Beyond Active Noun Tagging: Modeling Contextual Interactions for Multi-Class Active Learning
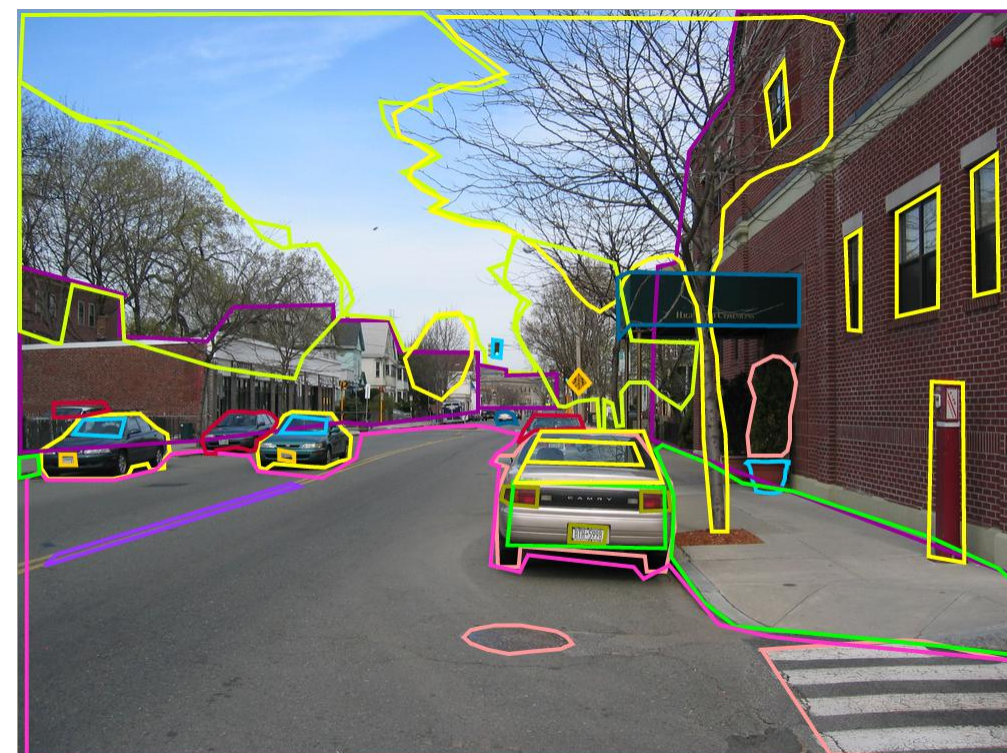## Behjat Siddiquie and Abhinav Gupta

Motivation

Scene Understanding
- Goal: Identify all the objects in a scene
- Appearance
  - Region features – color, texture, location and shape
- Relationships
  - Differential features
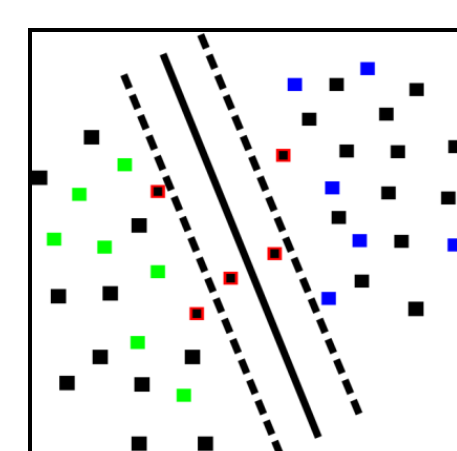 - *above, below, brighter, more green, more blue*

Learning a Model for Scene Understanding
- Learn O(n) appearance models
- Learn O(n²) relationship priors
- Heavy tailed distribution of object categories

Russell et al. IJCV, 2008
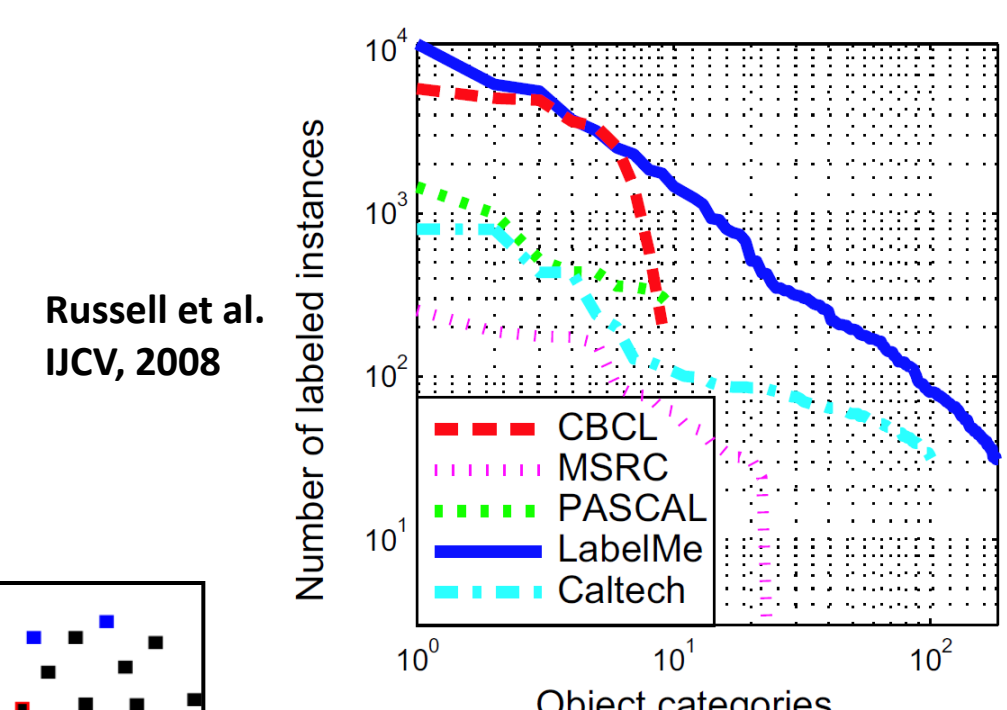
Active Learning
- Goal: minimize manual annotation
- Choose unlabeled samples which provide maximum information

