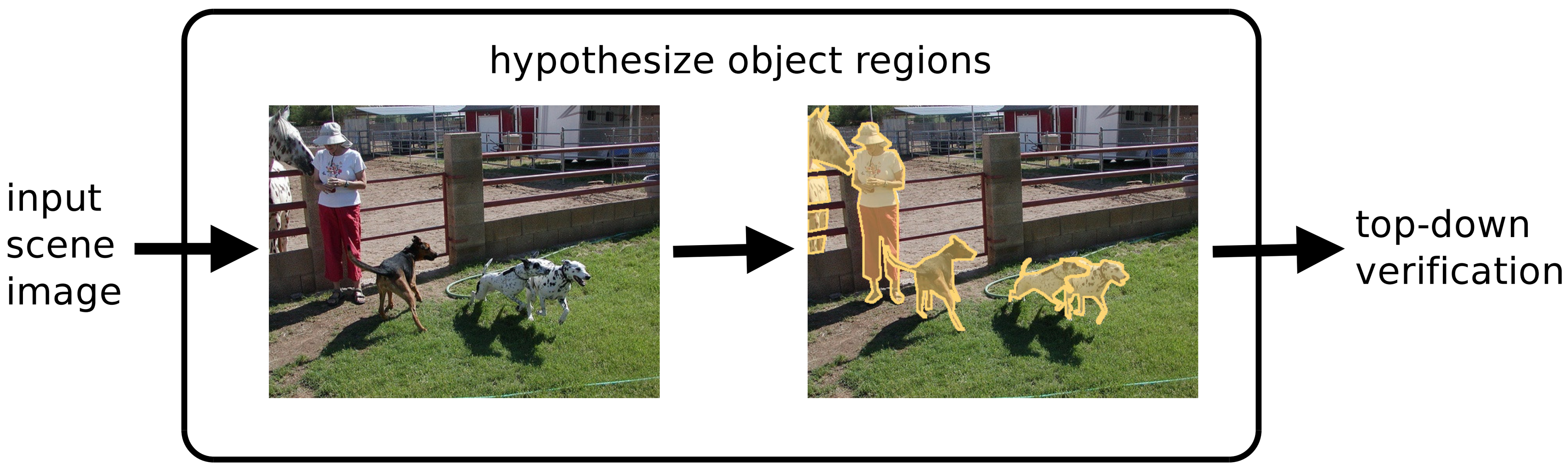
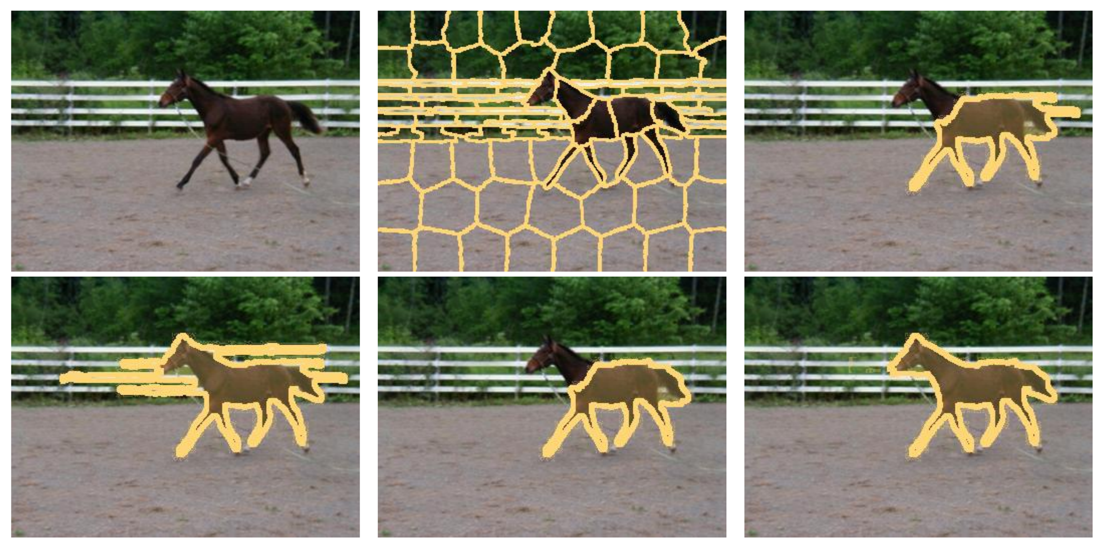
Multi-cue mid-level grouping

Tom Lee, Sanja Fidler, Sven Dickinson
University of Toronto

UNIVERSITY OF TORONTO

Overview





Motivation

- region proposals [CPMC, Selective Search] are far fewer than sliding windows
- yet, thousands of regions still contain a great deal of noise
- can we exploit mid-level grouping cues to improve precision?

