NAVER AI's Research: Towards Strong and Robust Deep Models

Sangdoo Yun @ Naver AI Lab

7 Feb, 2023

Biography

- Sangdoo Yun
- https://sangdooyun.github.io/
- Academia
  - 2006~2010: SNU ECE B.S.
  - 2011~2013: SNU ECE M.S.
  - 2013~2017: SNU ECE Ph.D.
  - 2021~now: Adjunct professor @ SNU AI
- Industry
  - 2018~now: Research scientist @ Naver AI Lab

    Leading Research group @ Naver AI Lab

Research interests

- (Almost) everything about vision models

Research interests

- (Almost) everything about vision models
- Model architecture: ReXNet[CVPR'21], PiT[ICCV'21]
- Optimizer: AdamP[ICLR'21]
- Robustness: ReBias[ICML'20], Shortcut learning[ICLR'22]
- Vision applications: Face, OCR[CVPR'19,ICCV'19,ECCV'22]

Skipped for this talk

[CVPR'19] Baek et al., Character Region Awareness for Text Detection
[ICCV'19] Baek et al., What Is Wrong with Scene Text Recognition Model Comparisons? Dataset and Model Analysis
[ICML'20] Bahng et al., Learning De-biased Representations with Biased Representations
[CVPR'21] Han et al., Rethinking Channel Dimensions for Efficient Model Design
[ICCV'21] Heo et al., Rethinking spatial dimensions of vision transformers
[ICLR'21] Heo et al., AdamP: Slowing Down the Slowdown for Momentum Optimizers on Scale-invariant Weights
[ECCV'22] Kim et al., Donut: Document Understanding Transformer without OCR
[ICLR'22] Scimeca et al., Which shortcut cues will dnns choose? a study from the parameter-space perspective

Research interests

- (Almost) everything about vision models
- How to “teach” vision models? ***data*** and ***supervision***
    - **Knowledge** distillation [ICCV’19a]
    - **Data** augmentation [ICCV’19b, CVPR’22a]
    - **Data** re-labeling [CVPR’21]
    - **Data** compression [ICML’22a, ICML’22b]
    - Weak **supervision** [ICCV’21, CVPR’22b]

[ICCV’19a] Heo et al., A Comprehensive Overhaul of Feature Distillation
[ICCV’19b] Yun et al., CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features
[CVPR’21] Yun et al., Re-labeling ImageNet: from Single to Multi-Labels, from Global to Localized Labels
[ICCV’21] Kim et al., Normalization Matters in Weakly Supervised Object Localization
[CVPR’22a] Park et al., The Majority Can Help The Minority: Context-rich Minority Oversampling for Long-tailed Classification
[CVPR’22b] Lee et al., Weakly Supervised Semantic Segmentation using Out-of-Distribution Data
[ICML’22a] Kim et al., Dataset Condensation via Efficient Synthetic-Data Parameterization
[ICML’22b] Lee et al., Dataset Condensation with Contrastive Signals

Research interests

- (Almost) everything about vision models
- How to “teach” vision models? ***data*** and ***supervision***
  - **Knowledge** distillation [ICCV’19a]
  - **Data** augmentation [ICCV’19b, CVPR’22a]
  - **Data** re-labeling [CVPR’21]
  - **Data** compression [ICML’22a, ICML’22b]
  - Weak **supervision** [ICCV’21, CVPR’22b]

[ICCV’19a] Heo et al., A Comprehensive Overhaul of Feature Distillation
[ICCV’19b] Yun et al., CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features
[CVPR’21] Yun et al., Re-labeling ImageNet: from Single to Multi-Labels, from Global to Localized Labels
[ICCV’21] Kim et al., Normalization Matters in Weakly Supervised Object Localization
[CVPR’22a] Park et al., The Majority Can Help The Minority: Context-rich Minority Oversampling for Long-tailed Classification
[CVPR’22b] Lee et al., Weakly Supervised Semantic Segmentation using Out-of-Distribution Data
[ICML’22a] Kim et al., Dataset Condensation via Efficient Synthetic-Data Parameterization
[ICML’22b] Lee et al., Dataset Condensation with Contrastive Signals

Towards Strong and Robust Deep Models

Our Vision models for NAVER Services

- We supply stronger vision models (than public ones) for NAVER services

OCR Service
Text detection
Text recognition

Face Service
Face detection
Face identification

•••

Spam Image Filtering
Image Retrieval
etc

How to get stronger models?

- Simple answers:

How to get stronger models?

- Simple answers:
- 1) Collect more pre-training datasets
- 2) Use computationally heavier architecture
- They are not cost efficient.

How to get stronger models?

- Simple answers:
- 1) Collect more pre-training datasets
- 2) Use computationally heavier architecture
- They are not cost efficient.

- Our research goal: obtain better model without extra cost!

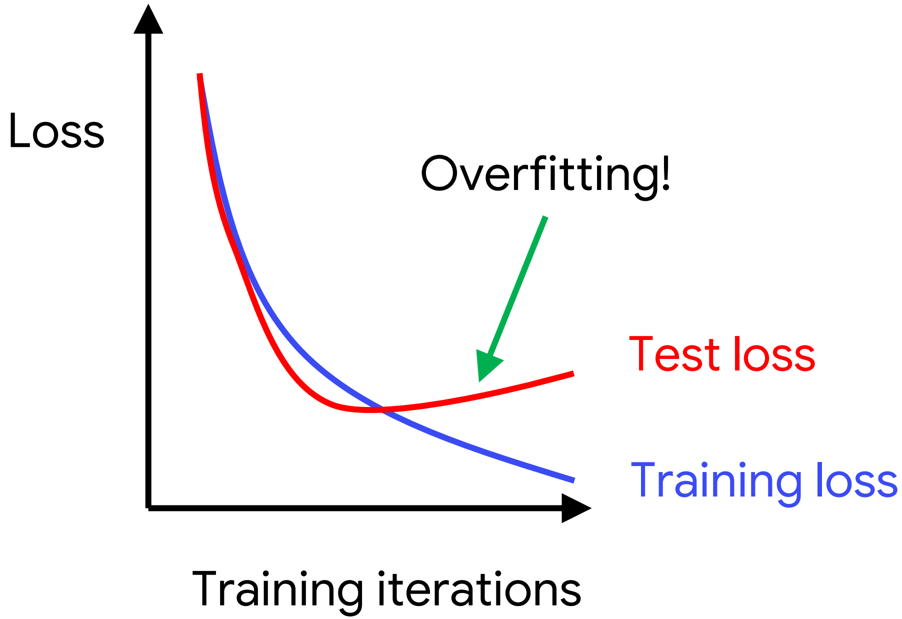
How to get stronger models?

- Simple answers:
- 1) ~~Collect more pre-training datasets~~
- ~~2) Use computationally heavier architecture~~
- They are not cost efficient.

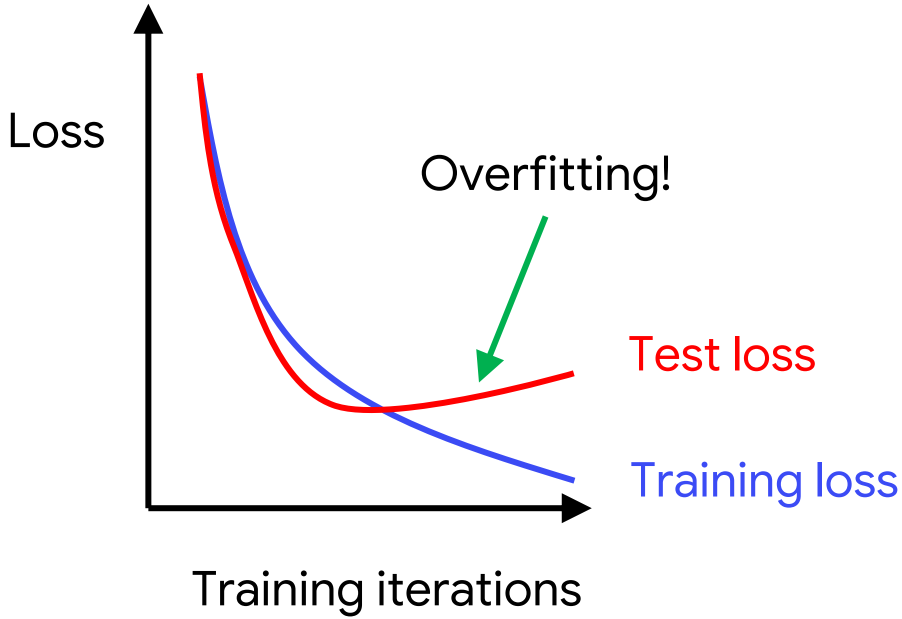
→ We use **better training strategy**

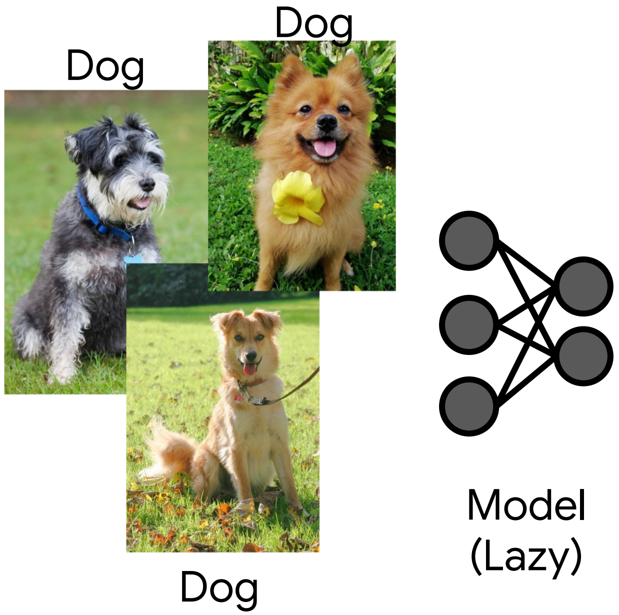
- Our research goal: obtain better model without extra cost!

Why training strategy matters?

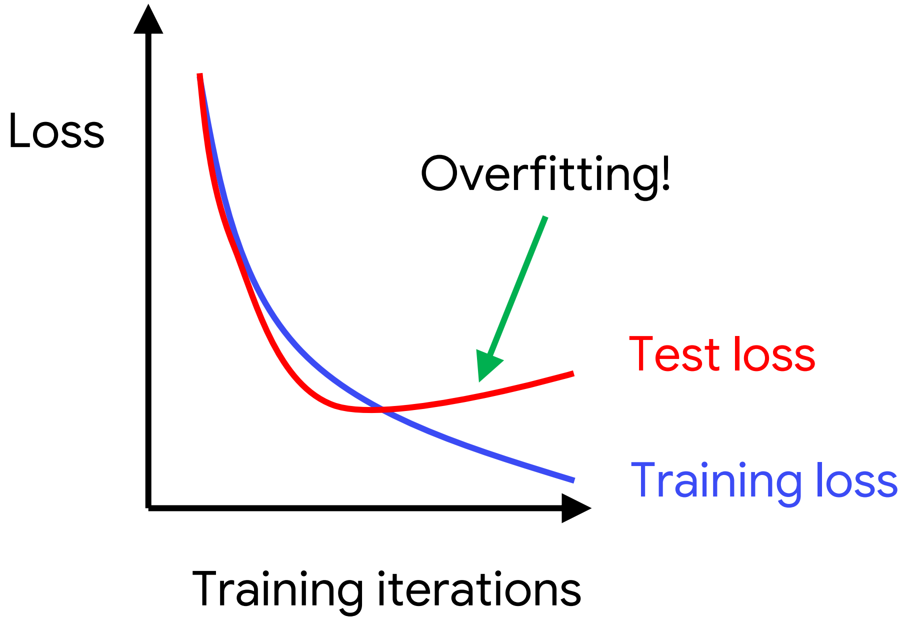


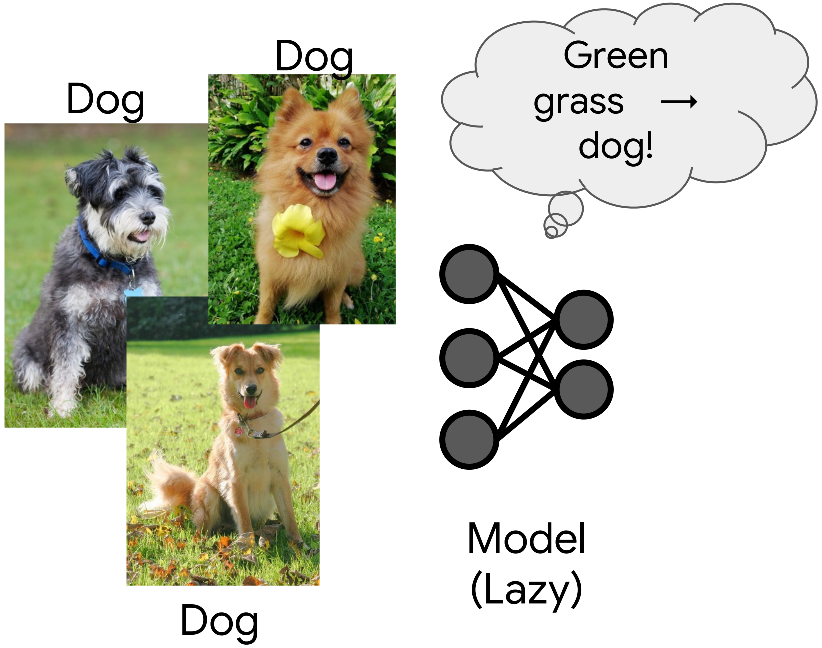
Why training strategy matters?