Our Contributions
- Active learning approach for building an appearance and relationship model for scene understanding
- Image Entropy for active selection
  - Account for appearance uncertainty as well as contextual uncertainty
- Linguistic Questions
  - Use high confidence detections as anchors to ask questions about the unknown objects in a scene
- Contextual Questions
  - Learn general contextual relationships directly from the annotator
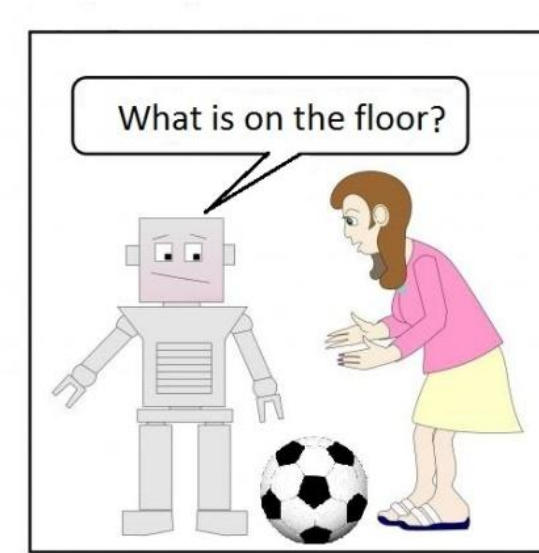
Approach
Image Entropy

Uncertainty based Approach
- Uncertainty is a measure of the goodness of the current model
- Model the uncertainty of the contextual relations between the regions in an image
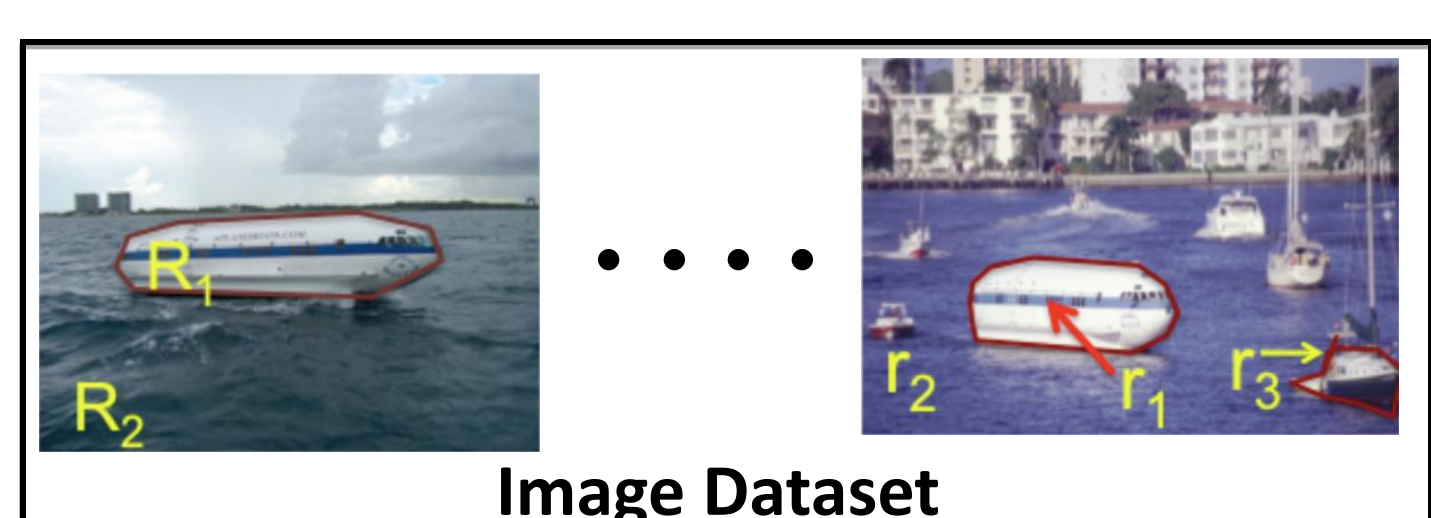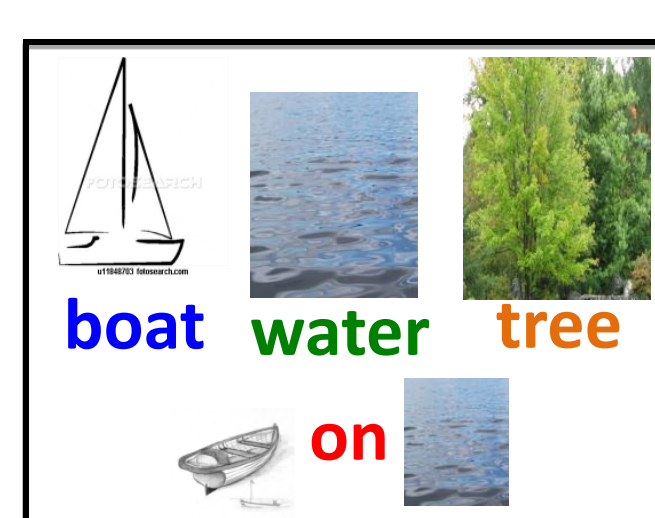

