南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Dual-Decoupling Learning and Metric-Adaptive Thresholding for Semi-Supervised Multi-Label Learning

Jia-Hao Xiao [1], Ming-Kun Xie [1], Heng-Bo Fan [1], Gang Niu [2], Masashi Sugiyama [2,3] and Sheng-Jun Huang [1]

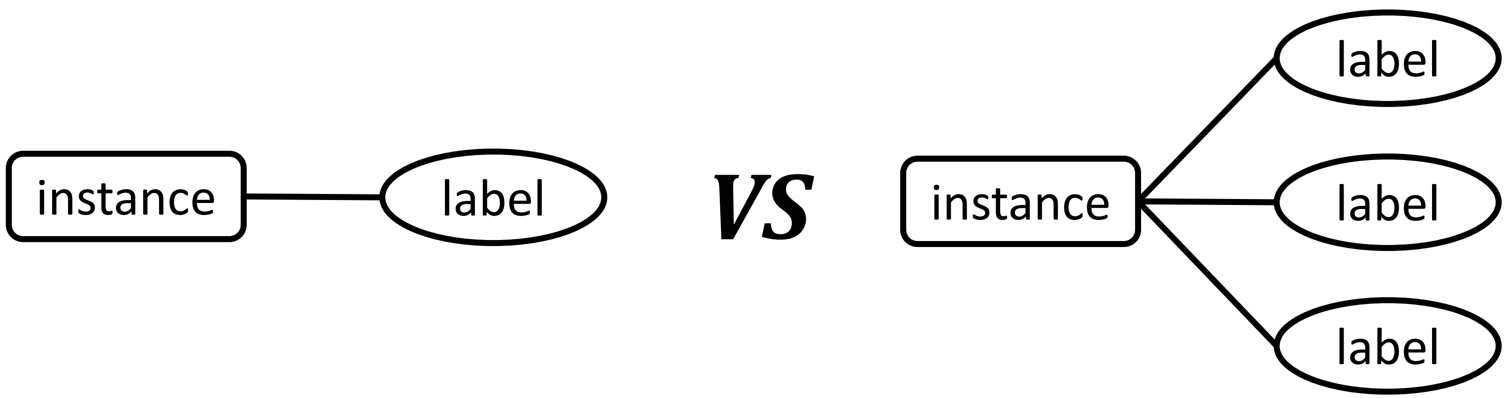
[1] Nanjing University of Aeronautics and Astronautics, Nanjing, China
[2] RIKEN Center for Advanced Intelligence Project, Tokyo, Japan
[3] The University of Tokyo, Tokyo, Japan

{jiahaoxiao, mkxie, fan_heng_bo, huangsj}@nuaa.edu.cn
gang.niu.ml@gmail.com   sugi@k.u-tokyo.ac.jp

ECCV 2024

Multi-Label Learning

- Multi-Label Learning vs. Ordinary Supervised Learning

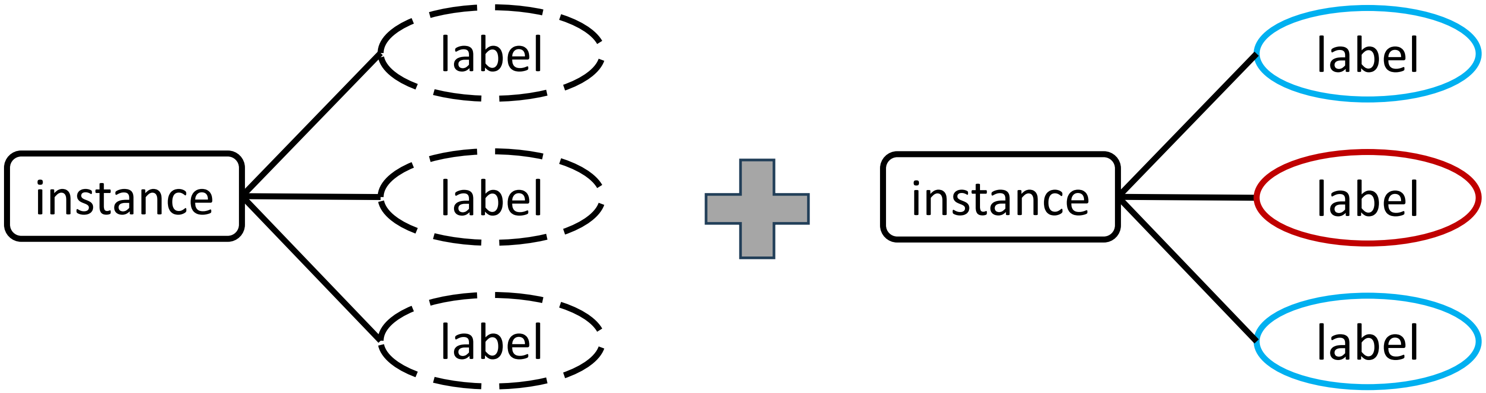


Ordinary supervised learning
(only one ground-truth label)

Multi-Label learning
(multiple ground-truth labels)

Semi-Supervised Multi-Label Learning

- A *Great* number of unlabeled data and *few* labeled data



Unlabeled data **Labeled data**

Motivation

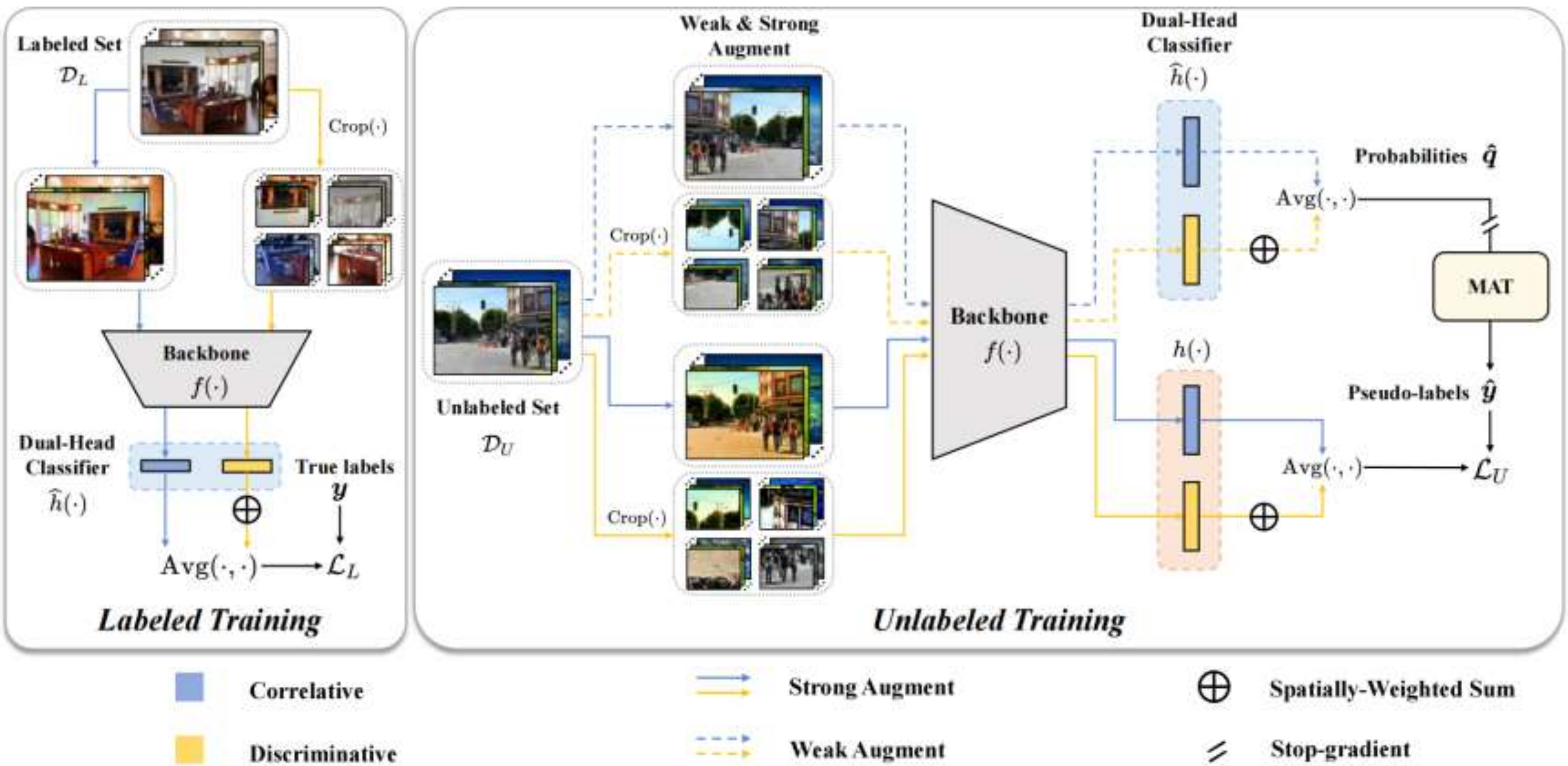
- How to generate ***high-quality*** pseudo-labels?
  - Depend on two key factors:
    ① Model Prediction;   ② Selection Strategy.
  - **Previous methods** merely focused on capturing the true class proportions, while neglecting the quality of model predictions.
  - **Our methods** aims to generate high-quality pseudo-labels from both:

    (a) develop the **D**ual-**D**ecoupling **L**earning (**D2L**) framework to obtain model predictions;

    (b) design the **M**etric-**A**daptive **T**hresholding (**MAT**) method to acquire proper thresholds.

Method - ①

- **D**ual-**D**ecoupling **L**earning (**D2L**) framework
  - (1) *Correlative and Discriminative Features Decoupling*
  - (2) *Generation and Utilization of Pseudo-Labels Decoupling*



Method - ②

- **M**etric-**A**daptive **T**hresholding (**MAT**) method

  Search the best threshold for each class.

$$\forall k \in [K], \quad \tau_k^\star = \arg\max_{\tau_k \in [0,1]} \mathcal{M}(\hat{Y}_k, Y_k)$$

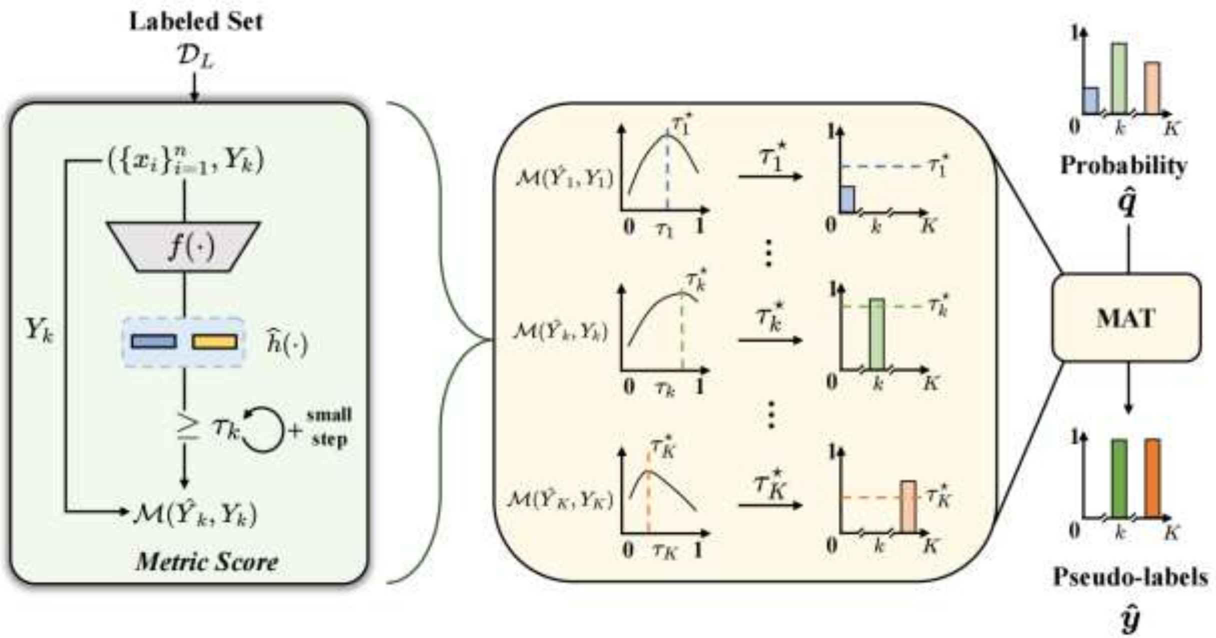


Fig. 5: An illustration of MAT. By feeding instances into the model $f(\cdot) \circ \widehat{h}(\cdot)$, we obtain the predictions. By adjusting τ_k, we can achieve the optimal pseudo-labeling performance $\mathcal{M}(\hat{Y}_k, Y_k)$.