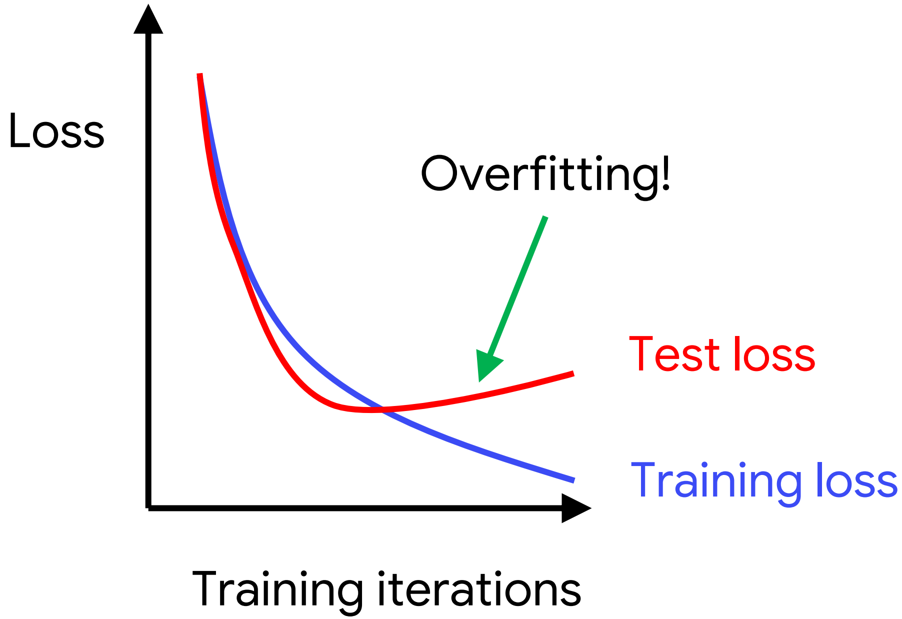


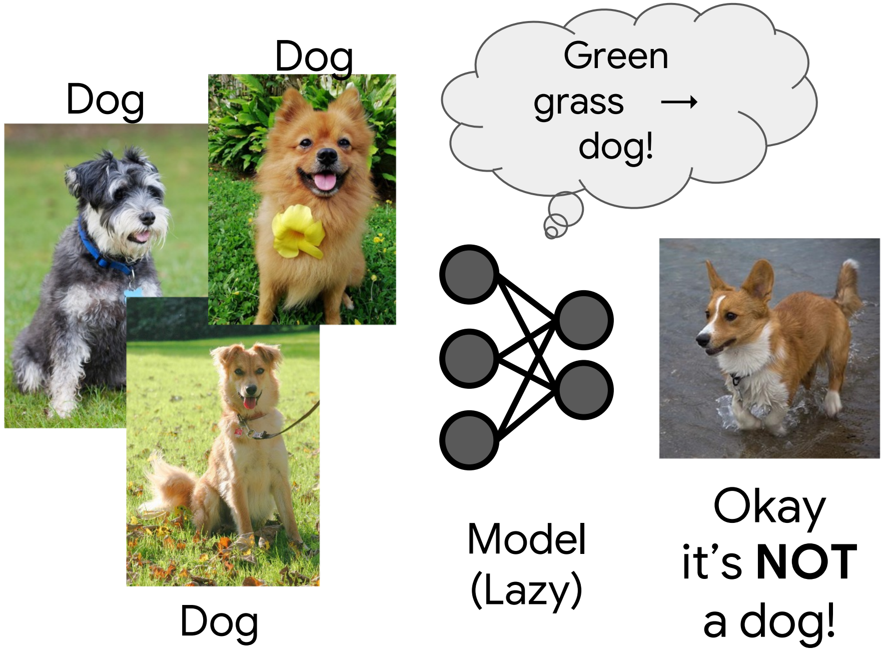
Why training strategy matters?





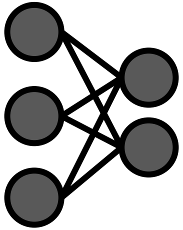
Why training strategy matters?





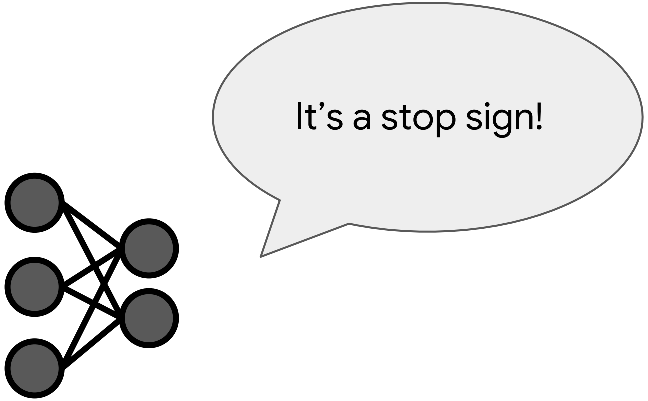
Why training strategy matters?

- Robustness



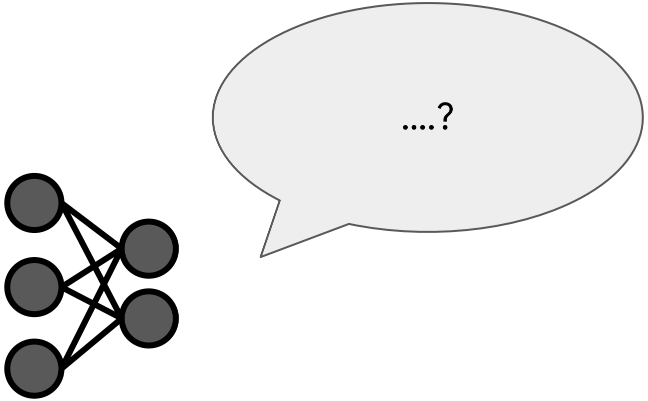
Why training strategy matters?

- Robustness

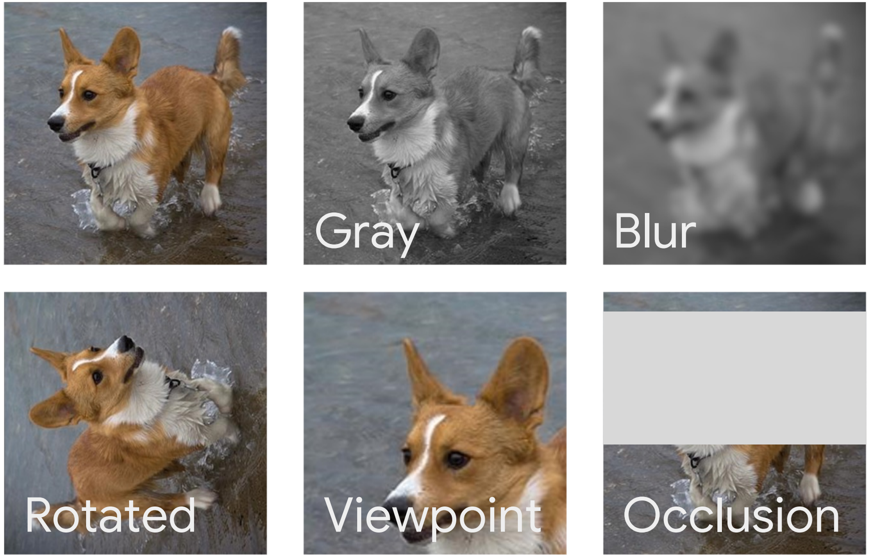


Why training strategy matters?

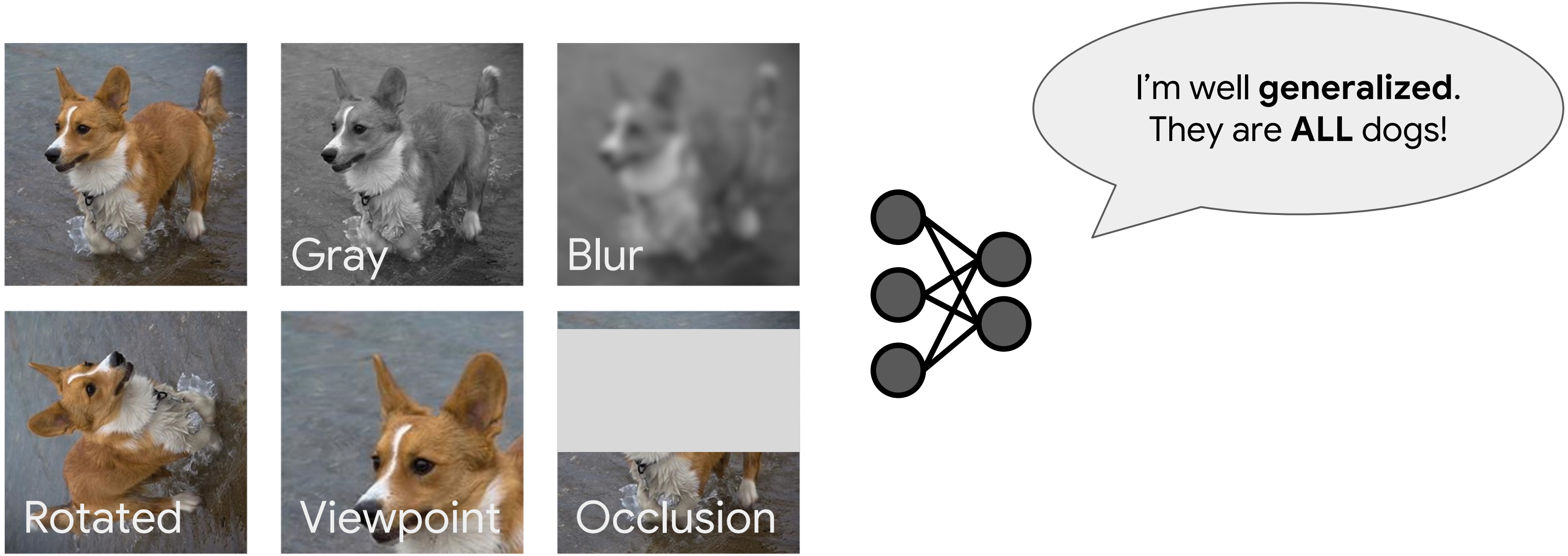
- Robustness



“Generalization”

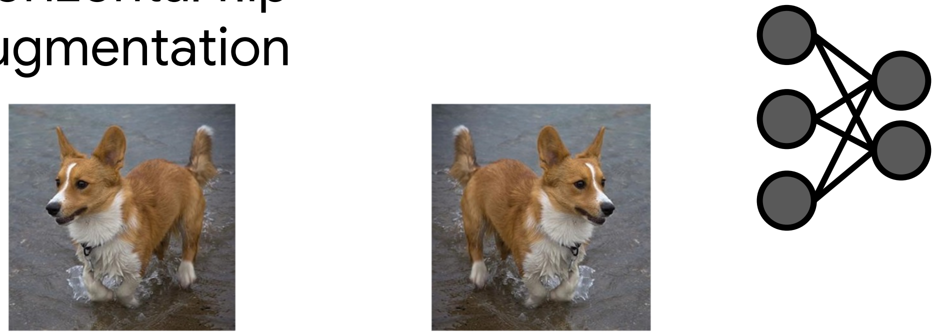


“Generalization”



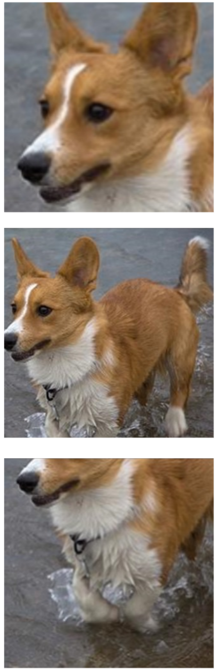
Data Augmentation: Simple but very effective solution

