

# Weakly and deeply supervised visual learning

CSIG青年科学家论坛

王兴刚

[www.xinggangw.info](http://www.xinggangw.info)

华中科技大学

# Annotation time of manual supervision

2



{motorbike, person} {motorbike (point),  
person (point)}

{motorbike (b-box),  
person (b-box)}

{motorbike (pixel labels),  
person (pixel labels)}

Annotation time:

1

2.4

10

78

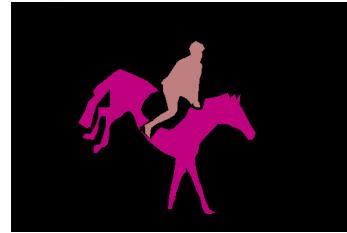
second per instance

Berman et al., What's the Point: Semantic Segmentation with Point Supervision, ECCV 16

# Image labels



Person, Horse

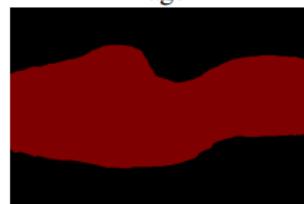


- Supervision: image (category) labels
- Target: Object detection, semantic segmentation etc

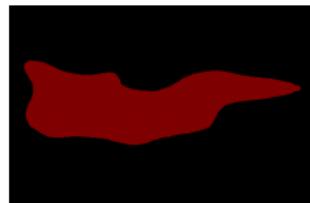
[Verbeek CVPR 07, Pendey, ICCV 11, Cinbis CVPR 14, Wang ECCV 14, Papandreou ICCV 15, Belien CVPR 15, Tang CVPR 17, Wei CVPR 17, Singh ICCV 17, Huang CVPR 18 etc.]

# Video labels

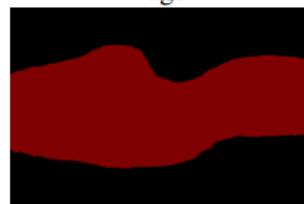
- Supervision: video (category) labels
- Target: Object detection, semantic segmentation etc



Image



DeepLab [8] (full)



EM-Adapt [28] (weak)



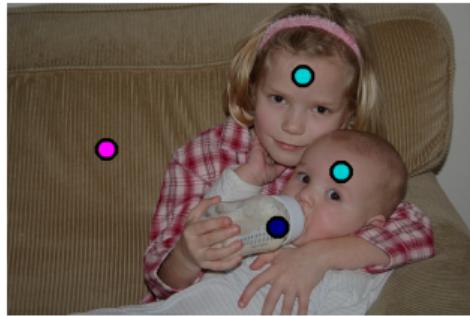
M-CNN (weak)

[Papazoglou, ICCV 13]

[Tokmakov ECCV 16]

# Clicks in object

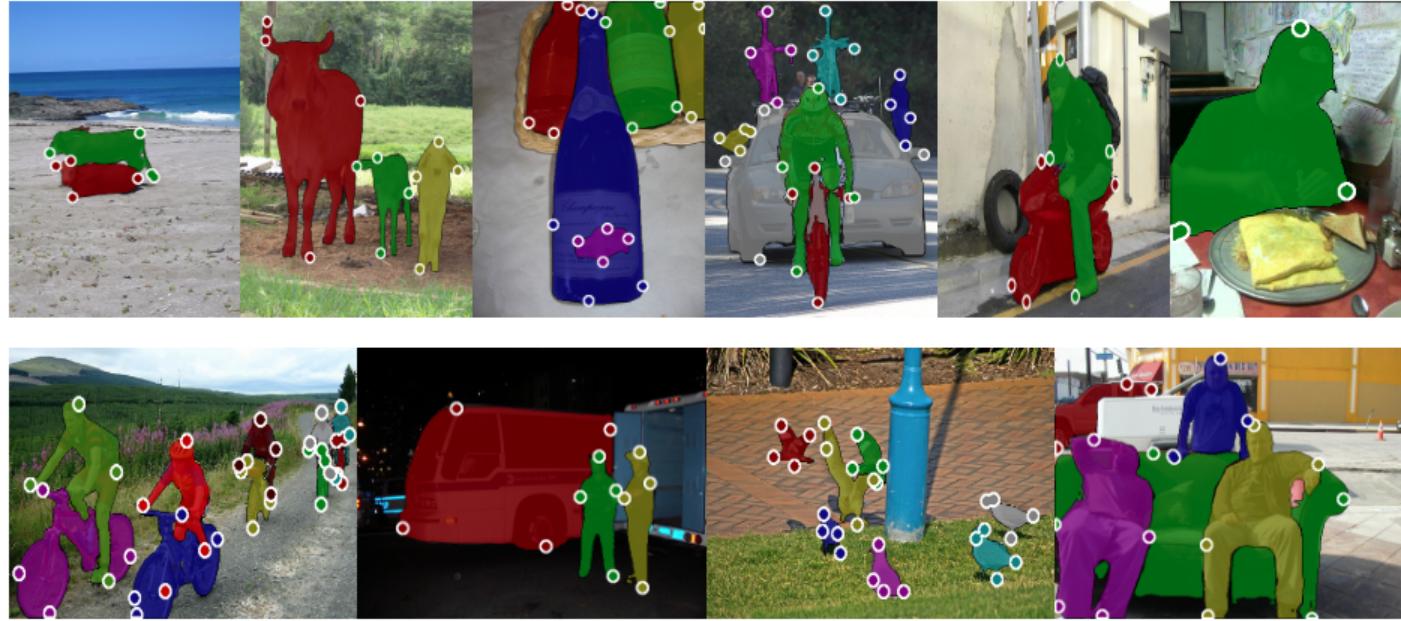
5



[Bearman ECCV 16]

- Supervision: one point each instance/category
- Target: Object detection, semantic segmentation etc.

# Extreme points

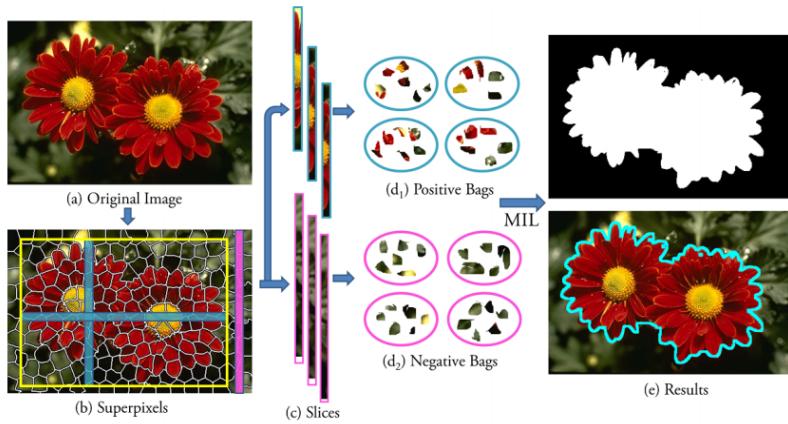


DEXTR [Maninis CVPR 18]

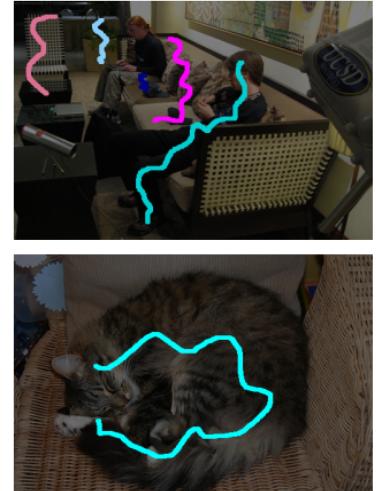
- Supervision: object extreme points
- Target: instance segmentation

[Padadopoulos ICCV 17, Maninis CVPR 18]

