模式分析与机器智能
工业和信息化部重点实验室
MIIT Key Laboratory of
Pattern Analysis & Machine Intelligence

ParNeC 模式识别与神经计算研究组
PAttern Recognition and NEural Computing

Boosting Multi-Label Image Classification with Complementary Parallel Self-Distillation

Jiazhi Xu[1], **Sheng Huang**[1*], **Fengtao Zhou**[1], **Luwen Huangfu**[2], **Daniel Zeng**[3] and **Bo Liu**[4]

[1]School of Big Data and Software Engineering, Chongqing University

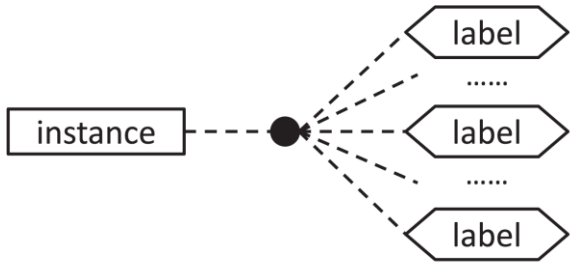
[2]Fowler College of Business & Center for Human Dynamics in the Mobile Age, San Diego State University, [3]Institute of Automation, Chinese Academy of Sciences, [4]JD.com

IJCAI 2022

Background

multi-class learning VS **multi-label learning**

(a) traditional supervised learning

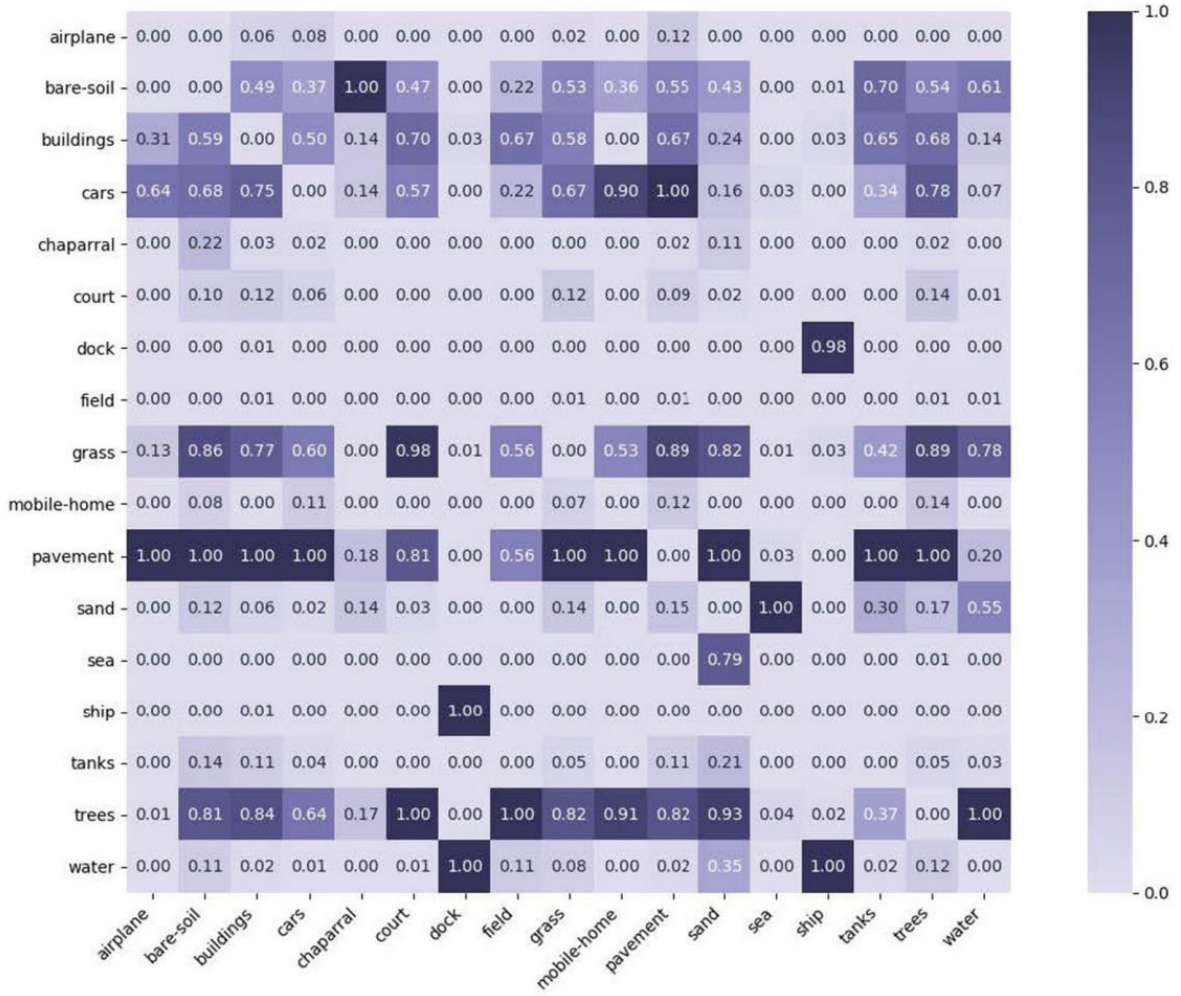


(b) single-instance multi-label learning

Background

Multi-Label Image Classification（MLIC）approaches usually exploit ***label correlations*** to achieve good performance.

label co-occurrences



Motivation

Emphasizing correlation may lead to model overfitting !!!