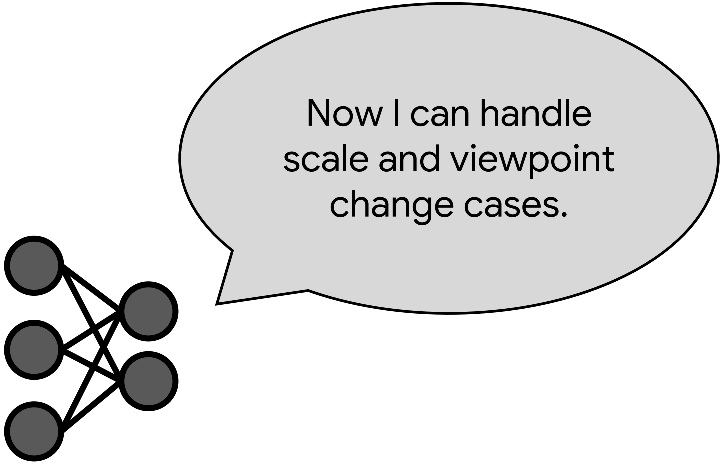
Horizontal flip
augmentation



Data Augmentation: Simple but very effective solution

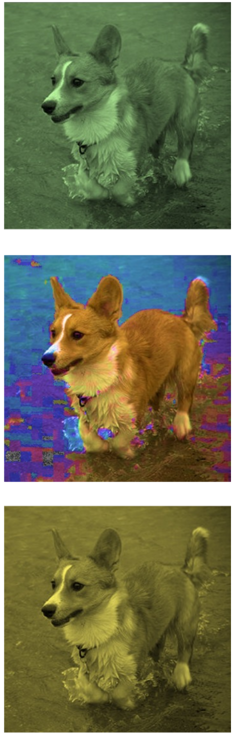
Random Crop augmentation

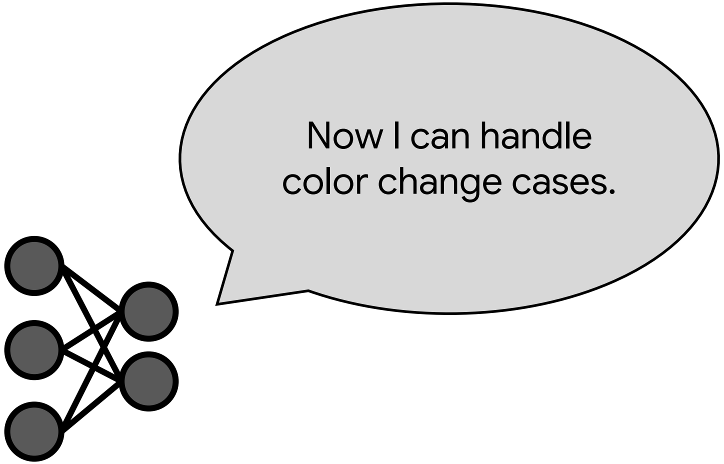




Data Augmentation: Simple but very effective solution

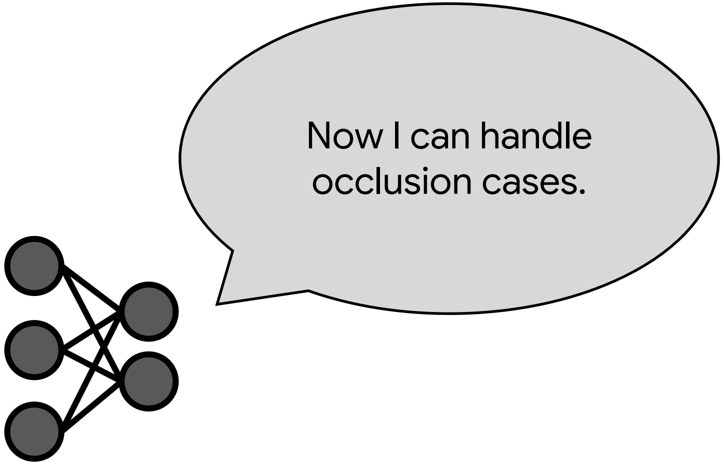
Color jittering / lighting augmentation





Data Augmentation: Simple but very effective solution

Random erasing[1],
Cutout[2]



✓ Good generalization ability

✗ Cannot utilize full image regions

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

Data Augmentation: Mixup

Mixup[1]
data augmentation

Dog

Cat

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

Data Augmentation: Mixup

Mixup[1]
data augmentation

Dog

Cat

Mix

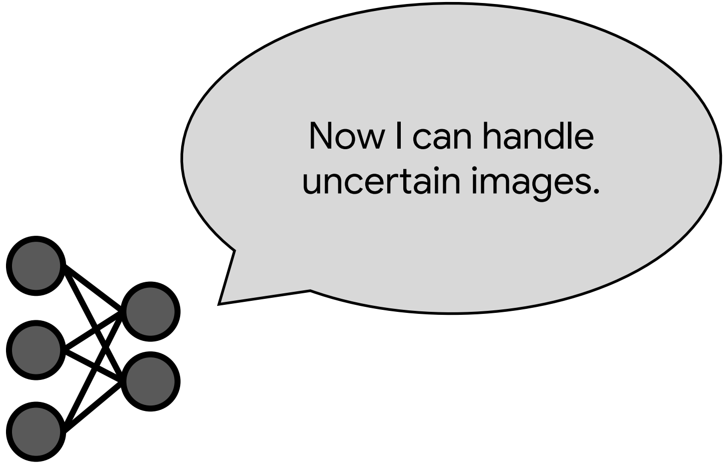
New image

Dog 50%
Cat 50%

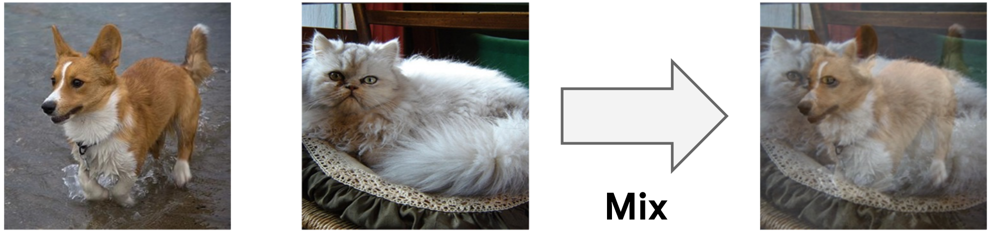
New label

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

Data Augmentation: Mixup



Mixup[1]
data augmentation



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

Data Augmentation: Mixup

✓ Good generalization ability

✓ Use full image region

✗ Locally unrealistic image

Mixup[1]
data augmentation

Dog

Cat

Mix

New image

Dog 50%
Cat 50%

New label

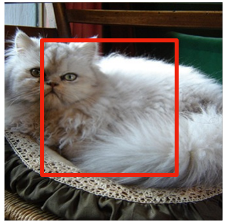
[1] Zhang et al., "mixup: Beyond empirical risk minimization.", ICLR 2018.

How about this?

Cat 1.0

Dog 1.0

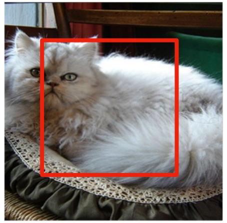
How about this?



Cat 1.0

Dog 1.0

How about this?



Cat 1.0

Dog 1.0

How about this?

New image

Dog 50%
Cat 50%

New label

How about this?

- We call this “**CutMix**”

New image

Dog 50%
Cat 50%

New label

New image

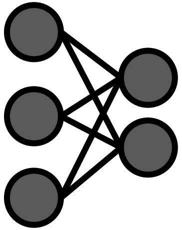
Dog 75%
Cat 25%

New label

...

How about this?

- We call this “**CutMix**”



Now I can handle both **occlusion** and **uncertain** cases.

New image

Dog 50%
Cat 50%

New label

New image

Dog 75%
Cat 25%

New label

...

ICCV’19 Oral Talk.

CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features.






Sangdoo Yun
Naver AI Lab

Dongyoon Han
Naver AI Lab

Seong Joon Oh
Naver AI Lab
(Univ. Tübingen)

Sanghyuk Chun
Naver AI Lab

Junsuk Choe
Naver AI Lab
(Sogang Univ.)

Youngjoon Yoo
Naver AI Lab

CutMix in a Nutshell

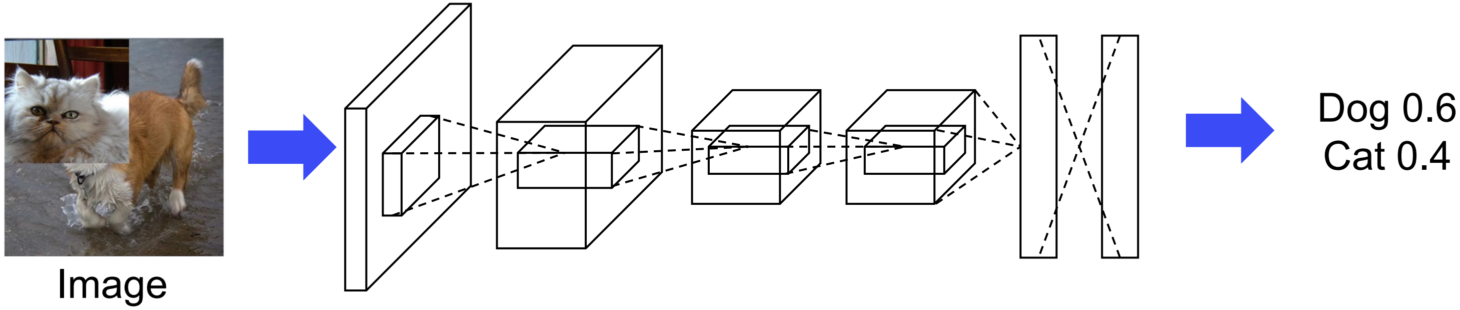
	Original	Cutout	Mixup	CutMix
Training Image				
Target Label	Dog 1.0	Dog 1.0	Dog 0.5 Cat 0.5	Dog 0.6 Cat 0.4

CutMix in a Nutshell

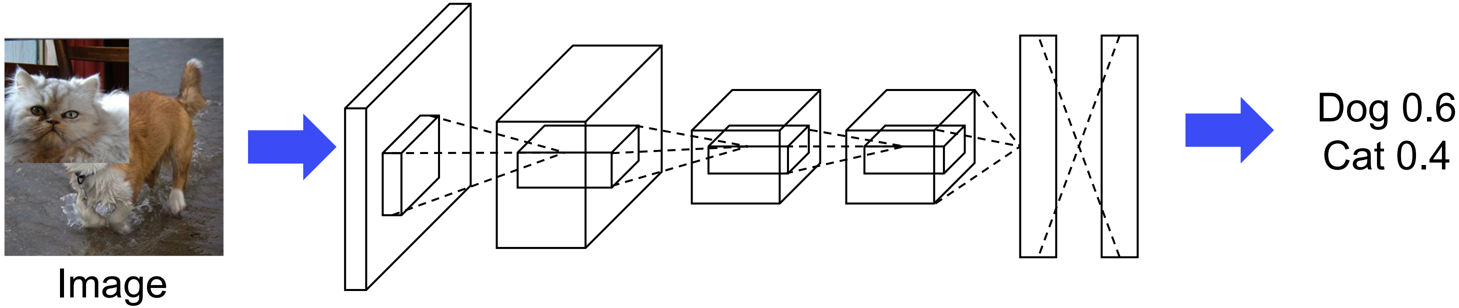
	Original	Cutout	Mixup	CutMix
Training Image				
Target Label	Dog 1.0	Dog 1.0	Dog 0.5 Cat 0.5	Dog 0.6 Cat 0.4

- ✓ Unlike Cutout, **CutMix uses full image region**
- ✓ Unlike Mixup, **CutMix makes realistic local image patches**
- ✓ **CutMix is simple**: only 20 lines of PyTorch code

CutMix training strategy



CutMix training strategy



The problem is changed from “image classification”
→ “**What**”, “**Where**”, and “**How large**” the objects are in the image.

There is a dog and a cat.

The cat is in the upper-left.
The dog is in the remaining region.

Dog with 60%
and cat with 40%

What does the model learn with CutMix?

Heatmap visualization*[1]*:
Where does the model recognize the object?

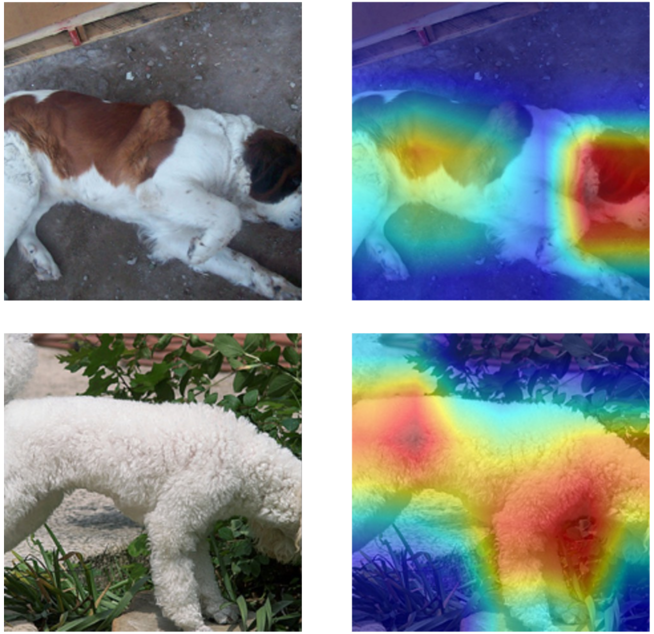
[1] Zhou et al., Learning Deep Features for Discriminative Localization, CVPR 2016.

What does the model learn with CutMix?

Heatmap visualization*[1]*:
Where does the model recognize the object?

Heatmap of St. Bernard

Heatmap of Poodle



[1] Zhou et al., Learning Deep Features for Discriminative Localization, CVPR 2016.

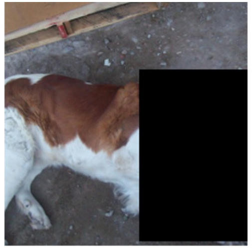
What does the model learn with CutMix?

