西湖大學
WESTLAKE UNIVERSITY

Mixup Data Augmentation for Computer Vision

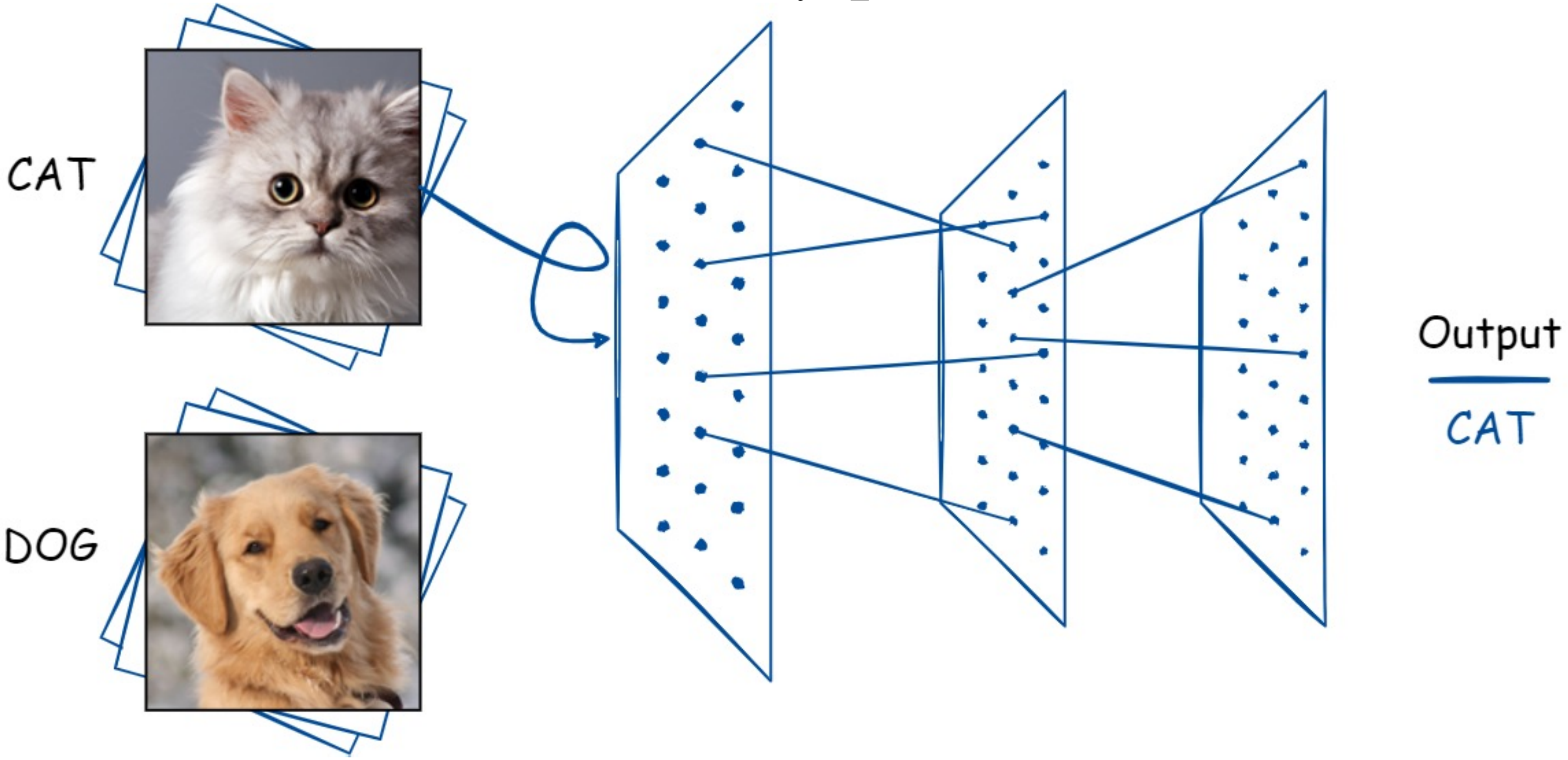
Siyuan Li

December, 2023

Preamble

- Learning a deep model

$$S = \{x_i, y_i\}_{i=1}^{n} \rightarrow Classifier,$$



Preamble

- **Mixup (Zhang et al. 2018) in Deep Learning**

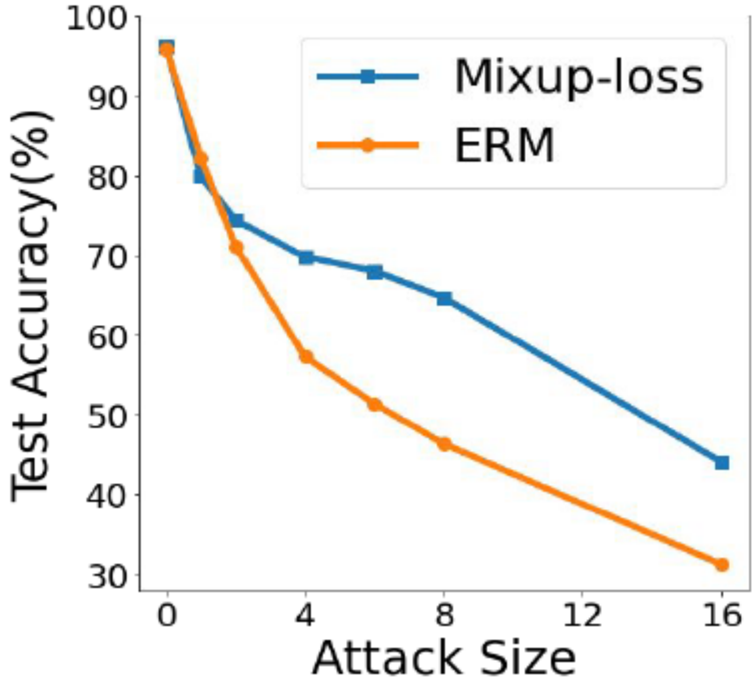
$$\tilde{S} = \{\tilde{x}_i, \tilde{y}_i\}_{i=1}^{n} \rightarrow Classifier,$$