Region Entropy   Image Entropy
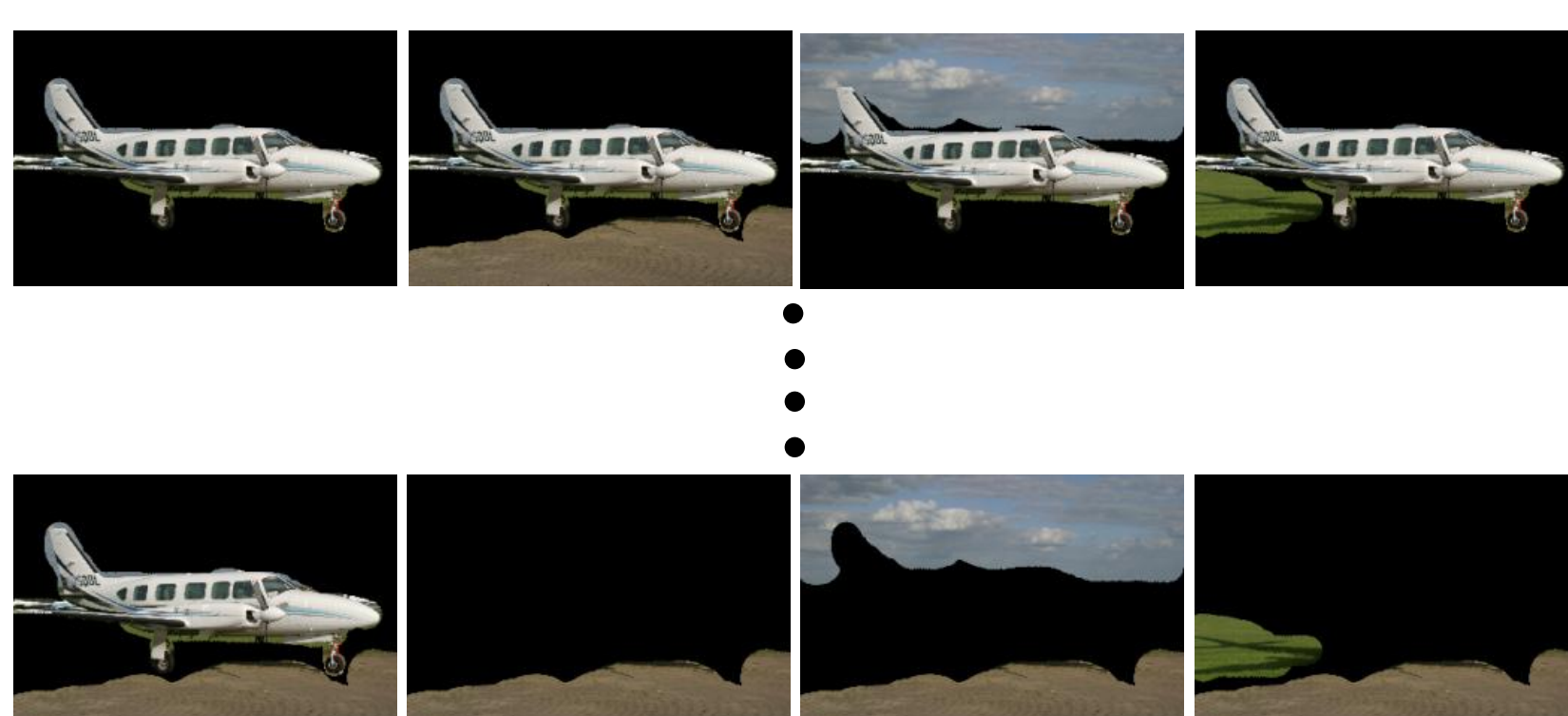
Image Dataset

Image Entropy
- Entropy over all possible label assignment probabilities

$$H(I) = \sum_{(n_1, n_2..)\in N} -P(n_1, n_2..|I) \log(P(n_1, n_2..|I))$$

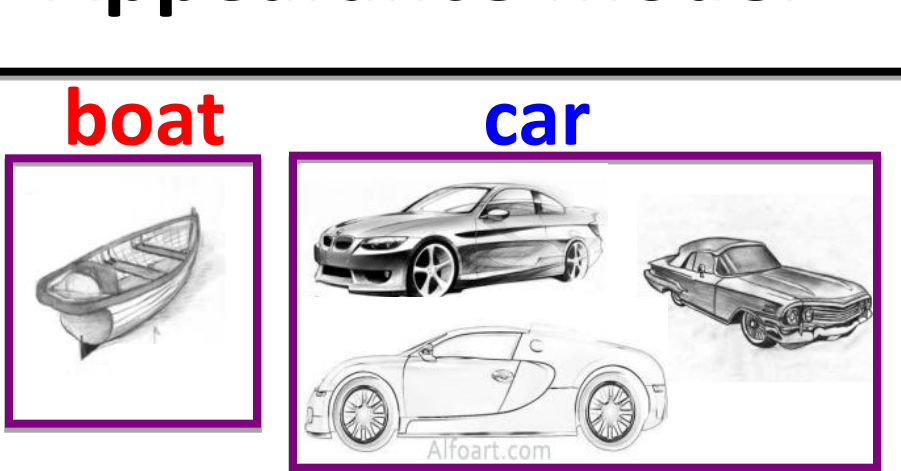#### Second order approximation
- Joint entropies of pairs of regions
- Captures contextual uncertainty as well

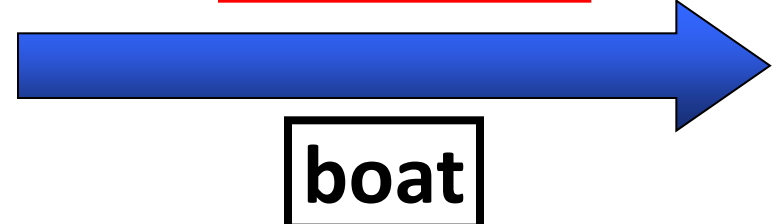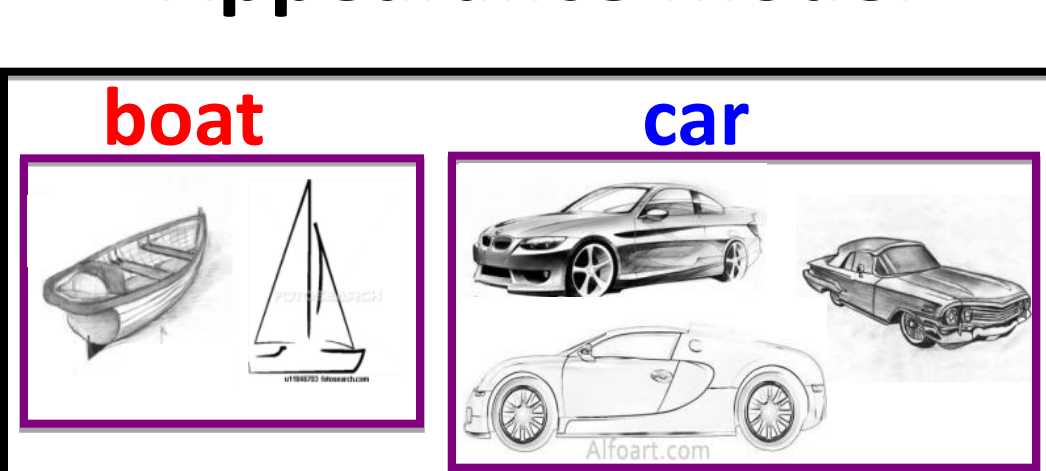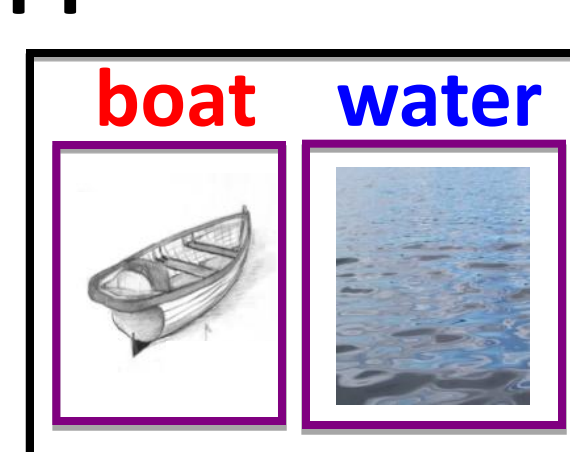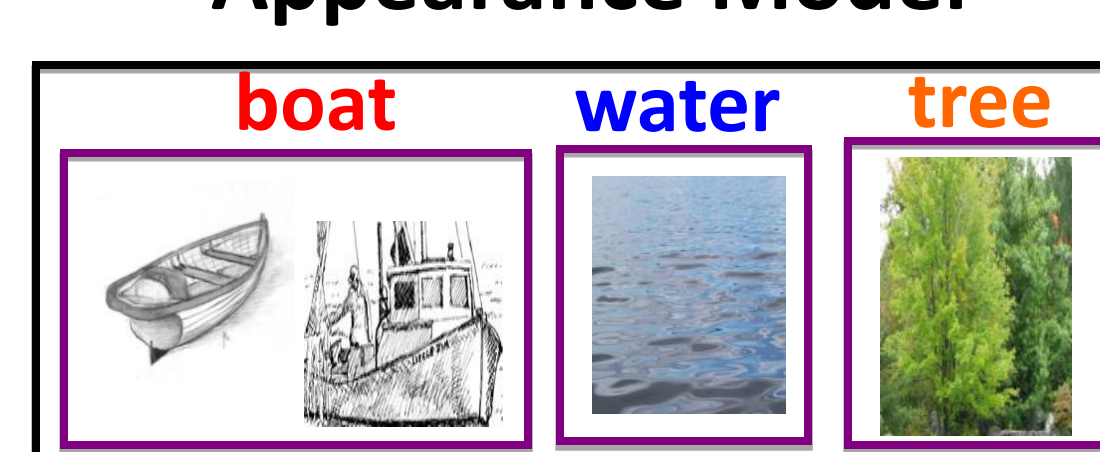### Region Labeling Questions

