AutoMix: Unveiling the Power of Mixup for Stronger Classifiers

ECCV 2022 Oral

PRESENTER: JIAHANG ZHANG

2023/02/12

Outline

Outline

Outline

Background

Mix up & Cut Mix

Original Image Dog | Mix up Dog & Cat | Cut out Dog | Cut Mix Dog & Cat

$$\tilde{y} = \lambda y_A + (1 - \lambda) y_B$$

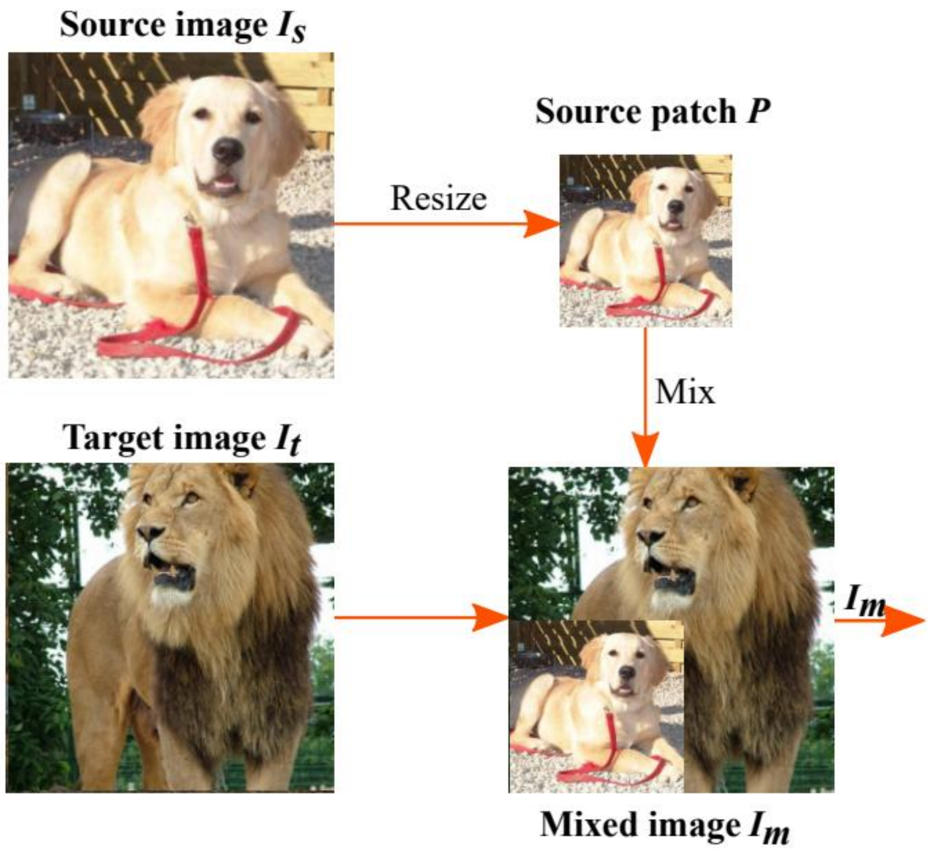
Background

Resize Mix

ResizeMix: Mixing Data with Preserved Object Information and True Labels

Jie Qin[1,2*], Jiemin Fang[3,4*], Qian Zhang[5], Wenyu Liu[4], Xingang Wang[2†], Xinggang Wang[4]