St. Bernard

Poodle

Mixup[1]

Cutout[2]

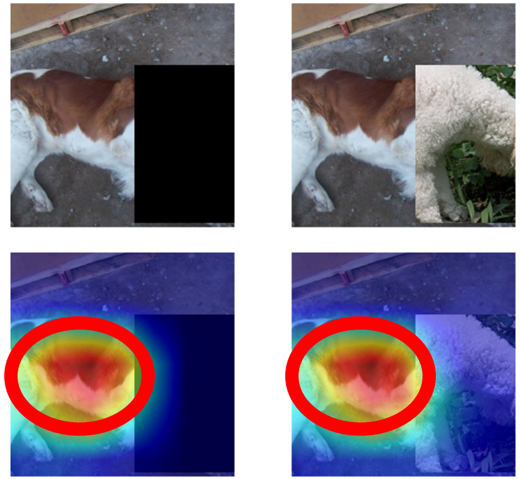


CutMix

[1] Zhang et al., “mixup: Beyond empirical risk minimization.”, ICLR 2018.
[2] Devries et al., “Improved regularization of convolutional neural networks with cutout”, arXiv 2017.

What does the model learn with CutMix?

Cutout[2] CutMix



Heatmap of
St. Bernard

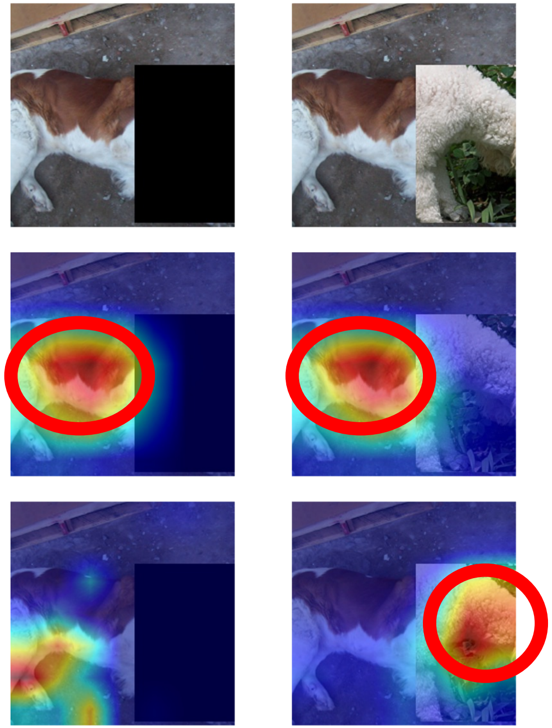
[1] Zhang et al., “mixup: Beyond empirical risk minimization.”, ICLR 2018.
[2] Devries et al., “Improved regularization of convolutional neural networks with cutout”, arXiv 2017.

What does the model learn with CutMix?

Cutout[2] CutMix

Heatmap of St. Bernard

Heatmap of Poodle



[1] Zhang et al., “mixup: Beyond empirical risk minimization.”, ICLR 2018.
[2] Devries et al., “Improved regularization of convolutional neural networks with cutout”, arXiv 2017.

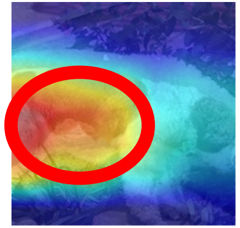
NAVER AI LAB

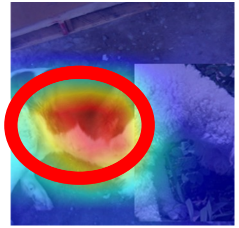
What does the model learn with CutMix?

Mixup[1]

CutMix

Heatmap of St. Bernard





[1] Zhang et al., “mixup: Beyond empirical risk minimization.”, ICLR 2018.
[2] Devries et al., “Improved regularization of convolutional neural networks with cutout”, arXiv 2017.

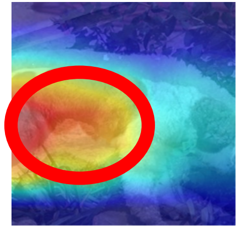
NAVER AI LAB

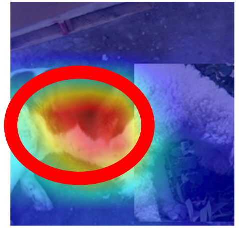
What does the model learn with CutMix?

Mixup[1]

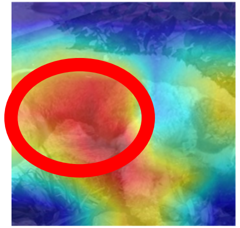
CutMix

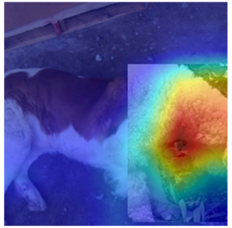
Heatmap of St. Bernard





Heatmap of Poodle





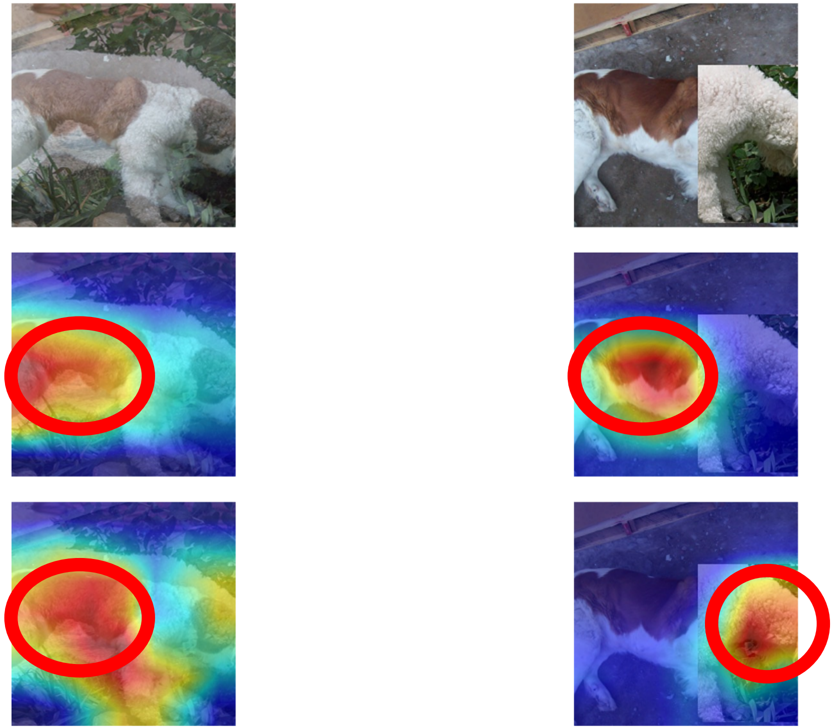
[1] Zhang et al., “mixup: Beyond empirical risk minimization.”, ICLR 2018.
[2] Devries et al., “Improved regularization of convolutional neural networks with cutout”, arXiv 2017.

What does the model learn with CutMix?

Mixup[1] | CutMix

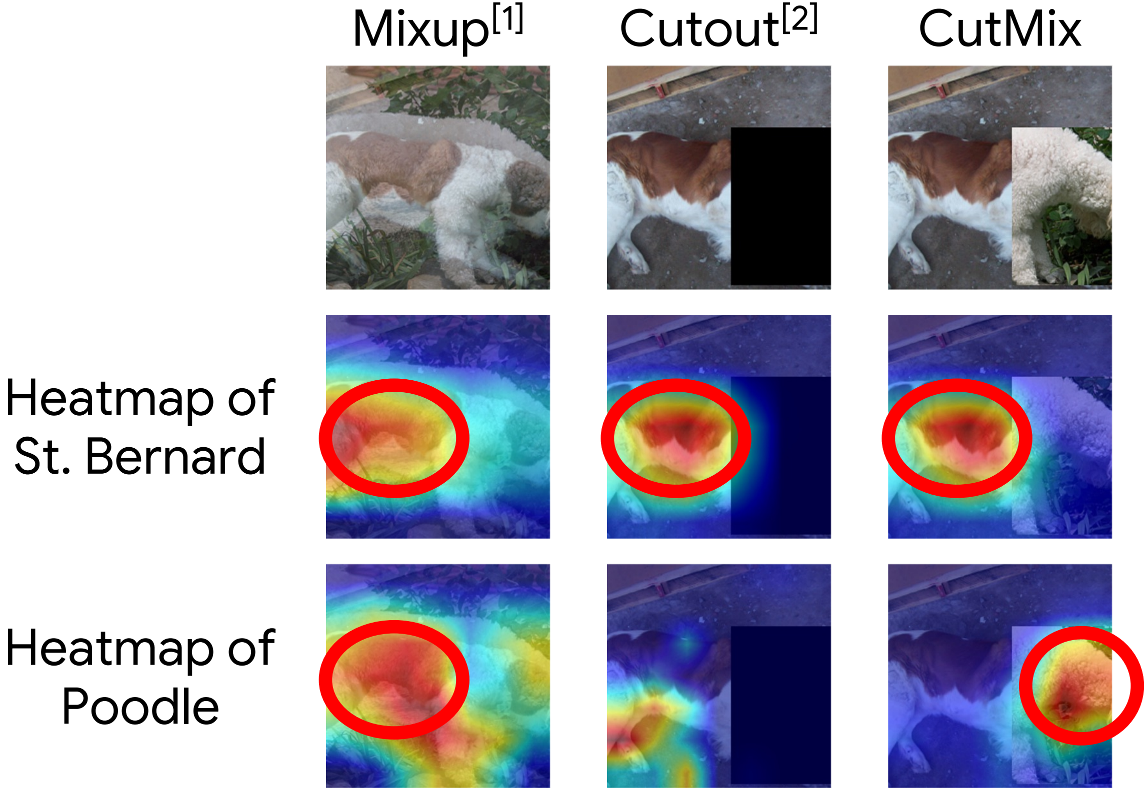
Heatmap of St. Bernard

Heatmap of Poodle



[1] Zhang et al., “mixup: Beyond empirical risk minimization.”, ICLR 2018.
[2] Devries et al., “Improved regularization of convolutional neural networks with cutout”, arXiv 2017.

What does the model learn with CutMix?



[1] Zhang et al., “mixup: Beyond empirical risk minimization.”, ICLR 2018.
[2] Devries et al., “Improved regularization of convolutional neural networks with cutout”, arXiv 2017.

Experiments

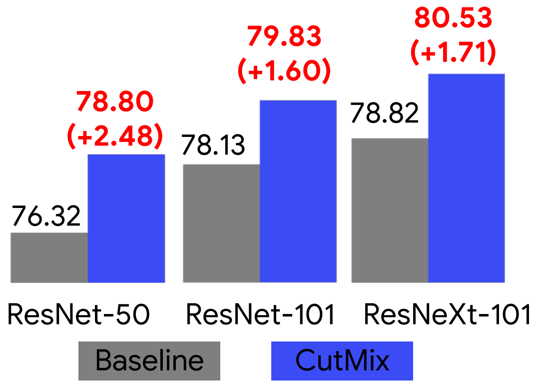
- ImageNet Classification

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

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



- ✓ Great improvement over baseline (+2%p).
- ✓ Outperforming existing methods.
- ✓ ResNet50 + CutMix ≈ ResNet152.

Experiments

- Transfer learning to ***object detection*** and ***image captioning.***

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)

✓ Great improvement on MS-COCO (+2%p): ResNet-50 → ResNet-101

✓ Choosing ***CutMix-pretrained*** model brings great performance gain

Experiments

- Improved robustness performance

	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)

Is CutMix still useful?

