CutMix

Group 2

GGH, KM, JYH

Outline

1. Introduction
2. Methods
3. Evaluation
4. Code Demo
5. Challenges
6. Conclusion

Introduction

Tackled Problem

Problem: CNN often focus too much on small region of input images

- Solutions for the performance drops being developed
  - e.g. Dropout: cuts out certain area of an image

Goal of CutMix

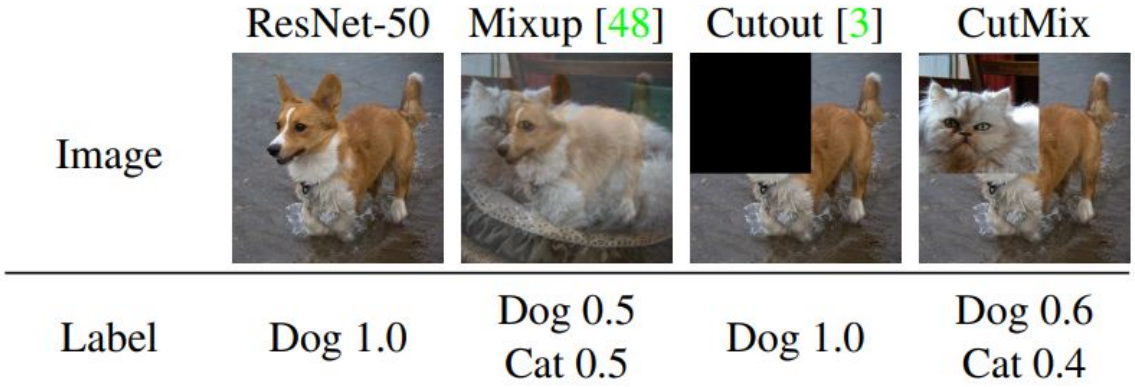
- Maximally utilize deleted regions,

  while taking advantage of better generalization using regional dropout

Introduction

Key concept

- “Cut & Paste”



Result

- Full objects considered for cues for classification
- Two objects are recognized in a single image
- Advantage of regional dropout
- Improvements on localization & state-of-an-art error

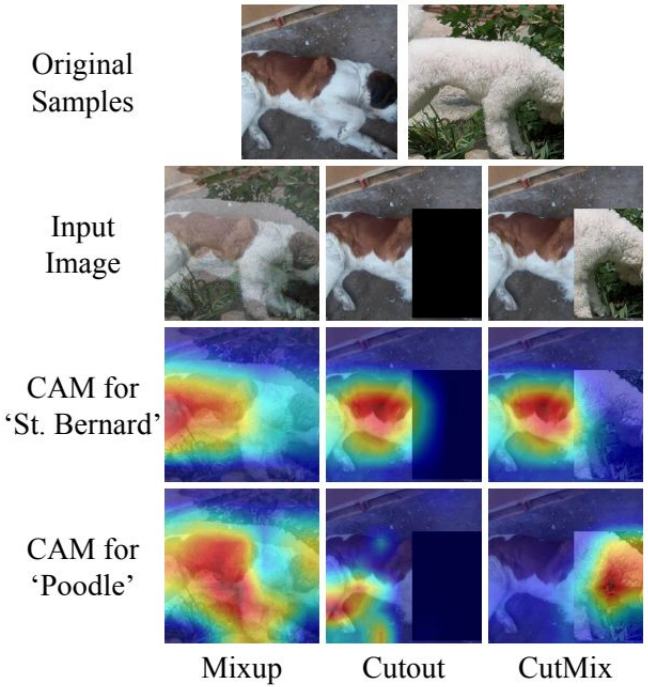
Methods

What does model learn with CutMix?

- Activation map (CAM)

Analysis on validation error

- ResNet50 for ImageNet Classification
- PyramidNet200 for CIFAR Classification
- Top1 Err, Top5 Err



*Terminologies

(1) Top-1 Error, Top-5 Error

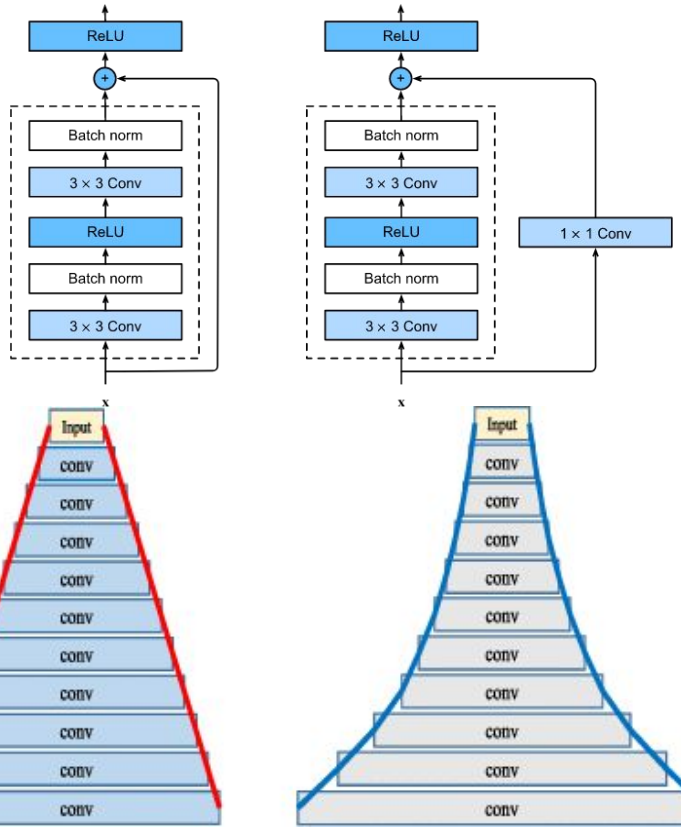
[cat]

Cat: 0.4
Dog: 0.26
Cup: 0.11
Male: 0.07
Bag: 0.02
.
.
.

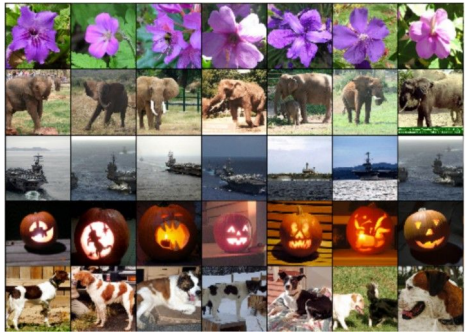
Top-1 Error

Top-5 Error

(2) ResNet50, PyramidNet200



(3) ImageNet, CIFAR



Example Dataset: **CIFAR10**

10 classes
50,000 training images
10,000 testing images

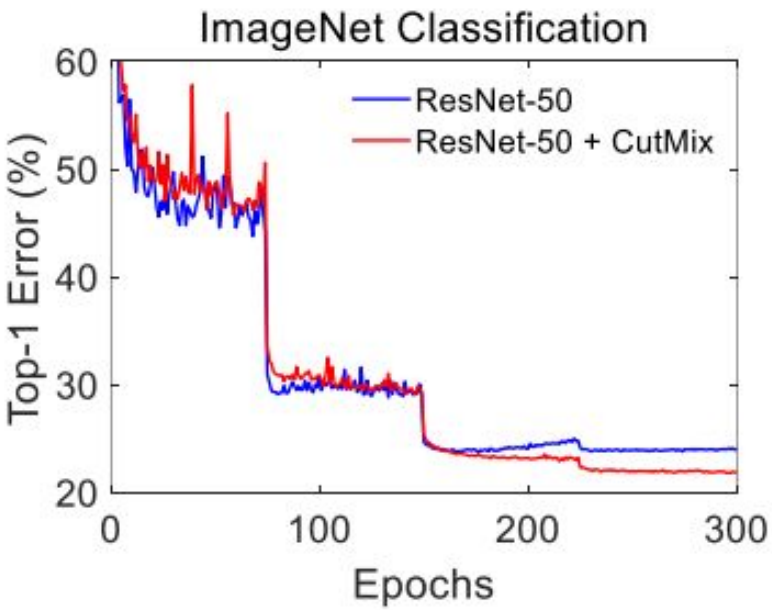
Alex Krizhevsky, "Learning Multiple Layers of Features from Tiny Images", Technical Report, 2009.