where

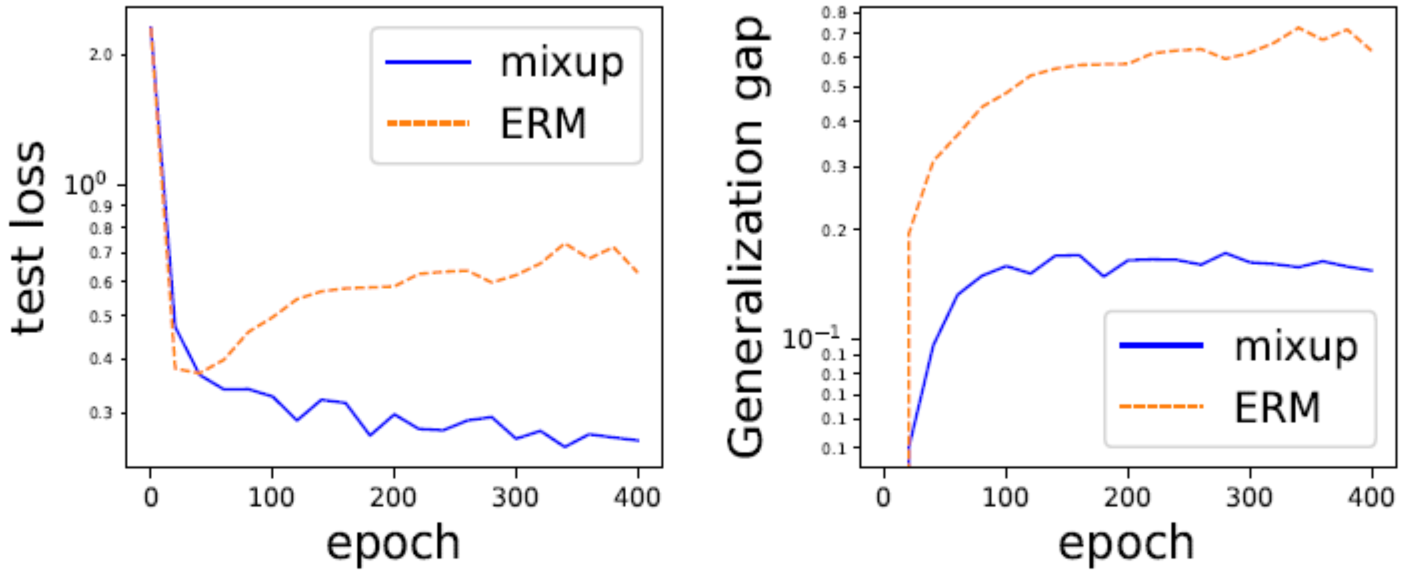
$$\tilde{x}_i = \lambda x_i + (1-\lambda)x_j, \tilde{y}_i = \lambda y_i + (1-\lambda)y_j, \lambda \sim Beta(\alpha, \beta) \in [0,1].$$

Preamble

- **Mixup Improves Generalization and Robustness (Zhang et al. 2021)**



(a) Robustness (Lamb et al. 2019)



(b) Generalization (Guo et al. 2019)

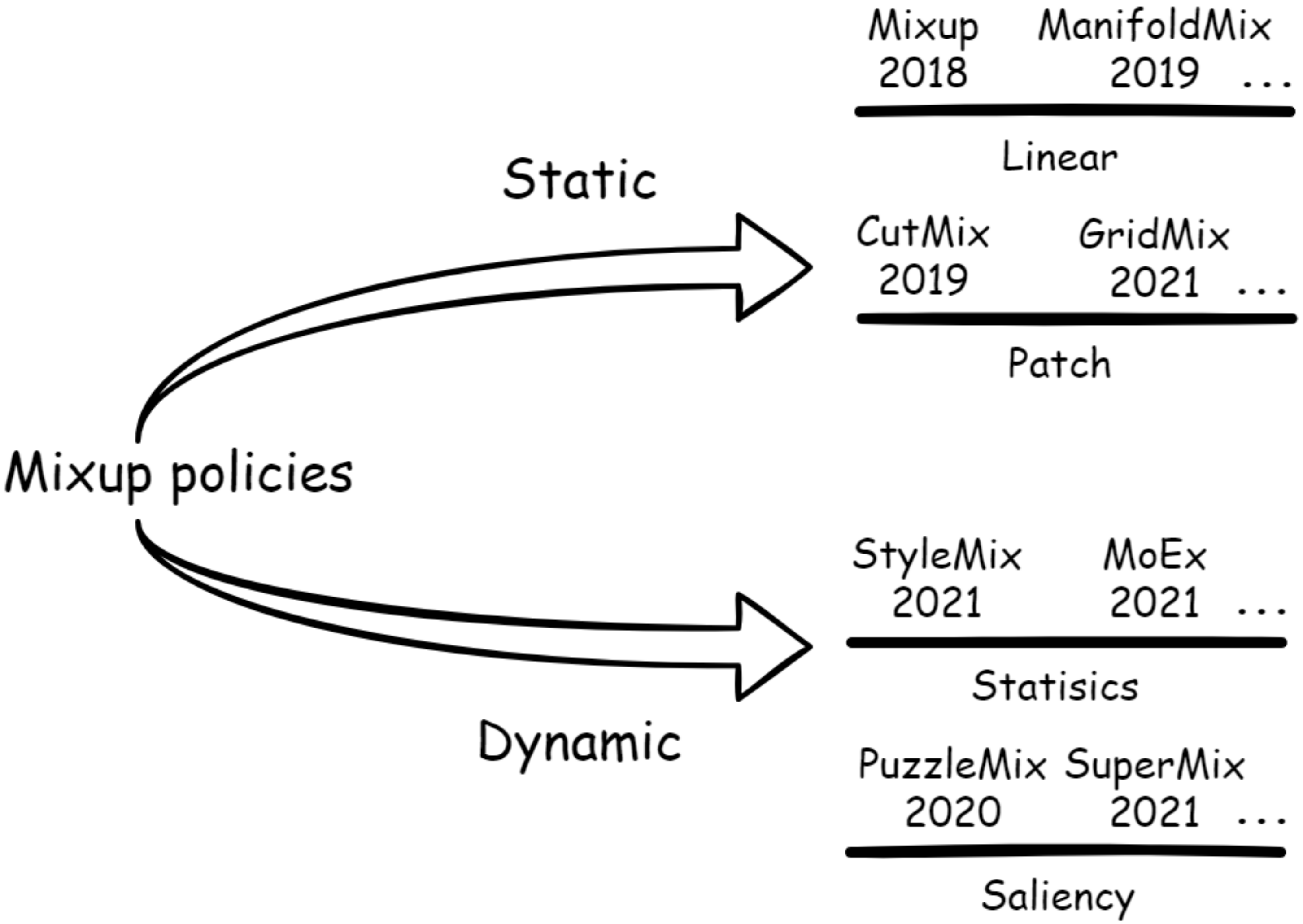
AutoMix: Unveiling the Power of Mixup for Stronger Classifiers

Zicheng Liu[1,2], Siyuan Li[1,2], Di Wu[1,2], Zihan Liu[1,2], Zhiyuan Chen[2], Lirong Wu[1,2] , and Stan Z. Li[2]

[1]Zhejiang University, [2]AI Lab, Westlake University,

Paper: https://arxiv.org/abs/2103.13027

Related Works



Heuristic random mixing methods
Low mixing precision.

Offline optimization mixing methods
High complexity.