Contributions

- we directly tackle the front-end generation of regions
- we use mid-level cues to eliminate false positive regions
- we learn the best combination of grouping cues

Grouping cues: non-accidental regularities

Energy function

- assign costs to binary superpixel labelings
- energy potentials express grouping cues

$$E(\mathbf{y};\mathbf{x}) = \sum_{\text{cue}} \mathbf{w}_{\text{cue}}^{\text{T}} \sum_{C} \phi_{\text{cue}}(\mathbf{y}_C;\mathbf{x})$$

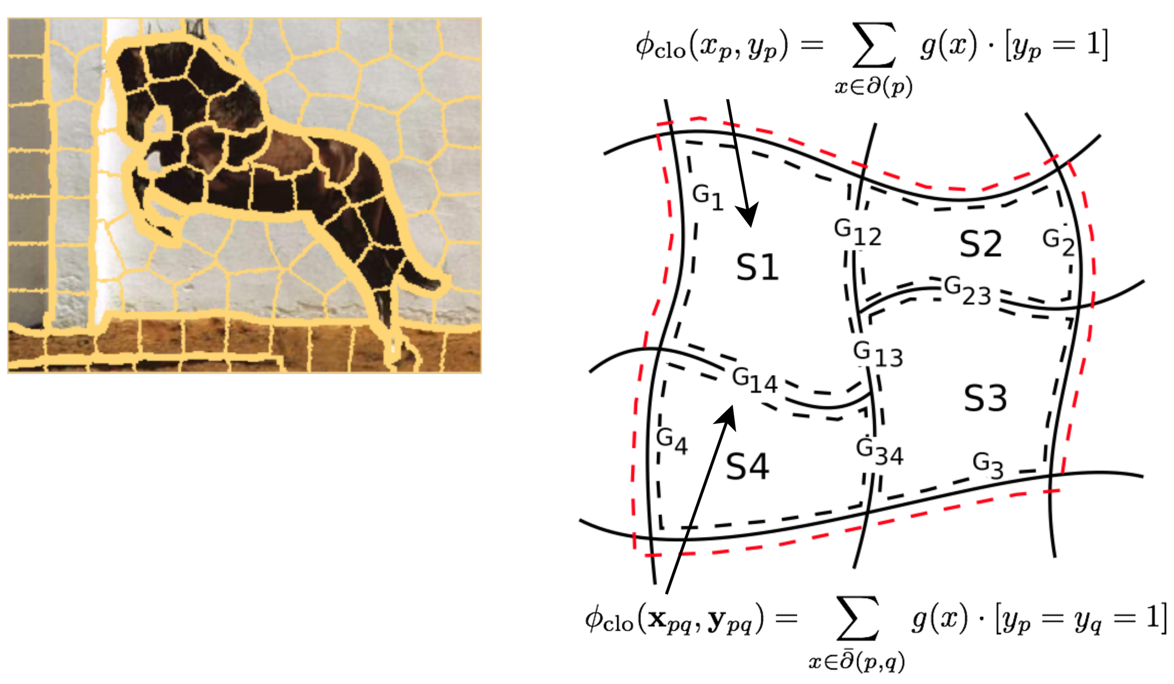
Appearance cue

- discourage splitting homogeneous regions
- use similarity in color and texture $s_{pq}^{\text{col}}, s_{pq}^{\text{text}}$

$$\phi_{\text{app}}(\mathbf{y};\mathbf{x}) = (s_{pq}^{\text{col}}, s_{pq}^{\text{text}}) \cdot [y_p \neq y_q]$$

Closure cue [Levinshtein et al., ECCV'10]

- discourage regions with gaps along boundary
- minimize $\text{Gap}(\mathbf{y};\mathbf{x}) = \sum_{x \in \partial(\mathbf{y})} g(x)$, rewritten as energy potentials



Symmetry cue [Lee et al., ICCV'13]

- discourage splitting regions in a symmetric part
- each symmetric part encourages cohesion of superpixels inside

- take each part's detection score into account

$$\phi_{\text{sym}}(\mathbf{y};\mathbf{x}) = \max_{s \in S(p,q)} \text{score}(s) \cdot [y_p \neq y_q]$$

