Long-tailed Object Detection Pretraining: Dynamic Rebalancing Contrastive Learning with Dual Reconstruction

Chen-Long Duan[1†], Yong Li[1†], Xiu-Shen Wei[2], Lin Zhao[1]*

November 05, 2024

[1]Nanjing University of Science and Technology,

[2]School of Computer Science and Engineering, and Key Laboratory of New Generation Artificial Intelligence Technology and Its Interdisciplinary Applications, Southeast University

Introduction

南京理工大学
NANJING UNIVERSITY OF SCIENCE & TECHNOLOGY

Deep learning has driven important progress!

Image classification

Autonomous driving

Games

Continue ……

南京理工大学
NANJING UNIVERSITY OF SCIENCE & TECHNOLOGY

Limitations of deep learning models

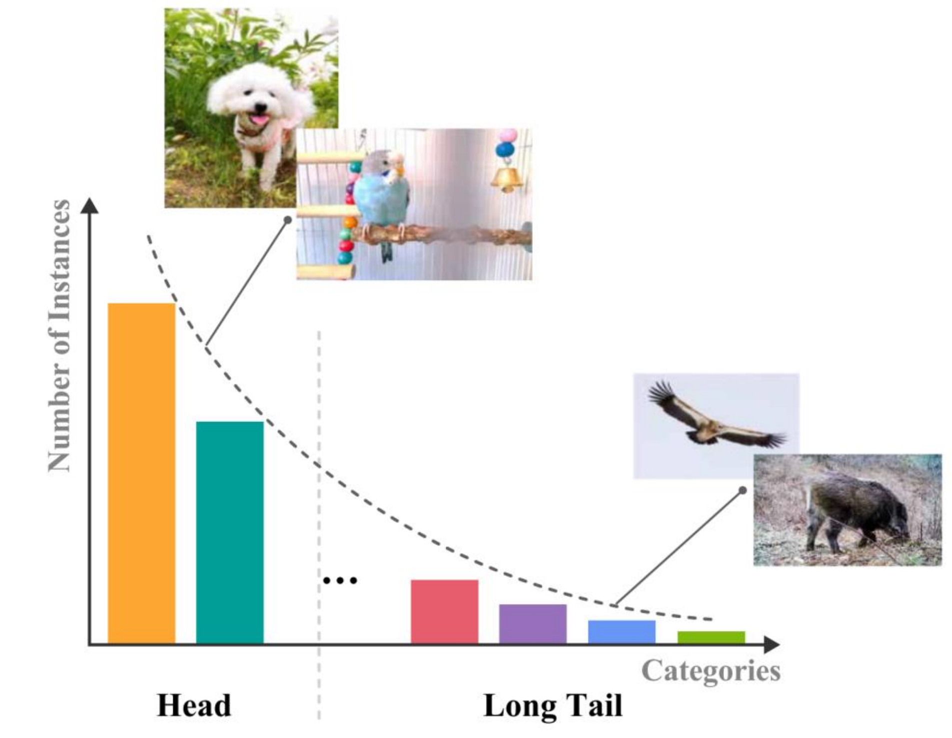
- Deep learning models often rely on large training datasets.
- Real-world data often exhibits a long-tailed distribution.
- Humans can learn from one or few examples, thanks to their rich prior knowledge.