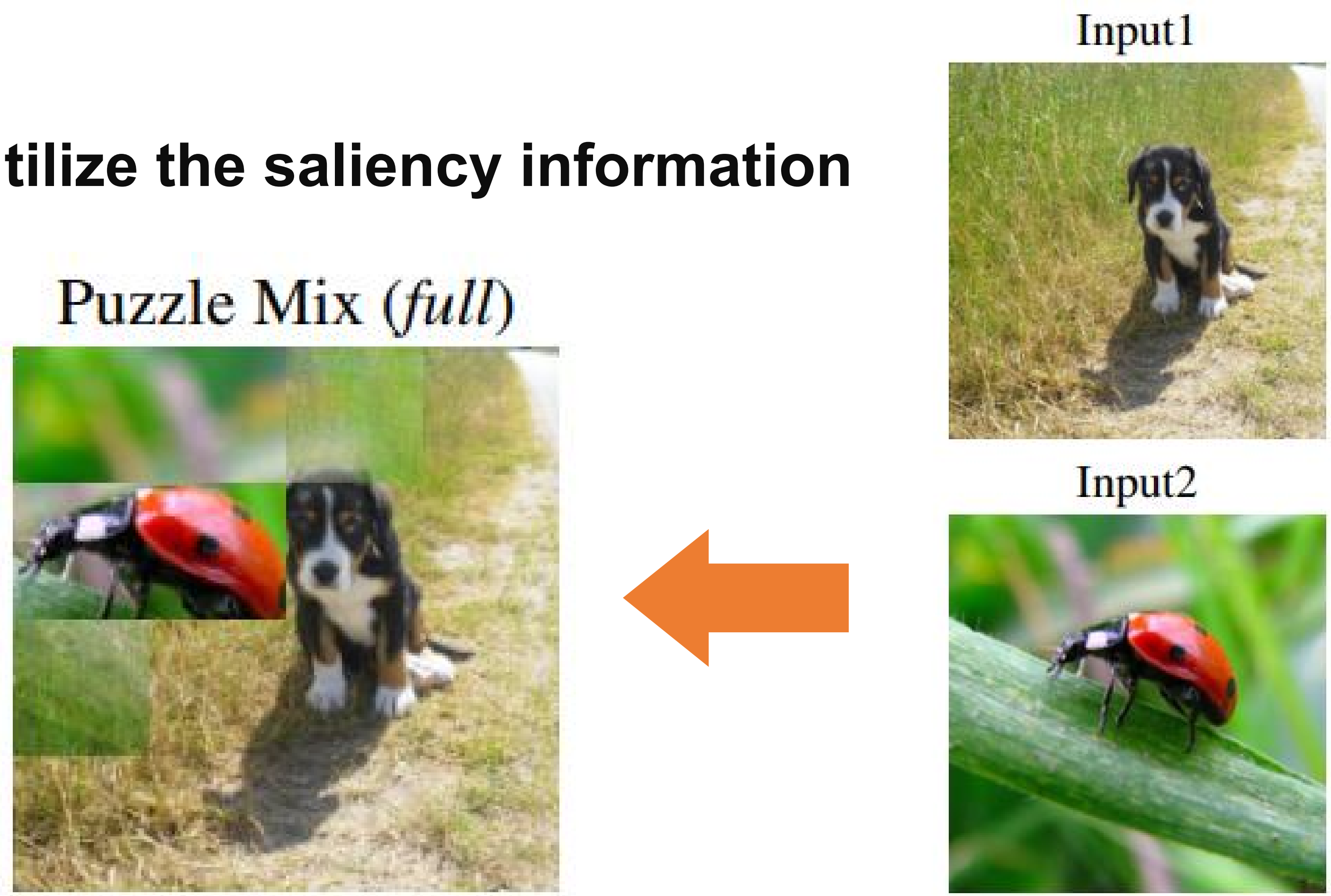
[1]School of Artificial Intelligence, University of Chinese Academy of Sciences
[2]Institute of Automation, Chinese Academy of Sciences
[3]Institute of Artificial Intelligence, Huazhong University of Science and Technology
[4]School of EIC, Huazhong University of Science and Technology [5]Horizon Robotics



$$l_m = \lambda l_s + (1-\lambda) l_t$$

Background

- **Puzzle Mix**
  - **Maximally utilize the saliency information**



Background

FMix: Enhancing Mixed Sample Data Augmentation

Ethan Harris, Antonia Marcu, Matthew Painter, Mahesan Niranjan, Adam Prügel-Bennett, Jonathon Hare