Appearance Model → Appearance Model

boat   car

boat

Linguistic Questions

Appearance Model
boat water   →   boat water tree

What is above the water?

1. Boat
2. Tree

Contextual Model
sky **above** grass
sky **bluer** than grass
⋮
car **above** road
→
sky **above** grass
sky **bluer** than grass
⋮
car **above** road
boat **above** water
tree **above** water
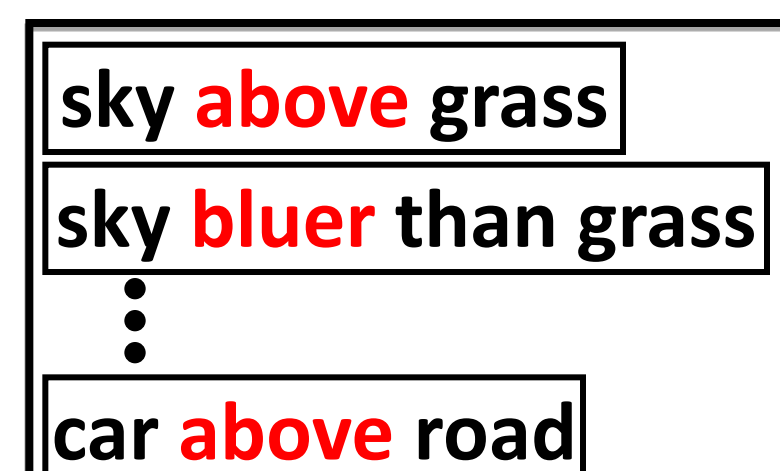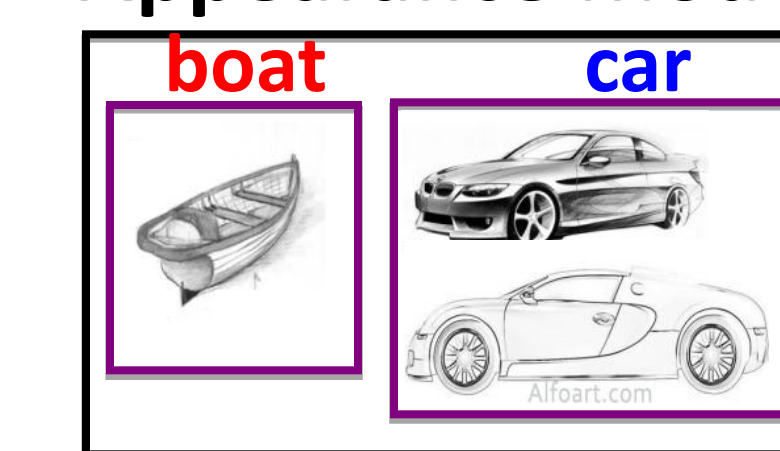
Contextual Questions

Appearance Model
boat car   →   boat car
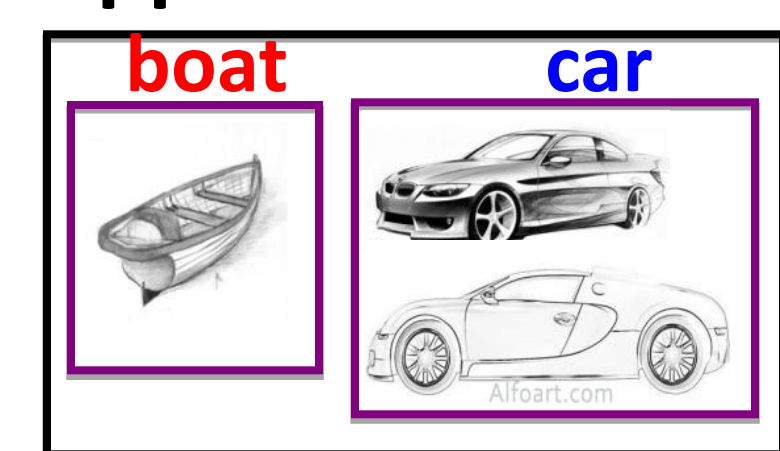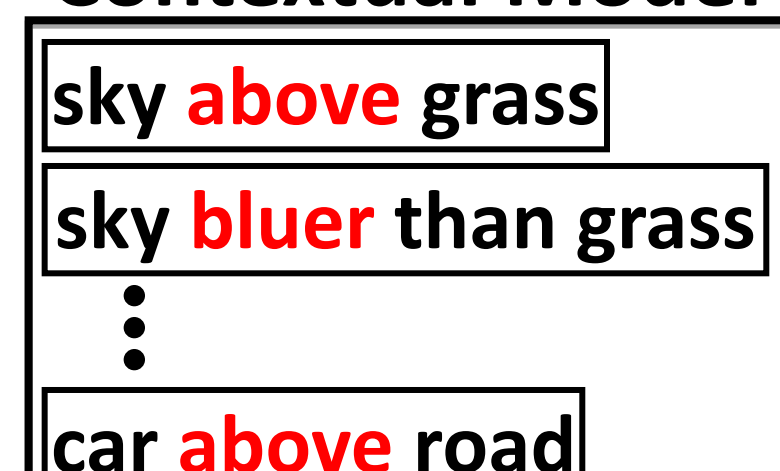
What is the relationship Between boat and water?