Problems

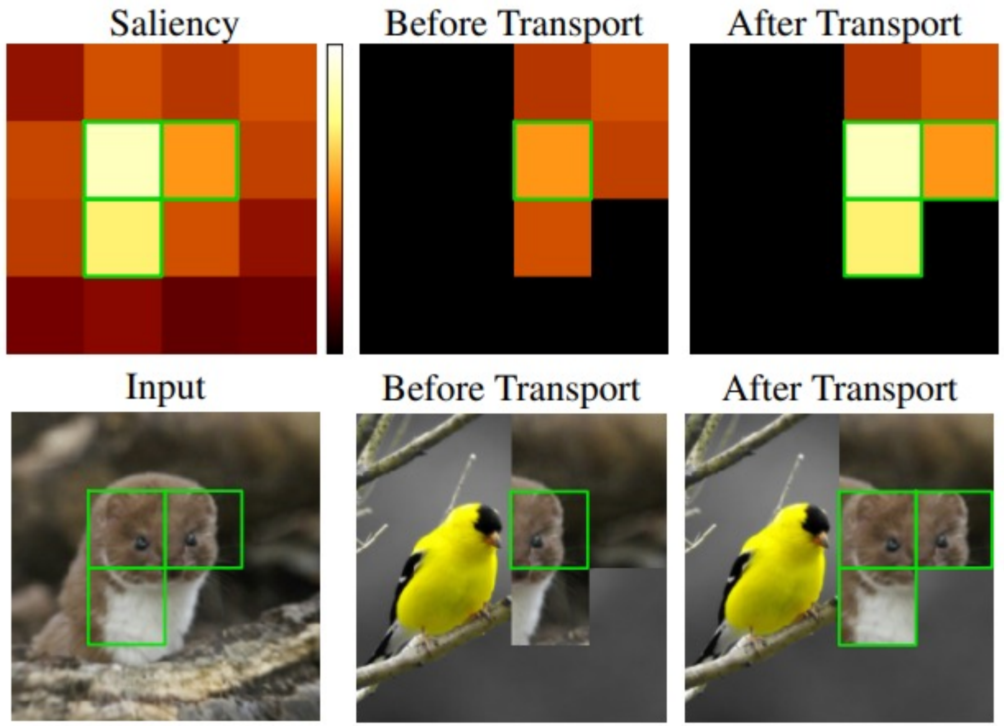
- Label Mismatch (CutMix etc.)

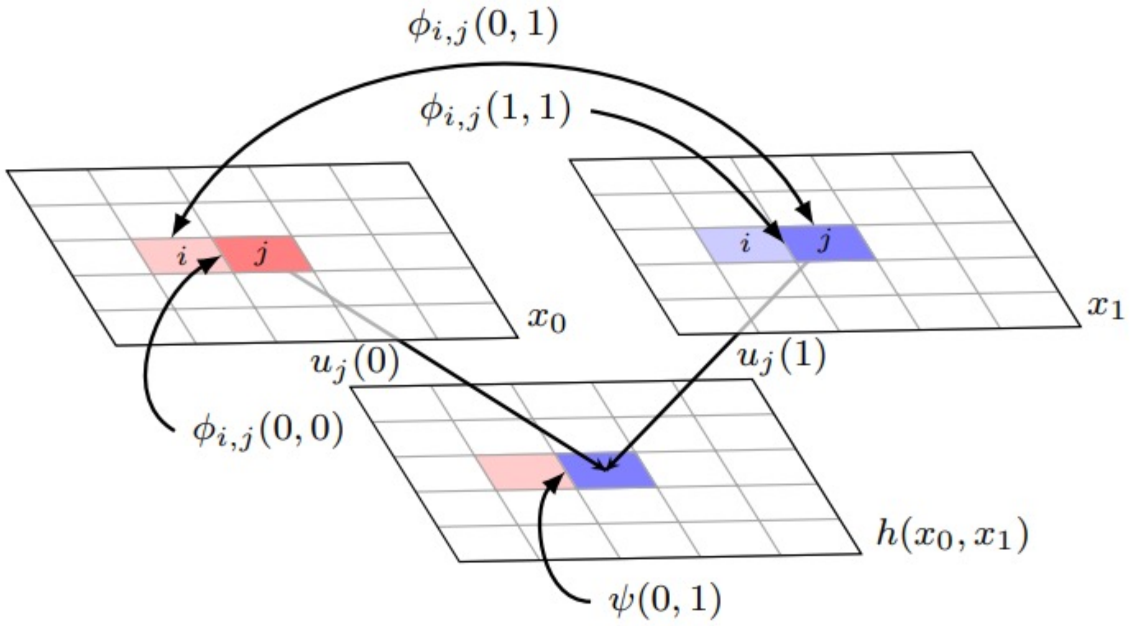
"Bird" "Panda" "Bird" + "Panda"

There is a mismatching between mixed samples and labels.

Problems

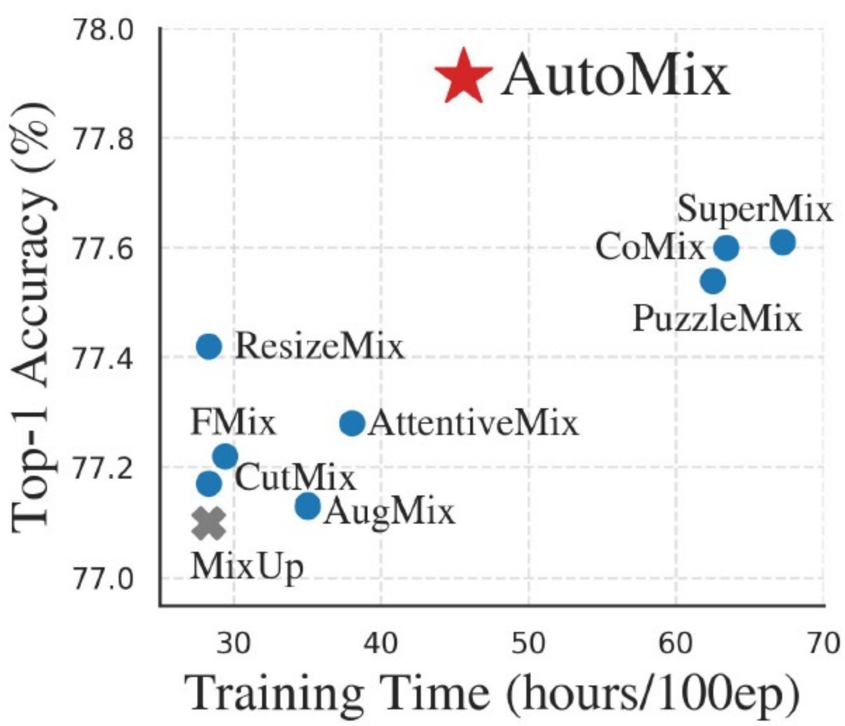
- High Complexity (PuzzleMix etc.)

