LEARNING DETECTION WITH DIVERSE PROPOSALS

SAMANEH AZADI[1], JIASHI FENG[2], TREVOR DARRELL[1]
[1]UC BERKELEY, [2]NATIONAL UNIVERSITY OF SINGAPORE

IEEE 2017 Conference on Computer Vision and Pattern Recognition
CVPR July 21-26 HONOLULU 2017

ABSTRACT AND MOTIVATION

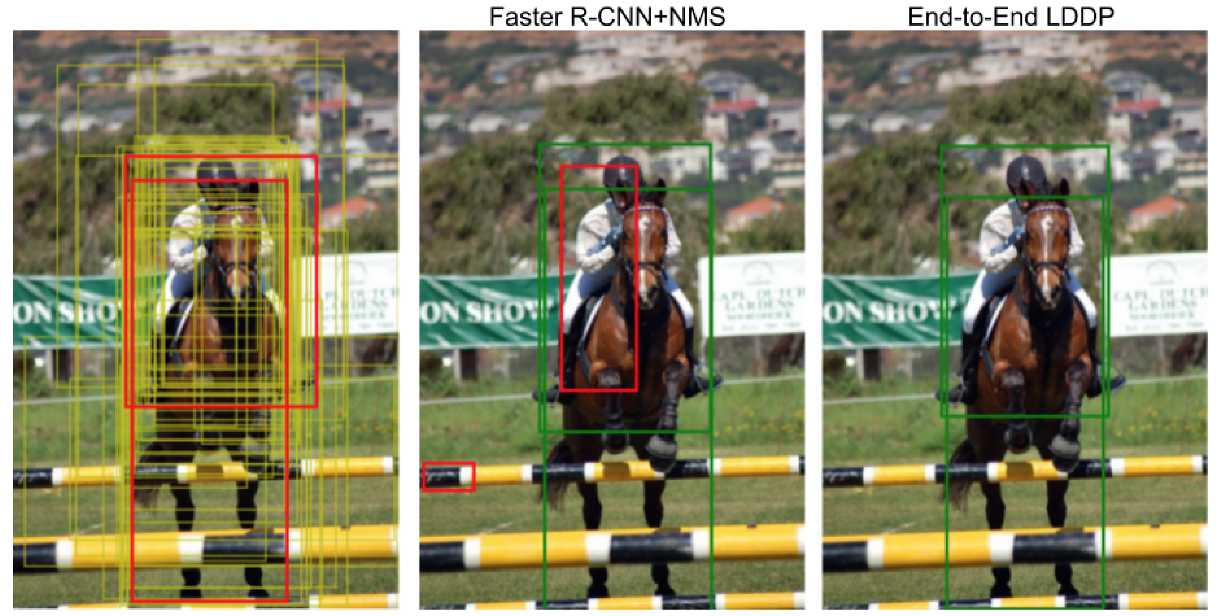
How do modern object detection architectures learn to localize objects?

