Multi-Label Image Classification via Knowledge Distillation from Weakly-Supervised Detection

Yongcheng Liu[1,2], Lu Sheng[3], Jing Shao[4], Junjie Yan[4], Shiming Xiang[1,2], Chunhong Pan[1]

[1]National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences

[2]School of Artificial Intelligence, University of Chinese Academy of Sciences

[3]CUHK-SenseTime Joint Lab, The Chinese University of Hong Kong *[4]SenseTime Research*

Project Page: https://yochengliu.github.io/MLIC-KD-WSD/

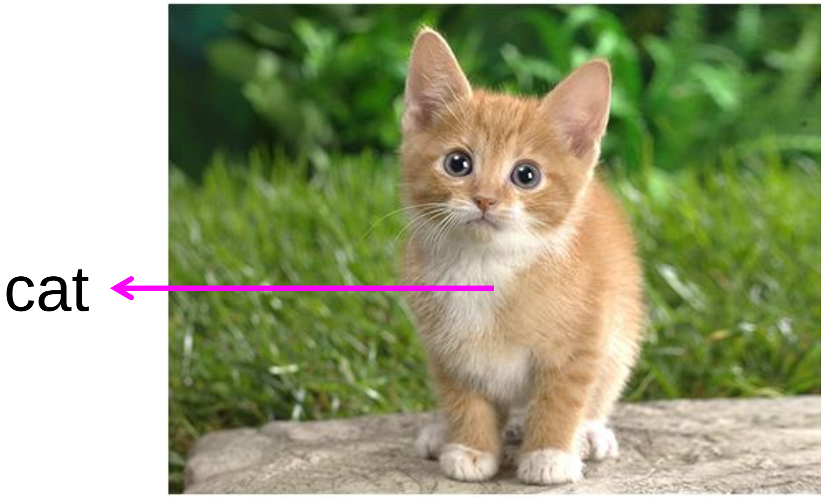
Outline

❶ **Introduction**

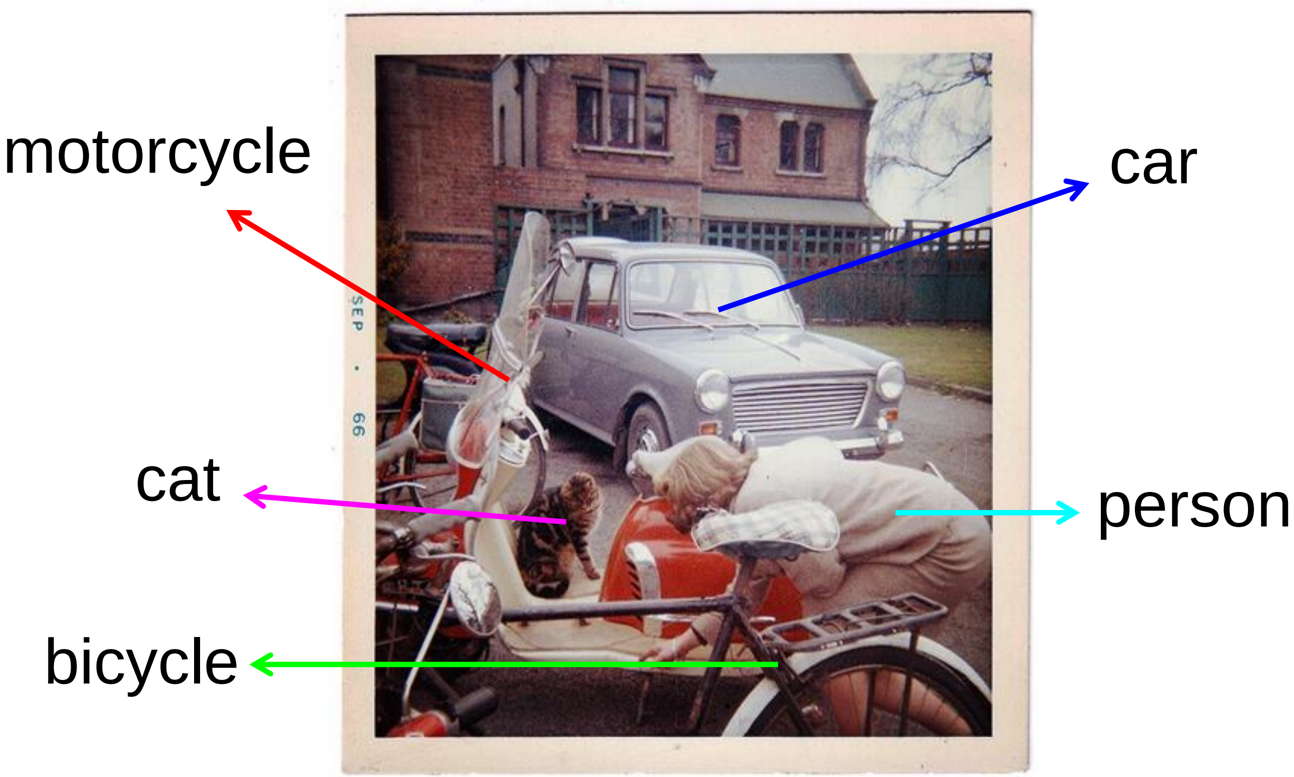
❷ **Related work**

❸ **MLIC-KD-WSD**

Introduction *Problem*



Single label, *e.g.*, ImageNet



Multiple labels

real-world images

general visual understanding

Introduction *Challenges*

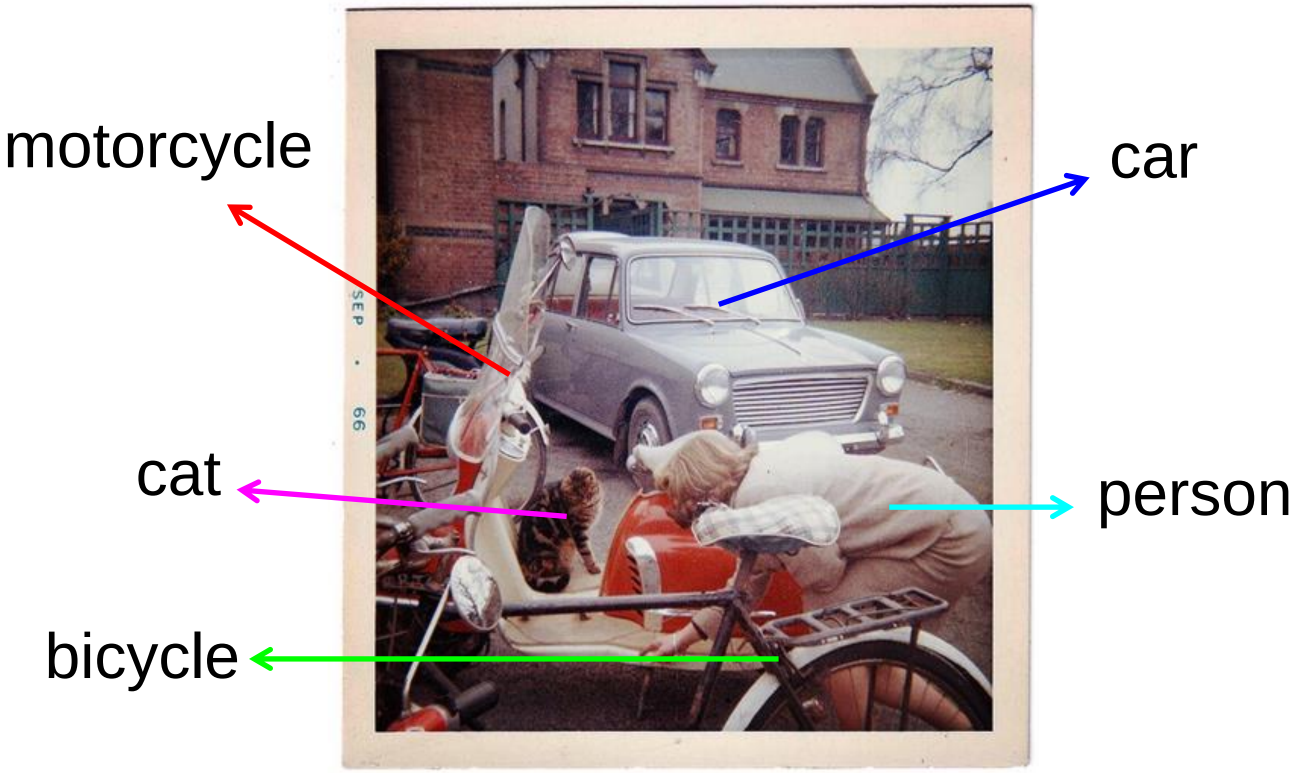
Visual recognition

- inter-class similarity
- intra-class variation

Thorough understanding

- classes ⟷ semantic regions
- class dependencies

  e.g., person & baby, cat & dog



Multiple labels

