Data Augmentation via Mixing Images

Dominik Lewy

Data Augmentation via Mixing Images

$$\tilde{x} = \mathbf{B} \odot x_1 + (\mathbf{1} - \mathbf{B}) \odot x_2$$

$$\tilde{y} = \lambda y_1 + (1 - \lambda) y_2$$

Initial approaches
- 2 Smart Augmentation
- 3 Pairing Samples

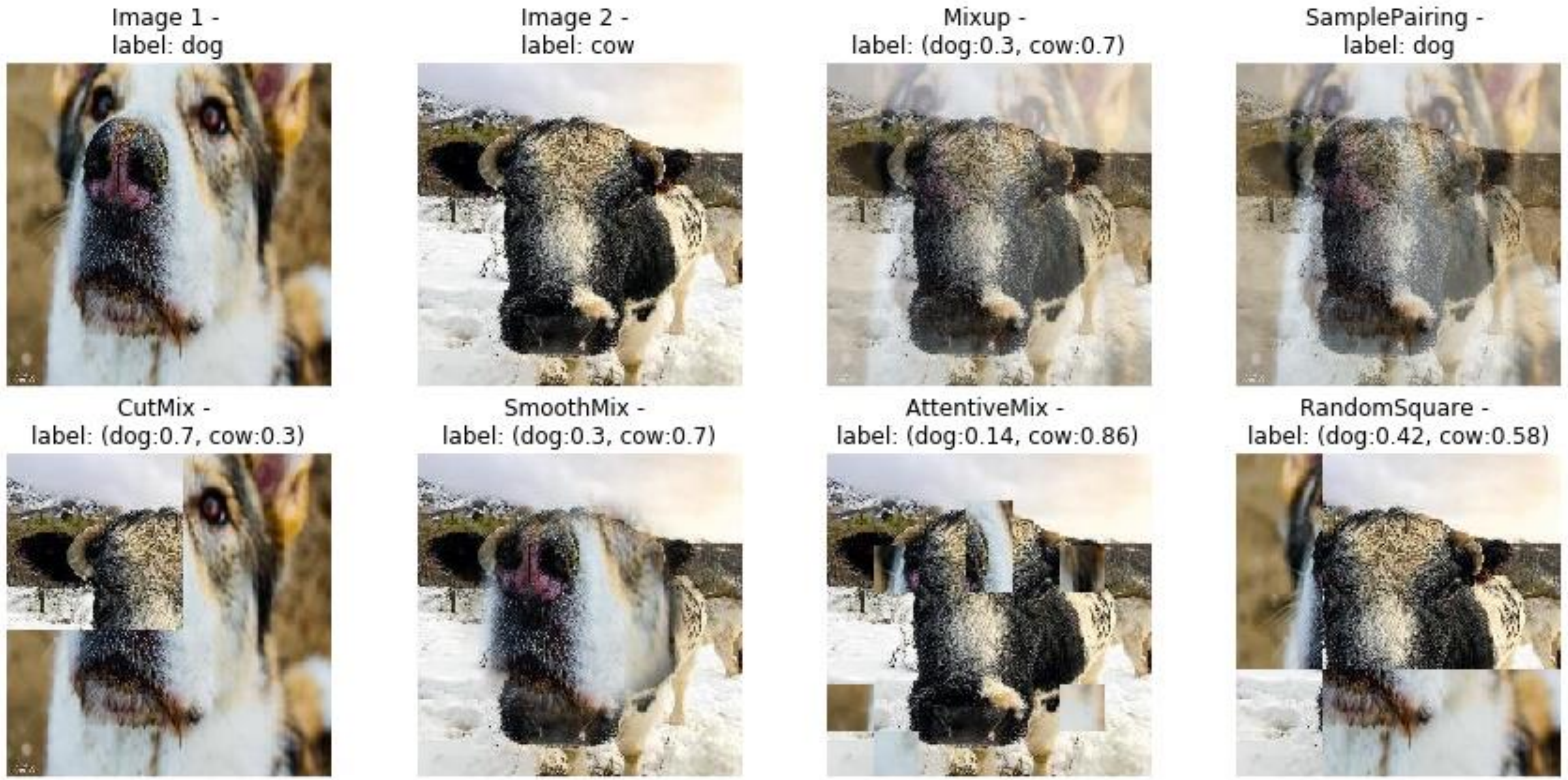
Canonical Paper
- 1 Mixup

Mixup improvements
- 5 Manifold Mixup
- 4 AdaMixup

Questioning linear mixing
- 6 Mixed Examples
- 7 Cutmix
- 10 RICAP

Cutmix improvements
- 8 SmoothMix
- 9 Attentive CutMix

Data Augmentation via Mixing Images

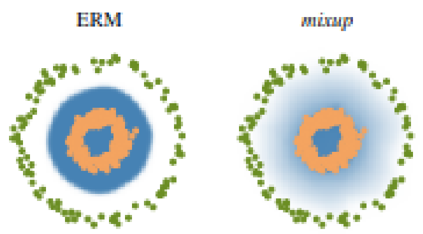


MIXUP

MIXUP - Zhang, Hongyi & Cisse, Moustapha & Dauphin, Yann & Lopez-Paz, David. (2017). mixup: Beyond Empirical Risk Minimization.

```
# y1, y2 should be one-hot vectors
for (x1, y1), (x2, y2) in zip(loader1, loader2):
    lam = numpy.random.beta(alpha, alpha)
    x = Variable(lam * x1 + (1. - lam) * x2)
    y = Variable(lam * y1 + (1. - lam) * y2)
    optimizer.zero_grad()
    loss(net(x), y).backward()
    optimizer.step()
```

(a) One epoch of *mixup* training in PyTorch.

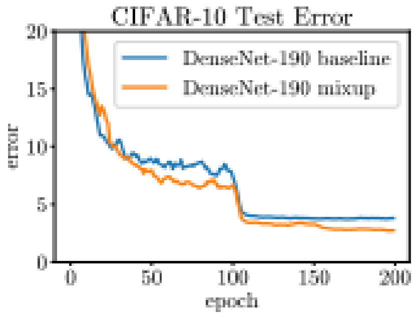


(b) Effect of *mixup* ($\alpha = 1$) on a toy problem. Green: Class 0. Orange: Class 1. Blue shading indicates $p(y = 1|x)$.

Figure 1: Illustration of *mixup*, which converges to ERM as $\alpha \to 0$.

Dataset	Model	ERM	*mixup*
CIFAR-10	PreAct ResNet-18	5.6	**4.2**
	WideResNet-28-10	3.8	**2.7**
	DenseNet-BC-190	3.7	**2.7**
CIFAR-100	PreAct ResNet-18	25.6	**21.1**
	WideResNet-28-10	19.4	**17.5**
	DenseNet-BC-190	19.0	**16.8**

(a) Test errors for the CIFAR experiments.



(b) Test error evolution for the best ERM and *mixup* models.

Figure 3: Test errors for ERM and *mixup* on the CIFAR experiments.

MIXUP

MIXUP - Zhang, Hongyi & Cisse, Moustapha & Dauphin, Yann & Lopez-Paz, David. (2017). mixup: Beyond Empirical Risk Minimization.

Label corruption	Method	Test error		Training error	
		Best	Last	Real	Corrupted
20%	ERM	12.7	16.6	0.05	0.28
	ERM + dropout ($p = 0.7$)	8.8	10.4	5.26	83.55
	mixup ($\alpha = 8$)	**5.9**	6.4	2.27	86.32
	mixup + dropout ($\alpha = 4, p = 0.1$)	6.2	**6.2**	1.92	85.02
50%	ERM	18.8	44.6	0.26	0.64
	ERM + dropout ($p = 0.8$)	14.1	15.5	12.71	86.98
	mixup ($\alpha = 32$)	11.3	12.7	5.84	85.71
	mixup + dropout ($\alpha = 8, p = 0.3$)	**10.9**	**10.9**	7.56	87.90
80%	ERM	36.5	73.9	0.62	0.83
	ERM + dropout ($p = 0.8$)	30.9	35.1	29.84	86.37
	mixup ($\alpha = 32$)	25.3	30.9	18.92	85.44
	mixup + dropout ($\alpha = 8, p = 0.3$)	**24.0**	**24.8**	19.70	87.67

Table 2: Results on the corrupted label experiments for the best models.

MIXUP

MIXUP - Zhang, Hongyi & Cisse, Moustapha & Dauphin, Yann & Lopez-Paz, David. (2017). mixup: Beyond Empirical Risk Minimization.

Method	Specification	Modified		Weight decay	
		Input	Target	10^{-4}	5×10^{-4}
ERM		✗	✗	5.53	5.18
mixup	AC + RP	✓	✓	**4.24**	4.68
	AC + KNN	✓	✓	4.98	5.26
mix labels and latent representations (AC + RP)	Layer 1	✓	✓	4.44	**4.51**
	Layer 2	✓	✓	4.56	4.61
	Layer 3	✓	✓	5.39	5.55
	Layer 4	✓	✓	5.95	5.43
	Layer 5	✓	✓	5.39	5.15

Table 5: Results of the ablation studies on the CIFAR-10 dataset. Reported are the median test errors of the last 10 epochs. A tick (✓) means the component is different from standard ERM training, whereas a cross (✗) means it follows the standard training practice. AC: mix between all classes. SC: mix within the same class. RP: mix between random pairs. KNN: mix between k-nearest neighbors (k=200). Please refer to the text for details about the experiments and interpretations.

SMART AUGMENTATION

SMART AUGMENTATION - Lemley, Joseph & Bazrafkan, Shabab & Corcoran, Peter. (2017). Smart Augmentation - Learning an Optimal Data Augmentation Strategy. IEEE Access. PP. 10.1109/ACCESS.2017.2696121.