# Scribbles in object



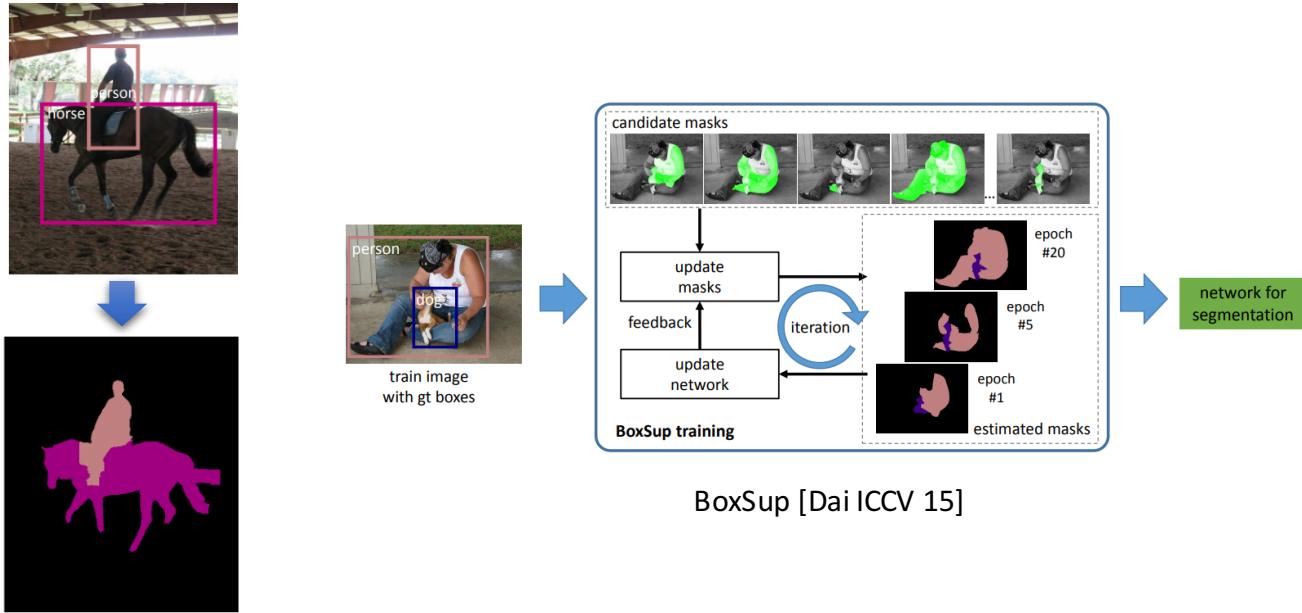
MIL Cut [Wu CVPR 14]



[Bearman ECCV 16]

- Supervision: scribbles/lines per instance
- Target: instance segmentation

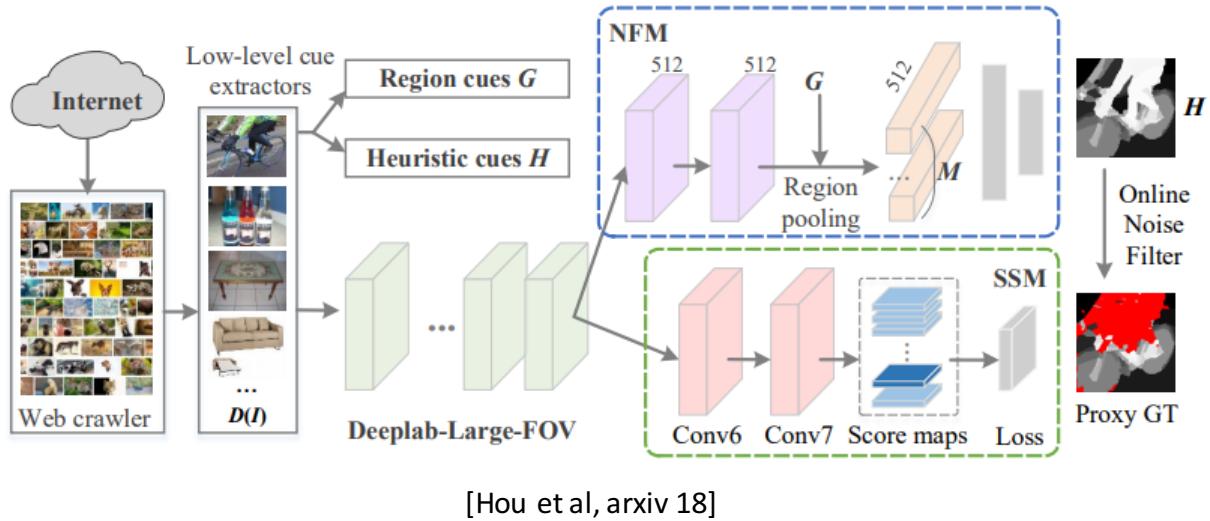
# Object bbox



- Supervision: object bounding boxes
- Target: instance segmentation

[Rother SIGGRAPH 04, Dai ICCV 15, Khoreva CVPR 17]

# Webly supervision

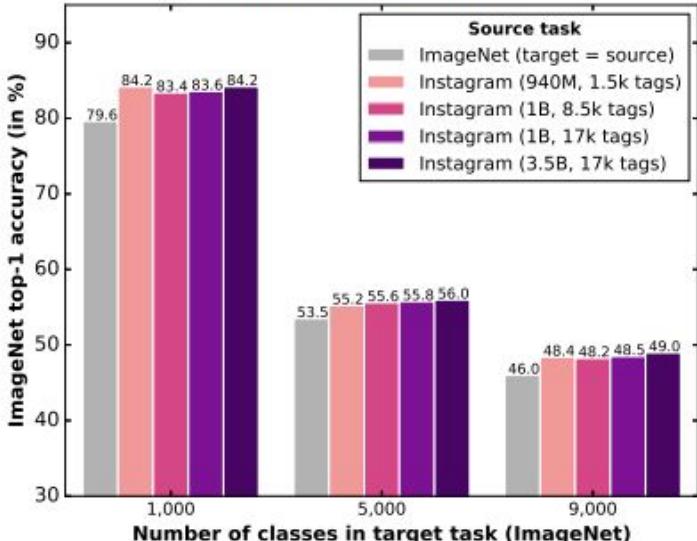


[Hou et al, arxiv 18]

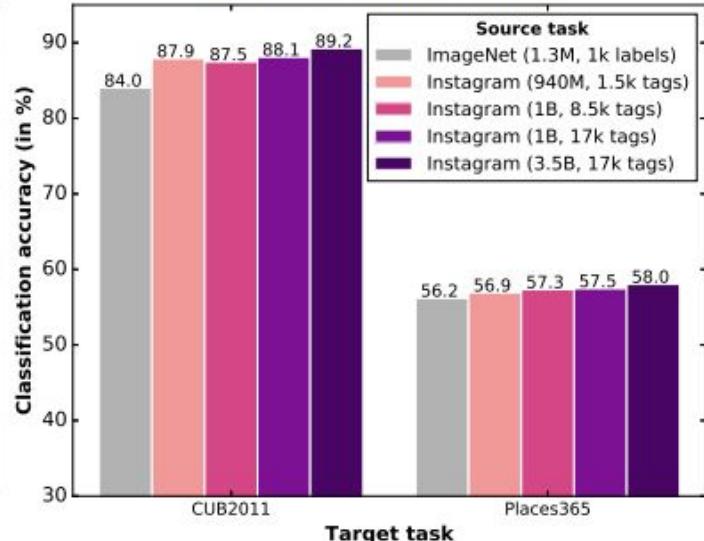
- Supervision: Keywords & search engines
- Target: semantic segmentation