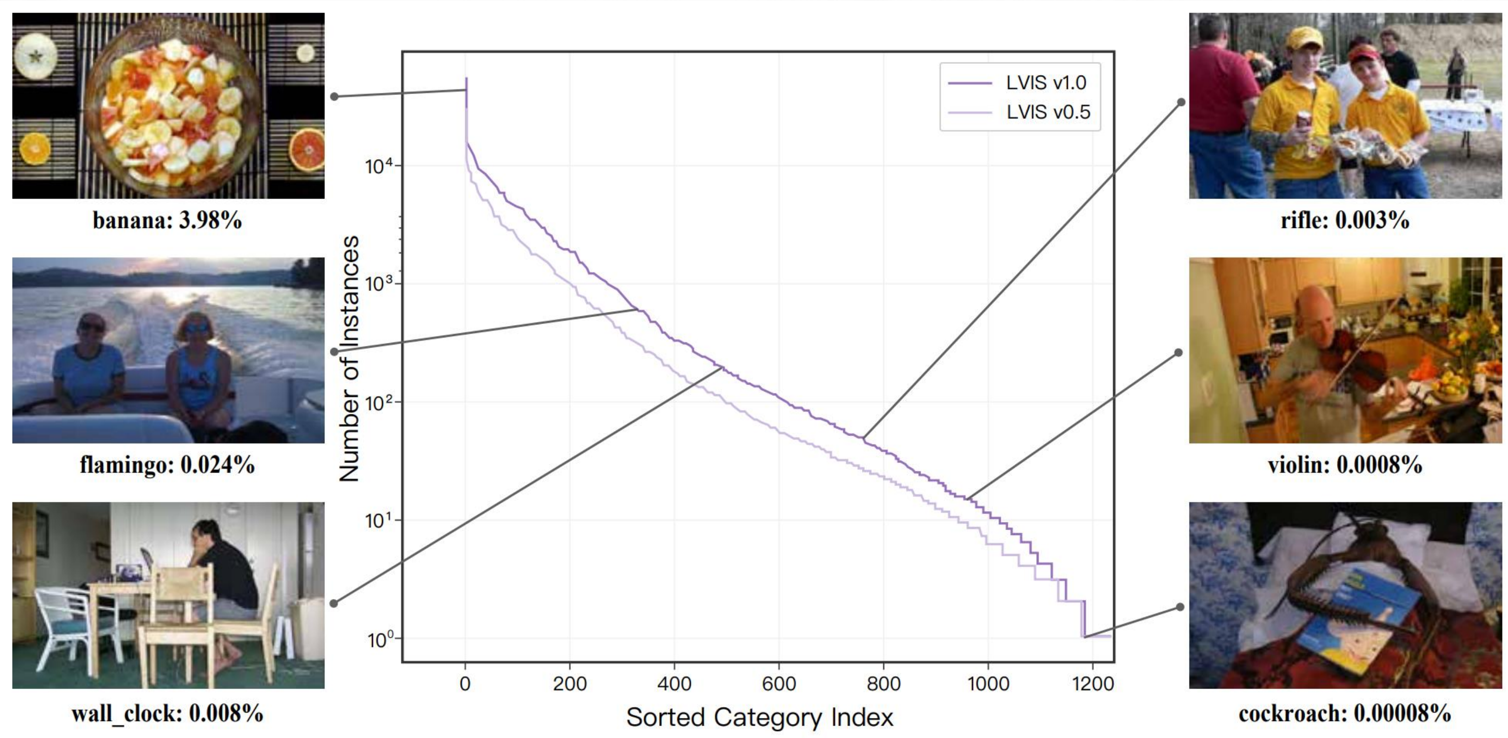
- **Pre-training for Object Detection**
- Long-tailed Object Detection
- Simplicity Bias

Method	Backbone	Neck	Head
Supervised Backbone	√	×	×
MoCo, SwAV, BYOL	√	×	×
DenseCL, DetCo, DetCon	√	×	×
PixPro, SoCo	√	√	×
UP-DETR, DETReg, AlignDet	√	√	√

Introduction (con't)

- Pre-training for Object Detection
- **Long-tailed Object Detection**
- Simplicity Bias



