Learning Detection with Diverse Proposals

Samaneh Azadi1, Jiashi Feng2, Trevor Darrell1
1UC Berkeley, 2National University of Singapore

Abstract and Motivation

How do modern object detection architectures learn to localize objects?

Training:
- Minimize deviations from ground truth,
- Ignore correlation between multiple proposals and different categories.

Inference:
- Use NMS to prune proposals,
- Ignore label- and instance-level relations between proposals.

Our proposed method, Learning Detection with Diverse Proposals (LDDP):
- Improves location and category specification of final detected bboxes through:
  - label-level contextual information,
  - spatial layout relationships between multiple proposals,
- Does not increase # of parameters of the network,
- Achieves superior performance over Faster R-CNN even with ~ 30% of the generated proposals.

Learning with Diverse Proposals

Determinantal Point Process (DPP):
- A point process \( P \) is called determinantal if:
  \[
  P_k(Y = Y) = \frac{\det[\Phi_{ij}]}{\det[\Phi_{ii}]},
  \]

Learnable DPP layer for object detection:

\[
\min_{P_0} L(n) = \log \prod_i P_i(X^i|X^<i) - \sum_i \left[ \log P_i(Y^i|X^i) - \log P_i(B|X^i) \right]
\]

\( Y \): precise and diverse set of boxes,
\( B \): background boxes,
\( X \): list of object proposals as output of the RPN network.

- Posterior probability \( P_i(Y^i|X^i) \) modeled as a DPP:
  \[
  L_{ij} = \phi_{ij} | \phi_{ij} \theta_{ij} = \log U_{ij} \times \sim \text{i}_{ij} \text{ if } i \in Y\]
  \[
  \text{IoU}(C_i, C_j) = \frac{\text{IC}(C_i \cup C_j)}{\text{IC}(C_i) + \text{IC}(C_j)}.
  \]
- Increase the scores of representative boxes in their ground-truth label and background boxes in background label
- e.g. quality of boxes for the first term \( \log P_i(Y^i|X^i) \):
  \[
  \Phi_i = \left\{ \begin{array}{ll}
  \text{IoU}_{gt} \times \exp(W_i^Tf_i) & \text{if } i \in Y \\
  \text{IoU}_{gt} \times \sum_{c \in C} \exp(W_i^Tf_i) & \text{if } i \notin Y
  \end{array} \right.
  \]

Inference with Diverse Proposals

- A greedy optimization algorithm based on a similar DPP,
- Quality of boxes, \( \Phi \), as per class prediction scores:
  \[
  \Psi_i = \frac{\exp(W_i^Tf_i)}{\sum_c \exp(W_i^Tf_c)}
  \]

Experiments

- Higher Precision and Recall:
  - Pascal VOC2007 test detection avg precision(%)
    \[
    \begin{array}{cccc}
    \text{Method} & \text{cat} & \text{cow} & \text{horse} & \text{sheep} & \text{...} & \text{mAP} \\
    \text{(FrRCNN, NMS)} & 71.5 & 66.3 & 61.5 & 76.7 & 53.3 & 60.45 \\
    \text{(LDDP, LDDP)} & 74.9 & 66.6 & 68.5 & 77.4 & 58.8 & 62.21
    \end{array}
    \]
  - MS COCO val detection avg precision and avg recall(%)
    \[
    \begin{array}{cccc}
    \text{Method} & \text{Avg Prec @ Area} & \text{Avg Prec @ Area} & \text{Avg Rec @ Area} \\
    \text{S M L} & \text{S M L} & \text{S M L} \\
    \text{(FrRCNN, NMS)} & 15.0 & 31.5 & 12.7 & 15.1 & 21.8 & 6.0 & 24.2 & 38.9 \\
    \text{(LDDP, LDDP)} & 15.5 & 32.2 & 13.4 & 15.8 & 24.7 & 6.8 & 27.3 & 43.2
    \end{array}
    \]
- Non-Redundant Diverse Proposals: