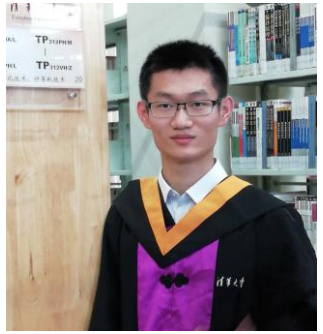
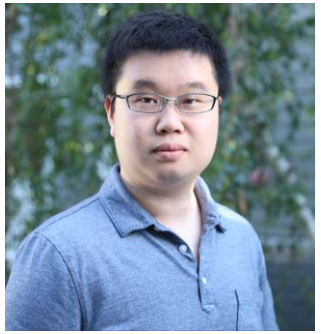
NEURAL INFORMATION PROCESSING SYSTEMS

AugMax: Adversarial Composition of Random Augmentations for Robust Training

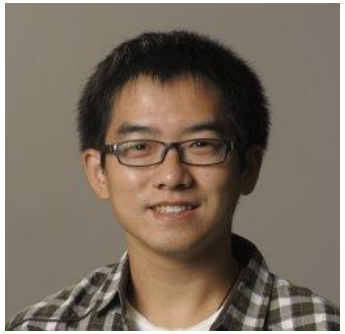

Haotao Wang[1]


Chaowei Xiao[2,3]


Jean Kossaifi[2]


Zhiding Yu[2]


Anima Anandkumar[2,4]


Zhangyang Wang[1]

[1]University of Texas at Austin, [2]NVIDIA, [3]Arizona State University, [4]California Institute of Technology

TEXAS The University of Texas at Austin

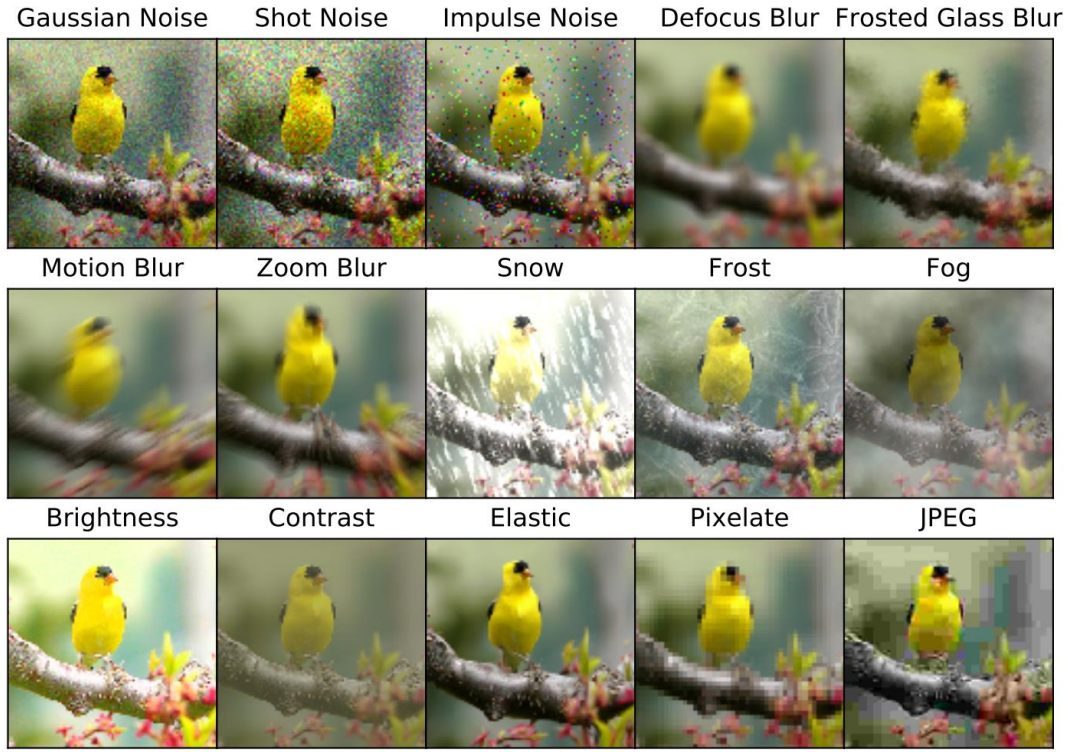
NVIDIA

NEURAL INFORMATION PROCESSING SYSTEMS

Problem Setting: Robust Training

Train on clean images

Generalize to



Evaluate robustness against unseen natural corruptions (e.g., ImageNet-C [1])

