CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features

Sangdoo Yun[1] Dongyoon Han[1] Seong Joon Oh[1] Sanghyuk Chun[1] Junsuk Choe[2] Youngjoon Yoo[1]

Clova AI Research[1] Yonsei University[2]

NAVER LINE Clova

ICCV 2019 Seoul, Korea

CutMix in a Nutshell

Q) What is **CutMix**?

- Blend two images and labels by **cut-and-paste** manner.

Q) What are the benefits of **CutMix**?

1. **Simple**, **intuitive**, and **effective**. (only 20 lines of PyTorch)
2. Vastly improves classifier's **accuracy**, **localization ability**.

Motivation and Related Works

- Goal: to learn **generalizable** and **localizable** features.
- Comparison with previous works:

- Unlike Cutout, **CutMix uses all input pixels** for training.
- Unlike Mixup, **CutMix** presents **realistic local image patches.**

[a] Devries et al., "Improved regularization of convolutional neural networks with cutout", arXiv 2017.
[b] Zhong et al., "Random erasing data augmentation", arXiv 2017.
[c] Zhang et al., "mixup: Beyond empirical risk minimization.", ICLR 2018.

Method

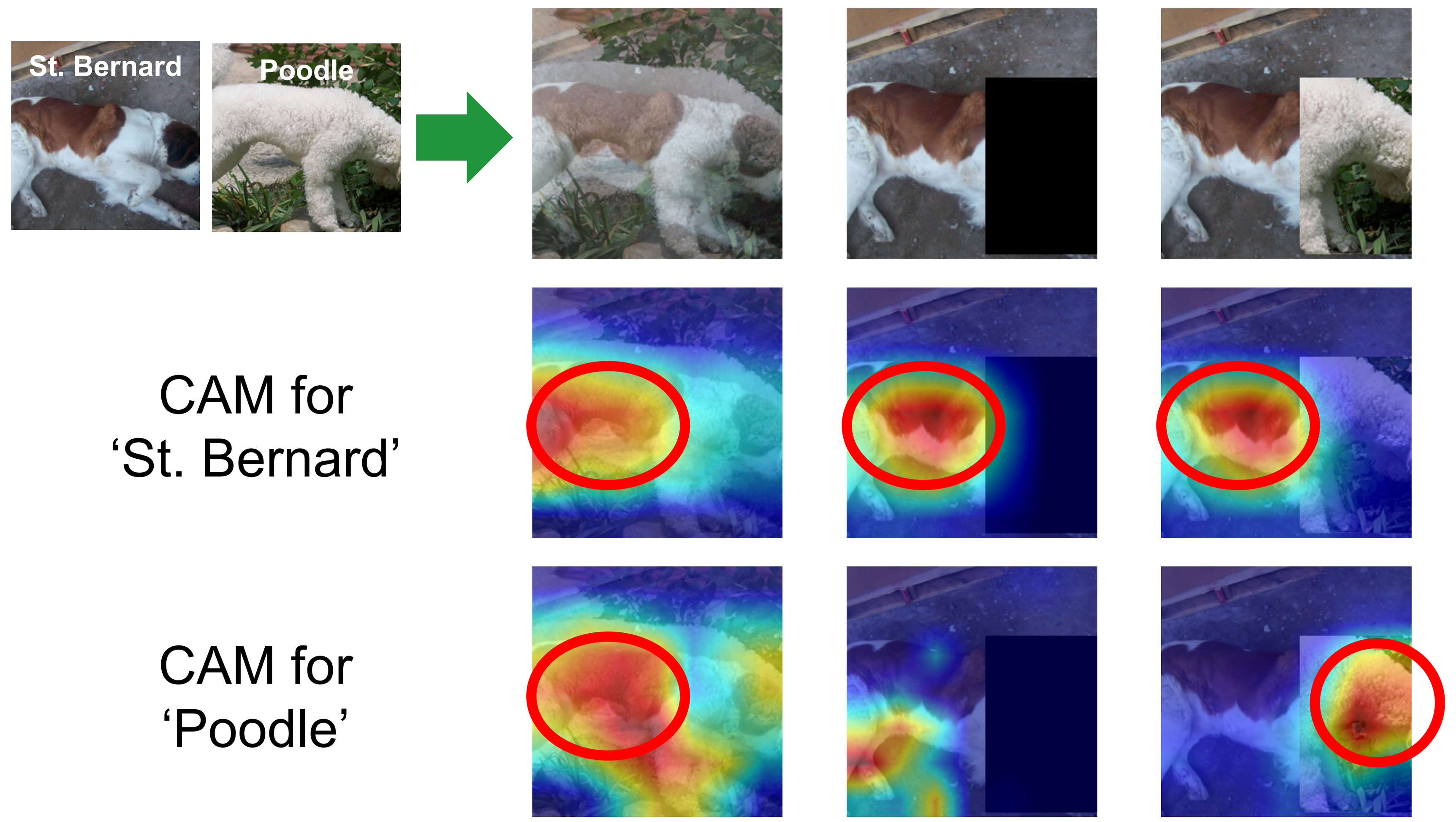
- How to training with CutMix:

1. For training samples $(x_A, y_A), (x_B, y_B)$ in a minibatch
2. Sample mixing ratio λ ~ Unif(0,1)
3. Select cropping mask $\mathbf{M}$ (the cropping area ratio = $1-\lambda$)
4. CutMix sample is generated by,

$$\tilde{x} = \mathbf{M} \odot x_A + (\mathbf{1} - \mathbf{M}) \odot x_B$$
$$\tilde{y} = \lambda y_A + (1 - \lambda) y_B,$$