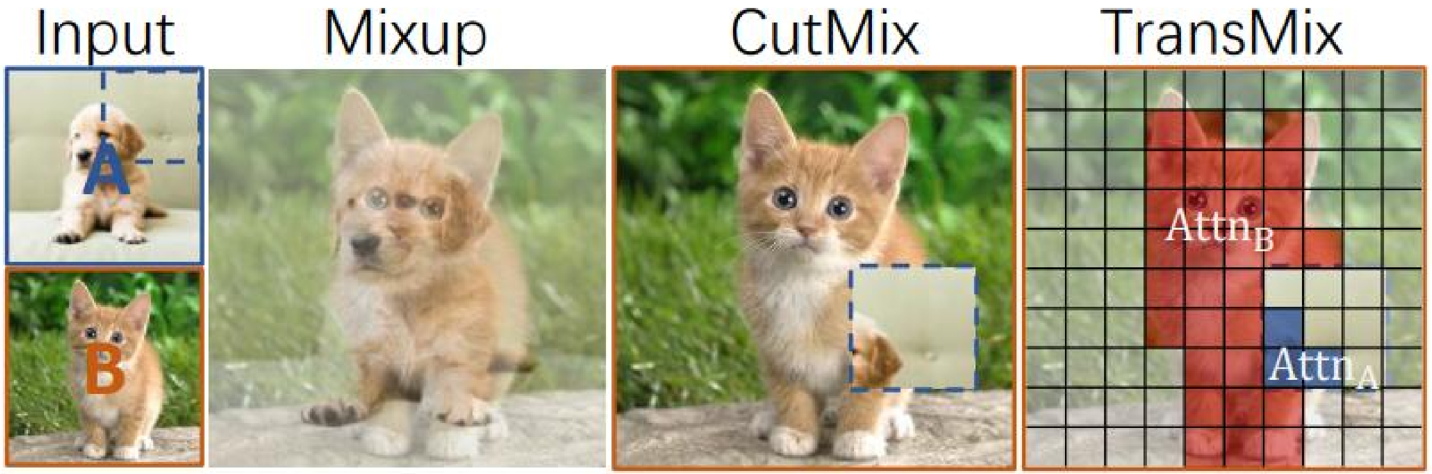
FMix

Background

TransMix: Attend to Mix for Vision Transformers

Jie-Neng Chen[1*] Shuyang Sun[2*] Ju He[1] Philip Torr[2] Alan Yuille[1] Song Bai[3]
[1]Johns Hopkins University [2]University of Oxford [3]ByteDance Inc.

- **TransMix**
  - **Attention-guided for ViTs**



Label $\quad \lambda \in [0,1] \quad \lambda = \dfrac{\mathrm{Area}_A}{\mathrm{Area}_{A\cup B}} \quad \lambda = \sum \mathrm{Attn}_A$

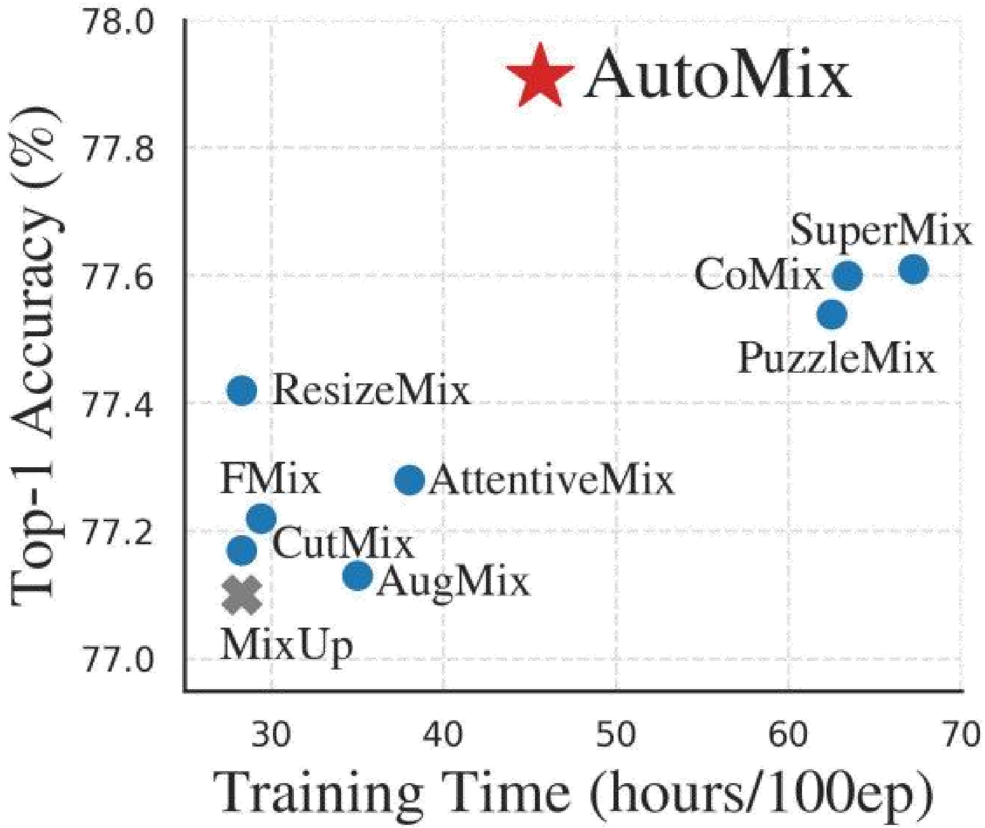
Background

- **Why Mix effective?**
  - **Data-dependent regulazation**
  - **Label smoothing**

Outline

Method

Motivation:



Method

- **Motivation:**
  - **How to design an accurate mixing policy to benefit the mixup classification objective?** ➡ *Label Mismatch Issue*
  - **How to solve generation-classification optimization problems efficiently?**

