RecursiveMix: Mixed Learning with History

Lingfeng Yang[1#], Xiang Li[2#], Borui Zhao[3], Renjie Song[3], and Jian Yang[1*]
[1]Nanjing University of Science and Technology , [2]Nankai University, [3]Megvii Technology

{yanglfnjust, csjyang}@njust.edu.cn, xiang.li.implus@nankai.edu.cn
zhaoborui.gm@gmail.com, songrenjie@megvii.com

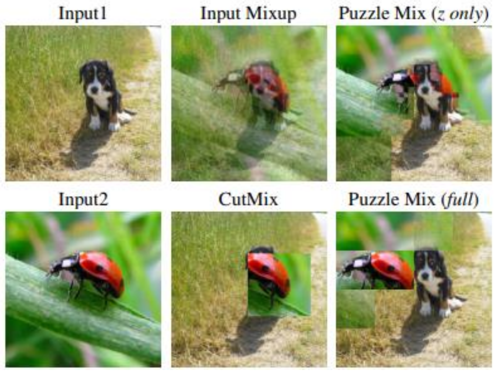
[*] Corresponding author. [#] Equal contributions. Research was done during Lingfeng's internship at Megvii

Background

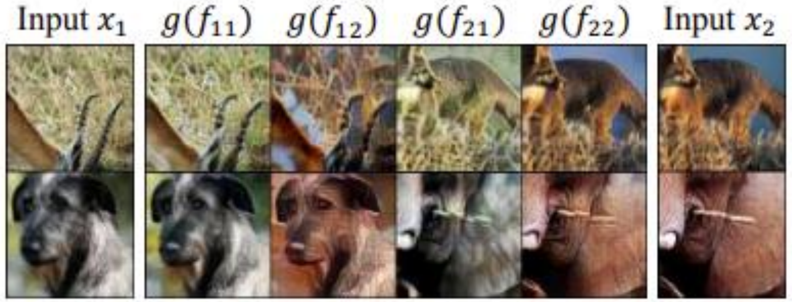
Mixup, ICLR 2018

CutMix, ICCV 2019

FMIX, Arxiv 2020



Puzzle Mix, ICML 2020



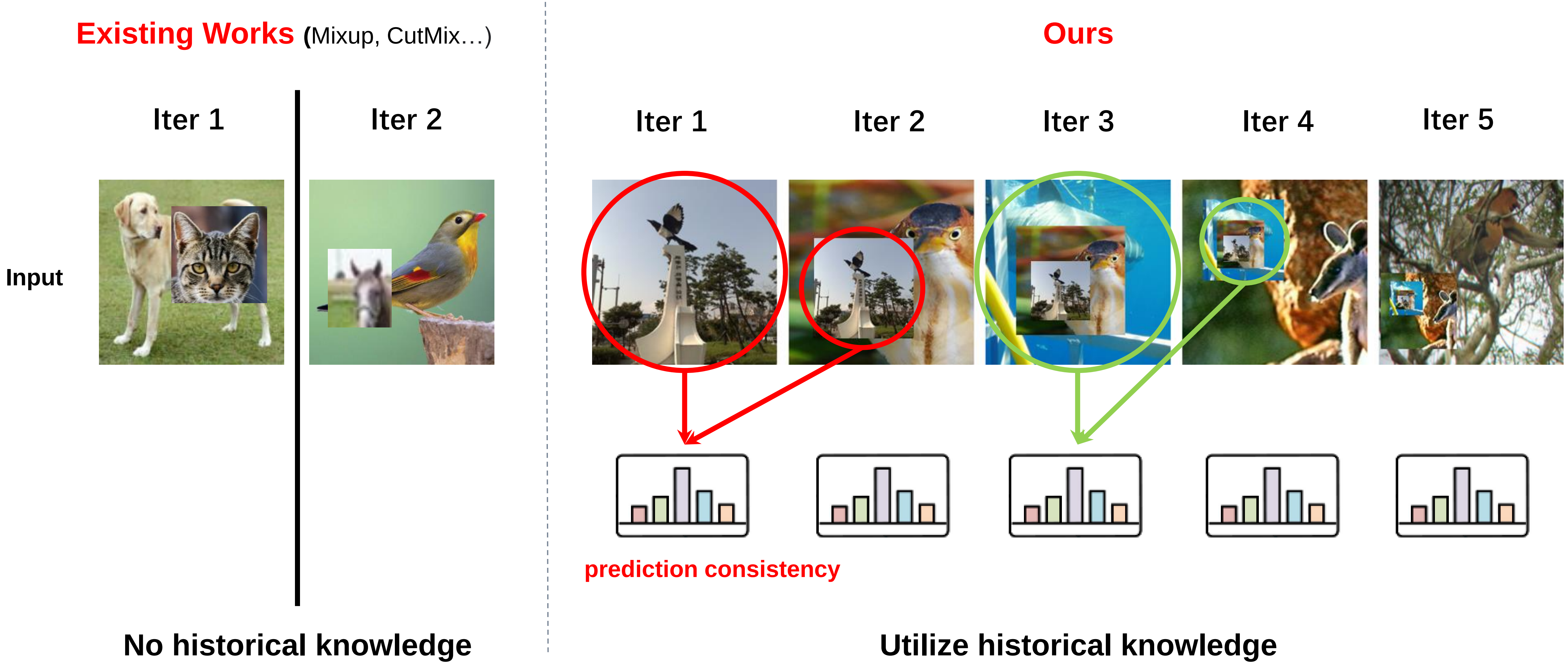
StyleMix, CVPR 2021

Mixed Sample Data Augmentation

$x_{mix}=mix_\lambda(x_1,x_2)$

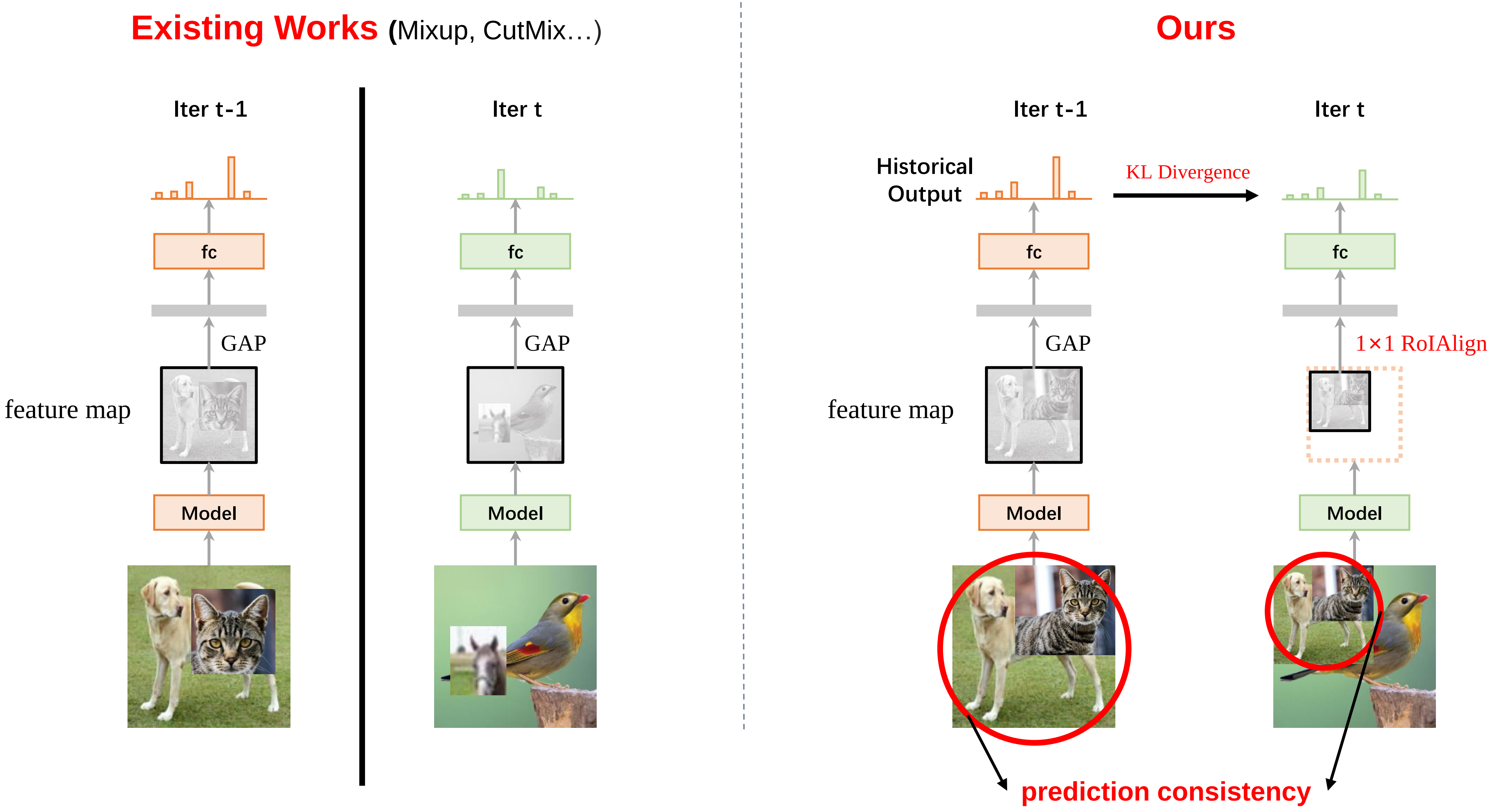
$y_{mix}=mix_\lambda(y_1,y_2)$

Background

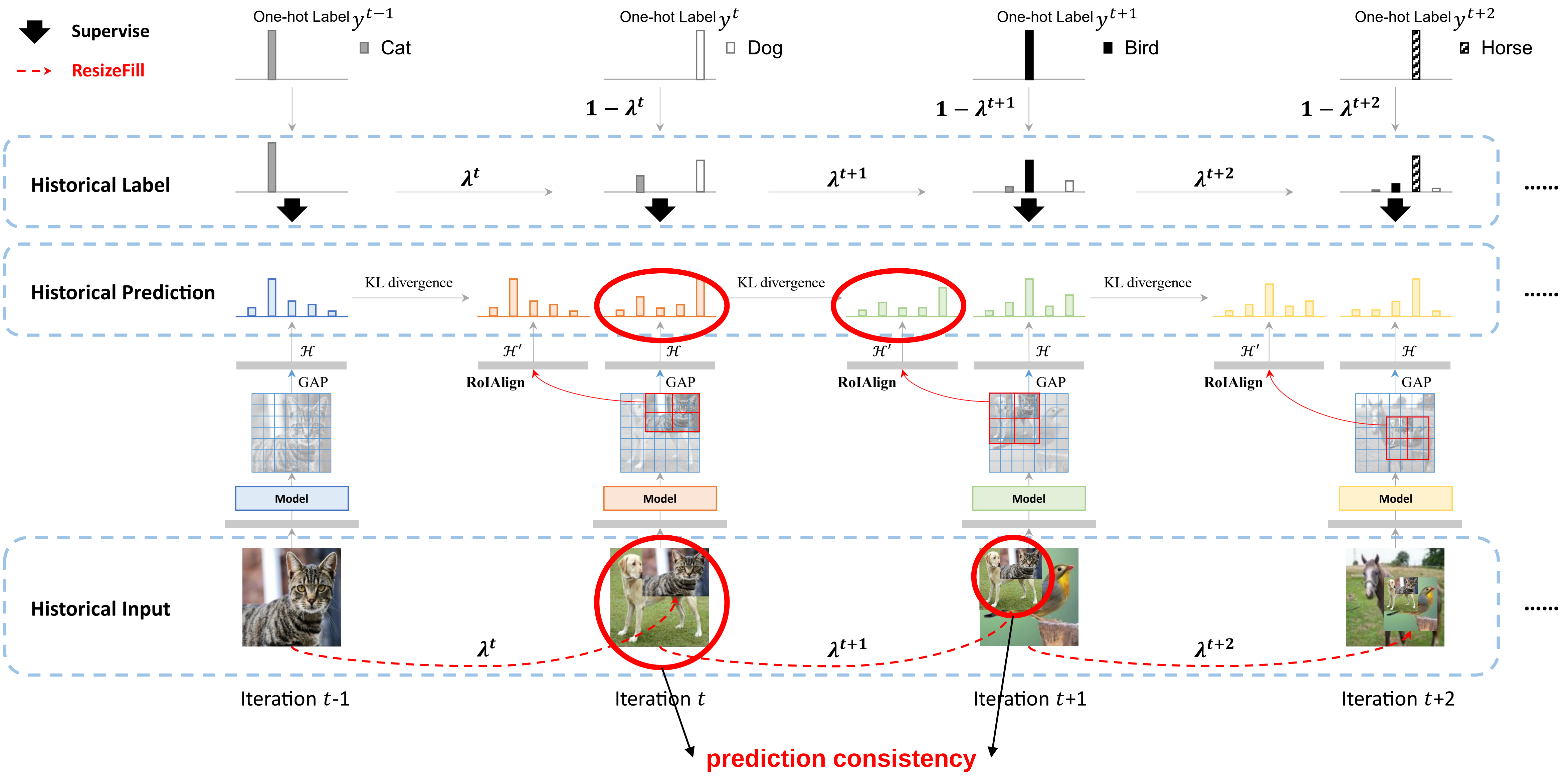


Existing Works (Mixup, CutMix…)
Ours
Iter 1
Iter 2
Input
Iter 1
Iter 2
Iter 3
Iter 4
Iter 5