Inference and learning

Hidden scale

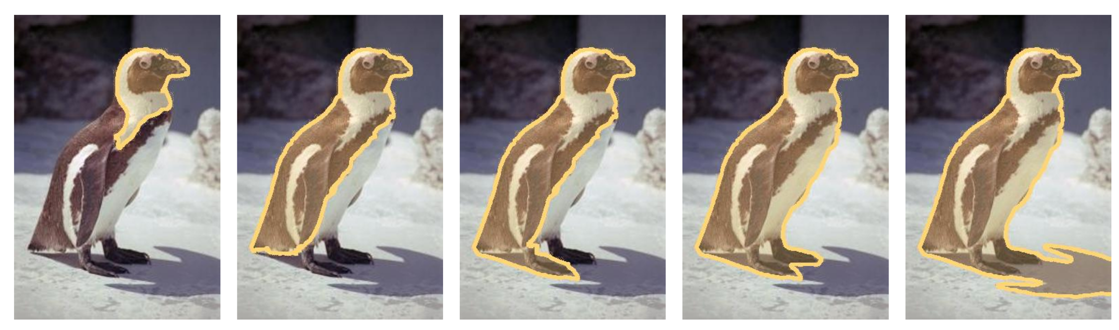
- use area potential $\phi_0(y_p)$ as a prior on scale
- introduce variable λ to adjust ϕ_0's influence

Parametric energy minimization

- minimize $E^{\lambda}(\mathbf{y})$ for multiple values of λ

$$E^{\lambda}(\mathbf{y}) = \mathbf{w}_1^{\text{T}} \sum_{p} (\lambda \cdot \phi_0 + \phi_1)(y_p) + \mathbf{w}_2^{\text{T}} \sum_{p,q} \phi_2(\mathbf{y}_{pq})$$

