ROYAL INSTITUTE
OF TECHNOLOGY

Strong Supervision from Weak Annotation: Interactive Training of Deformable Part Models

S. Branson, P. Perona, S. Belongie

Methods by level of supervision

- simple models, weak annotations
  - least effort, potentially fast learning, not best results
- complicated models, weak annotation
  - state-of-the-art performance (multiple instance learning, latent parts, latent structural SVM)
  - non-convex optimization
  - slow training
  - Hard to pinpoint error source (optimization error, inappropriate model or feature space, insufficient training data)

ROYAL INSTITUTE OF TECHNOLOGY

Methods by level of supervision

- Strong annotation
  - Very time consuming
  - Easy learning task (possible convex optimization)
  - Generalization guarrantee
  - Sensitive to quality and style of annotation

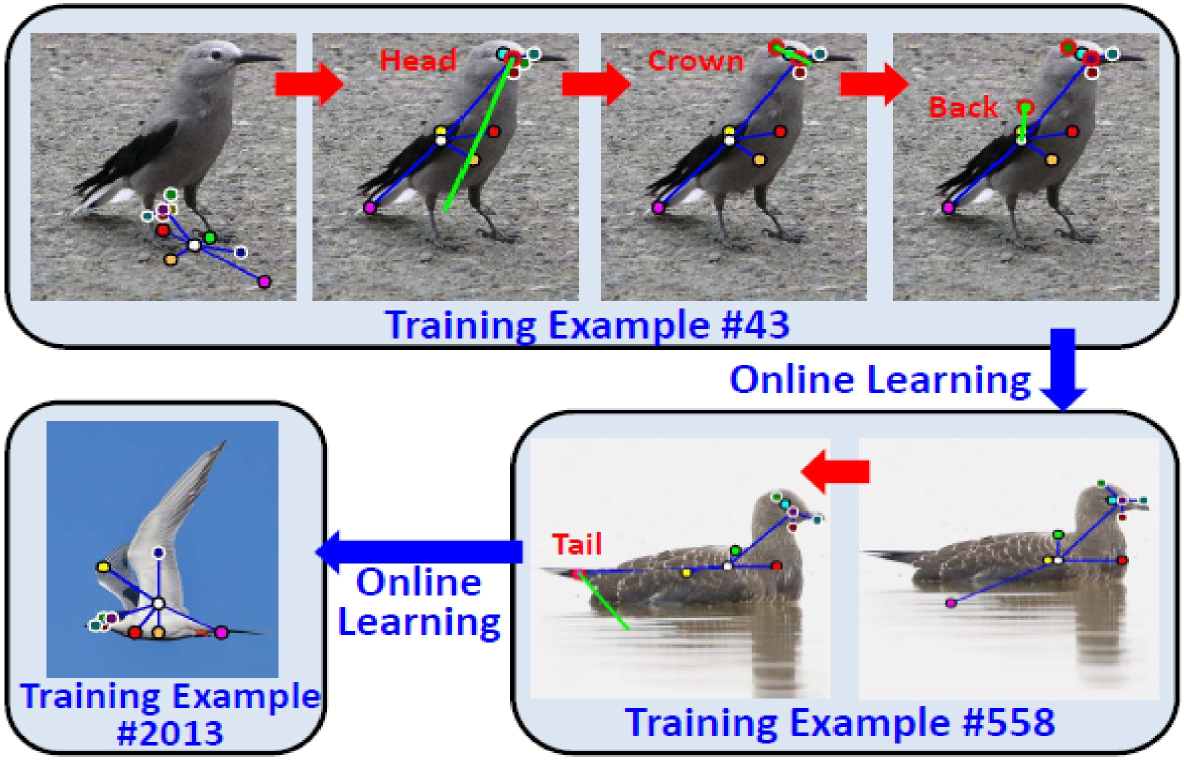
Qusetion: Is it possible to have strong supervision properties with weak annotation computational efficiency?

