A Meta-Theory of Boundary Detection Benchmarks

Xiaodi Hou
Computation and Neural Systems, Caltech

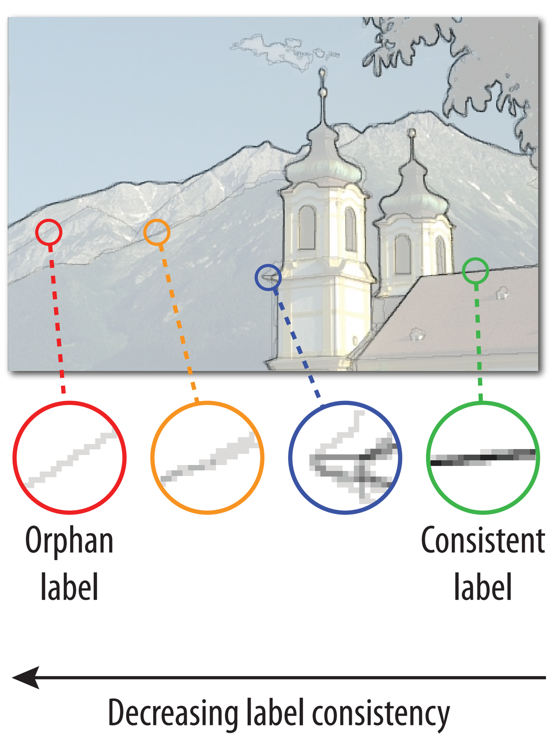
Alan Yuille
Department of Statistics, UCLA

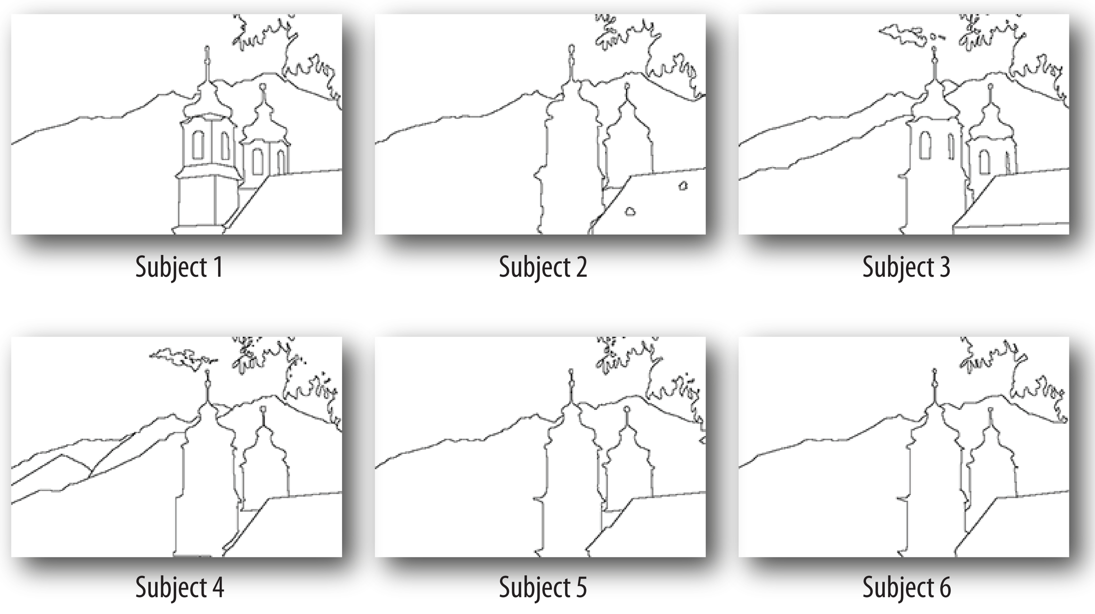
Christof Koch
Computation and Neural Systems, Caltech