prediction consistency
No historical knowledge
Utilize historical knowledge

Method


Existing Works (Mixup, CutMix…)
Iter t-1
Iter t
fc
GAP
feature map
Model
Ours
Iter t-1
Iter t
Historical Output
KL Divergence
fc
GAP
1×1 RoIAlign
feature map
Model
prediction consistency

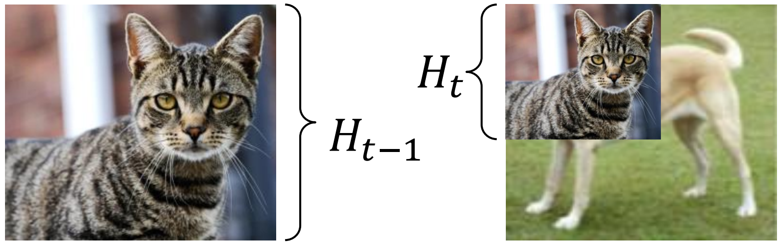
Method



Supervise
ResizeFill
One-hot Label y^{t-1}
Cat
One-hot Label y^{t}
Dog
One-hot Label y^{t+1}
Bird
One-hot Label y^{t+2}
Horse
$1-\lambda^t$
$1-\lambda^{t+1}$
$1-\lambda^{t+2}$
