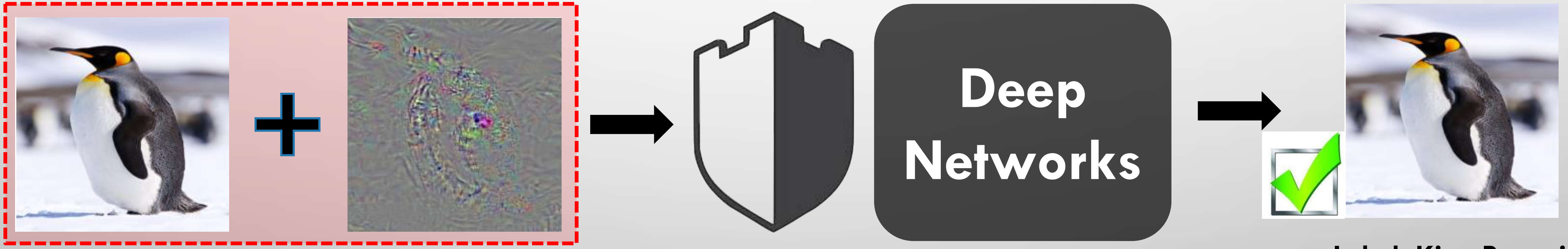
Adversarial Examples IMPROVE Image Recognition

Cihang Xie
Assistant Professor, UC Santa Cruz

Adversarial Examples Are **THREATS** to Deep Networks



Label: King Penguin

Icons made by Freepik from www.flaticon.com

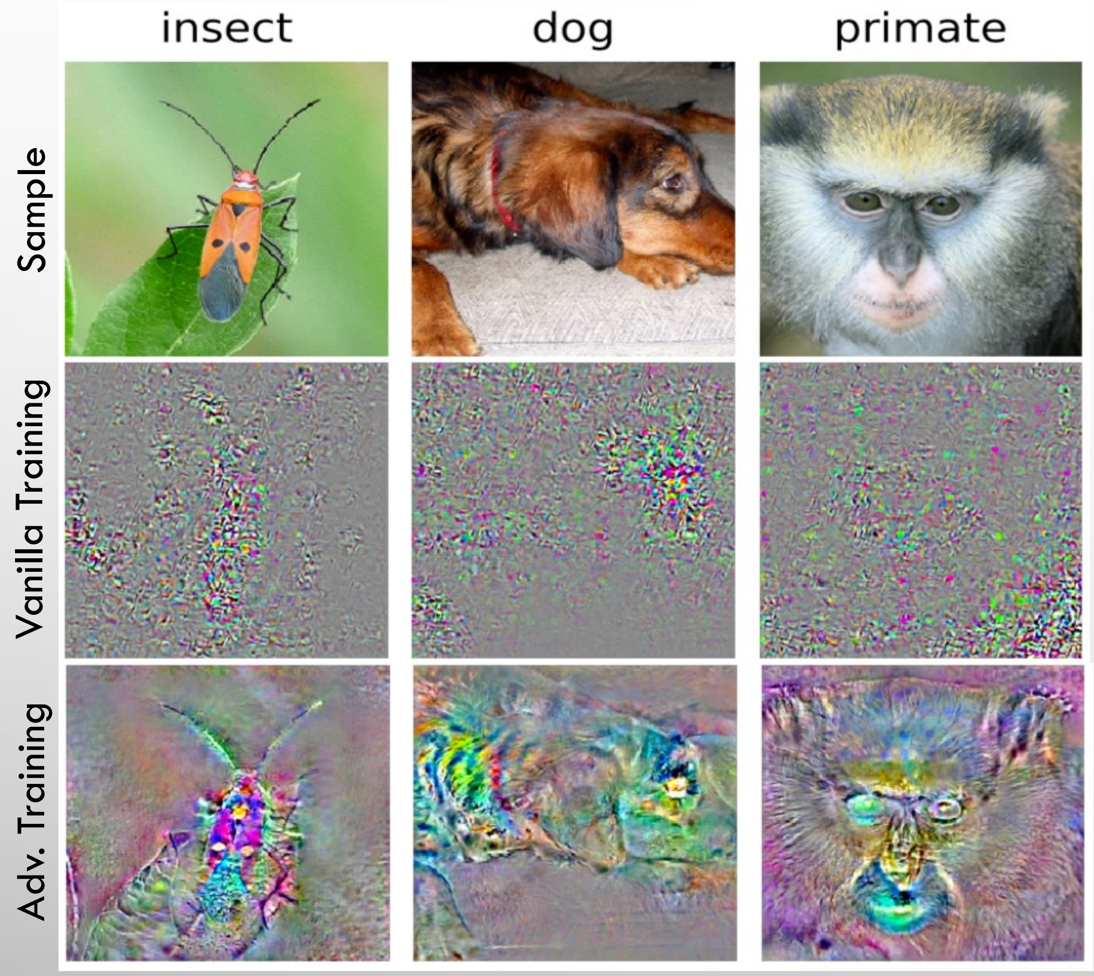
Can we use Adversarial Examples to **HELP** Deep Networks?

Label: King Penguin

Robust Learning Improves Generalization

- **Motivation**

Adversarial examples provide **VALUABLE & NEW** features



The loss gradient w.r.t. the input pixel of adversarially trained models is

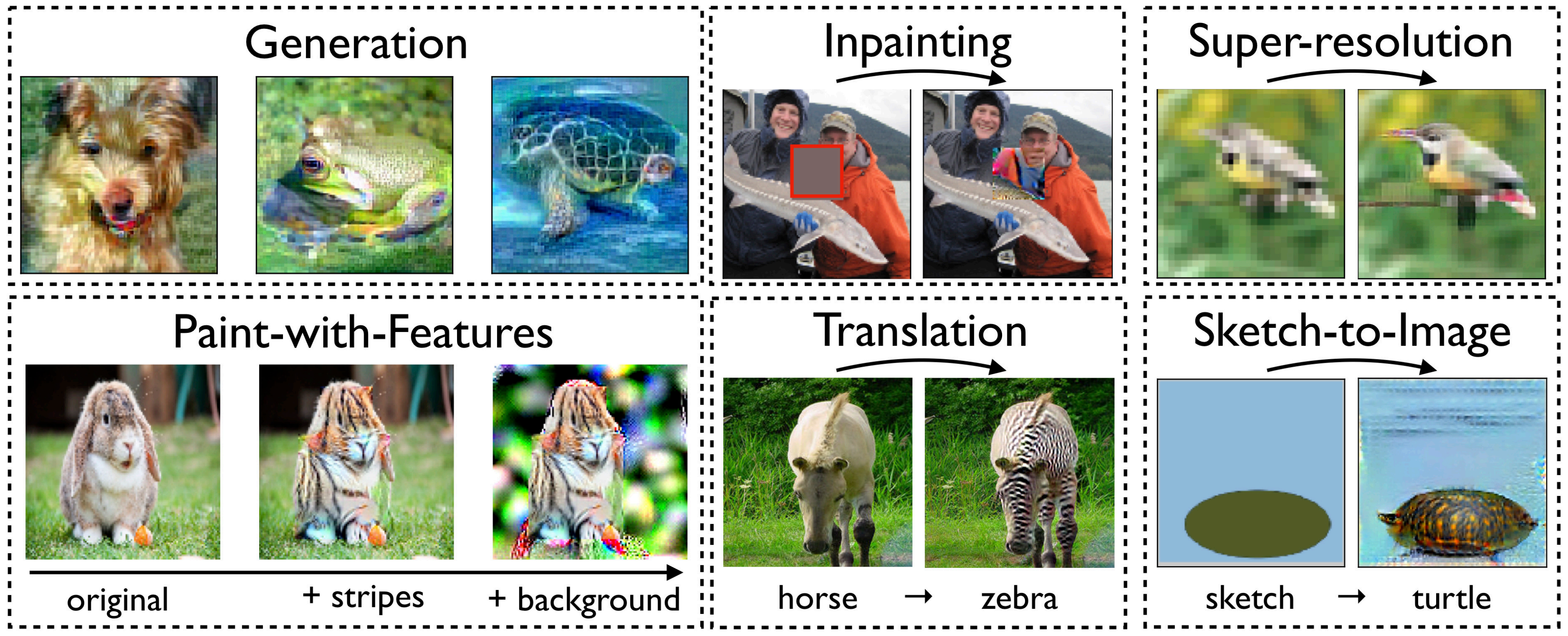
HUMAN-ALIGNED

[Tsipras et al. 2019]