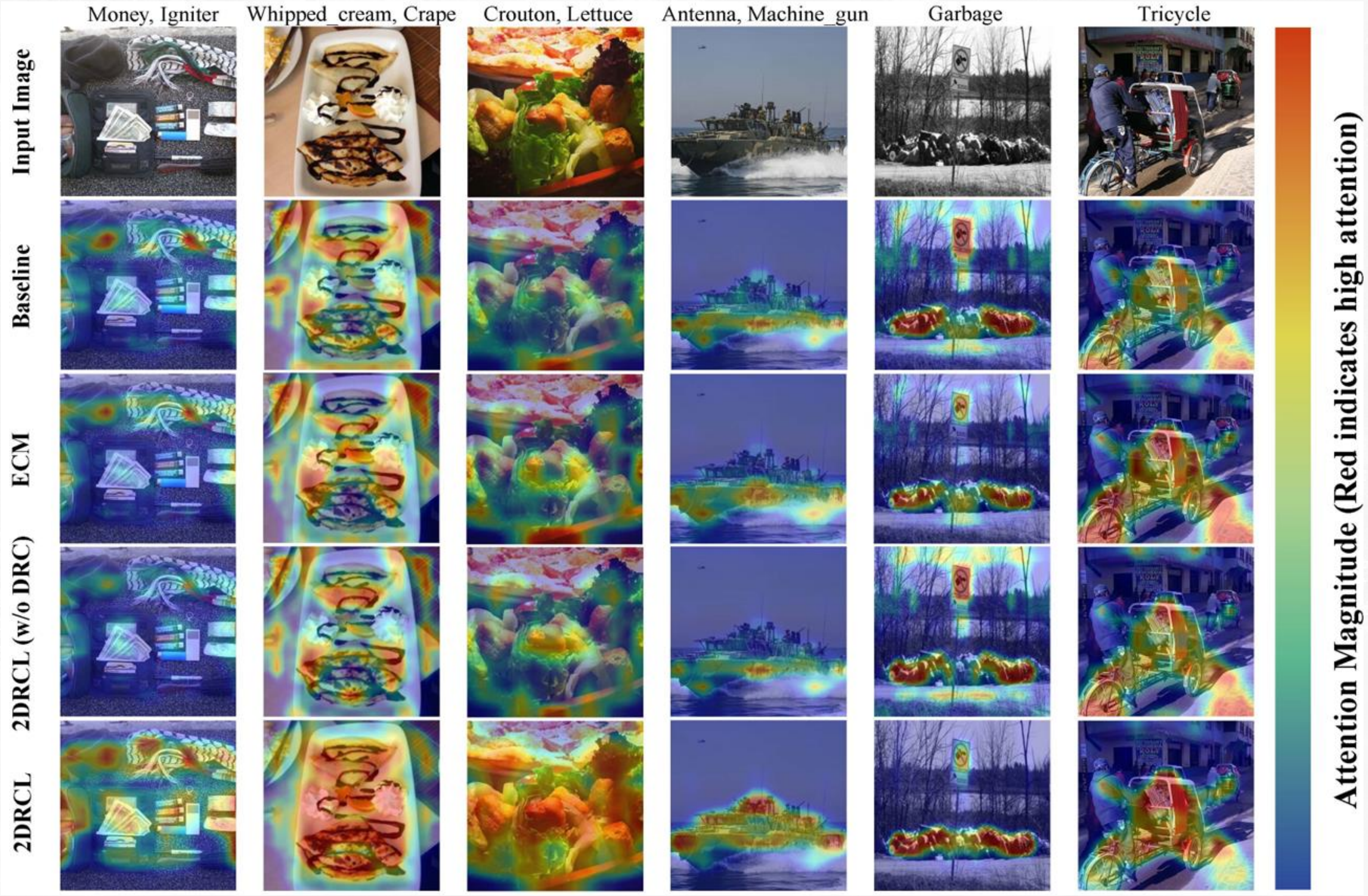


[L. Yang et al., International Journal of Computer Vision 2022]

Introduction (con't)

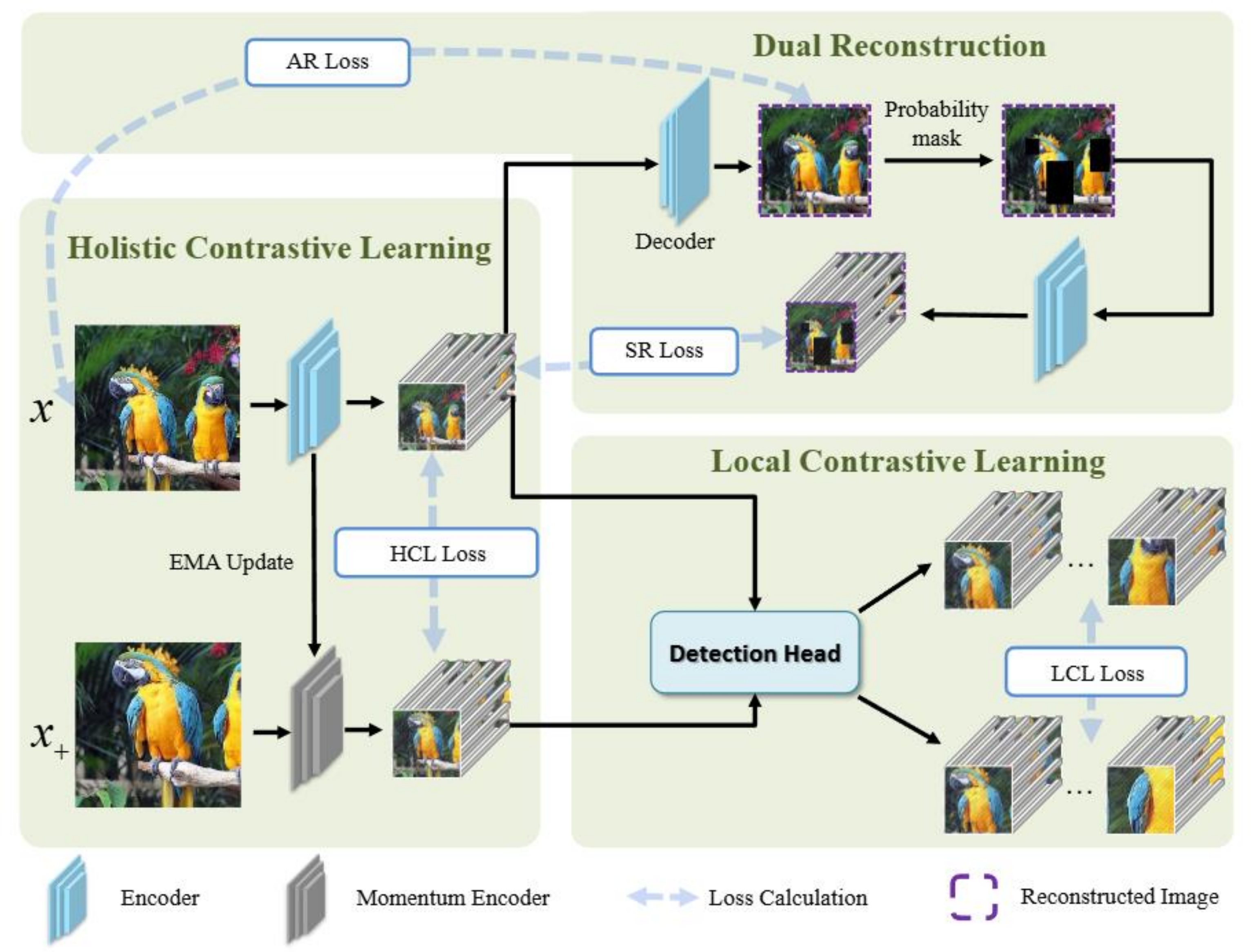
- Pre-training for Object Detection
- Long-tailed Object Detection
- **Simplicity Bias**

An **often-overlooked** but crucial challenge in long-tailed object detection is simplicity bias.