1. Boat **above** water
2. Water **bluer** than boat

Contextual Model
sky **above** grass
sky **bluer** than grass
⋮
car **above** road
→
sky **above** grass
sky **bluer** than grass
⋮
car **above** road
boat **above** water
water **bluer** than boat
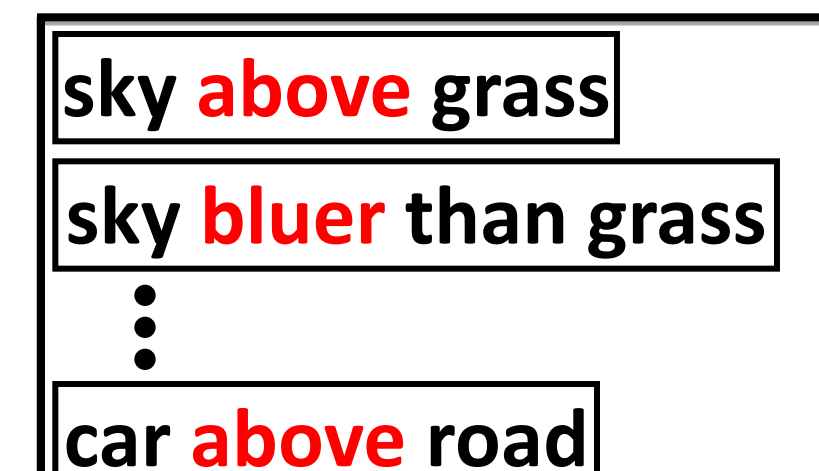
Expected Entropy Reduction

Region Labeling Questions

$H_{so}(I)$ $H_{so}(I|n_j = c)$

$$\Delta H_{so}(I, R_j) = \sum_{c\in\mathcal{C}} P(I_j|c, C_A)(H_{so}(I) - H_{so}(I|n_j = c))$$
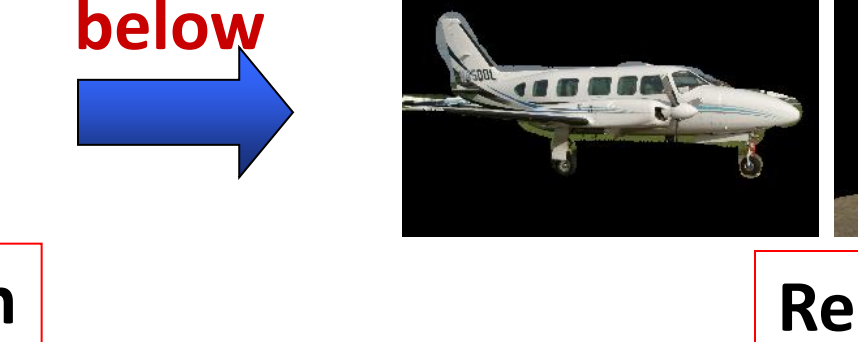
Linguistic Questions

What is **below** the sky?

airplane
grass
ground

$H_{so}(I)$ $H_{so}(I|n_{\mathcal{R}_q} \in \mathcal{C}_q)$
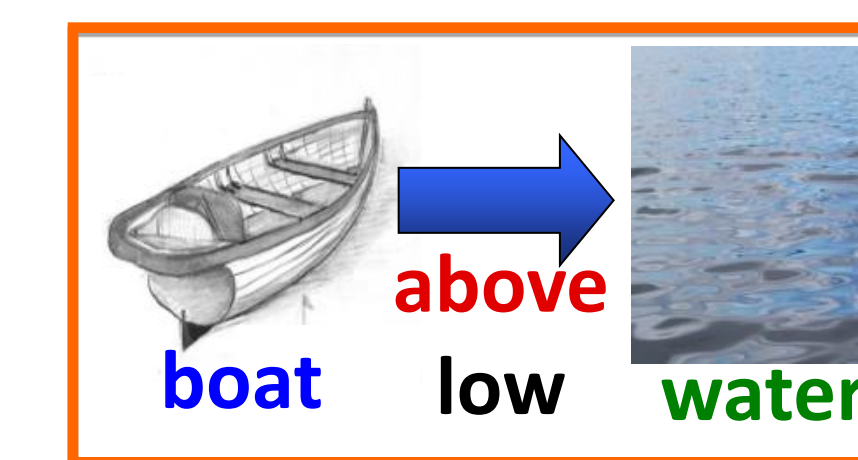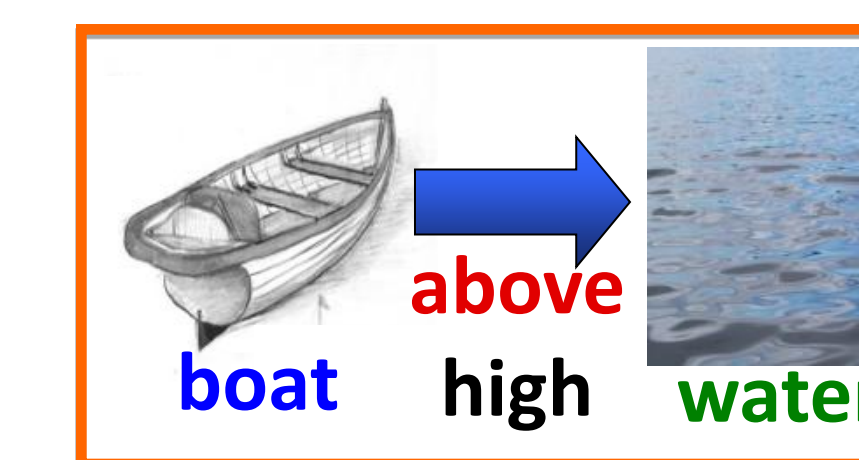
sky   **below**   Anchor Region   Relevant Regions

$$\Delta H_{so}(q, I) = \sum_{\mathcal{C}_q} P(\mathcal{C}_q|I)(H_{so}(I) - H_{so}(I|n_{\mathcal{R}_q} \in \mathcal{C}_q))$$