ROYAL INSTITUTE OF TECHNOLOGY

Yes! Interactive Labelling and Online Learning

- 1- Model part structures with structured models
- 2- Bring up a new image,
  - predict the part locations with current model
  - Correct the wrong locations
- 3-Update the learned model and go to 2

Interactive Labelling + Online Learning

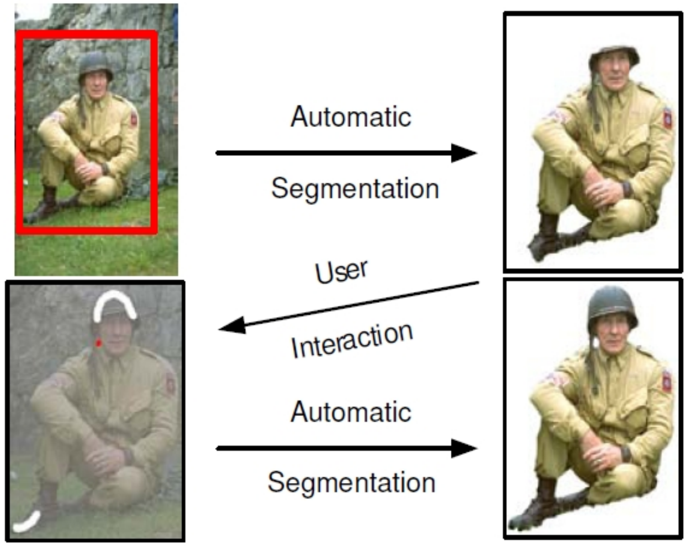


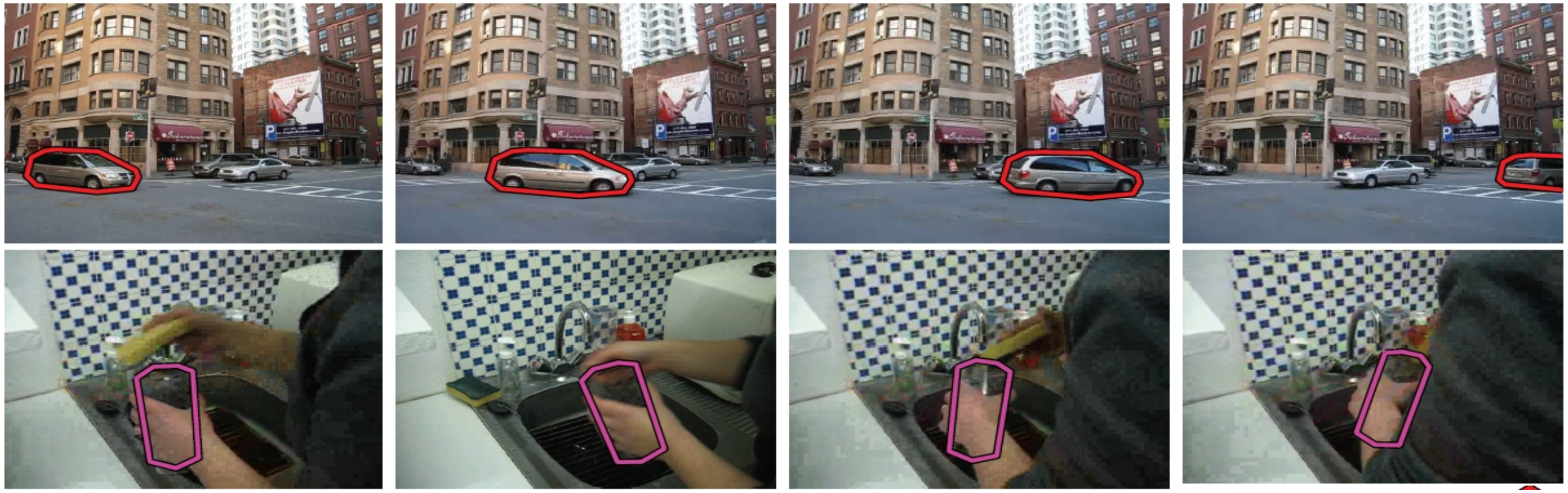
Interactive Labelling + Online Learning

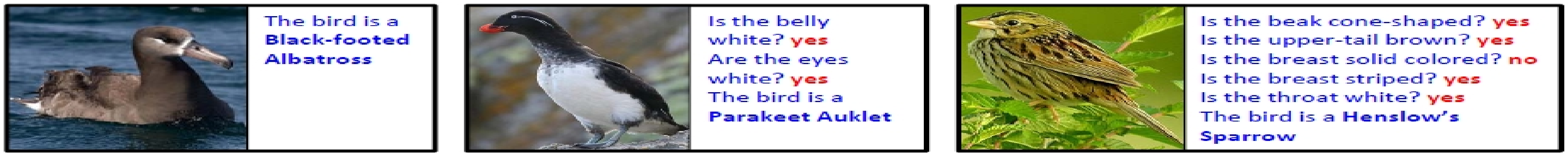
- Interactive Labelling
  - Real time detection
  - Easy update
  - Tree-structured deformable parts model with dynamic programming is a good choice!
- Online Learning
  - Fast model updating
  - Convex optimization
  - Stochastic gradient descent

Related works

- Interactive labelling
  - Grab cut (Segmentation)
  - Label me video
  - Visipedia (attributes)