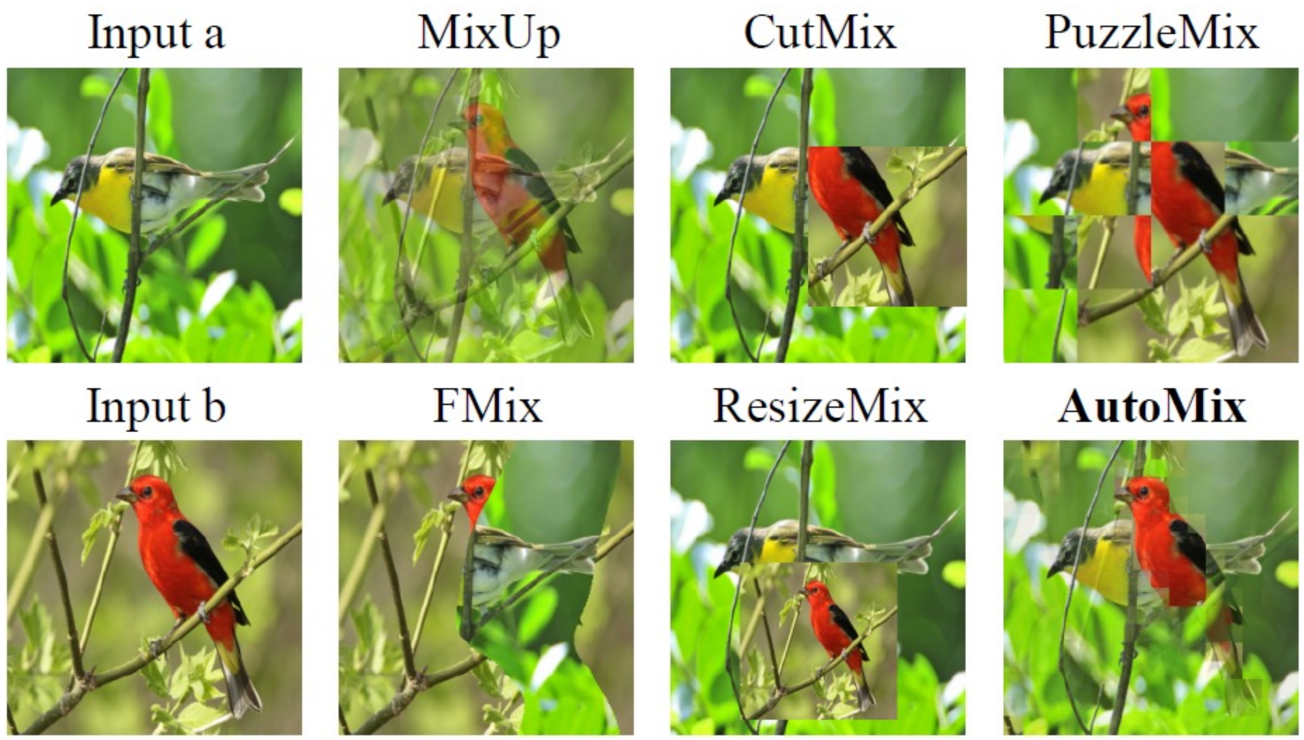




AutoMix

- How to balance precise mixing polices and complexity?





Solve the mixup problem in an end-to-end manner.

AutoMix: Reformulates Mixup

- Standard cross-entropy (CE) training

$$\ell_{CE}(f_\theta(x), y) = -y \log f_\theta(x).$$

- Standard mixup CE (MCE) training

$$\ell_{MCE} = \lambda \ell_{CE}(f_\theta(x_{mix}), y_i) + (1-\lambda)\ell_{CE}(f_\theta(x_{mix}), y_j).$$

- Mixup reformulation

$$\min_{\theta,\ \phi} \ell_{MCE}\Big(f_\theta\big(\underbrace{h_\phi(x_i, x_j, \lambda)}_{\textbf{Sample mixing}}\big), \underbrace{g(y_i, y_j, \lambda)}_{\text{Label mixing}}\Big).$$

Parameterize mixup function h as ϕ and optimize online with encoder f_θ.

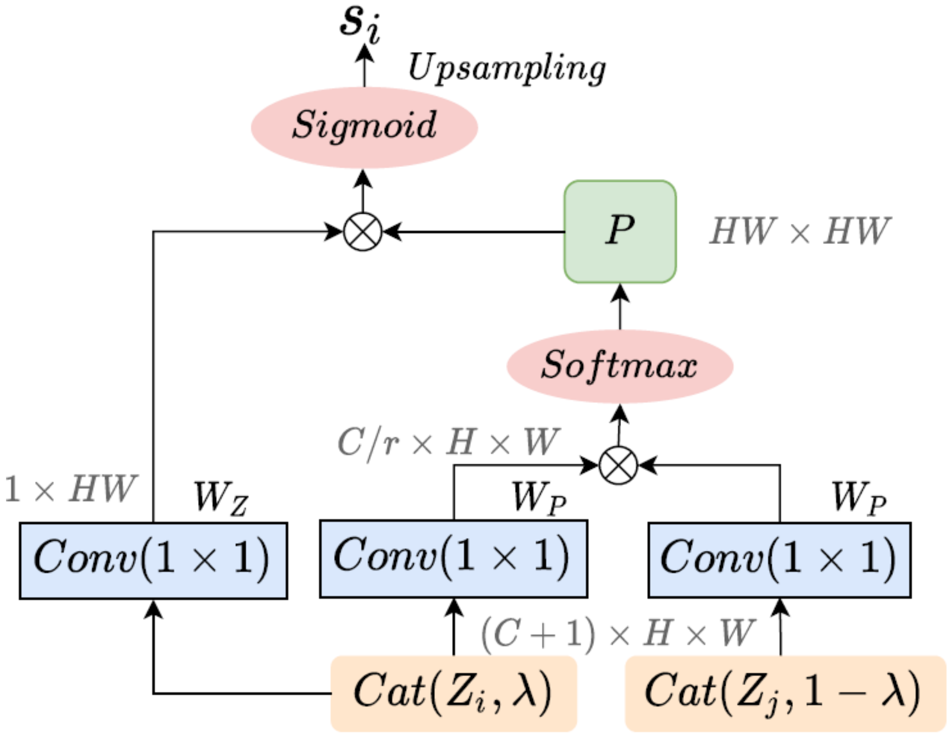
AutoMix: Mix Block

- How to capture the pixel-level pair-wise relationships?
  - Take deep feature Z as input.

