To detect the whole object:

1. Detect the distinctive part (head).
2. Segment the animal based on distinctive part.
3. Obtain the animal bounding box from the segmentation and the distinctive part.

### Overview

**Motivation**

Cats and dogs are very varied in their imaged shape

- Key observation: DefPMs are 
  - very poor at detecting the whole cat.
  - very good at detecting a distinctive part such as the head.

**Problem Overview**

Deformable Part Models (DefPMs) such as [Felzenszwalb et al. 09] are not deformable enough for very flexible objects such as cats and dogs.

We extend DefPMs by introducing Distinctive Parts Models (DisPMs), which combine DefPMs with segmentation and successfully detect highly deformable animals.

### Dataset and Annotations

- **Object**
- **Part**
- **Trimap**

- PASCAL VOC 2010 annotations: cats and dogs
- Bounding boxes
- Segmentations (trimaps)
- Additional head annotations
- Bounding boxes

### 1) Part Model

The distinctive part is modeled by a DefPM.

- Parts connected by springs
- HOG + LBP features
- Fast inference with dynamic programming
- Discriminatively trained by Latent variable SVM

### 2-a) Grab-Cut Segmentation

- Alternate two steps [Rother et al. 04]
- Estimate appearance model
- Minimize standard graph-cut energy function to assign a correct label $y_i$ to each pixel $x_i$

$$E(x, y) = \sum_{i} \log P(y_i | x_i) + \sum_{i<j} S(y_i, y_j | x)$$

- **Appearance Model**
  - Data term
    - from GMM estimated using head region and from global posterior probabilities.
    - Pairwise term from edge detector output
  - Model initialization
    - Foreground from head detection.
    - Background from predicted bounding box.
    - Model update re-estimating GMMs using the output from the previous step.

### 2-b) Modeling Edges

- Berkeley PB edge detector [Martin et al. 04]
- Edge detector response used as pairwise term.

$$S(y_i, y_j | x) = \gamma e^{-c(y_i(x)/\beta)}$$

- Correct Segmentation
- Incorrect Segmentation

### 2-c) Global Model

- Head seed insufficient to model object foreground.
- Posterior probabilities given by histograms learnt from ground truth segmentations over all training images.
- Used as data term for the first iteration of segmentation.

### 3) Post Processing

- Clean up: erode, dilate, and select component connected to head.
- Adjust predicted bounding box to be consistent with head detection.
- Rerank detections based on the head size.