# Hashtag

**Target task: ImageNet**



**Target task: CUB & Places**



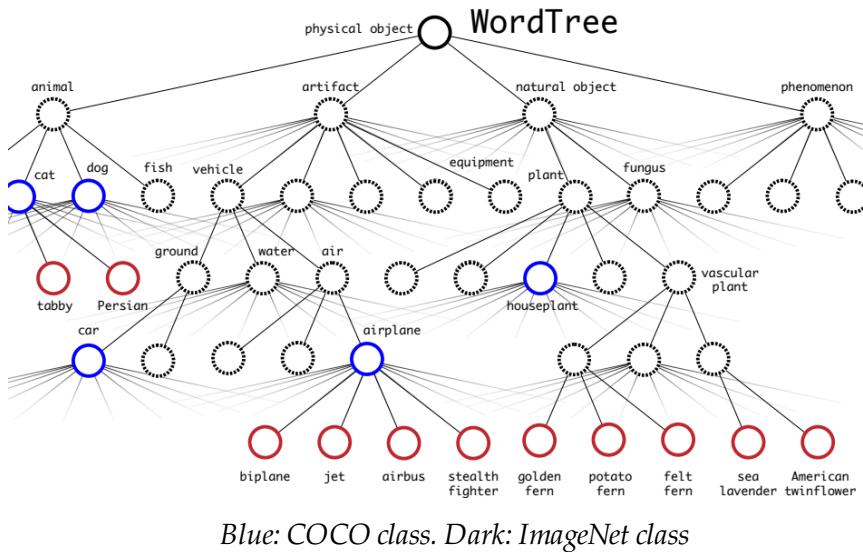
[Mahajan, ECCV 18]

- Supervision: 3.5 billion images with Instagram tags
- Target: a good pre-trained model

# Mixing full & weak supervision

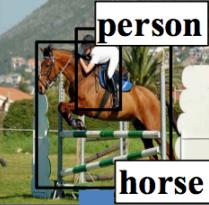
11

- Supervision: COCO (has bbox) + ImageNet (has image label)
- Target: object detection for **9000** classes

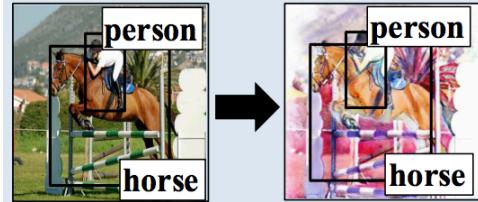


Blue: COCO class. Dark: ImageNet class

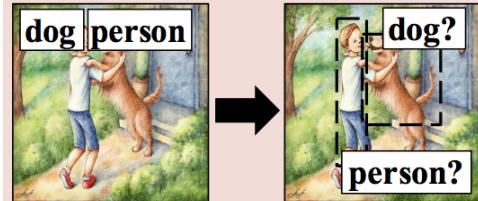
YOLO9000, CVPR 2017 Best Paper Honorable Mention [Redmon CVPR 17]

	Level of annotations	
	Image	Instance
Source domain		
Target domain		

## Domain transfer (DT)



## Pseudo-labeling (PL)



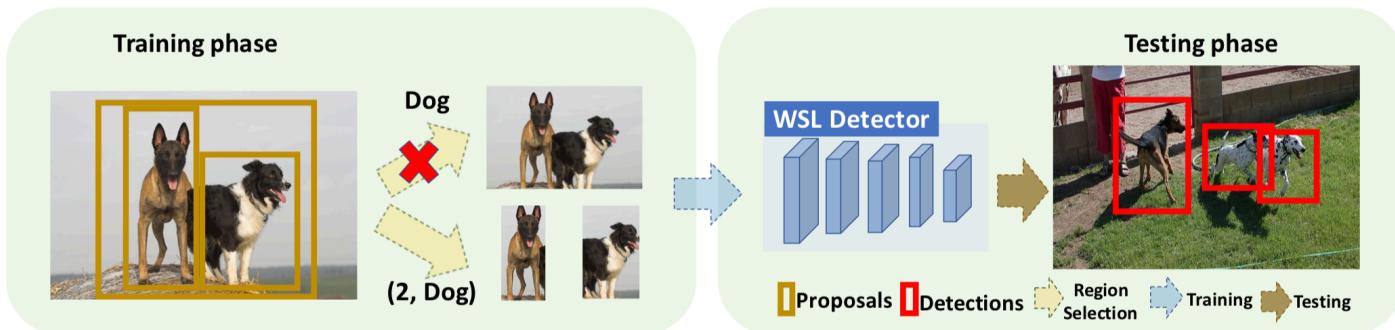
[Inoue, CVPR 18]

- Supervision: bbox in source domain + image label in target domain
- Target: bbox in target domain

# Count of object

13

- Supervision: counts of object per class
- Target: object detection

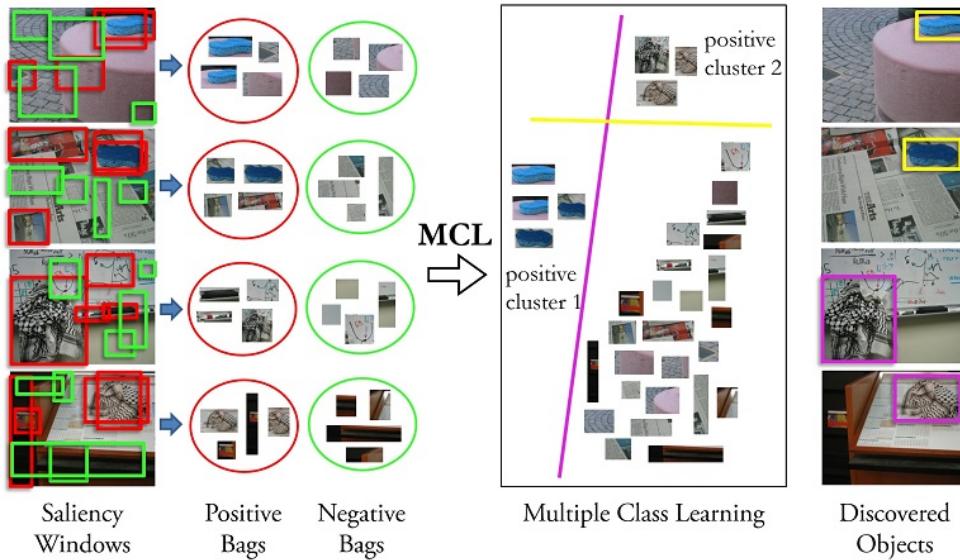


C-WSL [Wang, ECCV 18]

# Only number of classes

14

- Supervision: only number of classes
- Target: object bbox

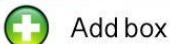
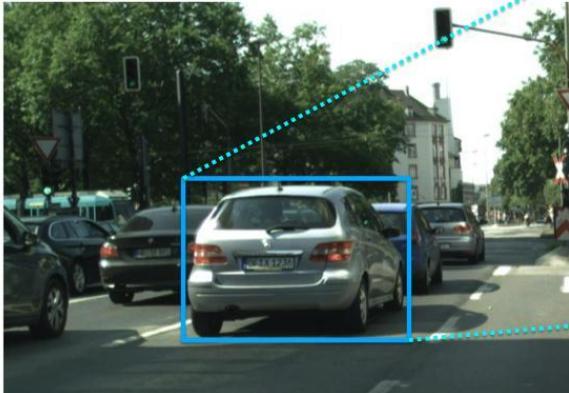


bMCL [Zhu, CVPR 12, PAMI 15]



## Polygon-RNN

Interactive Object Annotation Tool



Add box



To correct the prediction, drag and drop a point

Polygon-RNN, Honorable Mention Best Paper Award [Castrejon, CVPR 17]

Polygon-RNN cuts down the number of required annotation clicks by a factor of 4.74