Introduction *Applications*

Scene recognition

HUAWEI nova 3

高颜值 爱自拍

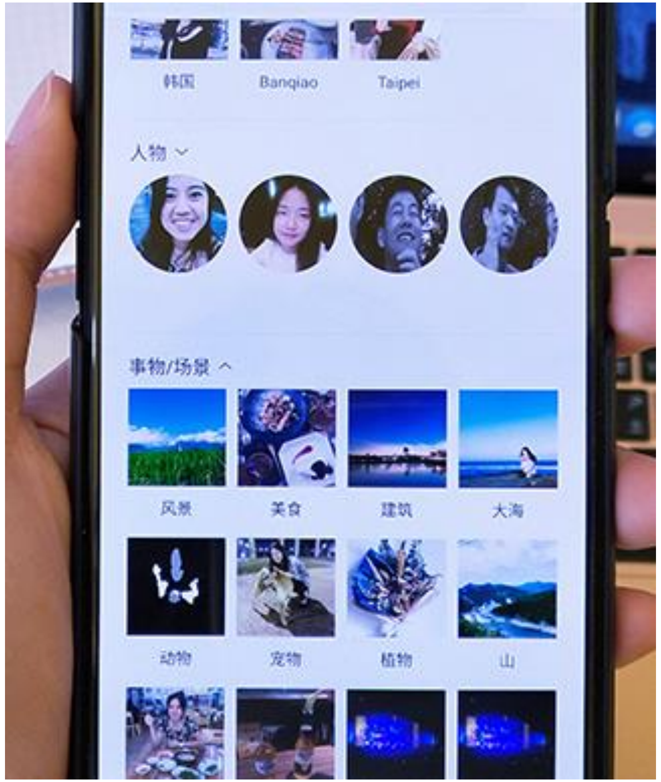
2400万四摄* AI实力派

*前置双摄像头为2400万+200万像素，后置双摄像头为2400万+1600万像素
图片仅供参考，请以实物为准。

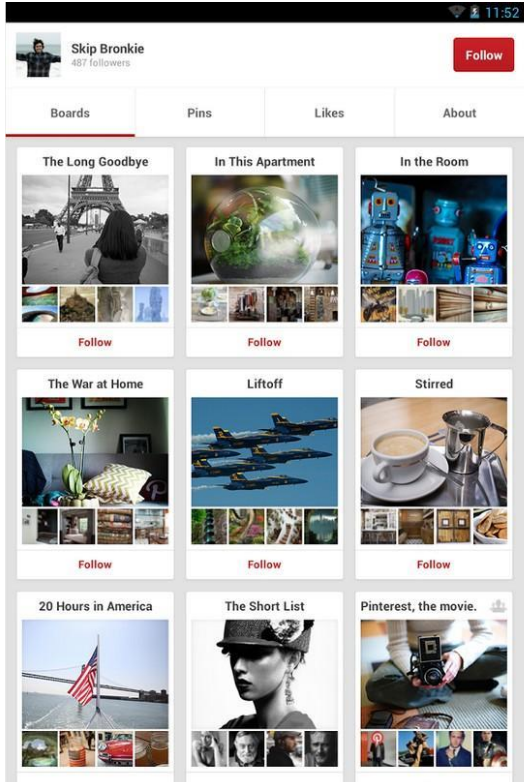
前置2400万 [2] AI美拍 [3]

前置2400万摄像头配合200万虚化镜头，支持放大2m×1.5m海报[4]。AI识别多种典型环境，自动配置拍照参数。

Smart albums

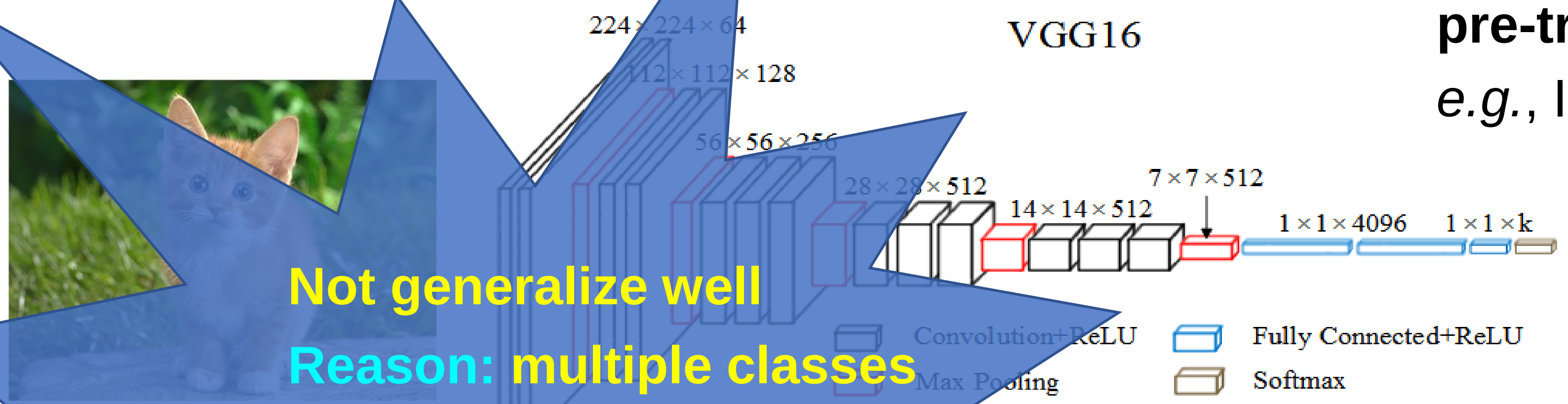


Social image analysis



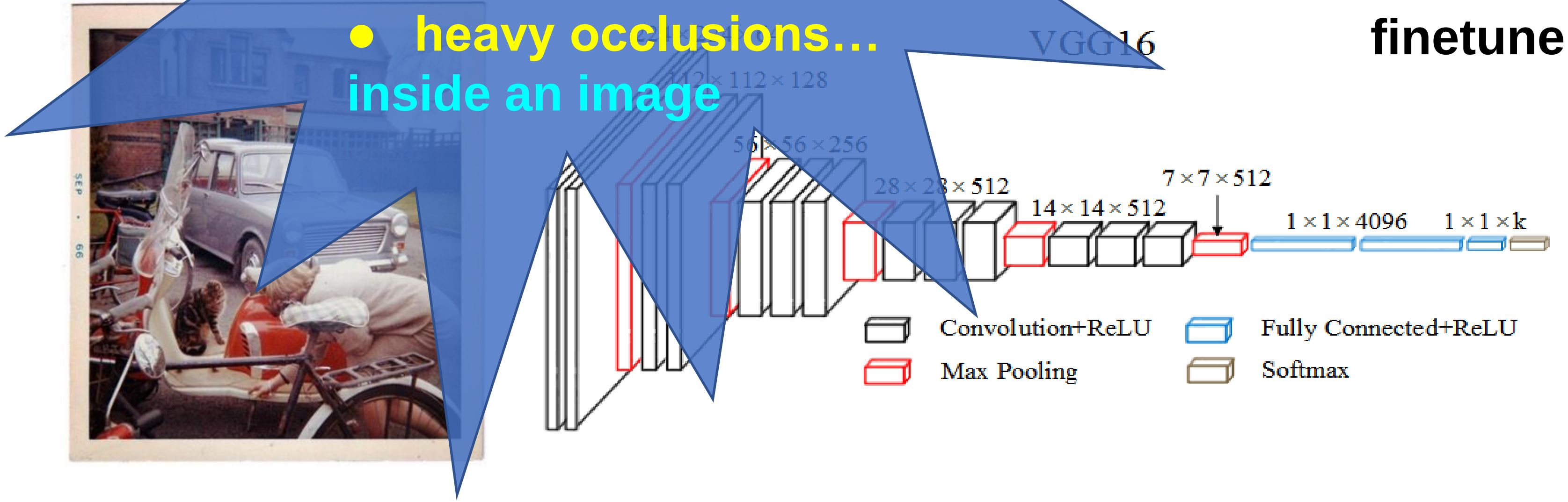
……

Related Work *Effortless method*



pre-train with single label,
e.g., ImageNet

cat



finetune with multiple labels

cat
bicycle
motorcycle
person
car

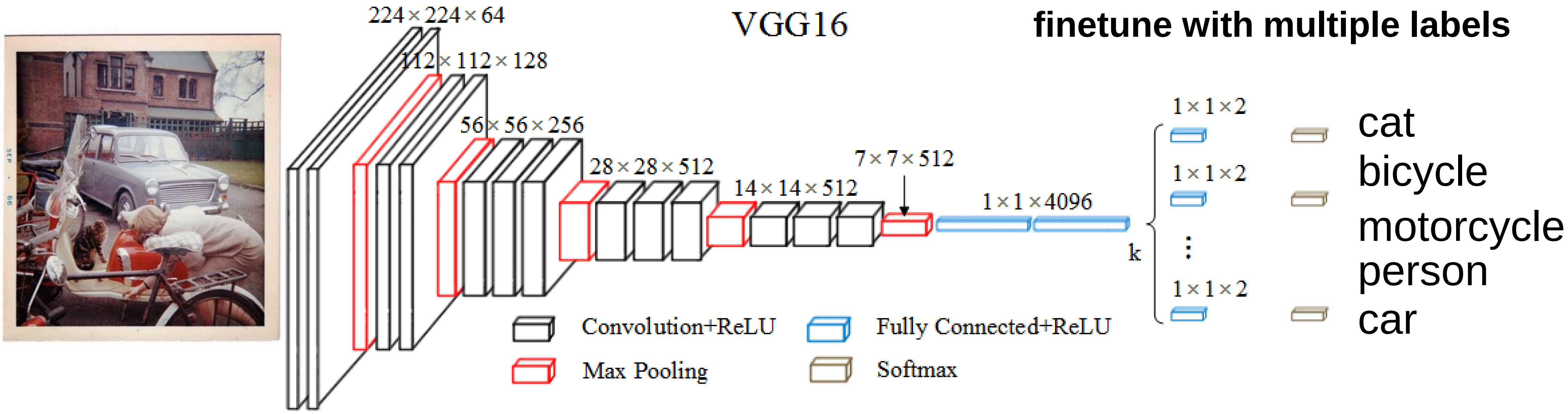
Not generalize well

Reason: multiple classes

- **different locations**
- **diverse scales**
- **heavy occlusions…**

inside an image

Related Work *Binary decomposition*



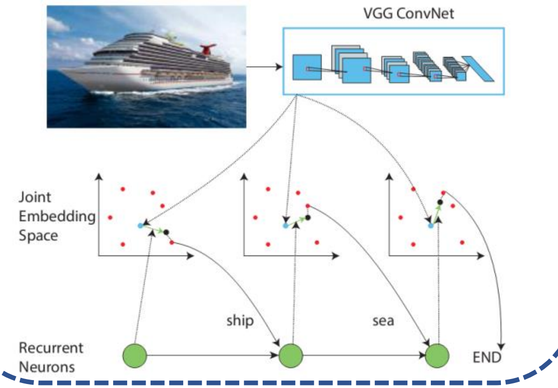
finetune with multiple labels

cat
bicycle
motorcycle
person
car

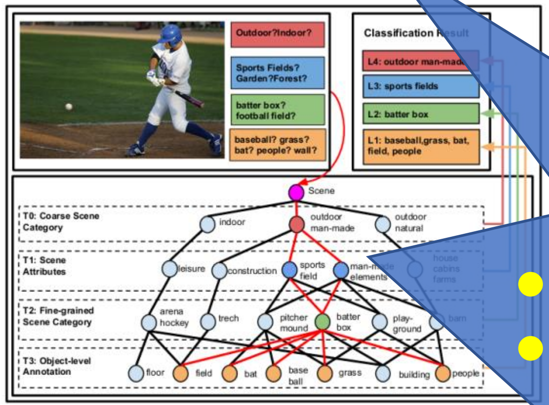
Ignoring **class dependencies**, *e.g.*, person & baby, cat & dog