Method - ①+②

Algorithm 1 Pseudo code of the proposed algorithm.

Input: Labeled data $\mathcal{D}_L = \{(\boldsymbol{x}_i)_{i=1}^N, Y\}$, Unlabeled data $\mathcal{D}_U = \{\boldsymbol{x}_j\}_{j=1}^M$, backbone $f(\cdot)$, two dual-head classifiers $\{\widehat{h}^g(\cdot), \widehat{h}^l(\cdot)\}$ and $\{h^g(\cdot), h^l(\cdot)\}$, metric function $\mathcal{M}(\cdot,\cdot)$, class number K, small step t.

Warm up the backbone $f(\cdot)$ and one classifier $\{\widehat{h}^g(\cdot), \widehat{h}^l(\cdot)\}$ on $\mathcal{D}_L$ with Eq. (1).
for each *epoch* **do**
    Input labeled data $\{\boldsymbol{x}_i\}_{i=1}^N$ into $f(\cdot)$ and $\{\widehat{h}^g(\cdot), \widehat{h}^l(\cdot)\}$ to get outputs $\{\hat{\boldsymbol{q}}_i\}_{i=1}^N$.
    for $\forall k \in [K], \tau_k = 0$ to 1 by t **do**
        Pseudo-label $\mathcal{D}_L$ in class k by τ_k, $\hat{Y}_k = \{\hat{y}_{ik}\}_{i=1}^N = \{\mathbb{I}(\hat{q}_{ik} \geq \tau_k)\}_{i=1}^N$.
        Select the τ_k which achieves the highest $\mathcal{M}(\hat{Y}_k, Y_k)$ as $\tau_k^\star$ (Eq. (6)).
    end for
    Pseudo-label $\mathcal{D}_U$ with Eq. (5), then train $f(\cdot)$, $\{\widehat{h}^g(\cdot), \widehat{h}^l(\cdot)\}$ and $\{h^g(\cdot), h^l(\cdot)\}$ on $\mathcal{D}_L$ and $\mathcal{D}_U$ together using the D2L framework as shown in Fig. 1.
end for

Experiments - Main Results

Results on VOC.

Method	BCE	ASL	LL-*	PLC	Top-*	IAT	ADSH	FM	DRML	CAP	Ours
$p = 0.01$	16.71	34.81	36.01	43.91	38.61	34.39	45.06	44.98	38.90	41.28	**49.09**
$p = 0.05$	67.95	71.46	75.79	74.49	75.77	73.24	75.37	75.11	61.77	76.16	**79.26**
$p = 0.10$	75.35	78.00	81.04	80.35	80.78	80.27	80.34	80.66	71.01	82.16	**84.06**
$p = 0.15$	78.19	79.69	82.36	82.35	82.65	82.39	82.80	82.63	72.98	83.48	**86.25**
$p = 0.20$	79.38	80.77	83.68	83.39	83.72	83.55	83.93	83.60	74.49	84.41	**87.16**

Results on COCO.

Method	BCE	ASL	LL-*	PLC	Top-*	IAT	ADSH	FM	DRML	CAP	Ours
$p = 0.01$	44.11	44.87	45.36	48.95	48.40	46.41	47.93	47.10	39.12	52.40	**56.59**
$p = 0.05$	58.90	59.12	59.33	59.85	58.25	60.34	60.75	59.94	53.60	62.43	**69.30**
$p = 0.10$	63.75	63.82	64.25	65.03	63.52	65.54	65.37	64.46	57.06	67.36	**73.06**
$p = 0.15$	65.91	66.10	66.69	67.62	66.11	67.88	67.70	66.79	58.53	69.11	**74.63**
$p = 0.20$	67.33	67.51	68.12	69.14	67.49	69.25	69.01	68.04	59.24	70.41	**75.70**

Results on NUS.