- Supervision: bbox + interactive key point
- Target: object polygon

# From the perspective of machine learning

16

**WSL [Zhou, 2018, National science review]:**

- Incomplete supervision
- Inaccurate supervision
- Inexact supervision



Full supervision



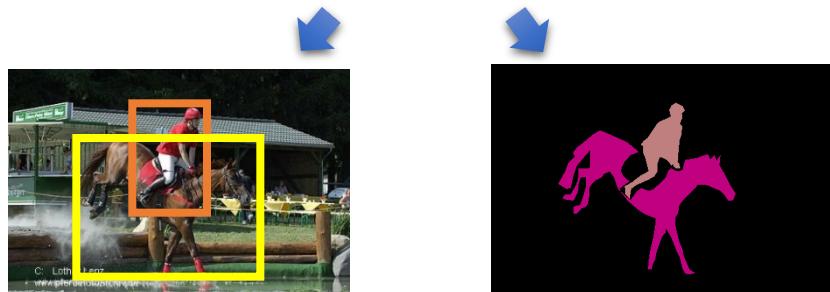
Incomplete supervision



Inaccurate supervision



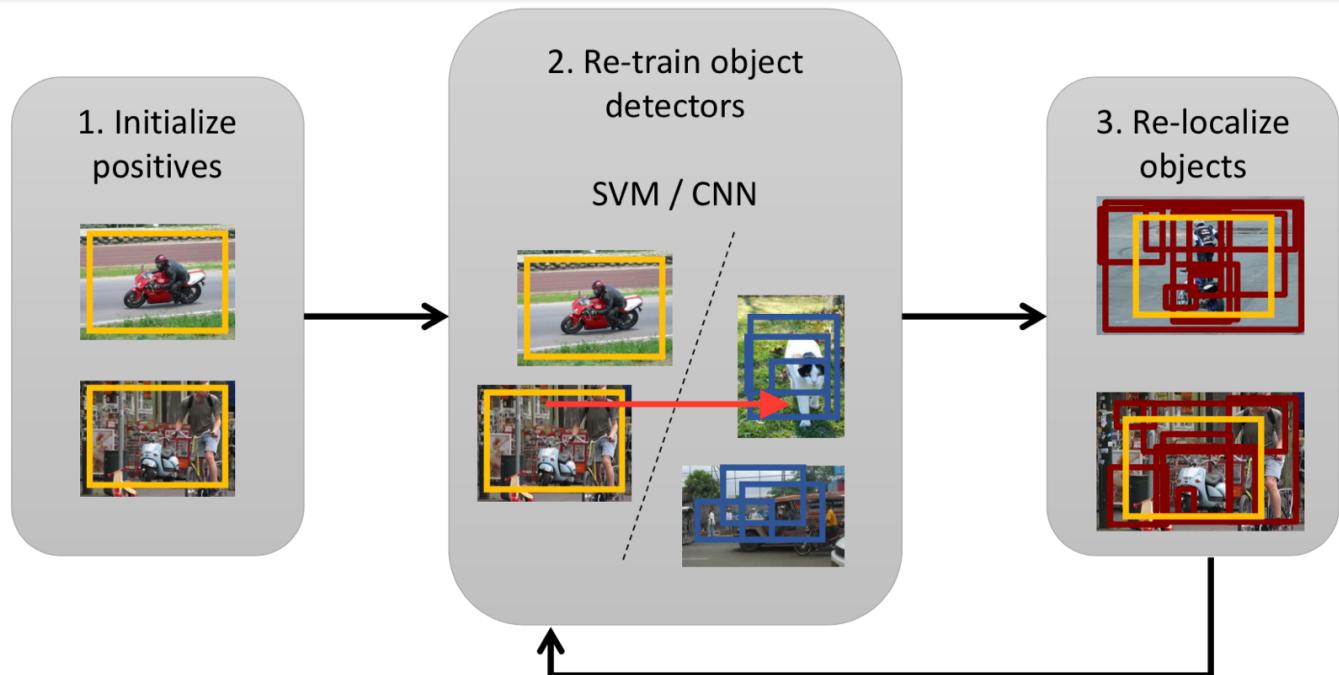
Inexact supervision



- Weakly supervised object detection
- Weakly supervised semantic segmentation

# Standard MIL pipeline

18



1. Window space (usually, using object proposals)
2. Initialization
3. Re-localization & Re-training

[Chum CVPR 07, Deselaers ECCV 10, Siva ICCV 11, Wang ICCV15, Bilen CVPR 15]

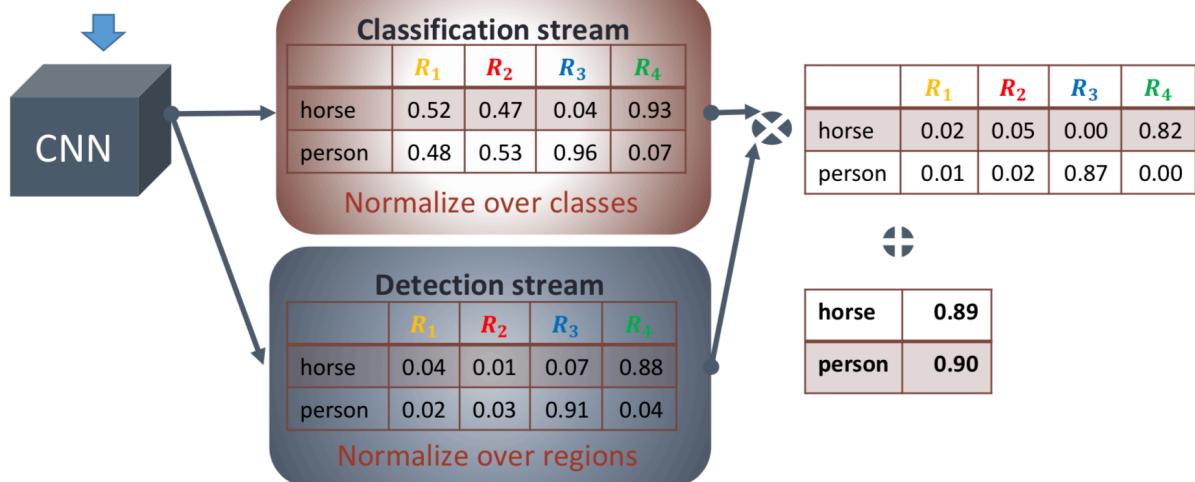
Slide credit: Vitto Ferrari

[Bilen CVPR 16]



Two stream architecture

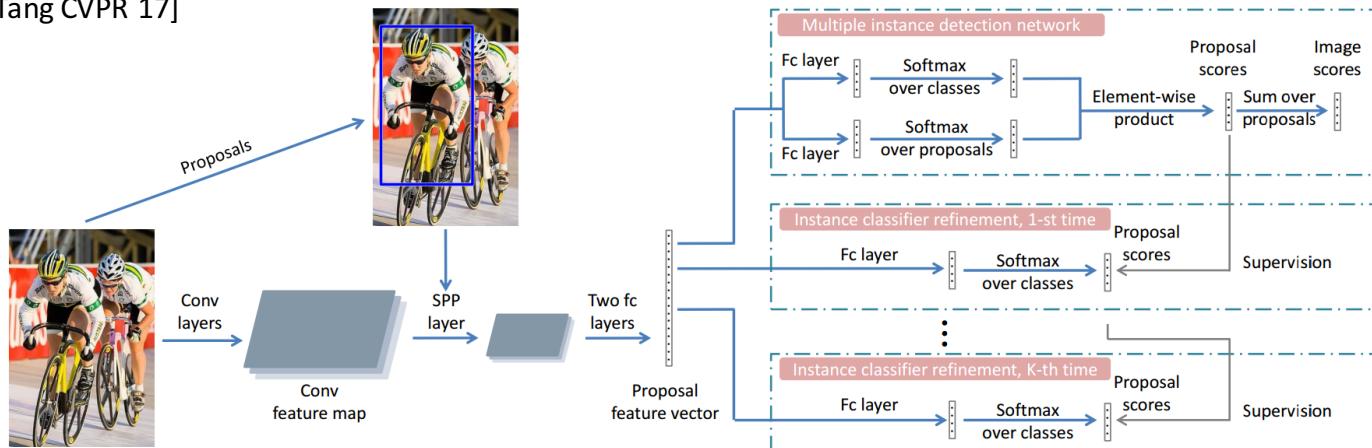
- Classification
- Detection



- ☺ End-to-end Region CNN for WSOD
- ☹ Normalization over classes hurts performance

