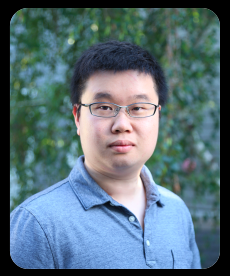
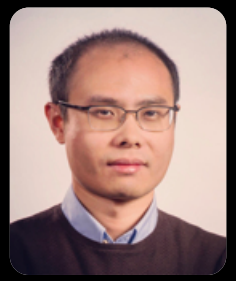
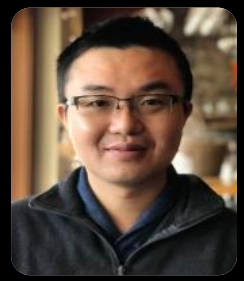
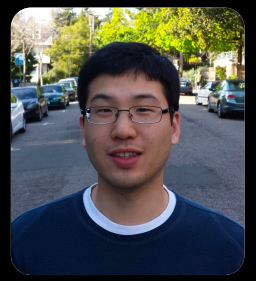
Instance-aware, Context-focused, and Memory-efficient Weakly Supervised Object Detection

Zhongzheng Ren, Zhiding Yu, Xiaodong Yang, Ming-Yu Liu,
Yong Jae Lee, Alexander G. Schwing, Jan Kautz