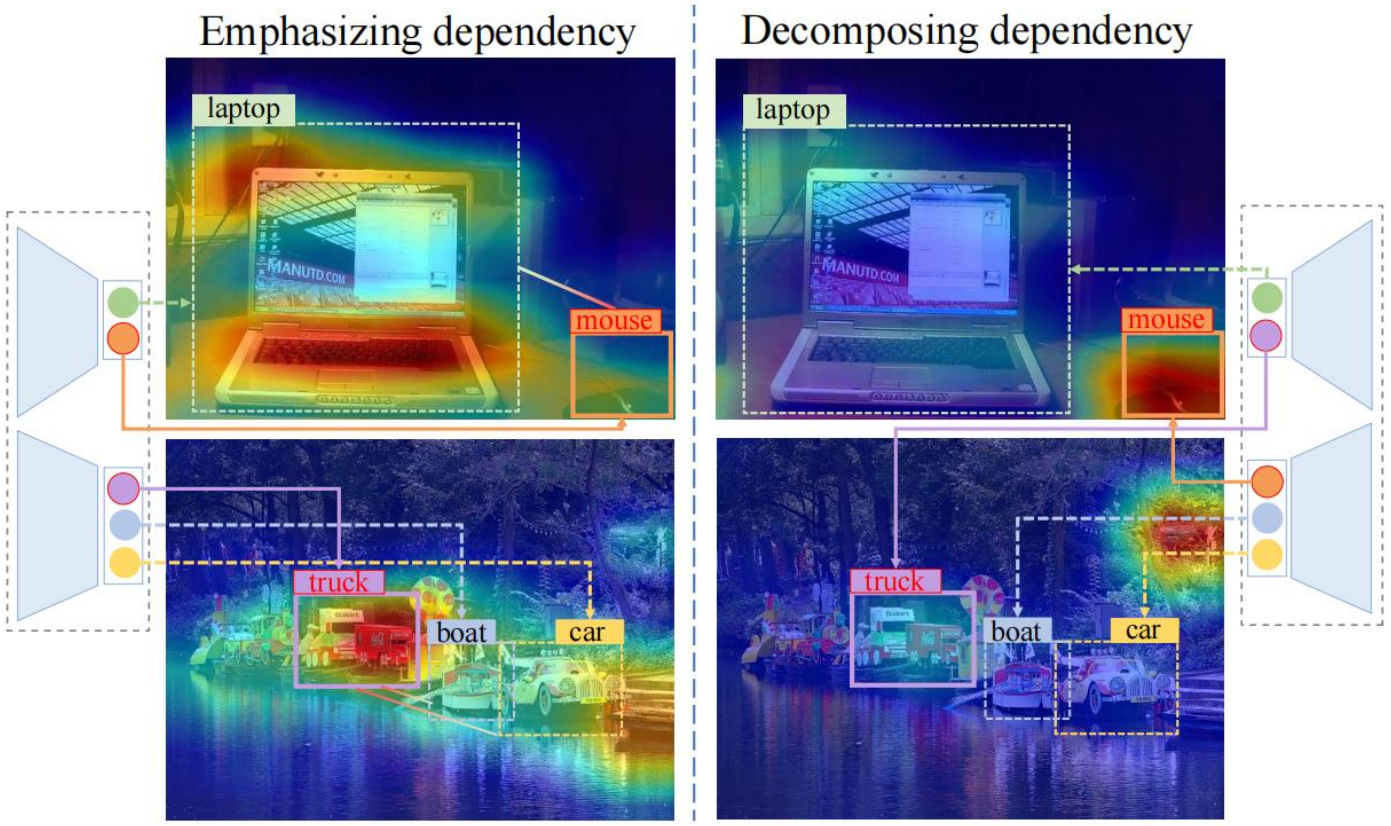


Figure 1: **Class Activation Map (CAM) of class *mouse* and *truck*.**

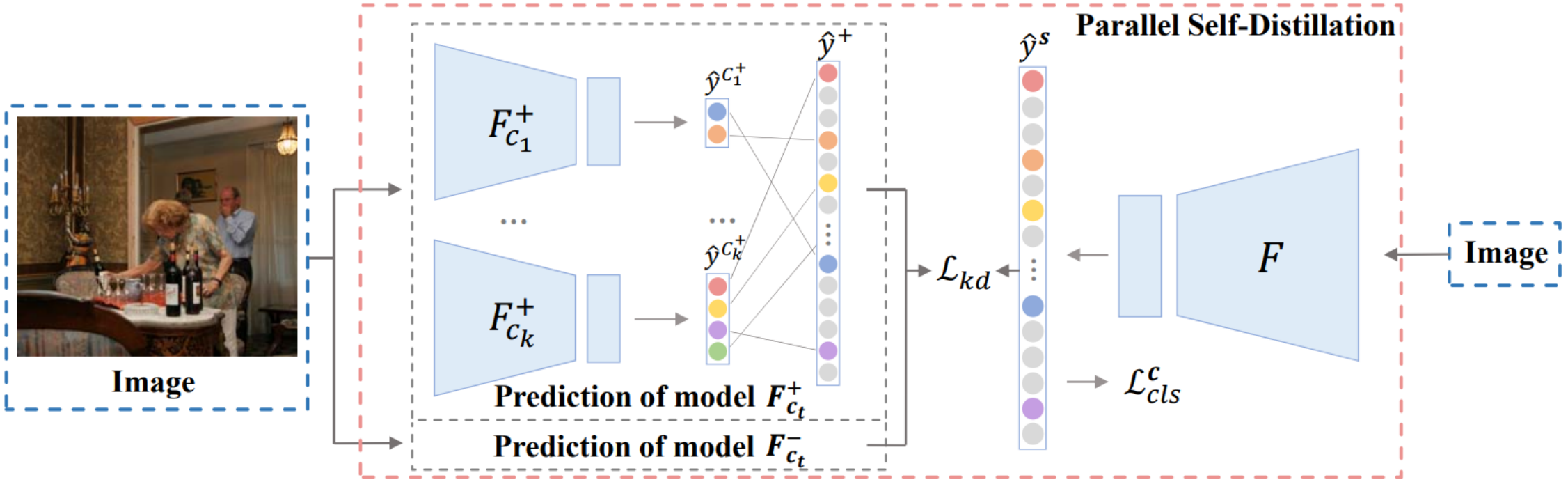
How to balance the learning of

label correlations

and

discriminative features ?

Parallel Self-Distillation (PSD) — leverages ***the proper task decomposition strategy*** to formulate a multiple sub-model parallelly training task, then distills these sub-models into a global model.