Contextual Questions
- Learn the correct relationship priors

boat **above** water low   boat **above** water high

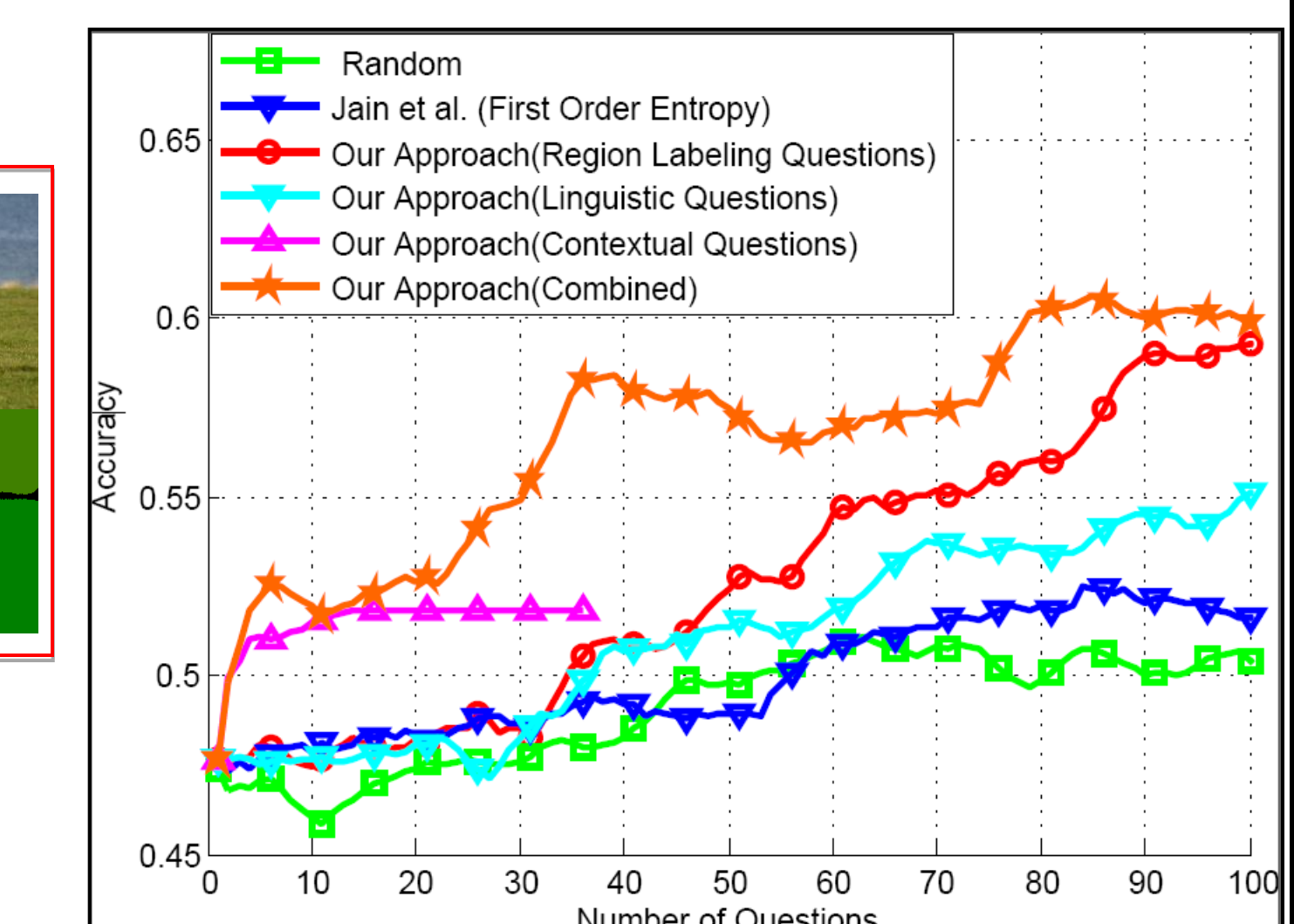$H_{so,low_{i,j,k}}(\mathcal{I}_U)$ $H_{so,high_{i,j,k}}(\mathcal{I}_U)$