Benchmarks of boundary detection

Human boundary labels are not always consistent

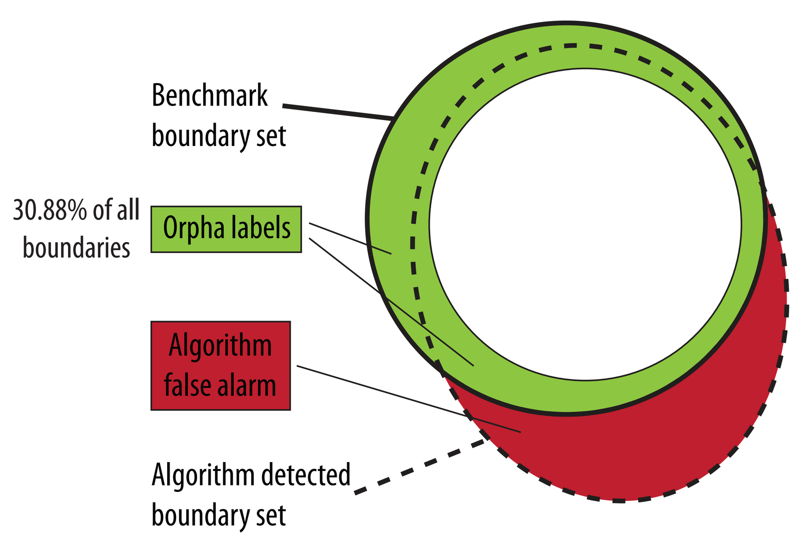
An example image in BSDS 300





Evaluating the risks of a benchmark

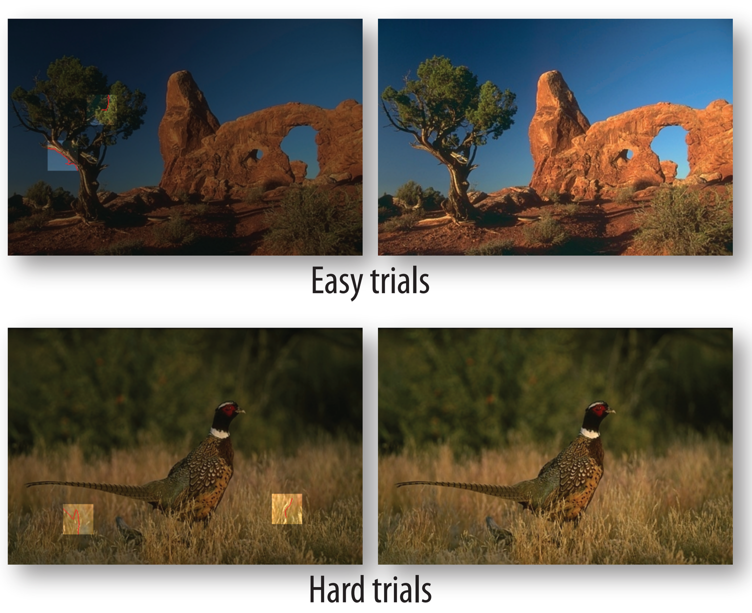
Algorithm false alarms v.s. human orphan labels



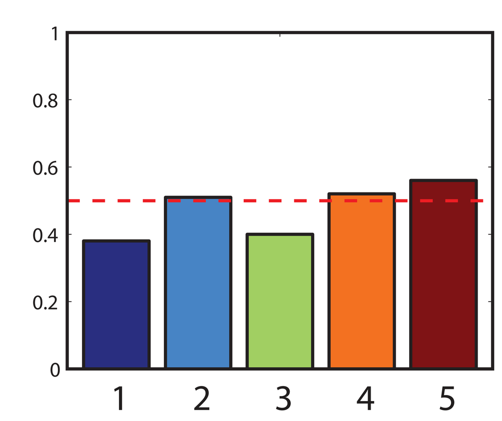
Red lines: algorithm false alarms. Green lines: human orphan labels