Robust Learning Improves Generalization

➢ **Motivation**

Adversarial examples provide **VALUABLE & NEW** features



Adversarially trained models are pretty good at tackle

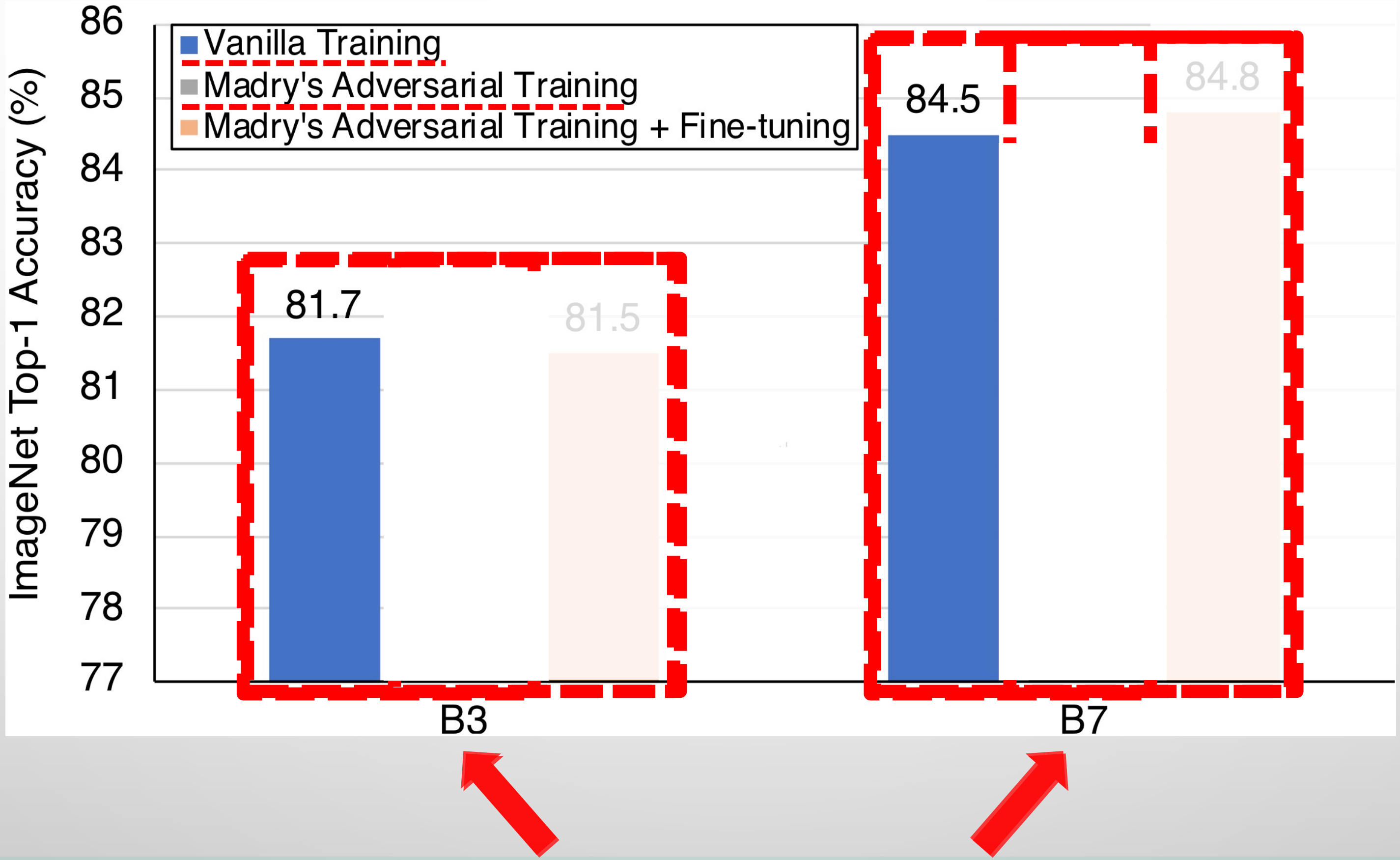
IMAGE SYNTHESIS TASKS

[Santurkar et al. 2019]

Robust Learning Improves Generalization

- **Motivation**

Using features from adversarial examples **ALONE** are **NOT ENOUGH**



Training **EXCLUSIVELY** on adversarial examples

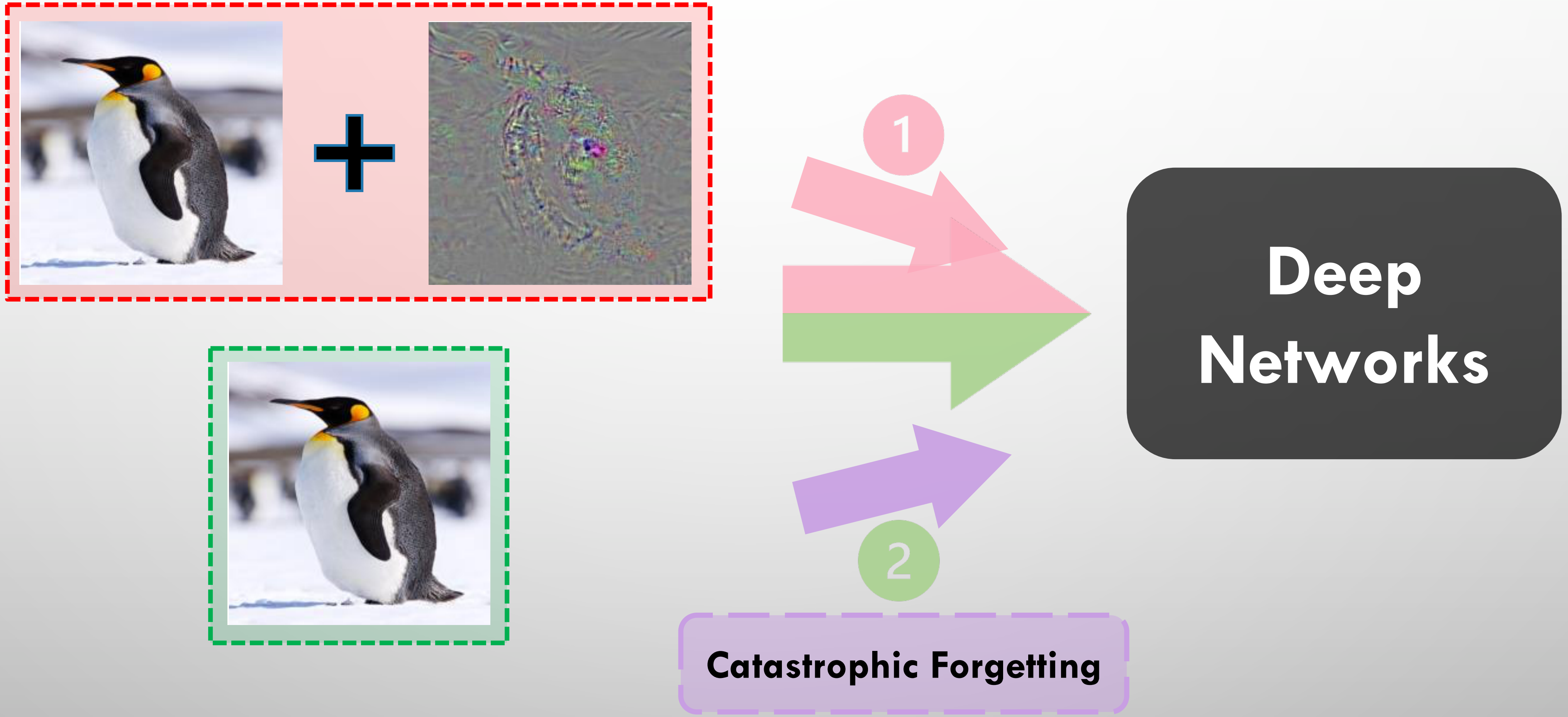
Simply **FINETUNING** with clean images **IMPROVES** performance on clean images