# Online instance classifier refinement (OICR) network 20

[Tang CVPR 17]

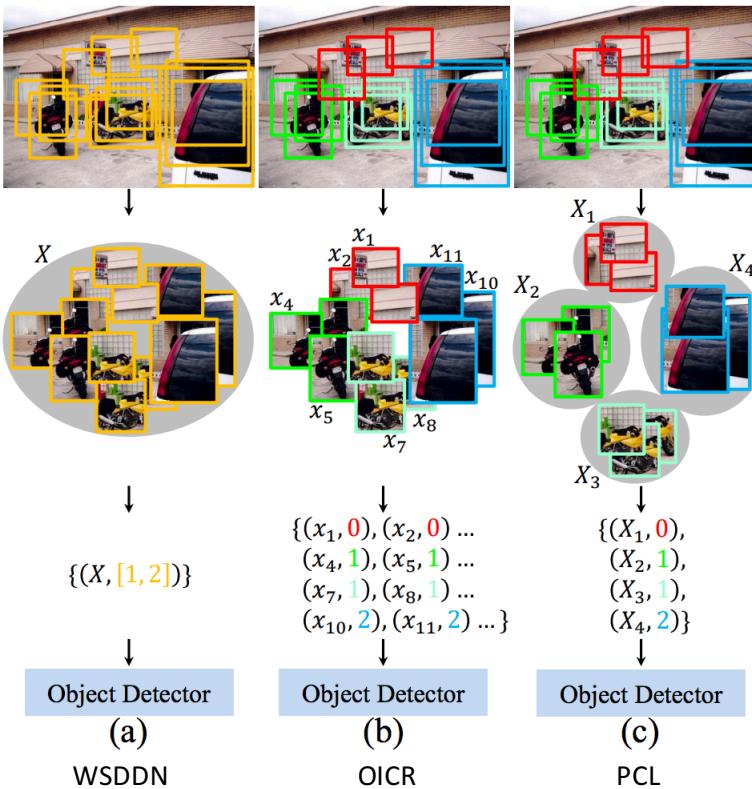


- Additional blocks (**instance classifiers**) for score propagation
  - **In-network supervision**
- 
- ☺ The positive proposals in one image are not sharing score
  - ☺ Performance significantly improves
- 
- ☹ The instance-level in-network supervision may not be correct

# Proposal cluster learning

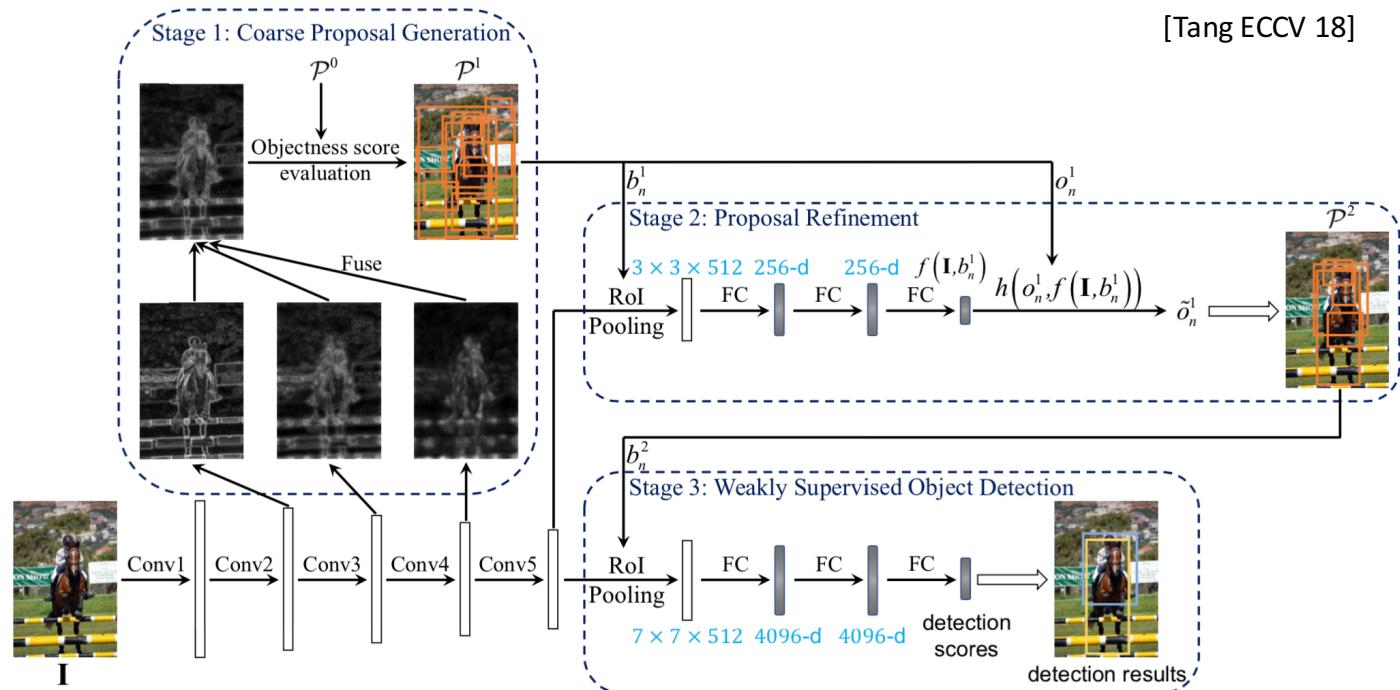
21

[Tang, arXiv:1807.03342v1, under revision of TPAMI]



- ☺ In-network supervision for proposal cluster is more robust
- ☺ MIL in MIL network (Bag in bag MIL)
- ☹ It still relies on hand-crafted object proposals