Method

Overview

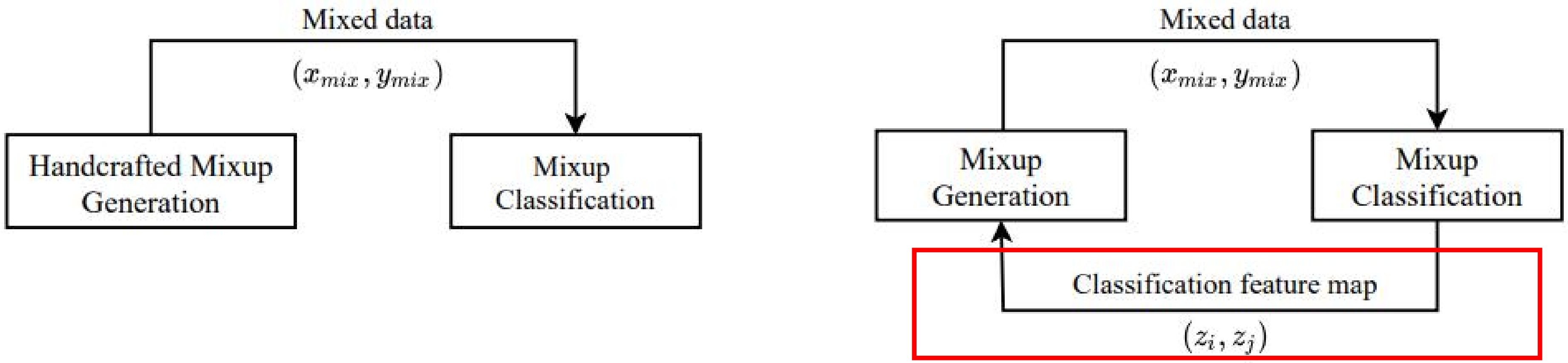


Fig. 2. The difference between AutoMix and offline approaches. **Left**: Offline mixup methods, where a fixed mixup policy generates mixed samples for the classifier to learn from. **Right**: AutoMix, where the mixup policy is trained with the feature map.

Method

- **Mixup policy for input** h_ϕ
- **Mixup policy for label** g

$$g(y_i, y_j, \lambda) = \lambda y_i + (1-\lambda) y_j \quad (1)$$

- **Optimization objective:**

$$\min_{\theta,\ \phi} \ell_{MCE}\Big(f_\theta(h_\phi(x_i, x_j, \lambda)), g(y_i, y_j, \lambda)\Big) \quad (2)$$

Method

- **Generate the pixel-level mask and obtain the mixed data**

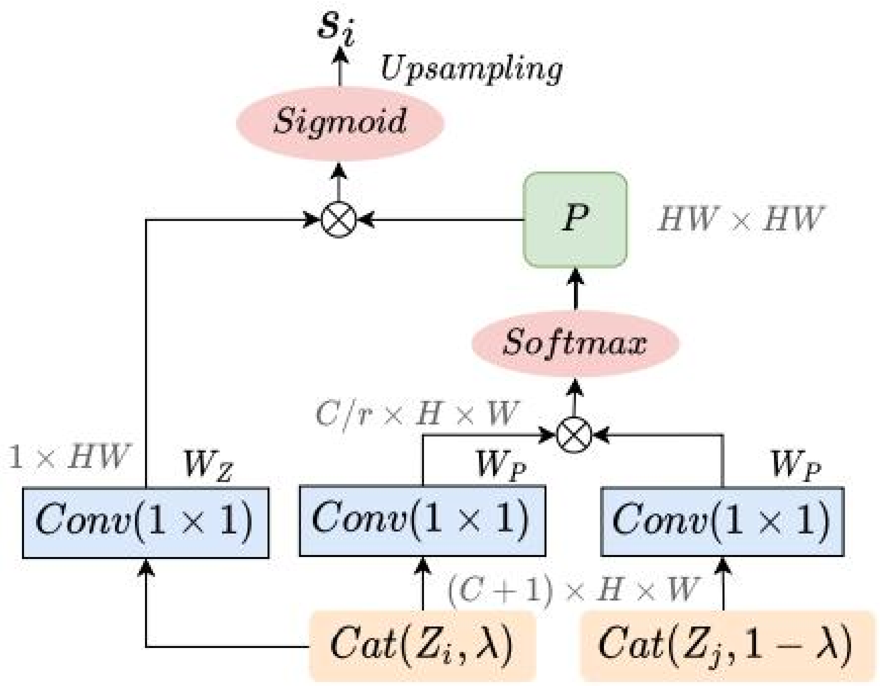
$$h_\phi(x_i, x_j, \lambda) = \mathcal{M}_\phi(z^l_{i,\lambda}, z^l_{j,1-\lambda}) \odot x_i + (1 - \mathcal{M}_\phi(z^l_{i,\lambda}, z^l_{j,1-\lambda})) \odot x_j \quad (3)$$

$$z^l_\lambda = \text{concat}(z, \lambda)$$

Method

■ **Generate the pixel-level mask and obtain the mixed data**

$$h_\phi(x_i, x_j, \lambda) = \mathcal{M}_\phi(z^l_{i,\lambda}, z^l_{j,1-\lambda}) \odot x_i + (1 - \mathcal{M}_\phi(z^l_{i,\lambda}, z^l_{j,1-\lambda})) \odot x_j \quad (3)$$