Two-way force choice experiment

Choose the stronger boundary segment from the two candidates



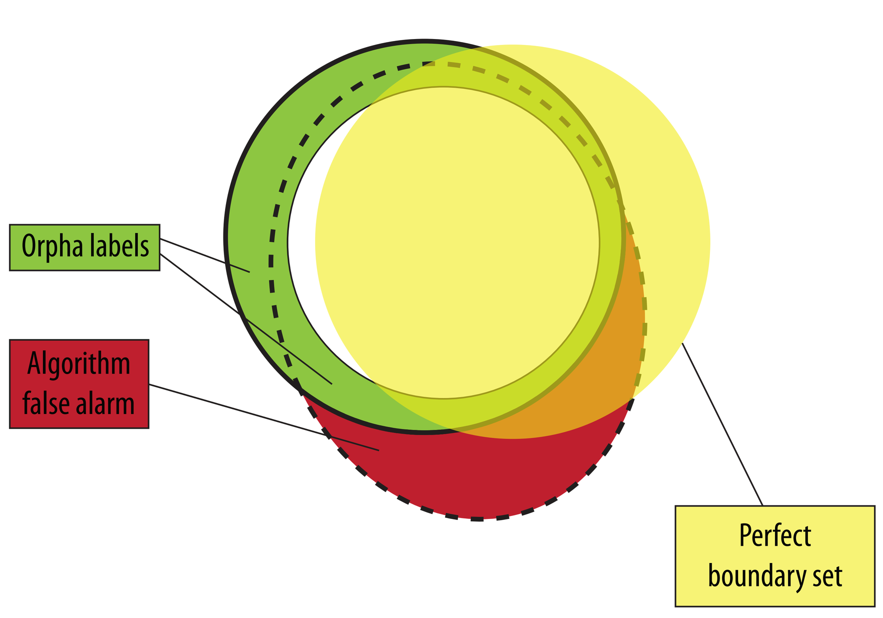
Experimental result of 5 subjects and 100 images:



Algorithm false alarms beats human orphan labels in 44% trials!!

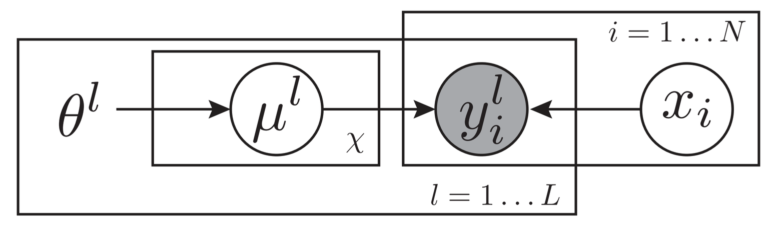
The risk of a boundary detection dataset

$$R(\mathcal{S},\mathcal{A}) = P(x_i < x_j \mid s_i \in \mathcal{S}, s_j \in \mathcal{A}\backslash\mathcal{S})$$



- s_i Boundary segment i
- x_i Perceptual strength of boundary segment i
- $\mathcal{S}$ Imperfect human-labeled boundary set
- $\mathcal{A}$ Algorithm (pB) detected boundary set
- $\tilde{\mathcal{S}}_\tau$ Risk-free perfect erfect boundary set