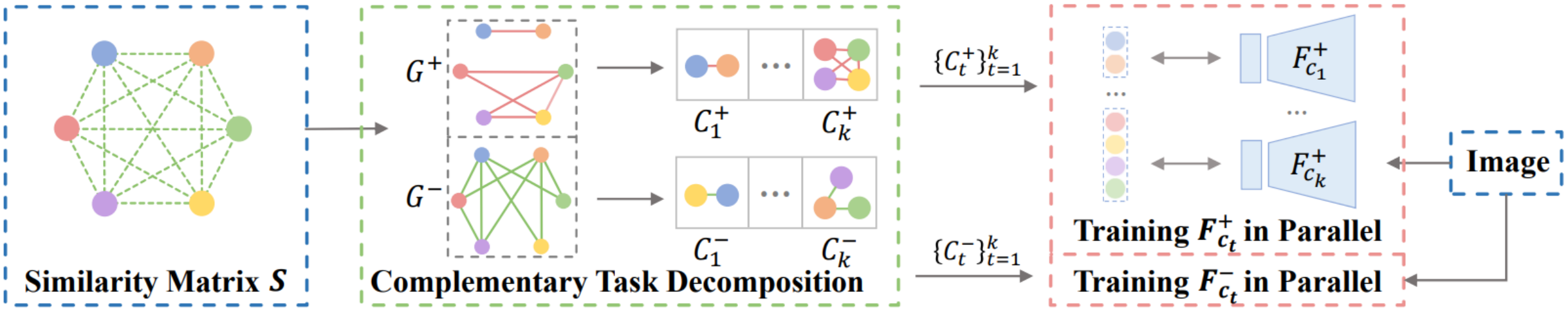
The superscripts + and – respectively represent co-occurrence and dis-occurrence branches.

Co-occurrence Graph Partition (CGP)

— assigns labels with co-occurrence into one task for learning the **label correlations.**

Dis-occurrence Graph Partition (DGP)

— assigns labels without co-occurrence into one task for learning the **discriminative features.**



Complementary Task Decomposition

S $m \times m$-dimensional

$$S_{ij} = e_{ij}/n_i \in [0, 1]$$

e_{ij} the amount of images containing both c_i and c_j

n_i the amount of images containing c_i

affinity matrices :

$$P = \begin{cases} P^+ = \dfrac{(\sqrt[\tau]{S} + \sqrt[\tau]{S}^T)}{2}, & G = G^+ \quad \textit{co-occurrence graph } G^+ \\ P^- = I - \dfrac{(\sqrt[\tau]{S} + \sqrt[\tau]{S}^T)}{2}, & G = G^- \quad \textit{dis-occurrence graph } G^- \end{cases}$$

Complementary Task Decomposition

$$\hat{F} \leftarrow \arg\min_{F} \text{Trace}(F^T D^{-\frac{1}{2}} L D^{-\frac{1}{2}} F), \text{ s.t. } F^T F = I$$

$D_{ii} = \sum_j P_{ij}$

degree matrix

$L = D - P$

Laplacian matrix

graph embedding of vertices (categories)

Parallel Self-Distillation

A teacher model :

$$\hat{F}_{C_t} \leftarrow \arg\min_{F_{C_t}} L_{cls}^{C_t}$$

Asymmetric Loss : (ASL)

$$L_{cls}^{C_t} = \sum_{x_i \in X_{C_t}} \sum_{c_j \in C_t} \begin{cases} (1-\hat{y}_i^{C_t}(j))^{\gamma+} \log(\hat{y}_i^{C_t}(j)), & y_i(j)=1 \\ \hat{\boldsymbol{y}}_i^{C_t}(j)^{\gamma-} \log(1-\hat{\boldsymbol{y}}_i^{C_t}(j)), & y_i(j)=0 \end{cases}$$

$$\max(\hat{y}_i^{C_t}(j) - \mu, 0)$$

γ_+ , γ_- are respectively the positive and negative focusing hyperparameters defifined in ASL

Parallel Self-Distillation

$\rho(\cdot)$ is a label merging and reshuffling operation

$$\hat{y}_i = \rho(\{\hat{F}_{C_t}(x_i)|x_i \in X_{C_t}\}_{t=1}^{k}) \qquad \hat{y}_i^{\boldsymbol{s}} = F(x_i)$$

Mean Square Error :

$$L_{kd} = \frac{1}{2}\sum_i \{||\hat{y}_i^{\boldsymbol{s}} - \hat{y}_i^{+}||_2^2 + ||\hat{y}_i^{\boldsymbol{s}} - \hat{y}_i^{-}||_2^2\}$$