5. Then, train a classifier using the CutMix sample $(\tilde{x}, \tilde{y})$

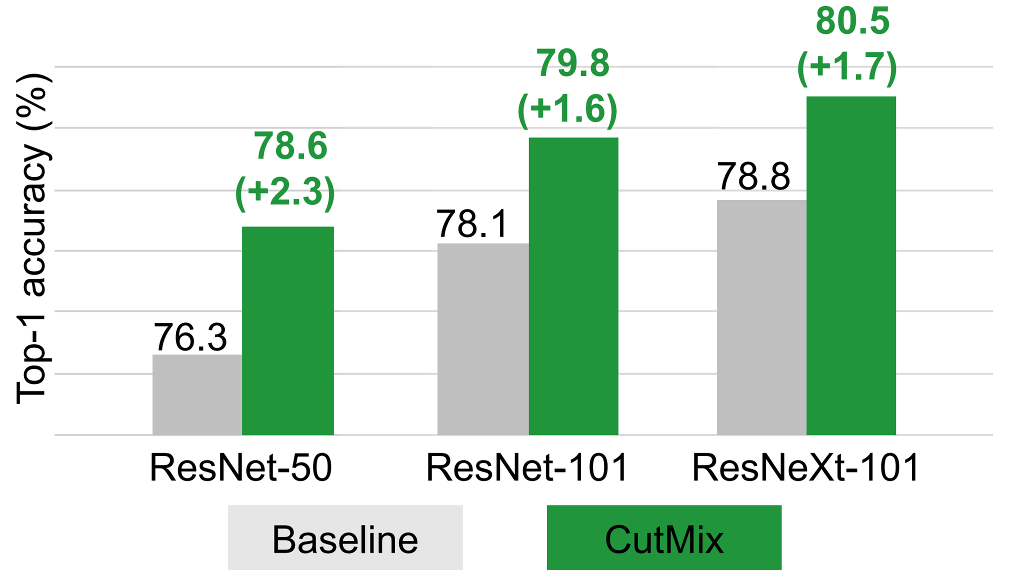
- Classification problem is changed to finding ***"what"***, ***"where"***, and ***"how large"*** the objects in image.



Experiments

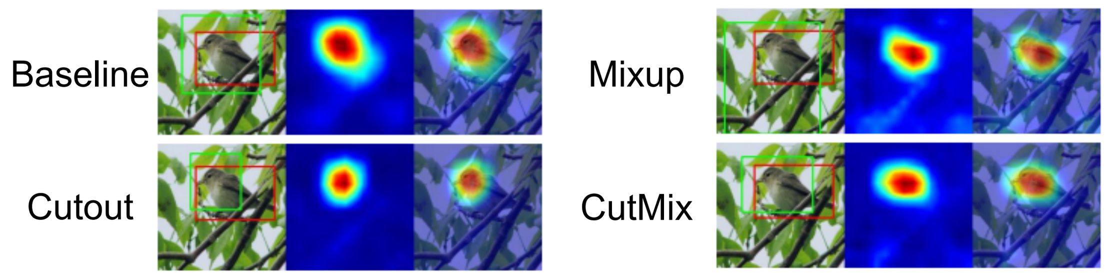
<ImageNet Classification>

Model	Top-1 Acc (%)	Top-5 Acc (%)
ResNet-50 (Baseline)	76.3	93.0
ResNet-50 + Cutout (arXiv'17)	77.1	93.3
ResNet-50 + StochDepth (ECCV'18)	77.5	93.7
ResNet-50 + Mixup (ICLR'18)	77.4	93.6
ResNet-50 + DropBlock (NeurIPS'18)	78.1	94.0
ResNet-50 + Manifold Mixup (ICML'19)	77.5	93.8
ResNet-50 + AutoAugment (CVPR'19)	77.6	93.8
ResNet-50 + CutMix	**78.6**	**94.1**
ResNet-152	78.3	94.1



<Weakly Supervised Object Localization>

Method	CUB200-2011 Loc Acc (%)	ImageNet Loc Acc (%)
ResNet-50	49.4	46.3
ResNet-50 + Mixup	49.3	46.0
ResNet-50 + Cutout	52.8	46.7
ResNet-50 + CutMix	**54.8**	**47.3**



<Robustness>

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

(After FGSM white-box attack)

<Ablation Study on CIFAR-100>

PyramidNet-200	Top-1 Acc (%)	Top-5 Acc (%)
Baseline	83.6	96.3
CutMix	**85.6**	**97.0**
Center Gaussian CutMix	84.1	96.6
Fixed-size CutMix	85.0	96.9
One-hot CutMix	84.1	96.7
Scheduled CutMix	85.3	96.8
Complete-label CutMix	84.8	96.9

<Transfer Learning with CutMix-pretrained Model>

Backbone Network	Pascal VOC Detection SSD (mAP)	Pascal VOC Detection Faster-RCNN (mAP)	MS-COCO Detection Faster-RCNN (mAP)	Image Captioning NIC (BLEU-4)
ResNet-50 (Baseline)	76.7 (+0.0)	75.6 (+0.0)	33.3 (+0.0)	22.9 (+0.0)
Mixup-pretrained	76.6 (-0.1)	73.9 (-1.7)	34.2 (+0.9)	23.2 (+0.3)
Cutout-pretrained	76.8 (+0.1)	75.0 (-0.6)	34.3 (+1.0)	24.0 (+1.1)
CutMix-pretrained	**77.6 (+0.9)**	**76.7 (+1.1)**	**35.2 (+1.9)**	**24.9 (+2.0)**

Conclusion

- Need a strong classifier? → Apply CutMix to your classifier.
- Better pretrained model? → Download CutMix-pretrained model.
- Our codes and models: github.com/clovaai/CutMix-PyTorch