The Graphical Model of Labeling

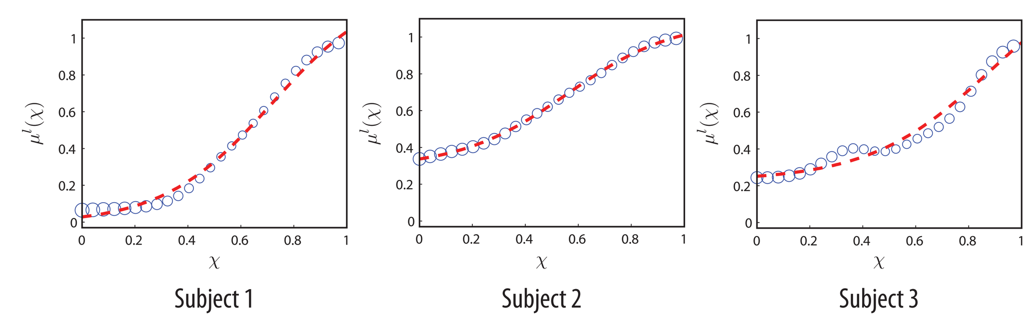


$$P(x_i) \sim \mathcal{U}(0,1)$$
$$\mu^l(\chi) = \mathrm{Sigmoid}_\theta(\chi)$$
$$P(y_i^l = 1 \mid \mu^l(\chi), x_i) = \int_\chi \phi_\sigma(x_i - \chi)\mu^l(\chi)\mathrm{d}\chi$$

- y_i^l Subject l's response to boundary segment i (observed)
- x_i Perceptual strength of boundary segment i (hidden)
- $\mu^l(\chi)$ Perceptual strength of boundary segment i (hidden)
- θ^l Perceptual strength of boundary segment i (hidden)

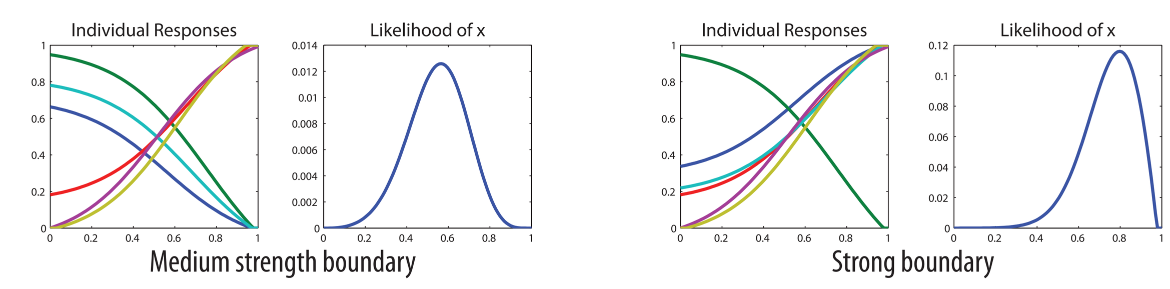
Estimating $\mu^l(\chi)$, θ^l

$$\mu^l(\chi)^\star = \sum_i y_i^l \phi_\sigma(x_i^\star - \chi)$$
$$\theta^{l\star} = \arg\min \int_\chi \left(\mathrm{Sigmoid}_{\theta^l}(\chi) - \mu^l(\chi)\right)^2 \mathrm{d}\chi$$