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

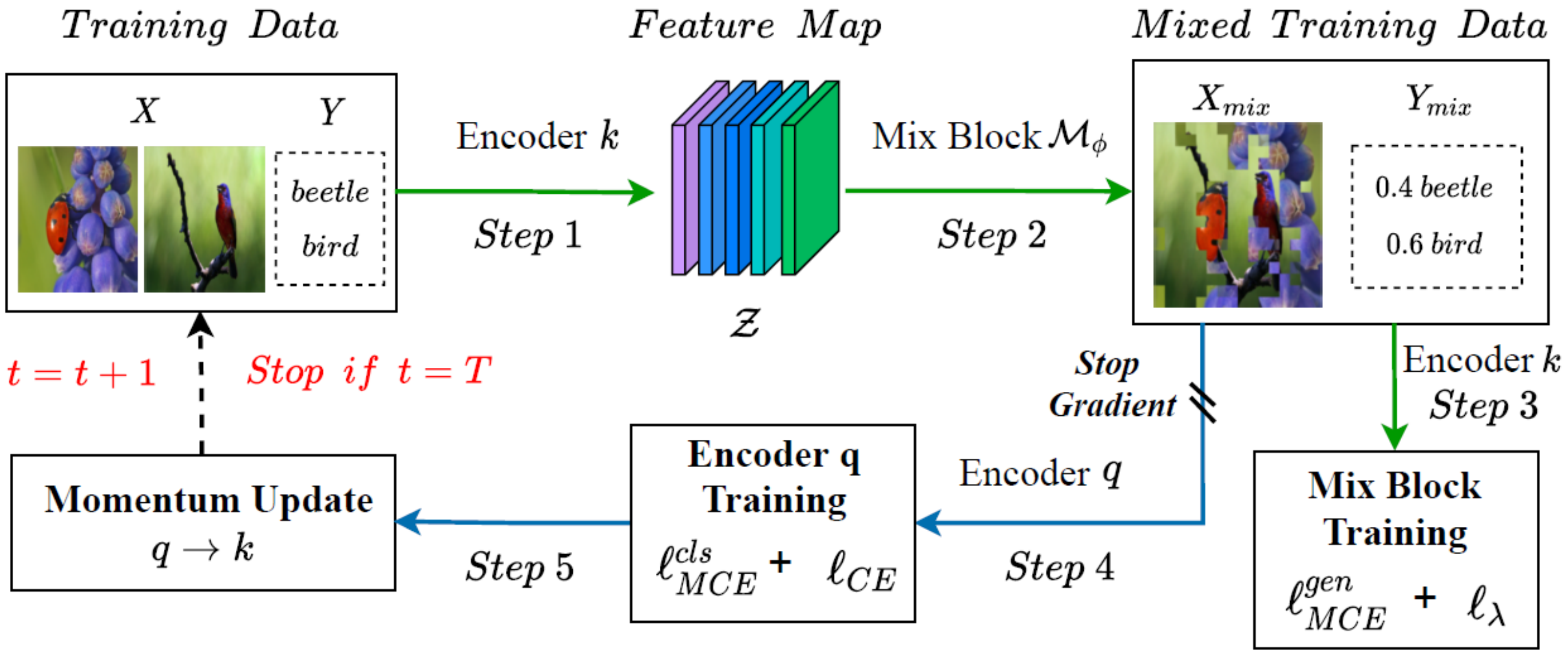
  - Cross-attention mechanism.

$$P(z^l_{i,\lambda}, z^l_{j,1-\lambda}) = \mathrm{softmax}\left(\frac{(W_P\, z^l_{i,\lambda})^T \otimes W_P\, z^l_{j,1-\lambda}}{C(z^l_{i,\lambda}, z^l_{j,1-\lambda})}\right),$$



AutoMix: Momentum Pipeline

- How to stabilize this bi-level optimization?



$$\ell_\lambda = \gamma \max\left(||\lambda - \frac{1}{HW}\sum_{h,w} s_{i,h,w}|| - \epsilon, 0\right), \quad \mathcal{L}(\theta,\phi) = \underbrace{\ell_{CE} + \ell_{MCE}^{cls}}_{classification} + \underbrace{\ell_{MCE}^{gen} + \ell_\lambda}_{generation}.$$

AutoMix: Momentum Pipeline

- Results

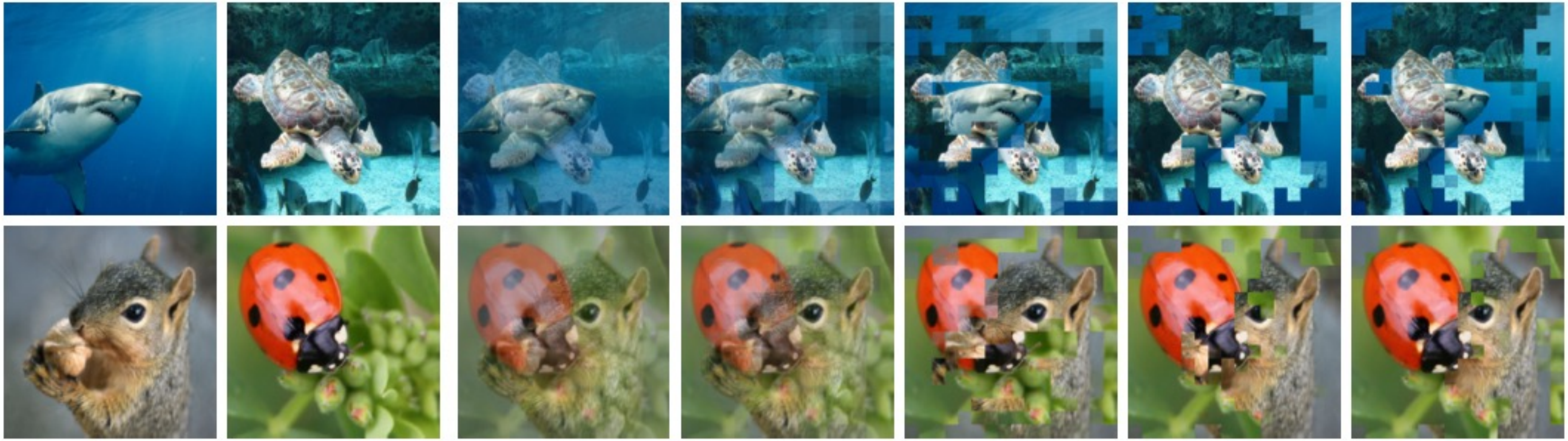


Image A | Image B | Initialization | Epoch 2 | Epoch 5 | Epoch 20 | End of training