Our 2DRCL

Overall framework of 2DRCL: Our 2DRCL framework integrates three key components: **Holistic-Local Contrastive Learning**, **Dynamic Rebalancing**, and **Dual Reconstruction**.



Our 2DRCL (con't)

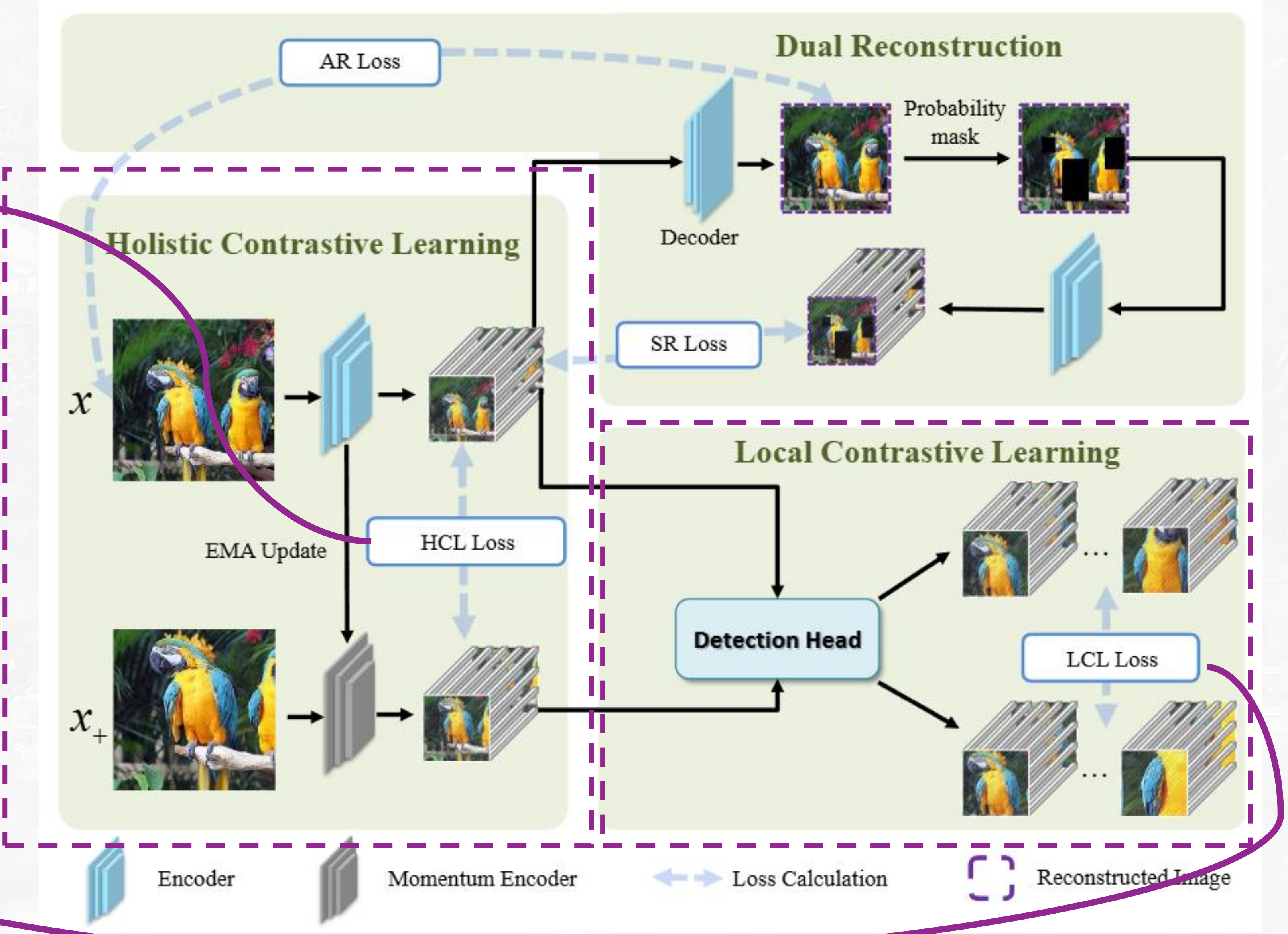
Holistic Contrastive Learning

$$\mathcal{L}_{HCL} = -\log \frac{\exp(z \cdot z_{+}/\tau)}{\exp(z \cdot z_{+}/\tau) + \sum_{i=1}^{K} \exp(z \cdot z_{i}/\tau)}$$