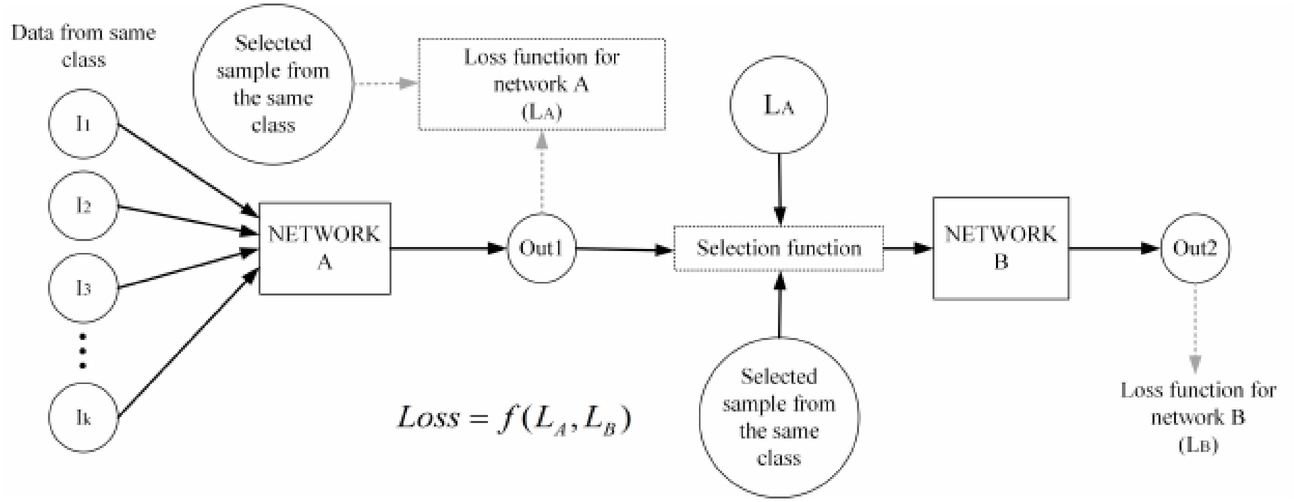


FIGURE 2. Diagram illustrating the reduced smart augmentation concept with just one network A.

FIGURE 3. The image on the left is created by a learned combination of the two images on the right. This type of image transformation helped increase the accuracy of network B. The image was not produced to be an ideal approximation of a face but instead, contains features that helped network B better generalize the concept of gender which is the task it was trained for.

PAIRING SAMPLES

PAIRING SAMPLES - Inoue, Hiroshi. (2018). Data Augmentation by Pairing Samples for Images Classification.

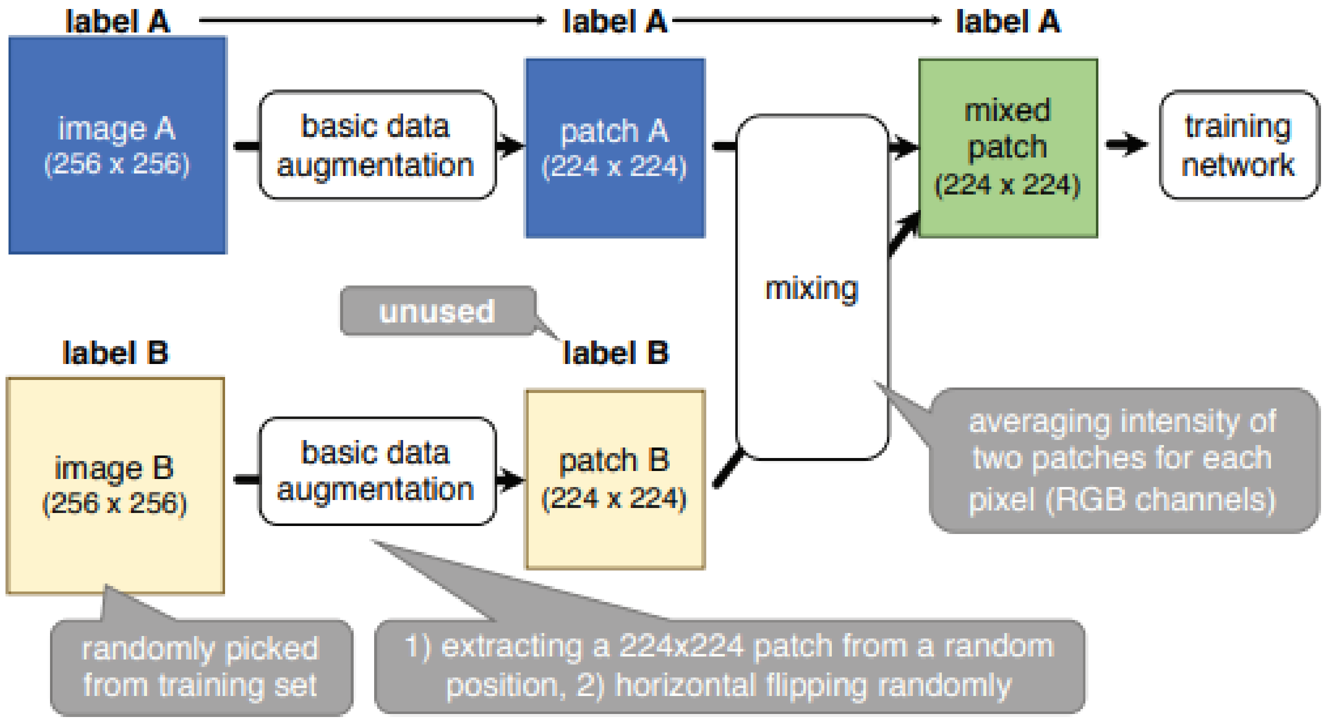


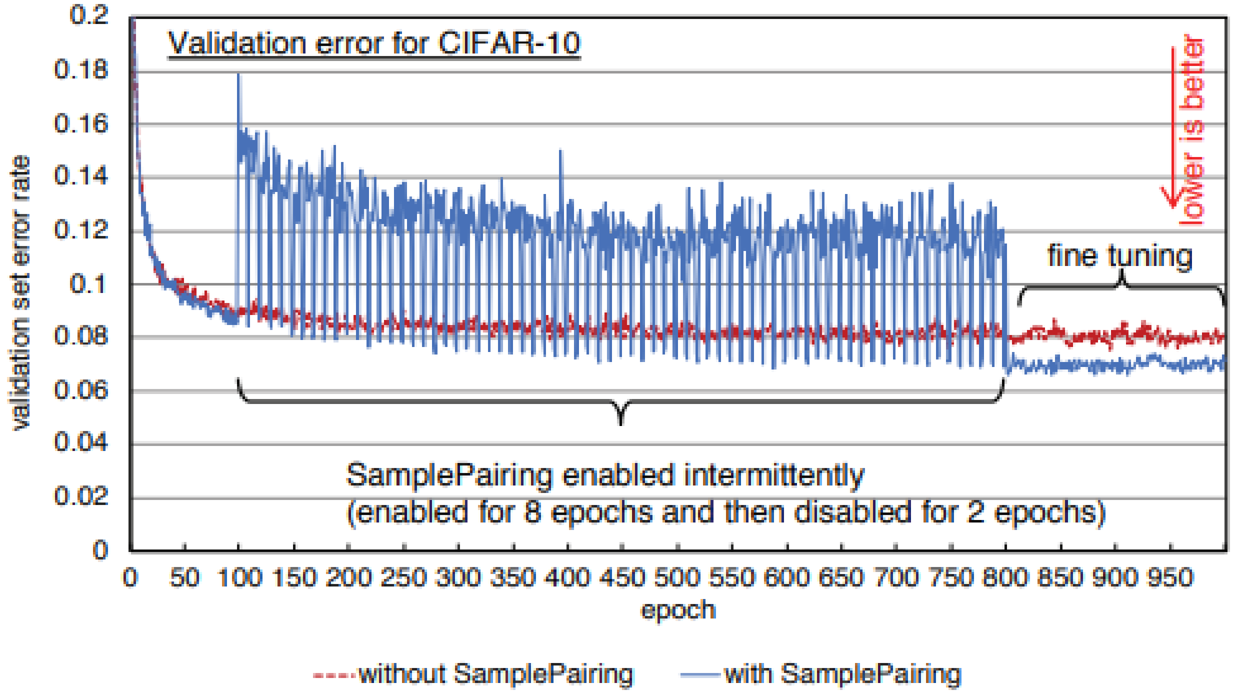
Figure 1. Overview of our SamplePairing data augmentation for ILSVRC dataset. For other datasets, the size of the original images is 32×32 and the size of the extracted patches is 28×28.

PAIRING SAMPLES

PAIRING SAMPLES - Inoue, Hiroshi. (2018). Data Augmentation by Pairing Samples for Images Classification.

Table 1. Training and validation sets error rates with and without our SamplePairing data augmentation.

Dataset		training error		validation error		
		without SamplePairing	with SamplePairing	without SamplePairing	with SamplePairing	reduction in error rate
CIFAR-10		0.55%	1.25%	8.22%	6.93%	-15.68%
CIFAR-100		5.78%	10.56%	30.5%	27.9%	-8.58%
SVHN		0.84%	2.08%	4.28%	4.15%	-3.05%
ILSVRC with 100 classes	top-1 error	0.95%	3.21%	26.21%	21.02%	-19.82%
	top-5 error	-	-	8.58%	6.11%	-28.74%
ILSVRC with 1000 classes	top-1 error	1.52%	17.58%	33.51%	29.01%	-13.46%
	top-5 error	-	-	13.15%	11.36%	-13.55%



- Complex training procedure

MANIFLOD MIXUP

MANIFLOD MIXUP - Verma, Vikas & Lamb, Alex & Beckham, Christopher & Najafi, Amir & Courville, Aaron & Mitliagkas, Ioannis & Bengio, Y.. (2018). Manifold Mixup: Learning Better Representations by Interpolating Hidden States.