TITLE	CITED BY	YEAR
Cutmix: Regularization strategy to train strong classifiers with localizable features S Yun, D Han, SJ Oh, S Chun, J Choe, Y Yoo Proceedings of the IEEE/CVF international conference on computer vision …	2209	2019

A ConvNet for the 2020s

Zhuang Liu[1,2*] Hanzi Mao[1] Chao-Yuan Wu[1] Christoph Feichtenhofer[1] Trevor Darrell[2] Saining Xie[1†]
[1]Facebook AI Research (FAIR) [2]UC Berkeley
Code: https://github.com/facebookresearch/ConvNeXt

epochs from the original 90 epochs for ResNets. We use the AdamW optimizer [46], data augmentation techniques such as Mixup [90], Cutmix [89], RandAugment [14], Random Erasing [91], and regularization schemes including Stochastic Depth [36] and Label Smoothing [69]. The complete set

ConvNext
CVPR 2022

Training data-efficient image transformers & distillation through attention

Hugo Touvron[*,†] Matthieu Cord[†] Matthijs Douze[*]
Francisco Massa[*] Alexandre Sablayrolles[*] Hervé Jégou[*]
[*]Facebook AI [†]Sorbonne University

were first adopted in the training procedure by Wightman [55]. Regularization like Mixup [60] and Cutmix [59] improve performance. We also use repeated

Vision Transformer
ICML 2021

Is CutMix still useful? Yes :)

TITLE	CITED BY	YEAR
Cutmix: Regularization strategy to train strong classifiers with localizable features S Yun, D Han, SJ Oh, S Chun, J Choe, Y Yoo Proceedings of the IEEE/CVF international conference on computer vision …	2209	2019

A ConvNet for the 2020s

Zhuang Liu[1,2*] Hanzi Mao[1] Chao-Yuan Wu[1] Christoph Feichtenhofer[1] Trevor Darrell[2] Saining Xie[1†]
[1]Facebook AI Research (FAIR) [2]UC Berkeley
Code: https://github.com/facebookresearch/ConvNeXt

epochs from the original 90 epochs for ResNets. We use the AdamW optimizer [46], data augmentation techniques such as Mixup [90], Cutmix [89], RandAugment [14], Random Erasing [91], and regularization schemes including Stochastic Depth [36] and Label Smoothing [69]. The complete set

ConvNext
CVPR 2022

Training data-efficient image transformers & distillation through attention

Hugo Touvron[*,†] Matthieu Cord[†] Matthijs Douze[*]
Francisco Massa[*] Alexandre Sablayrolles[*] Hervé Jégou[*]

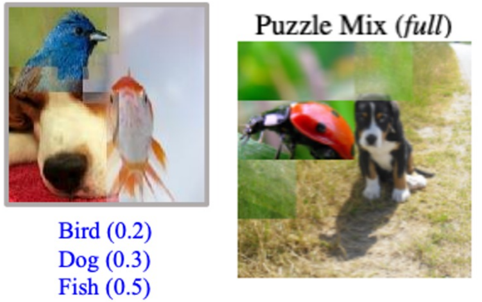
[*]Facebook AI [†]Sorbonne University

were first adopted in the training procedure by Wightman [55]. Regularization like Mixup [60] and Cutmix [59] improve performance. We also use repeated

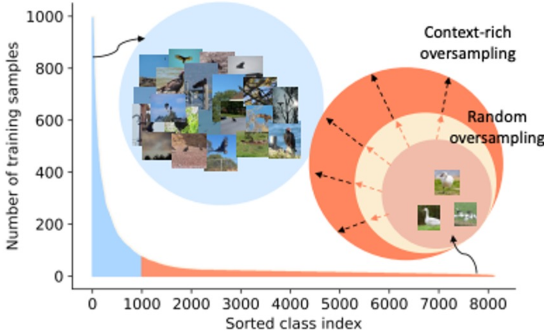
Vision Transformer
ICML 2021

Further studies

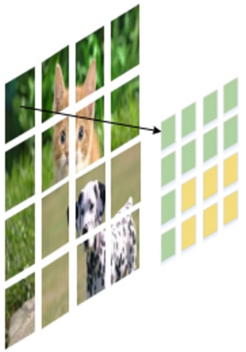
Video recognition [1]



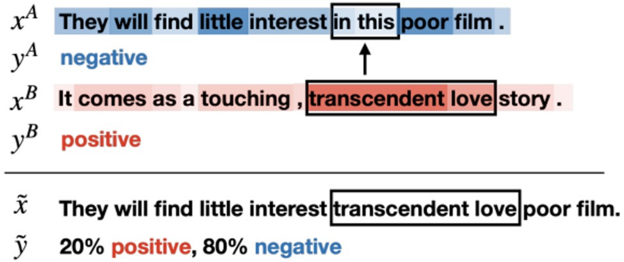
Better mixing strategy beyond random [2,3]



CutMix for imbalanced classification [4]



Token augmentation in ViT [5]



Text classification in NLP [6]

[1] Yun et al., VideoMix: Rethinking Data Augmentation for Video Classification, arXiv 2021
[2] Kim et al., Puzzle Mix: Exploiting Saliency and Local Statistics for Optimal Mixup, ICML 2020
[3] Kim et al., Co-Mixup: Saliency Guided Joint Mixup with Supermodular Diversity, ICLR 2021
[4] Park et al., The Majority Can Help The Minority: Context-rich Minority Oversampling for Long-tailed Classification, CVPR 2022.
[5] Jiang et al., All Tokens Matter: Token Labeling for Training Better Vision Transformers, NuerIPS 2021.
[6] Yoon et al., SSMix: Saliency-Based Span Mixup for Text Classification, Findings of ACL 2021.

Summary of CutMix

- CutMix makes robust and strong vision models
- Visit our website (codes & models): https://github.com/ClovaAI/CutMix-PyTorch

"ImageNet"[1][2] (ILSVRC 2012)

- More than 1M images for 1,000 object categories
- However, their annotations are ...

"monastery"

"Norwich terrier"

"stage"

From ImageNet to Image Classification: Contextualizing Progress on Benchmarks, ICML 2020.
https://slideslive.com/38928533/from-imagenet-to-image-classification-contextualizing-progress-on-benchmarks

[1] J. Deng, et al., ImageNet: A Large-Scale Hierarchical Image Database. CVPR 2009.
[2] Olga Russakovsky et al., ImageNet Large Scale Visual Recognition Challenge. IJCV 2015.

"ImageNet"[1][2] (ILSVRC 2012)

- More than 1M images for 1,000 object categories
- However, their annotations are ...

~~"monastery"~~ "church"

~~"Norwich terrier"~~ "Norfolk terrier"

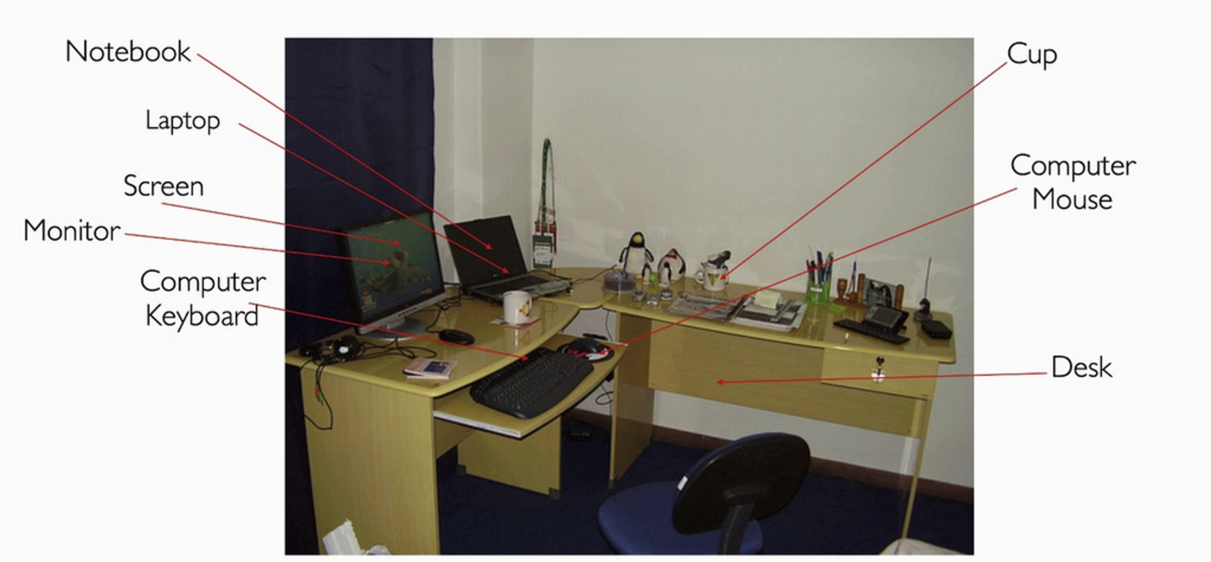
~~"stage"~~ "acoustic guitar"

From ImageNet to Image Classification: Contextualizing Progress on Benchmarks, ICML 2020.
https://slideslive.com/38928533/from-imagenet-to-image-classification-contextualizing-progress-on-benchmarks

[1] J. Deng, et al., ImageNet: A Large-Scale Hierarchical Image Database. CVPR 2009.
[2] Olga Russakovsky et al., ImageNet Large Scale Visual Recognition Challenge. IJCV 2015.

“ImageNet” (ILSVRC 2012)

- More than 1M images for 1,000 object categories
- However, their annotations are ...



ImageNet label: desk

Contextualizing Machine Accuracy on ImageNet, ICML 2020.
https://slideslive.com/38928496/contextualizing-machine-accuracy-on-imagenet

ImageNet’s Labeling issues

- Previous works [1,2,3] focus on **validation set (50,000 images)**
- Re-annotates multi-labels using **human labor**

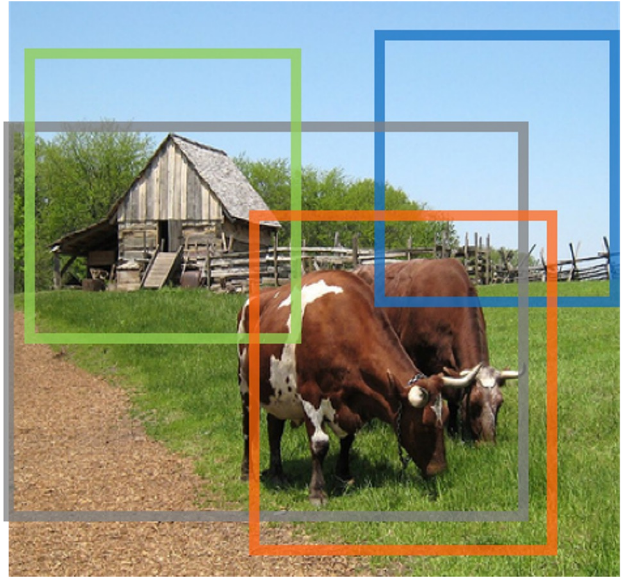
[1] From ImageNet to Image Classification: Contextualizing Progress on Benchmarks, ICML 2020.
[2] Contextualizing Machine Accuracy on ImageNet, ICML 2020.
[3] Are we done with ImageNet?, ArXiv 2020.

ImageNet’s Labeling issues: Training images

- How about training images? (1,280,000 images) - Can we ***re-label*** them?
- If we solve the labeling problems on training images, we might enhance models accuracy and robustness?

ImageNet's Labeling issues: Training images

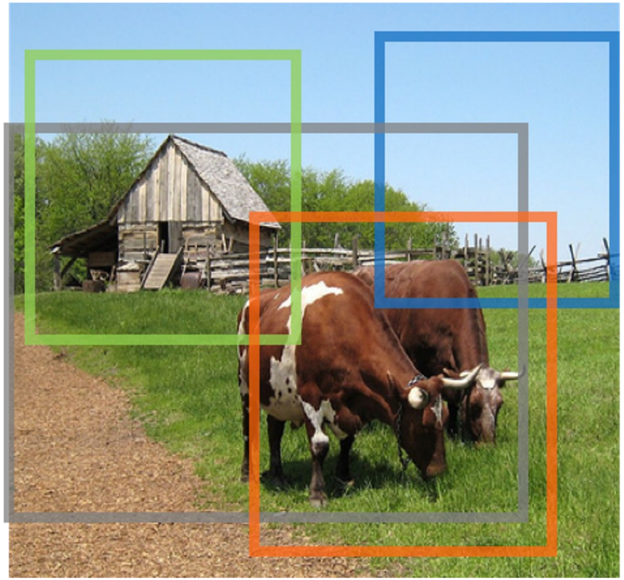
- During training, "Random Crop" intensifies the label noises