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

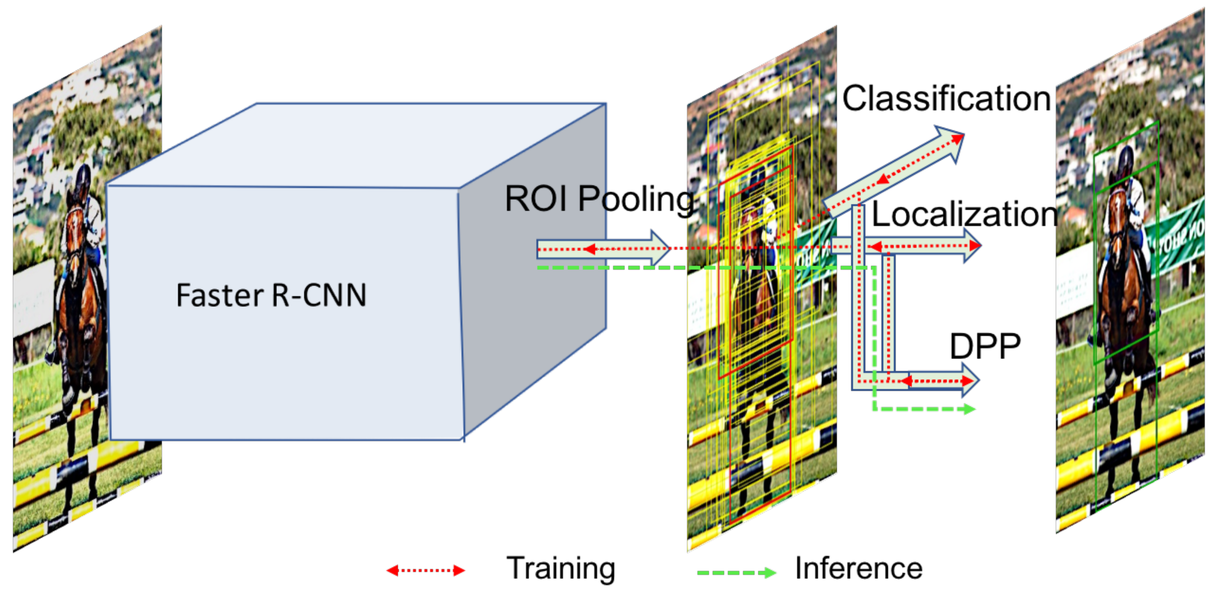
Inference:
- Use NMS to prune proposals,
- Ignore label- and instance-level relations between proposals.

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 $\sim 30\%$ of the generated proposals.



END-TO-END LDDP MODEL



LEARNING WITH DIVERSE PROPOSALS

Determinantal Point Process (DPP):

- A point process $\mathcal{P}$ is called determinantal if: $P_L(\mathbf{Y}=Y)=\frac{\det(L_Y)}{\det(L+I)}$.

Learn-able DPP layer for object detection:

$$\min_{\alpha} \mathcal{L}(\alpha)=\log\prod_i \frac{P_\alpha(Y^i|X^i)}{P_\alpha(B^i|X^i)}=\sum_i\left[\log P_\alpha(Y^i|X^i)-\log P_\alpha(B^i|X^i)\right]$$

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

- Posterior probability $P_\alpha(Y|X)$ modeled as a DPP:

$$L_{i,j} = \Phi_i^{1/2} S_{ij} \Phi_j^{1/2}, \quad S_{ij}=\text{IoU}_{ij}\times \text{sim}_{ij}$$

$$\text{IoU}_{ij} = \frac{A_i\cap A_j}{A_i\cup A_j}, \quad \text{sim}_{ij}=\frac{2IC(l_{cs}(C_i,C_j))}{IC(C_i)+IC(C_j)}.$$

- 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 $\log P_\alpha(Y|X)$:

$$\Phi_i=\begin{cases}\text{IoU}_{i,gt^i}\times \exp\{W_{gt}^T f_i\}, & \text{if } i\in Y\\ \text{IoU}_{i,gt^i}\times \sum_{c\neq 0}\exp\{W_c^T f_i\} & \text{if } i\notin Y\end{cases}$$

INFERENCE WITH DIVERSE PROPOSALS

- A greedy optimization algorithm based on a similar DPP,
- Quality of boxes, Φ, as per class prediction scores: $\Phi_i=\frac{\exp\{W_c^T f_i\}}{\sum_{c'}\exp\{W_{c'}^T f_i\}}$