Method

- **Generate the pixel-level mask and obtain the mixed data**

$$h_\phi(x_i, x_j, \lambda) = \mathcal{M}_\phi(z^l_{i,\lambda}, z^l_{j,1-\lambda}) \odot x_i + (1 - \mathcal{M}_\phi(z^l_{i,\lambda}, z^l_{j,1-\lambda})) \odot x_j \quad (3)$$

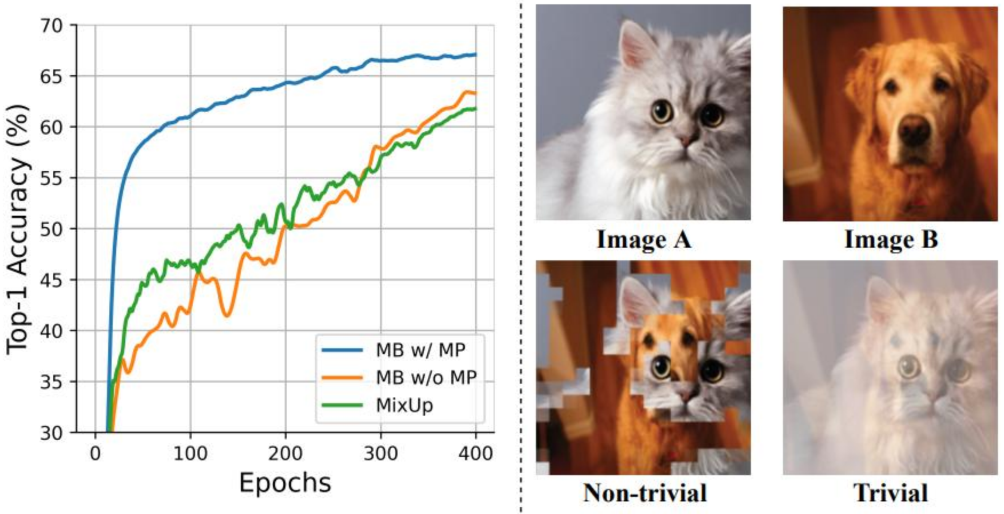
- **Objective for** $\mathcal{M}_\phi$

$$\ell_\lambda = \gamma \max\left(||\lambda - \frac{1}{HW}\sum_{h,w} s_{i,h,w}|| - \epsilon, 0\right) \quad (4)$$