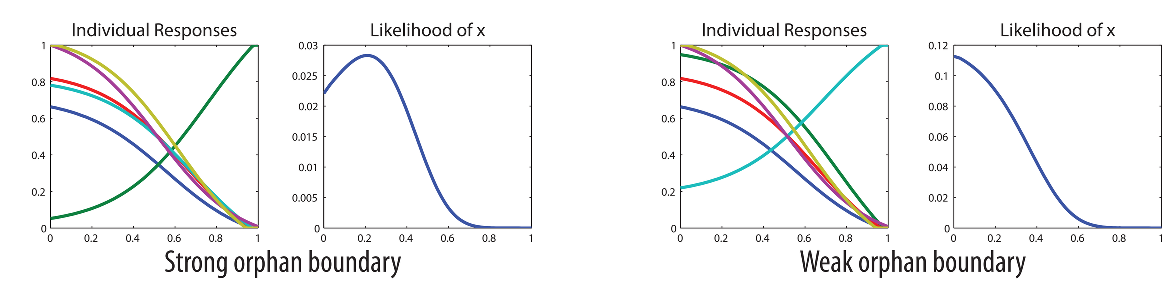


Estimating x_i

$$x_i^\star = \arg\max_{x_i} \prod_l P(y_i^l \mid \mu^l, x_i)$$

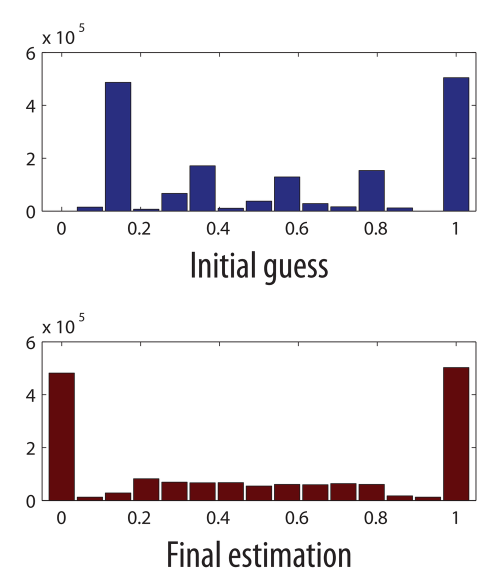


Some orphan labels annotated by (unnecessarily) detailed labelers can be filtered out by the majority of more determined labelers.



Experiment Results

Distributions of perceptual strength



Risks

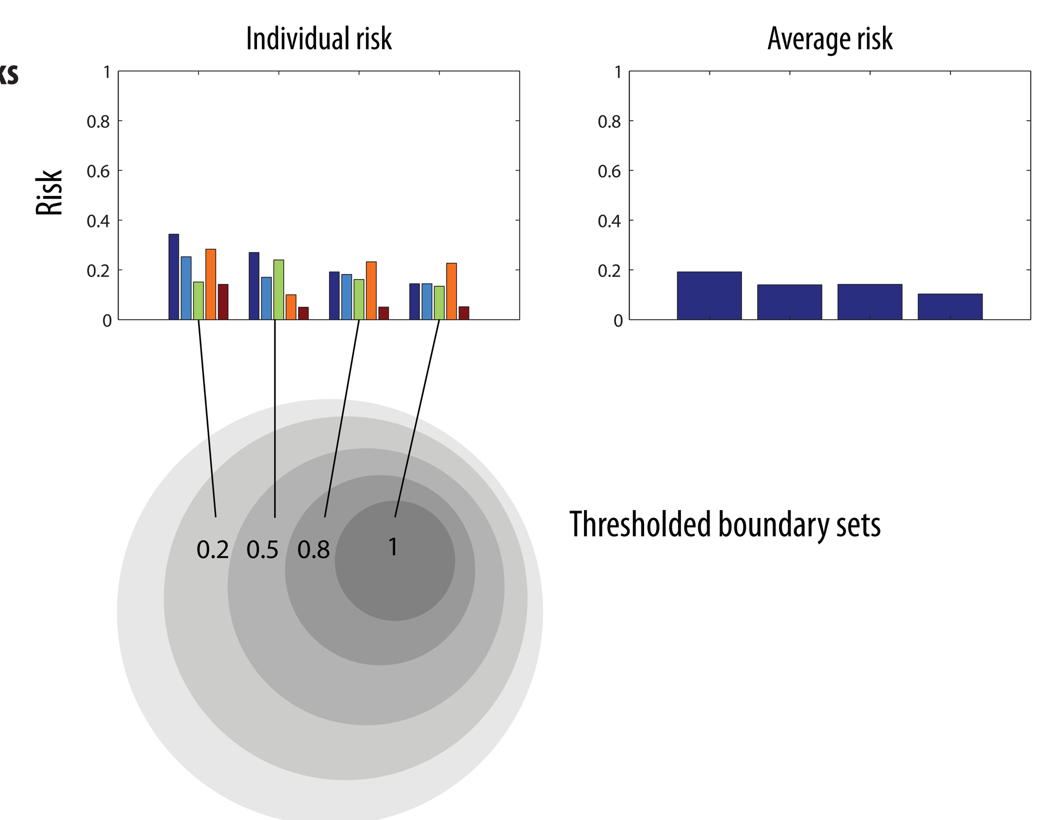


Image results