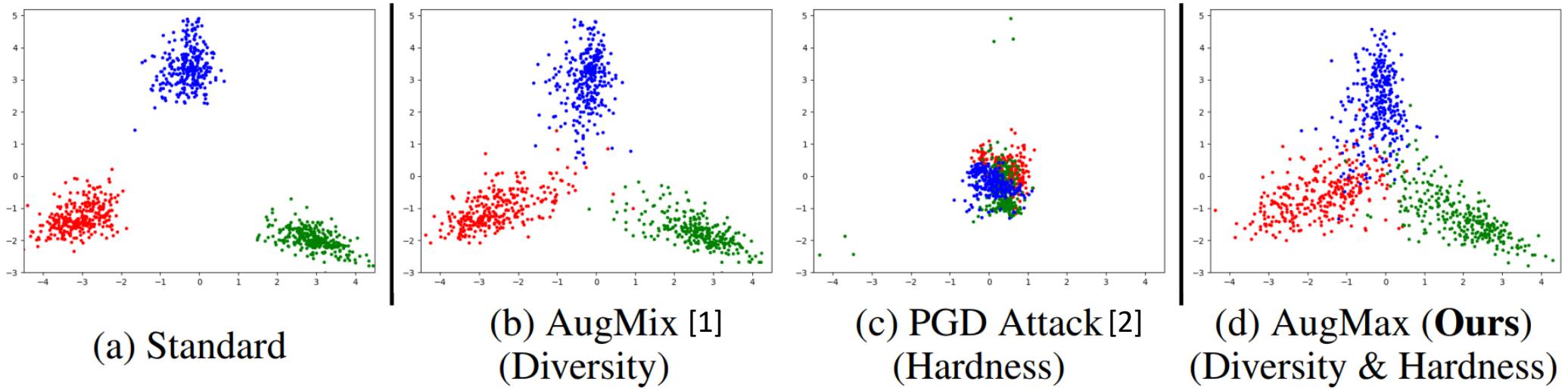
TEXAS The University of Texas at Austin

NVIDIA

[1] Hendrycks, D. and Dietterich, T. Benchmarking neural network robustness to common corruptions and perturbations. *ICLR*, 2019.

NEURAL INFORMATION PROCESSING SYSTEMS

Motivation: Unification between Diversity and Hardness



(a) Standard

(b) AugMix [1] (Diversity)

(c) PGD Attack [2] (Hardness)

(d) AugMax (**Ours**) (Diversity & Hardness)

Figure 1: Features of augmented images fed to the network during training. Different colors represent images of different classes.

- Previous methods: Leverage either diversity or hardness to improve robustness.
- AugMax: Unify diversity and hardness in a single framework.