Steps:

1. Select random layer K
2. Process two random minibatches of data until layer K
3. Perform mixup on the intermediate minibatches
4. Continue the forward pass with mixed sample
5. Output is used to compute loss and update al network parameters

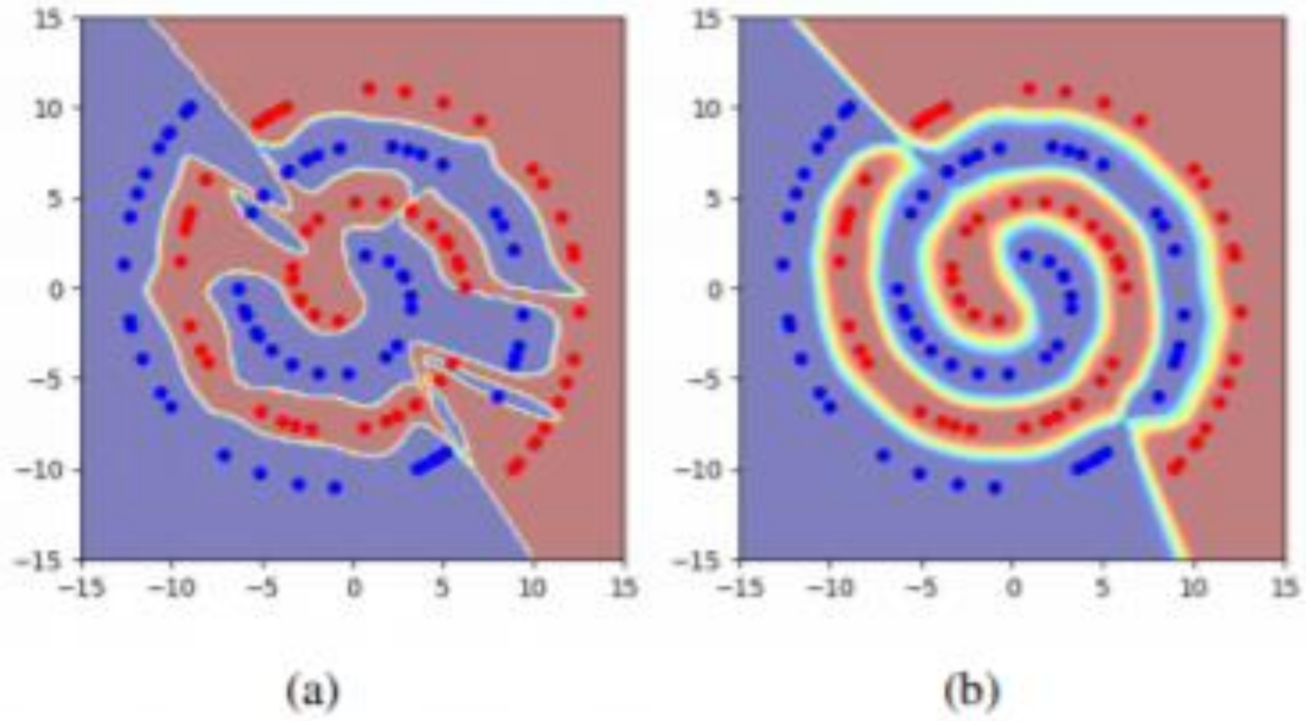


Figure 1: An experiment on a network trained on the 2D spiral dataset with a 2D bottleneck hidden representation in the middle of the network. Manifold mixup has three effects on learning when compared to vanilla training. First, it smoothens decision boundaries (from a. to b.). Second, it improves the arrangement of hidden representations and encourages broader regions of low-confidence predictions (from d. to e.). Black dots are the hidden representation of the inputs sampled uniformly from the range of the input space. Third, it flattens the representations (c. at layer 1, f. at layer 3). Figure 2 shows that these effects are not accomplished by other well-studied regularizers (input mixup, weight decay, dropout, batch normalization, and adding noise to the hidden representations).

MANIFLOD MIXUP

MANIFLOD MIXUP - Verma, Vikas & Lamb, Alex & Beckham, Christopher & Najafi, Amir & Courville, Aaron & Mitliagkas, Ioannis & Bengio, Y.. (2018). Manifold Mixup: Learning Better Representations by Interpolating Hidden States.

Table 1: Classification errors on (a) CIFAR-10 and (b) CIFAR-100. We include results from (Zhang et al., 2018)† and (Guo et al., 2016)‡. Standard deviations over five repetitions.

PreActResNet18	Test Error (%)	Test NLL
No Mixup	4.83 ± 0.066	0.190 ± 0.003
AdaMix‡	3.52	NA
Input Mixup†	4.20	NA
Input Mixup ($\alpha = 1$)	3.82 ± 0.048	0.186 ± 0.004
Manifold Mixup ($\alpha = 2$)	$\underline{2.95 \pm 0.046}$	$\underline{0.137 \pm 0.003}$
PreActResNet34		
No Mixup	4.64 ± 0.072	0.200 ± 0.002
Input Mixup ($\alpha = 1$)	2.88 ± 0.043	0.176 ± 0.002
Manifold Mixup ($\alpha = 2$)	$\underline{2.54 \pm 0.047}$	$\underline{0.118 \pm 0.002}$
Wide-Resnet-28-10		
No Mixup	3.99 ± 0.118	0.162 ± 0.004
Input Mixup ($\alpha = 1$)	2.92 ± 0.088	0.173 ± 0.001
Manifold Mixup ($\alpha = 2$)	$\underline{2.55 \pm 0.024}$	$\underline{0.111 \pm 0.001}$

(a) CIFAR-10

PreActResNet18	Test Error (%)	Test NLL
No Mixup	24.01 ± 0.376	1.189 ± 0.002
AdaMix‡	20.97	n/a
Input Mixup†	21.10	n/a
Input Mixup ($\alpha = 1$)	22.11 ± 0.424	1.055 ± 0.006
Manifold Mixup ($\alpha = 2$)	$\underline{20.34 \pm 0.525}$	$\underline{0.912 \pm 0.002}$
PreActResNet34		
No Mixup	23.55 ± 0.399	1.189 ± 0.002
Input Mixup ($\alpha = 1$)	20.53 ± 0.330	1.039 ± 0.045
Manifold Mixup ($\alpha = 2$)	$\underline{18.35 \pm 0.360}$	$\underline{0.877 \pm 0.053}$
Wide-Resnet-28-10		
No Mixup	21.72 ± 0.117	1.023 ± 0.004
Input Mixup ($\alpha = 1$)	18.89 ± 0.111	0.927 ± 0.031
Manifold Mixup ($\alpha = 2$)	$\underline{18.04 \pm 0.171}$	$\underline{0.809 \pm 0.005}$

(b) CIFAR-100

MANIFLOD MIXUP

MANIFLOD MIXUP - Verma, Vikas & Lamb, Alex & Beckham, Christopher & Najafi, Amir & Courville, Aaron & Mitliagkas, Ioannis & Bengio, Y.. (2018). Manifold Mixup: Learning Better Representations by Interpolating Hidden States.

Table 5: Test accuracy *Manifold Mixup* for different sets of eligible layers $\mathcal{S}$ on CIFAR.

$\mathcal{S}$	CIFAR-10	CIFAR-100
$\{0,1,2\}$	97.23	79.60
$\{0,1\}$	96.94	78.93
$\{0,1,2,3\}$	96.92	80.18
$\{1,2\}$	96.35	78.69
$\{0\}$	96.73	78.15
$\{1,2,3\}$	96.51	79.31
$\{1\}$	96.10	78.72
$\{2,3\}$	95.32	76.46
$\{2\}$	95.19	76.50
$\{\}$	95.27	76.40

Table 6: Test accuracy (%) of Input Mixup and *Manifold Mixup* for different α on CIFAR-10.

α	Input Mixup	*Manifold Mixup*
0.5	96.68	96.76
1.0	96.75	97.00
1.2	96.72	97.03
1.5	96.84	97.10
1.8	96.80	97.15
2.0	96.73	97.23

MANIFLOD MIXUP

MANIFLOD MIXUP - Verma, Vikas & Lamb, Alex & Beckham, Christopher & Najafi, Amir & Courville, Aaron & Mitliagkas, Ioannis & Bengio, Y.. (2018). Manifold Mixup: Learning Better Representations by Interpolating Hidden States.

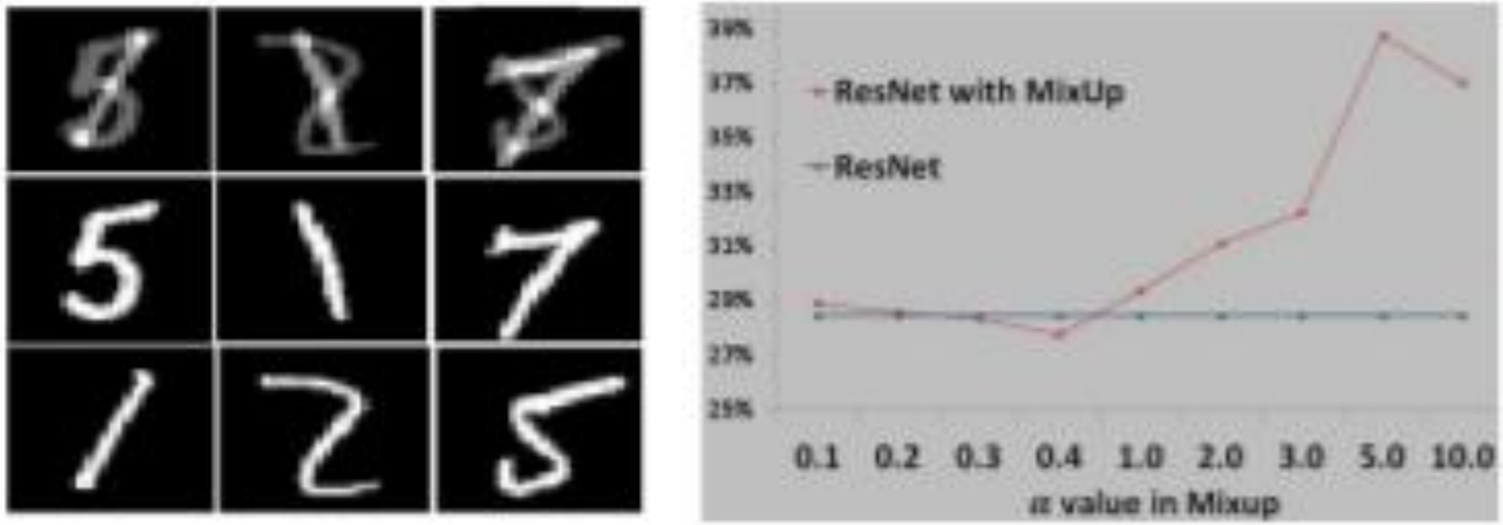
Table 7: Test accuracy on white-box FGSM adversarial examples on CIFAR-10 and CIFAR-100 (using a PreActResNet18 model) and SVHN (using a WideResNet20-10 model). We include the results of (Madry et al., 2018)†.