Historical Label
λ^t
λ^{t+1}
λ^{t+2}
......
Historical Prediction
KL divergence
KL divergence
KL divergence
......
$\mathcal{H}$
$\mathcal{H}'$
GAP
RoIAlign
Model
Historical Input
......
Iteration t-1
Iteration t
Iteration t+1
Iteration t+2
prediction consistency

Method

Resize and paste



$$\lambda = \text{Uniform}(0, \boldsymbol{\alpha})$$
$$H_t = \sqrt{\lambda} \cdot H_{t-1}$$

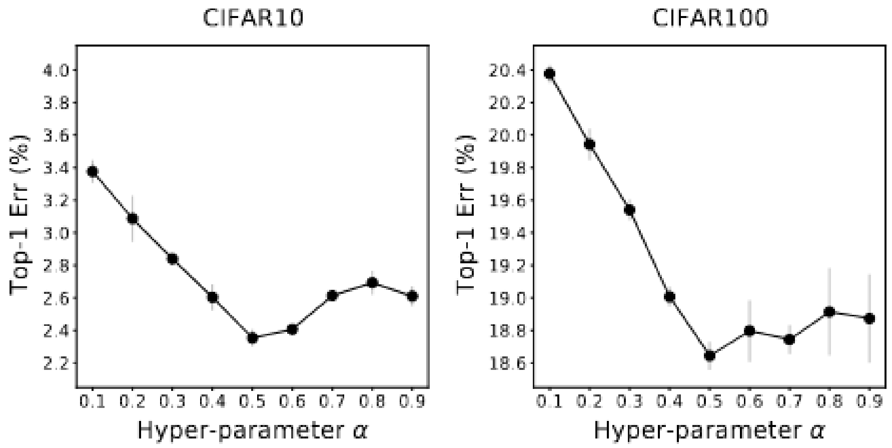


Figure: Ablation study on $\boldsymbol{\alpha}$.