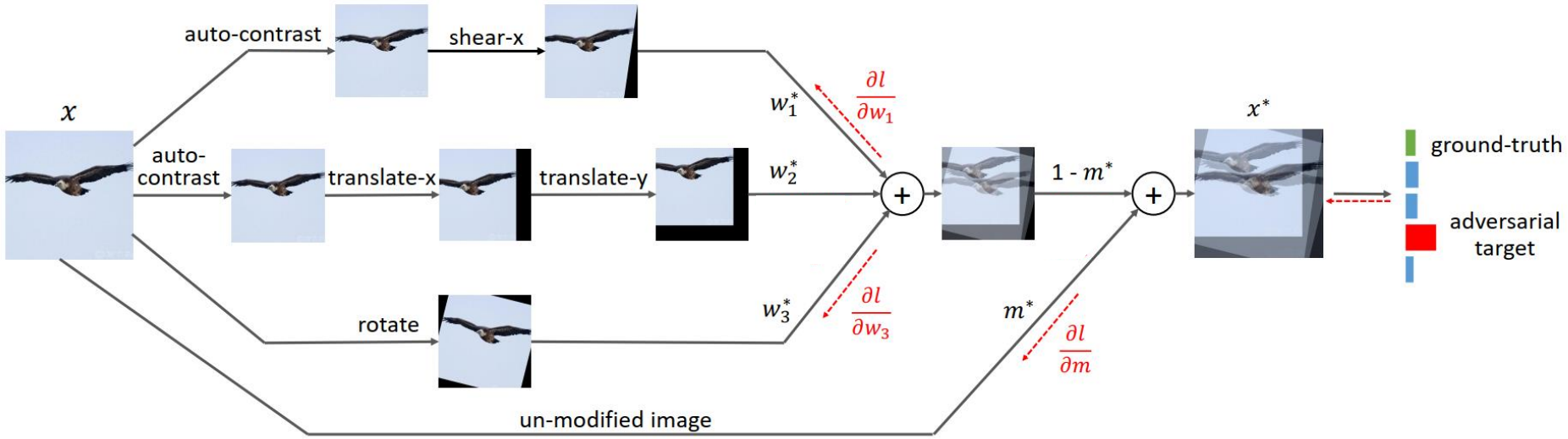
TEXAS The University of Texas at Austin

NVIDIA

[1] Hendrycks, D. et al. AugMix: A simple data processing method to improve robustness and uncertainty. *ICLR*, 2020.
[2] Madry, A. et al. Towards deep learning models resistant to adversarial attacks. *ICLR*, 2018.

NEURAL INFORMATION PROCESSING SYSTEMS

Overall Framework of AugMax



AugMax

Method	Mixing parameters $(\boldsymbol{w}^*, m^*)$
AugMix [1]	Randomly selected
AugMax	Adversarially learned

$$\min_{\boldsymbol{\theta}} \mathbb{E}_{(\boldsymbol{x},\boldsymbol{y})\sim\mathcal{D}} \frac{1}{2}[\mathcal{L}(f(\boldsymbol{x}^*);\boldsymbol{\theta}),\boldsymbol{y}) + \mathcal{L}(f(\boldsymbol{x});\boldsymbol{\theta}),\boldsymbol{y})] + \lambda\mathcal{L}_c(\boldsymbol{x},\boldsymbol{x}^*)$$

$$\text{s.t. } \boldsymbol{x}^* = g(\boldsymbol{x}; m^*, \boldsymbol{w}^*); \quad \boldsymbol{w}^* = \sigma(\boldsymbol{p}^*); \quad m^*, \boldsymbol{p}^* = \underset{m\in[0,1],\boldsymbol{p}\in\mathbb{R}^b}{\arg\max} \mathcal{L}(f(g(\boldsymbol{x}; m, \sigma(\boldsymbol{p}));\boldsymbol{\theta}),\boldsymbol{y})$$

Adversarially selected

f: classifier (CNN) with parameter $\boldsymbol{\theta}$
g: data augmentation pipeline
$\mathbf{x}, \mathbf{y}$: data and label
$\boldsymbol{w}^*, m^*$: mixing parameters
σ: softmax function
L, L_c: Loss functions
λ: hyper-parameter