CIFAR-10	FGSM
No Mixup	36.32
Input Mixup ($\alpha = 1$)	71.51
Manifold Mixup ($\alpha = 2$)	<u>77.50</u>
PGD training (7-steps)†	56.10
CIFAR-100	**FGSM**
Input Mixup ($\alpha = 1$)	40.7
Manifold Mixup ($\alpha = 2$)	44.96
SVHN	**FGSM**
No Mixup	21.49
Input Mixup ($\alpha = 1$)	56.98
Manifold Mixup ($\alpha = 2$)	65.91
PGD training (7-steps)†	<u>72.80</u>

ADAMIXUP

ADAMIXUP - Guo, Hongyu & Mao, Yongyi & Zhang, Richong. (2019). MixUp as Locally Linear Out-of-Manifold Regularization. Proceedings of the AAAI Conference on Artificial Intelligence. 33. 3714-3722. 10.1609/aaai.v33i01.33013714.

Figure 1: Left: Linearly interpolated images (top row) from the original images (bottom two rows). Right: Performance of MixUp on a reduced Cifar100 data set (reduced to containing 20% of data samples) vs various values of α.



ADAMIXUP

ADAMIXUP - Guo, Hongyu & Mao, Yongyi & Zhang, Richong. (2019). MixUp as Locally Linear Out-of-Manifold Regularization. Proceedings of the AAAI Conference on Artificial Intelligence. 33. 3714-3722. 10.1609/aaai.v33i01.33013714.

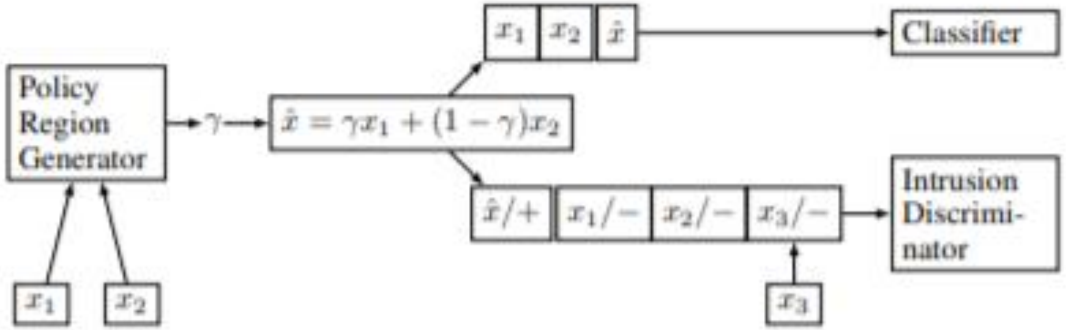


Figure 3: Fold-2 AdaMixUp for a single triplet (x_1, x_2, x_3). Each batch is implemented to contain multiple such triplets. "+" and "-" indicate positive and negative examples, respectively.

$$\mathcal{L}_{\text{total}} := \mathcal{L}_{\mathcal{D}}(H) + \mathcal{L}_{\mathcal{D}'}(H, \{\pi_k\}) + \mathcal{L}_{\text{intr}}(\{\pi_k\}, \varphi)$$

ADAMIXUP

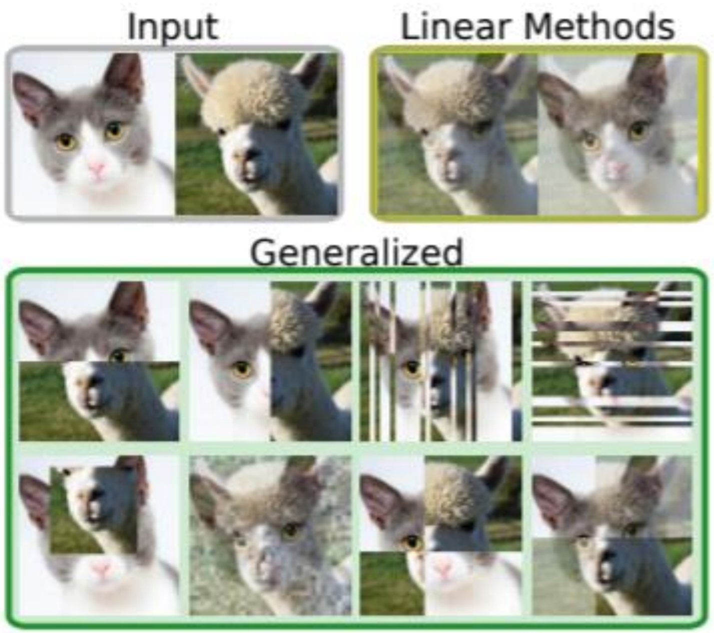
ADAMIXUP - Guo, Hongyu & Mao, Yongyi & Zhang, Richong. (2019). MixUp as Locally Linear Out-of-Manifold Regularization. Proceedings of the AAAI Conference on Artificial Intelligence. 33. 3714-3722. 10.1609/aaai.v33i01.33013714.

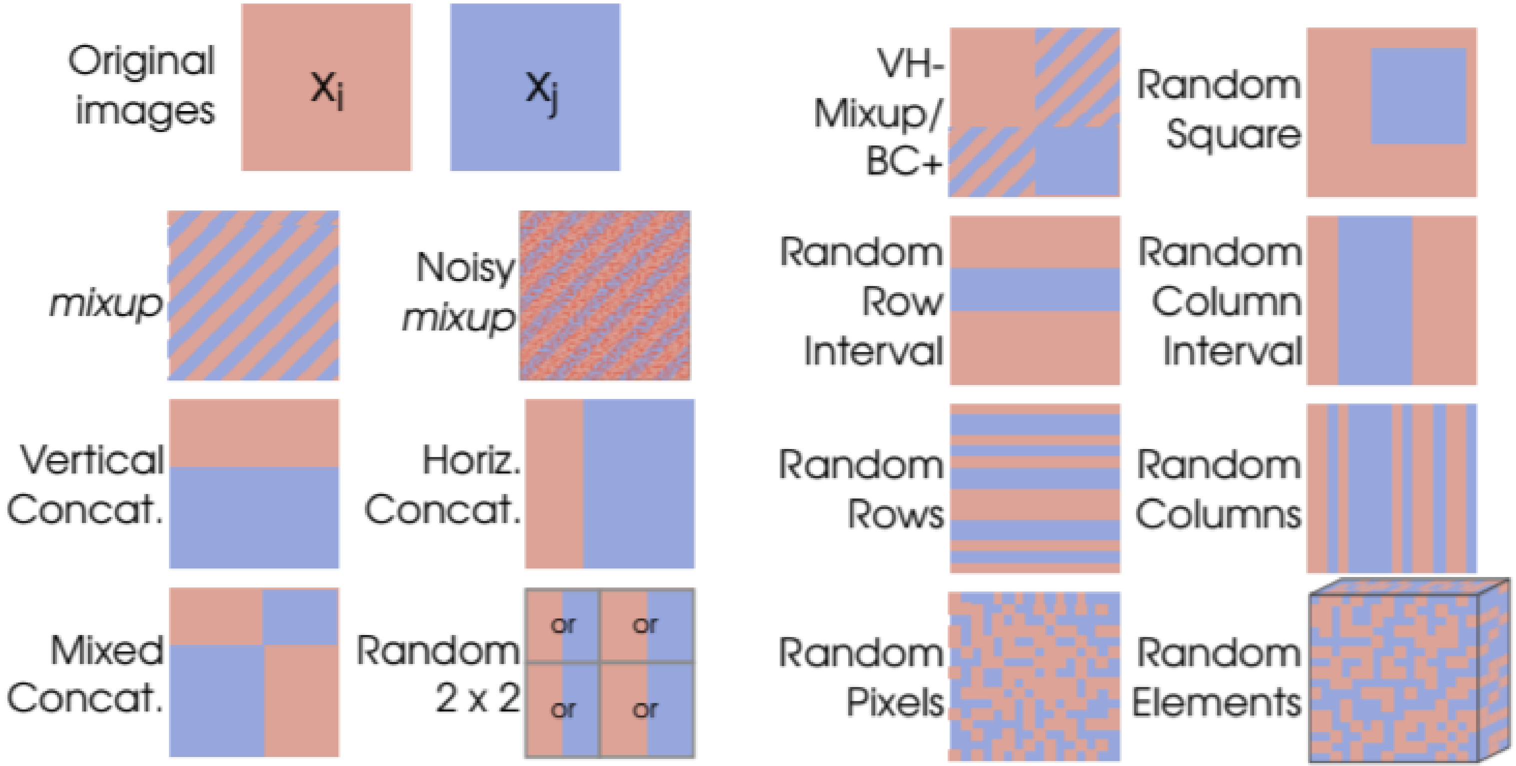
Data Set	Baseline	MixUp	Ada MixUp	Relative Impro. (%)
mnist	0.52	<u>0.57</u>	0.49	5.77
fashion	7.37	6.92	6.21	15.74
svhn	4.50	3.80	3.12	30.67
cifar10	5.53	4.24	3.52	36.35
cifar100	25.6	21.14	20.97	18.09
cifar10-S	7.68	<u>7.88</u>	6.85	10.81
cifar100-S	28.47	<u>29.39</u>	26.72	6.15
ImageNet-R top1	53.00	<u>54.89</u>	49.17	7.22
ImageNet-R top5	29.41	<u>31.02</u>	25.78	12.34

Table 1: Error rates (%) obtained by the testing methods.

GENERALIZED MIXED-EXAMPLE

GENERALIZED MIXED-EXAMPLE – Summers, Cecilia & Dinneen, Michael. (2018). Improved Mixed-Example Data Augmentation.





GENERALIZED MIXED-EXAMPLE

GENERALIZED MIXED-EXAMPLE – Summers, Cecilia & Dinneen, Michael. (2018). Improved Mixed-Example Data Augmentation.

