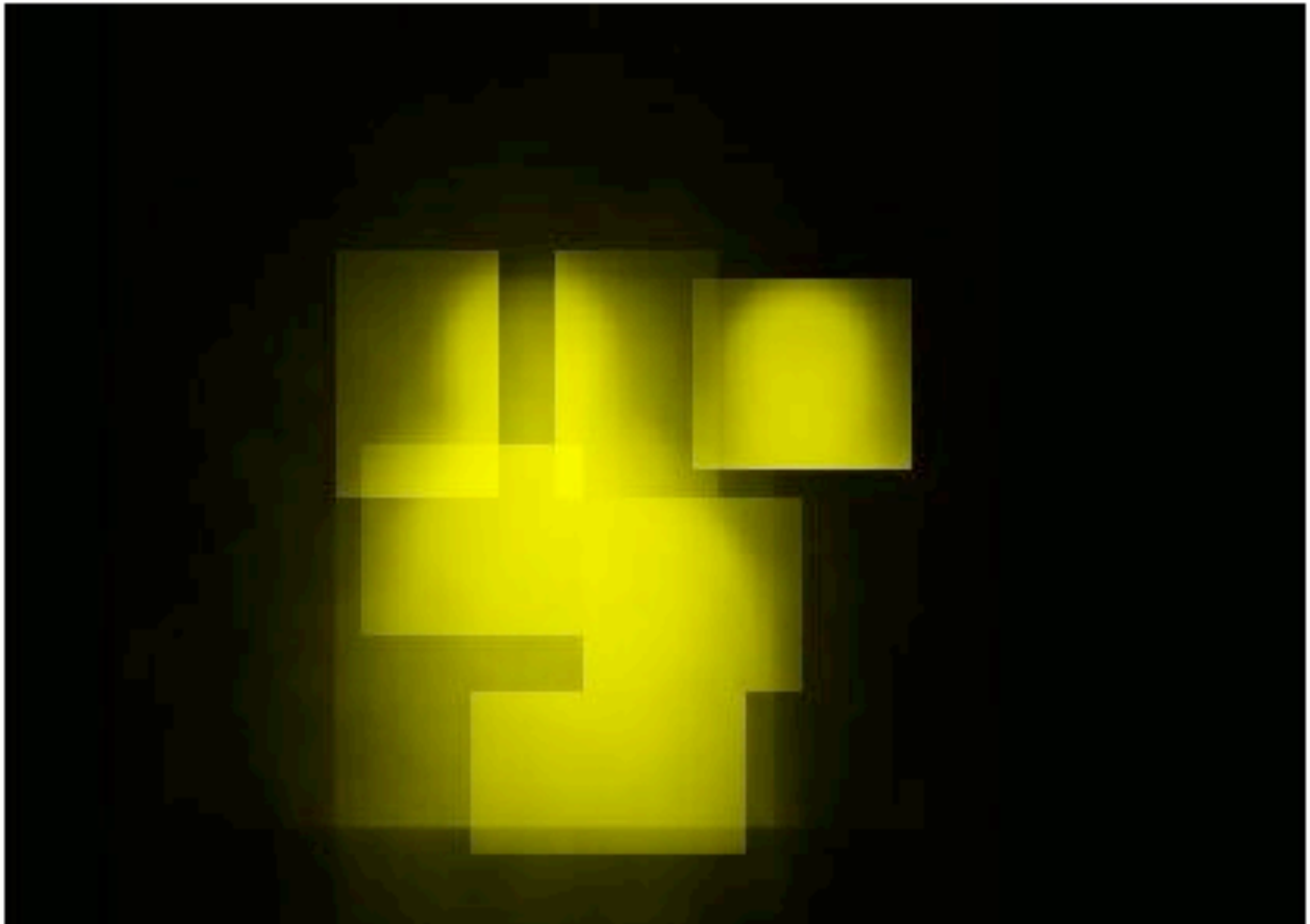
Compositing detections



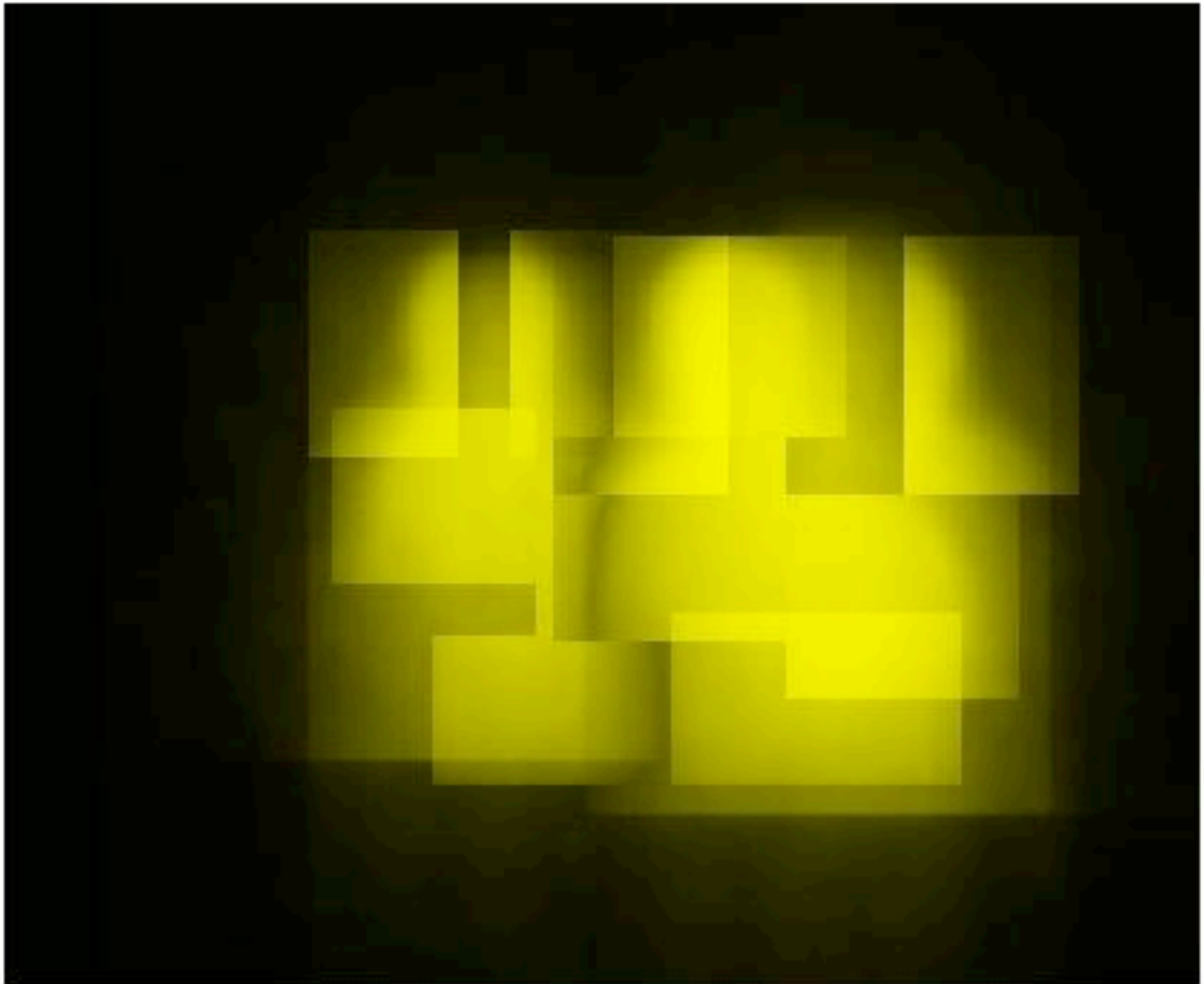
Compositing detections



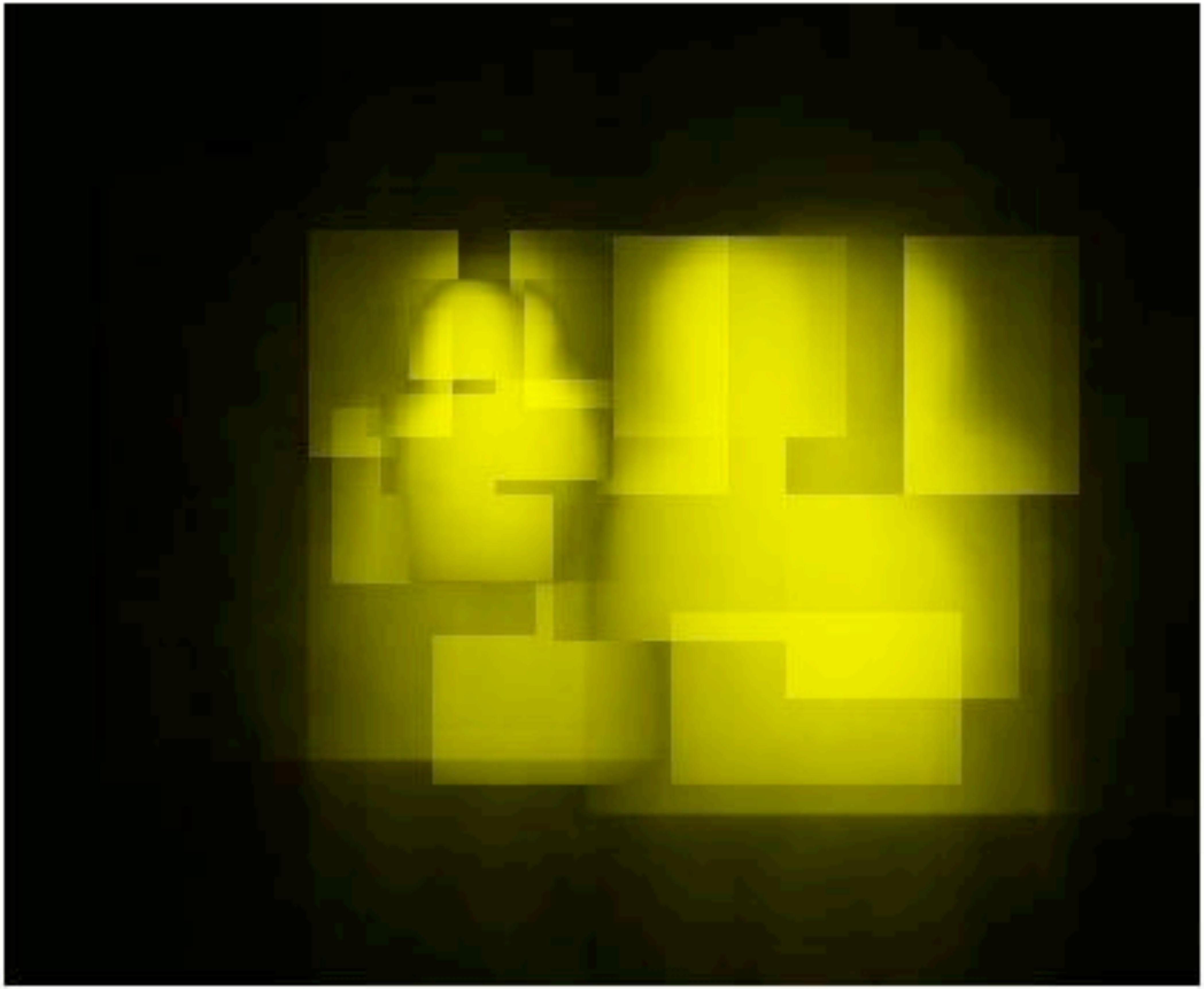
Composite parts in front of parent