Can we do **BETTER?**

Robust Learning Improves Generalization

- **Our Solution**

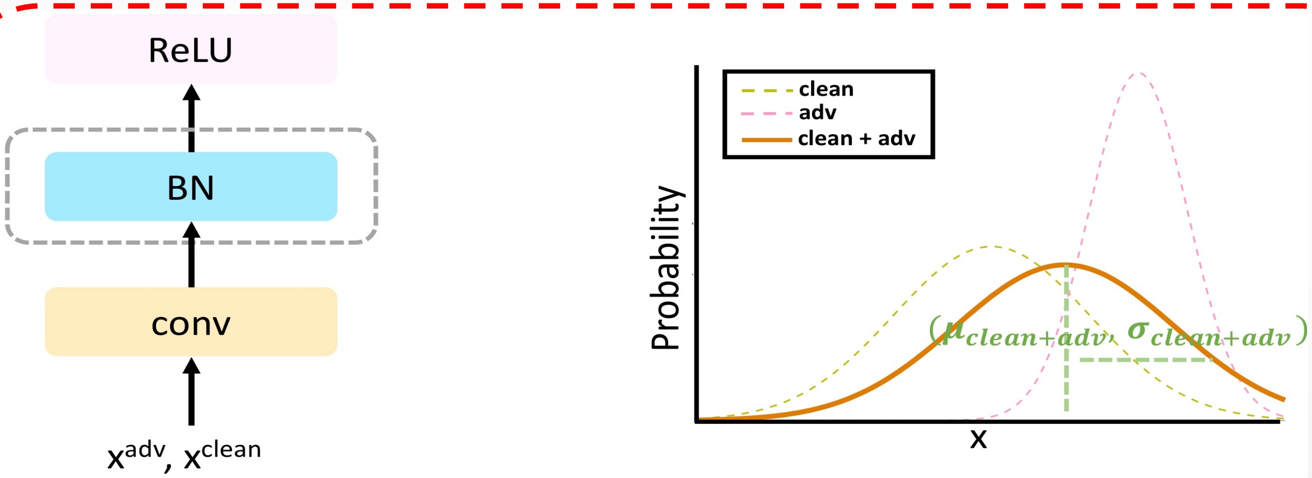
  JOINT TRAINING But with Distinction



Robust Learning Improves Generalization

- **Our Solution**

Joint Training <u>BUT WITH DISTINCTION</u>



(a) Traditional BN

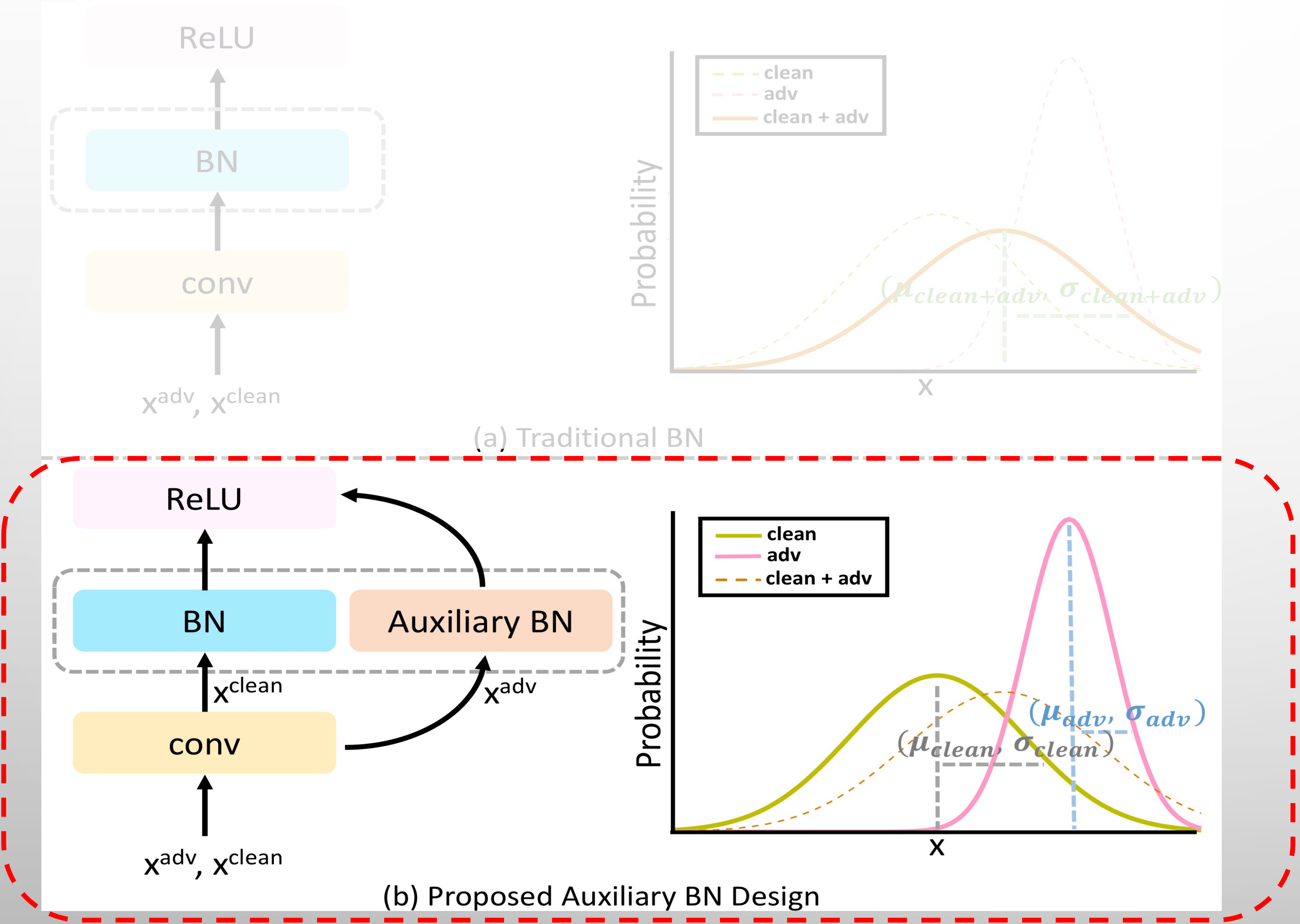
Traditional BN

The statistics estimation at BN may be **<u>CONFUSED</u>** when facing a mixture distribution

Robust Learning Improves Generalization

- **Our Solution**

Joint Training <u>BUT WITH DISTINCTION</u>



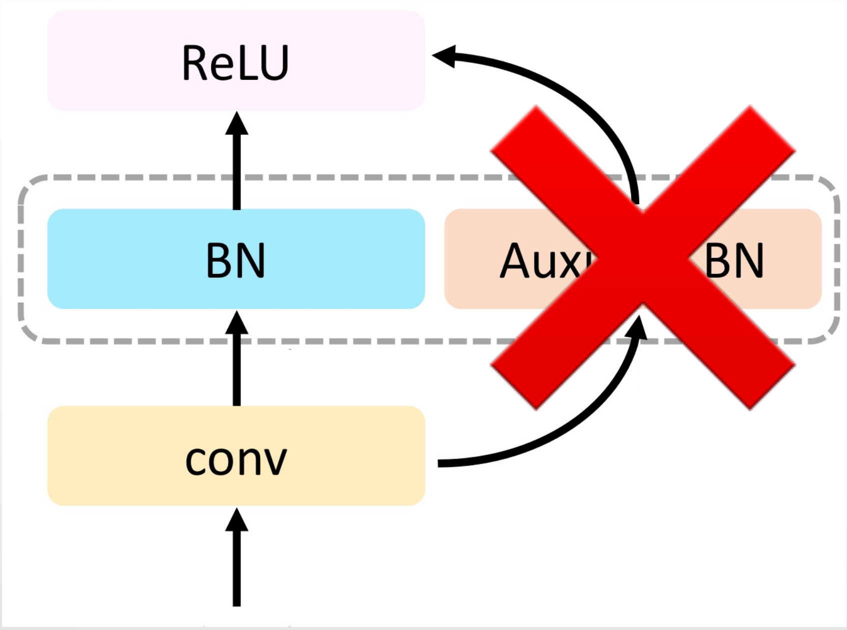
Proposed BN

Auxiliary BN guarantees that data from different distributions are **<u>NORMALIZED SEPARATELY</u>**

Robust Learning Improves Generalization

- **Our Solution**

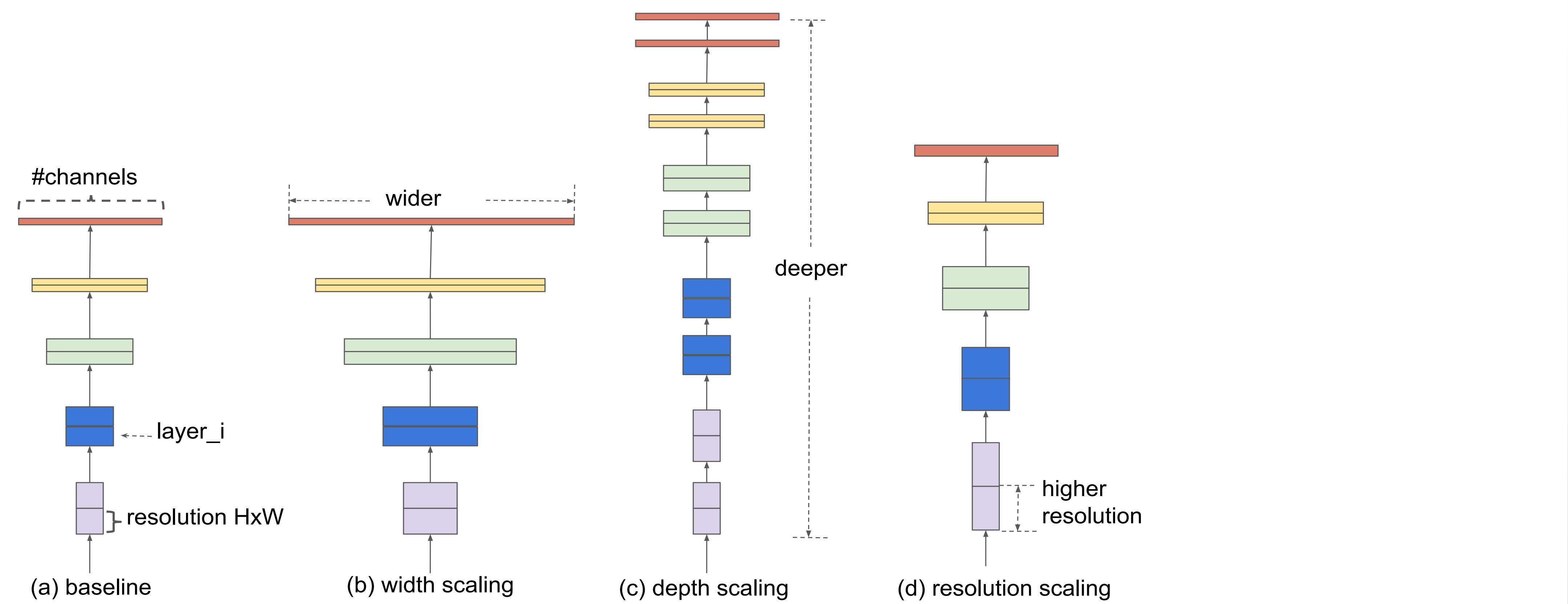
  Adversarial Propagation (**AdvProp**)