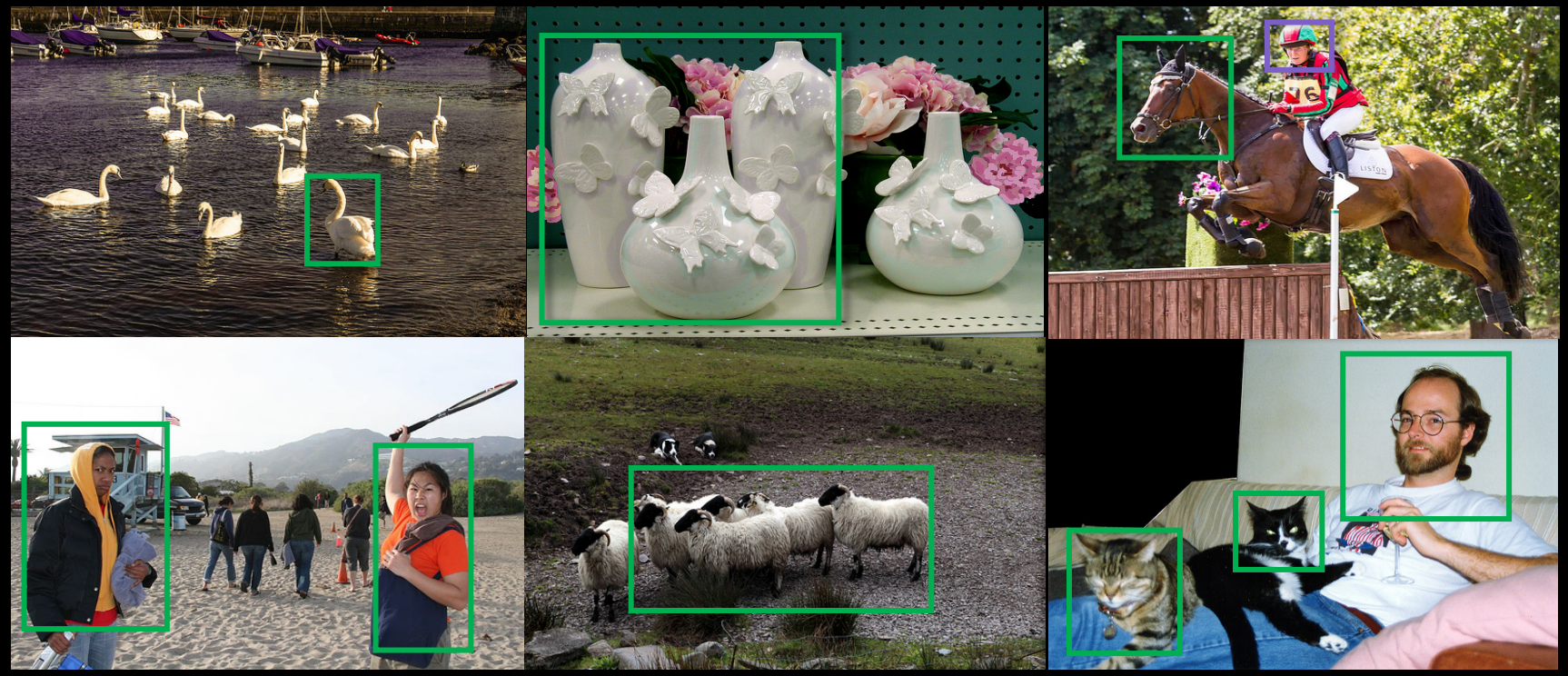
CVPR 2020

Weakly Supervised Object Detection

- Learn object detectors using image-level class labels
- Prior work suffers from several issues:

Missing Instances *Grouped Instances* *Part Domination* *Huge memory consumption*



Weakly Supervised Object Detection

- Learn object detectors using image-level class labels
- Prior work suffers from several issues:

Missing Instances *Grouped Instances* → *Multiple instance self-training*

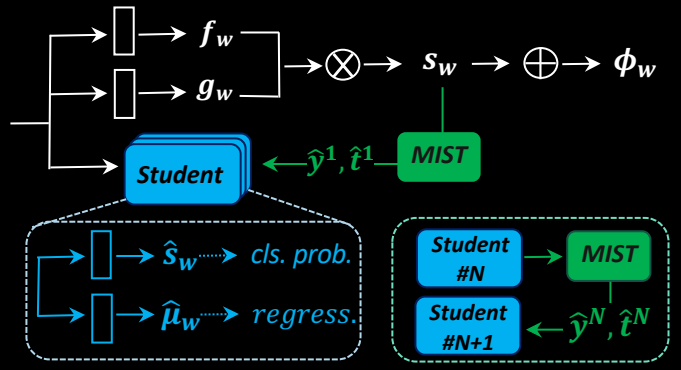
Part Domination → *Concrete DropBlock*

Huge memory consumption → *Sequential batch back propagation*

Weakly Supervised Object Detection

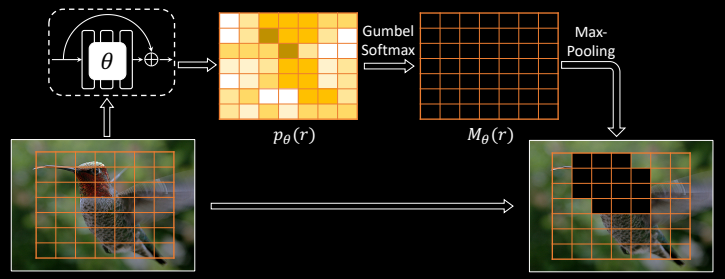
Multi-Inst. Self-Training

- ***Spatial-diversity***: *select multiple diverse yet representative ROIs as pseudo-boxes*
- **Regression**: *introduce regression to WSOD tasks*



