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 (c_i)_{i=1}^n appearance models
- Learn (r_{ij})_{i,j=1}^n relationship priors
- Heavy tailed distribution of object categories

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

Uncertainty based Approach
- Entropy over all possible label assignment probabilities
- \( H(I) = \sum_{(n_1,n_2) \in \mathcal{N}} -P(n_1,n_2,I) \log(P(n_1,n_2,I)) \)

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

Region Labeling Questions
- Appearance Model
- Linguistic Questions
- Contextual Questions

Expected Entropy Reduction
Region Labeling Questions
- \( H_{\text{old}}(I) \)
- \( H_{\text{old}}(I|n_j = c) \)
- \( \Delta H_{\text{old}}(I,R_j) = \sum_{c \in \mathcal{C}} P(I|c,R_j) (H_{\text{old}}(I) - H_{\text{old}}(I|n_j = c)) \)

Linguistic Questions
- \( H_{\text{old}}(I) \)
- \( H_{\text{old}}(I|n_R \in \mathcal{C}_j) \)
- \( \Delta H_{\text{old}}(q,I) = \sum_{c_i} P(C_i|I) (H_{\text{old}}(I) - H_{\text{old}}(I|n_{R_i} \in \mathcal{C}_j)) \)

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

Stanford Dataset
Dataset
- 715 images (413 training, 300 test)
- Annotated using Amazon Mechanical Turk
- 8 categories and multi-label images

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