Related Work *Class dependencies*

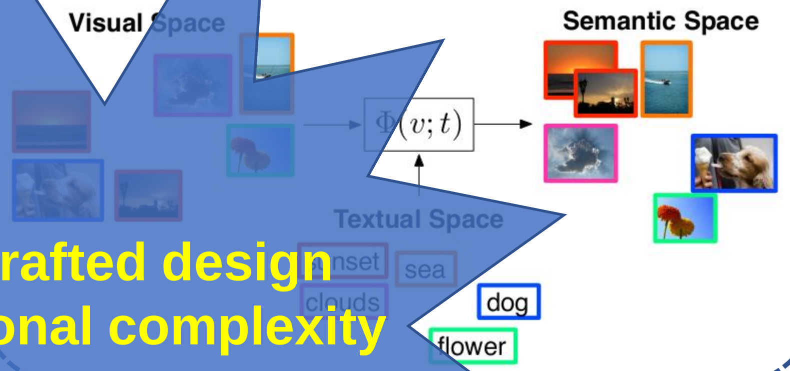
CNN-RNN, CVPR2016



Label relation, CVPR2016

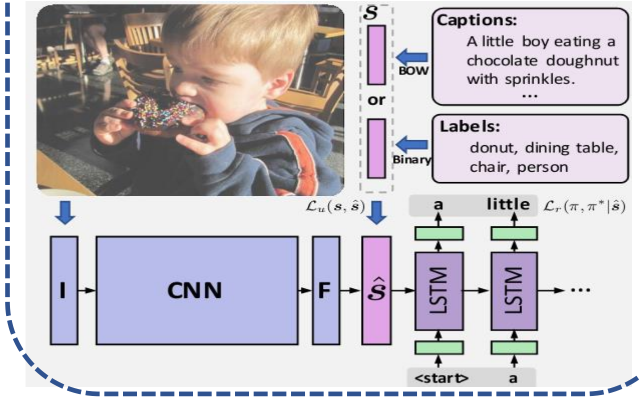


Semantic label transfer, PR2017

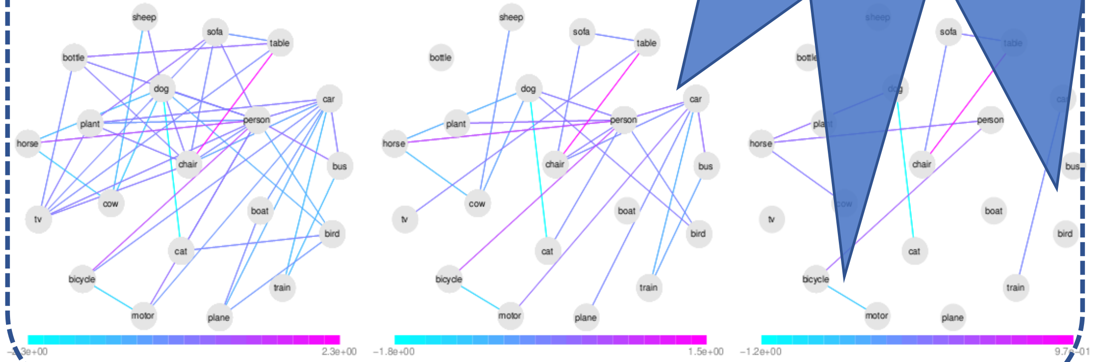


- handcrafted design
- additional complexity

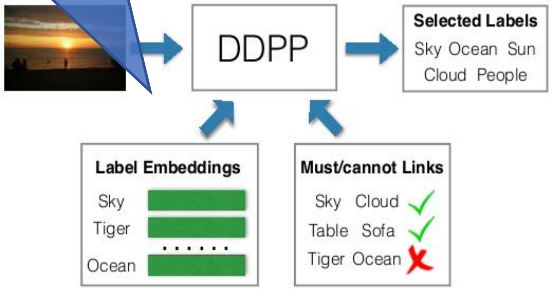
Regularised CNN-RNN, CVPR2017


s
BOW
or
Binary
Captions:
A little boy eating a chocolate doughnut with sprinkles.
...
Labels:
donut, dining table, chair, person
$\mathcal{L}_u(s,\hat{s})$
$\mathcal{L}_r(\pi,\pi^*|\hat{s})$
I
CNN
F
$\hat{s}$
LSTM
LSTM
a
little
<start>
a