Evaluation

baseline: ResNet-50
dataset: ImageNet
300 epochs, batch size = 256
lr = 0.1, decay = 0.1
metric: top-1 error, top-5 error

1) ImageNet Classification

Model	# Params	Top-1 Err (%)	Top-5 Err (%)
ResNet-152*	60.3 M	21.69	5.94
ResNet-101 + SE Layer* [15]	49.4 M	20.94	5.50
ResNet-101 + GE Layer* [14]	58.4 M	20.74	5.29
ResNet-50 + SE Layer* [15]	28.1 M	22.12	5.99
ResNet-50 + GE Layer* [14]	33.7 M	21.88	5.80
ResNet-50 (Baseline)	25.6 M	23.68	7.05
ResNet-50 + Cutout [3]	25.6 M	22.93	6.66
ResNet-50 + StochDepth [17]	25.6 M	22.46	6.27
ResNet-50 + Mixup [48]	25.6 M	22.58	6.40
ResNet-50 + Manifold Mixup [42]	25.6 M	22.50	6.21
ResNet-50 + DropBlock* [8]	25.6 M	21.87	5.98
ResNet-50 + Feature CutMix	25.6 M	21.80	6.06
ResNet-50 + CutMix	25.6 M	**21.40**	**5.92**

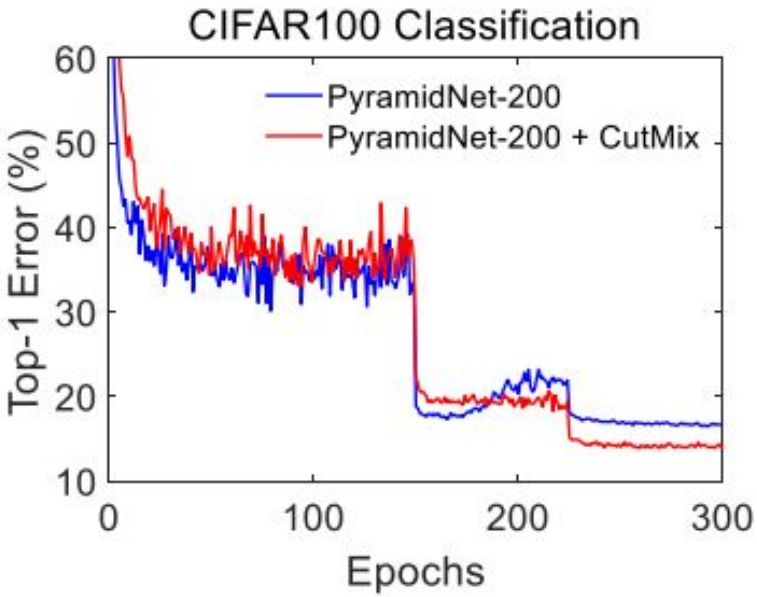


Evaluation

2) CIFAR-100 Classification

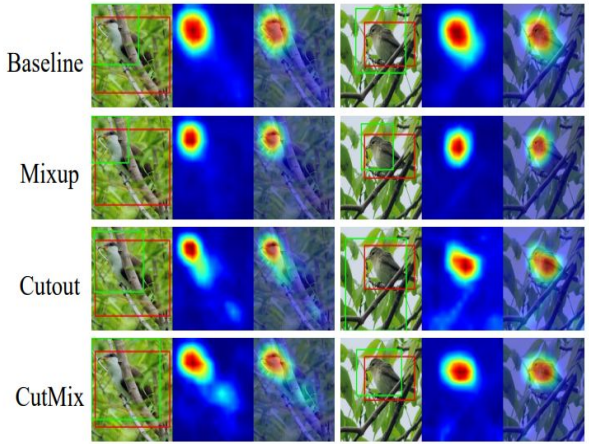
baseline: PyramidNet-200
dataset: CIFAR-100
300 epochs, batch size = 64
lr = 0.25, decay = 0.1
metric: top-1 error, top-5 error

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**

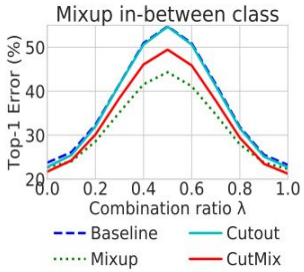


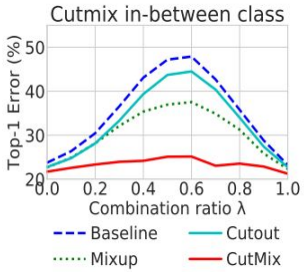
Evaluation

3) Weakly Supervised Object Localization
4) CutMix-ImageNet pre-trained model for object detection and image captioning
5) Robustness & Uncertainty