Only Main BN is used at the INFERENCE stage

Robust Learning Improves Generalization

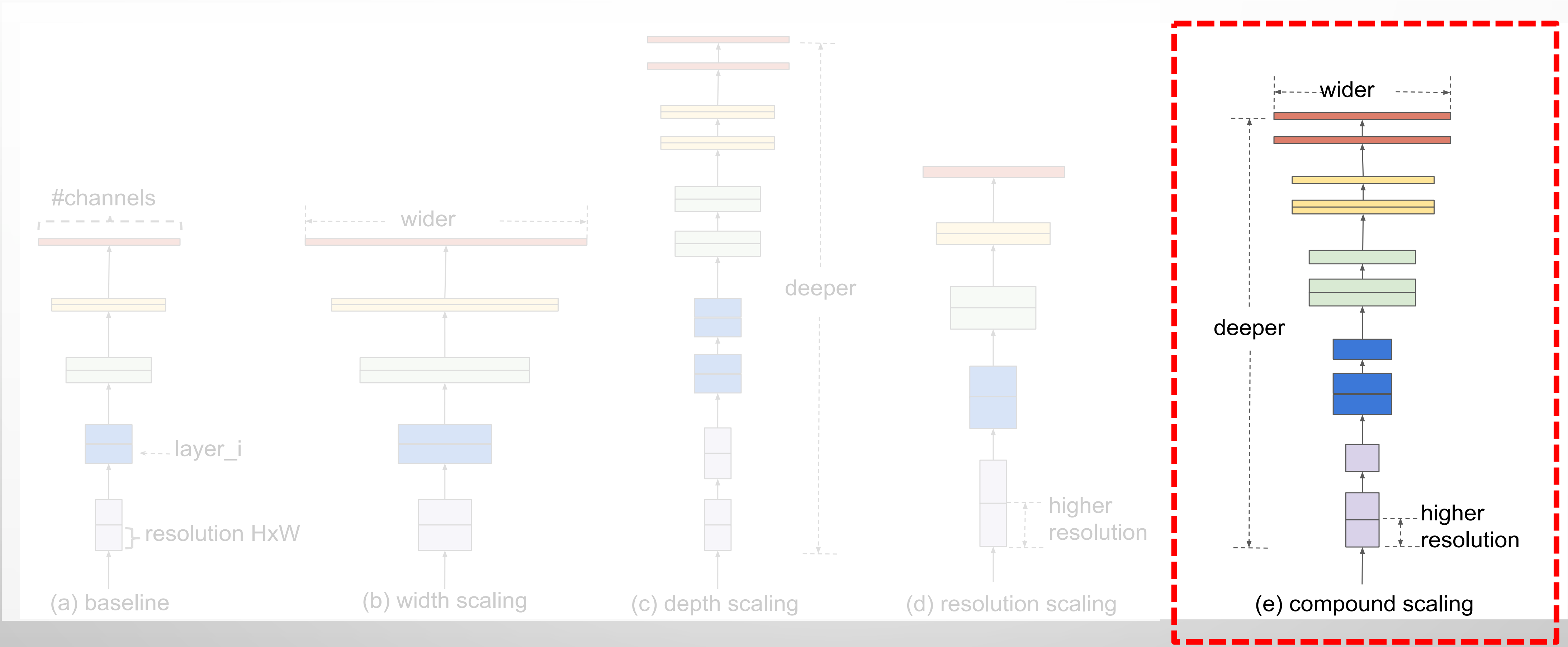
- ➢ **Background --- EfficientNet**



We already know THREE important scaling factors

Robust Learning Improves Generalization

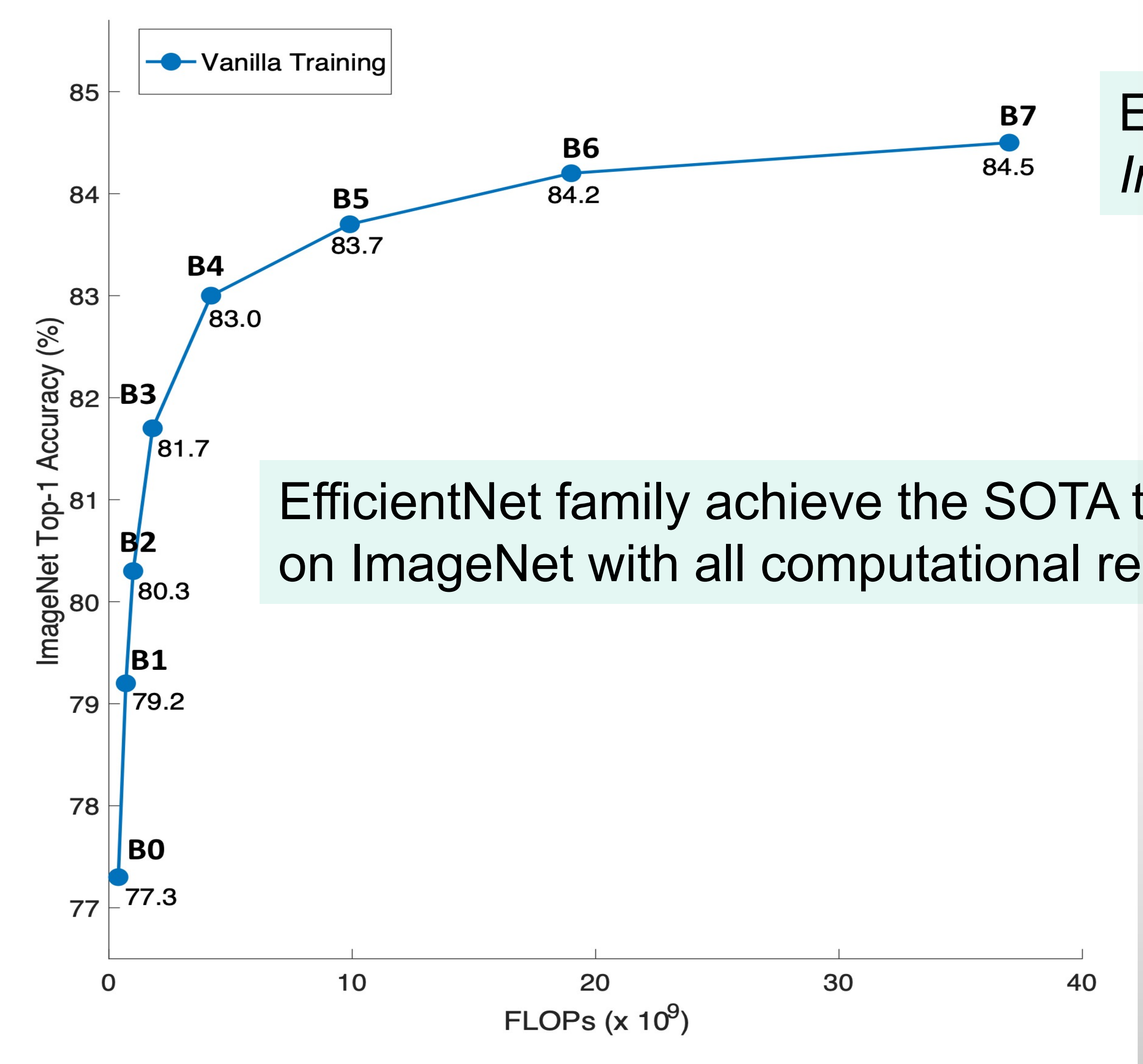
- **Background --- EfficientNet**



A Better SCALING-UP Policy

Robust Learning Improves Generalization

➢ **Results on ImageNet**

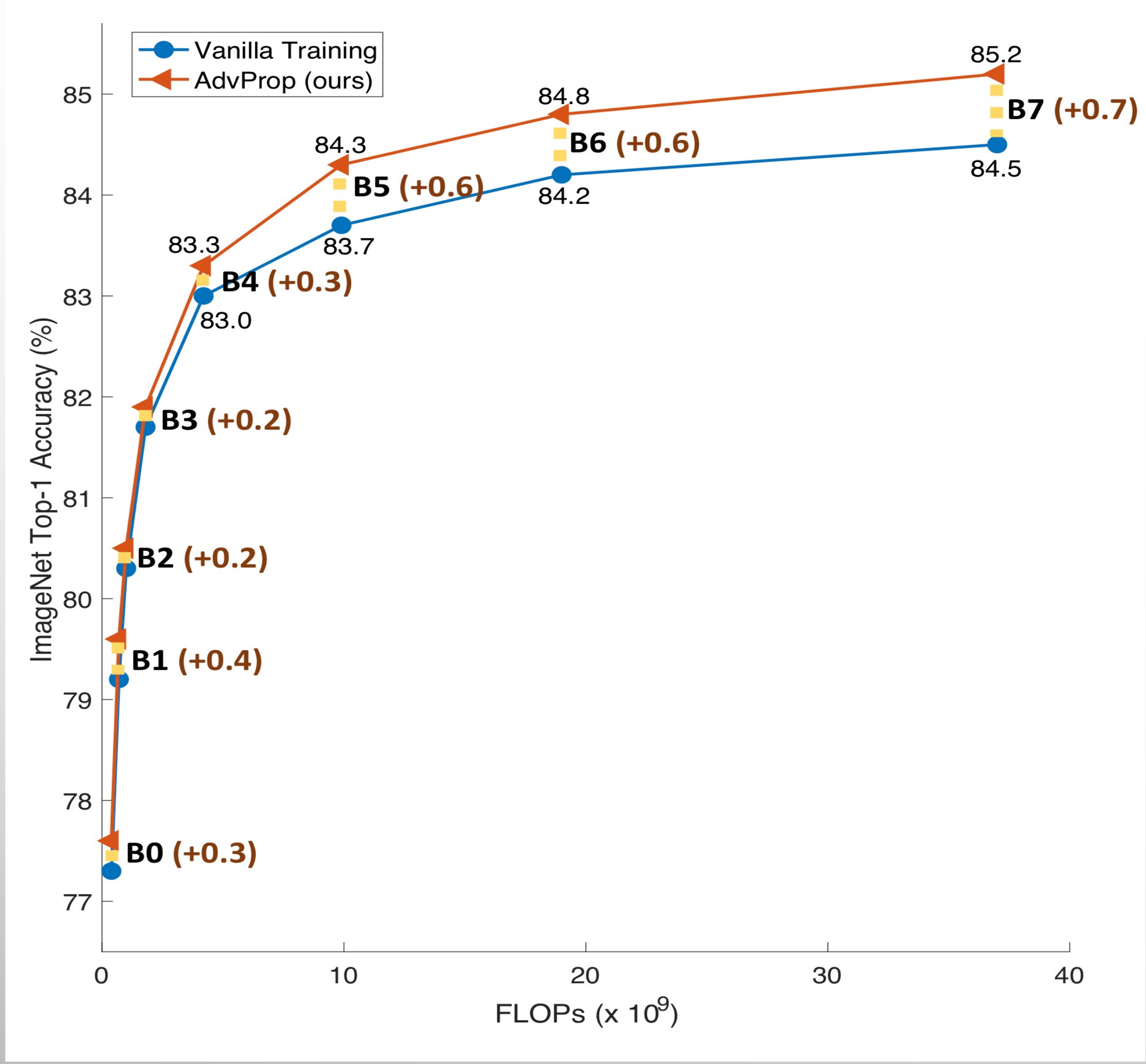


EfficientNet-B7's **84.5%** top-1 accuracy on *ImageNet* is the previous SOTA

EfficientNet family achieve the SOTA top-1 accuracy on ImageNet with all computational regimes

Robust Learning Improves Generalization

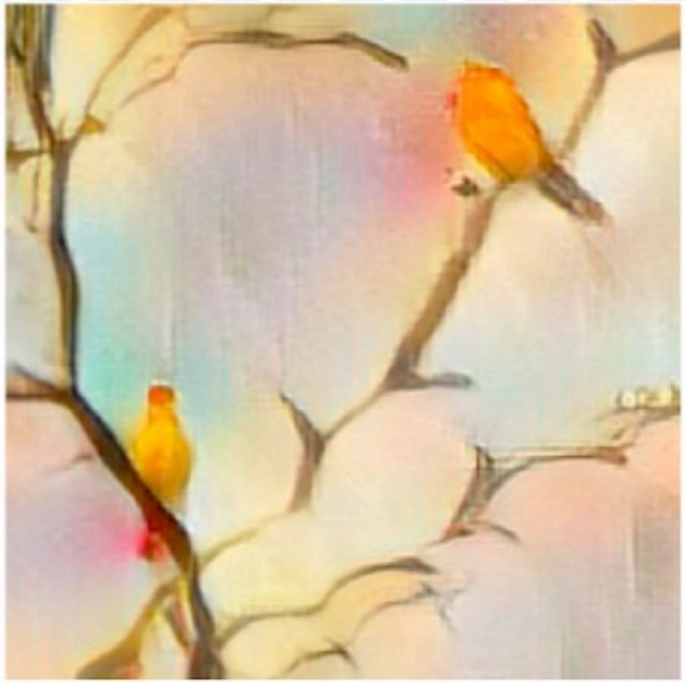
➢ **Results on ImageNet**