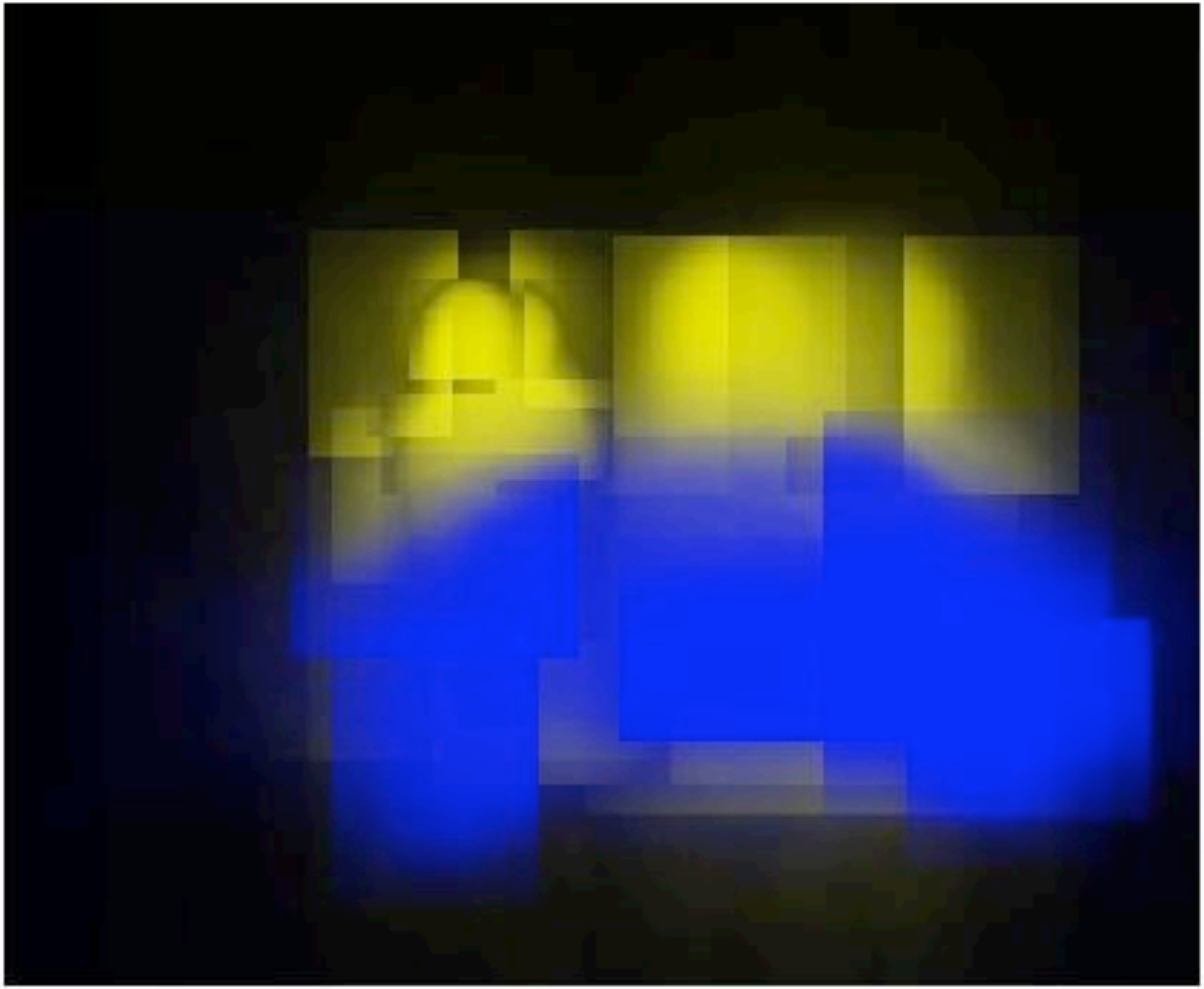
Composite parts in front of parent



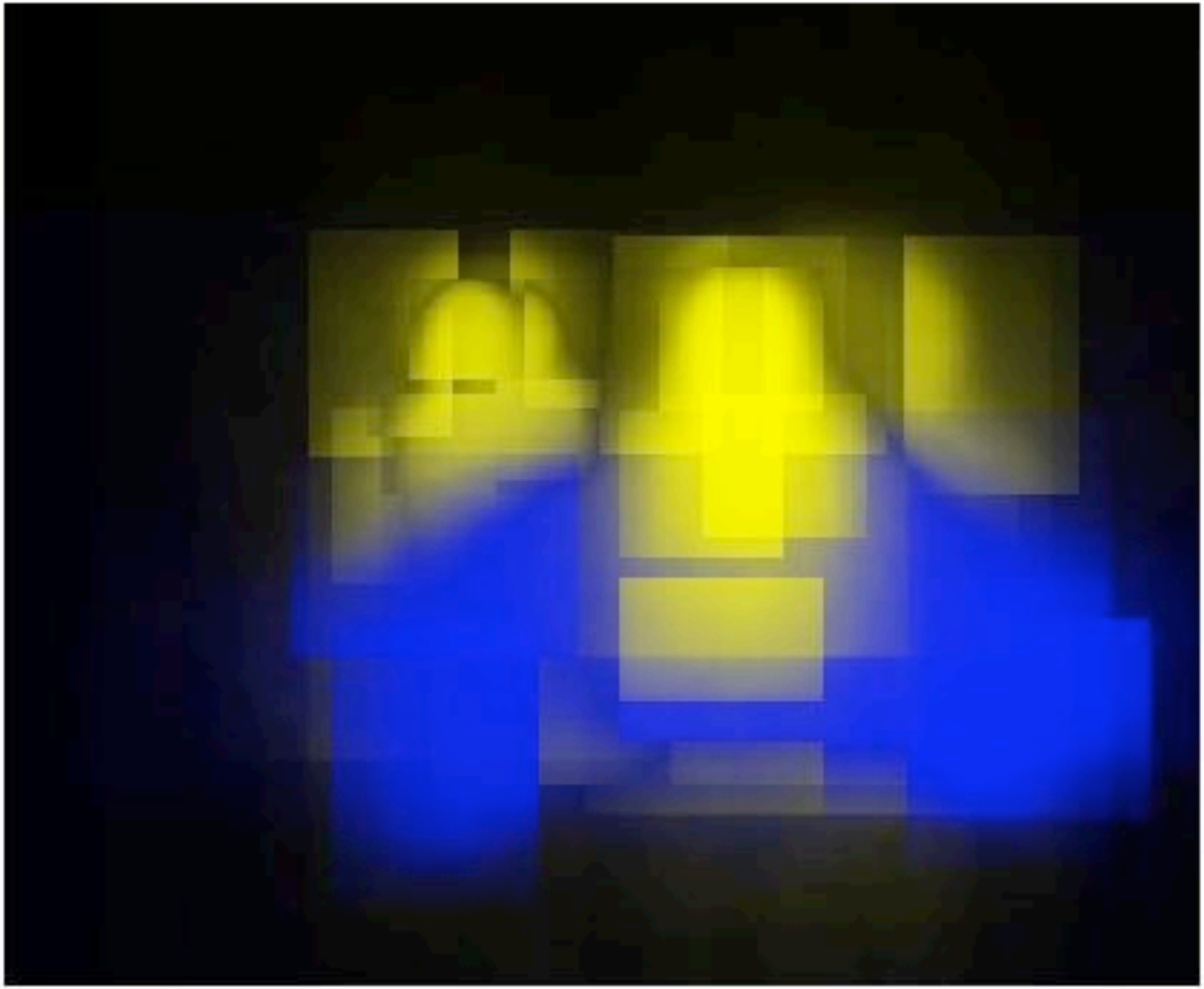
Composite parts in front of parent



Composite parts in front of parent