Backbone Network	ImageNet Cls Top-1 Error (%)	Detection		Image Captioning	
		SSD [24] (mAP)	Faster-RCNN [30] (mAP)	NIC [43] (BLEU-1)	NIC [43] (BLEU-4)
ResNet-50 (Baseline)	23.68	76.7 (+0.0)	75.6 (+0.0)	61.4 (+0.0)	22.9 (+0.0)
Mixup-trained	22.58	76.6 (-0.1)	73.9 (-1.7)	61.6 (+0.2)	23.2 (+0.3)
Cutout-trained	22.93	76.8 (+0.1)	75.0 (-0.6)	63.0 (+1.6)	24.0 (+1.1)
CutMix-trained	21.40	**77.6 (+0.9)**	**76.7 (+1.1)**	**64.2 (+2.8)**	**24.9 (+2.0)**





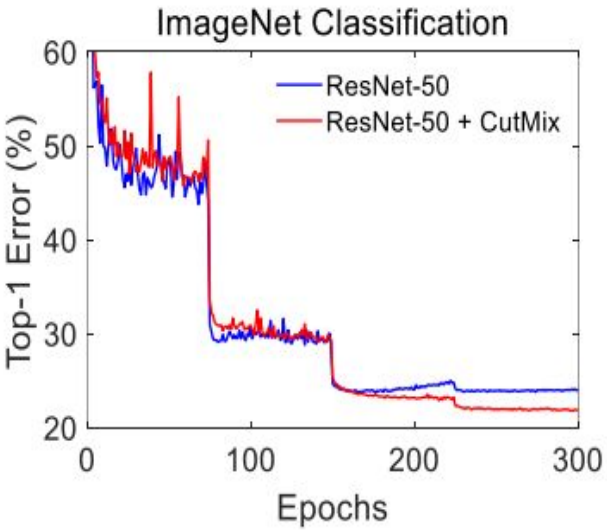
Method	TNR at TPR 95%	AUROC	Detection Acc.
Baseline	26.3 (+0)	87.3 (+0)	82.0 (+0)
Mixup	11.8 (-14.5)	49.3 (-38.0)	60.9 (-21.0)
Cutout	18.8 (-7.5)	68.7 (-18.6)	71.3 (-10.7)
CutMix	**69.0 (+42.7)**	**94.4 (+7.1)**	**89.1 (+7.1)**