AdvProp improves EfficientNet-B7's top-1 accuracy by **0.7% (85.2%)**

Robust Learning Improves Generalization

➢ *Out-of-Distribution* **Generalization**



Networks	ImageNet-C	ImageNet-A	Stylized-ImageNet
EfficientNet-B7	53.1%	37.7%	21.8%
+ AdvProp	**58.2% (+5.1%)**	**44.7% (+7.0%)**	**26.6% (+4.8%)**
ResNet-50	**40.7%**	**3.1%**	**8.0%**

Robust Learning Improves Generalization

➢ **Comparing to the Prior Art**

~10X LESS parameters

3,000X LESS training data

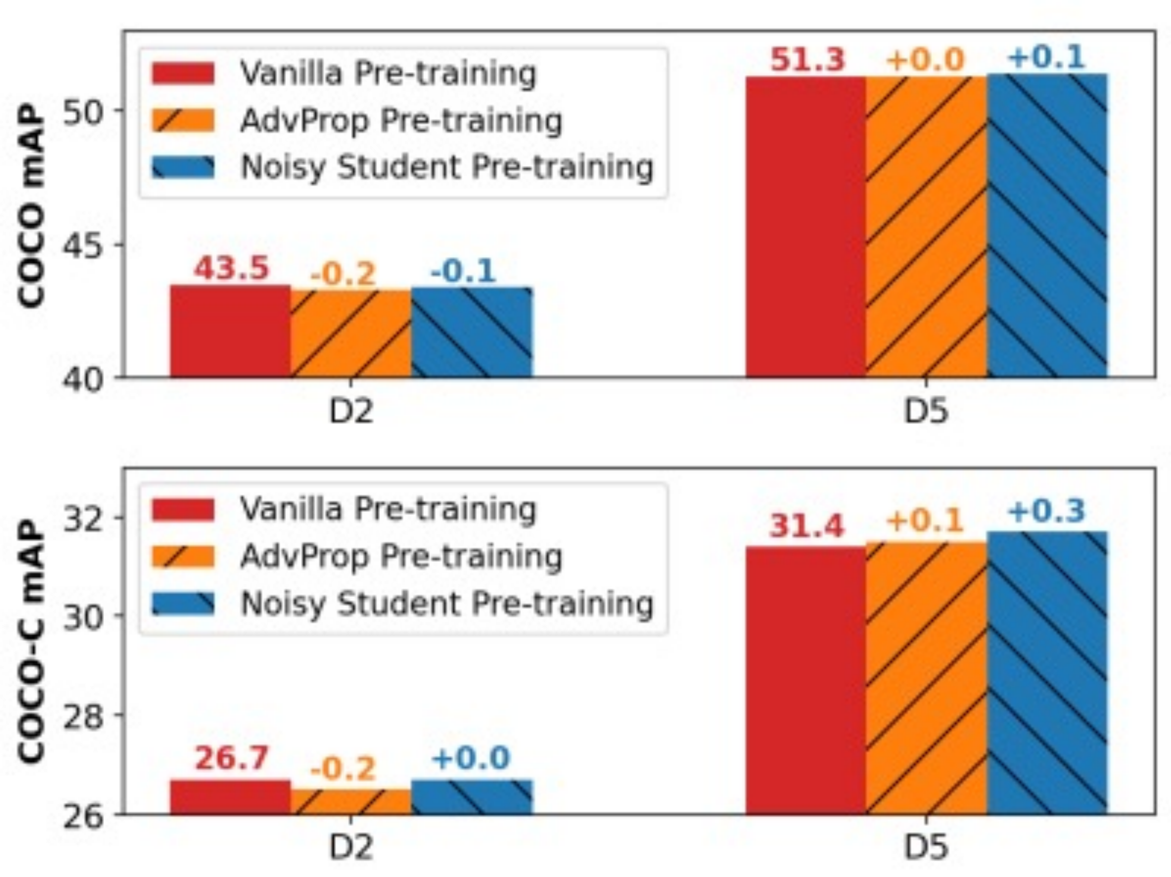
BETTER performance

	# Params	Extra Data	Top-1 Acc.
EfficientNet-B8 + AdvProp	88M	✗	**85.5%**
ResNeXt-101 32x48d [20]	829M	3000× more	85.4%

Robust Learning Improves Generalization

- Improving Object Detection [Chen et al. CVPR'21]