Criterion

$$\mathcal{L} = \mathcal{L}_{CE}(\tilde{x}^t, \tilde{y}^t) + \boldsymbol{\omega} \lambda^t \mathcal{L}_{KL}(\tilde{p}^t_{roi}, p^h)$$

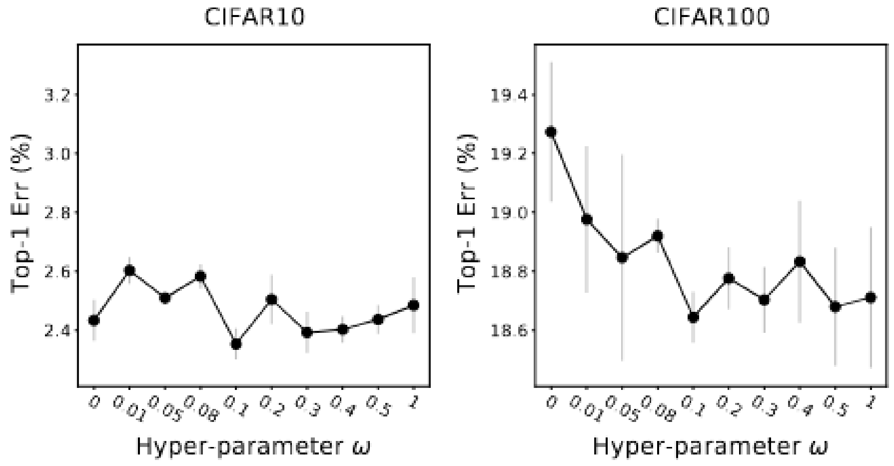


Figure: Ablation study on $\boldsymbol{\omega}$.

Analysis

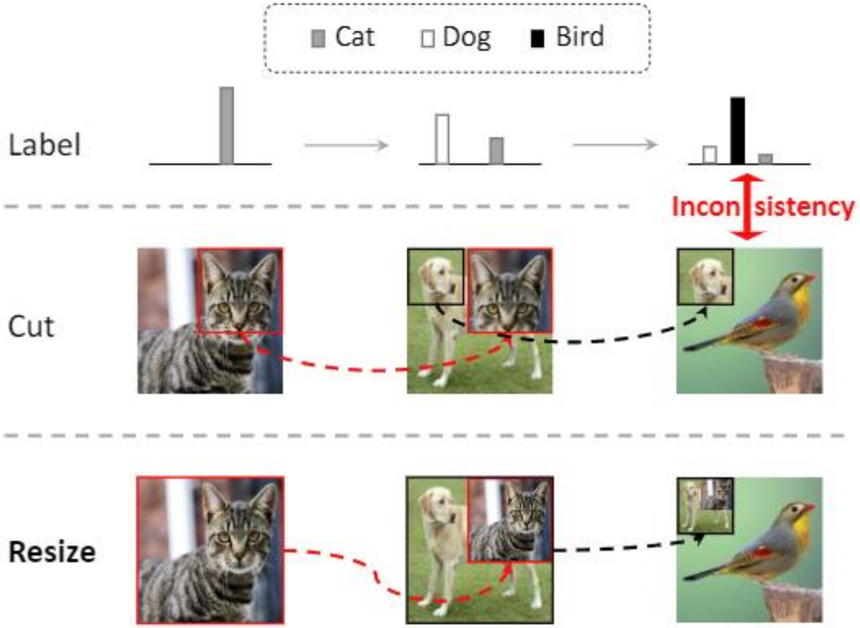


Figure: “Cut” may lead to inconsistency while
“Resize” concretely preserve the consistency.