Local Contrastive Learning

$$\mathcal{L}_{LCL} = -\log \frac{\exp\left(z_{bb} \cdot z_{bb_{+}}/\tau\right)}{\exp\left(z_{bb} \cdot z_{bb_{+}}/\tau\right) + \sum_{i=1}^{K} \exp\left(z_{bb} \cdot z_{bb_{i}}/\tau\right)}$$

Holistic-Local Contrastive Learning



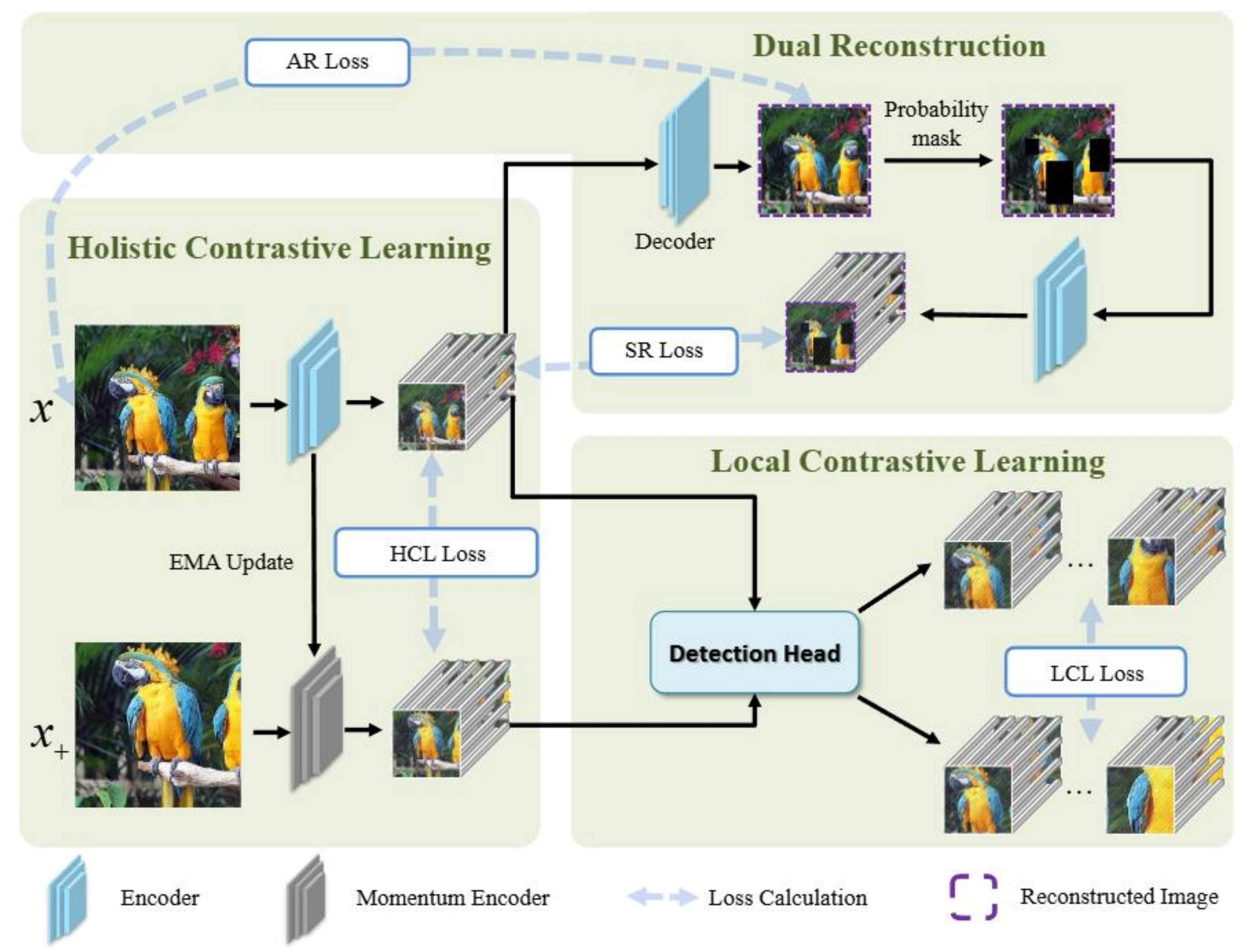
Our 2DRCL (con't)

Dynamic Rebalancing

$$f_c = \frac{f_c^{im} \cdot f_c^{in}}{\alpha_d f_c^{im} + (1-\alpha_d) f_c^{in}}$$