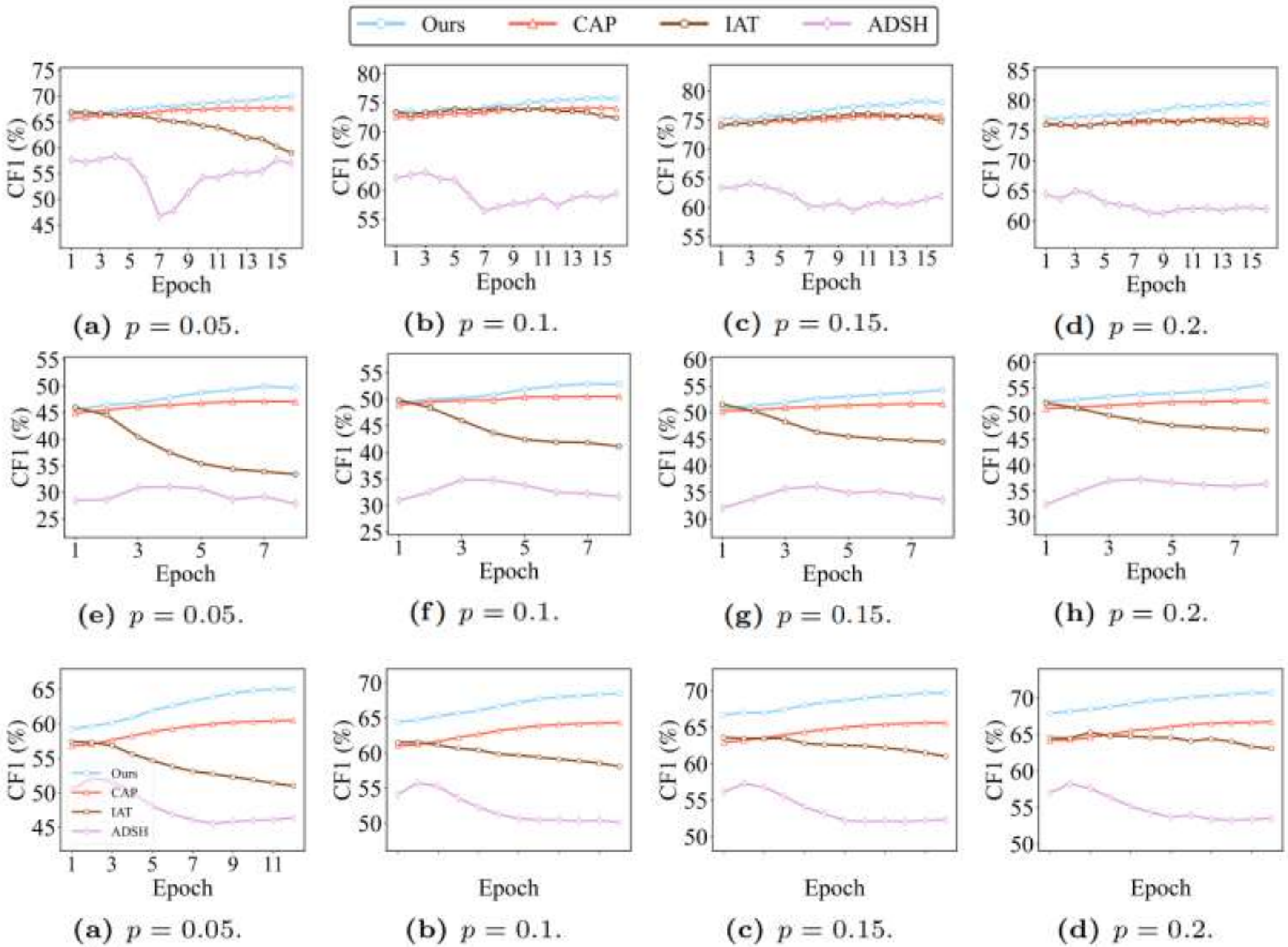
Method	BCE	ASL	LL-*	PLC	Top-*	IAT	ADSH	FM	DRML	CAP	Ours
$p = 0.01$	29.58	30.51	20.70	33.59	26.84	26.28	33.13	32.10	17.40	24.75	**38.09**
$p = 0.05$	41.09	42.87	40.20	43.55	40.99	42.58	43.94	43.12	30.61	44.82	**46.86**
$p = 0.10$	45.39	46.50	44.95	47.51	45.07	46.60	47.28	46.65	35.09	48.24	**50.25**
$p = 0.15$	47.30	48.42	47.32	49.75	47.43	48.76	49.22	48.74	37.91	49.90	**51.61**
$p = 0.20$	48.36	49.65	48.31	50.71	48.49	49.62	49.93	49.59	39.98	51.06	**52.64**