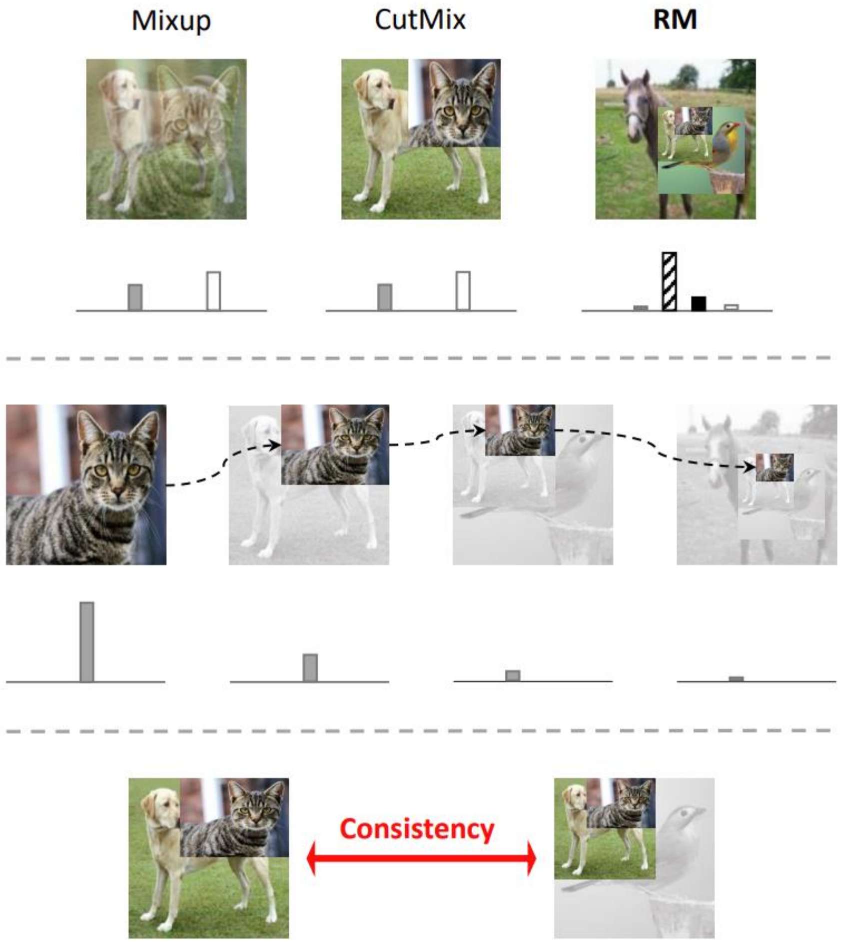
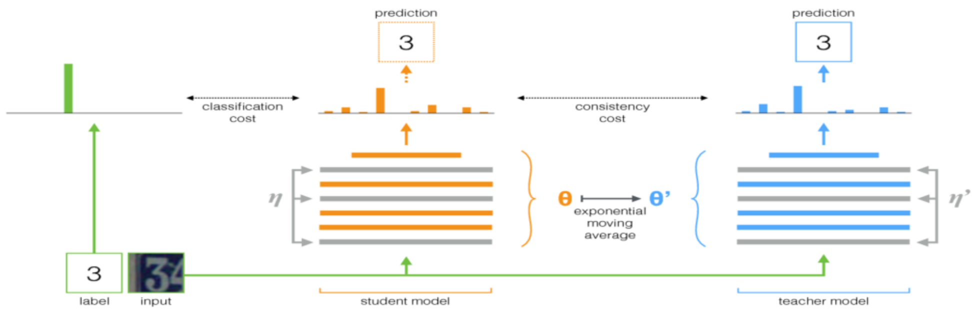


Figure: 1) Richer supervisions. 2) Multi-scale/-space views.
3) Explicit learning on the spatial semantic consistency.