$$\alpha_d = \frac{T}{T_{max}}$$

$$r_c = \max\left(1, \sqrt{t/f_c}\right)$$

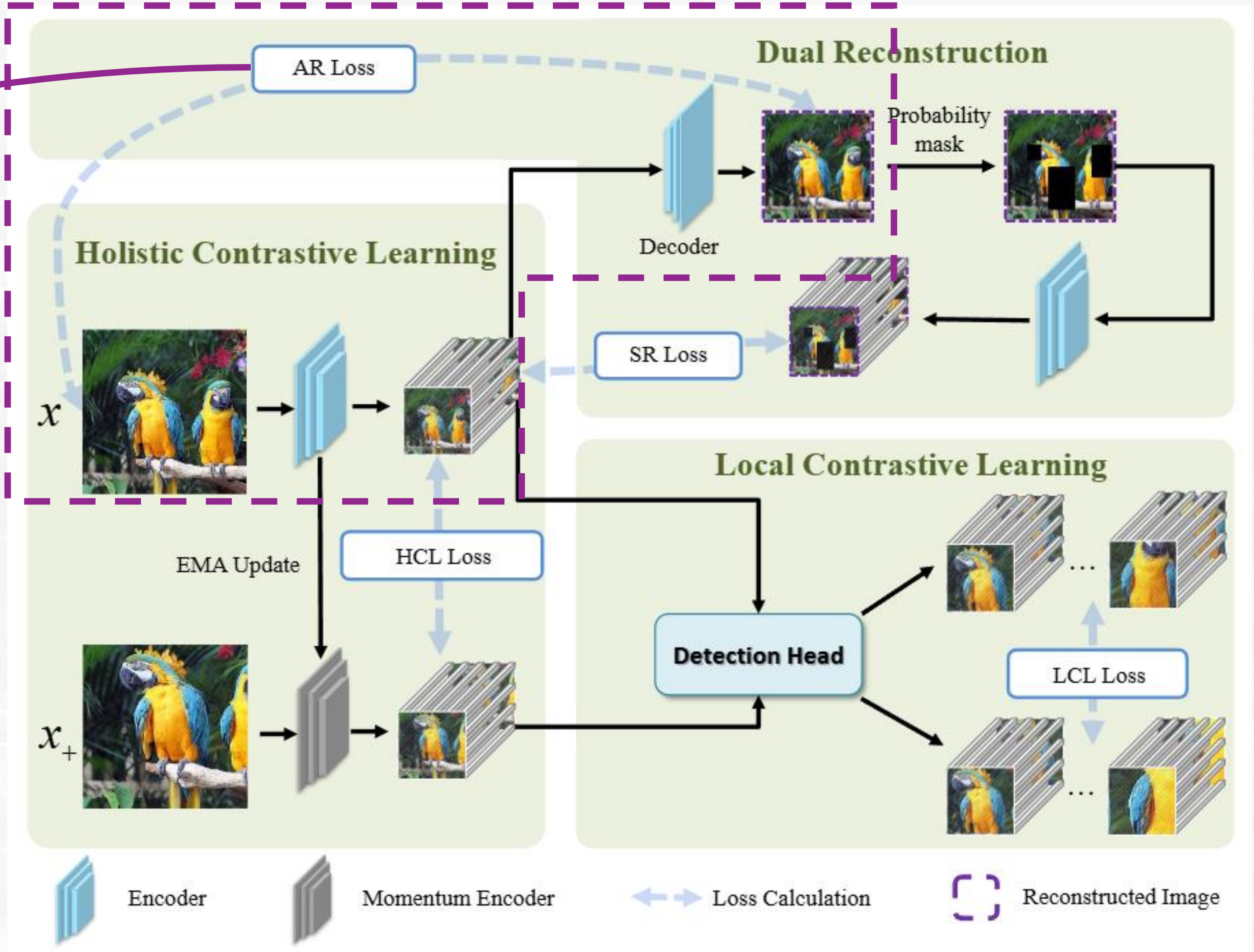


Our 2DRCL (con't)