ImageNet Label: ox

ImageNet's Labeling issues: Training images

- During training, "Random Crop" intensifies the label noises



ImageNet Label: ox

ox 1.00

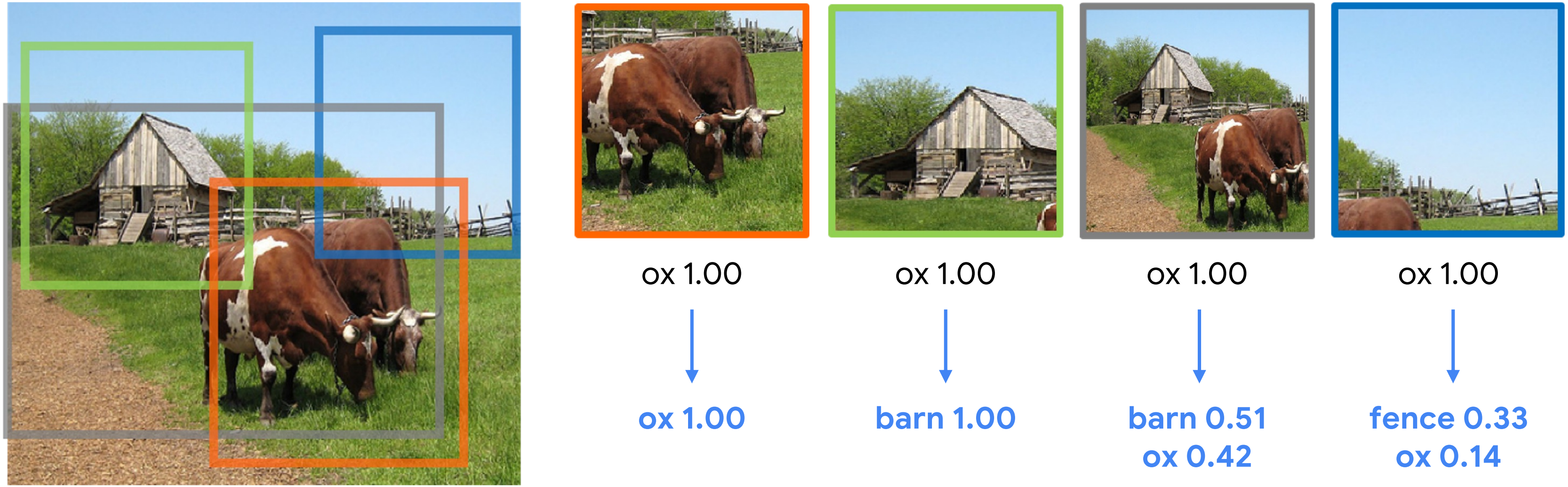
ox 1.00

ox 1.00

ox 1.00

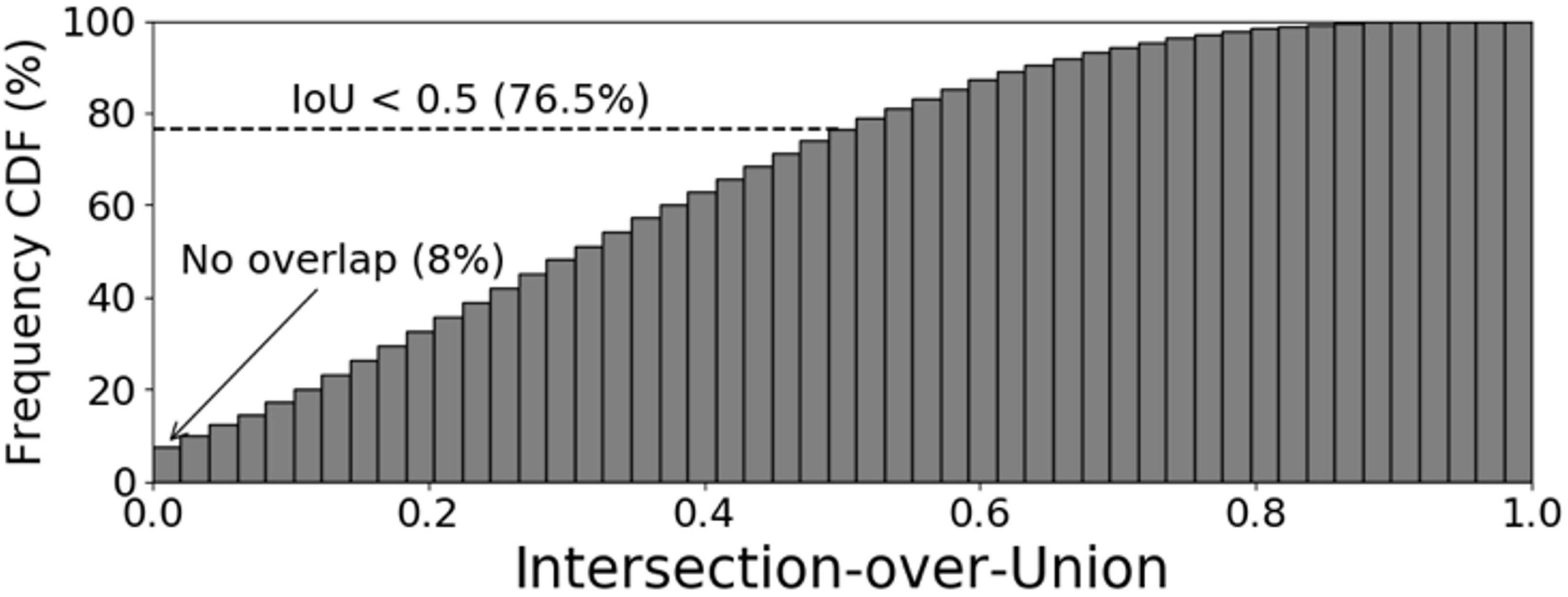
ImageNet's Labeling issues: Training images

- During training, "Random Crop" intensifies the label noises



ImageNet Label: ox

Random crop analysis



CVPR'21

Re-labeling ImageNet: from Single to Multi-Labels, from Global to Localized Labels






Sangdoo Yun
Naver AI Lab

Seong Joon Oh
Naver AI Lab
(Univ. Tübingen)

Byeongho Heo
Naver AI Lab

Dongyoon Han
Naver AI Lab

Junsuk Choe
Naver AI Lab
(Sogang Univ.)

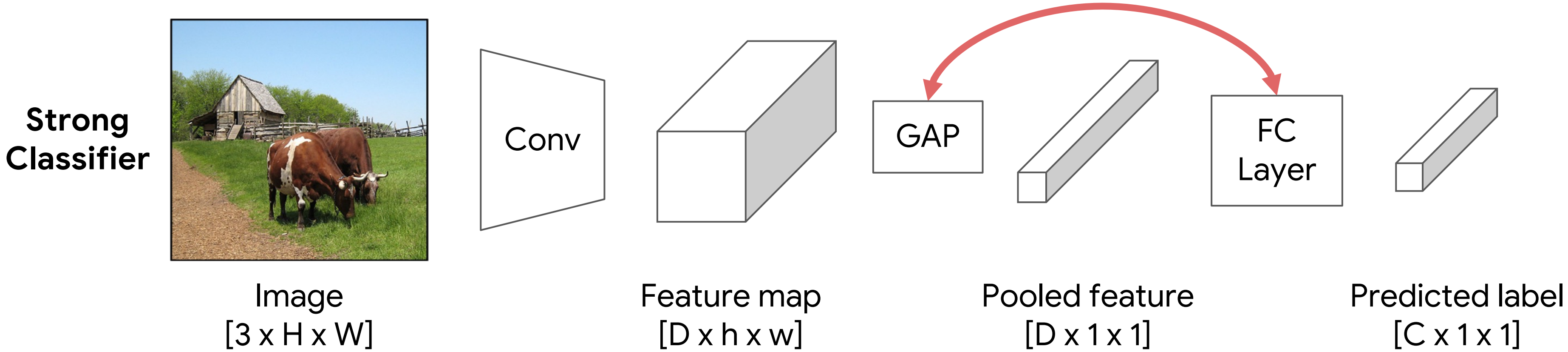
Sanghyuk Chun
Naver AI Lab

Our solution (**ReLabel**)

- Our goal: **(1) Multi-label, (2) Localized label**
- Re-labeling using “machine annotator” (or, pseudo-labeling)
- Machine annotator: state-of-the-art classifier trained with extra source data

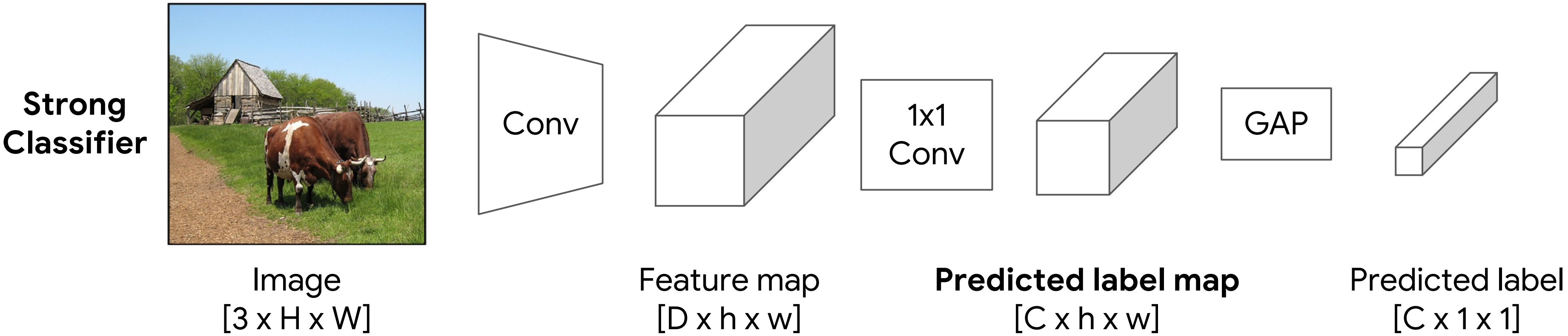
Our solution (**ReLabel**)

- Our goal: **(1) Multi-label, (2) Localized label**
- Re-labeling using “machine annotator” (or, pseudo-labeling)
- Machine annotator: state-of-the-art classifier trained with extra source data



Our solution (**ReLabel**)

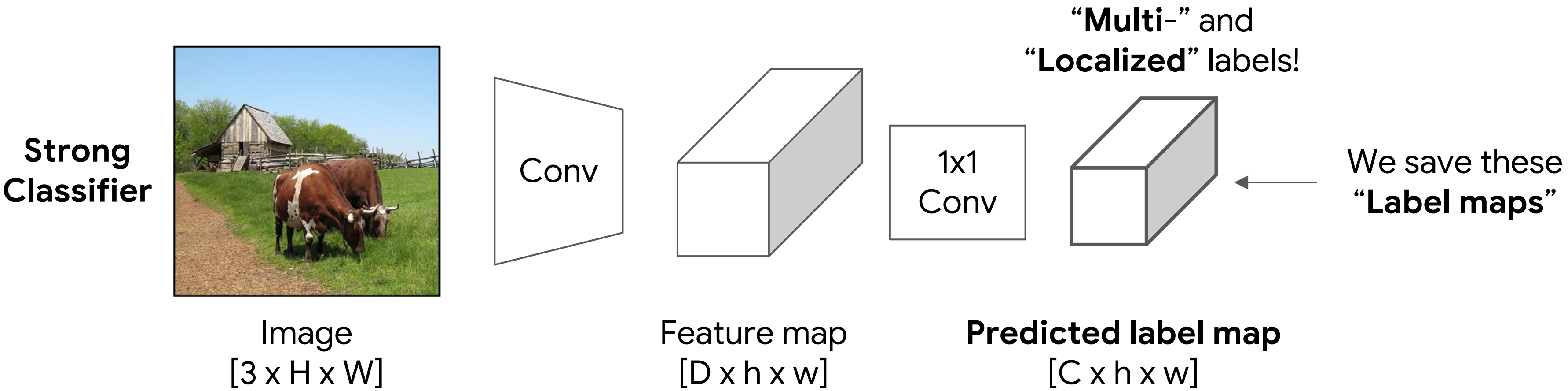
- Our goal: **(1) Multi-label, (2) Localized label**
- Re-labeling using “machine annotator” (or, pseudo-labeling)
- Machine annotator: state-of-the-art classifier trained with extra source data



Long et al., Fully convolutional networks for semantic segmentation. CVPR 2015.
Zhou et al.,Learning deep features for discriminative localization, CVPR 2016.

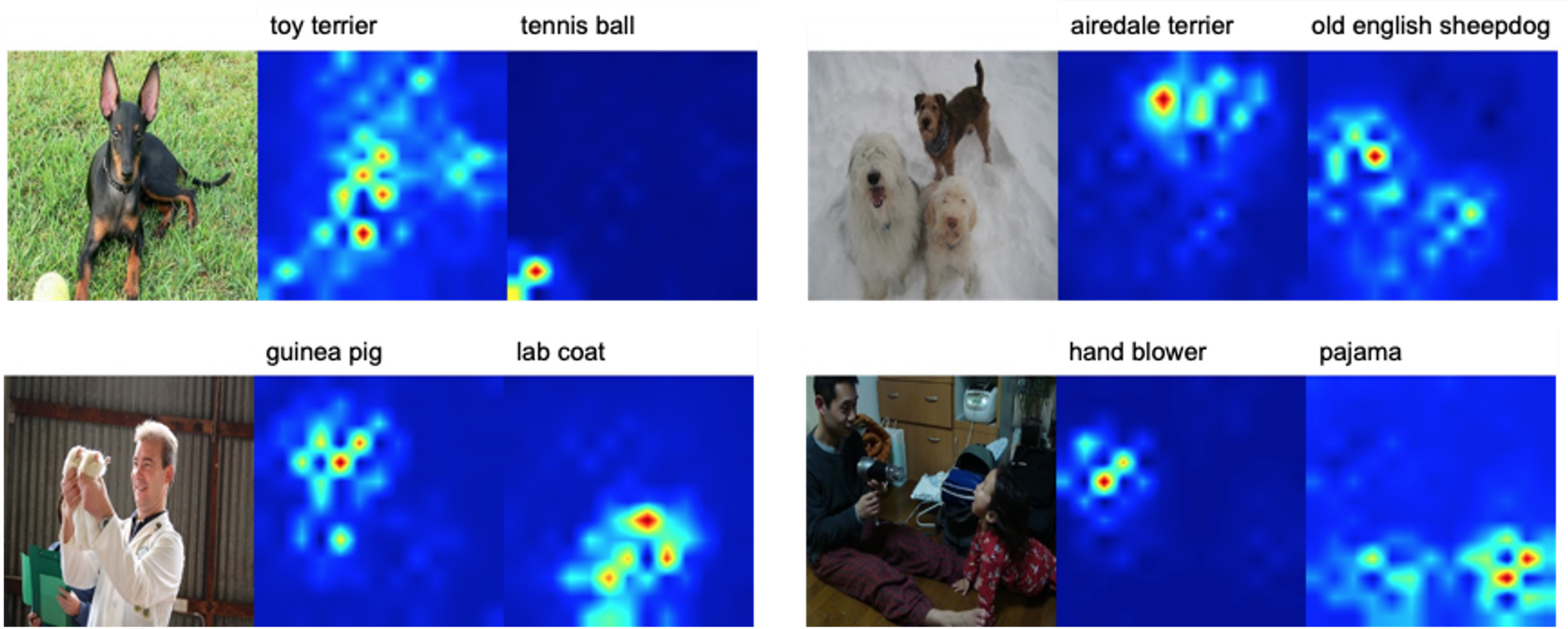
Our solution (**ReLabel**)