Composite parts in front of parent



Inference: coordinate descent

product
over pixels

$$P(Z, X|C, D) = \prod_{i=1}^{N} P(x_i|C_{z_i}) P(z_i|D)$$

color
likelihood

bbox
prior

$$f(C, Z) = -\log P(Z, X|C, D)$$

Step 1: $\underset{C}{\operatorname{argmin}} f(C, Z)$

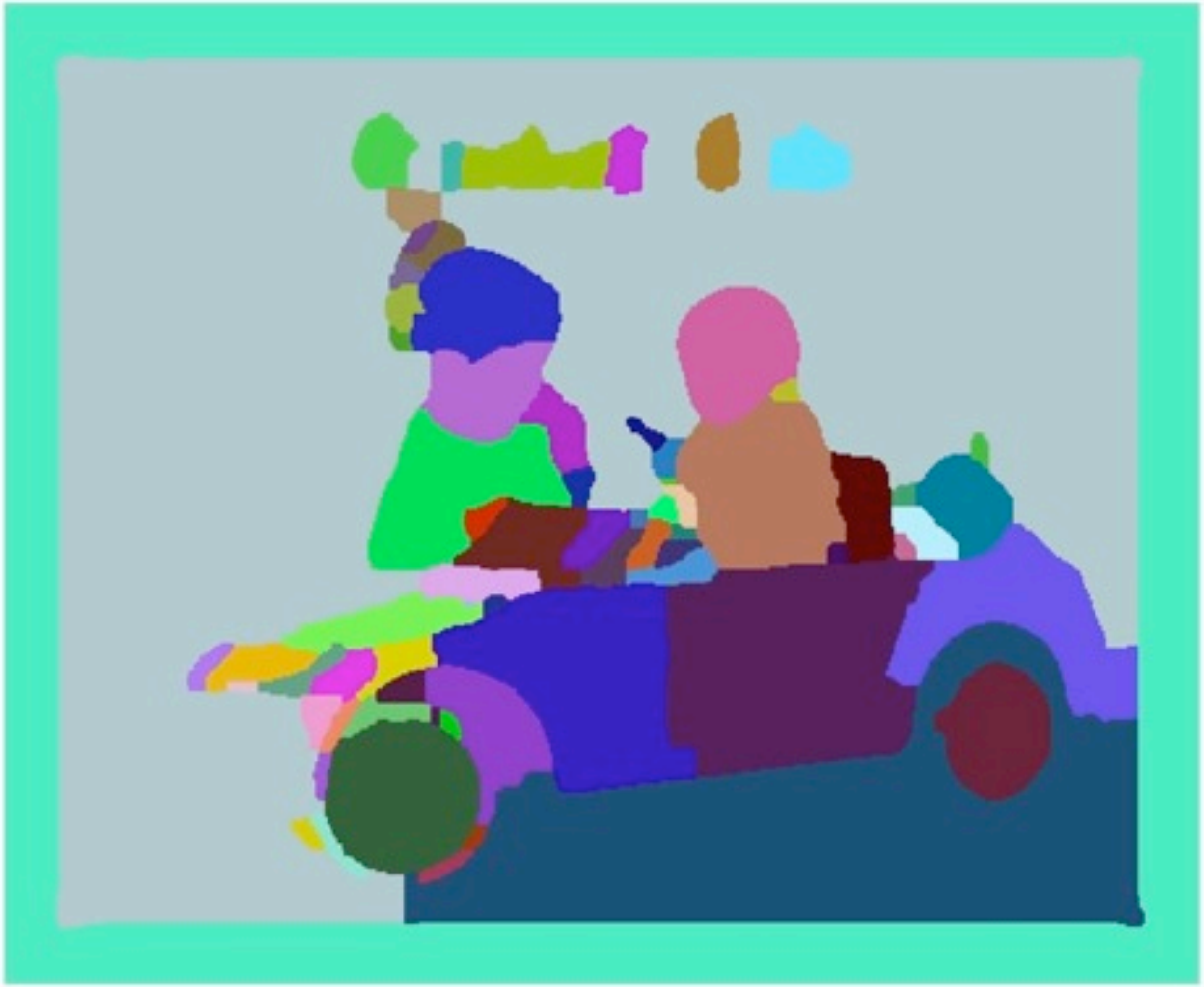
(fit color histogram
models to segments & bg)

Step 2: $\underset{Z \in \mathcal{Z}}{\operatorname{argmin}} f(C, Z)$

(per-pixel segmentation given
bbox prior & color likelihood)

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

Tune superpixel code for 200 superpixels



What's making our model work?