Dual Reconstruction

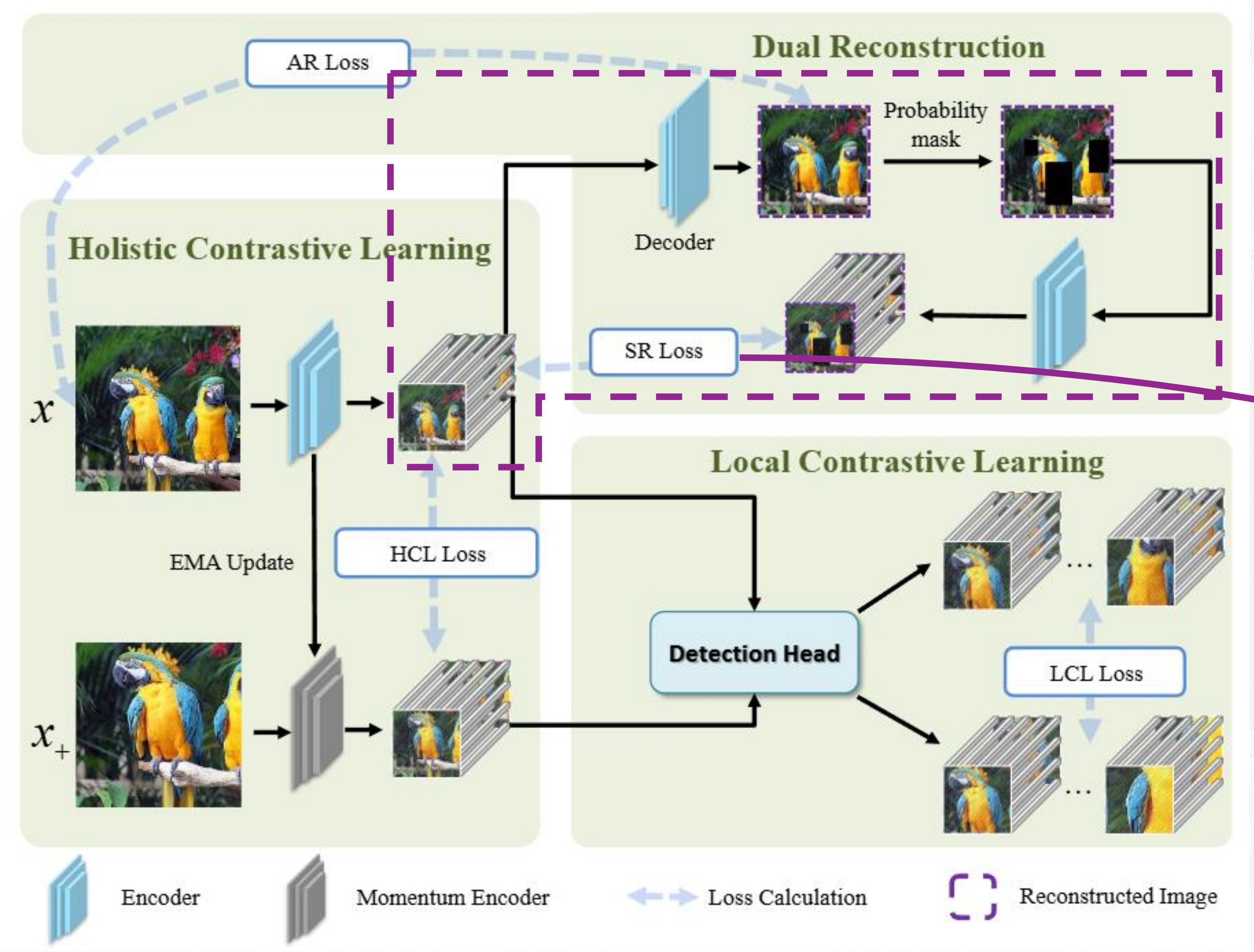
Appearance Reconstruction

$$\mathcal{L}_{AR} = \| x - g(f(x)) \|_2^2$$



Our 2DRCL (con't)

Dual Reconstruction



Appearance Reconstruction

$$\mathcal{L}_{AR} = \| x - g(f(x)) \|_2^2$$

Semantic Reconstruction

$$\mathcal{L}_{SR} = \sum_{p=1}^{P} \| f(x) - f(\mathcal{M}(g(f(x)))) \|_2^2$$

Main results comparisons (COCO)

Table 1: Comparisons with state-of-the-art methods on COCO (Mask R-CNN with R50-FPN).

Backbone Initialization	Methods	AP^{bb}	AP^{bb}_{50}	AP^{bb}_{75}	AP^{mk}	AP^{mk}_{50}	AP^{mk}_{75}
From scratch	DenseCL [49]	39.6	59.3	43.3	-	-	-
	Self-EMD [34]	40.4	61.1	43.7	37.4	56.5	**39.7**
	SoCo [50]	40.6	61.1	44.4	-	-	-
	SlotCon [55]	41.0	61.1	45.0	-	-	-
ImageNet pre-trained backbone	Surpervised	38.3	58.0	42.1	34.3	54.9	36.6
	AlignDet [25]	39.4	59.2	43.2	35.3	56.1	37.7
	Ours	**41.4**	**61.3**	**45.8**	**37.4**	**57.2**	39.4

Main results comparisons (LVIS v1.0)

Table 3: Comparisons with state-of-the-art methods on LVIS v1.0 with a 2× schedule.

(a) Faster R-CNN with R50-FPN.

Method	AP^{bb}	AP^{bb}_r	AP^{bb}_c	AP^{bb}_f
BCE [40]	19.5	1.6	16.6	30.6
RFS [11]	24.2	14.2	22.3	30.6
DropLoss [19]	21.8	5.2	21.8	29.1
PCB [17]	23.0	6.2	21.5	32.2
EQLv2 [42]	25.4	15.8	23.5	31.7
Seesaw [45]	26.4	16.8	25.1	32.2
BAGS [27]	23.7	14.2	22.2	29.6
ACSL [48]	22.2	9.9	21.3	28.5
LOCE [9]	25.1	15.7	24.2	30.1
BACL [38]	26.1	16.0	25.7	30.9
ECM [22]	26.7	17.5	25.7	32.2
Ours	**27.3**	**18.6**	**25.8**	**32.6**