Experiments

- Small-scale Datasets

Method	CIFAR-10 R-18	CIFAR-10 RX-50	CIFAR-100 R-18	CIFAR-100 RX-50	CIFAR-100 WRN-28-8	Tiny-ImageNet R-18	Tiny-ImageNet RX-50
Vanilla	95.50	96.23	78.04	81.09	81.63	61.68	65.04
MixUp	96.62	97.30	79.12	82.10	82.82	63.86	66.36
CutMix	96.68	97.01	78.17	81.67	84.45	65.53	66.47
ManifoldMix	96.71	97.33	80.35	82.88	83.24	64.15	67.30
SaliencyMix	96.53	97.18	79.12	81.53	84.35	64.60	66.55
FMix*	96.58	96.76	79.69	81.90	84.21	63.47	65.08
PuzzleMix	97.10	97.27	81.13	82.85	85.02	65.81	67.83
Co-Mixup	97.15	97.32	81.17	82.91	85.05	65.92	68.02
ResizeMix*	96.76	97.21	80.01	81.82	84.87	63.74	65.87
AutoMix	**97.34**	**97.65**	**82.04**	**83.64**	**85.18**	**67.33**	**70.72**
Gain	**+0.19**	**+0.32**	**+0.87**	**+0.76**	**+0.13**	**+1.41**	**+2.70**

Experiments

- ImageNet

Methods	PyTorch 100 epochs R-18	R-34	R-50	R-101	RX-101	PyTorch 300 epochs 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**

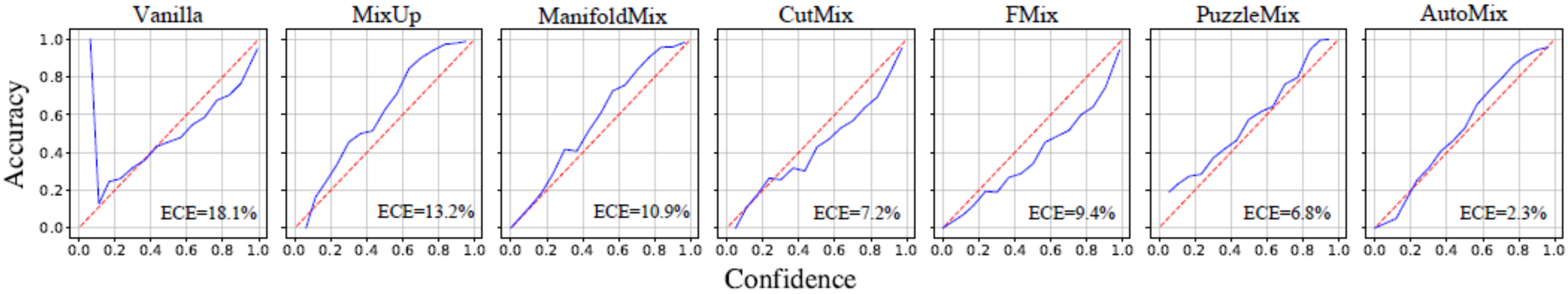
Experiments

- Fine-grained classification

Method	CUB-200 R-18	CUB-200 RX-50	FGVC-Aircraft R-18	FGVC-Aircraft RX-50	iNat2017 R-50	iNat2017 RX-101	iNat2018 R-50	iNat2018 RX-101	Place205 R-18	Place205 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

Experiments

- Calibration



- Weakly supervised object localization

Backbone	Vanilla	Mixup	CutMix	FMix*	PuzzleMix	Co-Mixup	**Ours**
R-18	49.91	48.62	51.85	50.30	53.95	54.13	**54.46**
RX-50	53.38	50.27	57.16	59.80	59.34	59.76	**61.05**

Experiments

- Robustness and transfer learning

	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**

Robustness

Methods	VOC mAP	COCO mAP	COCO AP^{bb}_{50}	COCO 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**

Transfer Learning
(object detection)

Experiments

- Ablation Study
  - Are the modules in Mix Block effective?
  - How many gains can Mix Block bring without EMA and CE?
  - Is AutoMix robust to hyperparameters?

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

	R-18			RX-50		
	Acc(%)	Params	Time	Acc(%)	Params	Time
Mixup	63.86	11.27	20	66.36	23.38	113
l_1	67.30	11.38	67	70.70	23.80	413
l_2	67.27	11.39	41	70.43	23.86	252
l_3	**67.33**	11.44	34	**70.72**	24.84	196
l_4	67.32	11.64	28	70.67	27.99	174

ImageNet-1k		
MixUp	CutMix	$\mathcal{M}_\phi$
69.98	68.95	70.04
-	-	70.41
70.13	70.02	70.45
70.10	**70.04**	**70.50**

Harnessing Hard Mixed Samples with Decoupled Regularizer

Zicheng Liu[1,2], Siyuan Li[1,2], Ge Wang[1,2], Chen Tan[1,2], Lirong Wu[1,2] , and Stan Z. Li[2]

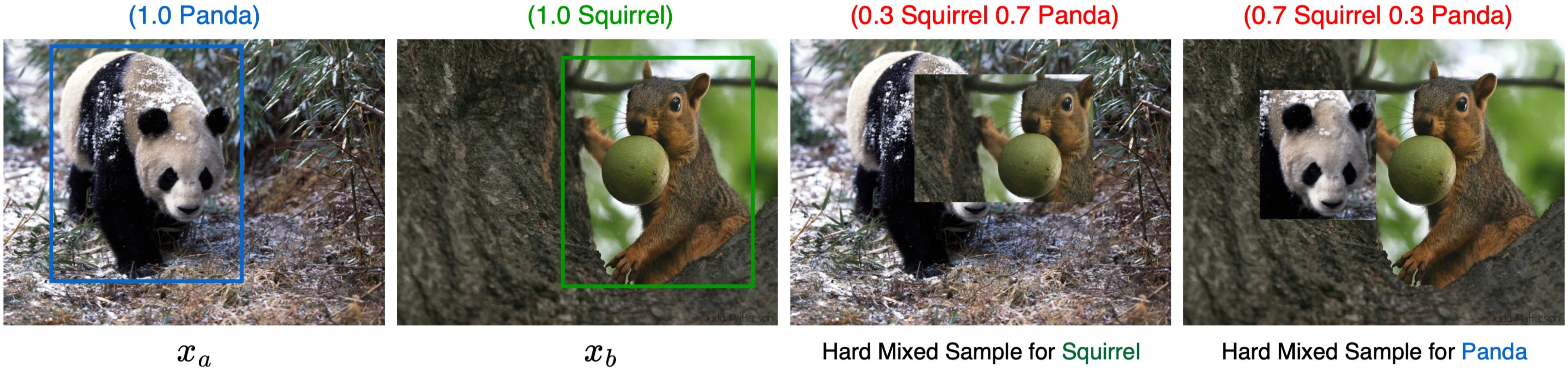
[1]Zhejiang University, [2]AI Lab, Westlake University,

Paper: https://arxiv.org/abs/2203.10761

Problems

NEURAL INFORMATION PROCESSING SYSTEMS

- Hard Mixed Samples (CutMix etc.)



There is a semantic mismatching between mixed samples and labels.

We hope to improve the prediction confidence in these cases.