Code Demo

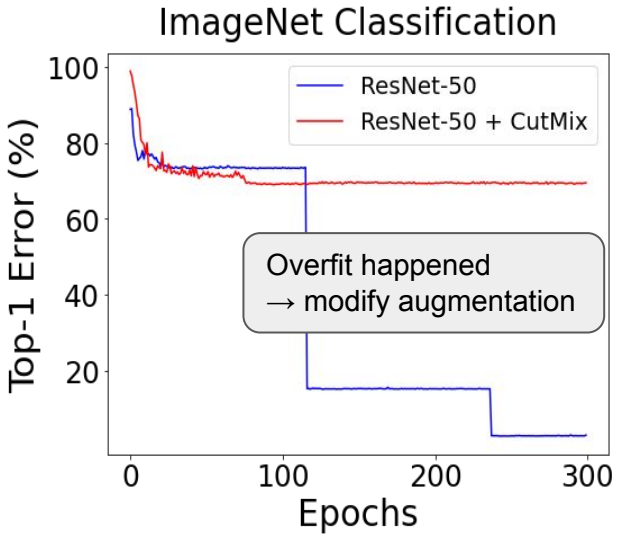
Code Reproducing Results

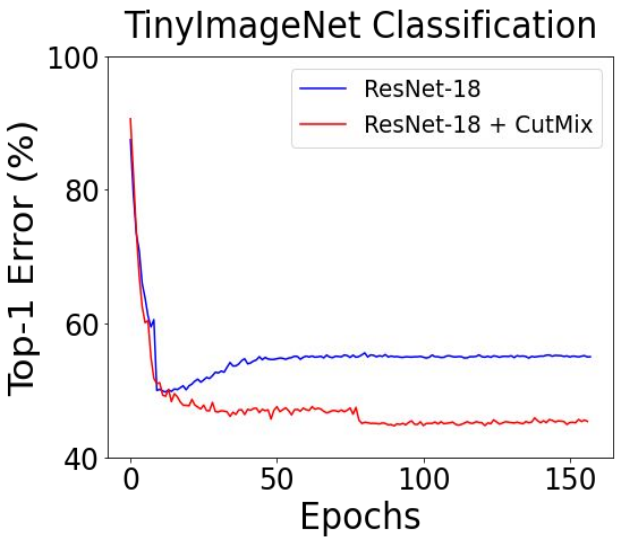
1) ImageNet Classification

< Result in Paper >



< Our Result >

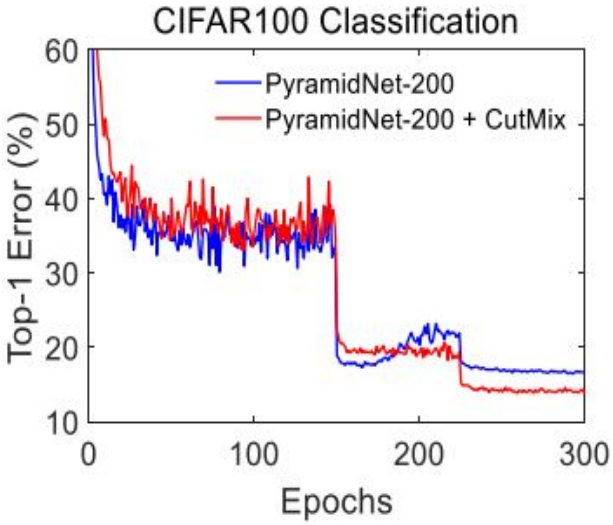




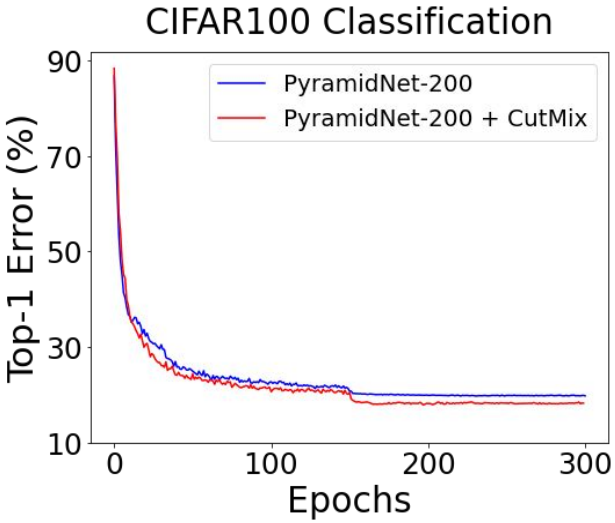
Code Reproducing Results

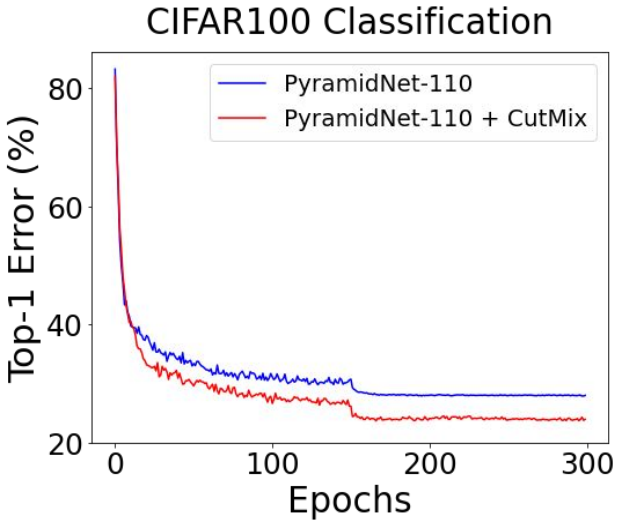
2) CIFAR-100 Classification

< Result in Paper >

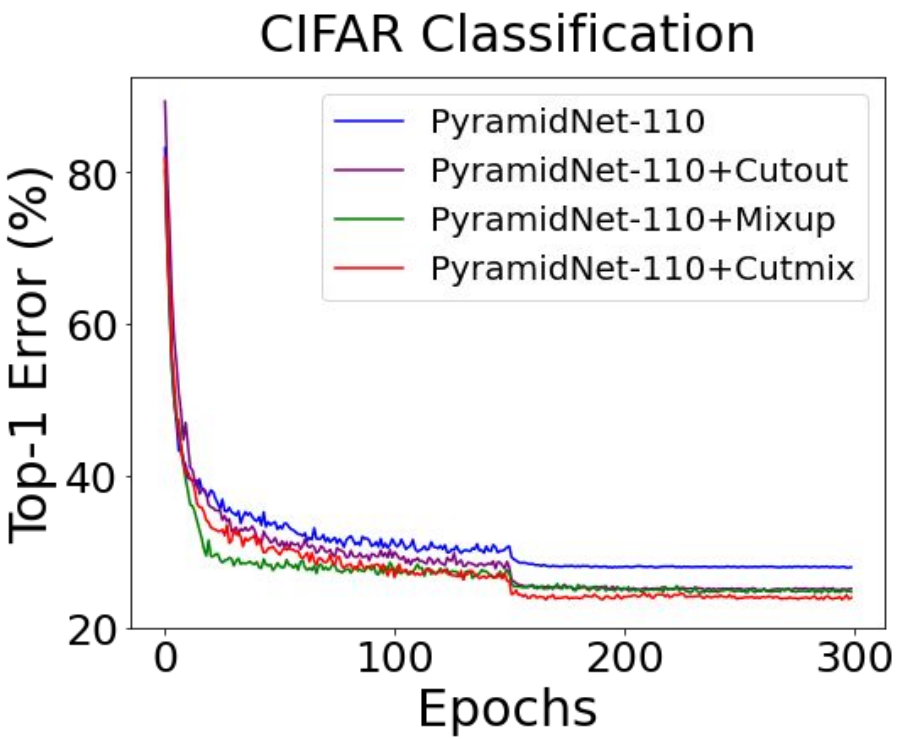


< Our Result >



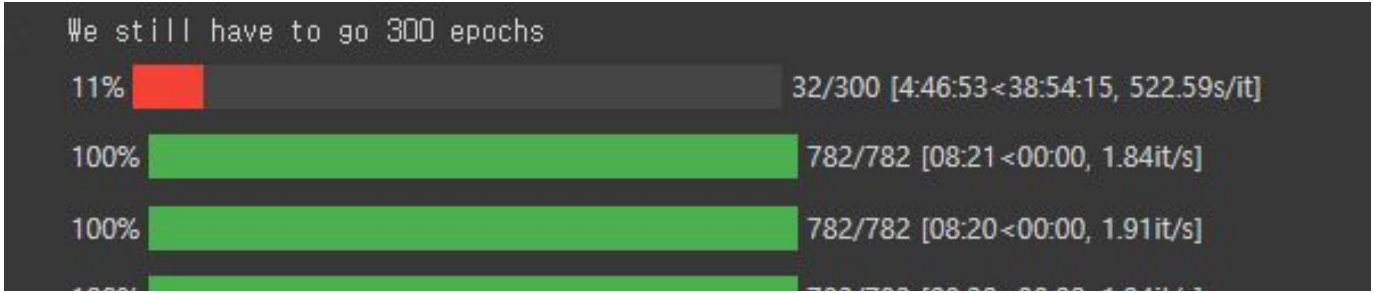


Code Reproducing Results

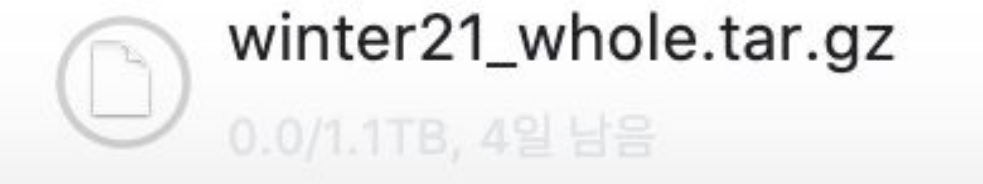


Challenges

1) GPU runtime limitation



2) Too large dataset to proceed training



ImageNet : 1.1TB

Conclusion

- To solve the information loss of existing regional dropout
- CutMix = the augmentation strategy that attaches a patch of another image to the cut part
- Achievements
  - ImageNet classifier : accuracy ↑
  - CIFAR-100 classifier : accuracy ↑
  - Weakly Supervised Object Localization : accuracy ↑
  - Object detection & Image captioning by transfer learning : performance ↑
  - Robustness & Uncertainty ↑
- Limitation: performance varies by dataset
- Extension: black-and-white images