# Weakly supervised region proposal network 22

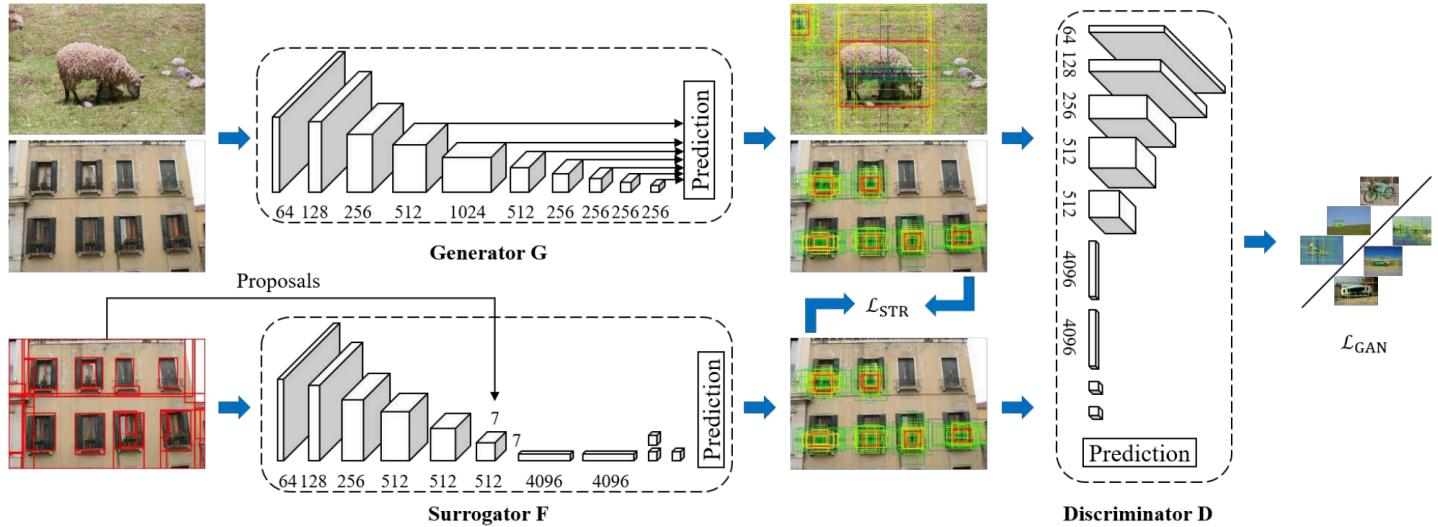


- Generating object proposals from neural activations
- Confirming that CNN contains rich localization information even under weak supervision
- The first weakly supervised region proposal network (wsRPN)

# Generative adversarial learning

23

[Shen CVPR 18]



- ☺ Training SSD by WSOD using GAN loss
- ☺ Fast inference speed using SSD
- ☺ Accurate WSOD by adversarial learning

# Performance

24

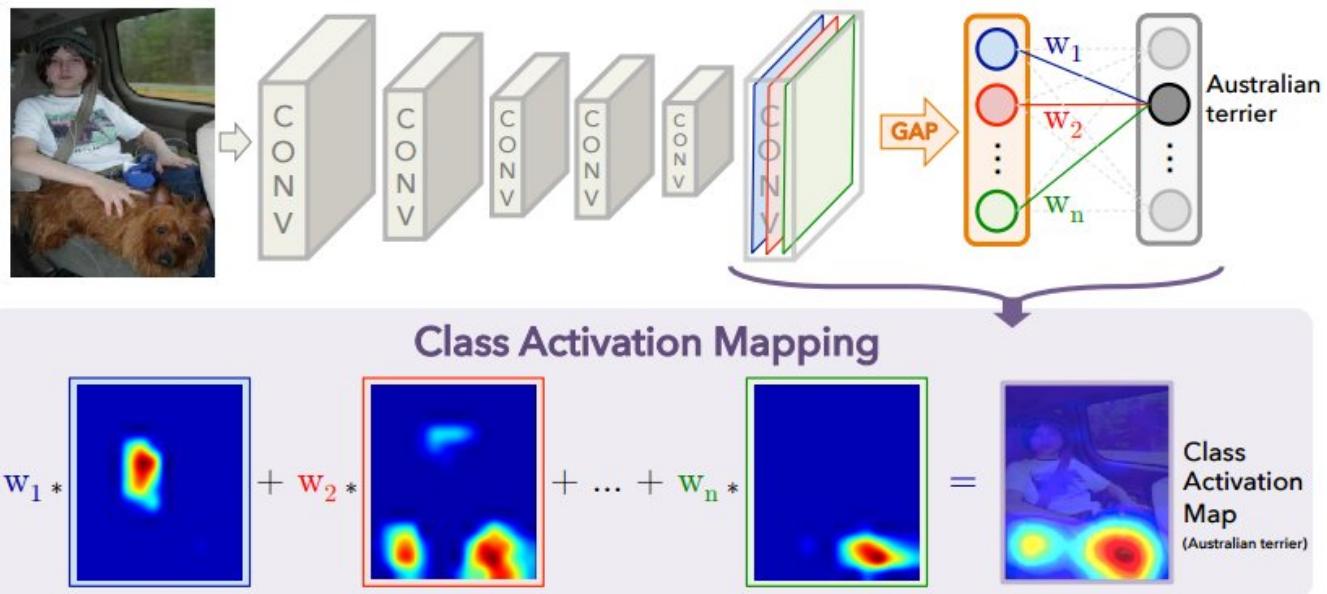


WSOD performance (mAP on PASCAL VOC 2007 test)

# Class activation maps

25

[Zhou CVPR 16]

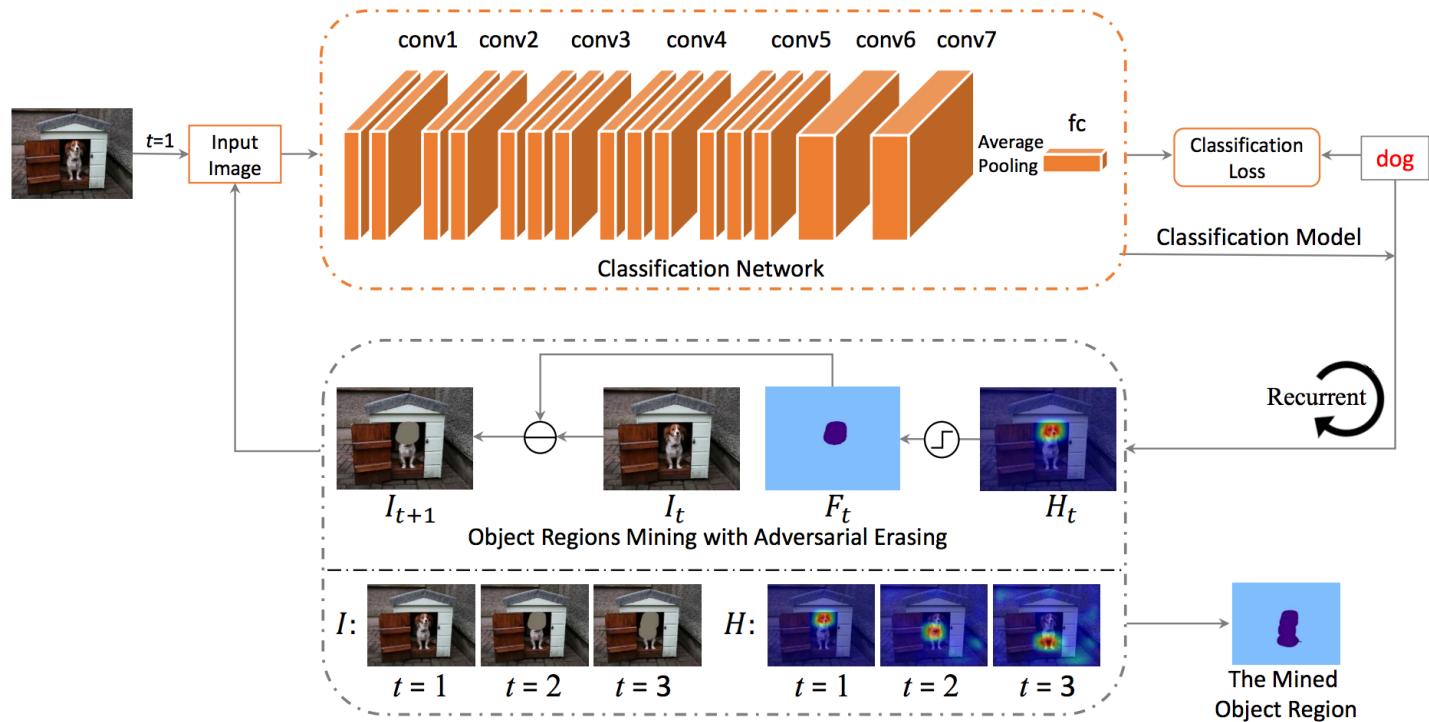


- ☺ Finding discriminative regions by Global Average Pooling in a CNN trained using image labels
- ☺ A very insightful work for understanding CNN