Conditional graphical lasso, CVPR2016

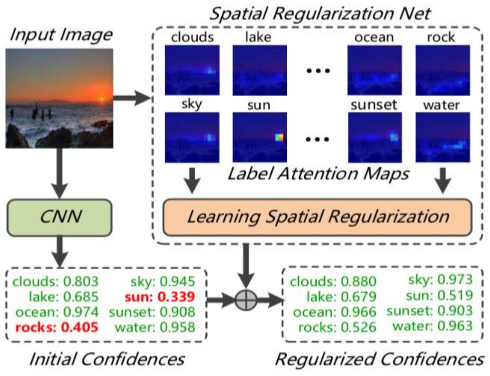


DDPP, ICCV2017

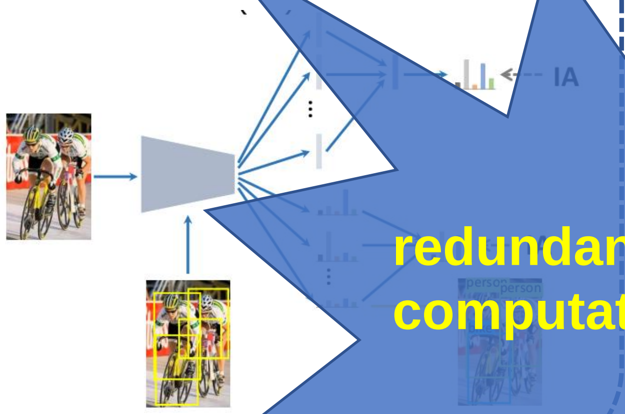


Related Work *Localize semantic regions*

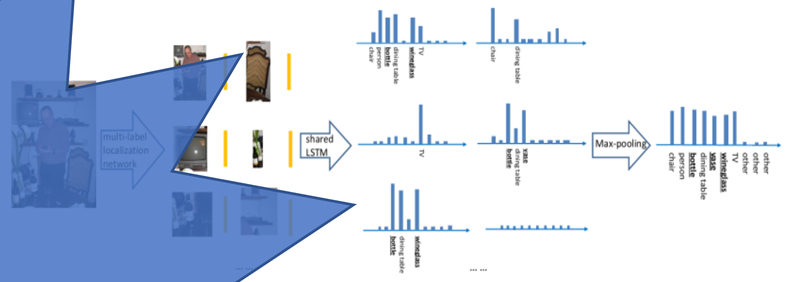
SRN, CVPR2017



Patch learning, PR2017

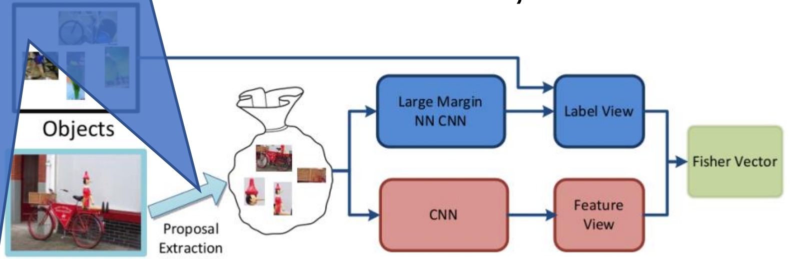


Regional dependencies, 2017

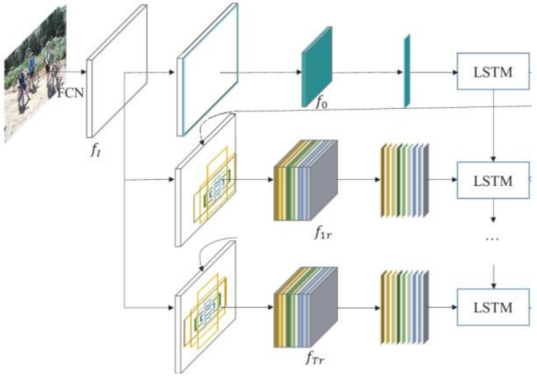


redundant computational cost

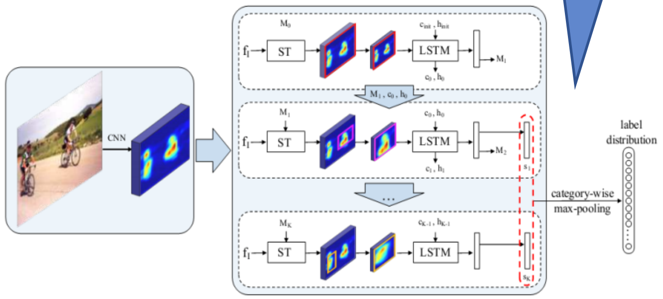
Box Annotation, CVPR2016



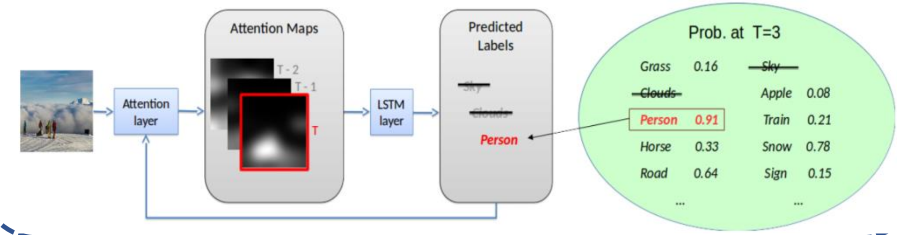
Attentional RL, AAAI2018



Attentional regions, ICCV2017

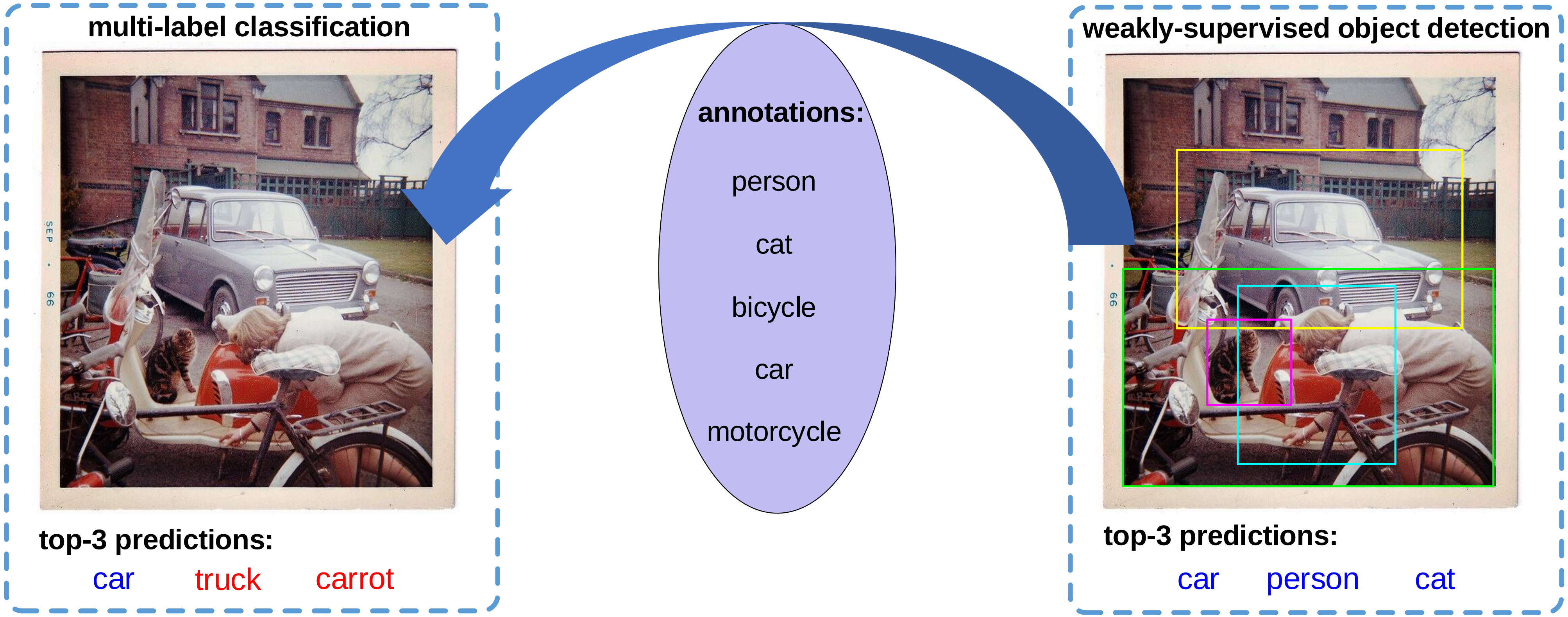


RNN with visual attention, AAAI2018



MLIC-KD-WSD *Motivation*

knowledge distillation



multi-label classification

top-3 predictions:

car truck carrot

annotations:

person

cat

bicycle

car

motorcycle

weakly-supervised object detection

top-3 predictions:

car person cat

poor localization

for multiple instances

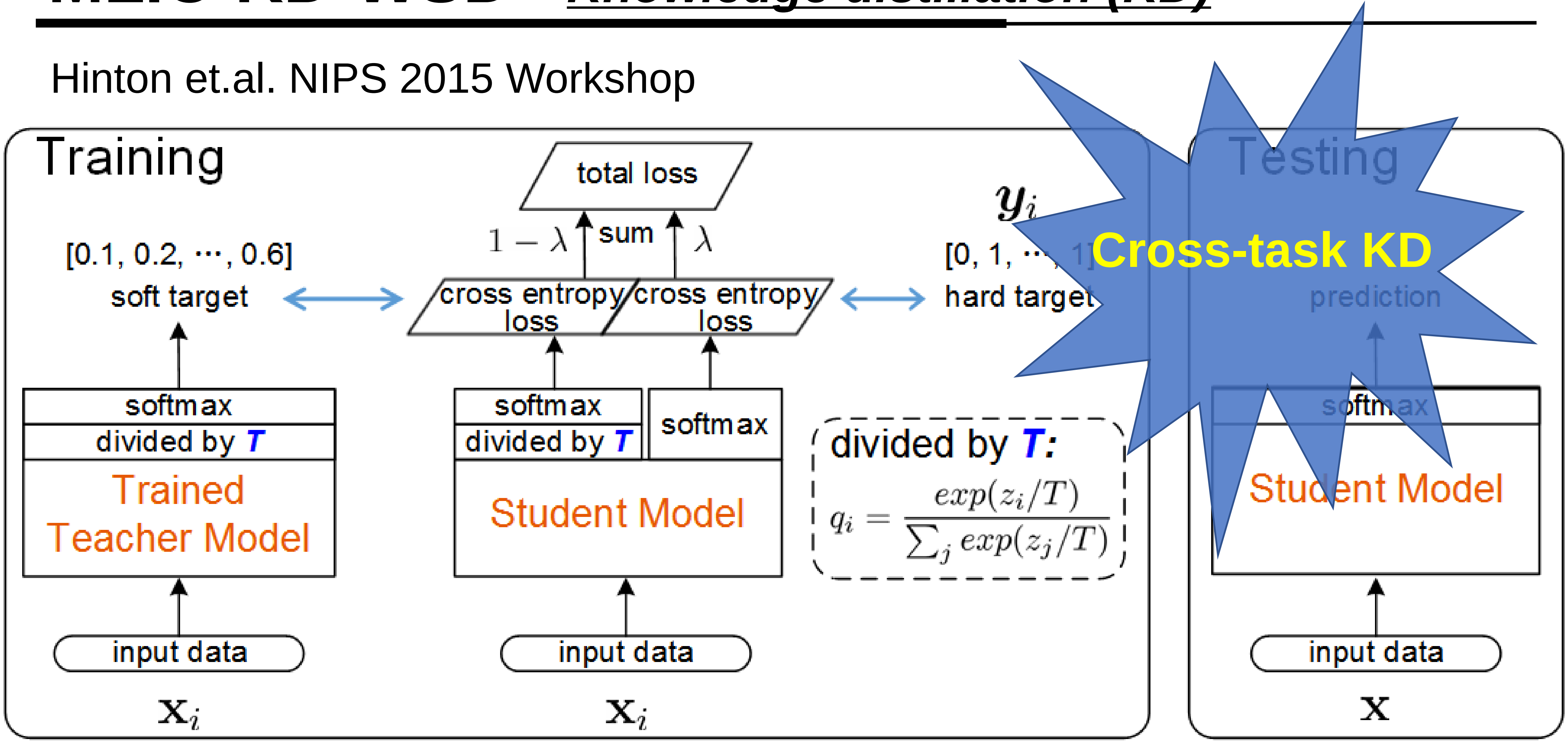
locate semantic regions

object-relevant **informative**

image-level annotation

MLIC-KD-WSD *Knowledge distillation (KD)*

Hinton et.al. NIPS 2015 Workshop



divided by *T*:

$$q_i = \frac{exp(z_i/T)}{\sum_j exp(z_j/T)}$$

Cross-task KD

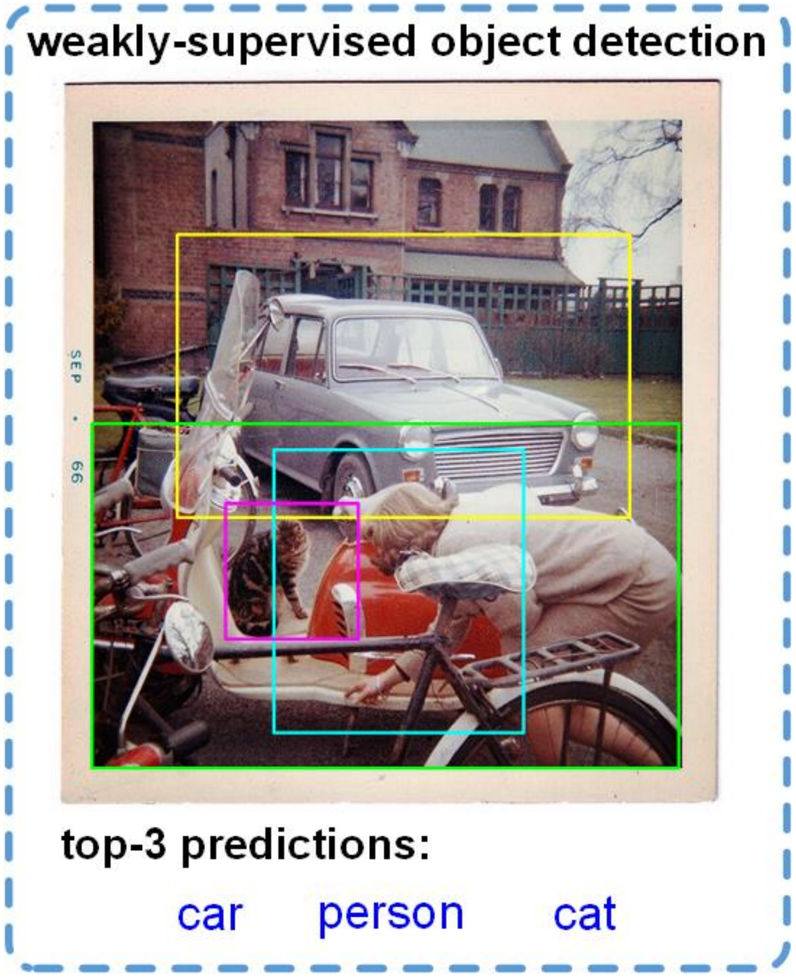
MLIC-KD-WSD *Cross-task KD*

What is the unique knowledge in WSD?

✓ localization of semantic instances

✓ object-level features

✓ class dependencies



MLIC-KD-WSD *Cross-task KD*

Step 1:

Weakly-Supervised Detection

Step 2:

Knowledge Distillation

Stage 1:

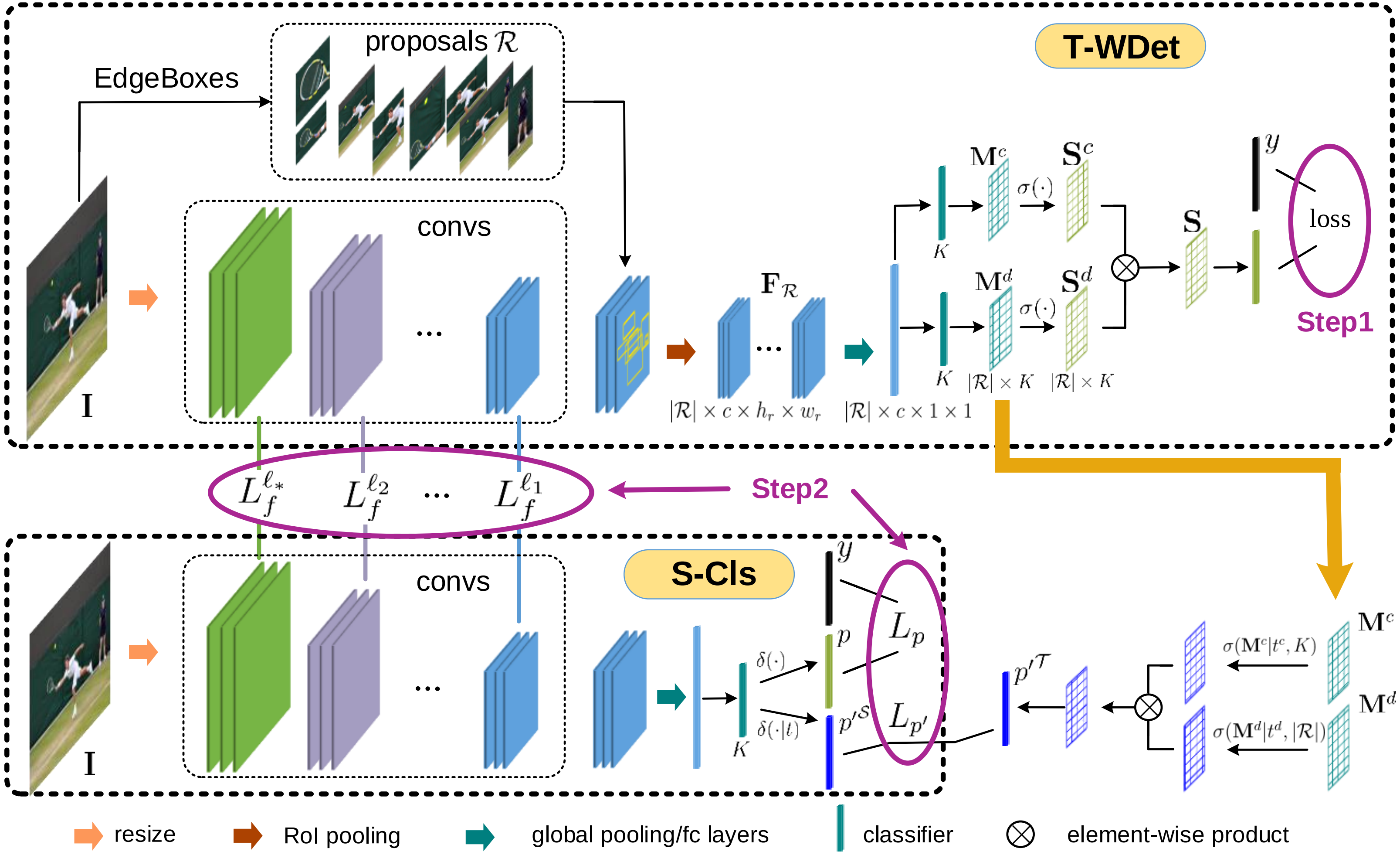
Feature-level transfer

Only update convs' params

Stage 2:

Prediction-level transfer

Update all params



MLIC-KD-WSD *Cross-task KD*

Stage 1:

Feature-level transfer

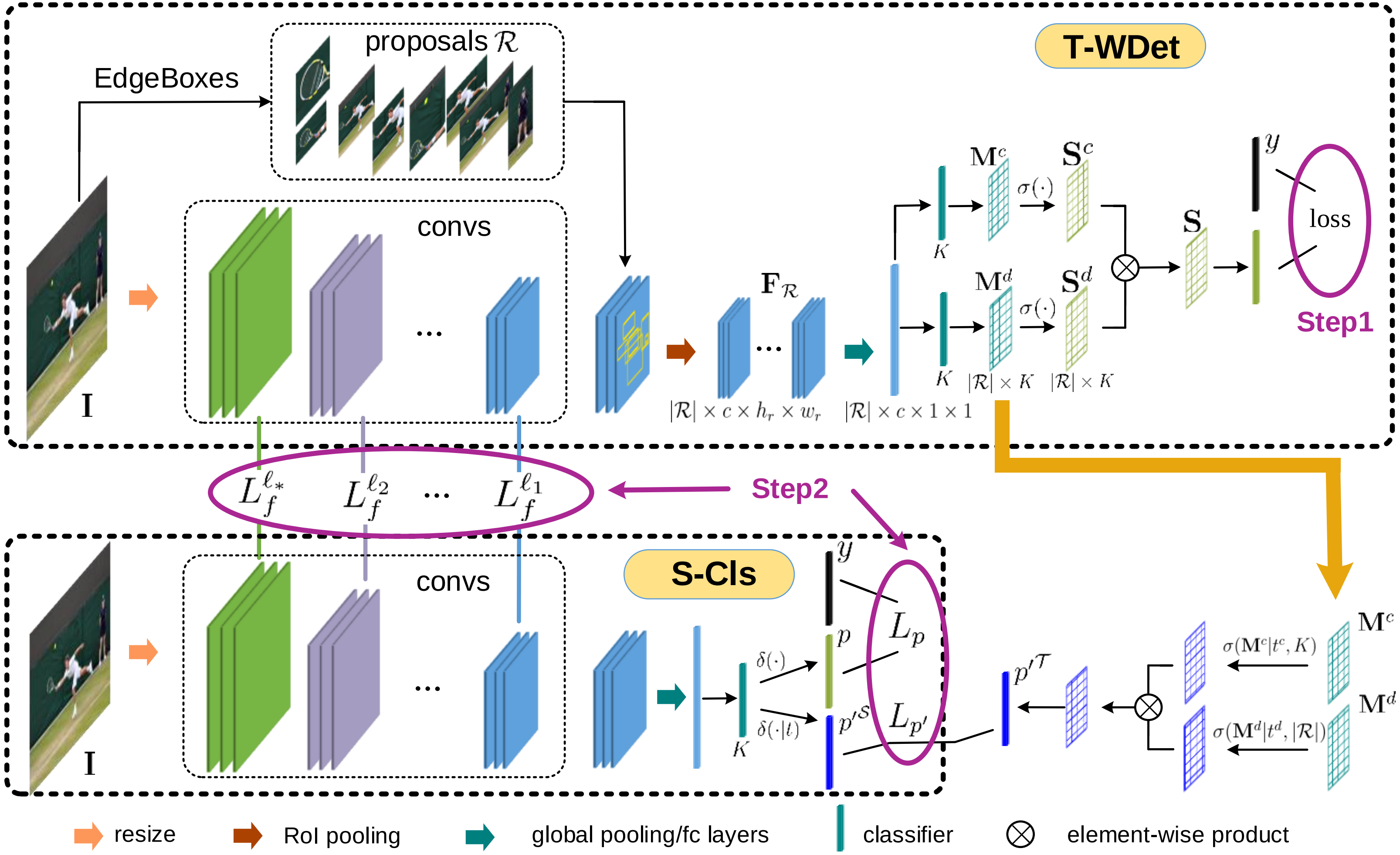
only update convs' params

- ✓ localization of instances
- ✓ object-level features

$$L_f(\mathbf{w}^{\mathcal{S}}_{\text{conv}}) = \frac{1}{2N}\sum\nolimits_n \frac{1}{|\mathcal{R}'_n|}\|\mathbf{F}^{\mathcal{T}}_{\mathcal{R}'_n} \ominus \mathbf{F}^{\mathcal{S}}_{\mathcal{R}'_n}\|_2^2$$

$$\mathbf{F}^{\mathcal{T}}_{\mathcal{R}'_n} = \mathbb{C}_{R\in\mathcal{R}'_n}\left[s'_R \odot \phi_{\text{RoI}}(\mathbf{F}^{\mathcal{T}}_{\text{conv}}; R)\right],$$

$$\mathbf{F}^{\mathcal{S}}_{\mathcal{R}'_n} = \mathbb{C}_{R\in\mathcal{R}'_n}\left[s'_R \odot \phi_{\text{RoI}}(\Psi(\mathbf{F}^{\mathcal{S}}_{\text{conv}})|\mathbf{w}^{\mathcal{S}}_{\text{conv}}; R)\right]$$