$$\Delta H_{so}(r_k, n_i, n_j) = \max \begin{cases} H_{so}(\mathcal{I}_U) - H_{so,high_{i,j,k}}(\mathcal{I}_U) \\ H_{so}(\mathcal{I}_U) - H_{so,low_{i,j,k}}(\mathcal{I}_U) \\ 0 \end{cases}$$
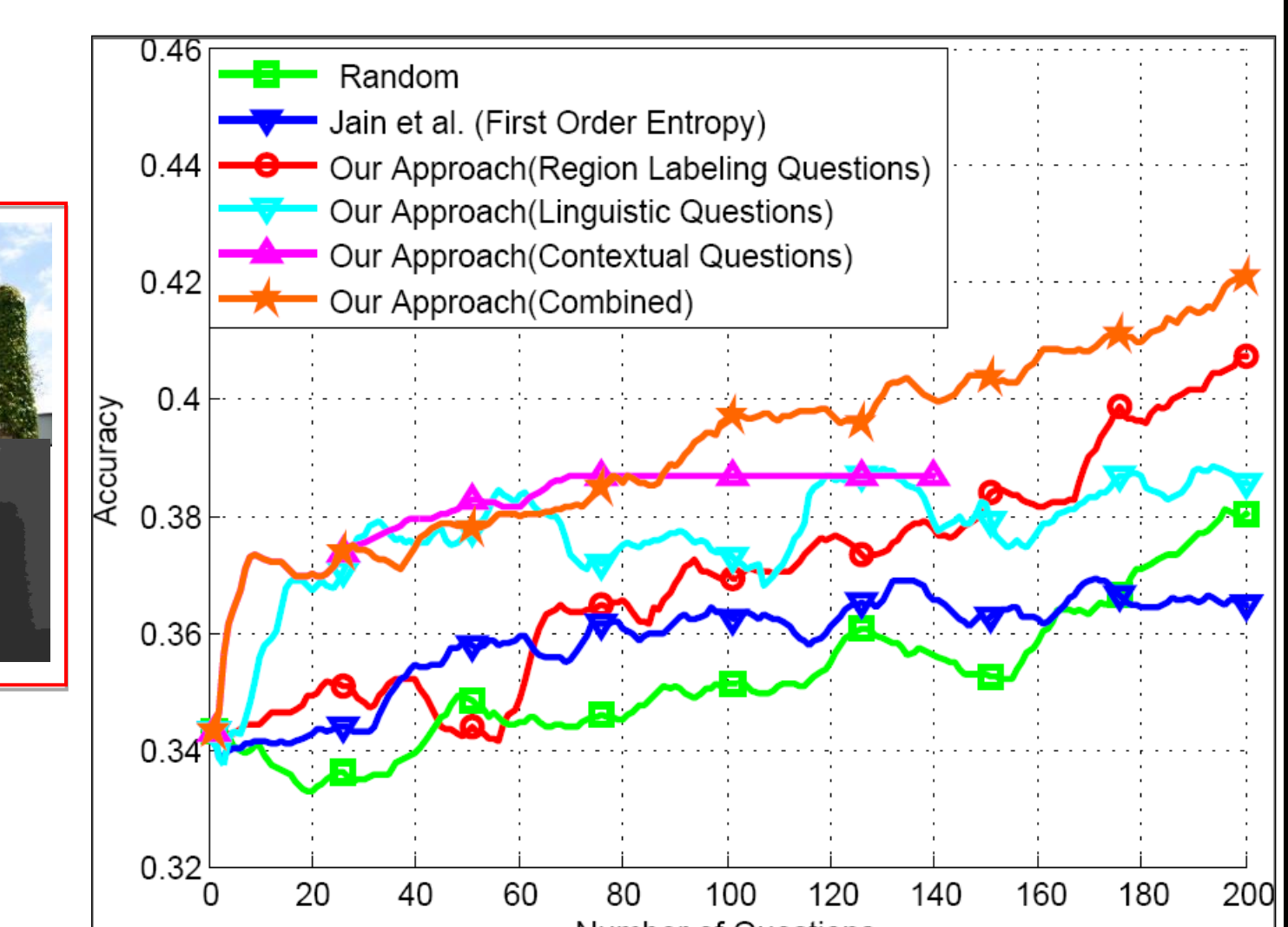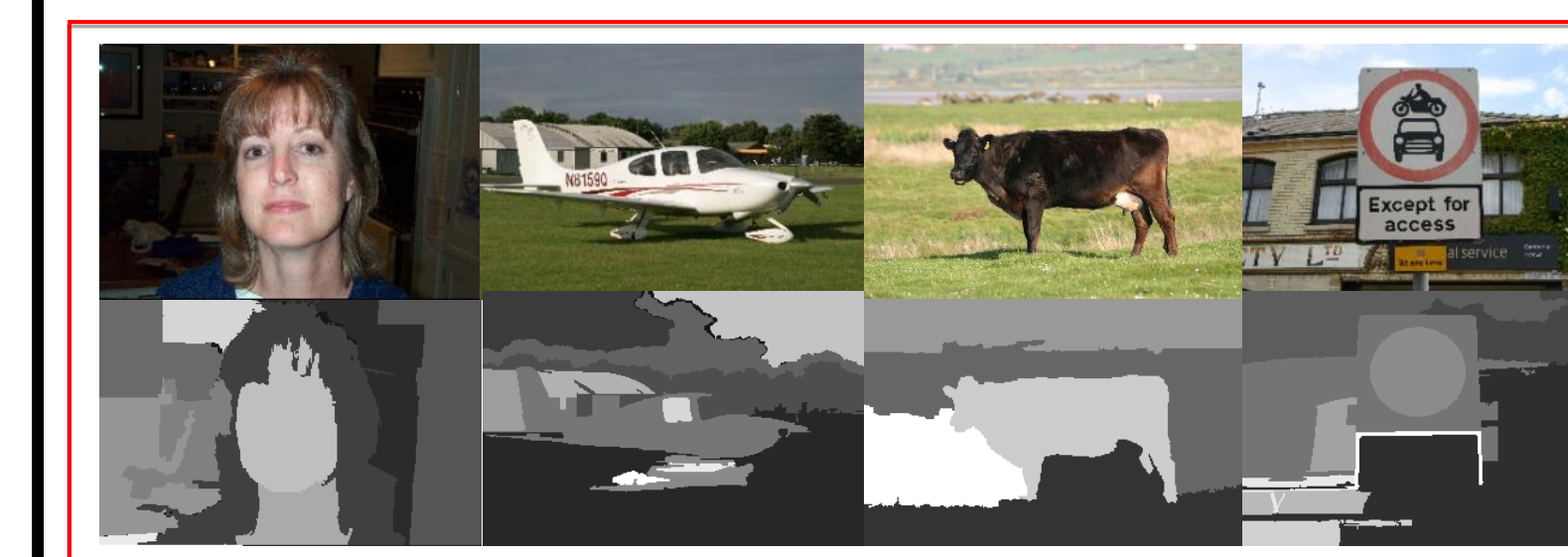
Results

