Technical Report - Relating Things and Stuff by High-Order Potential Modeling

Byung-soo Kim*[1], Min Sun*[1], Pushmeet Kohli[2], and Silvio Savarese[1]

[1]University of Michigan, Ann Arbor, U.S.A
[2]Microsoft Research Cambridge, UK
* indicates equal contributions

1 Associating Object Instances with Segmentation Mask

Segmentation mask is a $\mathcal{V}_j$, which is the set of image elements associated with the object. In specific, $\{x_i\} \in \mathcal{V}_j$ is the set which satisfies $x_i \cap \mathrm{BBox}(j) \geq 50\%$, where x_i are super-pixels.

2 Loss Function for Structural SVM Learning

For the experiment on Stanford dataset, since the performance is measured by the average classification accuracy across different object categories, we define the following loss function. The overall loss function $\triangle(X, Y; X^n, Y^n)$ is decomposed into sum of the segmentation loss $\triangle(X; X^n)$ and the detection loss $\triangle(Y; Y^n)$.

The segmentation loss $\triangle(X; X^n)$ is defined as

$$\triangle(X; X^n) = \frac{1}{Q} \sum_{i \in \mathcal{V}} \mathbf{1}\{x_i \neq x_i^n\} c_x(l_i) \ , \tag{1}$$

where $\mathcal{V}$ captures the indices for the set of segments, $\mathbf{1}\{STATEMENT\}$ is 1 if the $STATEMENT$ is true, $c_x(l_i)$ is the object category l_i dependent cost (used to re-weight the loss contributed from different object categories), and $Q = \sum_{i \in \mathcal{V}} c_x(l_i)$. Therefore, the overall segmentation loss can be decomposed into a sum over local loss for each segment $\frac{1}{Q}\mathbf{1}\{x_i \neq x_i^n\} c_x(l_i)$.

The detection loss $\triangle(Y; Y^n)$ is defined as

$$\triangle(Y; Y^n) = \frac{1}{M} \sum_{i \in \mathcal{B}} \mathbf{1}\{y_i \neq y_i^n\} c_y(l_i) \ , \tag{2}$$

where $\mathcal{B}$ captures the indices for the set of detections, $M = \sum_{i \in \mathcal{B}} c_y(l_i)$. Similarly, the overall detection loss can be decomposed into a sum over local loss for each detection $\frac{1}{M}\mathbf{1}\{y_i \neq y_i^n\} c_y(l_i)$.