CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features

Abstract

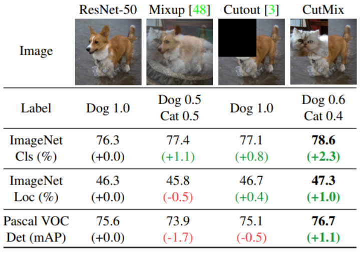
cnn
less discriminative parts of objects -> generalize better / better object localization capabilities.
regional dropout bad because -> remove informative pixels / information loss / inefficiency during training.

CutMix -> patches are cut and pasted among training images where the ground truth labels are also mixed proportionally to the area of the patches.
efficient use of training pixels / regularization effect of regional dropout
when used as a pretrained model, results in consistent performance
improves the model robustness against input corruptions and its out-of-distribution detection performances.

Introduction

기존 - including data augmentation / regularization techniques
to prevent a CNN from focusing too much on a small set of intermediate activations or on a small region on input images, random feature removal regularizations have been proposed.
ex) dropout
attend to the entire object region -> feature removal strategies -> generalization / localization
기존 dropout은 zeroed-out or filled with random noise -> greatly reducing the proportion of informative pixels
data hungry
maximally utilize the deleted regions / better generalization and localization -> CutMix
replace the removed regions with a patch from another image (See Table 1).
truth labels are also mixed proportionally to the number of pixels of combined images.
there is no uninformative pixel during training, making training efficient, while retaining the advantages of regional dropout to attend to non-discriminative parts of objects.

enhance localization ability
CutMix shares similarity with Mixup
Mixup samples tend to be unnatural
CutMix overcomes the problem by replacing the image region with a patch from another training image.



	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)**

CutMix also enhances the model robustness and alleviates the over-confidence issue [13, 22] of deep networks.

Related Works

Regional Dropout

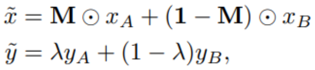
Synthesizing Training Data

Mixup

Tricks for training deep networks
Methods such as weight decay,
dropout [34], and Batch Normalization
noises to the internal features of CNNs [17, 8, 46] or adding
extra path to the architecture

CutMix

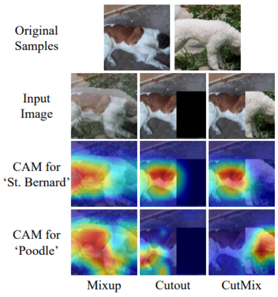
Algorithm

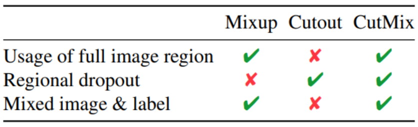


$$\tilde{x} = \mathbf{M} \odot x_A + (\mathbf{1} - \mathbf{M}) \odot x_B$$
$$\tilde{y} = \lambda y_A + (1 - \lambda) y_B,$$

simple
negligible computational overhead

Discussion





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

What does model learn with CutMix?
Cutout - lets a model focus on less discriminative parts of the object
CutMix - Cutout + localize

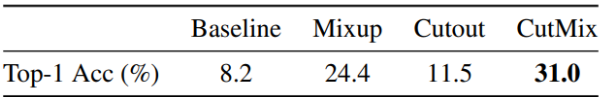
Analysis on validation error

stabilizing

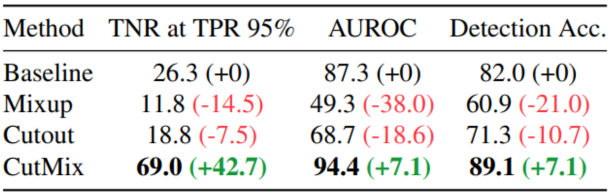
diverse training samples reduce overfitting.

Experiments

Robustness and Uncertainty



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



Method	TNR at TPR 95%	AUROC	Detection Acc.
Baseline	26.3 (+0)	87.3 (+0)	82.0 (+0)
Mixup	11.8 (-14.5)	49.3 (-38.0)	60.9 (-21.0)
Cutout	18.8 (-7.5)	68.7 (-18.6)	71.3 (-10.7)
CutMix	**69.0 (+42.7)**	**94.4 (+7.1)**	**89.1 (+7.1)**

Conclusion

strong classification and localization

easy to implement / no computational overhead