MSRC Dataset

Ground Truth Segmentations
- Combined method better than individual questions
- Image Entropy outperforms region entropy
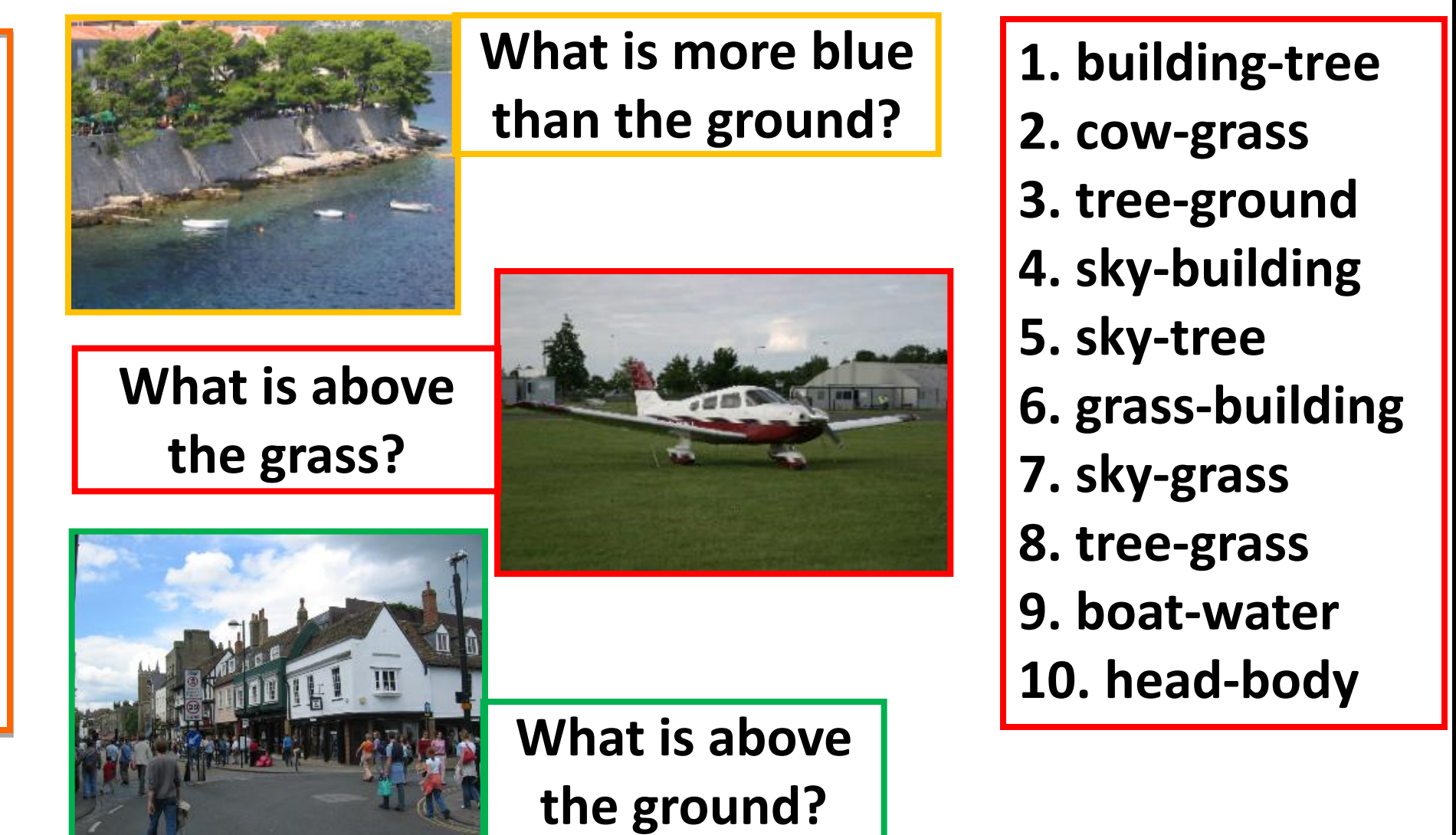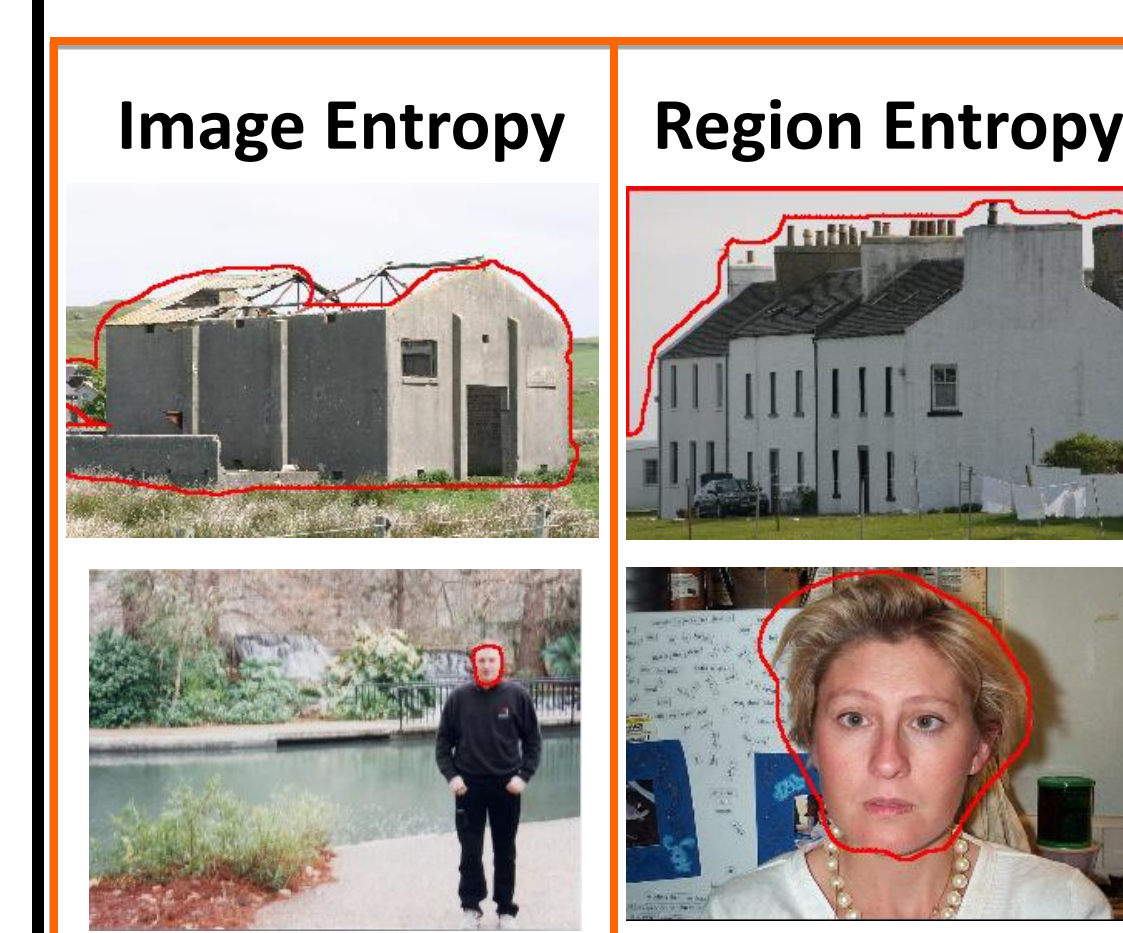- Order of Questions

Automatic Segmentations
- Difficulty in labeling during training and test phases
- Lower accuracy compared to Ground Truth
- Slow rate of increase of accuracy

Sample Questions
- Region Labeling Questions - Selects regions providing information about other regions
- Linguistic Questions - High confidence regions (e.g. "sky", "grass" & "ground") used as anchors
- Contextual Questions - Relevant pairs of object categories

Image Entropy   Region Entropy

What is more blue than the ground?

What is above the grass?

What is above the ground?

1. building-tree
2. cow-grass
3. tree-ground
4. sky-building
5. sky-tree
6. grass-building
7. sky-grass
8. tree-grass
9. boat-water
10. head-body

Stanford Dataset

Dataset
- 715 images (415 training, 300 test)
- Annotated using Amazon Mechanical Turk
- 8 categories and multi-label images

Results
- Linguistic questions – good anchors
- Contextual Questions – good initial priors