	Method	Error (%)
CIFAR-10	ResNet-18	5.4
	mixup[27]	4.3
	BC+[24]	4.2
	Rand. Elems.	6.2
	Rand. Pixels	5.7
	Rand. Col. Int.	5.1
	Rand. Cols	4.8
	Horiz. Concat.	4.7
	Rand. Rows	4.6
	Noisy Mixup	4.5
	Rand. Row. Int.	4.5
	Vert. Concat.	4.4
	Mixed. Concat.	4.4
	Rand. Square.	4.3
	Rand. 2×2	*4.1*
	VH-BC+	*3.8*
	VH-Mixup	*3.8*

	Method	Error (%)
CIFAR-100	ResNet-18	23.6
	mixup[27]	21.3
	BC+[24]	21.1
	Rand. Elems.	24.2
	Rand. Pixels	24.0
	Rand. Cols	22.4
	Noisy Mixup	21.8
	Horiz. Concat.	21.7
	Rand. Col. Int.	21.4
	Rand. Square.	*20.9*
	Rand. Rows	*20.9*
	Mixed. Concat.	*20.9*
	Vert. Concat.	*20.8*
	Rand. 2×2	*20.4*
	Rand. Row. Int.	*20.1*
	VH-BC+	*19.9*
	VH-Mixup	*19.7*

CUTMIX

CUTMIX – Yun, Sangdoo & Han, Dongyoon & Chun, Sanghyuk & Oh, Seong Joon & Yoo, Youngjoon & Choe, Junsuk. (2019). CutMix: Regularization Strategy to Train Strong Classifiers With Localizable Features. 6022-6031. 10.1109/ICCV.2019.00612.

	ResNet-50	Mixup [48]	Cutout [3]	CutMix
Image				
Label	Dog 1.0	Dog 0.5 Cat 0.5	Dog 1.0	Dog 0.6 Cat 0.4
ImageNet Cls (%)	76.3 (+0.0)	77.4 (+1.1)	77.1 (+0.8)	**78.6** (+2.3)
ImageNet Loc (%)	46.3 (+0.0)	45.8 (-0.5)	46.7 (+0.4)	**47.3** (+1.0)
Pascal VOC Det (mAP)	75.6 (+0.0)	73.9 (-1.7)	75.1 (-0.5)	**76.7** (+1.1)

	Mixup	Cutout	CutMix
Usage of full image region	✔	✘	✔
Regional dropout	✘	✔	✔
Mixed image & label	✔	✘	✔

CUTMIX

CUTMIX – Yun, Sangdoo & Han, Dongyoon & Chun, Sanghyuk & Oh, Seong Joon & Yoo, Youngjoon & Choe, Junsuk. (2019). CutMix: Regularization Strategy to Train Strong Classifiers With Localizable Features. 6022-6031. 10.1109/ICCV.2019.00612.

PyramidNet-200 ($\tilde{\alpha}$=240) (# params: 26.8 M)	Top-1 Err (%)	Top-5 Err (%)
Baseline	16.45	3.69
+ StochDepth [17]	15.86	3.33
+ Label smoothing (ϵ=0.1) [38]	16.73	3.37
+ Cutout [3]	16.53	3.65
+ Cutout + Label smoothing (ϵ=0.1)	15.61	3.88
+ DropBlock [8]	15.73	3.26
+ DropBlock + Label smoothing (ϵ=0.1)	15.16	3.86
+ Mixup (α=0.5) [48]	15.78	4.04
+ Mixup (α=1.0) [48]	15.63	3.99
+ Manifold Mixup (α=1.0) [42]	16.14	4.07
+ Cutout + Mixup (α=1.0)	15.46	3.42
+ Cutout + Manifold Mixup (α=1.0)	15.09	3.35
+ ShakeDrop [46]	15.08	2.72
+ CutMix	14.47	2.97
+ CutMix + ShakeDrop [46]	**13.81**	**2.29**

Table 5: Comparison of state-of-the-art regularization methods on CIFAR-100.

Model	# Params	Top-1 Err (%)	Top-5 Err (%)
PyramidNet-110 ($\tilde{\alpha}$ = 64) [11]	1.7 M	19.85	4.66
PyramidNet-110 + CutMix	1.7 M	**17.97**	**3.83**
ResNet-110 [12]	1.1 M	23.14	5.95
ResNet-110 + CutMix	1.1 M	**20.11**	**4.43**

Table 6: Impact of CutMix on lighter architectures on CIFAR-100.

PyramidNet-200 ($\tilde{\alpha}$=240)	Top-1 Error (%)
Baseline	3.85
+ Cutout	3.10
+ Mixup (α=1.0)	3.09
+ Manifold Mixup (α=1.0)	3.15
+ CutMix	**2.88**

Table 7: Impact of CutMix on CIFAR-10.

CUTMIX

CUTMIX – Yun, Sangdoo & Han, Dongyoon & Chun, Sanghyuk & Oh, Seong Joon & Yoo, Youngjoon & Choe, Junsuk. (2019). CutMix: Regularization Strategy to Train Strong Classifiers With Localizable Features. 6022-6031. 10.1109/ICCV.2019.00612.

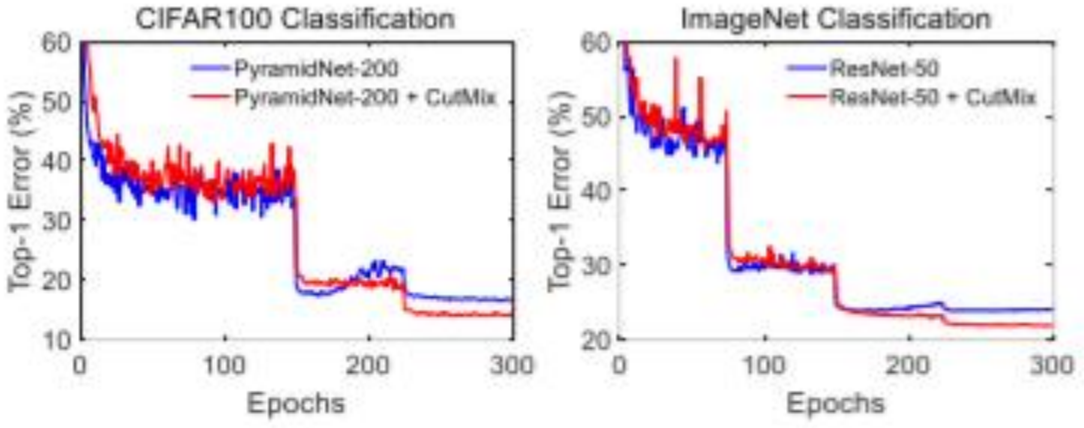
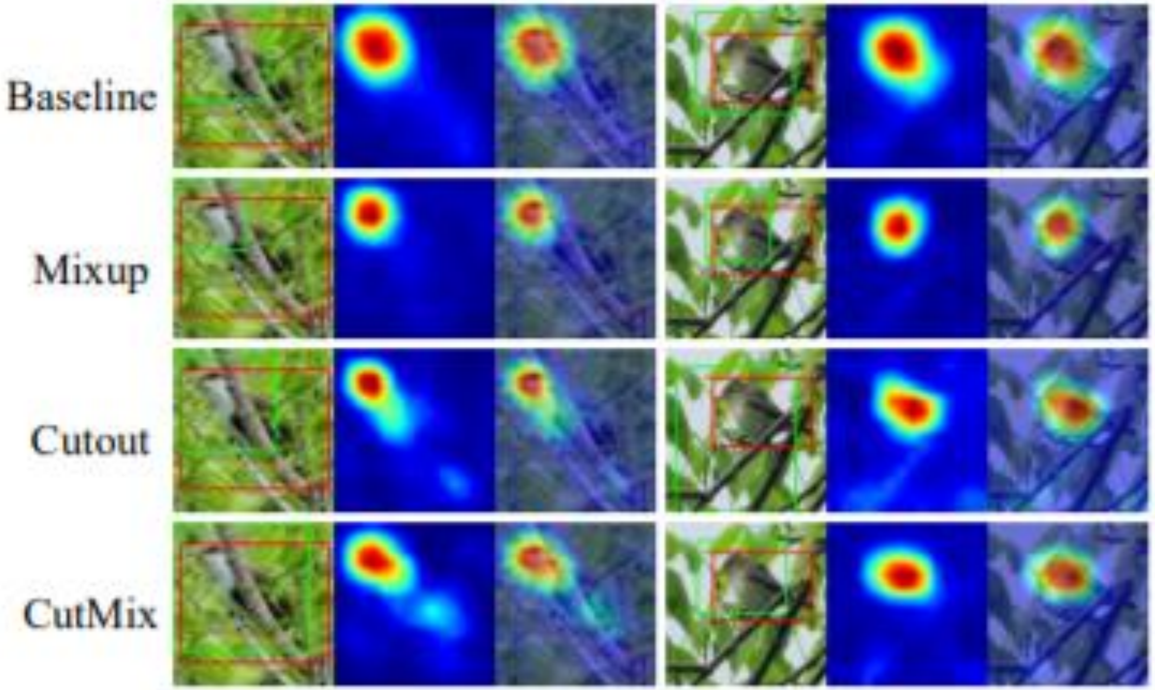


Figure 2: Top-1 test error plot for CIFAR100 (left) and ImageNet (right) classification. Cutmix achieves lower test errors than the baseline at the end of training.