Method

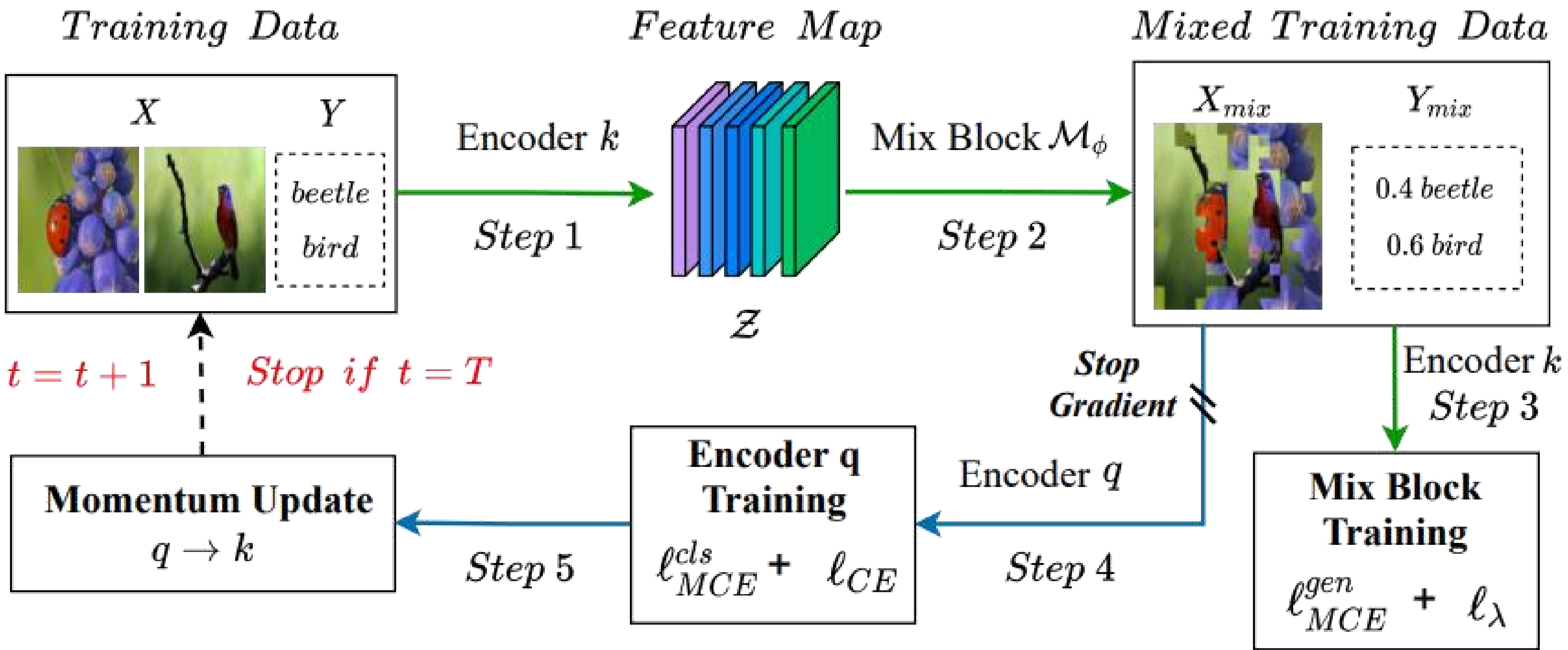
- A trick for training $\mathcal{M}_\phi$
  - Direct optimizing the two sub-tasks leads to gradient entanglement problem

$$\nabla_\phi \mathcal{L}^{cls}_{MCE} \propto \nabla_\phi h_\phi(x_i, x_j, \lambda) \odot f'_\theta(h_\phi(x_i, x_j, \lambda)) \quad (5)$$



Method

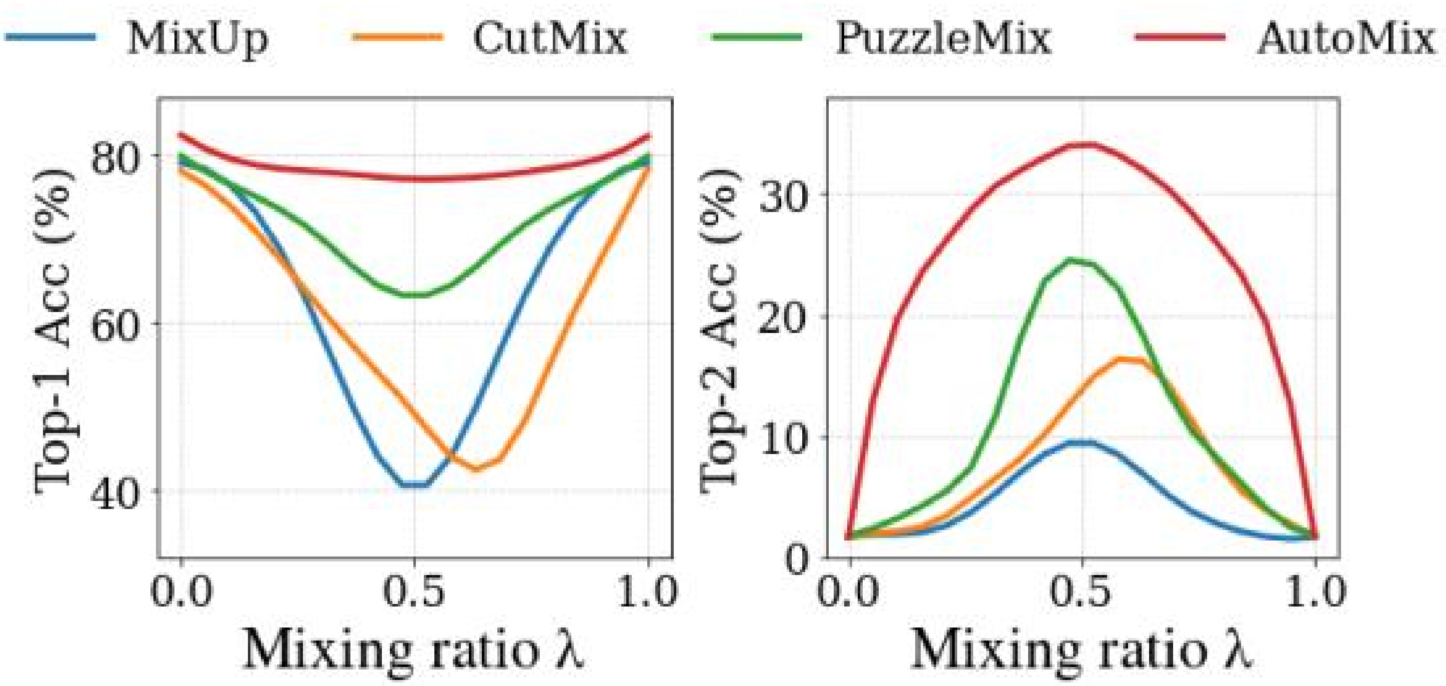
- A trick for training $\mathcal{M}_\phi$
  - To this end, Momentum Pipeline are proposed for bi-level optimization
  - Inspired by the self-supervised learning