Pre-training then fine-tuning paradigm

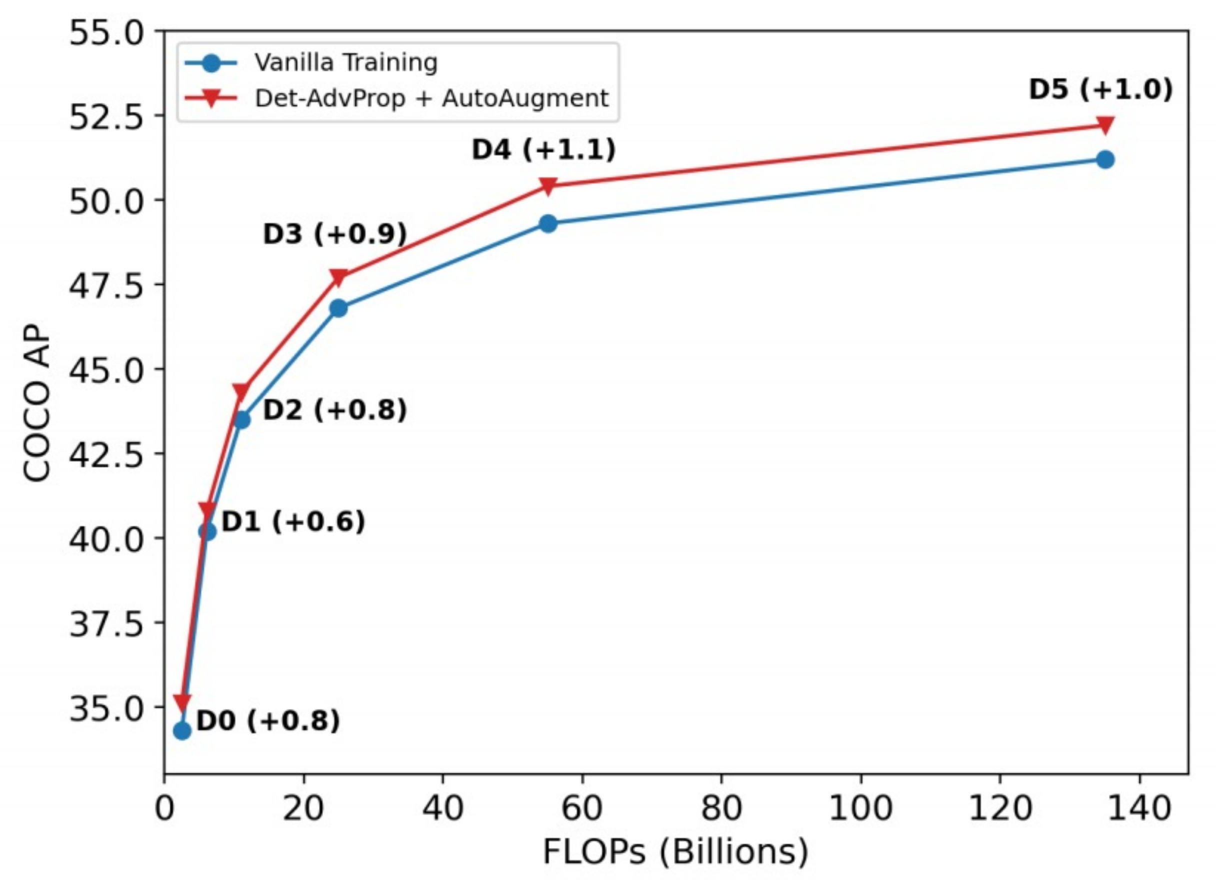


Finetuning **DIFFERENT** pre-trained models yields **SIMILAR** performance on both accuracy and robustness

directly augmenting down-stream object detection task

Robust Learning Improves Generalization

➢ **Improving Object Detection [Chen et al. CVPR'21]**

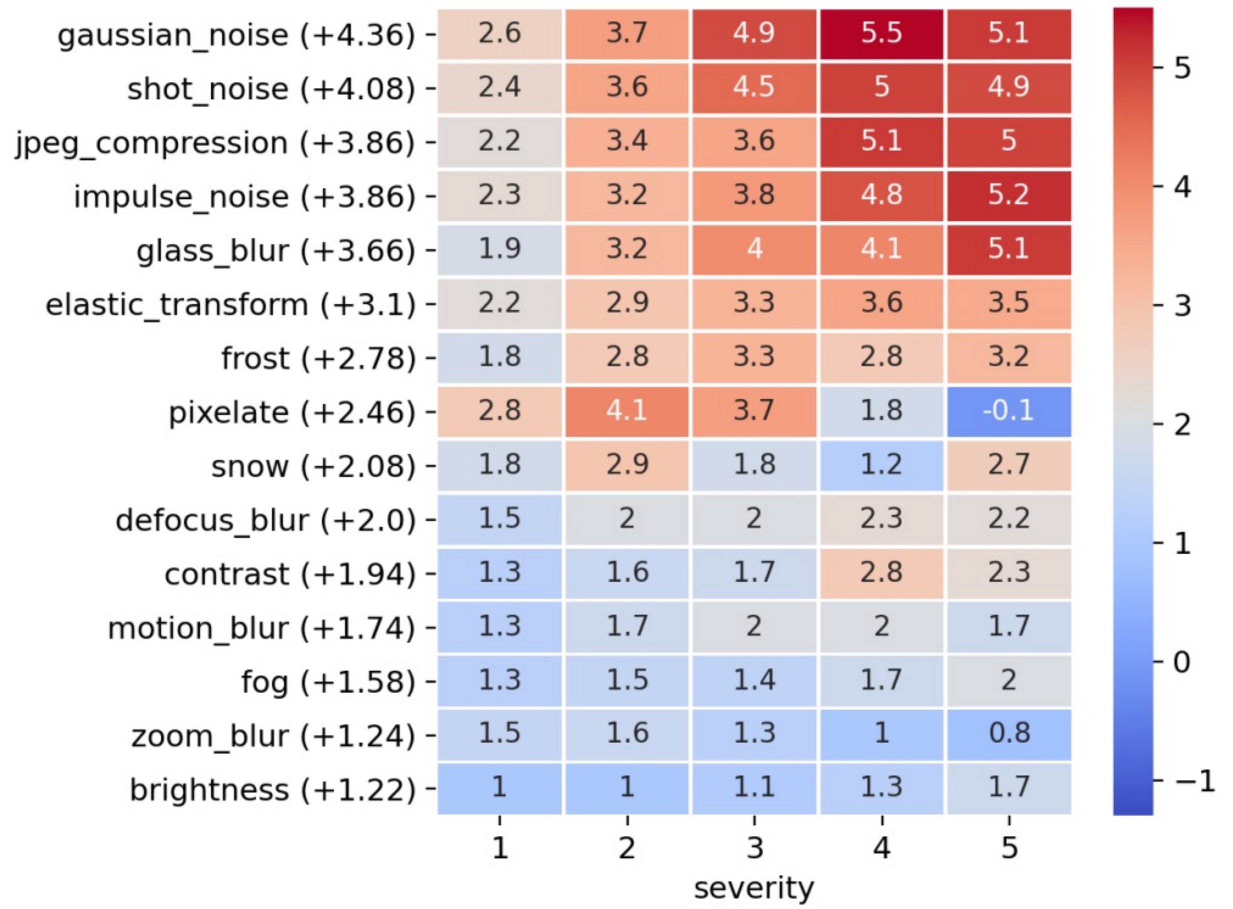


- Boost COCO accuracy up to 1.1 mAP
- Larger improvement on bigger model
- Adapt to single-class detection

Class	Object Size	# Images	Vanilla	Auto-Augment	Det-AdvProp (ours)
Donut	Small	1,585	25.4	23.9 (-1.5)	**28.7 (+3.3)**
Person	Medium	66,808	58.2	58.0 (-0.2)	**58.5 (+0.3)**
Truck	Large	6,377	28.1	25.5 (-2.6)	**28.7 (+0.6)**