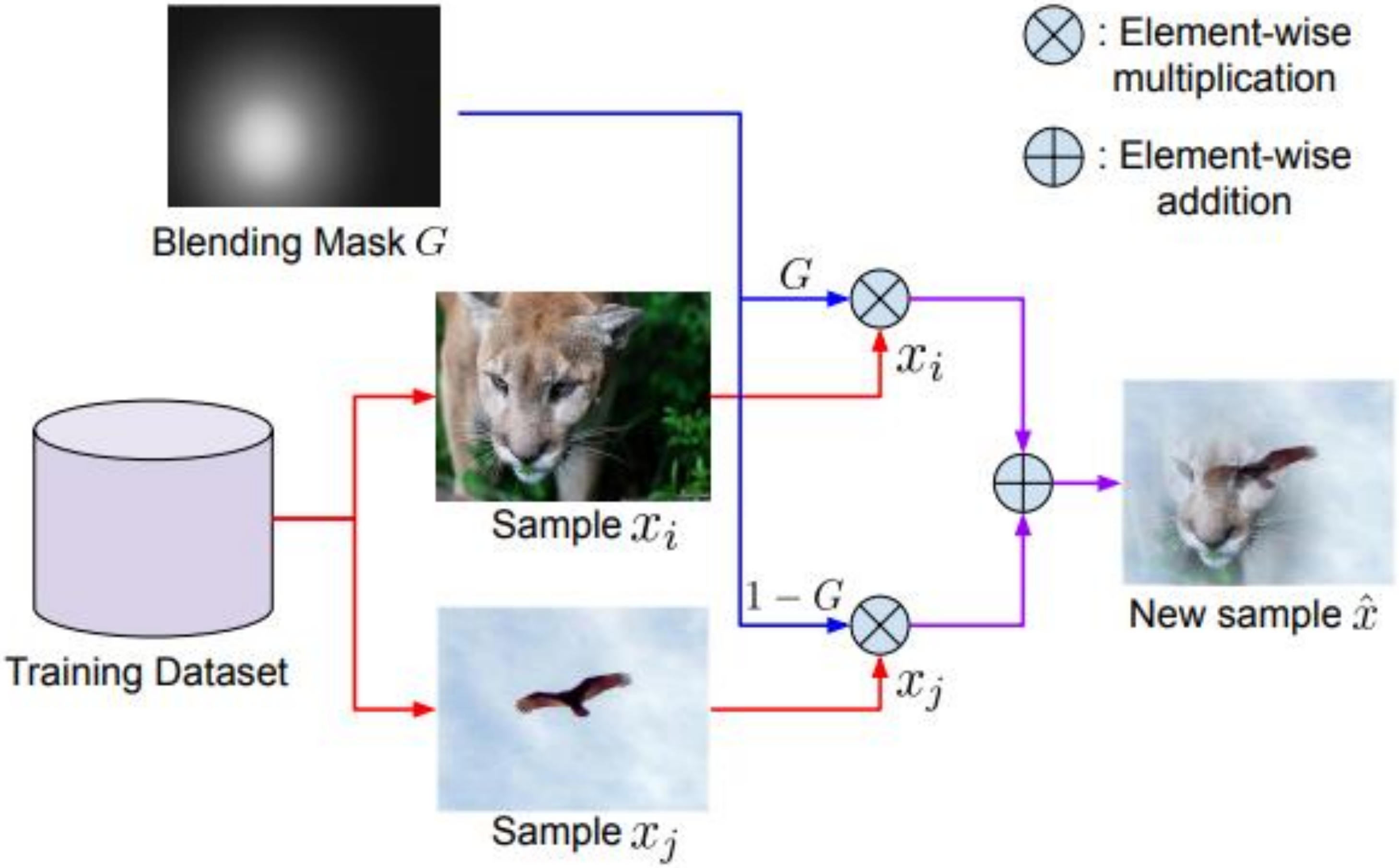


	Baseline	Mixup	Cutout	CutMix
Top-1 Acc (%)	8.2	24.4	11.5	**31.0**

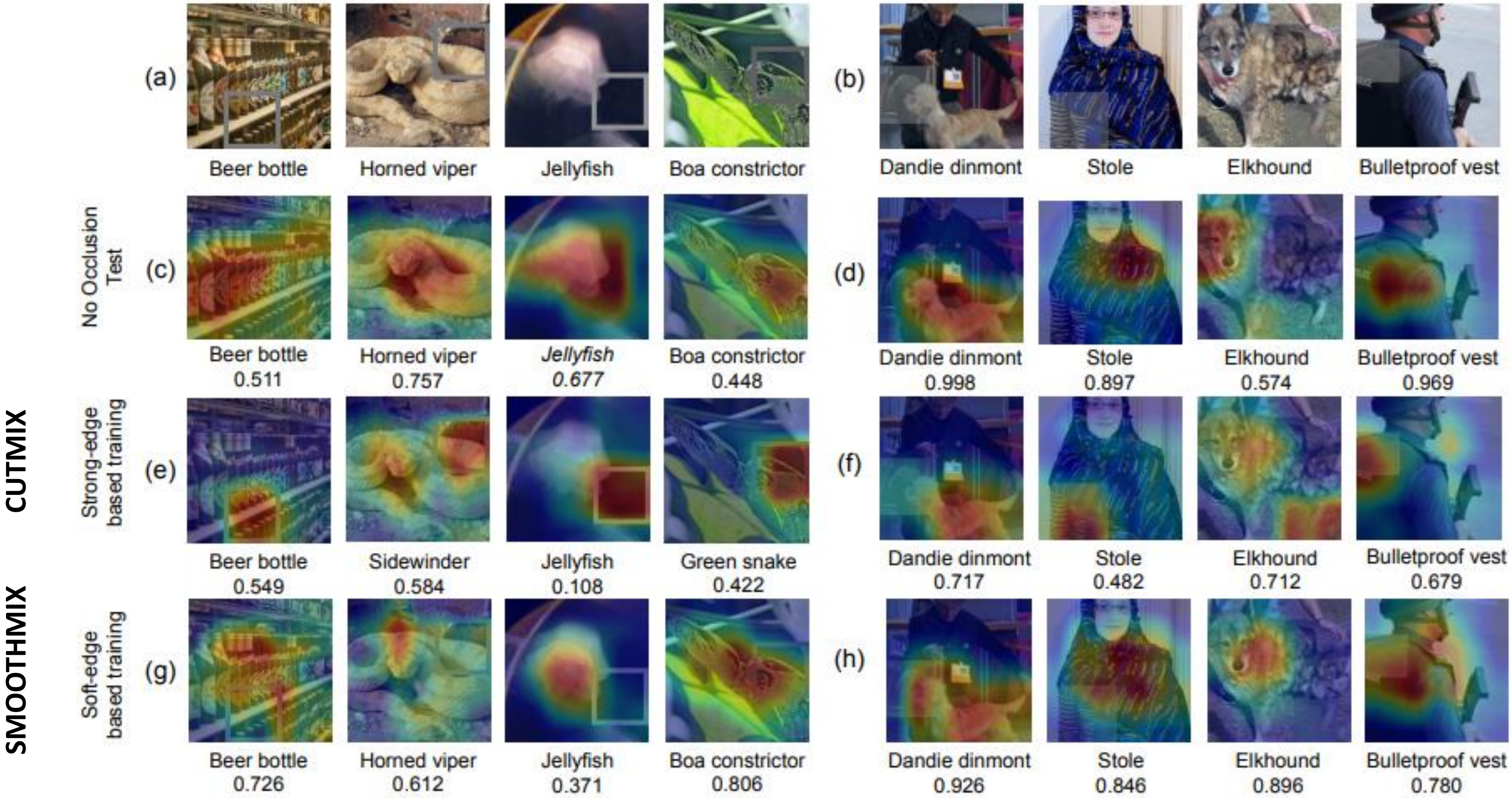
Table 11: Top-1 accuracy after FGSM white-box attack on ImageNet validation set.

SMOOTHMIX

SMOOTHMIX – Jin Ha Lee, Muhammad Zaigham Zaheer, Marcella Astrid, Seung-Ik Lee: SmoothMix: a Simple Yet Effective Data Augmentation to Train Robust Classifiers. CVPR Workshops 2020: 3264-3274



SMOOTHMIX – strong edge problem



SMOOTHMIX

SMOOTHMIX – Jin Ha Lee, Muhammad Zaigham Zaheer, Marcella Astrid, Seung-Ik Lee: SmoothMix: a Simple Yet Effective Data Augmentation to Train Robust Classifiers. CVPR Workshops 2020: 3264-3274

Model	Top-1 ERR(%)	Top-5 ERR(%)
Baseline(PyramidNet-200) [18]	16.45	3.69
+ Stochdepth [28]	16.73	3.37
+ Cutout [10]	16.53	3.65
+ DropBlock [15]	15.73	3.26
+ Mixup [58]	15.63	3.99
+ Manifold Mixup [51]	16.14	4.07
+ Shakedrop [55]	15.08	**2.72**
+ Cutmix [56]	**14.47**	2.97
+ $SmoothMix_S$	14.74	3.3
+ $SmoothMix_C$	**14.47**	2.99

Table 1. Image classification results on CIFAR-100 dataset. Best and second-best are highlighted as bold and italic respectively.

Model	Top-1 ERR(%)
Baseline(PyramidNet-200 [18])	3.85
+ Cutout [10]	3.1
+ Mixup [58]	3.09
+ Manifold mixup [51]	3.15
+ Cutmix [56]	**2.88**
+ $SmoothMix_C$	2.98

Table 2. Image classification results on CIFAR-10 dataset.

Model	Top-1 ERR(%)	Top-5 ERR(%)
Baseline(Resnet-50) [22]	23.68	7.05
+ Cutout [10]	22.93	6.66
+ Hide-and-Seek [45]	22.8	x
+ StochDepth [28]	22.46	6.27
+ Mixup [58]	22.58	6.4
+ Manifold Mixup [51]	22.5	6.21
+ AutoAugment [7]	22.4	x
+ AugMix [25]	22.47	6.06
+ Drop block [15]	21.87	5.98
+ Cutmix [56]	**21.4**	**5.92**
+ $SmoothMix_C$	22.34	6.37