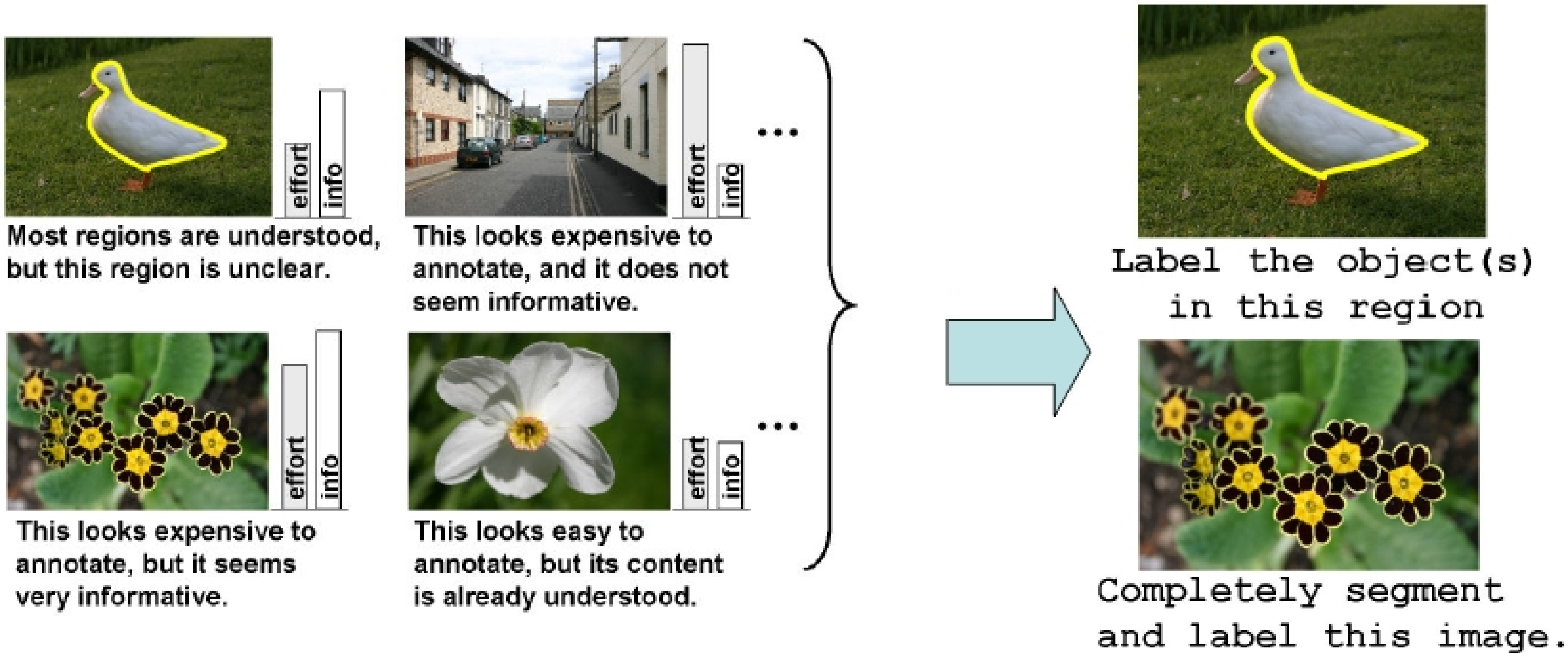






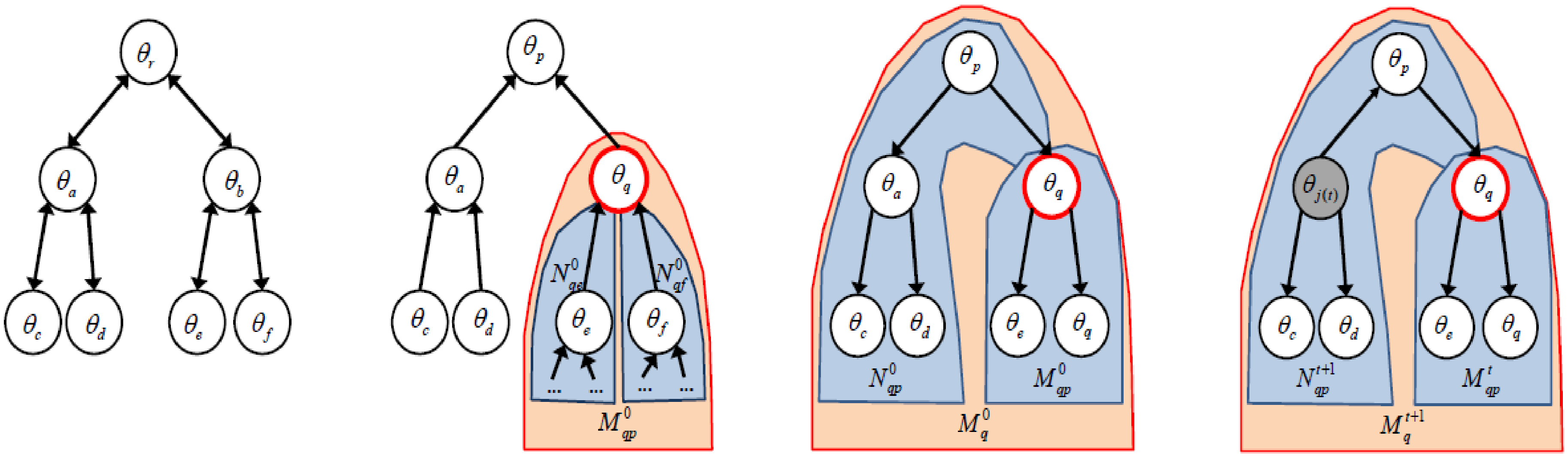
Related works

- Active learning
  - Intelligent computer decide which image to annotate
  - More savings than interactive labeling
  - In comparison to strong supervision
    - Higher computational complexity
    - Fewer theoretical guarantees



Model and Interactive UI

$$s(\Theta; x) = \sum_{p \in V} \psi_p(\theta_p; x) + \sum_{(p,q) \in E} \lambda_{pq}(\theta_p, \theta_q)$$



(a) Hierarchical Part Model (b) Bottom-Up Preprocessing (c) Top-Down Preprocessing (d) Propagate User Response

Learning framework

- Strong Convex formualtion

$$F_n(\mathbf{w}) = \frac{\lambda}{2}\|\mathbf{w}\|^2 + \frac{1}{n}\sum_{i=1}^{n}\ell_i(\mathbf{w}) \tag{16}$$

$$\ell_i(\mathbf{w}) = \max_y \left(\mathbf{w}\cdot\Phi(x_i, y) - \mathbf{w}\cdot\Phi(x_i, y_i) + \Delta(y_i, y)\right)$$

- Gradient computable
- by one inference

$$\begin{aligned}\bar{y}_i &= \max_y \left(\mathbf{w}\cdot\Phi(x_i, y) + \Delta(y_i, y)\right)\\ \nabla\ell_i &= \Phi(x_i, \bar{y}_i) - \Phi(x_i, y_i)\end{aligned}$$

- Stochastic gradient descent
- Process an image at each step
  - Pegasus!