Analysis

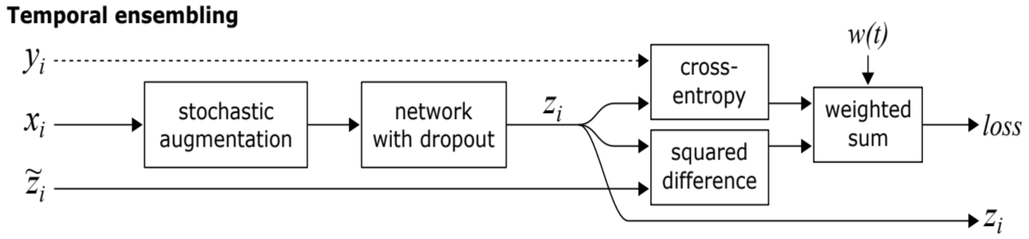
Existing Contrastive Learning Methods

Additional computation cost



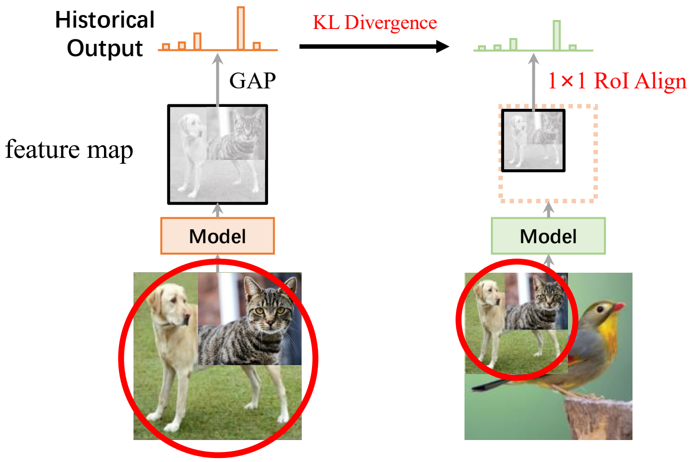
Mean teachers, NeurIPS 2017

Consume large memory



Temporal Ensemble , ICLR 2017

Ours



The additional computation/memory cost is negligible

ResNet-50 (300 epochs)	Memory	Flops	#P (deploy)	Top-1 Err (%)
Baseline	5.74 G	4.12 G	25.56 M	23.68
+ Mixup	5.74 G	4.12 G	25.56 M	22.58
+ CutMix	5.74 G	4.12 G	25.56 M	21.40
+ RM (ours)	5.74 G	4.12 G	25.56 M	**20.80**

Ablation Study

Classification

Model	RS	HS	CL	Top-1 Err (%)
PyramidNet	—	—	—	16.67
+CutMix [1]				15.59
+RM (ours)	√			15.36
	√	√		14.81
	√	√	√	**14.65**

Table: "RS": Resize strategy. "HS": Historical mix. "CL": Consistency loss.

Downstream

Detector	CL	AP	AP_{50}	AP_{75}
ATSS [2]		41.1	59.4	44.5
	√	**41.5**	**59.9**	**45.1**
GFL [3]		41.4	59.4	44.9
	√	**41.9**	**60.2**	**45.6**

Table: Object detection

Segmentor	CL	mIoU	mAcc	aAcc
PSPNet [4]		41.09	51.72	79.99
	√	**41.73**	**52.47**	**80.01**
UperNet [5]		41.88	**52.79**	79.94
	√	**42.30**	52.61	**80.14**

Table: Semantic segmentation

[1] Cutmix: Regularization strategy to train strong classifiers with localizable features. Yun S et al. ICCV 2019
[2] Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection, Zhang S et al. CVPR 2020
[3] Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection, Li X et al. NeurIPS 2020
[4] Pyramid scene parsing network, Zhao H et al. CVPR 2017
[5] Unified perceptual parsing for scene understanding, Xiao T et al. ECCV 2018.

Results

CIFAR10

PyramidNet-200 (300 epochs)	Top-1 Err (%)
Baseline	3.85
+ Label Smoothing	3.74
+ DropBlock	3.27
+ Stochastic Depth	3.11
+ Cutout	3.10
+ Mixup (α=1.0)	3.09
+ Manifold Mixup (α=1.0)	3.15
+ CutMix	2.88
+ MoEx	3.44
+ StyleCutMix (auto-γ)	2.55
+ RM (ours)	**2.35**

CIFAR100

Model (200 epochs)	Type	Top-1 Err (%)
ResNet-18	Baseline	21.70
	+ Mixup	20.99
	+ CutMix	19.61
	+ RM (ours)	**18.64**
ResNet-34	Baseline	20.62
	+ Mixup	19.19
	+ CutMix	17.89
	+ RM (ours)	**17.15**
DenseNet-121	Baseline	19.51
	+ Mixup	17.71
	+ CutMix	17.21
	+ RM (ours)	**16.22**
DenseNet-161	Baseline	18.78
	+ Mixup	16.84
	+ CutMix	16.64
	+ RM (ours)	**15.54**
PyramidNet-164	Baseline	16.67
	+ Mixup	16.02
	+ CutMix	15.59
	+ RM (ours)	**14.65**

ImageNet

ResNet-50 (300 epochs)	Top-1 Err (%)	Top-5 Err (%)
Baseline	23.68	7.05
+ Cutout	22.93	6.66
+ Stochastic Depth	22.46	6.27
+ Mixup	22.58	6.40
+ Manifold Mixup	22.50	6.21
+ DropBlock	21.87	5.98
+ Feature CutMix	21.80	6.06
+ CutMix	21.40	5.92
+ PuzzleMix	21.24	5.71
+ MoEx	21.90	6.10
+ CutMix + MoEx	20.90	5.70
+ RM (ours)	**20.80**	**5.42**