Experiments - Performance of Pseudo-labeling

PASCAL VOC 2012

NUS-WIDE

MS-COCO 2014



(a) $p = 0.05.$ (b) $p = 0.1.$ (c) $p = 0.15.$ (d) $p = 0.2.$

(e) $p = 0.05.$ (f) $p = 0.1.$ (g) $p = 0.15.$ (h) $p = 0.2.$

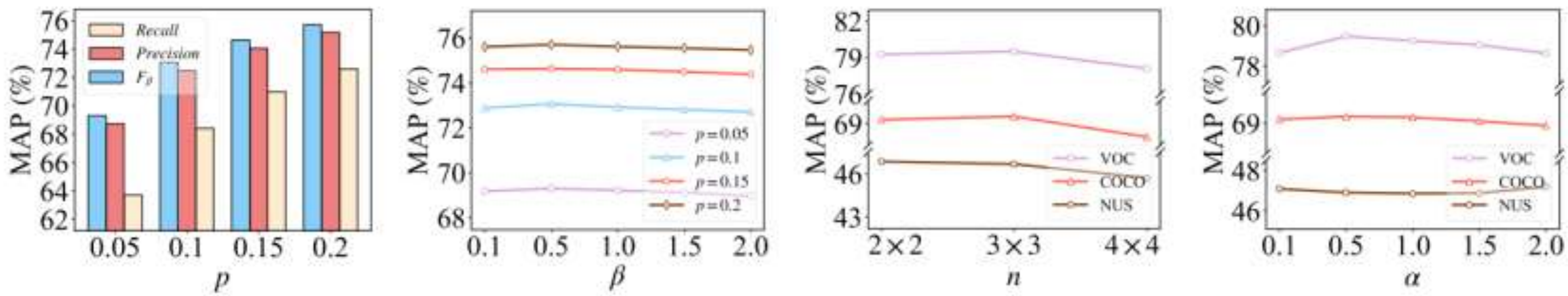
(a) $p = 0.05.$ (b) $p = 0.1.$ (c) $p = 0.15.$ (d) $p = 0.2.$

Experiments - Reproducibility and Resource Consumption

Table 4: Mean and standard deviation of mAP(%) in CAP and our method, on three datasets, along with the time/memory comparison. 'Time' is the training time per epoch, including the process of threshold updating, 'GPU' is the max memory allocated during training phase.

Methods	CAP			Ours		
Datasets	VOC	COCO	NUS	VOC	COCO	NUS
$p = 0.05$	77.15±0.58	63.11±0.35	45.30±0.30	81.45±1.50	70.15±0.48	47.42±1.00
$p = 0.10$	82.54±0.20	67.96±0.32	48.89±0.37	85.65±0.92	73.65±0.34	51.01±0.43
$p = 0.15$	83.95±0.24	69.92±0.41	50.53±0.54	87.02±0.67	75.18±0.31	52.15±0.46
$p = 0.20$	85.04±0.32	71.23±0.42	51.82±0.43	87.83±0.38	76.21±0.30	53.37±0.43
Time	0.9min	10.3min	12.2min	1.7min	21.8min	38.9min
GPU	11.1G			14.2G		

Experiments - Parameter Sensitivity Analyses



(a) The ablation study on metric $\mathcal{M}(\cdot,\cdot)$.

(b) The ablation study on β in metric F_β.

(c) The ablation study on number of patches n.

(d) The ablation study on temperature α.

Fig. 3: The analyses of parameters in D2L and MAT: (a-b) The results of various metric functions $\mathcal{M}(\cdot,\cdot)$ used in MAT and different β values used in metric F_β, at $p = \{0.05, 0.1, 0.15, 0.2\}$ on COCO; (c-d) The analyses of two parameters, number of patches n and temperature α in D2L framework, at $p = 0.05$ on three datasets. The parameter analyses under other settings will be presented in Appendix E.

Experiments - Ablation Study

Table 2: Mean average precision (mAP %) of the baseline incorporated with different components, on datasets VOC and COCO. The baseline here indicates the method CAP (the results of the first row, without any components).