Robust Learning Improves Generalization

➢ **Improving Object Detection [Chen et al. CVPR’21]**



- COCO-C: 15 corruptions and 5 severity
- Significantly improve robustness
- Larger gain under stronger corruption strength

Robust Learning Improves Generalization

➢ **Improving Object Detection [Chen et al. CVPR'21]**

Model	mAP	AP50	AP75
Effic			
+ Au			
+ De			
Effic			
+ Au			
+ De			
Effic			
+ Au			
+ De			
Effic			
+ Au			
+ De			
Effic			
+ Au			
+ Det-AdvProp (ours)	**67.5 (+0.5)**	**86.6 (+0.6)**	**74.0 (+1.0)**
EfficientDet-D5	67.4	86.9	73.8
+ AutoAugment	67.6 (+0.2)	87.2 (+0.3)	74.2 (+0.4)
+ Det-AdvProp (ours)	**68.2 (+0.8)**	**87.6 (+0.7)**	**74.7 (+0.9)**

Friday, June 25, 2021 6:00 AM – 8:30 AM

(10200) Robust and Accurate Object Detection via Adversarial Learning

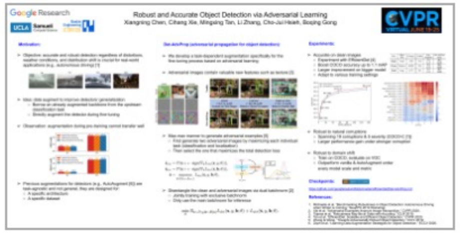
Xiangning Chen, Cihang Xie, Mingxing Tan, Li Zhang, Cho-Jui Hsieh, Boqing Gong

Presenting Author(s)

XC

Xiangning Chen

Google, UCLA



every

Robust Learning Improves Generalization

- **Shape-Texture Debiased Training [Li et al. ICLR'21]**

(a) Texture image

81.4%	**Indian elephant**
10.3%	indri
8.2%	black swan

(b) Content image

71.1%	**tabby cat**
17.3%	grey fox
3.3%	Siamese cat

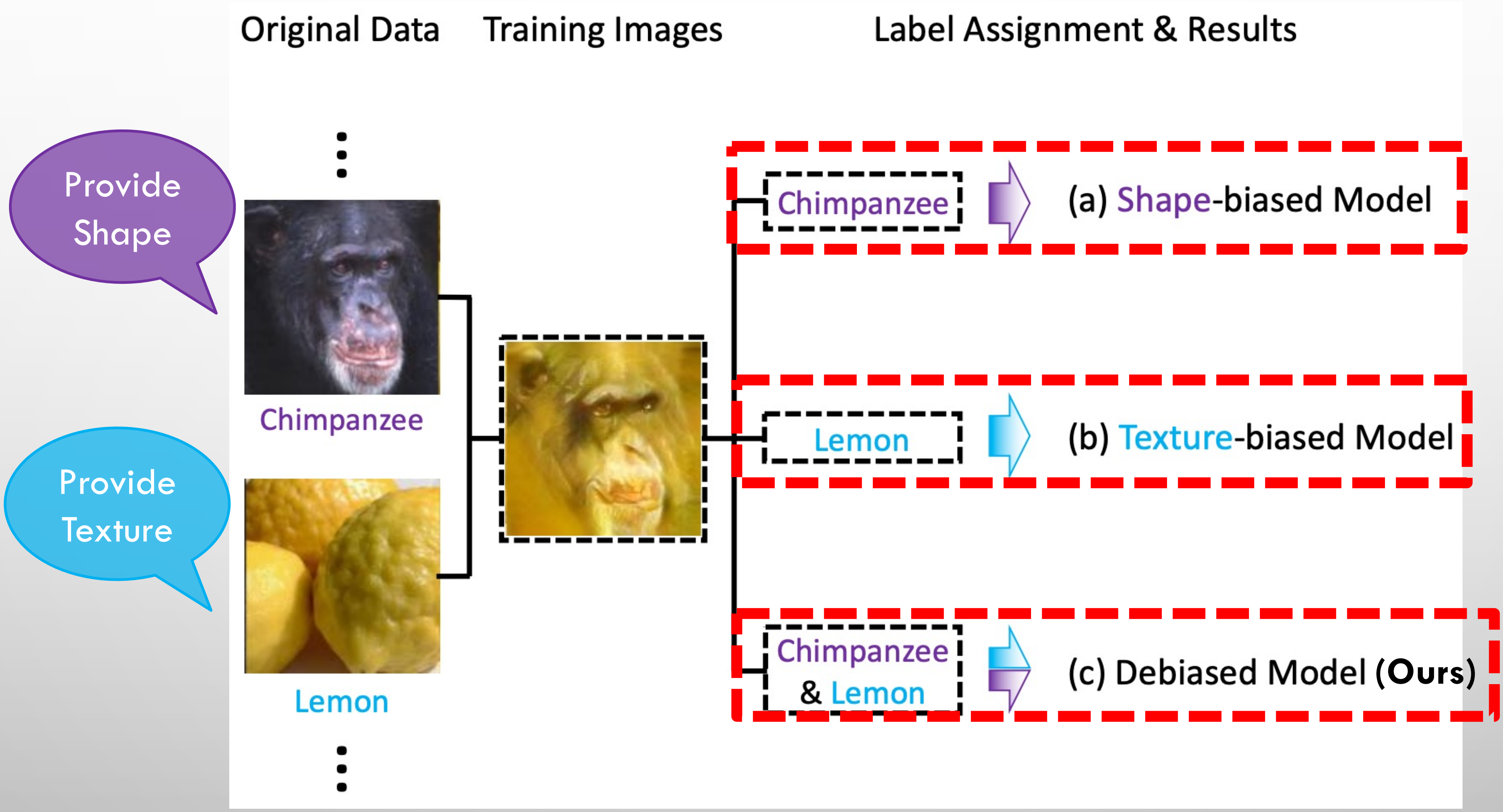
(c) Texture-shape cue conflict

63.9%	**Indian elephant**
26.4%	indri
9.6%	black swan

ImageNet-trained CNNs are biased towards texture [Geirhos et al. 2019]

Robust Learning Improves Generalization

- **Shape-Texture Debiased Training [Li et al. ICLR'21]**



Robust Learning Improves Generalization

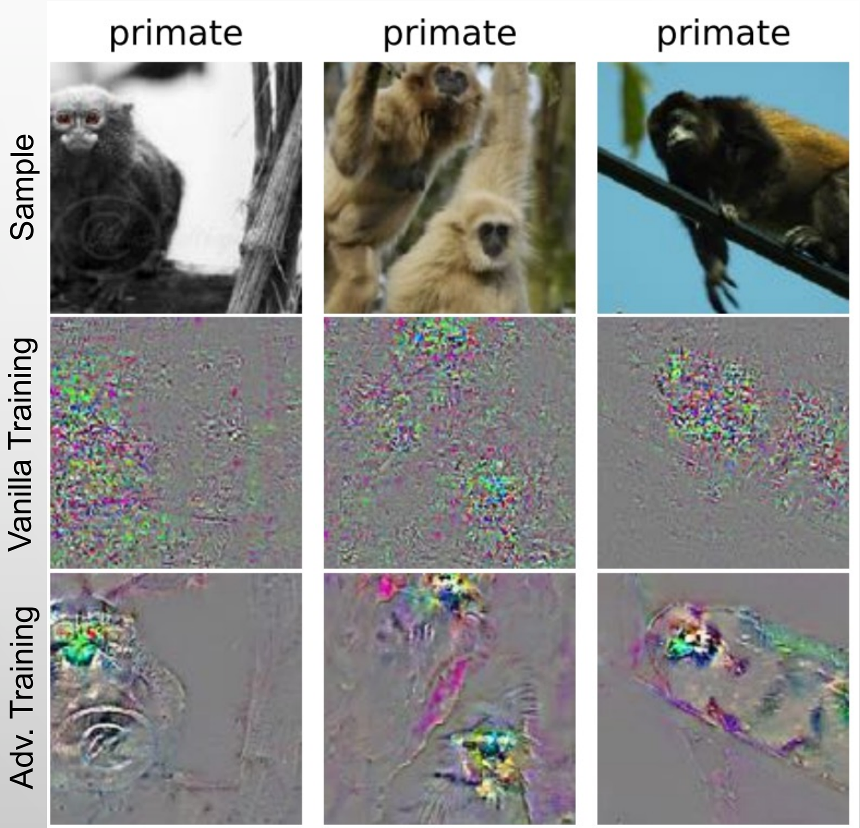
➢ **Shape-Texture Debiased Training [Li et al. ICLR'21]**