- Our goal: **(1) Multi-label, (2) Localized label**
- Re-labeling using “machine annotator” (or, pseudo-labeling)
- Machine annotator: state-of-the-art classifier trained with extra source data



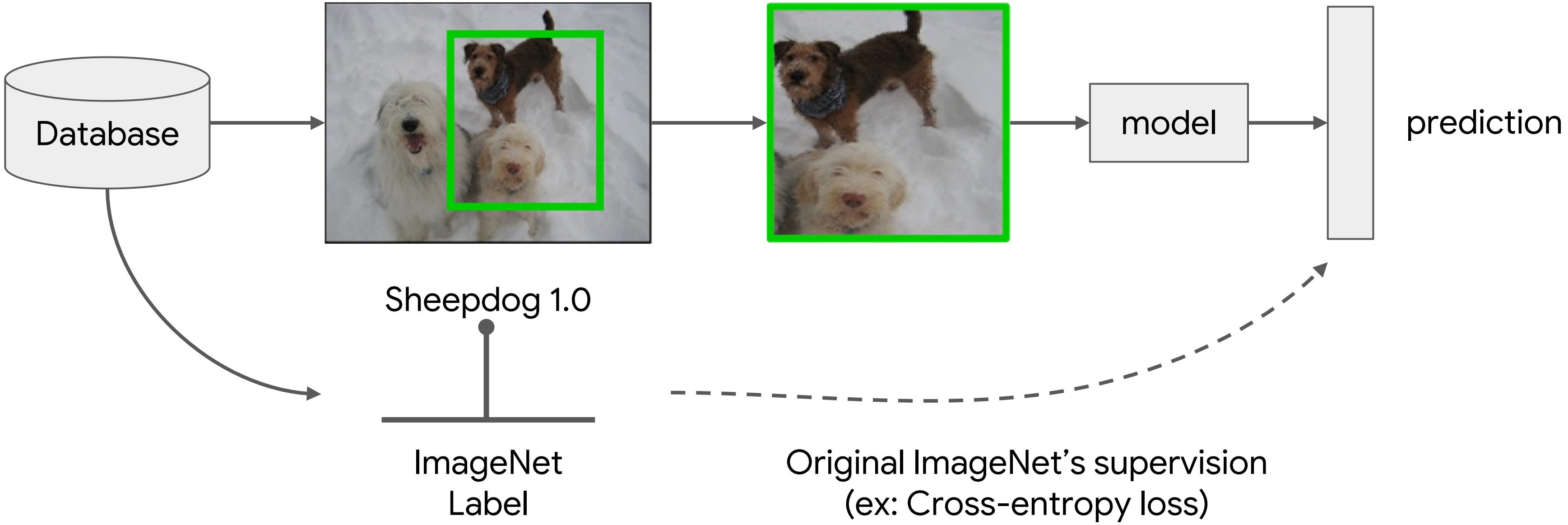
Long et al., Fully convolutional networks for semantic segmentation. CVPR 2015.
Zhou et al.,Learning deep features for discriminative localization, CVPR 2016.

Label Map Examples

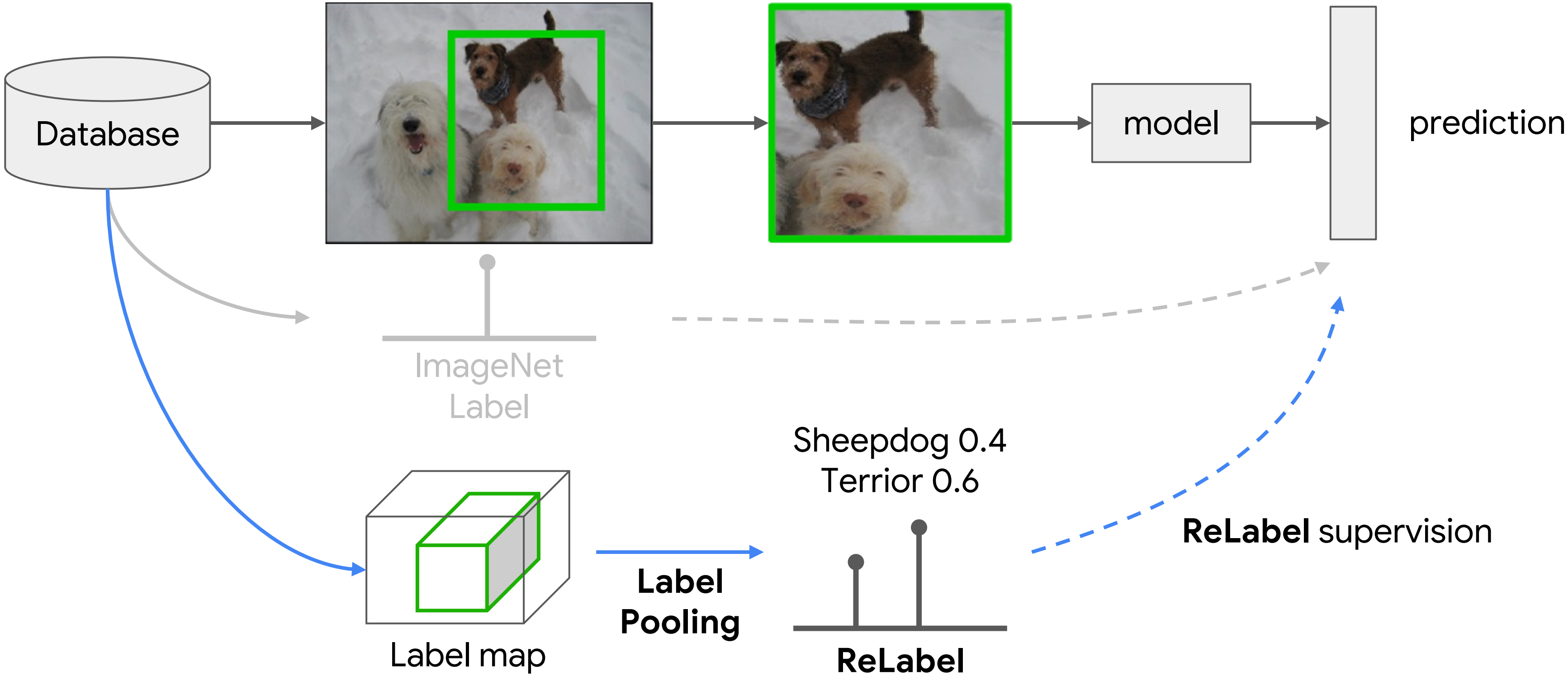


How to train?