Theoretical properties

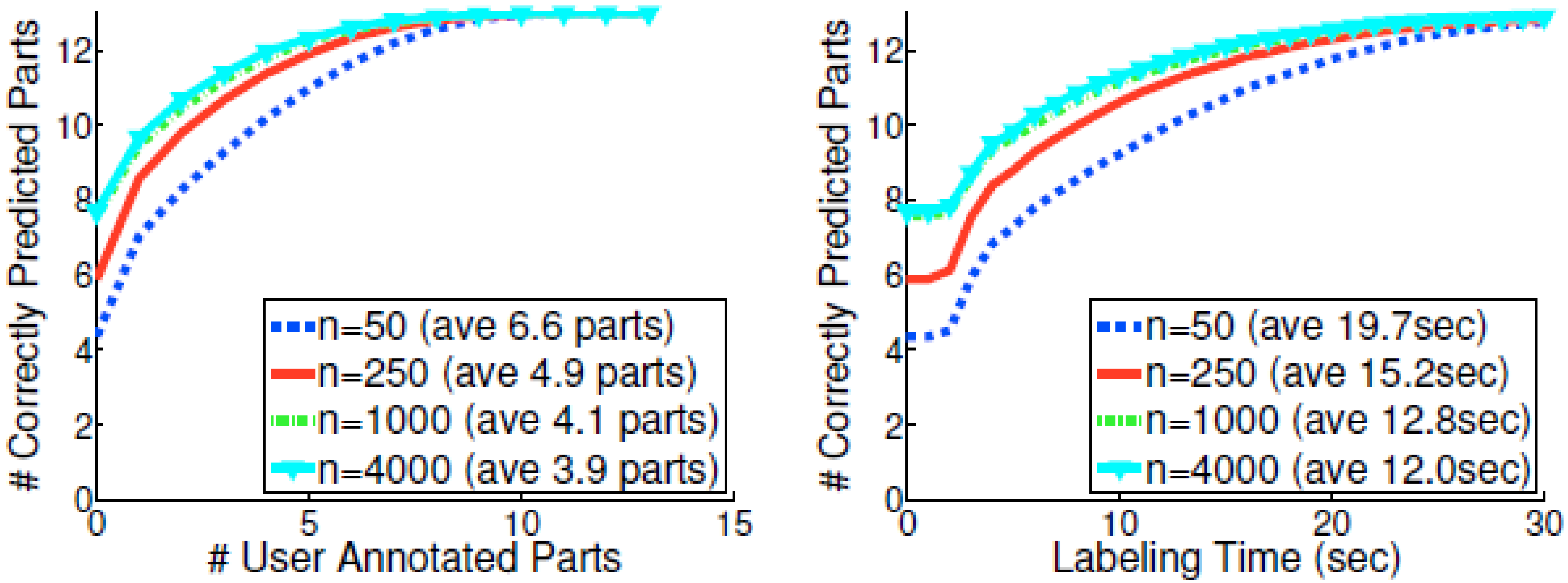
- Pegasos
  - Faster convergence rate than linear SVM
  - Performance guarantee with the number of iterations
  - Training time does not increase with increasing number of images (for a specific performance)
  - Slower steps than linear svm
    - Inference at each step
- Interactive labelling
  - Loss function is defined as the number of misplaced part
  - Number of annotation is bounded!

ROYAL INSTITUTE
OF TECHNOLOGY

Some results...

Tables

- 50 images:
  - 6.6 / 13 correction.
  - 19.7 seconds
- 4000 images:
  - 3.9 / 13 correction
  - 12 seconds



(a) Part Prediction By # Corrections (b) Part Prediction By Time

ROYAL INSTITUTE OF TECHNOLOGY

Conclusion

- Framework for large scale annotation
- Simulataneous learning of structured models
- Nice theoretical properties, seen in practice

KTH VETENSKAP OCH KONST

ROYAL INSTITUTE OF TECHNOLOGY

Cross-category Object Recognition (CORE)

- University of Illinois at Urbana-Champaign

- more detailed models and for exploring cross-category generalization in object recognition

CORE – Data overview

- Images from ImageNet, thus coming with object hierarchy
- Binary attributes
  - Pose
  - Sorrounding context
  - Viewpoint
  - Etc.