# Adversarial erasing network

26

[Wei CVPR 17]

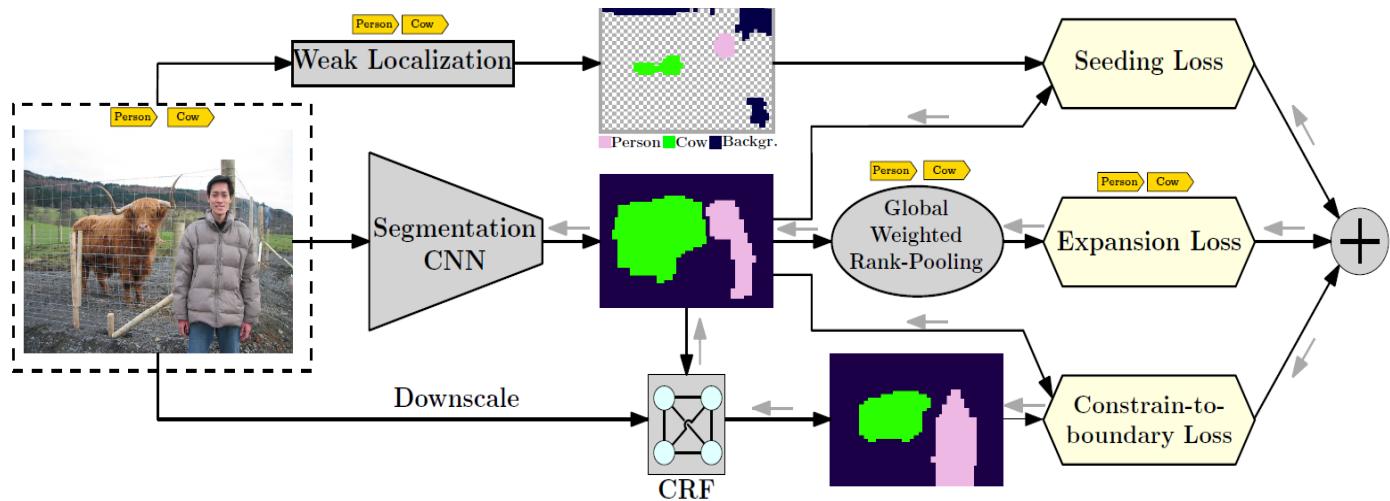


- ☺ Adversarial erasing finds dense and complete object regions
- ☺ Very impressive WSSS results

# Seed, Expand and Constrain (SEC)

27

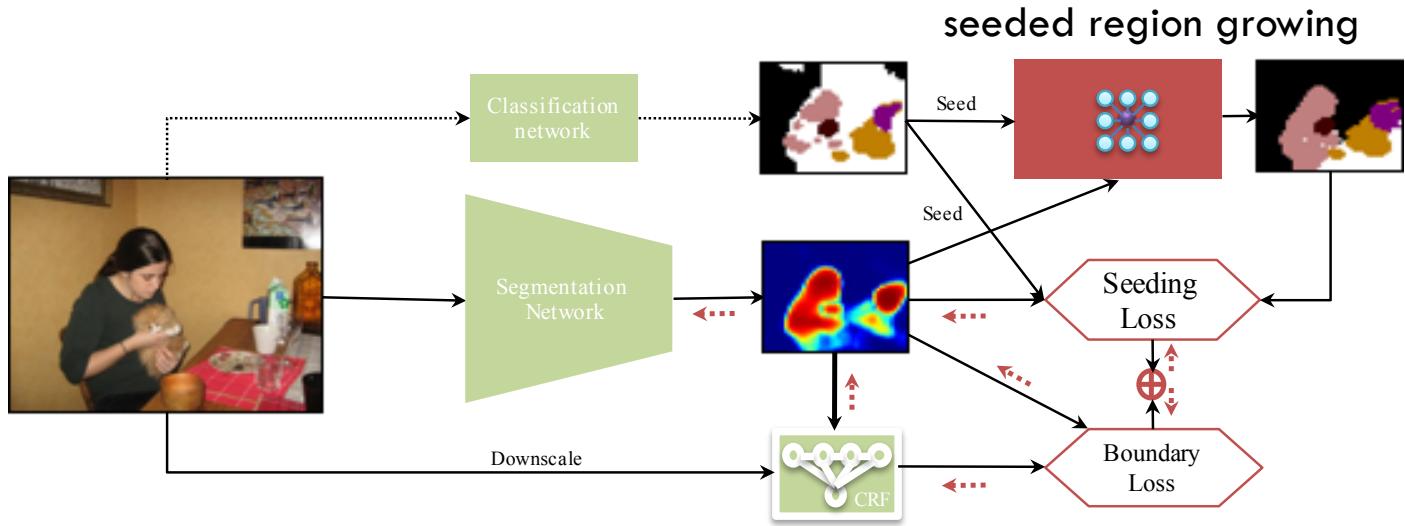
[Kolesnikov ECCV 16]



- ☺ Seed with weak localization cues
- ☺ Expand with image labels
- ☺ Constrain to object boundary using CRF

# Deep seeded region grow (DSRG) network 28

[Huang CVPR 18]

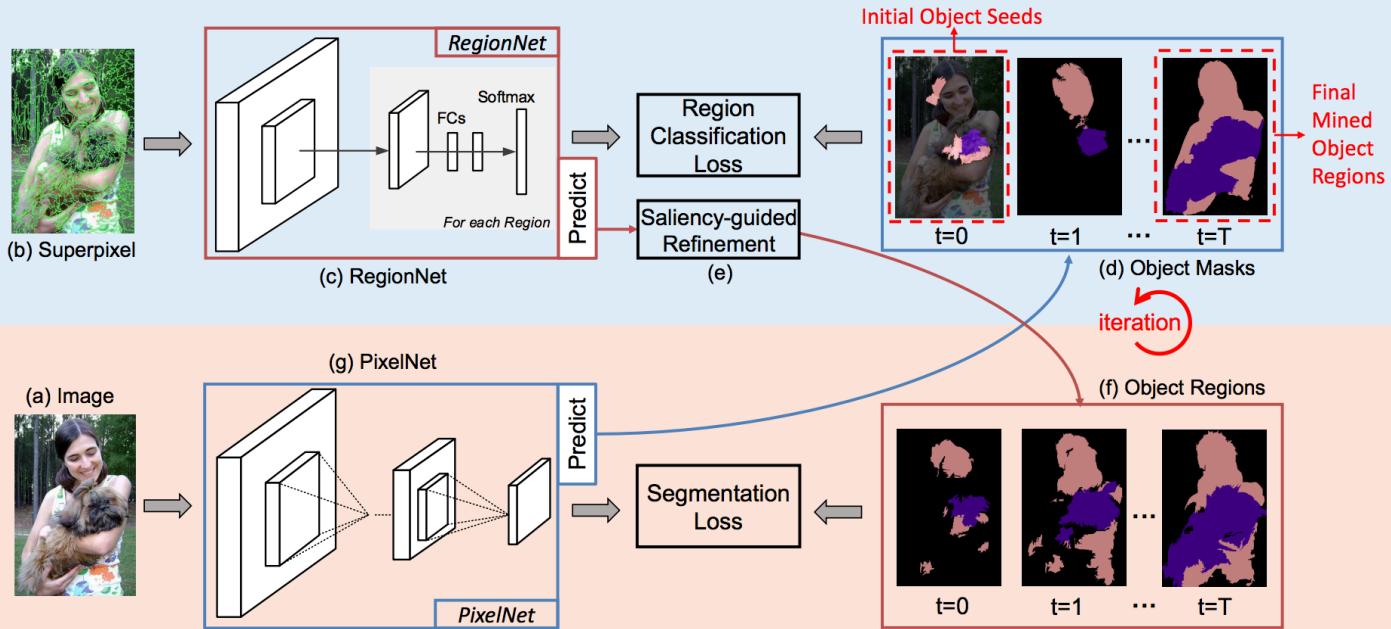


- ☺ Region growing for complete and dense object regions
- ☺ A segmentation network generates new pixel labels by itself