Final optimization goal :

$$\hat{F} \leftarrow \arg\min_F L_{cls}^{\mathcal{C}} + L_{kd}$$

Experiment

MS-COCO

Methods	Backbone	Resolution	mAP	CP	CR	CF1	OP	OR	OF1
ResNet-101 [He *et al.*, 2016]	ResNet101	224×224	78.3	80.2	66.7	72.8	83.9	70.8	76.8
DSDL [Zhou *et al.*, 2021b]	ResNet101	448×448	81.7	84.1	70.4	76.7	85.1	73.9	79.1
CPCL [Zhou *et al.*, 2021a]	ResNet101	448×448	82.8	85.6	71.1	77.6	86.1	74.6	79.9
ML-GCN [Chen *et al.*, 2019c]	ResNet101	448×448	83.0	85.1	72.0	78.0	85.8	75.4	80.3
KSSNet [Liu *et al.*, 2018]	ResNet101	448×448	83.7	84.6	73.2	77.2	87.8	76.2	81.5
MS-CMA [You *et al.*, 2020]	ResNet101	448×448	83.8	82.9	74.4	78.4	84.4	77.9	81.0
MCAR [Gao and Zhou, 2021]	ResNet101	448×448	83.8	85.0	72.1	78.0	88.0	73.9	80.3
Q2L-R101 [Liu *et al.*, 2021]	ResNet101	448×448	84.9	84.8	74.5	79.3	86.6	76.9	81.5
ResNet101*(baseline)	ResNet101	448×448	81.6	80.6	72.7	76.4	83.7	76.7	80.0
Ours + ResNet101	ResNet101	448×448	83.1	83.5	73.6	78.2	84.8	77.3	80.9
ResNet101 + TF*	ResNet101	448×448	84.3	87.4	71.6	78.7	87.9	75.2	81.0
Ours + ResNet101 + TF	ResNet101	448×448	**85.2**	84.9	75.5	**79.9**	85.6	78.5	**81.9**
Q2L-R101*	ResNet101	448×448	84.0	82.0	**75.8**	78.8	83.3	**78.8**	81.0
Ours + Q2L-R101	ResNet101	448×448	84.9	**88.4**	71.7	79.2	**89.3**	74.8	81.4
SSGRL [Chen *et al.*, 2019a]	ResNet101	576×576	83.8	**89.9**	68.5	76.8	**91.3**	70.8	79.7
C-Trans [Lanchantin *et al.*, 2021]	ResNet101	576×576	85.1	86.3	74.3	79.9	87.7	76.5	81.7
ADD-GCN [Ye *et al.*, 2020]	ResNet101	576×576	85.2	84.7	75.9	80.1	84.9	79.4	82.0
Q2L-R101 [Liu *et al.*, 2021]	ResNet101	576×576	86.5	85.8	76.7	81.0	87.0	78.9	82.8
ResNet101 + TF*	ResNet101	576×576	85.9	88.6	73.4	80.3	88.8	76.8	82.4
Ours + ResNet101 + TF	ResNet101	576×576	**86.7**	83.5	**79.0**	**81.2**	84.5	**81.4**	**82.9**
TResL [Ridnik *et al.*, 2021]	TResNetL	448×448	86.6	87.2	76.4	81.4	88.2	79.2	81.8
Q2L-TResL [Liu *et al.*, 2021]	TResNetL	448×448	**87.3**	**87.6**	76.5	81.6	**88.4**	79.2	81.8
TResL*(baseline)	TResNetL	448×448	86.2	85.0	77.5	81.1	85.6	80.4	82.9
Ours + TResL	TResNetL	448×448	**87.3**	85.5	**78.9**	**82.1**	85.7	**81.5**	**83.7**
ML-GCN [Nguyen *et al.*, 2021]	ResNeXt50-SWSL	448×448	86.2	85.8	77.3	81.3	86.2	79.7	82.8
MGTN [Nguyen *et al.*, 2021]	ResNeXt50-SWSL	448×448	87.0	86.1	77.9	81.8	87.7	79.4	83.4
ResNeXt50*(baseline)	ResNeXt50-SWSL	448×448	86.7	85.8	77.8	81.6	86.9	80.3	83.5
Ours + ResNeXt50	ResNeXt50-SWSL	448×448	**87.7**	**86.9**	**78.6**	**82.5**	**87.6**	**80.9**	**84.1**