	CLEAN Top-1 Acc.↑	IMAGENET-A Top-1 Acc.↑	IMAGENET-C mCE↓	S-IMAGENET Top-1 Acc.↑	FGSM Top-1 Acc.↑
ResNet-50	76.4	2.0	75.0	7.4	17.1
Debiased	76.9**(+0.5)**	3.5**(+1.5)**	67.5**(-7.5)**	17.4**(+10.0)**	27.4**(+10.3)**
ResNet-101	77.9	5.6	69.8	9.9	23.1
Debiased	78.9**(+1.0)**	9.1**(+3.5)**	62.2**(-7.6)**	22.0**(+12.1)**	34.4**(+11.3)**
ResNet-152	78.6	7.4	67.2	11.3	25.2
Debiased	79.8**(+1.2)**	12.6**(+5.2)**	58.9**(-8.3)**	22.4**(+11.1)**	39.6**(+14.4)**

Robust Learning Improves Generalization

Takeaways

➢ Adversarially learned features are **VALUABLE**



Qualitative Evidence
[Tsipras et al. 2019]

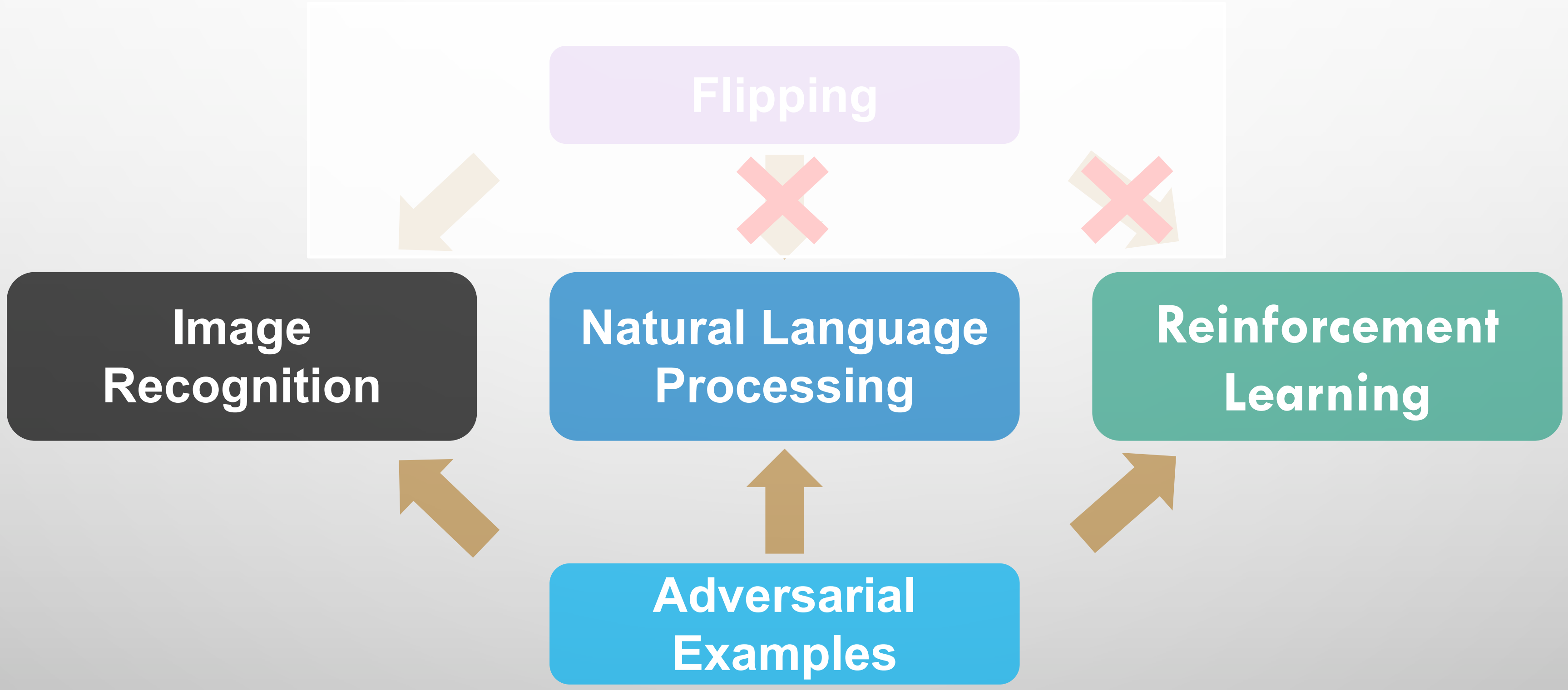
Quantitative Evidence
[Xie et al. 2020]

Robust Learning Improves Generalization

Takeaways

- Adversarially learned features are VALUABLE

- Adversarial examples can serve as a GENERAL data augmentation method



Robust Learning Improves Generalization

Takeaways

- Adversarially learned features are **VALUABLE**
- Adversarial examples can serve as a **GENERAL** data augmentation method
- **DISENTANGLED LEARNING** is important when inputs come from different distributions

Robust Learning Improves Generalization

Takeaways

Shape cue: cat
Texture cue: elephant

	IMAGENET-A Top-1 Acc. ↑	IMAGENET mCE ↓
ResNet-152	7.4	67.2
+ Debiased	12.6 (+5.2)	58.9 (-8

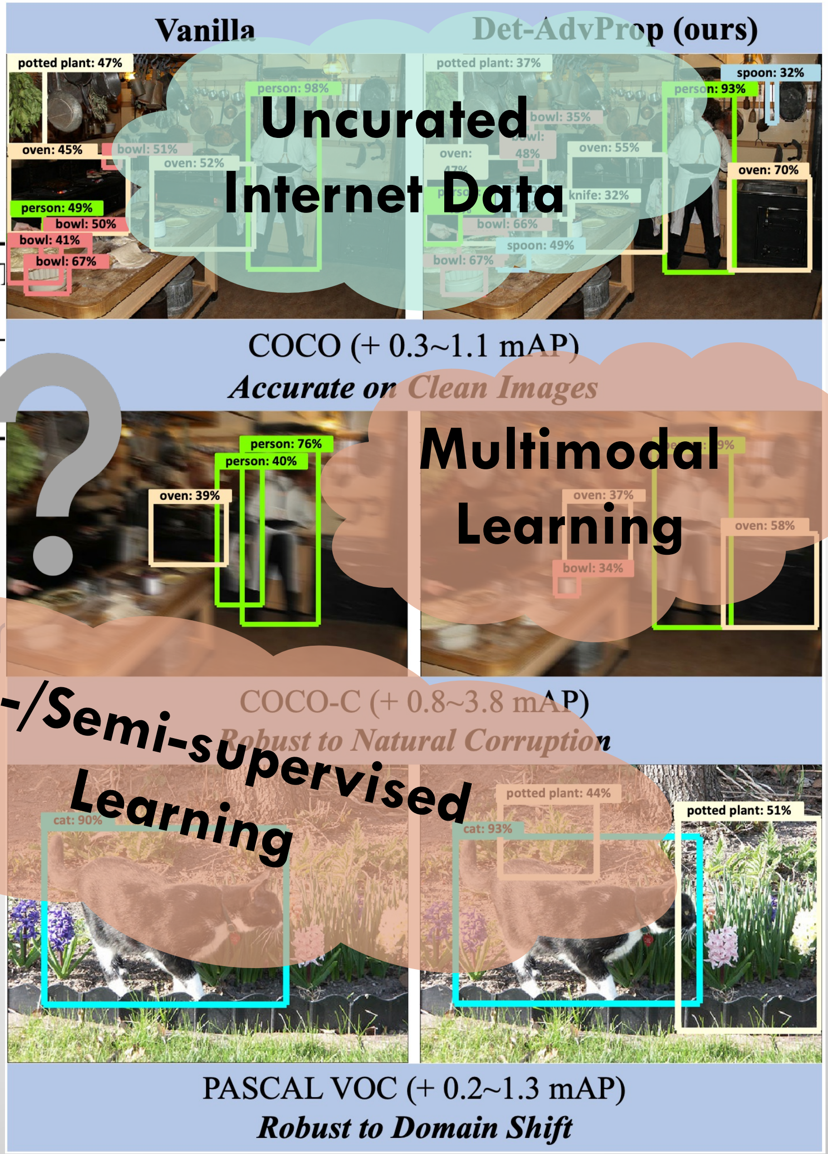
Synthetic Data

Few-shot Learning

Uncurated Internet Data

Multimodal Learning

Un-/Semi-supervised Learning



Style Transfer

Shape-Texture (ICLR'21)

Adversarial Perturbation

AdvProp (CVPR'20)

Det-AdvProp (CVPR'21)

Classification Segmentation Detection • • •

ACKNOWLEDGEMENT

➢ COLLABORATORS

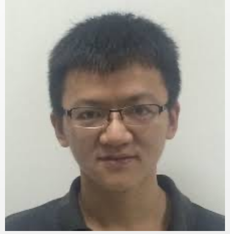










➢ Sponsor

2020

Open Philanthropy

2021 - 2024

We're HIRING

- Multiple Positions for (Remote) Summer Interns & Visiting Students

Email: cixie@ucsc.edu


Q
A