Outline

Experiments

Quality of the proposed method



Experiments

Classification

Table 2. Top-1 accuracy (%)↑ of image classification based on ResNet variants on ImageNet-1k using PyTorch-style 100-epoch and 300-epoch training procedures.

	PyTorch 100 epochs					PyTorch 300 epochs			
Methods	R-18	R-34	R-50	R-101	RX-101	R-18	R-34	R-50	R-101
Vanilla	70.04	73.85	76.83	78.18	78.71	71.83	75.29	77.35	78.91
MixUp	69.98	73.97	77.12	78.97	79.98	71.72	75.73	78.44	80.60
CutMix	68.95	73.58	77.17	78.96	80.42	71.01	75.16	78.69	80.59
ManifoldMix	69.98	73.98	77.01	79.02	79.93	71.73	75.44	78.21	80.64
SaliencyMix	69.16	73.56	77.14	79.32	80.27	70.21	75.01	78.46	80.45
FMix*	69.96	74.08	77.19	79.09	80.06	70.30	75.12	78.51	80.20
PuzzleMix	70.12	74.26	77.54	79.43	80.53	71.64	75.84	78.86	80.67
ResizeMix*	69.50	73.88	77.42	79.27	80.55	71.32	75.64	78.91	80.52
AutoMix	**70.50**	**74.52**	**77.91**	**79.87**	**80.89**	**72.05**	**76.10**	**79.25**	**80.98**
Gain	**+0.38**	**+0.26**	**+0.37**	**+0.44**	**+0.34**	**+0.22**	**+0.26**	**+0.34**	**+0.31**

Experiments

Classification

Table 2. Top-1 accuracy (%)↑ of image classification based on ResNet variants on ImageNet-1k using PyTorch-style 100-epoch and 300-epoch training procedures.

	PyTorch 100 epochs					PyTorch 300 epochs			
Methods	R-18	R-34	R-50	R-101	RX-101	R-18	R-34	R-50	R-101
Vanilla	70.04	73.85	76.83	78.18	78.71	71.83	75.29	77.35	78.91
MixUp	69.98	73.97	77.12	78.97	79.98	71.72	75.73	78.44	80.60
CutMix	68.95	73.58	77.17	78.96	80.42	71.01	75.16	78.69	80.59
ManifoldMix	69.98	73.98	77.01	79.02	79.93	71.73	75.44	78.21	80.64
SaliencyMix	69.16	73.56	77.14	79.32	80.27	70.21	75.01	78.46	80.45
FMix*	69.96	74.08	77.19	79.09	80.06	70.30	75.12	78.51	80.20
PuzzleMix	70.12	74.26	77.54	79.43	80.53	71.64	75.84	78.86	80.67
ResizeMix*	69.50	73.88	77.42	79.27	80.55	71.32	75.64	78.91	80.52
AutoMix	**70.50**	**74.52**	**77.91**	**79.87**	**80.89**	**72.05**	**76.10**	**79.25**	**80.98**
Gain	**+0.38**	**+0.26**	**+0.37**	**+0.44**	**+0.34**	**+0.22**	**+0.26**	**+0.34**	**+0.31**

Experiments

Classification

Table 4. Top-1 accuracy (%)↑ on ImageNet-1k based on ViTs and ConvNeXt using DeiT training settings.

Methods	DeiT-S	Swin-T	ConvNeXt-T
DeiT	79.80	81.28	82.10
MixUp	79.65	81.01	80.88
CutMix	79.78	81.20	81.57
AttentiveMix	77.63	77.27	78.19
SaliencyMix	79.88	81.37	81.33
FMix*	77.37	79.60	81.04
PuzzleMix	80.45	81.47	81.48
ResizeMix*	78.61	81.36	81.64
TransMix†	80.70	**81.80**	-
AutoMix	**80.78**	**81.80**	**82.28**
Gain	**+0.08**	+0.00	**+0.18**

Outline

Discussion

- Review the proposed *AutoMix* framework
  - Optimize both the mixed sample generation task and the mixup classification task in a momentum training pipeline.
  - Without adding cost to inference, AutoMix can generate various masks adaptively.