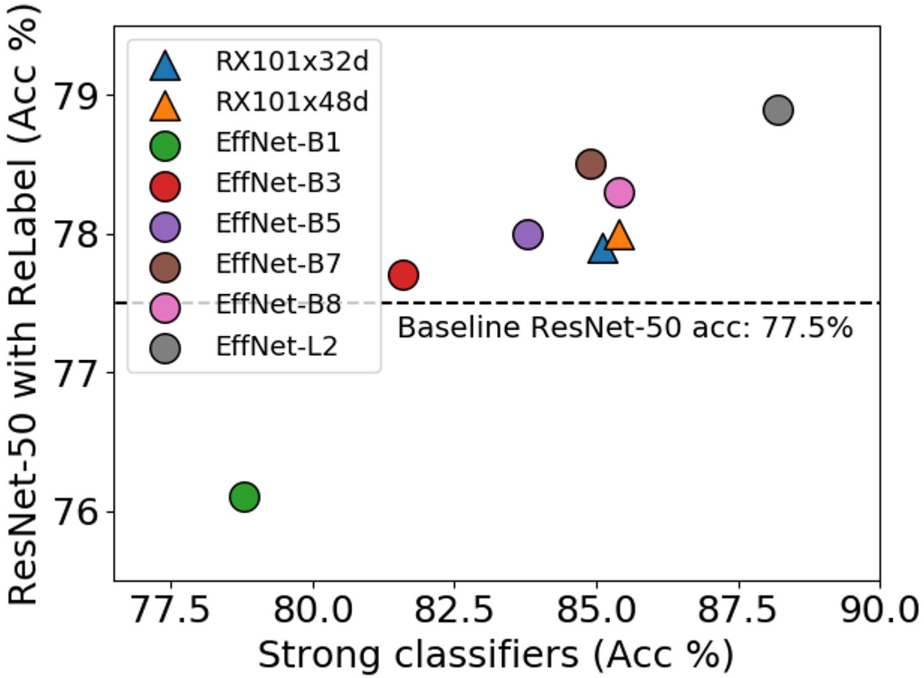
How to train? – Original supervision



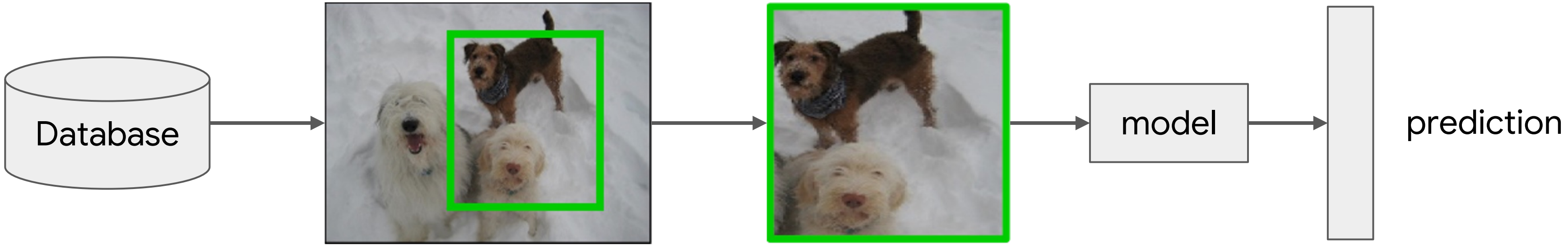
How to train? – **ReLabel** supervision



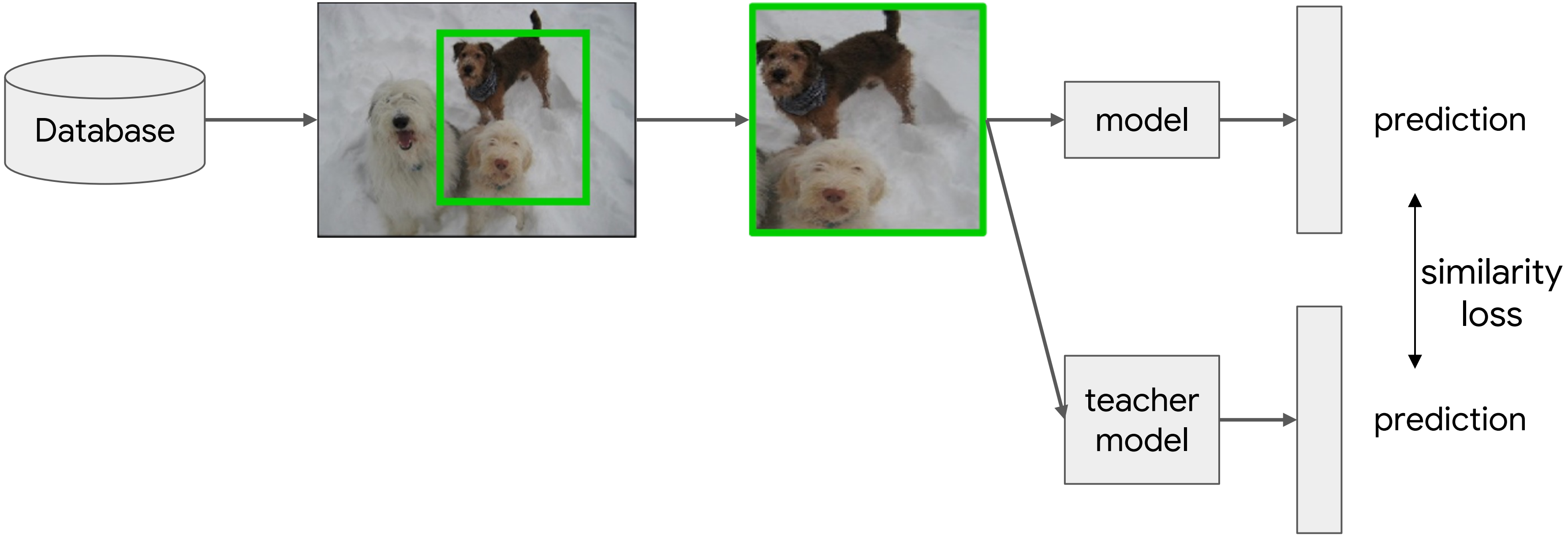
Re-labeling ImageNet: Analysis



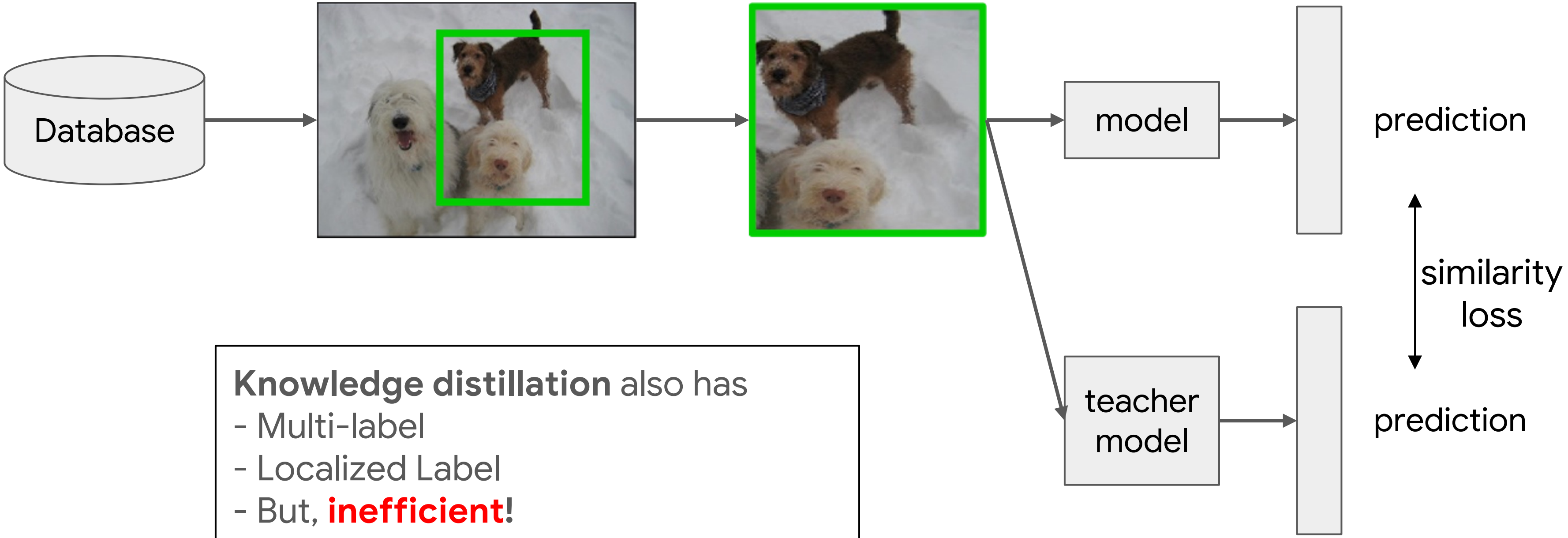
Comparison with **knowledge distillation**



Comparison with **knowledge distillation**



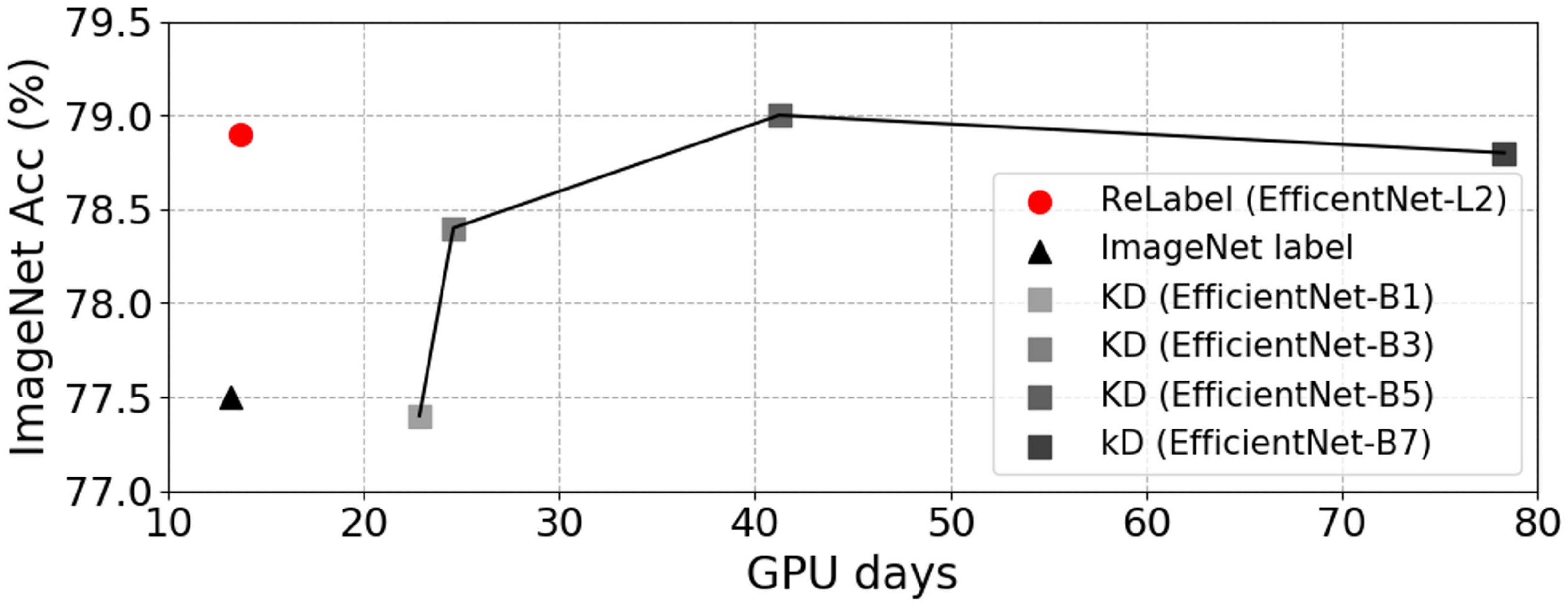
Comparison with **knowledge distillation**

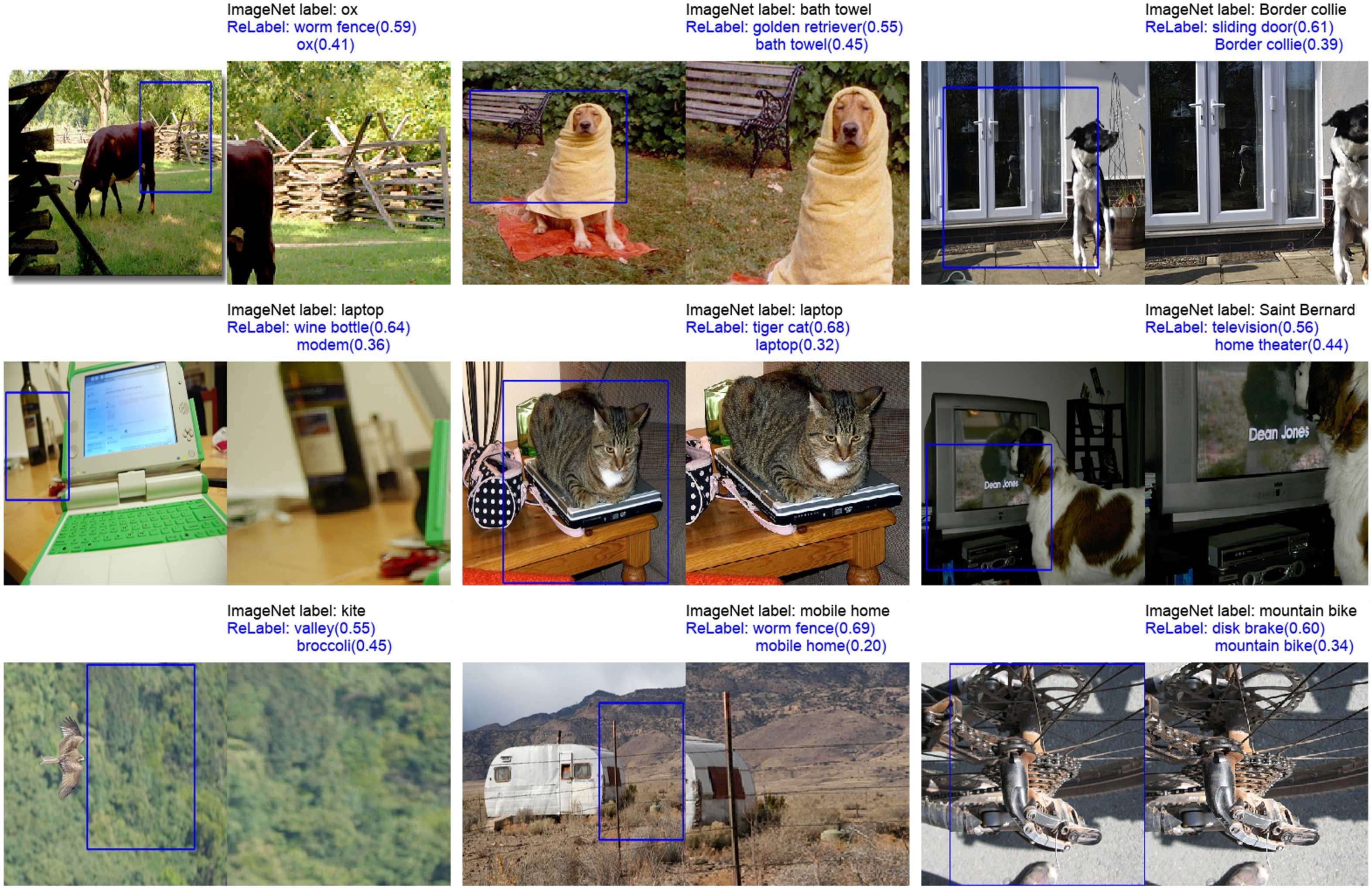


Knowledge distillation also has
- Multi-label
- Localized Label
- But, **inefficient!**

Re-labeling ImageNet: Analysis

- Comparison with knowledge distillation




ImageNet label: ox
ReLabel: worm fence(0.59)
ox(0.41)
ImageNet label: bath towel
ReLabel: golden retriever(0.55)
bath towel(0.45)
ImageNet label: Border collie
ReLabel: sliding door(0.61)
Border collie(0.39)
ImageNet label: laptop
ReLabel: wine bottle(0.64)
modem(0.36)
ImageNet label: laptop
ReLabel: tiger cat(0.68)
laptop(0.32)
ImageNet label: Saint Bernard
ReLabel: television(0.56)
home theater(0.44)
ImageNet label: kite
ReLabel: valley(0.55)
broccoli(0.45)
ImageNet label: mobile home
ReLabel: worm fence(0.69)
mobile home(0.20)
ImageNet label: mountain bike
ReLabel: disk brake(0.60)
mountain bike(0.34)

Experiments

- Results on ImageNet benchmark (single, multi-label evaluation)

Network	Supervision	ImageNet single-label	ImageNetV2 [34] single-label	ReaL [2] multi-label	Shankar *et al*. [37] multi-label
ResNet-50	Original	77.5	79.0	83.6	85.3
ResNet-50	Label smoothing (ϵ=0.1) [43]	78.0	79.5	84.0	84.7
ResNet-50	Label cleaning [2]	78.1	79.1	83.6	85.2
ResNet-50	**ReLabel**	**78.9**	**80.5**	**85.0**	**86.1**

Experiments

- Results on ImageNet benchmark (various architectures)

Architecture	Resources		Supervision	
	Params	Flops	Vanilla	ReLabel
ResNet-18	11.7M	1.8B	71.7	72.5 (+0.8)
ResNet-50	25.6M	3.8B	77.5	78.9 (+1.4)
ResNet-101	44.7M	7.6B	78.1	80.7 (+2.6)
EfficientNet-B0	5.3M	0.4B	77.4	78.0 (+0.6)
EfficientNet-B1	7.8M	0.7B	79.2	80.3 (+1.1)
EfficientNet-B2	9.2M	1.0B	80.3	81.0 (+0.7)
EfficientNet-B3	12.2M	1.8B	81.7	82.5 (+0.8)

Experiments

- Results on ImageNet benchmark (towards SOTA)

Model	ImageNet top1 (%)
ResNet-50	77.5
+ **ReLabel**	78.9 (+1.4)
+ **ReLabel** + CutMix	80.2 (+2.7)
+ **ReLabel** + CutMix + Extra data	81.2 (+3.7)
ResNet-101	78.1
+ **ReLabel**	80.7 (+2.6)
+ **ReLabel** + CutMix	81.6 (+3.5)

Experiments

- Robustness

Models	FGSM	ImageNet-A	ImageNet-C	BCG
ResNet-50	25.7	4.9	27.9	25.9
+ ReLabel	31.3 (+5.6)	7.1 (+2.2)	28.1 (+0.2)	34.6 (+8.7)
+ CutMix	42.4 (+16.7)	11.4 (+6.5)	47.5 (+19.6)	34.1 (+8.2)
+ Extra data	45.0 (+19.3)	24.8 (+19.9)	54.2 (+26.3)	36.0 (+10.1)

Experiments

- Transfer learning

	Food-101 [3]	Stanford Cars [24]	DTD [6]	FGVC Aircraft [31]	Oxford Pets [33]
ResNet-50 (Baseline)	87.98	92.64	75.43	85.09	93.92
ResNet-50 (**ReLabel**-trained)	88.12	92.73	75.74	88.89	94.28

	Faster-RCNN	Mask-RCNN	
	bbox AP	bbox AP	mask AP
ResNet-50 (Baseline)	37.7	38.5	34.7
ResNet-50 (**ReLabel**-trained)	38.2	39.1	35.2

Summary of ReLabel

- We propose a re-labeling strategy, ReLabel for ImageNet training data.
- ReLabel improves the model performance with 3% extra computation.
- Our re-labeled ImageNet, models, and codes: https://github.com/naver-ai/relabel_imagenet.

Thank you