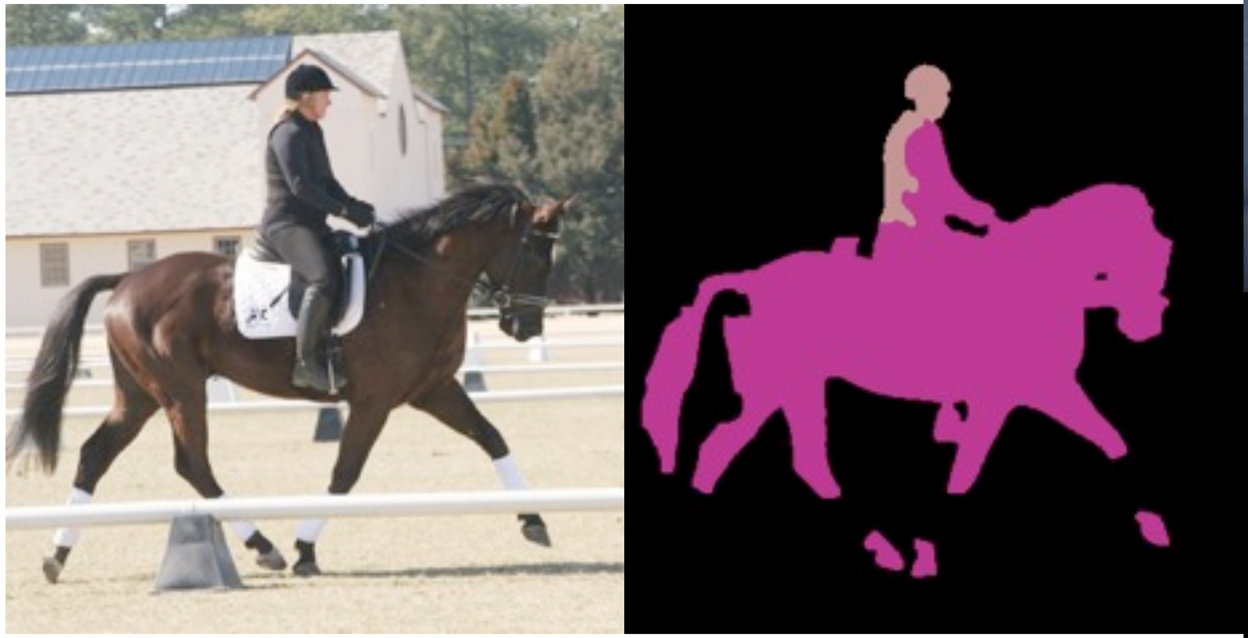
	- color	-parts	-sp	all
bg	78.51	78.55	76.44	78.67
plane	31.80	37.55	33.08	38.22
bicycle	24.66	25.81	19.94	26.26
bird	3.69	4.35	3.93	4.38
boat	8.85	10.69	10.07	10.67
bottle	36.04	35.13	36.39	36.02
bus	42.61	42.98	39.45	43.16
car	41.14	38.48	36.95	40.18
cat	14.99	14.90	15.68	15.50
chair	2.59	5.02	6.18	5.43
cow	10.22	11.17	10.03	10.95
table	6.03	7.14	5.96	7.10
dog	7.26	8.11	6.48	8.20
horse	13.25	13.67	12.59	14.40
mbike	23.59	24.56	21.45	25.17
person	29.73	33.70	33.06	35.41
plant	16.77	19.35	17.70	19.14
sheep	15.41	16.87	15.51	17.22
sofa	10.03	10.63	10.16	10.48
train	30.36	31.02	26.63	31.41
tv	41.22	39.43	38.81	41.04
avg	23.27	24.24	22.69	24.71