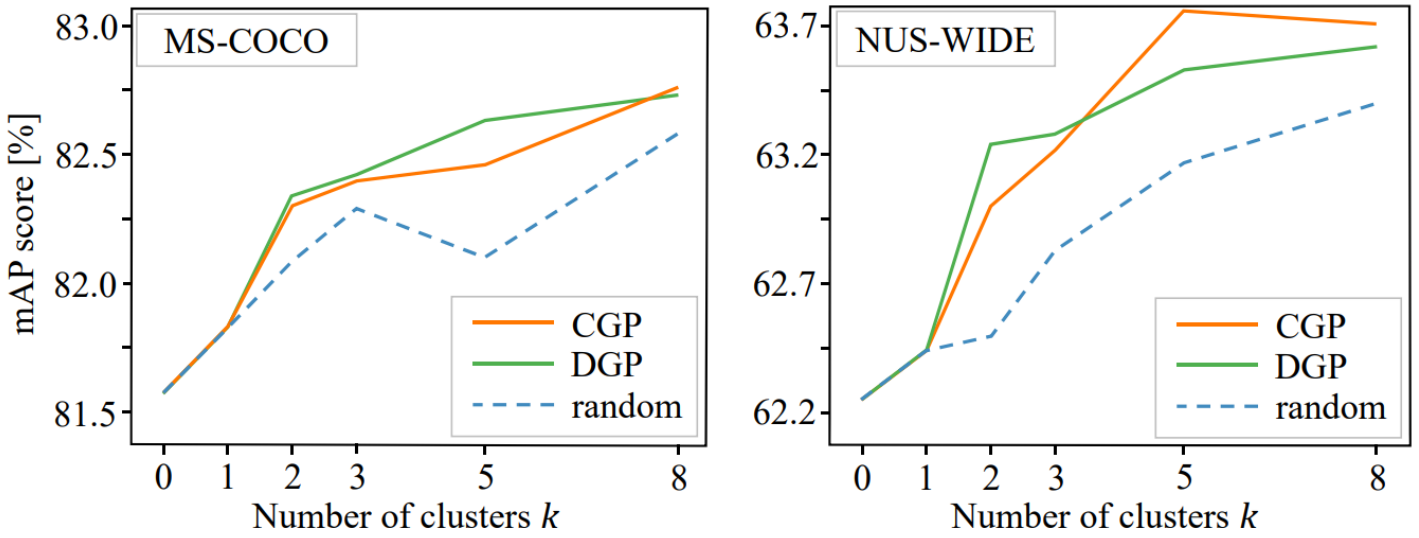
Experiment

NUS-WIDE

Methods	mAP	CF1	OF1
MS-CMA [You *et al.*, 2020]	61.4	60.5	73.8
SRN [Zhu *et al.*, 2017]	62.0	58.5	73.4
CPCL [Zhou *et al.*, 2021a]	62.3	59.2	73.0
CADM [Chen *et al.*, 2019b]	62.8	60.7	74.1
Q2L-R101 [Liu *et al.*, 2021]	65.0	63.1	75.0
ResNet101+TF*	64.1	62.8	74.9
Ours+ResNet101+TF	**65.8**	**64.0**	**75.3**
TResL [Ridnik *et al.*, 2021]	65.2	63.6	75.0
Q2L-TResL [Liu *et al.*, 2021]	66.3	64.0	75.0
TResL*(baseline)	64.7	63.7	75.0
Ours+TResL	**66.5**	**64.6**	**75.5**

Ablation

Different k with different strategies on both MS-COCO and NUS-WIDE:

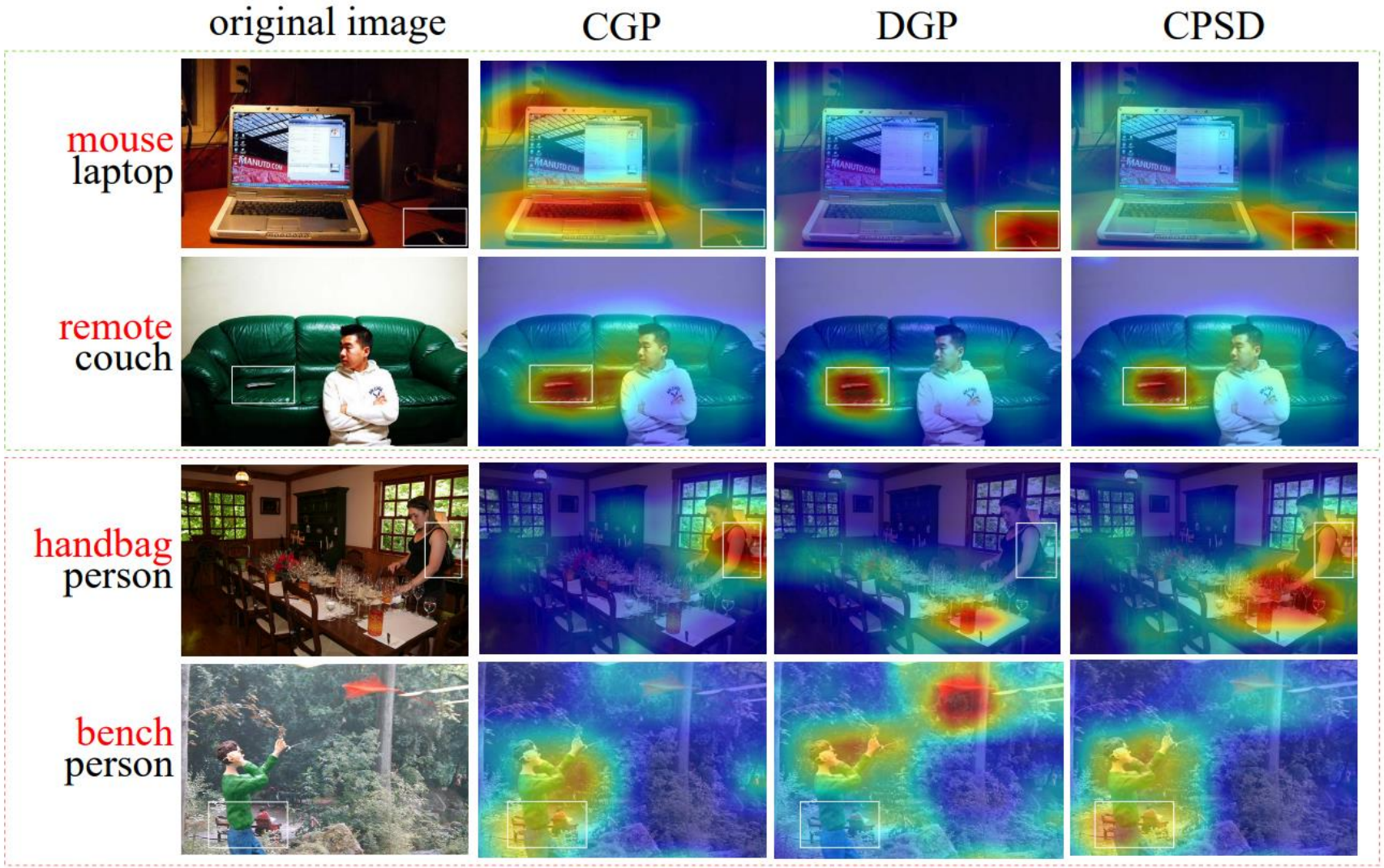


Component ablation study of CPSD:

	R101	TResL	ResX50	Q2L-R101	R101+TF
baseline	81.6	86.2	86.7	84.0	84.3
+ SD	81.9	86.5	86.9	84.2	84.4
+ CGPD	82.4	86.6	87.3	84.6	84.7
+ DGPD	82.7	86.8	87.4	84.3	84.5
+ CPSD	83.1	87.3	87.7	84.9	85.2

Experiment

Activation map visualizations of images under different decomposition strategies :



Thanks