MLIC-KD-WSD *Cross-task KD*

Stage 2:

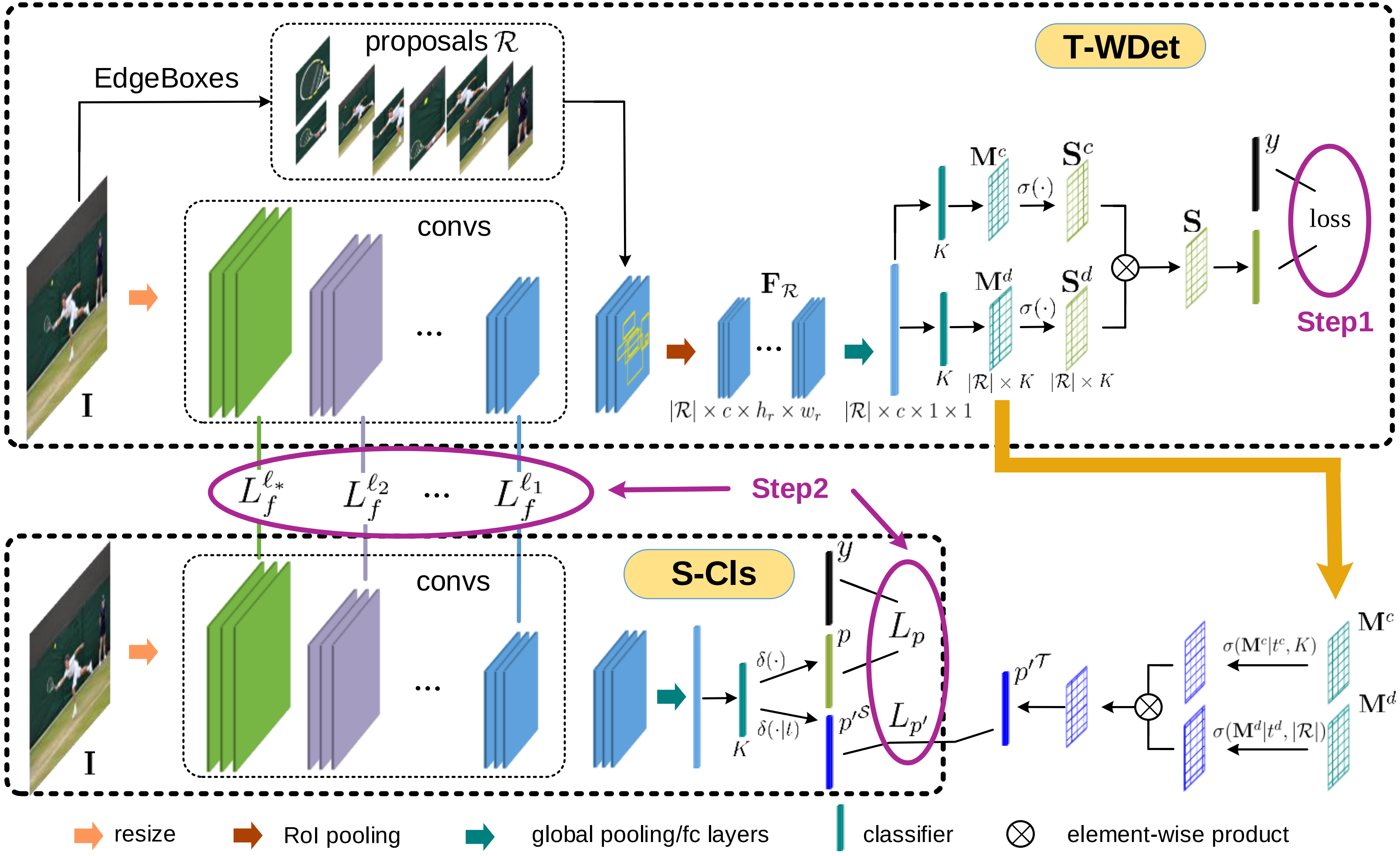
Prediction-level transfer

update all params

✓ class dependencies

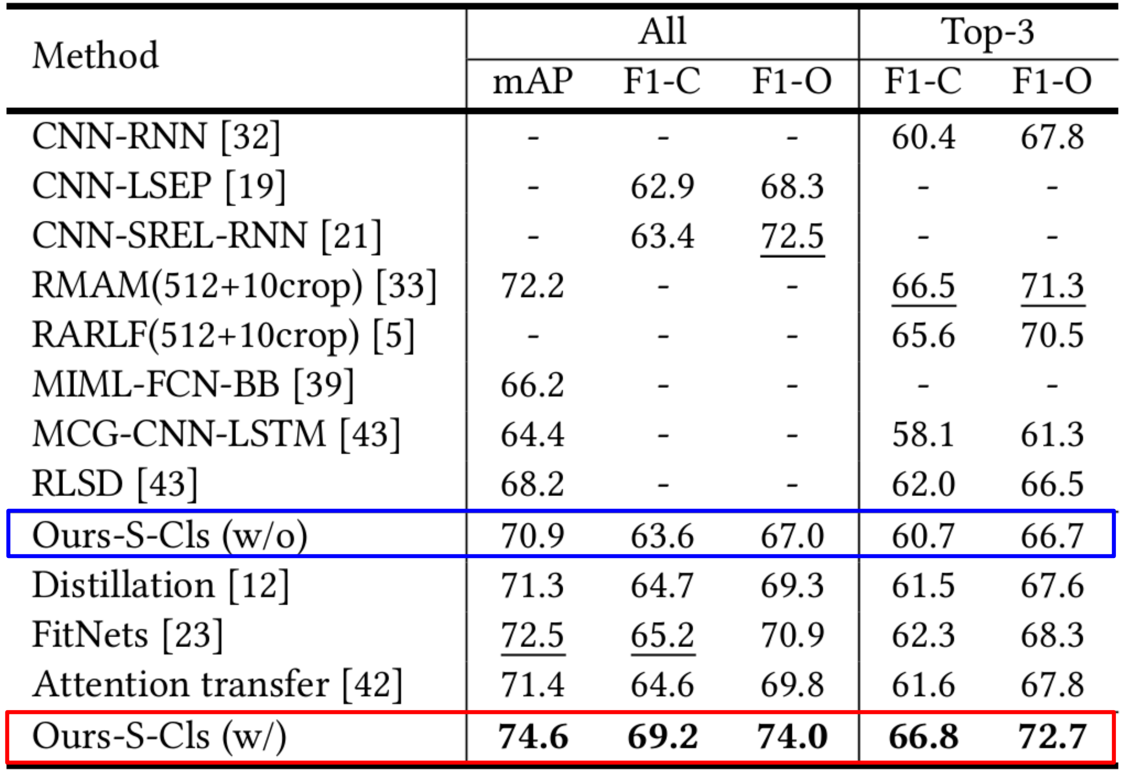
$$L_{p'}(\mathbf{w}^{\mathcal{S}}) = \frac{1}{2N}\sum_n \|p'^{\mathcal{T}} - p'^{\mathcal{S}}(\mathbf{w}^{\mathcal{S}})\|_2^2$$

$$L_p(\mathbf{w}^{\mathcal{S}}) = -\frac{1}{N}\sum_n [y\log p + (1-y)\log(1-p)]$$



