Improving Object Detection with Deep Convolutional Networks via Bayesian Optimization and Structured Prediction

Yuting Zhang†, Kihyuk Sohn†, Ruben Villegas†, Gang Pan†, Honglak Lee†

1. Revisiting R-CNN

- Convolutional neural network (CNN) for object detection
- The state-of-the-art "Regions with CNN" (R-CNN)

R-CNN: Regions with CNN features

2. R-CNN with structured objective (StructObj)

- Training set: \{x, y\}, \(x\) is the input image, and \(y\) is its annotation (if applicable).
- Linear classifier: \(f(x,y) = w^T x\).
- Structured objective: \(\min_{w} \sum_{(x,y) \in A} \phi(x,y, f(x,y))\)
  - Classification constraints
  - Localisation constraints

\[ \phi(x,y, f(x,y)) = \begin{cases} 0 & \text{if } f(x,y) \text{ is correct} \\ \|y - f(x,y)\| & \text{otherwise} \end{cases} \]

3. Fine-grained search (FGS) for bounding boxes via Bayesian optimization

4. Experiments

- a. mAP on VOC 2007 & 2012 test set
- b. FGS efficiency
- c. Localization accuracy on VOC 2007 test set
- d. Examples with large improvement on VOC 2007 test set


References