References

Dongyoon Han, Jiwhan Kim, Junmo Kim; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5927-5935

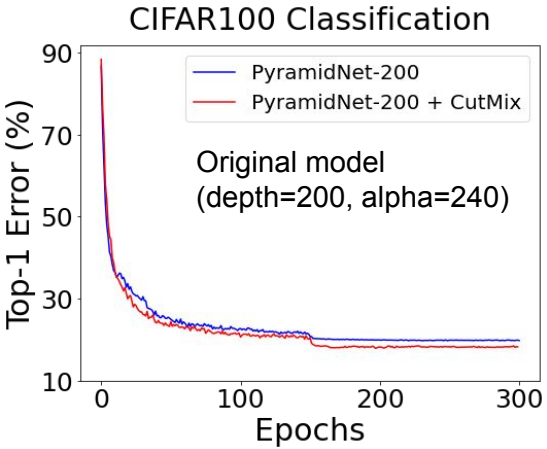
Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778

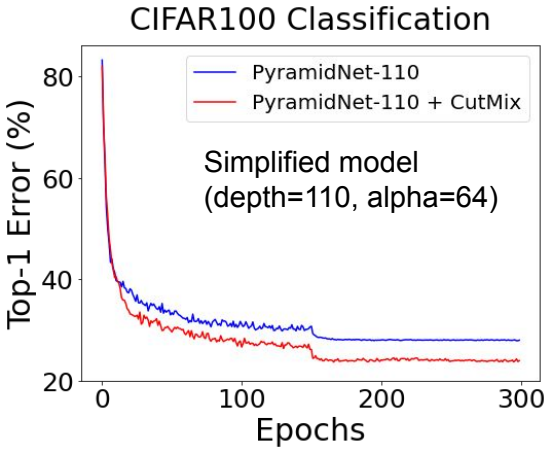
Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, Youngjoon Yoo; Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 6023-6032

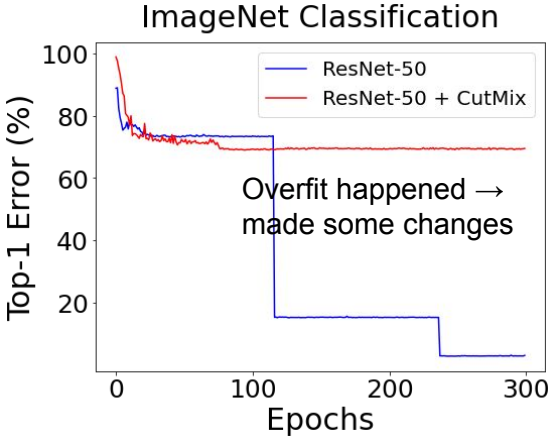
Wikipedia contributors. (2021, December 29). *Canadian Institute for Advanced Research*. Wikipedia. https://en.wikipedia.org/wiki/Canadian_Institute_for_Advanced_Research

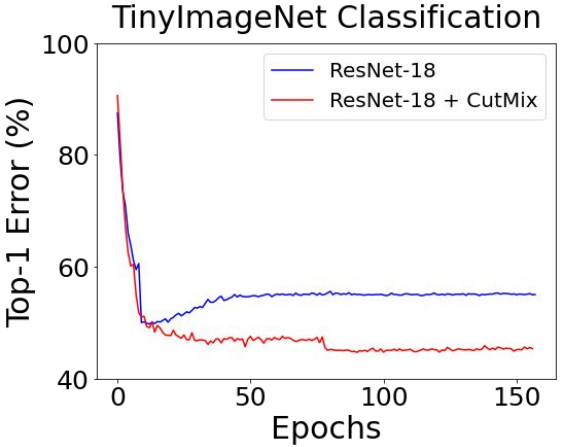
Wikipedia contributors. (2022, May 10). *ImageNet*. Wikipedia. https://en.wikipedia.org/wiki/ImageNet