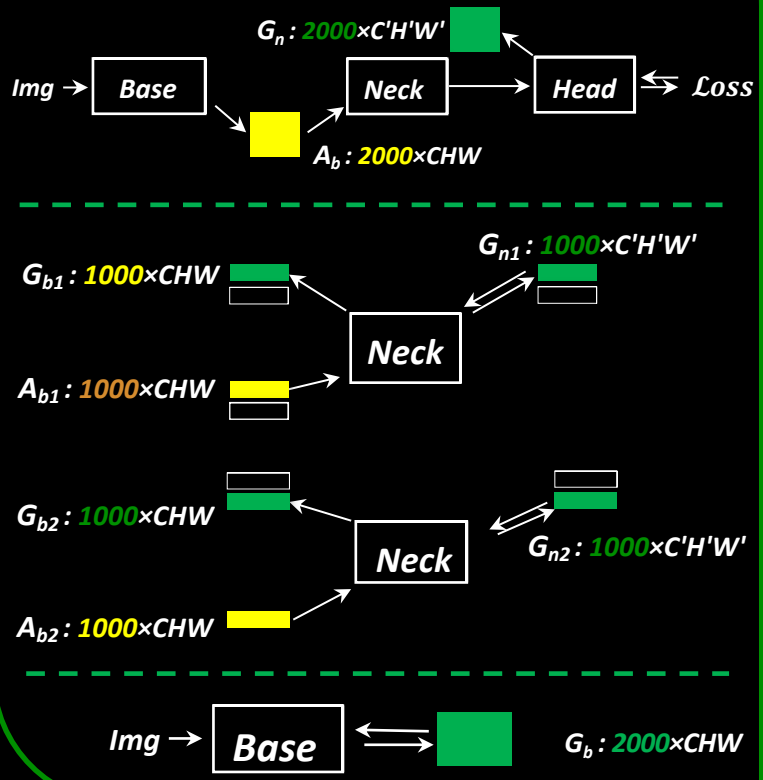
Concrete DropBlock

- **Parametric**: *learning to drop*
- **Data-driven**: *diff. areas (location, size) for diff. objects*
- **Structured**: *drop square blocks*



$$w^*, \theta^* = \arg\min_{w} \max_{\theta} \sum_{I} \mathscr{L}_{img}(w, \theta) + \mathscr{L}_{roi}(w, \theta)$$

Seq. Batch Back-Prop.



Instance-aware *Context-focused* *Memory-efficient*

Results: SOTA on multiple benchmarks

COCO

Dataset	2014-Val.		Test	
Metrics	AP	AP50	AP	AP50
Fast R-CNN	18.9	38.6	19.3	39.3
Faster R-CNN	21.2	41.5	21.5	42.1
WSDDN	-	-	-	11.5
WCCN	-	-	-	12.3
PCL	8.5	19.4	-	-
C-MIDN	9.6	21.4	-	-
WSOD2	10.8	22.7	-	-
Ours	**11.4**	**24.3**	**12.1**	**24.8**
Ours (R-50)	**12.6**	**26.1**	**-**	**-**
Ours (R-101)	**13.0**	**26.3**	**-**	**-**

PASCAL VOC

Dataset	VOC-07	VOC-12
Metric	AP50	AP50
Fast R-CNN	66.9	65.7
Faster R-CNN	69.9	67.0
WSDDN	34.8	-
PCL	43.5	40.6
C-MIL	50.5	46.7
Pred-Net	52.9	48.4
C-MIDN	52.6	50.2
WSOD2	53.6	47.2
Ours	**54.9**	**52.1**
Ours (07+12)	**54.9**	**-**

ImageNet-VID

New task: video WSOD

Backbone	VGG16	ResNet
Metric	AP	
Supervised	61.7	80.5
WSDDN	24.2	21.9
OICR	34.8	40.5
Ours	**36.6**	**45.7**
Ours+flow	**38.3**	**46.9**