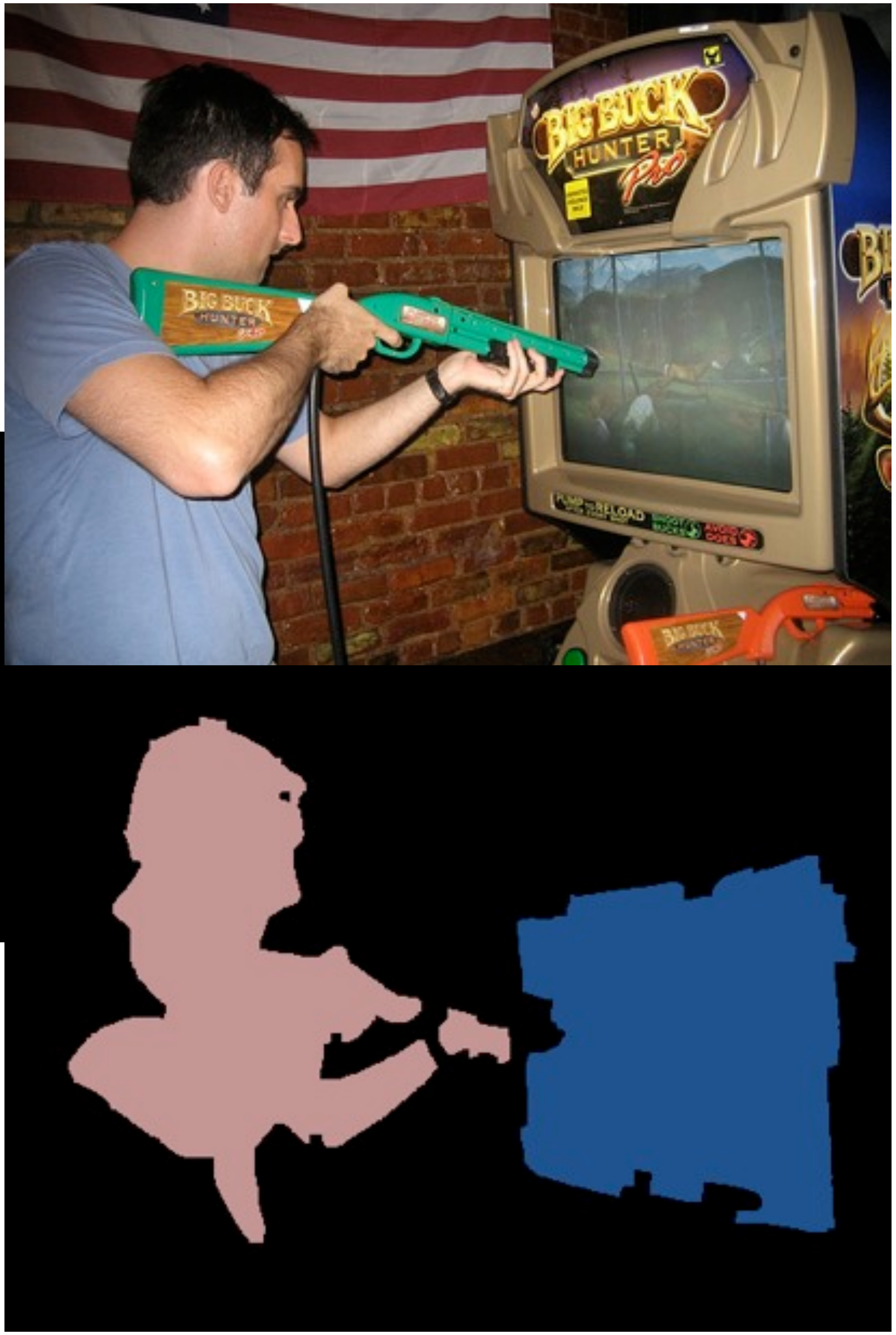
Superpixels really help bikes

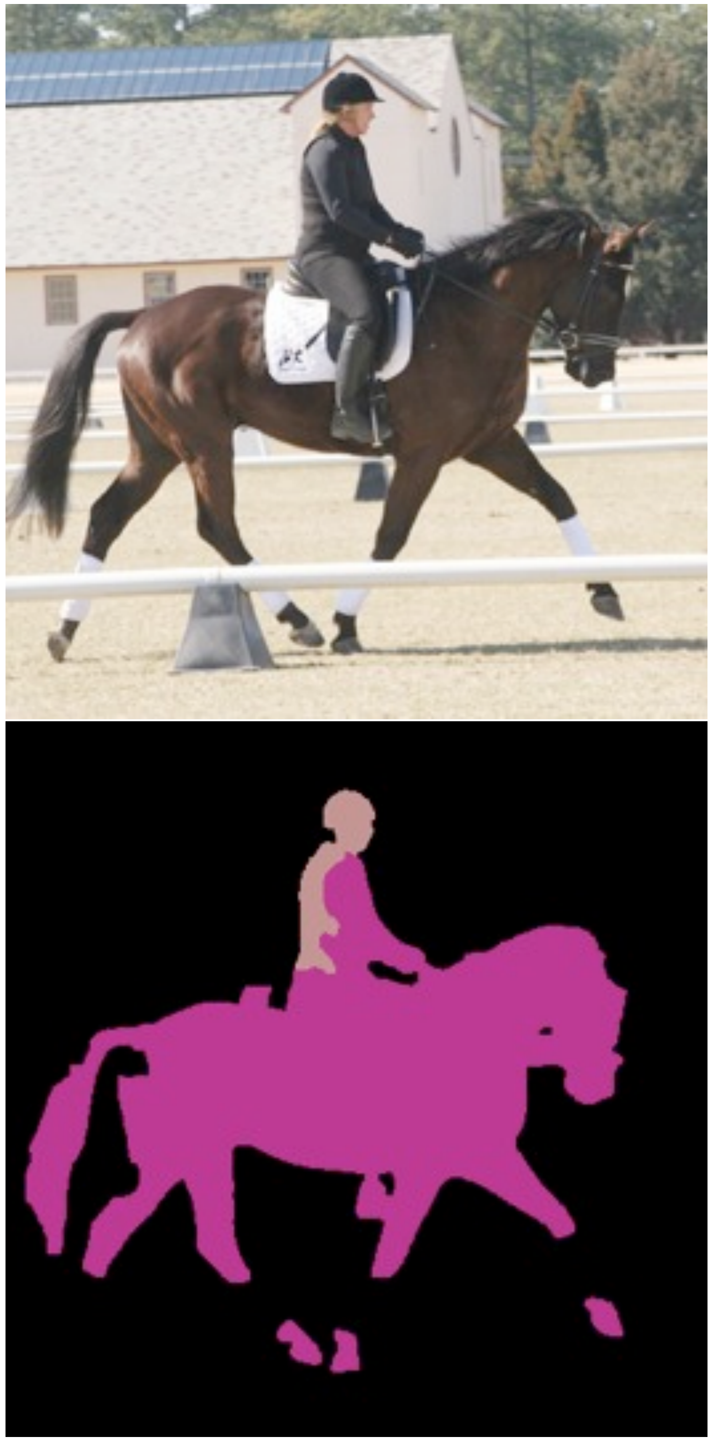
Color really helps person

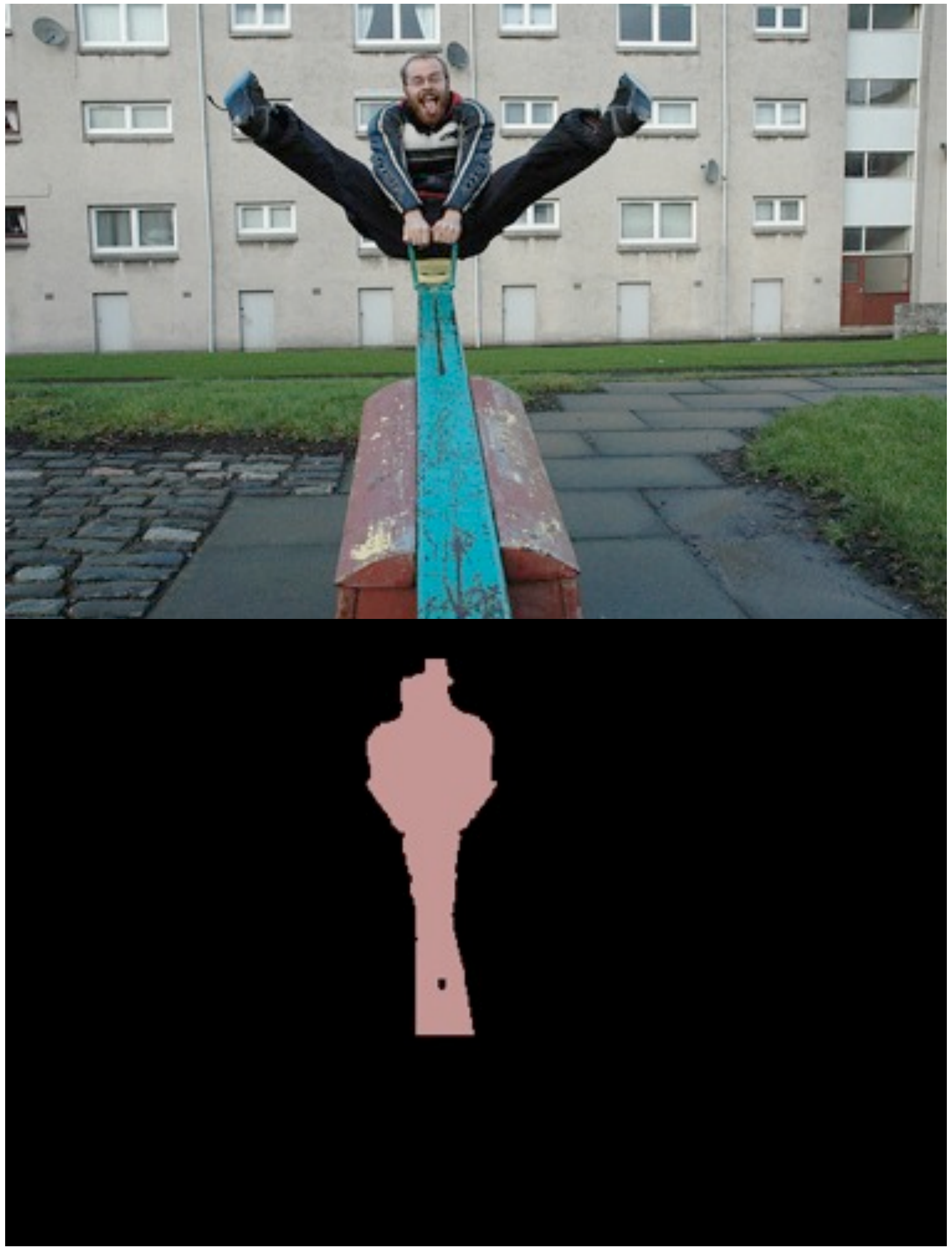
Parts provide small but noticeable improvement

...more examples...









WESTERN KENTUCKY
GLOB

QUALITY
SCHWINN
CHICAGO
Schwinn Built
Biocycles

Thanks!