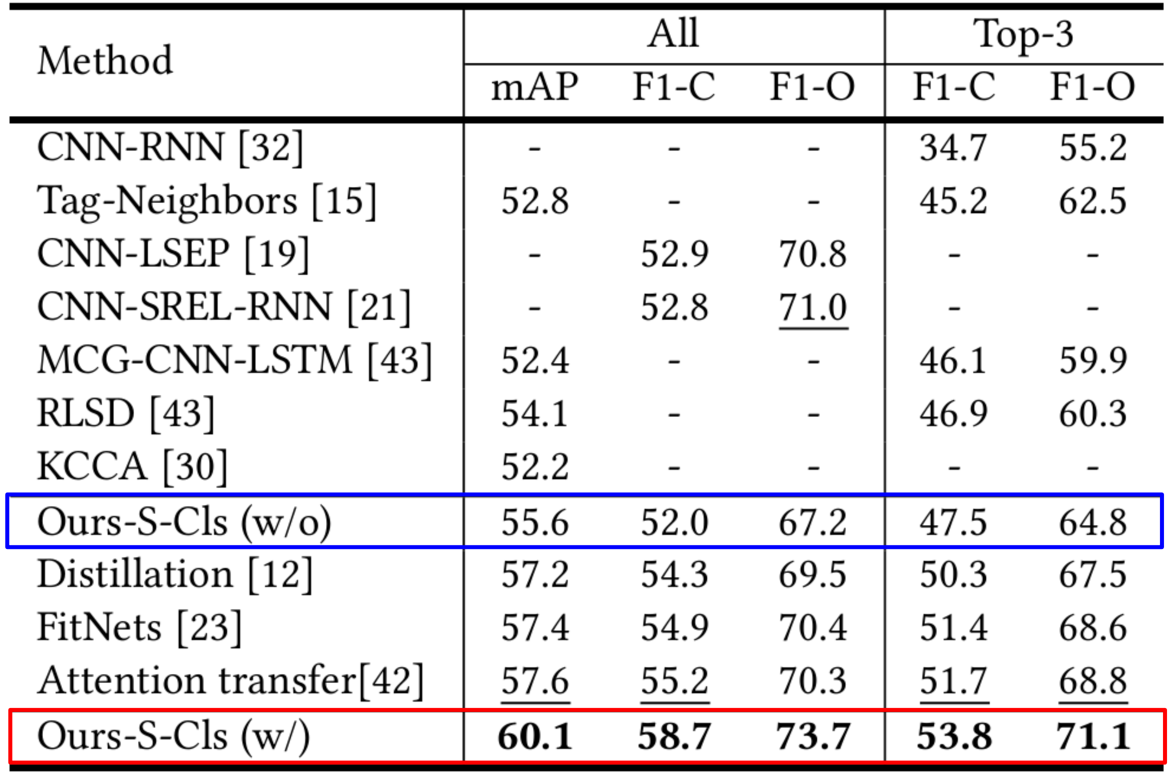
MLIC-KD-WSD *Quantitative results*

MS-COCO: object label



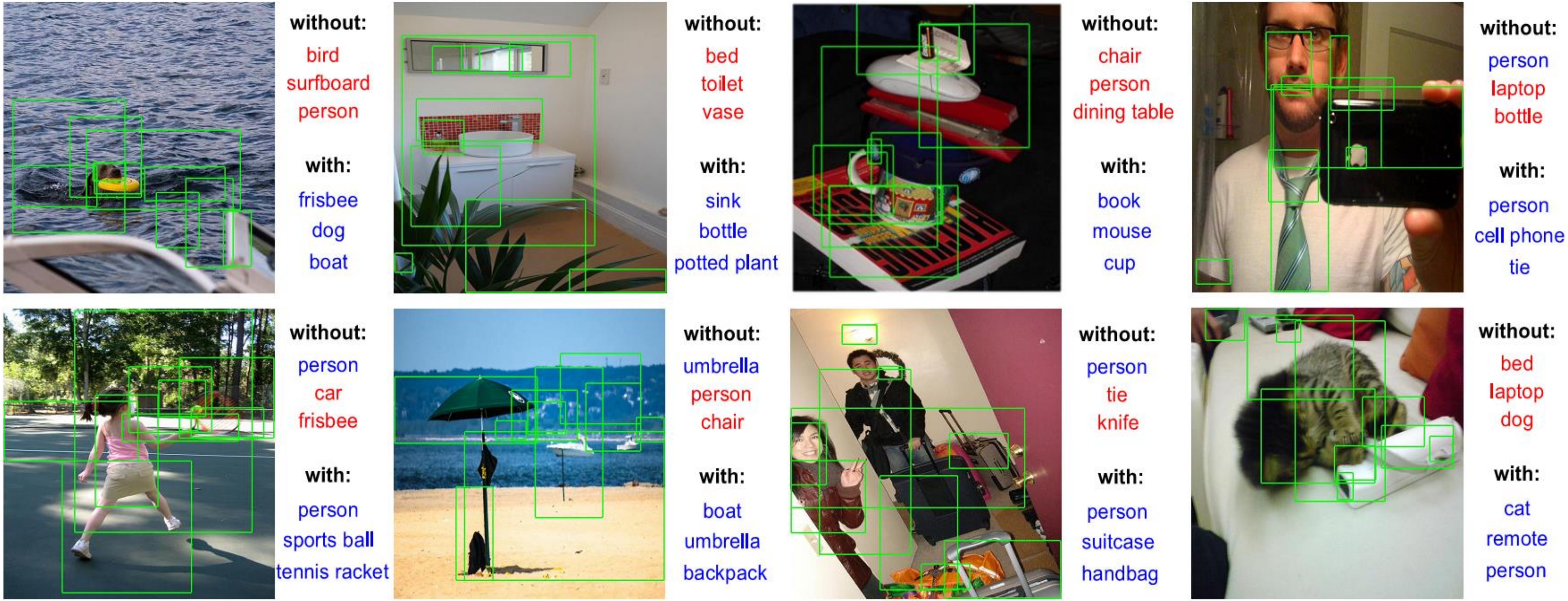
Method	All			Top-3	
	mAP	F1-C	F1-O	F1-C	F1-O
CNN-RNN [32]	-	-	-	60.4	67.8
CNN-LSEP [19]	-	62.9	68.3	-	-
CNN-SREL-RNN [21]	-	63.4	72.5	-	-
RMAM(512+10crop) [33]	72.2	-	-	66.5	71.3
RARLF(512+10crop) [5]	-	-	-	65.6	70.5
MIML-FCN-BB [39]	66.2	-	-	-	-
MCG-CNN-LSTM [43]	64.4	-	-	58.1	61.3
RLSD [43]	68.2	-	-	62.0	66.5
Ours-S-Cls (w/o)	70.9	63.6	67.0	60.7	66.7
Distillation [12]	71.3	64.7	69.3	61.5	67.6
FitNets [23]	72.5	65.2	70.9	62.3	68.3
Attention transfer [42]	71.4	64.6	69.8	61.6	67.8
Ours-S-Cls (w/)	**74.6**	**69.2**	**74.0**	**66.8**	**72.7**

NUS-WIDE: concept label

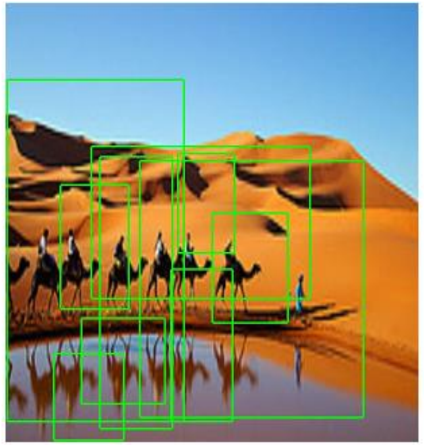


Method	All			Top-3	
	mAP	F1-C	F1-O	F1-C	F1-O
CNN-RNN [32]	-	-	-	34.7	55.2
Tag-Neighbors [15]	52.8	-	-	45.2	62.5
CNN-LSEP [19]	-	52.9	70.8	-	-
CNN-SREL-RNN [21]	-	52.8	71.0	-	-
MCG-CNN-LSTM [43]	52.4	-	-	46.1	59.9
RLSD [43]	54.1	-	-	46.9	60.3
KCCA [30]	52.2	-	-	-	-
Ours-S-Cls (w/o)	55.6	52.0	67.2	47.5	64.8
Distillation [12]	57.2	54.3	69.5	50.3	67.5
FitNets [23]	57.4	54.9	70.4	51.4	68.6
Attention transfer[42]	57.6	55.2	70.3	51.7	68.8
Ours-S-Cls (w/)	**60.1**	**58.7**	**73.7**	**53.8**	**71.1**