Preliminary

- Mixed Cross-Entropy Underutilizes Mixup

$$(\nabla_{z_{(a,b)}}\mathcal{L}_{MCE})^i = \begin{cases} -\lambda + \frac{\exp(z^i_{(a,b)})}{\sum_c \exp(z^c_{(a,b)})}, & i = a \\ -(1-\lambda) + \frac{\exp(z^i_{(a,b)})}{\sum_c \exp(z^c_{(a,b)})}, & i = b \\ \frac{\exp(z^i_{(a,b)})}{\sum_c \exp(z^c_{(a,b)})}, & i \neq a, b \end{cases}$$

The confidence of mixed classes is forced to be related to λ.

Could we preserve the smoothness and achieve more confidence?

Decoupled regularizer

- Softmax Degrades Confidence in Mixup (winner takes all).

$$\sigma(z_{(a,b)})^i = \frac{\exp(z^i_{(a,b)})}{\sum_c \exp(z^c_{(a,b)})}.$$

- Decoupled Softmax (remove the competitor).

$$\phi(z_{(a,b)})^{i,j} = \frac{\exp(z^i_{(a,b)})}{\cancel{\exp(z^j_{(a,b)})} + \sum_{c\neq j} \exp(z^c_{(a,b)})}.$$

- The Mixup with Decoupled Regulerizer.

$$\mathcal{L}_{DM(CE)} = -\Big(\underbrace{y^T_{(a,b)} \log(\sigma(z_{(a,b)}))}_{\mathcal{L}_{MCE}} + \eta \underbrace{y^T_{[a,b]} \log(\phi(z_{(a,b)}))y_{[a,b]}}_{\mathcal{L}_{DM}}\Big).$$

Asymmetrical Strategy

- Reliable connection to augment data.

$$\hat{x}_{(a,b)} = \lambda x_a + (1-\lambda)u_b; \quad \hat{y}_{(a,b)} = \lambda y_a + (1-\lambda)v_b.$$

Notice: u_b is unlabeled data and λ is fixed less than 0.5.

- The Decoupled Mixup.

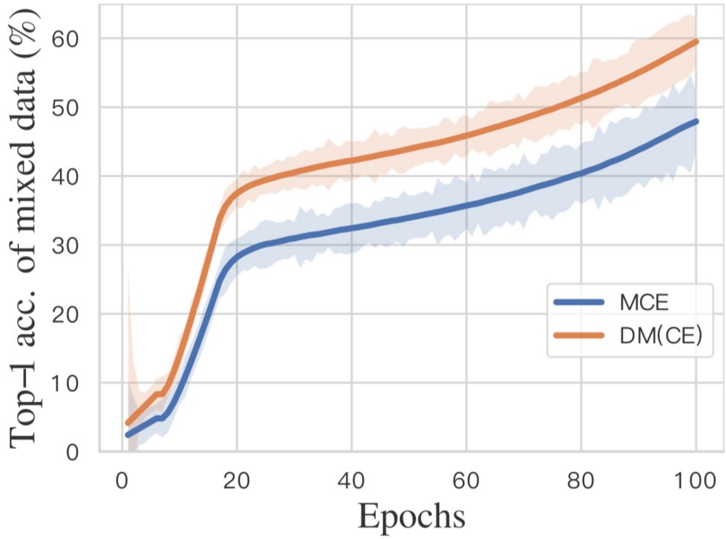
$$\hat{\mathcal{L}}_{DM} = y_a^T \log\left(\phi(z_{(a,b)})\right) y_b,$$

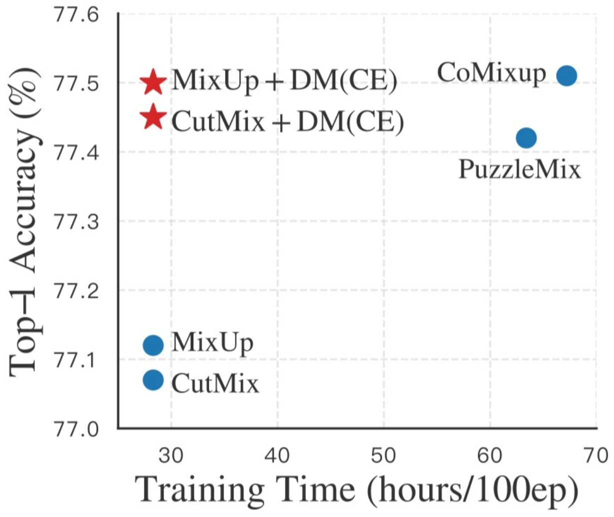
Notice: we only retained the labeled part.

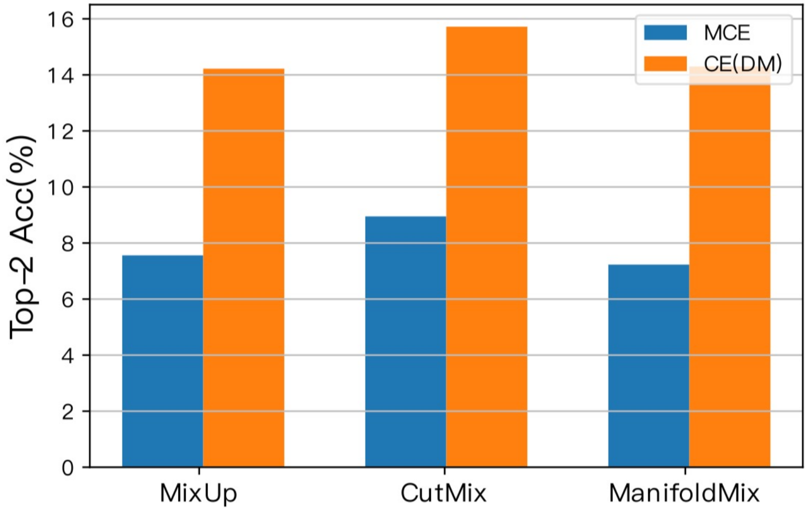
Fully utilize labeled data by applying our decoupled mechanism.

Practical Consequences

- Make What Should be Certain More Certain.
  - The model trained with decoupled mixup mostly doubled the top-2 mixup accuracy.
- Enhance the Training Efficiency.
  - Boosting performance without extra computation.