EXPERIMENTS

- **Higher Precision and Recall:**

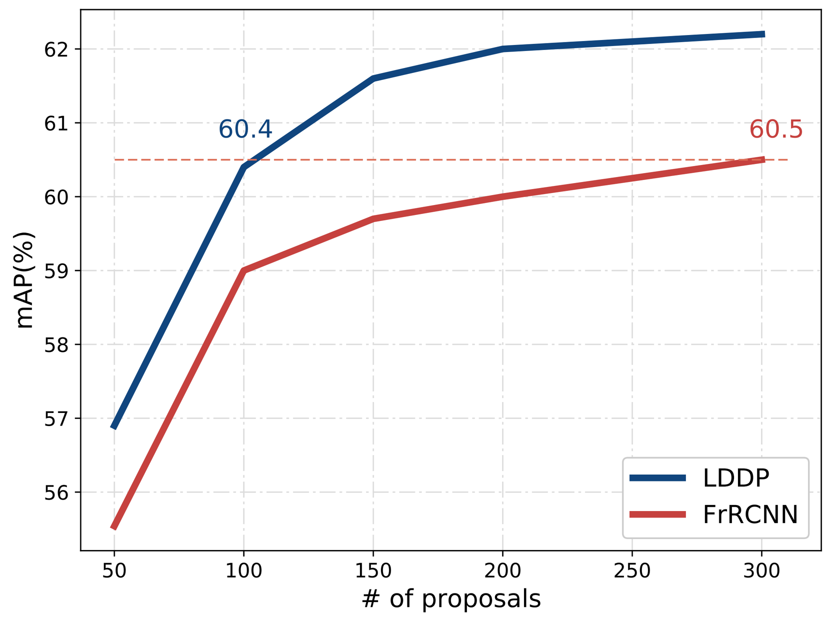
Pascal VOC2007 test detection avg precision(%)

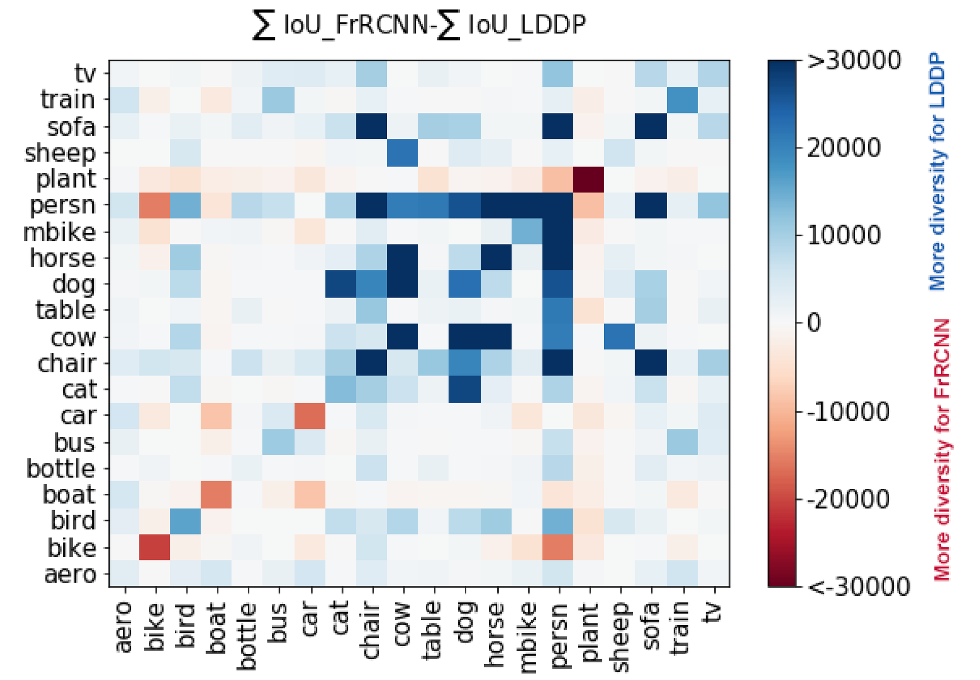
Method	cat	cow	dog	horse	sheep	···	mAP
(FrRCNN, NMS)	71.5	66.3	61.5	76.7	53.3		60.45
(LDDP, LDDP)	**74.9**	**66.6**	**68.5**	**77.4**	**58.8**		**62.21**

MS COCO val detection avg precision and avg recall(%)

Method	Avg Prec @ IoU 0.5-0.95	Avg Prec @ IoU 0.5	Avg Prec @ IoU 0.75	Avg Prec @ Area S	Avg Prec @ Area M	Avg Prec @ Area L	Avg Rec @ Area S	Avg Rec @ Area M	Avg Rec @ Area L
(FrRCNN, NMS)	15.0	31.5	12.7	**3.6**	15.1	23.8	6.0	24.2	38.9
(LDDP, LDPP)	**15.5**	**32.2**	**13.4**	3.5	**15.8**	**24.7**	**6.8**	**27.3**	**43.2**

- **Non-Redundant Diverse Proposals:**





Code is available at https://github.com/azadis/LDDP

VISUALIZATION