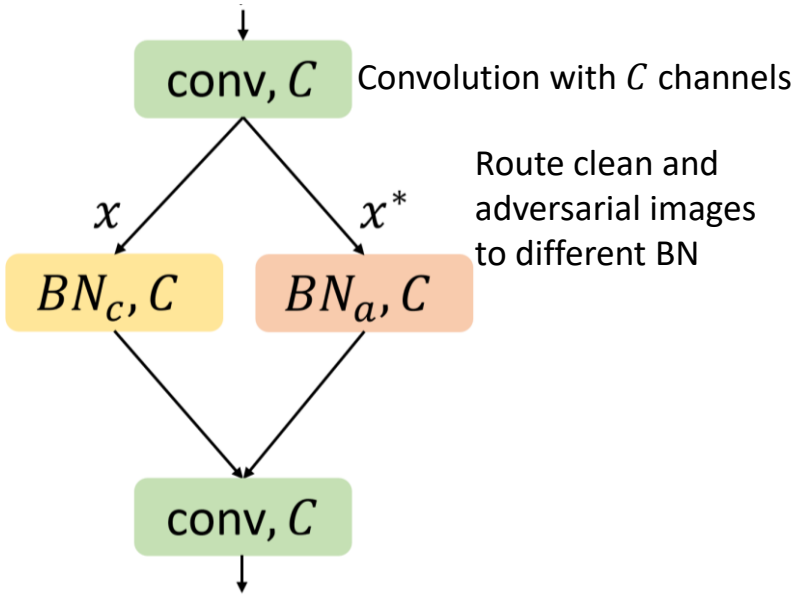
TEXAS The University of Texas at Austin

NVIDIA

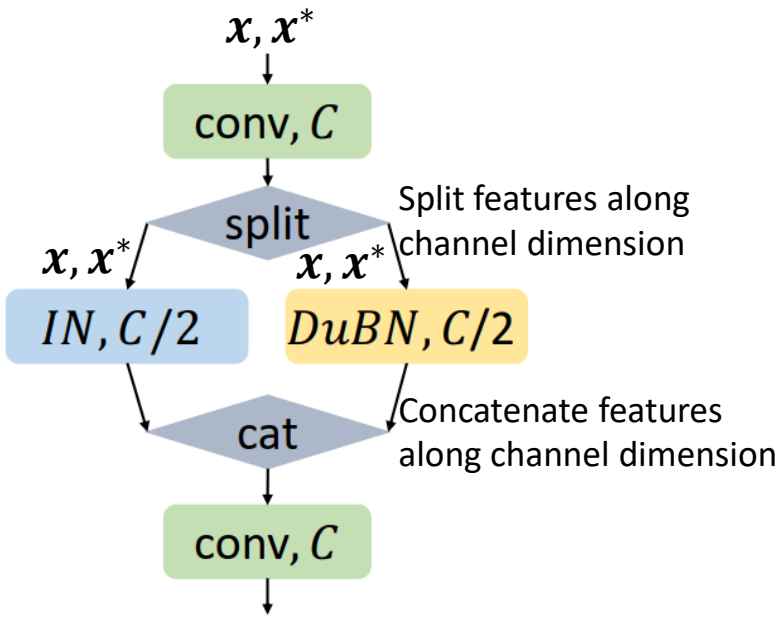
[1] Hendrycks, D. et al. AugMix: A simple data processing method to improve robustness and uncertainty. *ICLR*, 2020.

DuBIN: Disentangled Normalization for Heterogeneous Features

$\boldsymbol{x}$: Clean images
$\boldsymbol{x}^*$: Adversarial images



DuBN [1]



DuBIN

Table 1: BN statistics of different layers in WRN40-2 with DuBN or DuBIN trained on CIFAR100 using AugMax training.

Layer		Block 1 layer 2	Block 1 layer 3	Block 2 layer 2
AugMax-DuBN	$\overline{\sigma}_c^2$	0.0369	0.0450	0.0301
	$\overline{\sigma}_a^2$	0.0469	0.0585	0.0382
AugMax-DuBIN	$\overline{\sigma}_c^2$	0.0306	0.0403	0.0264
	$\overline{\sigma}_a^2$	0.0348	0.0466	0.0292

Both $\overline{\sigma}_c^2$ and $\overline{\sigma}_a^2$ (the variance of BN_c and BN_a) are smaller in the DuBIN network than in its DuBN counterpart.

IN in DuBIN can reduce the feature diversity that BN needs to model, by sharing the burden of encoding instance-level diversity.

[1] Xie, C. and Yuille, A. Intriguing properties of adversarial Training. *ICLR*, 2020.

NEURAL INFORMATION PROCESSING SYSTEMS

Main Results: Robustness against Natural Corruptions

Evaluation results on CIFAR10 and CIFAR10-C.

Model	Metric	Normal	AugMix	AugMax-DuBIN
ResNet18	SA (%)	95.56	**95.79**	95.76 (-0.03)
	RA (%)	74.75	89.49	**90.36** (+0.87)
WRN40-2	SA (%)	94.78	95.67	**95.68** (+0.01)
	RA (%)	73.71	89.01	**90.67** (+1.66)
ResNeXt29	SA (%)	95.60	96.25	**96.39** (+0.14)
	RA (%)	71.70	89.08	**92.11** (+3.03)

Evaluation results on CIFAR100 and CIFAR100-C.

Model	Metric	Normal	AugMix	AugMax-DuBIN
ResNet18	SA (%)	77.99	78.23	**78.69** (+0.46)
	RA (%)	48.46	62.67	**65.75** (+3.08)
WRN40-2	SA (%)	76.19	**77.03**	76.80 (-0.23)
	RA (%)	46.80	64.56	**66.35** (+1.79)
ResNeXt29	SA (%)	79.95	78.58	**80.70** (+2.12)
	RA (%)	47.76	65.37	**68.86** (+3.49)

Evaluation results on ImageNet and ImageNet-C.