Thanks!

Table 5. Top-1 accuracy (%)↑ of various algorithms based on ResNet variants on fine-grained and scenic classification datasets.

	CUB-200		FGVC-Aircraft		iNat2017		iNat2018		Places205	
Method	R-18	RX-50	R-18	RX-50	R-50	RX-101	R-50	RX-101	R-18	R-50
Vanilla	77.68	83.01	80.23	85.10	60.23	63.70	62.53	66.94	59.63	63.10
MixUp	78.39	84.58	79.52	85.18	61.22	66.27	62.69	67.56	59.33	63.01
CutMix	78.40	85.68	78.84	84.55	62.34	67.59	63.91	69.75	59.21	63.75
ManifoldMix	79.76	86.38	80.68	86.60	61.47	66.08	63.46	69.30	59.46	63.23
SaliencyMix	77.95	83.29	80.02	84.31	62.51	67.20	64.27	70.01	59.50	63.33
FMix*	77.28	84.06	79.36	86.23	61.90	66.64	63.71	69.46	59.51	63.63
PuzzleMix	78.63	84.51	80.76	86.23	62.66	67.72	64.36	70.12	59.62	63.91
ResizeMix*	78.50	84.77	78.10	84.08	62.29	66.82	64.12	69.30	59.66	63.88
AutoMix	**79.87**	**86.56**	**81.37**	**86.72**	**63.08**	**68.03**	**64.73**	**70.49**	**59.74**	**64.06**
Gain	+0.11	+0.18	+0.61	+0.12	+0.42	+0.31	+0.37	+0.37	+0.08	+0.15

Table 6. Top-1 accuracy (%)↑ and FGSM error (%)↓ on CIFAR-100 based on ResNeXt-50 (32x4d) trained 400 epochs.

	Clean Acc(%)↑	Corruption Acc(%)↑	FGSM Error(%)↓
Vanilla	80.24	51.71	63.92
MixUp	82.44	58.10	56.60
CutMix	81.09	49.32	76.84
AugMix	81.18	66.54	55.59
PuzzleMix	82.76	57.82	63.71
AutoMix	**83.13**	**58.35**	**55.34**

Table 7. Trasfer learning of object detection task with Faster-RCNN on Pascal VOC and COCO datasets.

	VOC	COCO		
Methods	mAP	mAP	AP^{bb}_{50}	AP^{bb}_{75}
Vanilla	81.0	38.1	59.1	41.8
Mixup	80.7	37.9	59.0	41.7
CutMix	81.9	38.2	59.3	42.0
PuzzleMix	81.9	38.3	59.3	42.1
ResizeMix	82.1	38.4	59.4	42.1
AutoMix	**82.4**	**38.6**	**59.5**	**42.2**

Table 6. Top-1 accuracy (%)↑ and FGSM error (%)↓ on CIFAR-100 based on ResNeXt-50 (32x4d) trained 400 epochs.

	Clean Acc(%)↑	Corruption Acc(%)↑	FGSM Error(%)↓
Vanilla	80.24	51.71	63.92
MixUp	82.44	58.10	56.60
CutMix	81.09	49.32	76.84
AugMix	81.18	66.54	55.59
PuzzleMix	82.76	57.82	63.71
AutoMix	**83.13**	**58.35**	**55.34**

Table 9. Ablation of modules in MixBlock.

module	Tiny-ImageNet R-18	Tiny-ImageNet RX-50
(random grids)	64.40	66.83
+cross attention	66.87	69.76
+λ embedding	67.15	70.41
+ℓ_λ	**67.33**	**70.72**

Table 10. Ablation of the proposed momentum pipeline (MP) and the cross-entropy loss l_{CE} (CE) based on ResNet-18.

	CIFAR-100			Tiny-ImageNet			ImageNet-1k		
modules	MixUp	CutMix	$\mathcal{M}_\phi$	MixUp	CutMix	$\mathcal{M}_\phi$	MixUp	CutMix	$\mathcal{M}_\phi$
(none)	79.12	78.17	79.46	63.39	64.40	64.84	69.98	68.95	70.04
+MP(m=0)	-	-	81.75	-	-	67.05	-	-	70.41
+MP	**80.82**	79.57	81.93	66.02	**65.72**	67.19	**70.13**	70.02	70.45
+MP+CE	80.41	**79.64**	**82.04**	**66.10**	65.05	**67.33**	70.10	**70.04**	**70.50**