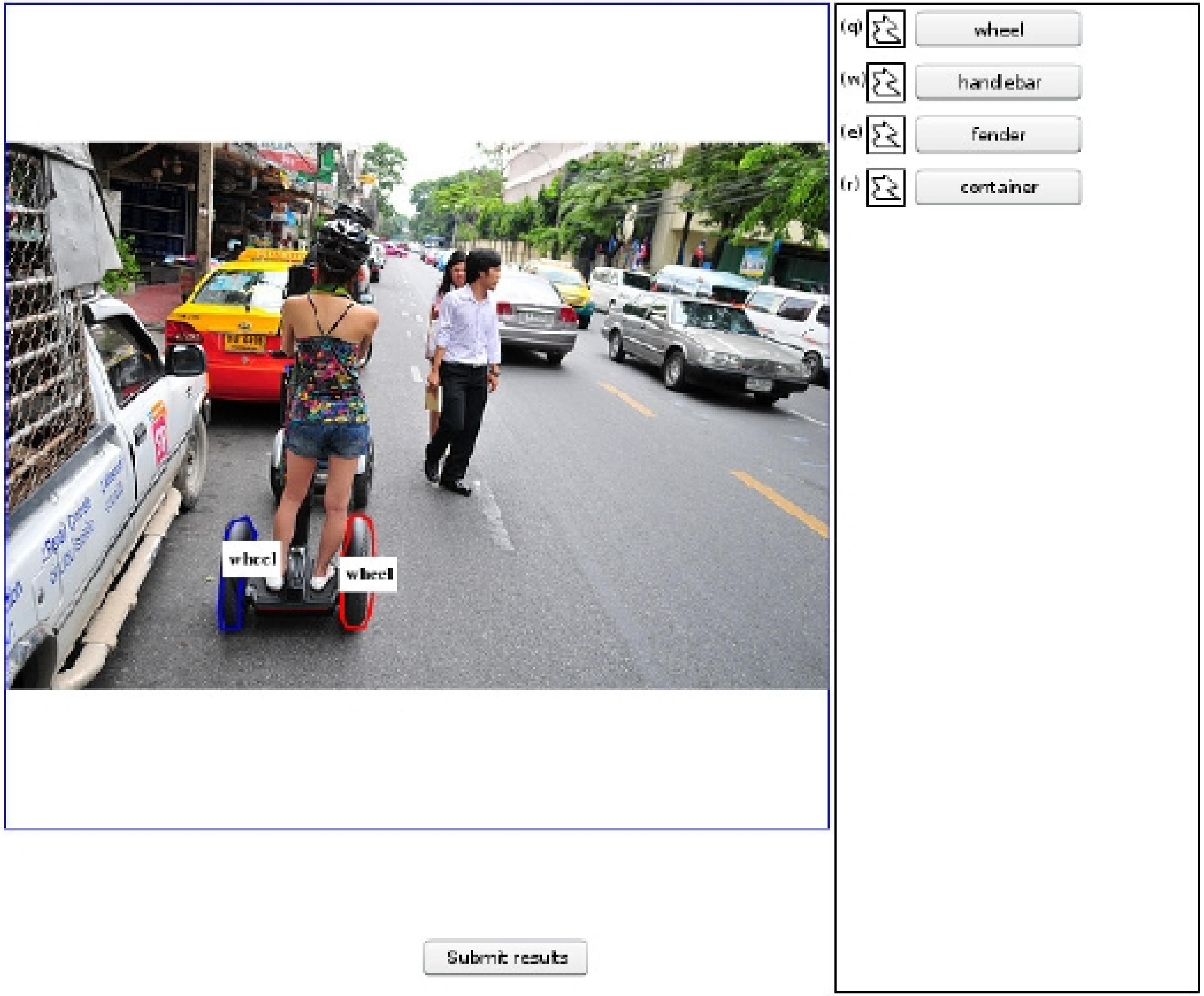
Property	Value	Unsure
Objects Top	○ Not Visible ○ Visible	☐
Objects Bottom	○ Not Visible ○ Visible	☐
Objects Front	○ Not Visible ○ Visible	☐
Objects Back	○ Not Visible ○ Visible	☐
Objects Left Side	○ Not Visible ○ Visible	☐
Objects Right Side	○ Not Visible ○ Visible	☐
Context	○ On Land ○ Flying	☐
Another object in this image	☐ Visible	☐
Visibility	○ Fully Visible ○ Partially Visible	☐
Unusual	○ Typical Example ○ Something Unusual ○ Too Small	☐