MLIC-KD-WSD *Results on MS-COCO*

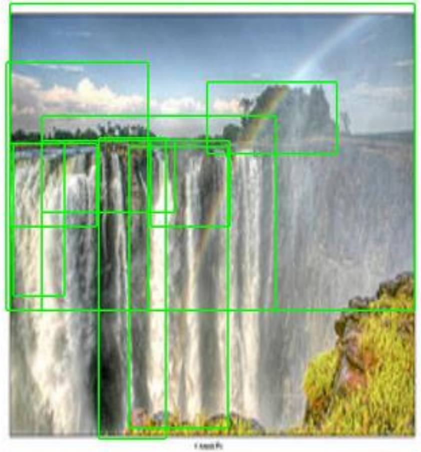


MLIC-KD-WSD *Results on NUS-WIDE*



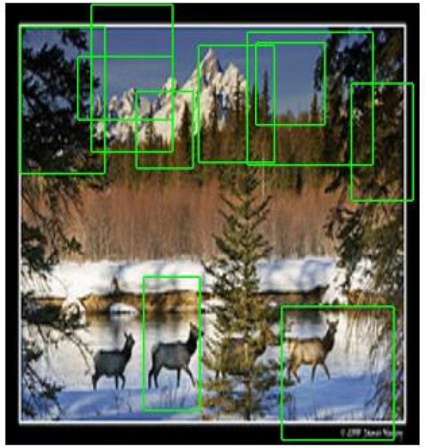
without:
sky
horses
clouds

with:
animal
reflection
sand



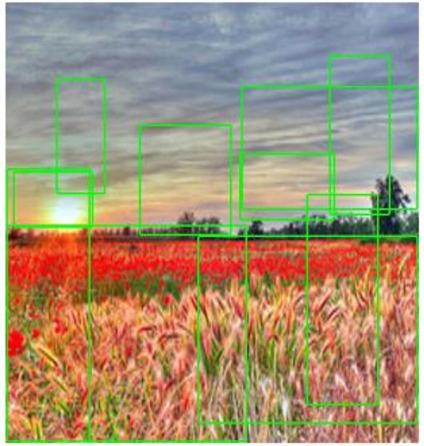
without:
sky
grasss
flowers

with:
waterfall
sky
rainbow



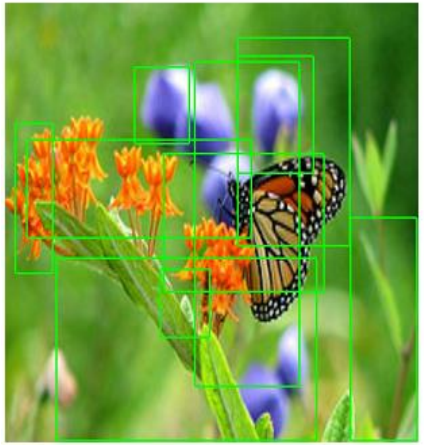
without:
reflection
cow
ocean

with:
mountain
sky
animal



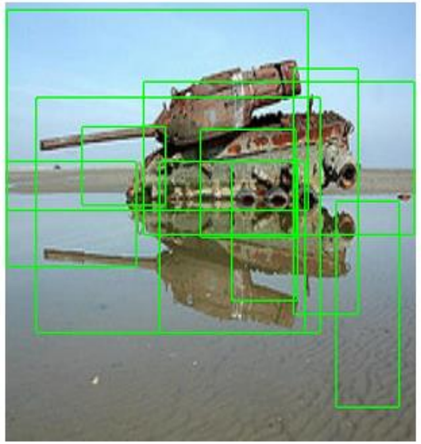
without:
sky
food
birds

with:
sky
flowers
sunset



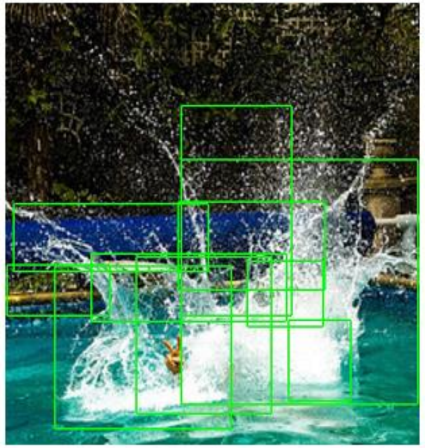
without:
plants
rainbow
food

with:
animal
flowers
leaf



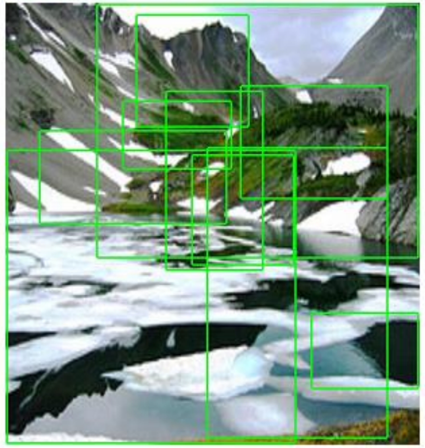
without:
sky
birds
boats

with:
military
sky
water



without:
lake
waterfall
plants

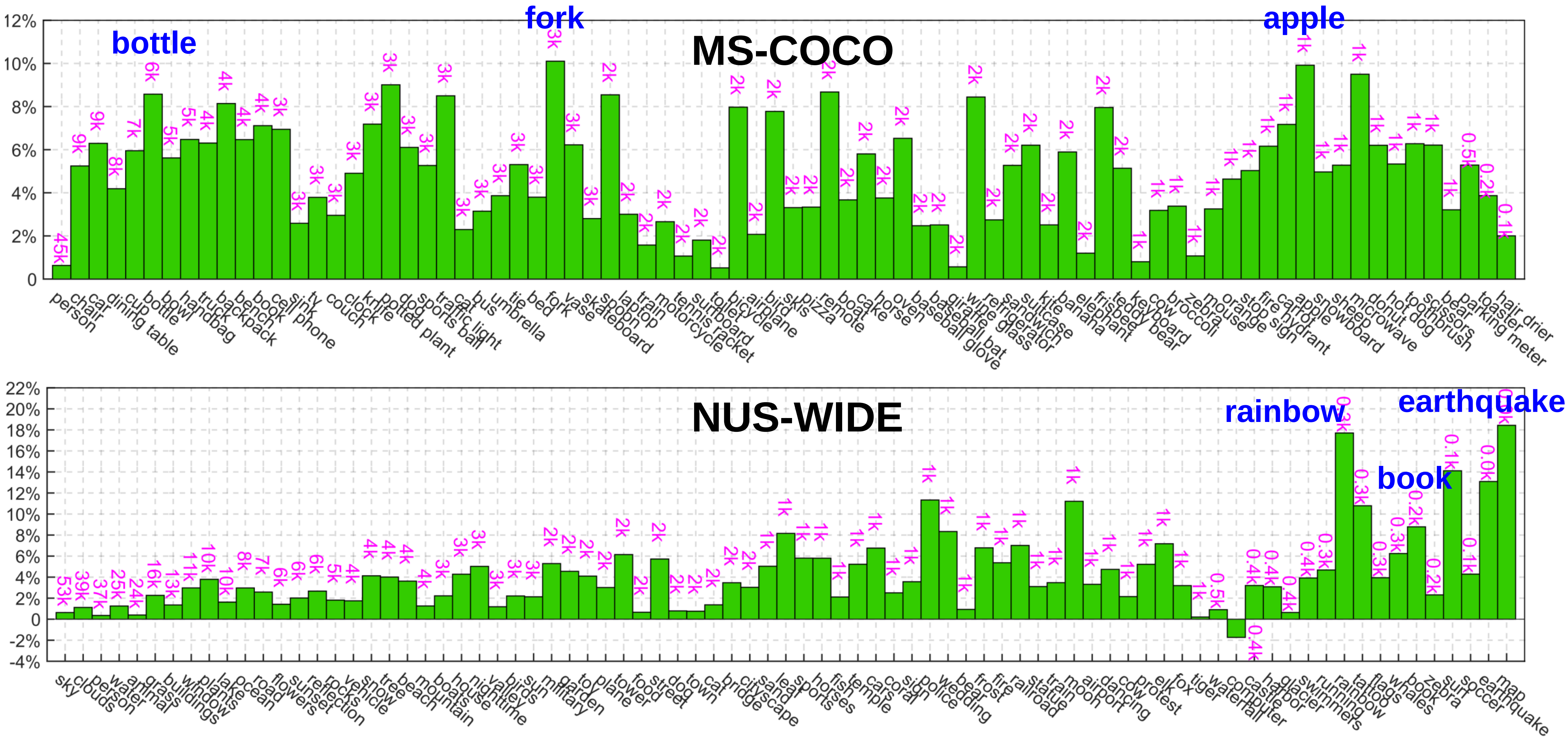
with:
water
sports
swimmers



without:
garden
earthquake
waterfall

with:
rocks
water
valley

MLIC-KD-WSD ***Robustness***



References

- Jiang Wang, Yi Yang, Junhua Mao, Zhiheng Huang, Chang Huang, Wei Xu: CNN-RNN: A Unified Framework for Multi-label Image Classification. CVPR 2016: 2285-2294
- Hexiang Hu, Guang-Tong Zhou, Zhiwei Deng, Zicheng Liao, Greg Mori: Learning Structured Inference Neural Networks with Label Relations. CVPR 2016: 2960-2968
- Uricchio et.al. Automatic Image Annotation via Label Transfer in the Semantic Space. Pattern Recognition 71: 144-157 (2017)
- Pengtao Xie, Ruslan Salakhutdinov, Luntian Mou, Eric P. Xing: Deep Determinantal Point Process for Large-Scale Multi-label Classification. ICCV 2017: 473-482
- Feng Liu, Tao Xiang, Timothy M. Hospedales, Wankou Yang, Changyin Sun: Semantic Regularisation for Recurrent Image Annotation. CVPR 2017: 4160-4168
- Qiang Li, Maoying Qiao, Wei Bian, Dacheng Tao: Conditional Graphical Lasso for Multi-label Image Classification. CVPR 2016: 2977-2986
- Junjie Zhang, Qi Wu, Jian Zhang, Chunhua Shen, Jianfeng Lu: Kill Two Birds With One Stone: Weakly-Supervised Neural Network for Image Annotation and Tag Refinement. *AAAI* 2018: 7550-7557

References

- Feng Zhu, Hongsheng Li, Wanli Ouyang, Nenghai Yu, Xiaogang Wang: Learning Spatial Regularization with Image-Level Supervisions for Multi-label Image Classification. *CVPR* 2017: 2027-2036
- Peng Tang, Xinggang Wang, Zilong Huang, Xiang Bai, Wenyu Liu: Deep patch learning for weakly supervised object classification and discovery. Pattern Recognition 71: 446-459 (2017)
- Junjie Zhang, Qi Wu, Chunhua Shen, Jian Zhang, Jianfeng Lu: Multi-Label Image Classification with Regional Latent Semantic Dependencies. CoRR abs/1612.01082 (2017)
- Shang-Fu Chen, Yi-Chen Chen, Chih-Kuan Yeh, Yu-Chiang Frank Wang: Order-Free RNN With Visual Attention for Multi-Label Classification. AAAI 2018: 6714-6721
- Zhouxia Wang, Tianshui Chen, Guanbin Li, Ruijia Xu, Liang Lin: Multi-label Image Recognition by Recurrently Discovering Attentional Regions. ICCV 2017: 464-472
- Tianshui Chen, Zhouxia Wang, Guanbin Li, Liang Lin: Recurrent Attentional Reinforcement Learning for Multi-Label Image Recognition. AAAI 2018: 6730-6737
- Geoffrey E. Hinton, Oriol Vinyals, Jeffrey Dean: Distilling the Knowledge in a Neural Network. NIPS workshop 2015

Thanks for your attention !

Q & A