Representation of Figure 2









Representation of Table 3

Model	Top-1 Err(%)	Top-5 Err(%)	Note
ResNet-50 (Baseline)	3.11	2.00	
ResNet-50 + Cutout	71.37	48.71	Overfit
ResNet-50 + CutMix	69.52	46.77	Overfit

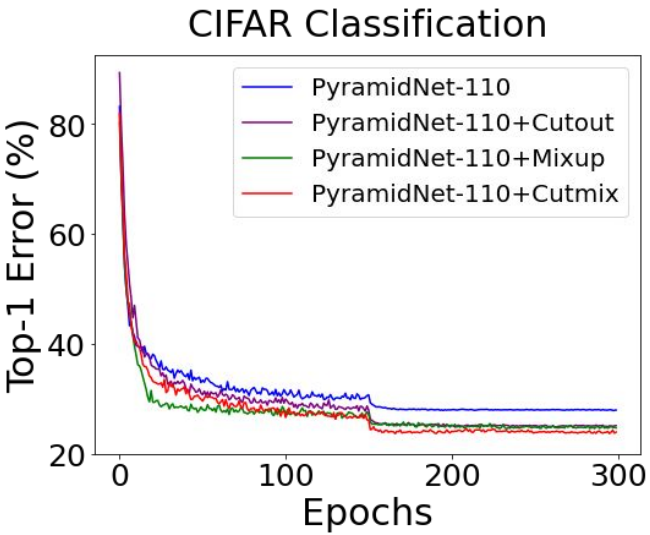
Model	Top-1 Err(%)	Top-5 Err(%)
ResNet-18 (Baseline)	50.13	24.47
ResNet-18 + CutMix	45.16	22.45

NOTE: We've suffered dramatic overfitting during Cutout and CutMix in ResNet-50.
So we stopped training with Mixup, and changed models to ResNet18.
For ResNet18, we trained only baseline and CutMix.

Representation of Table 5

Model	Top-1 Err(%)	Top-5 Err(%)
PyramidNet-110 (Baseline)	27.99	8.20
PyramidNet-110 +Cutout	25.14	6.77
PyramidNet-110 +Mixup	24.72	7.55
PyramidNet-110 +CutMix	23.96	6.81

Model	Top-1 Err(%)	Top-5 Err(%)
PyramidNet-200 (Baseline)	19.79	5.21
PyramidNet-200 +CutMix	18.25	4.59



Appendix: Experiments with FashionMNIST

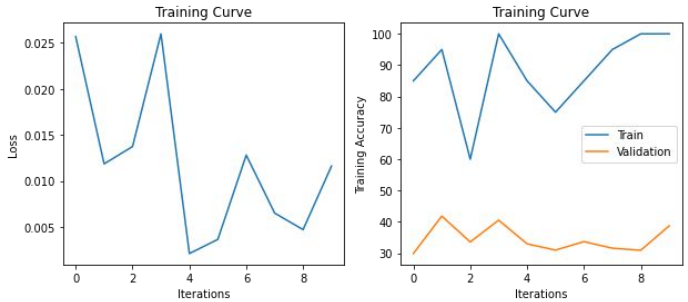


Fig 1. Applying 4 TL methods

- Dataset Similarity(Low), Dataset Size(Small)
- Dataset Similarity(Low), Dataset Size(Large)
- Dataset Similarity(High), Dataset Size(Small)
- Dataset Similarity(High), Dataset Size(Large)

Fig 2. Changing transformation

- Normalize only
- Norm + Horizontal Flip + Vertical Flip
- HFlip + VFlip + Random Rotation

Dataset: FashionMNIST

Test Results

- CutMIx: O, X
- Transform:
  - A: Normalization only
  - B: Normalization + RandomHFlip + RandomVFlip + Randomrotation
- Results ordered by Train acc, Val acc, Test acc

Test 1. lr=0.001, iter=10, batch_size=128, weight_decay=0.01

	Transform A	Transform B
CutMix Y	93.97%, 90.59%, 87.33%	91.57%, 86.37%, 40.8%
CutMix N	97.73%, 90.84%, 88.27%	96.94%, 86.55%, 42.81%