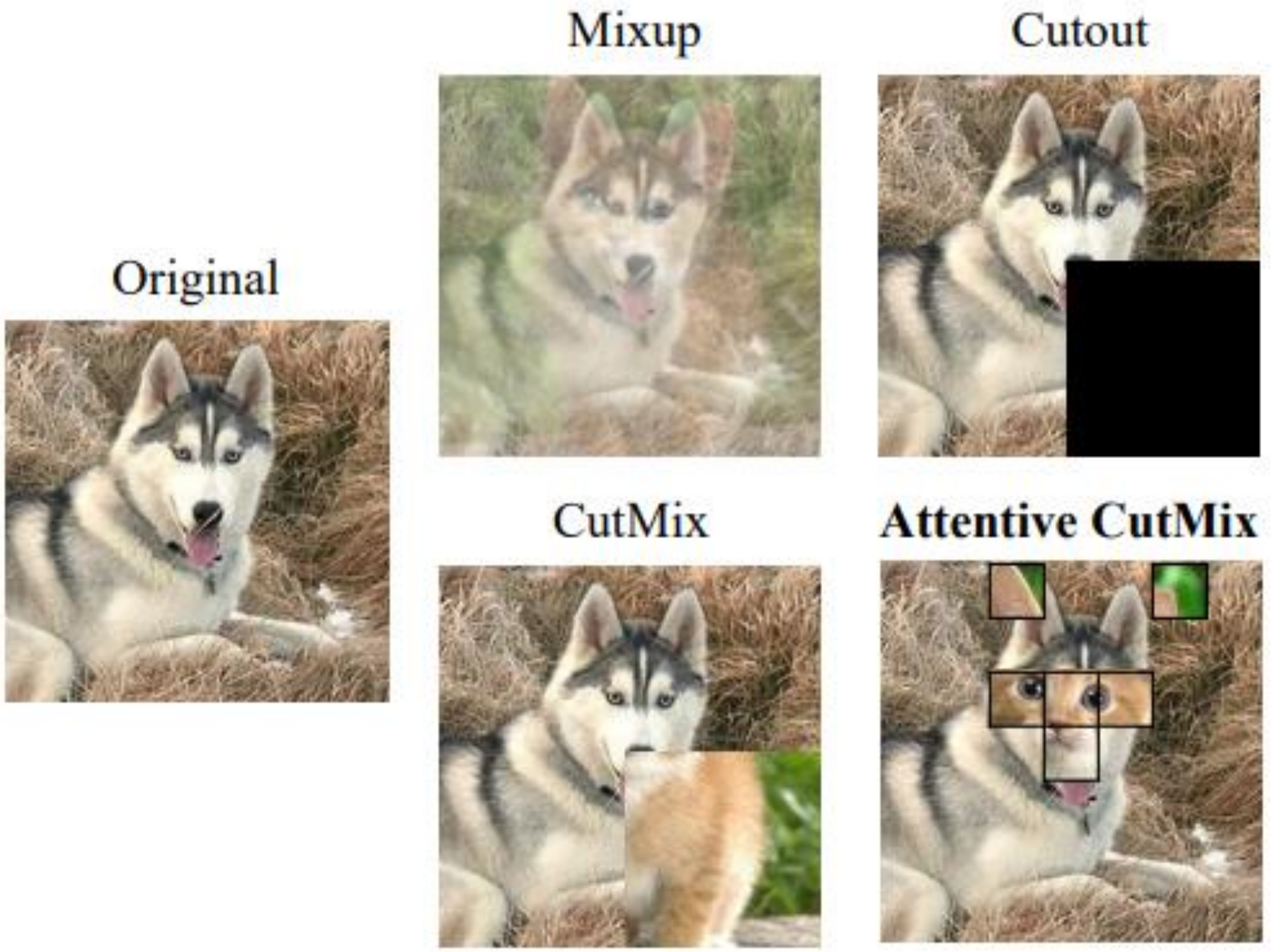
Table 3. Image classification results on Imagenet dataset.

Corruption type		Baseline[56]	$SmoothMix_C$
Noise	Gaussian_noise	92.38	**88.4**
	Shot_noise	85.29	**80.67**
	Impulse_noise	**83.13**	91.94
	Speckle_noise	82.99	**79.49**
Blur	Defocus_blur	29.78	**30.5**
	Glass_blur	**74.47**	79.71
	Motion_blur	36.27	**34.52**
	Zoom_blur	**33.05**	35.07
	Gaussian_blur	39.53	**39.05**
Weather	Brightness	18.89	**18.01**
	Fog	26.19	**21.52**
	Frost	40.73	**35.41**
	Snow	32.06	**30.12**
	Spatter	**23.82**	31.05
Digital	Saturate	26.9	**25.81**
	Pixelate	46.53	**42.86**
	Contrast	33.72	**26.06**
	Elastic_transform	**31.85**	32.37
	Jpeg_compression	54.1	**49.6**
	Average	46.93	**45.90**

Table 4. Image classification results on CIFAR-100-C corruption dataset.

ATTENTIVE CUTMIX

ATTENTIVE CUTMIX - Devesh Walawalkar, Zhiqiang Shen, Zechun Liu, Marios Savvides: Attentive Cutmix: An Enhanced Data Augmentation Approach for Deep Learning Based Image Classification. ICASSP 2020: 3642-3646



ATTENTIVE CUTMIX

ATTENTIVE CUTMIX - Devesh Walawalkar, Zhiqiang Shen, Zechun Liu, Marios Savvides: Attentive Cutmix: An Enhanced Data Augmentation Approach for Deep Learning Based Image Classification. ICASSP 2020: 3642-3646

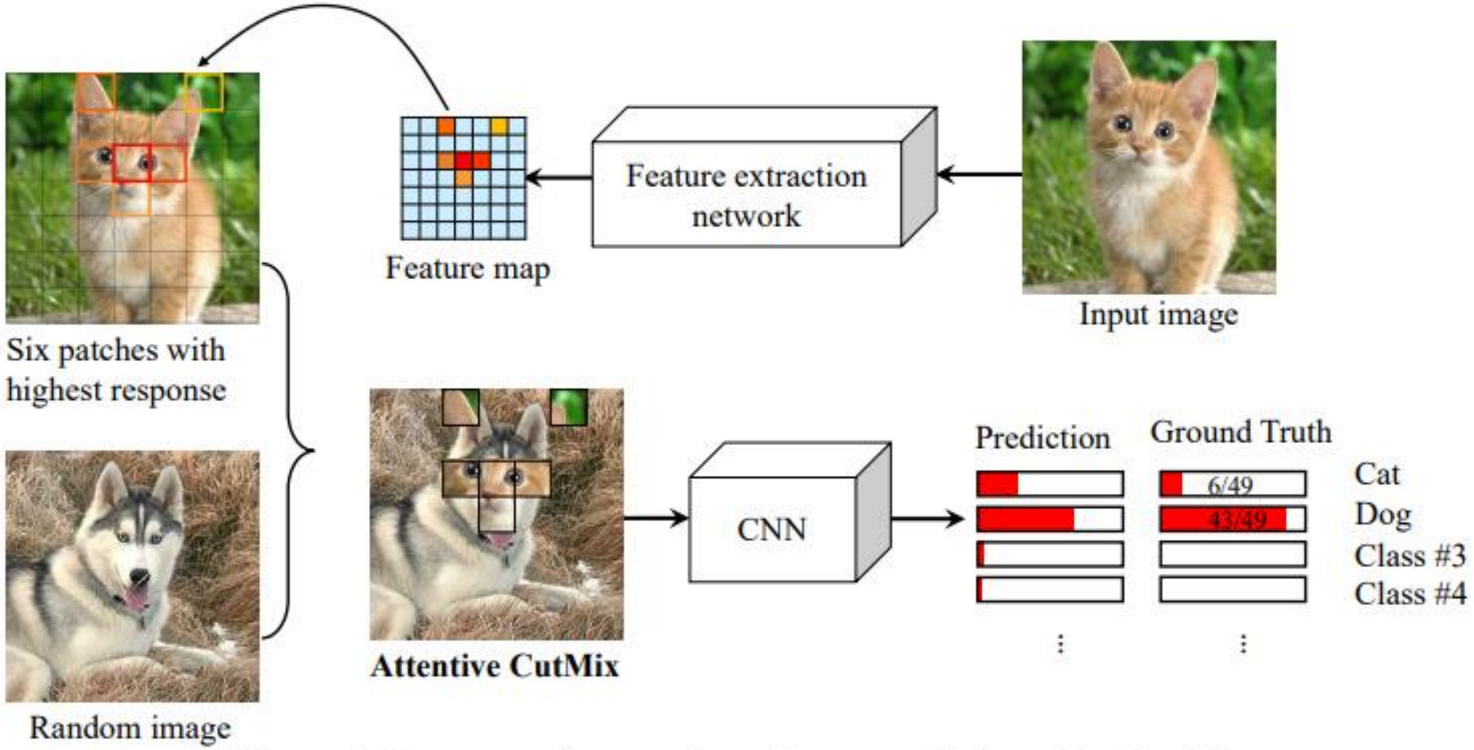


Figure 2: Framework overview of proposed *Attentive CutMix*.

ATTENTIVE CUTMIX

ATTENTIVE CUTMIX - Devesh Walawalkar, Zhiqiang Shen, Zechun Liu, Marios Savvides: Attentive Cutmix: An Enhanced Data Augmentation Approach for Deep Learning Based Image Classification. ICASSP 2020: 3642-3646

	CIFAR-10 (%)			
Method	**Baseline**	**Mixup**	**CutMix**	**Attentive CutMix**
ResNet-18	84.67	88.52	87.92	**88.94**
ResNet-34	87.12	88.70	88.75	**90.40**
ResNet-101	90.47	91.89	92.13	**93.25**
ResNet-152	92.45	94.21	94.35	**94.79**
DenseNet-121	85.65	87.56	87.98	**88.34**
DenseNet-169	87.67	89.12	89.23	**90.45**
DenseNet-201	91.21	93.21	93.45	**94.16**
DenseNet-264	92.78	94.20	94.34	**94.83**
EfficientNet - B0	87.45	88.07	88.67	**88.94**
EfficientNet - B1	90.12	90.99	91.37	**92.10**
EfficientNet - B6	92.74	93.76	93.28	**93.92**
EfficientNet - B7	94.95	95.11	95.25	**95.86**

	CIFAR-100 (%)			
Method	**Baseline**	**Mixup**	**CutMix**	**Attentive CutMix**
ResNet-18	63.14	64.40	65.90	**67.16**
ResNet-34	65.54	67.83	68.32	**70.03**
ResNet-101	68.24	70.76	71.32	**72.86**
ResNet-152	71.49	74.81	73.21	**75.37**
DenseNet-121	65.12	66.84	67.62	**69.23**
DenseNet-169	66.42	68.24	69.58	**71.34**
DenseNet-201	70.28	72.89	73.57	**74.65**
DenseNet-264	73.51	76.49	75.23	**77.58**
EfficientNet - B0	64.67	65.78	66.95	**67.48**
EfficientNet - B1	66.89	68.23	68.12	**68.96**
EfficientNet - B6	71.34	73.56	73.75	**74.82**
EfficientNet - B7	75.67	77.21	77.57	**78.52**

RICAP

RICAP - Takahashi, Ryo & Matsubara, Takashi & Uehara, Kuniaki. (2019). Data Augmentation using Random Image Cropping and Patching for Deep CNNs. IEEE Transactions on Circuits and Systems for Video Technology. PP. 1-1. 10.1109/TCSVT.2019.2935128.