Supervised Learning

Datasets	CIFAR-100						Tiny-ImageNet			
	R-18		RX-50		WRN-28-8		R-18		RX-50	
Methods	MCE	**DM(CE)**	MCE	**DM(CE)**	MCE	**DM(CE)**	MCE	**DM(CE)**	MCE	**DM(CE)**
Mixup	79.12	**80.44**	82.10	**82.96**	82.82	**83.51**	63.86	**65.07**	66.36	**67.70**
CutMix	78.17	**79.39**	81.67	**82.39**	84.45	**84.88**	65.53	**66.45**	66.47	**67.46**
ManifoldMix	80.35	**81.05**	82.88	**83.15**	83.24	**83.72**	64.15	**65.45**	67.30	**68.48**
FMix	79.69	**80.12**	81.90	**82.74**	84.21	**84.47**	63.47	**65.34**	65.08	**66.96**
ResizeMix	80.01	**80.26**	81.82	**82.96**	**84.87**	84.72	63.74	**64.33**	65.87	**68.56**
Avg. Gain		+0.78		+0.77		+0.34		+1.18		+1.62

Benchmarking on toy datasets

	R-18		R-34		R-50	
Methods	MCE	**DM(CE)**	MCE	**DM(CE)**	MCE	**DM(CE)**
Vanilla	70.04	-	73.85	-	76.83	-
Mixup	69.98	**70.20**	73.97	**74.26**	77.12	**77.41**
CutMix	68.95	**69.26**	73.58	**73.88**	77.07	**77.32**
ManifoldMix	69.98	**70.33**	73.98	**74.25**	77.01	**77.30**
FMix	69.96	**70.26**	74.08	**74.34**	77.19	**77.38**
ResizeMix	69.50	**69.90**	73.88	**74.00**	77.42	**77.65**
Avg. Gain		+0.32		+0.24		+0.25

ConvNets on ImageNet

	DeiT-S		Swin-T	
Methods	MCE	**DM(CE)**	MCE	**DM(CE)**
DeiT	79.80	**80.37**	81.28	**81.49**
Mixup	79.65	**80.04**	80.71	**80.97**
CutMix	79.78	**80.20**	80.83	**81.05**
FMix	79.41	**79.89**	80.37	**80.54**
ResizeMix	79.93	**80.03**	80.94	**81.01**
Avg. Gain		+0.39		+0.19

ViTs on ImageNet

Semi-supervised Learning

Methods	Losses	CIFAR-10 250	CIFAR-10 4000	CIFAR-100 400	CIFAR-100 2500	CIFAR-100 10000
Pseudo-Labeling	CE	53.51±2.20	84.92±0.19	12.55±0.85	42.26±0.28	63.45±0.24
MixMatch	CE+Con	86.37±0.59	93.34±0.26	32.41±0.66	60.24±0.48	72.22±0.29
ReMixMatch	CE+Con+Rot	93.70±0.05	95.16±0.01	57.15±1.05	73.87±0.35	79.08±0.27
MixMatch+DM	CE+Con+DM	89.16±0.71	95.15±0.68	35.72±0.53	62.51±0.37	74.70±0.28
UDA	CE+Con	94.84±0.06	95.71±0.07	53.61±1.59	72.27±0.21	77.51±0.23
FixMatch	CE+Con	95.14±0.05	95.79±0.08	53.58±0.82	71.97±0.16	77.80±0.12
FlexMatch	CE+Con+CPL	95.02±0.09	95.81±0.01	**60.06**±1.62	73.51±0.20	78.10±0.15
FixMatch+Mixup	CE+Con+MCE	95.05±0.23	95.83±0.19	50.61±0.73	72.16±0.18	78.75±0.14
FixMatch+DM	CE+Con+DM	**95.23**±0.09	**95.87**±0.11	59.75±0.95	**74.12**±0.23	**79.58**±0.17
Average Gain		+1.44	+0.95	+4.74	+2.30	+2.13

Training from scratch.

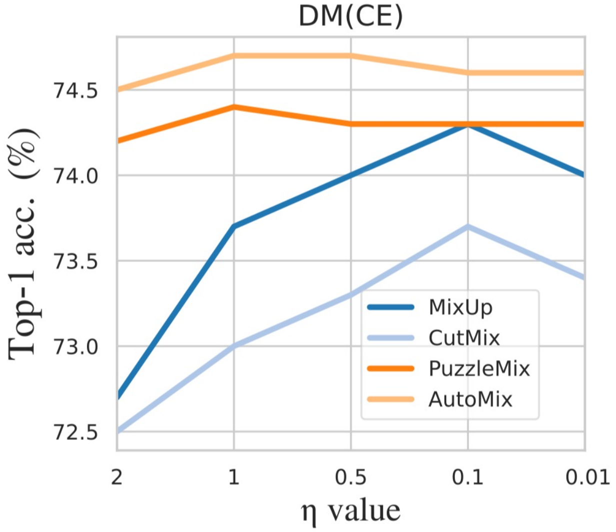
- Ablation Study

Methods	15%	30%	50%	100%
Self-Tuning	57.82	69.12	73.59	75.08
+MCE	63.36	72.81	75.73	76.67
+MCE+AS($\lambda \geq 0.5$)	59.04	69.67	74.89	75.96
+MCE+AS($\lambda \leq 0.5$)	62.97	72.46	75.40	76.34
+DM(CE)+AS($\lambda \leq 0.5$)	**66.17**	**74.25**	**77.68**	**78.52**

Components are effective.

Methods	CUB-200 15%	CUB-200 30%	CUB-200 50%	FGVC-Aircraft 15%	FGVC-Aircraft 30%	FGVC-Aircraft 50%	Stanford-Cars 15%	Stanford-Cars 30%	Stanford-Cars 50%
Fine-Tuning	45.25±0.12	59.68±0.21	70.12±0.29	39.57±0.20	57.46±0.12	67.93±0.28	36.77±0.12	60.63±0.18	75.10±0.21
+DM	50.04±0.17	61.39±0.24	71.87±0.23	43.15±0.22	61.02±0.15	70.38±0.18	41.30±0.16	62.65±0.21	77.19±0.19
BSS	47.74±0.23	63.38±0.29	72.56±0.17	40.41±0.12	59.23±0.31	69.19±0.13	40.57±0.12	64.13±0.18	76.78±0.21
Co-Tuning	52.58±0.53	66.47±0.17	74.64±0.36	44.09±0.67	61.65±0.32	72.73±0.08	46.02±0.18	69.09±0.10	80.66±0.25
+DM	**54.96**±0.65	**68.25**±0.21	**75.72**±0.37	**49.27**±0.83	**65.60**±0.41	**74.89**±0.17	**51.78**±0.34	**74.15**±0.29	**83.02**±0.26
Self-Tuning	64.17±0.47	75.13±0.35	80.22±0.36	64.11±0.32	76.03±0.25	81.22±0.29	72.50±0.45	83.58±0.28	88.11±0.29
+Mixup	62.38±0.32	74.65±0.24	81.46±0.27	59.38±0.31	74.65±0.26	81.46±0.27	70.31±0.27	83.63±0.23	88.66±0.21
+DM	**73.06**±0.38	**79.50**±0.35	**82.64**±0.24	**67.57**±0.27	**80.71**±0.25	**84.82**±0.26	**81.69**±0.23	**89.22**±0.21	**91.26**±0.19
Avg. Gain	+5.95	+2.77	+1.34	+5.65	+4.52	+2.65	+7.22	+4.22	+2.35

Fine-tuning.



DM is robust to hyper-parameters

西湖大學
WESTLAKE UNIVERSITY

Thank you!

Code: OpenMixup

Awesome-Mixup

Homepage

lisiyuan@westlake.edu.cn