- parametric maxflow returns a linear number of solutions



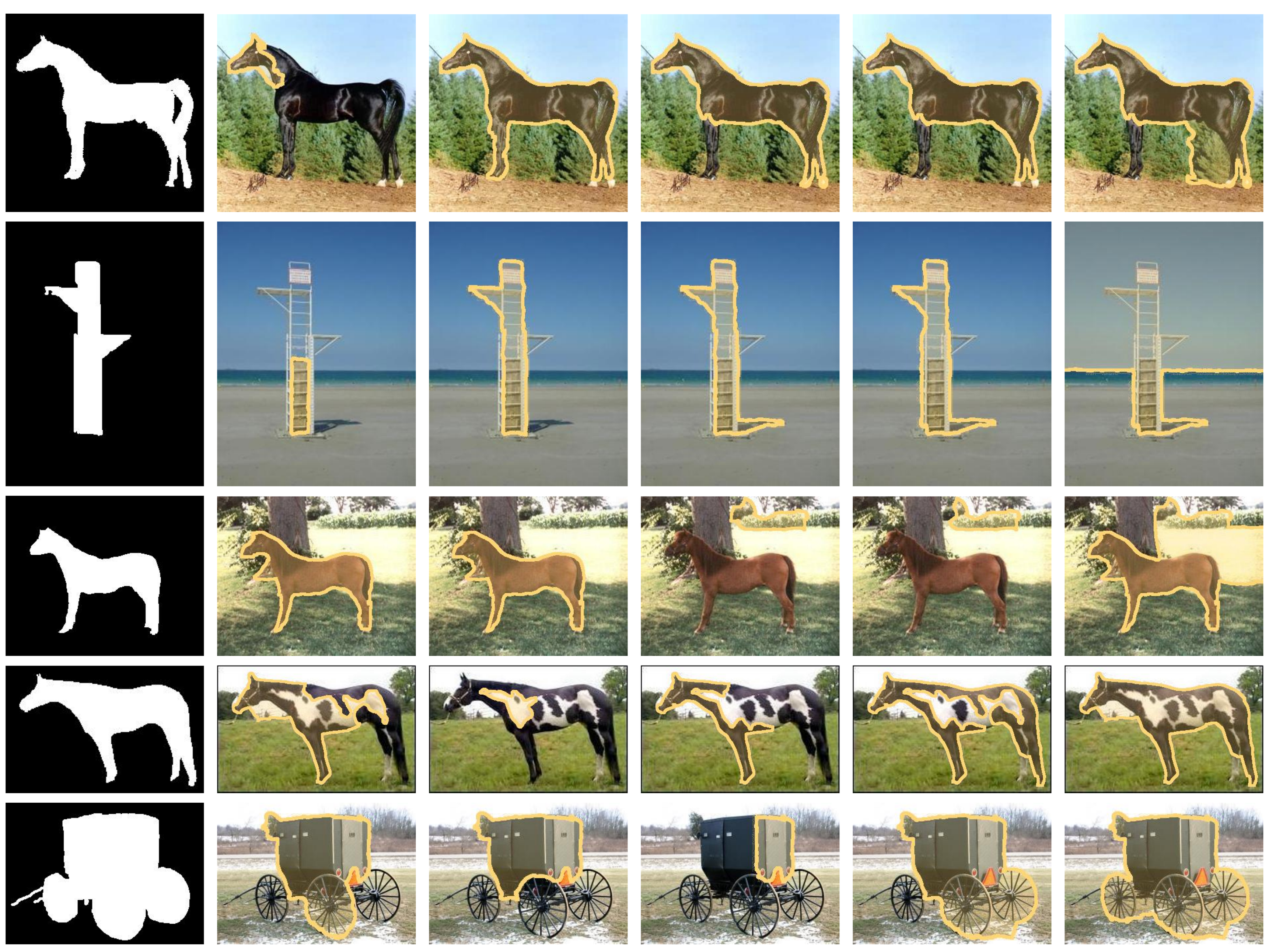
Learning

- solve a standard Structured SVM problem
- fix λ uniformly for loss function

Results

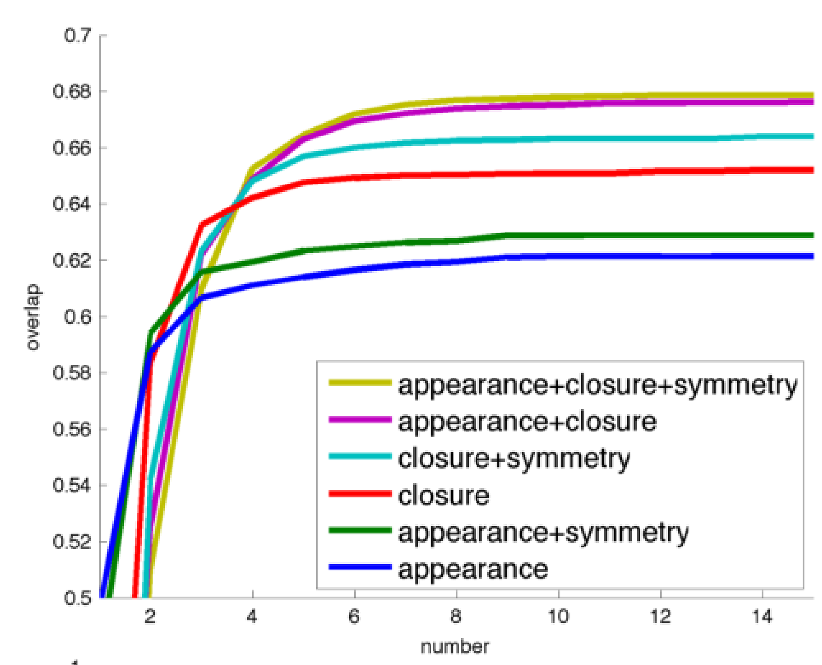
Mid-level cue transfer

- mid-level grouping cues are shared by all object classes
- cues trained on one object can be applied to other objects
- dataset: Weizmann Segmentations (WSD)



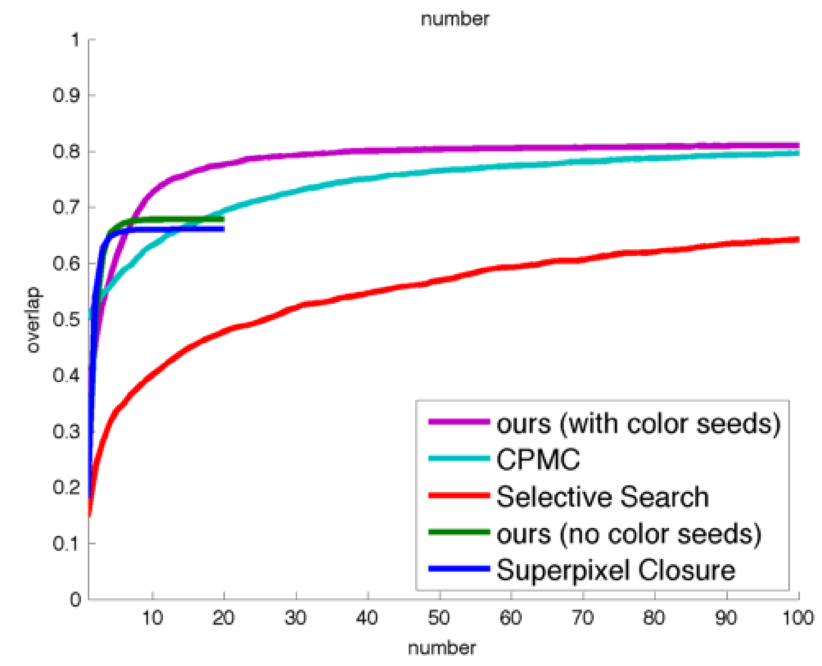
Cue combination

- determine the effect of combining grouping cues
- appearance, closure, and symmetry each contribute to higher quality regions



Region proposals

- compare our method with CPMC and Selective Search
- we improve precision when budget is limited to 100 proposals
- dataset: Weizmann Horses (WHD)



Conclusions

- mid-level regularities are ubiquitous and apply to all object classes
- grouping cues have the greatest impact when combined