# Iteratively Mining Common Object Features

29

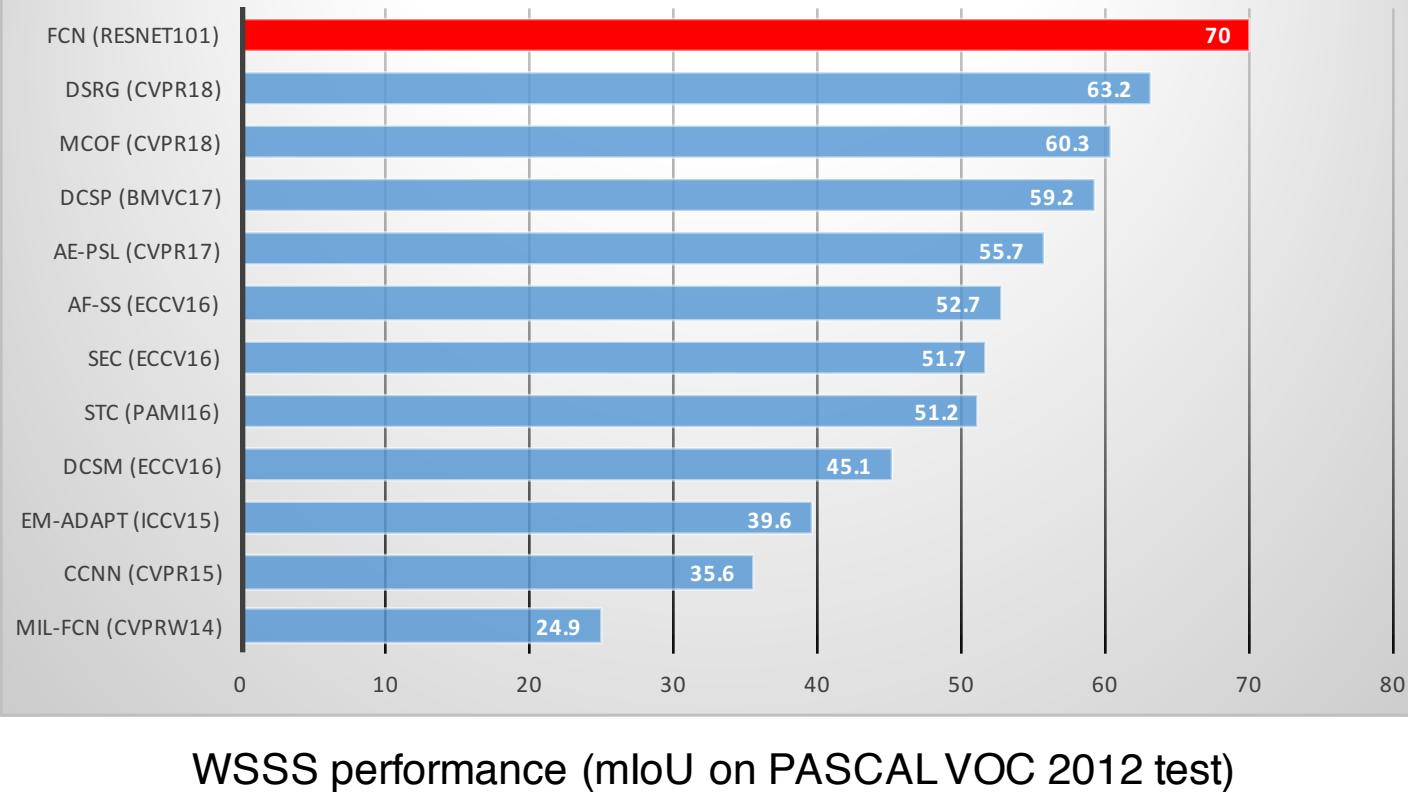
[Wang CVPR 18]

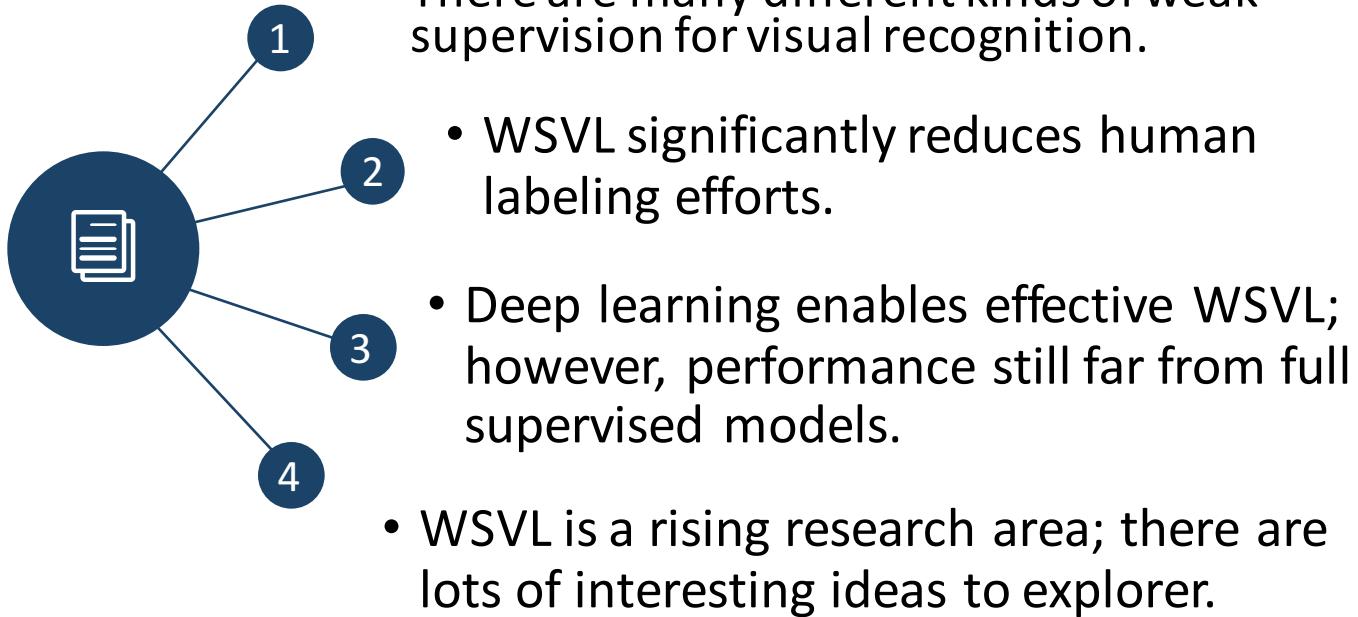


- ☺ Mining common features between region(super-pixel)-level classification network and pixel-level segmentation network

# Performance

30





- CVPR tutorial: **Weakly supervised learning for computer vision**, by Hakan Belen, Rodrigo Benenson, Jasper Uijlings, <https://hbilen.github.io/wsl-cvpr18.github.io>
- Source codes
  - WSSDN: <https://github.com/hbilen/WSDDN>
  - CAM: <http://cnnlocalization.csail.mit.edu>
  - OICR/PCL: <https://github.com/ppengtang/oicr/tree/pcl>
  - SEC: <https://github.com/kolesman/SEC>
  - DSRG: <https://github.com/speedinghzl/DSRG>

Thanks a lot your attention!

[www.xinggangw.info](http://www.xinggangw.info)