Results

Object detection

Detector	Pretrain Backbone	AP	AP_{50}	AP_{75}
ATSS	ResNet-50	39.4	57.6	42.8
	+ CutMix	40.1	58.4	43.4
	+ RM (ours)	**41.5**	**59.9**	**45.1**
	PVTv2-B1	39.3	57.2	42.5
	+ CutMix	41.8	60.3	45.5
	+ RM (ours)	**42.3**	**61.0**	**45.6**
GFL	ResNet-50	40.2	58.4	43.3
	+ CutMix	41.3	59.5	44.6
	+ RM (ours)	**41.9**	**60.2**	**45.6**
	PVTv2-B1	40.2	58.1	43.2
	+ CutMix	42.1	60.7	45.5
	+ RM (ours)	**43.0**	**61.6**	**46.5**

Semantic segmentation

Segmentor	Pretrain Backbone	mIoU	mAcc	aAcc
PSPNet	ResNet-50	40.90	51.11	79.52
	+ CutMix	40.96	51.16	79.93
	+ RM (ours)	**41.73**	**52.47**	**80.01**
	PVTv2-B1	36.48	46.26	76.79
	+ CutMix	37.99	48.70	77.50
	+ RM (ours)	**38.67**	**49.40**	**77.93**
UperNet	ResNet-50	40.40	51.00	79.54
	+ CutMix	41.24	51.79	79.69
	+ RM (ours)	**42.30**	**52.61**	**80.14**
	PVTv2-B1	39.94	50.75	79.02
	+ CutMix	41.73	52.99	80.02
	+ RM (ours)	**43.26**	**54.21**	**80.36**

Results

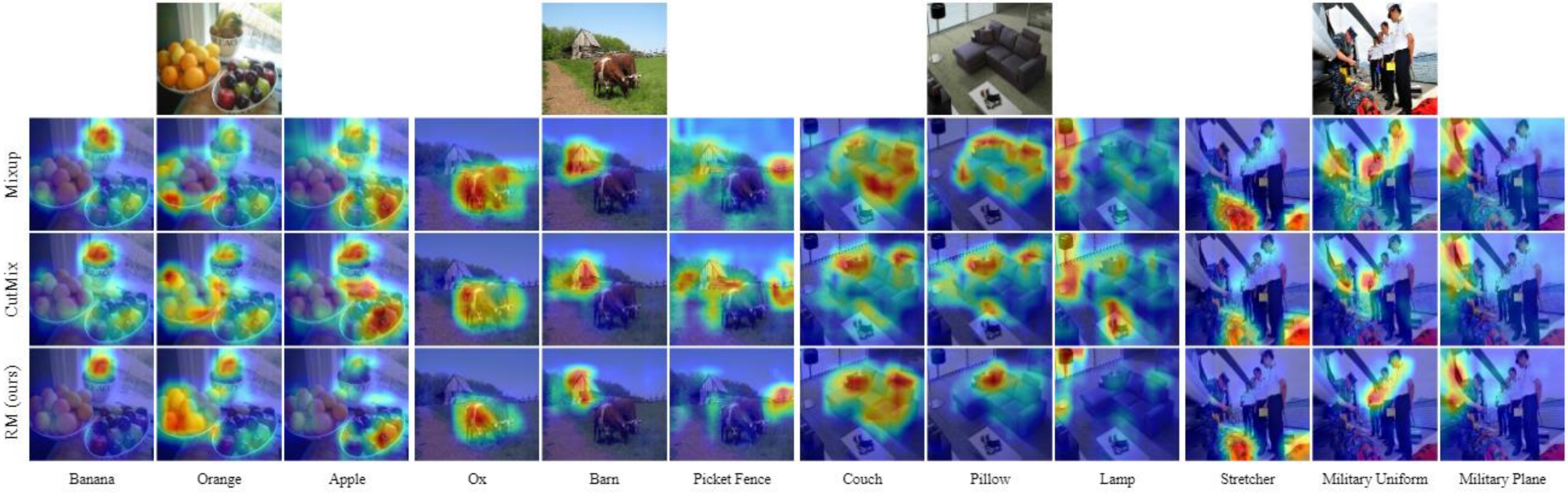


Figure: CAM visualization on natural samples with multiple labels.

- We propose recursive mix (RM) data augmentation, which constructs training pairs with identical inputs to learn spatial semantic consistency using historical prediction knowledge.
- RM shows better performance on image classification as well as various downstream tasks.

Thank you!

Codes and pretrained models are available at
https://github.com/implus/RecursiveMix