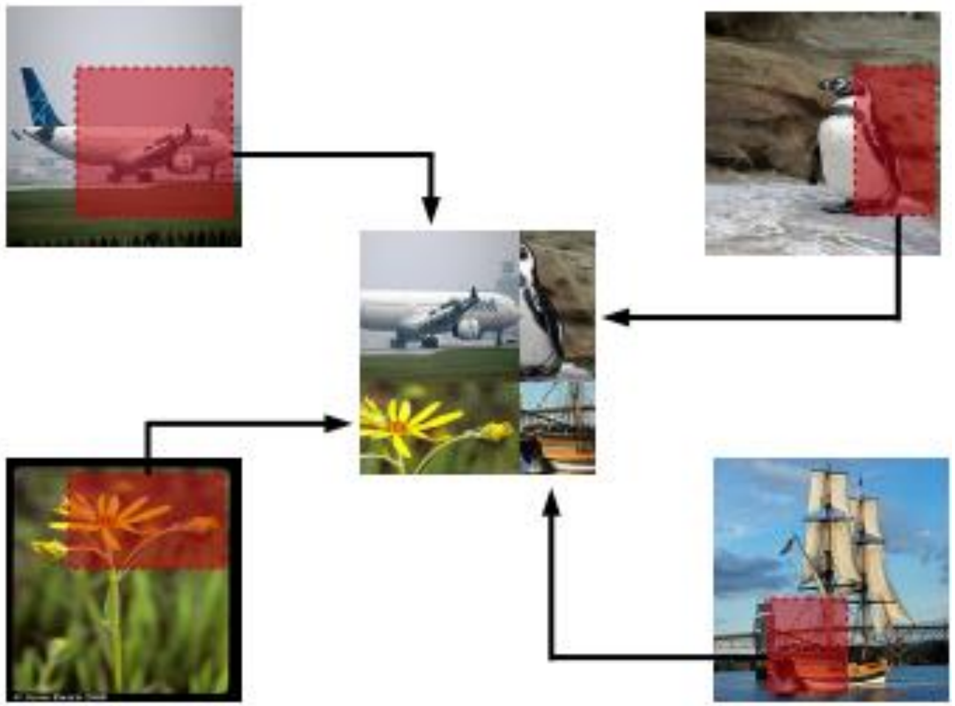


Fig. 1. Conceptual explanation of the proposed *random image cropping and patching* (*RICAP*) data augmentation. Four training images are randomly cropped as denoted by the red shaded areas, and patched to construct a new training image (at center). The size of the final image is identical to that of the original one (e.g., 32 × 32 for the CIFAR dataset [8]). These images are collected from the training set of the ImageNet dataset [24].

RICAP

RICAP - Takahashi, Ryo & Matsubara, Takashi & Uehara, Kuniaki. (2019). Data Augmentation using Random Image Cropping and Patching for Deep CNNs. IEEE Transactions on Circuits and Systems for Video Technology. PP. 1-1. 10.1109/TCSVT.2019.2935128.

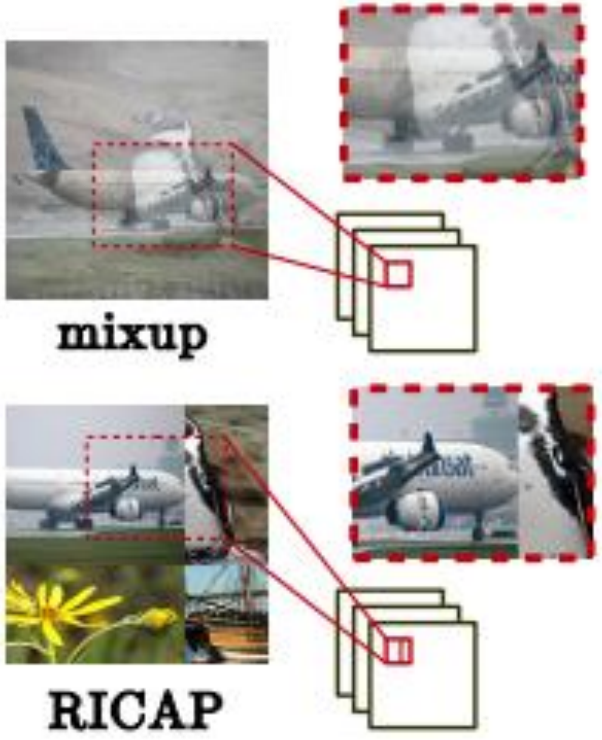


Fig. 4. Comparison between images processed by RICAP and mixup.

TABLE I
TEST ERROR RATES USING WIDERESNET ON THE CIFAR DATASET.

Method	CIFAR-10	CIFAR-100
Baseline	3.89	18.85
+ dropout ($p = 0.2$)	4.65 $\pm$0.08†	21.27 $\pm$0.19†
+ cutout (16×16)	3.08 $\pm$0.16	18.41 $\pm$0.27
+ random erasing	3.08 $\pm$0.05	17.73 $\pm$0.15
+ mixup ($\alpha = 1.0$)	3.02 $\pm$0.04†	17.62 $\pm$0.25†
+ RICAP ($\beta = 0.3$)	**2.85** $\pm$0.06	**17.22** $\pm$0.20

† indicates the results of our experiments.

RICAP

RICAP - Takahashi, Ryo & Matsubara, Takashi & Uehara, Kuniaki. (2019). Data Augmentation using Random Image Cropping and Patching for Deep CNNs. IEEE Transactions on Circuits and Systems for Video Technology. PP. 1-1. 10.1109/TCSVT.2019.2935128.

TABLE III
TEST ERROR RATES ON CIFAR-10.

Method	DenseNet-BC 190-40	Pyramidal ResNet 272-200	Shake-Shake 26 2x96d
Baseline	3.46	3.31 $\pm$0.08	2.86
+ dropout ($p = 0.2$)	4.56 †	4.06 †	3.79 †
+ cutout (8×8)	2.73 $\pm$0.06†	2.84 $\pm$0.05†	2.56 $\pm$0.07
+ mixup ($\alpha = 1.0$)	2.73 $\pm$0.08†	2.57 $\pm$0.09†	2.32 $\pm$0.11†
+ RICAP ($\beta = 0.3$)	**2.69** $\pm$0.12	**2.51** $\pm$0.02	**2.19** $\pm$0.08

† indicates the results of our experiments.

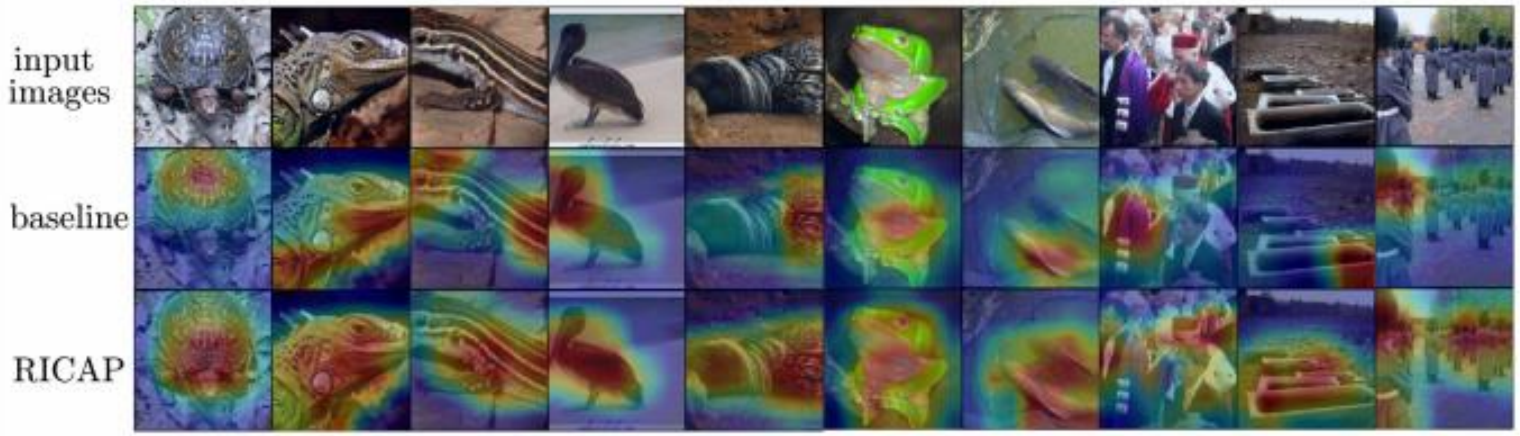


Fig. 7. Class Activation Mapping (CAM) [25] of WideResNet 28-10. The top row shows the input images. The middle row shows the CAM of WideResNet 28-10 without RICAP denoted as *baseline*. The bottom row shows the CAM of WideResNet 28-10 with RICAP.

RICAP

RICAP - Takahashi, Ryo & Matsubara, Takashi & Uehara, Kuniaki. (2019). Data Augmentation using Random Image Cropping and Patching for Deep CNNs. IEEE Transactions on Circuits and Systems for Video Technology. PP. 1-1. 10.1109/TCSVT.2019.2935128.

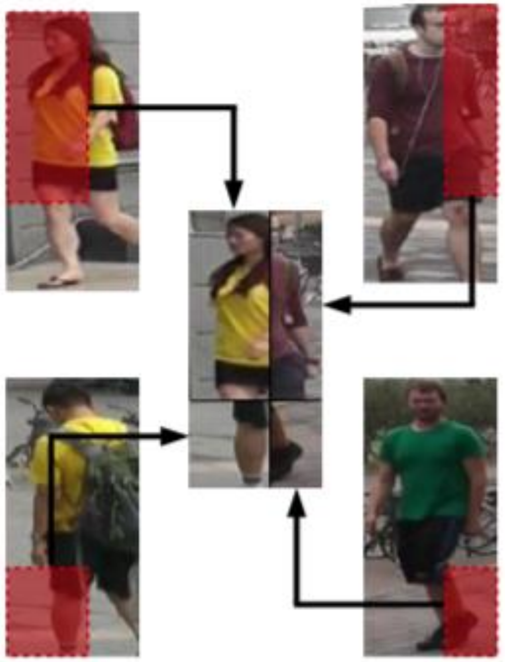


Fig. 10. Conceptual explanation of the proposed *fixed image cropping and patching* (*FICAP*) data augmentation. For application of RICAP to person re-identification task, we replaced the random cropping with fixed cropping because of the feature vector matching for image-to-image retrieval.

Fig. 11. Detection examples of MS-COCO test images by the baseline YOLOv3 (top row) and a TOLOv3 trained with RICAP (bottom row). Without RICAP, a zebra hidden in a tree was detected as two different objects in the left image, side-by-side buses were not detected in the middle image, and a horse tail was misdetected as the dog in the right image, respectively. RICAP solved these issues, indicating that RICAP makes YOLOv3 be robust to the occlusion.

The end.
Thank you!