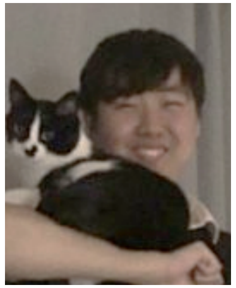
CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features

Sangdoo Yun
Clova AI
Naver

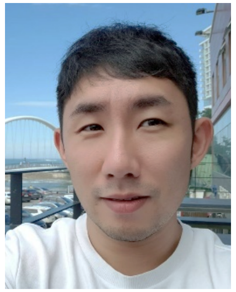
Dongyoon Han
Clova AI
Naver

Seong Joon Oh
Clova AI
LINE+

Sanghyuk Chun
Clova AI
Naver



Junsuk Choe*
Yonsei
University

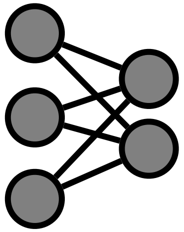


Youngjoon Yoo
Clova AI
Naver

* Intern at Clova.

NAVER Clova

Image Classification



Dog = 1.0

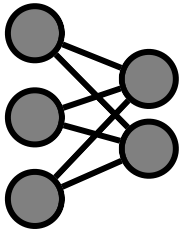
Input Image

Model

Target

Image Classification

- Regional dropout strategy for “occlusion-robust” classifier[a, b]



Dog = 1.0

Input Image

Model

Target

✓ Good generalization ability

✘ Cannot utilize full image regions

[a] Devries et al., “Improved regularization of convolutional neural networks with cutout”, arXiv 2017.
[b] Zhong et al., “Random erasing data augmentation”, arXiv 2017.

CutMix in a Nutshell

Target Label
Cat = 0.4
Dog = 0.6

- Cut and paste two images and labels.
- In this way, the classifier learns “what” and “where” objects are in the image.

CutMix in a Nutshell

	Original	Cutout[a]	Mixup[c]	**CutMix**

Label	Dog 1.0	Dog 1.0	Dog 0.5 Cat 0.5	Dog 0.6 Cat 0.4

✓ Unlike Cutout, **CutMix uses all input pixels** for training.

✓ Unlike Mixup, CutMix presents **realistic local image patches.**

✓ CutMix is simple: only 20 lines of pytorch code.

[c] Zhang et al., “mixup: Beyond empirical risk minimization.”, ICLR 2018.

Generalizability

- ImageNet validation set accuracies

Model	Top-1 Err (%)	Top-5 Err (%)
ResNet-50 (Baseline)	23.68	7.05
ResNet-50 + Cutout (arXiv'17)	22.93	6.66
ResNet-50 + StochDepth (ECCV'18)	22.46	6.27
ResNet-50 + Mixup (ICLR'18)	22.58	6.40
ResNet-50 + DropBlock (NeurIPS'18)	21.87	5.98
ResNet-50 + Manifold Mixup (ICML'19)	22.50	6.21
ResNet-50 + AutoAugment (CVPR'19)	22.40	6.20
ResNet-50 + CutMix	**21.60**	**5.90**
ResNet-152	21.69	5.94

✓ Great improvement over baseline **(+2%p)**.

✓ Outperforming existing methods.

✓ **ResNet50 + CutMix ≈ ResNet152**.

Localizability

- Weakly-supervised object localization (WSOL) on CUB and ImageNet.

Method	CUB200-2011 Loc Acc (%)	ImageNet Loc Acc (%)
ResNet-50	49.41	46.30
ResNet-50 + Mixup	49.30	45.84
ResNet-50 + Cutout	52.78	46.69
ResNet-50 + CutMix	**54.81**	**47.25**

✓ CutMix encourages detection of less discriminative object parts.

✓ Great improvement on localization task.

Transfer Learning

- CutMix-pretrained model is utilized as a backbone network.

Backbone Network	Pascal VOC Detection		MS-COCO Detection	Image Captioning
	SSD (mAP)	Faster-RCNN (mAP)	Faster-RCNN (mAP)	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)**

✓ **+2%p** improvements on MS-COCO ≈ **ResNet50→ResNet101 backbone change**.

✓ Choosing CutMix-pretrained model brings great performance improvement.

Conclusion and Take-home Messages

- Need to train a strong classifier?

  → Apply CutMix regularizer to your classifier.
- Need a better pretrained model for localization-related task?

  → Download our CutMix-pretrained model.

- Our codes and pre-trained models are available online:
  https://github.com/clovaai/CutMix-PyTorch

  (We are planning to release Tensorflow & MXNet codes and models.)

CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features

Please visit our poster (#6) for more experiments and analysis!

NAVER Clova