MAT	D2L		VOC				COCO			
	CDD	GUD	p=0.05	p=0.10	p=0.15	p=0.20	p=0.05	p=0.10	p=0.15	p=0.20
			76.16	82.16	83.48	84.41	62.43	67.36	69.11	70.41
✓			76.87	82.59	84.29	85.16	65.03	68.87	70.54	71.54
✓	✓		77.11	83.48	85.72	86.55	66.07	70.72	72.92	74.26
✓	✓	✓	79.26	84.06	86.25	87.16	69.30	73.06	74.63	75.70

Experiments - Case Study

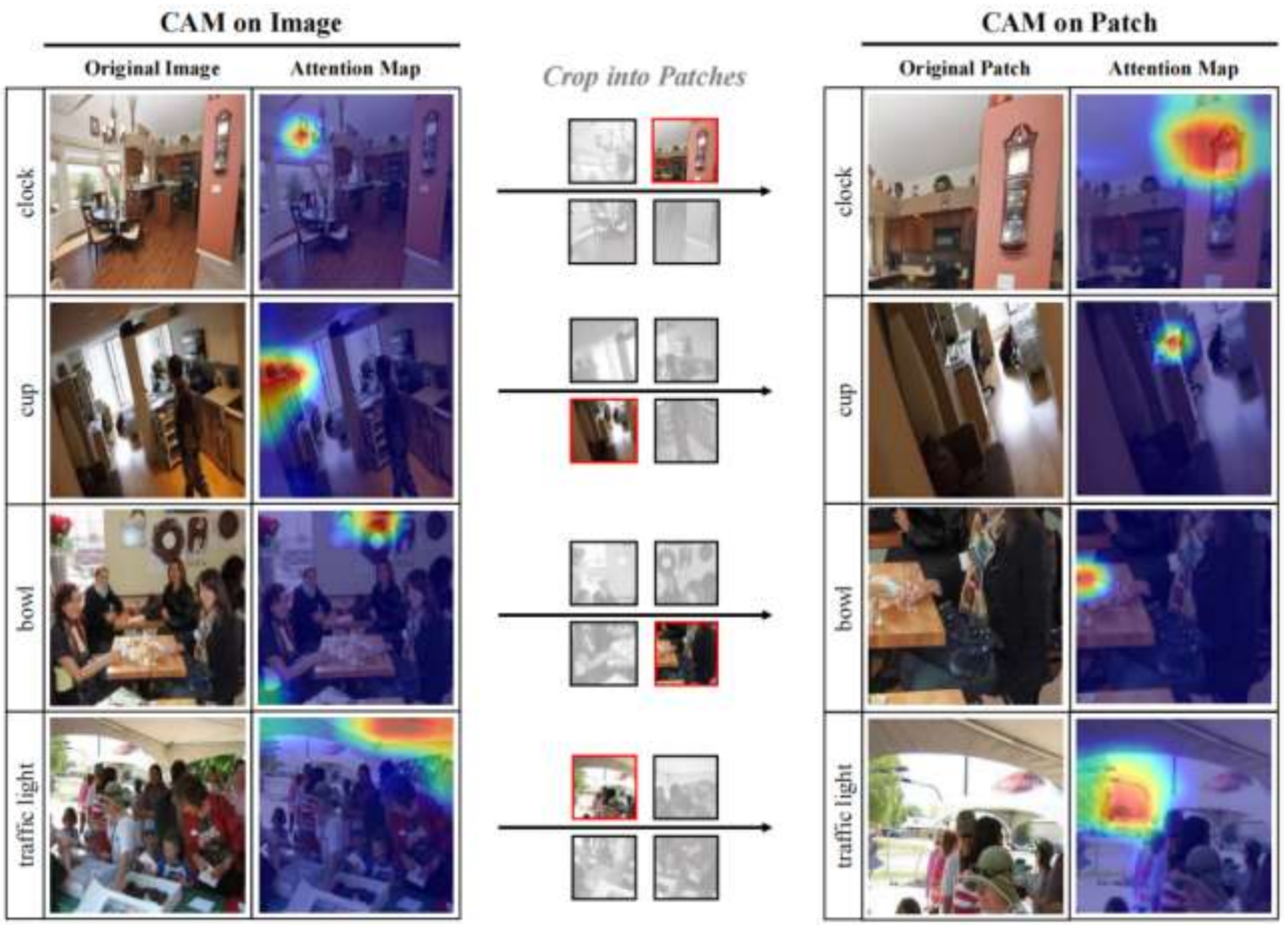


Fig. 4: Visualization of attention maps on COCO. Each patch is cropped from the original image starting from the beginning of a row. The class label attached in front of every original image or cropped patch is activated in the attention map.

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

東京大学
THE UNIVERSITY OF TOKYO

THANKS