Submit Annotation

(a) Binary Attributes

CORE – data overview

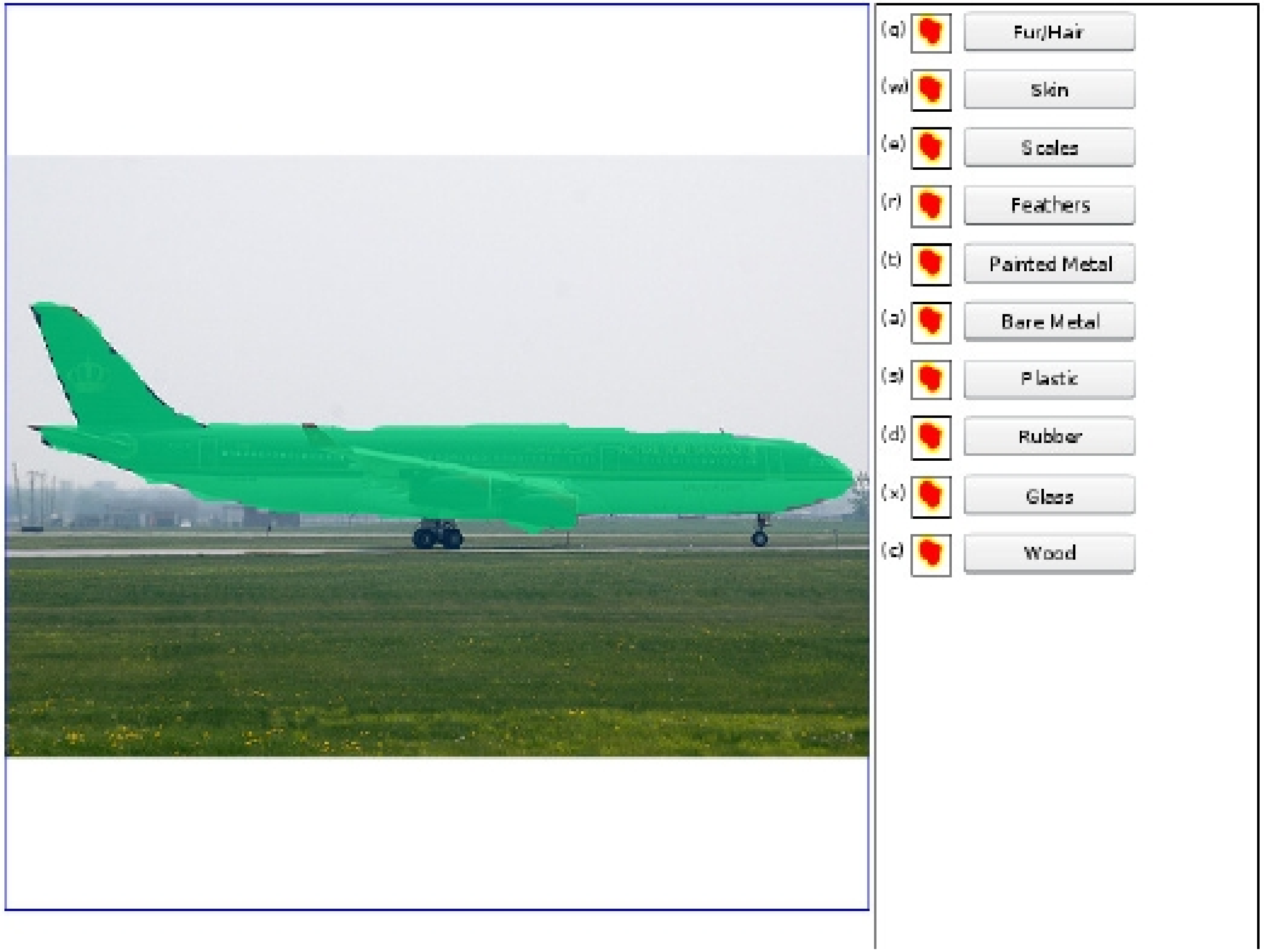
- Ploygon labels
  - Objects
  - Pre-defined parts of a category



(b) Polygons

CORE – data overview

- Segmentation mask
  - Materials



Quality Measures

- Way of collecting images
- Which attributes
  - Easily annotatable
  - Unsure buttun
- Quality assurance methods...