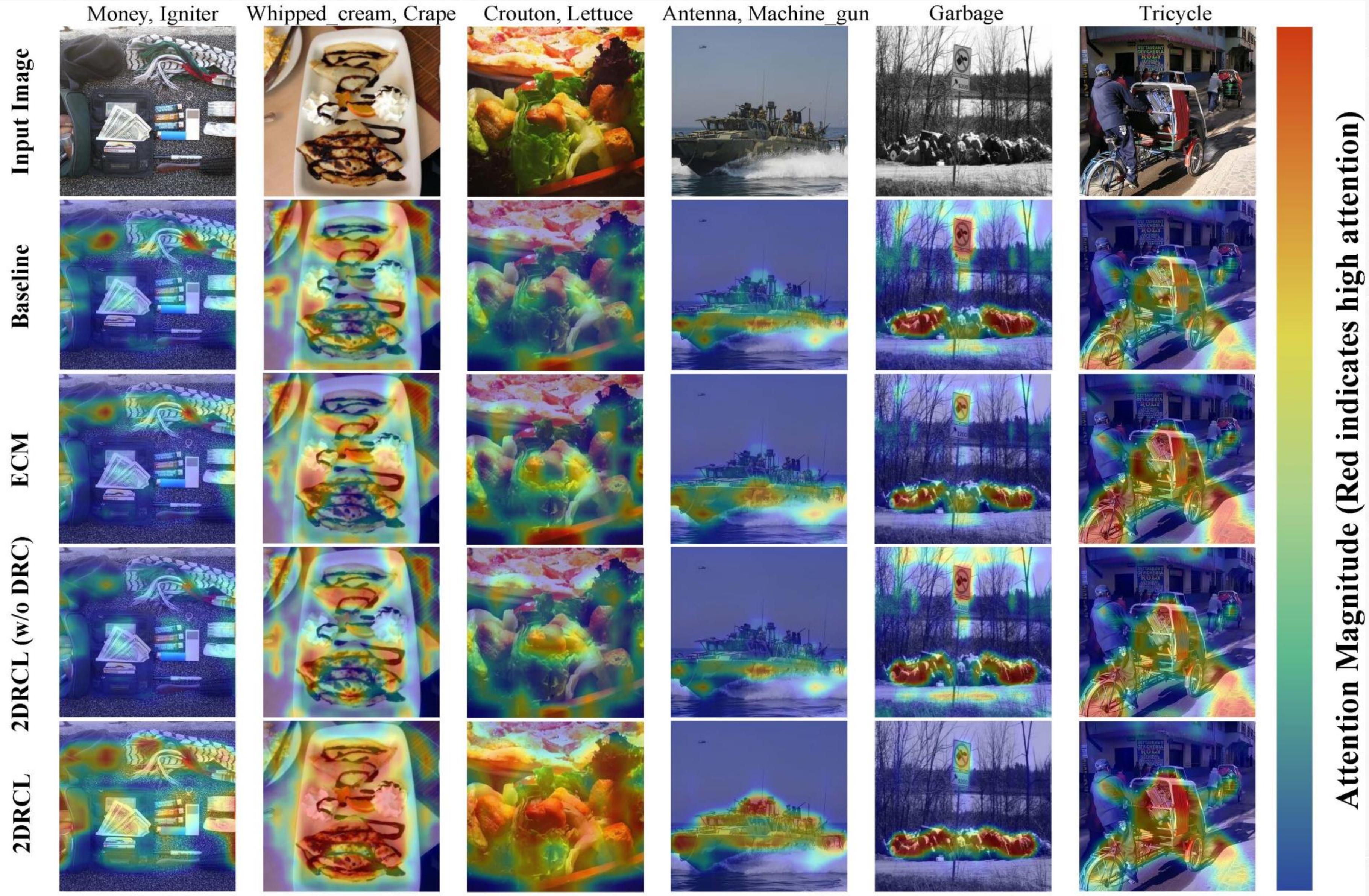
(b) Mask R-CNN with ResNet-50/101.

Backbone	Method	AP	AP_r	AP_c	AP_f	AP^{bb}
R50-FPN	CE	18.7	0.4	16.5	29.3	19.7
	RFS [11]	23.7	14.2	22.9	29.3	24.7
	EQLv2 [42]	25.2	17.4	24.1	29.9	26.0
	LOCE [9]	26.6	18.5	26.2	30.7	27.4
	SeeSaw [45]	26.9	19.6	26.8	30.5	27.3
	ECM [22]	27.4	19.7	27.0	31.1	27.9
	Ours	**27.7**	**20.4**	**27.1**	**31.4**	**28.3**
R101-FPN	CE	25.5	16.6	24.5	30.6	26.6
	EQLv2 [42]	27.2	20.6	25.9	31.4	27.9
	SeeSaw [45]	28.2	20.3	28.1	31.8	29.0
	ECM [22]	28.7	**21.9**	28.4	32.2	29.4
	Ours	**28.8**	21.1	**28.7**	**32.3**	**29.6**

Experiments (con't)

南京理工大学
NANJING UNIVERSITY OF SCIENCE & TECHNOLOGY

Simplicity Bias Analyses



南京理工大学
NANJING UNIVERSITY OF SCIENCE & TECHNOLOGY

Thanks all!

團结 獻身 求是 創新