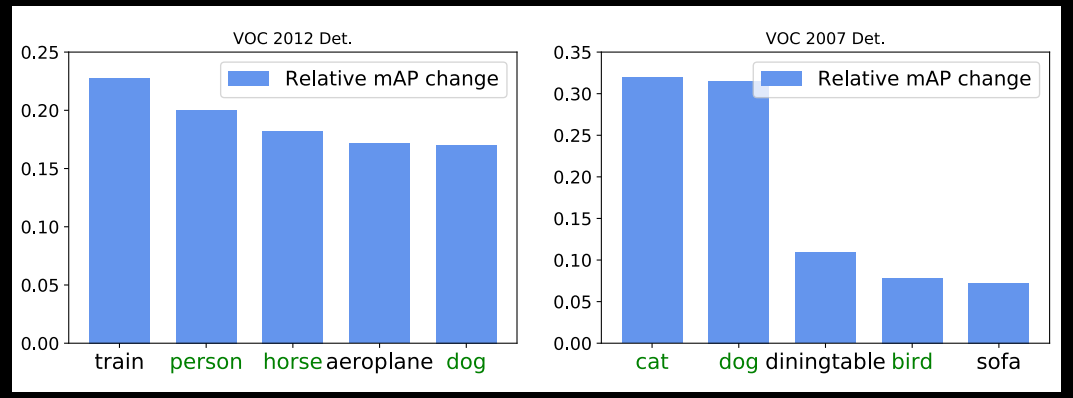
Effectiveness of each module

Dataset	VOC-07		VOC-12	
Task	CorLoc	Det.	CorLoc	Det.
Baseline	60.8	42.5	-	-
+PCL	62.7	43.5	63.2	40.6
+MIST no reg.	62.9	48.3	65.1	-
+MIST	64.9	51.4	66.7	-
+Img Spa. Dropout	64.3	51.1	65.9	-
+ROI Spa. Dropout	66.8	52.4	67.3	-
+DropBlock	67.1	52.9	68.4	-
Ours	68.8	54.9	70.9	52.1

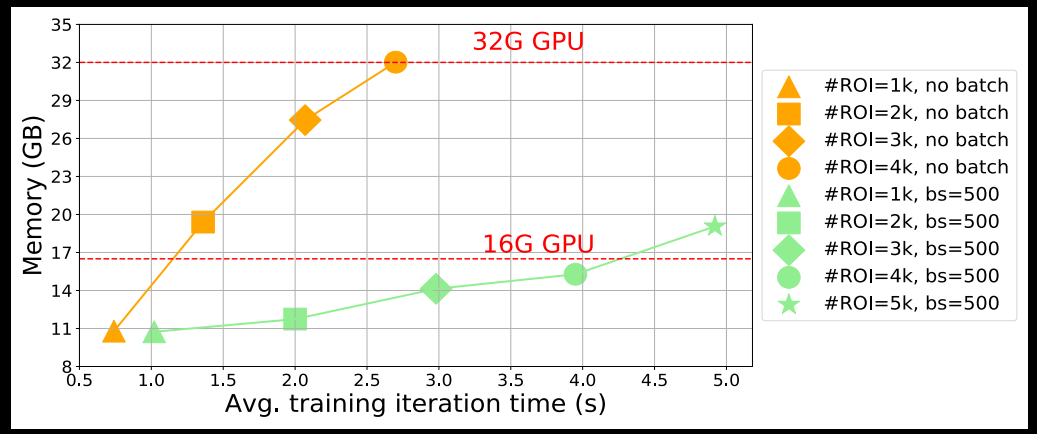
Ablation study.

Metrics	AR	AP10	AR100	ARS	ARM	ARL
w/o MIST	18.6	30.6	32.5	8.8	25.8	38.9
w/ MIST	20.5	37.8	43.9	15.0	34.8	51.7

Has instance ambiguity been addressed?



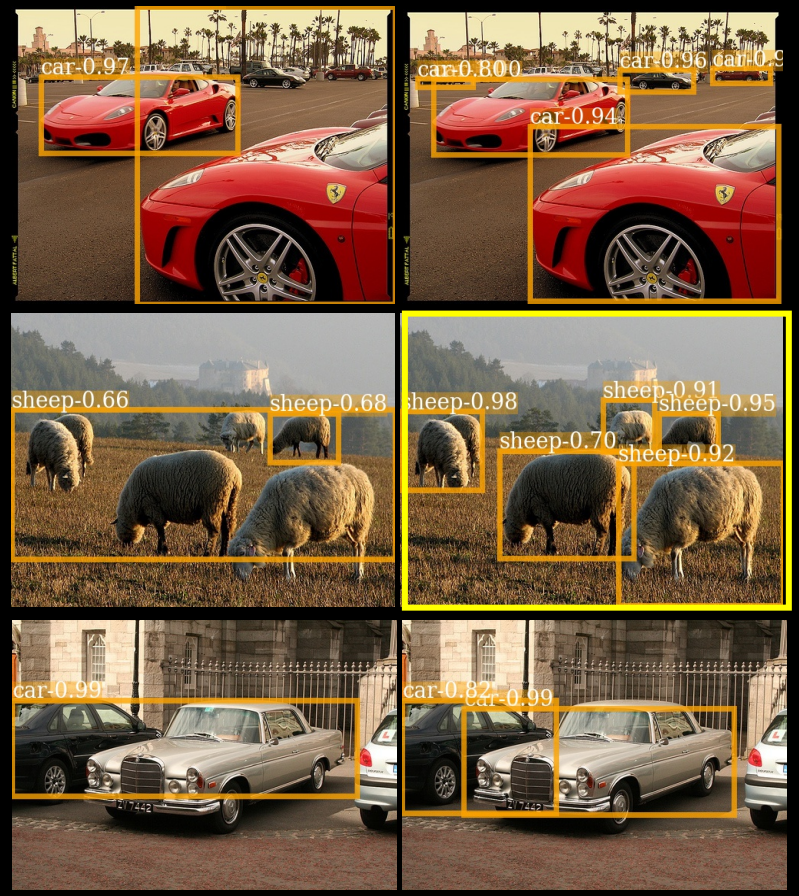
Has part domination been addressed?



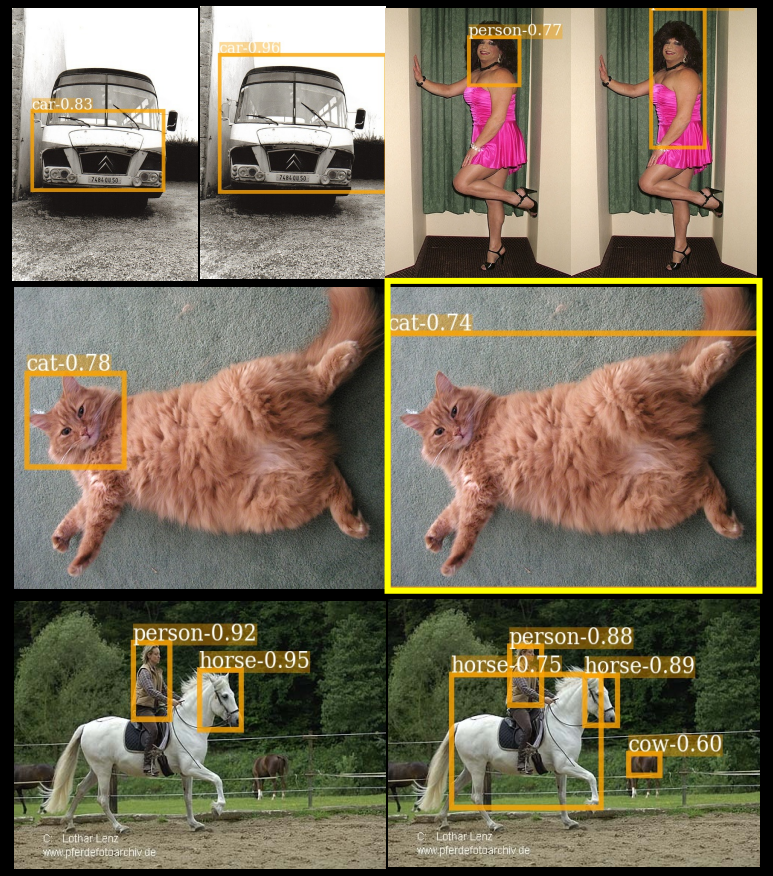
Has memory consumption been reduced?

Results

Missing Instance



Grouped Instance



Part Domination

Thanks for watching!