Method	SA (%, ↑)	RA (%, ↑)	mCE (%, ↓)
Normal	69.83	30.91	87.47
AugMix	**68.06**	34.58	83.08
AugMax-DuBIN	67.62 (-0.44)	**35.01** (+0.43)	**82.56** (-0.52)
DeepAugment + AugMix	**65.32**	45.84	69.29
DeepAugment + AugMax-DuBIN	64.43 (-0.89)	**46.55** (+0.71)	**68.47** (-0.82)

Evaluation results on Tiny ImageNet and Tiny ImageNet-C.

Method	SA (%, ↑)	RA (%, ↑)	mCE (%, ↓)
Normal	61.64	23.91	100.00
AugMix	61.79	36.85	83.04
AugMax-DuBIN	**62.21** (+0.42)	**38.67** (+1.82)	**80.72** (-2.32)
DeepAugment + AugMix	59.59	40.67	78.28
DeepAugment + AugMax-DuBIN	**59.72** (+0.13)	**40.99** (+0.32)	**77.83** (-0.45)

TEXAS The University of Texas at Austin

NVIDIA

NEURAL INFORMATION PROCESSING SYSTEMS

Ablation Study: Analysis of Different Normalization Layers

Table 8: Ablation results on DuBIN. RA (%) on CIFAR10-C and CIFAR100-C with WRN40-2 backbone are reported.

Method	CIFAR10-C	CIFAR100-C
AugMix-BN	89.01 (± 0.03)	64.56 (± 0.04)
AugMix-IBN	89.17 (± 0.24)	63.94 (± 0.28)
AugMix-DuBN	89.11 (± 0.21)	64.07 (± 0.38)
AugMix-DuBIN	89.74 (± 0.35)	65.02 (± 0.58)
AugMax-DuBN	89.60 (± 0.62)	65.06 (± 0.28)
AugMax-DuBIN	**90.67** (± 0.16)	**66.31** (± 0.24)

Observation 1:
- AugMax-DuBN outperforms both AugMix-BN and AugMix-DuBN.
- AugMax-DuBIN outperforms both AugMix-IBN and AugMix-DuBIN.

AugMax results in more robustness than AugMix as a data augmentation method.

Observation 2: DuBIN can also help improve robustness when combined with AugMix.

DuBIN may potentially be applied to other diversity augmentation methods for general performance improvement.

TEXAS The University of Texas at Austin

NVIDIA

NEURAL INFORMATION PROCESSING SYSTEMS

Ablation Study: Comparison with Different Diversity and Hardness Strategies

Table 10: Results of different augmentation strategies on CIFAR100(-C) with ResNet18.

Method	Strategy	SA (%)	RA (%)
Normal	-	77.99	48.46
AugMix [5]	(diversity)	78.23	62.67
PGDAT [6]	(hardness)	60.94	47.71
FAT [26]	(hardness)	61.51	48.70
AdvMax	(hardness)	56.61	38.65
AugMix+PGDAT	(diversity & adversity)	61.68	51.39
AdvMix	(diversity & adversity)	72.36	56.89
AugMax-DuBIN	(diversity & adversity)	**78.52**	**64.02**

- AdvMix: applying adversarial attacks on the augmentation hyperparameters (e.g., rotation angles) while randomly selecting the mixing parameters.
- AdvMax: applying adversarial attacks on both the augmentation hyperparameters and the mixing weights.

Observation 1: AugMax-DuBIN outperforms all methods from either diversity or hardness group.

Diversity and hardness are indeed two complementary dimensions and a proper combination of them can boost model robustness.

Observation 2: The other two naive baselines jointly considering diversity and hardness achieve poor performance.

It is nontrivial to design a method achieving good balance between diversity and hardness.

TEXAS The University of Texas at Austin

NVIDIA

NEURAL INFORMATION PROCESSING SYSTEMS

Results: Robustness against Other Distributional Shifts

Table 11: Robustness against other distribution shifts. Accuracy (%) on CIFAR10.1 and CIFAR10-STA are evaluated on ResNeXt29 trained on CIFAR10.

Method	CIFAR10.1	CIFAR10-STA
Normal	88.90	30.30
AugMix	88.90	54.30
AugMax-DuBIN	**90.64**	**63.20**

- CIFAR10.1: New test images sampled to minimize distributional shifts to original CIFAR10 [1].
- CIFAR10-STA: CIFAR10 test set under Spatial Transform Adversarial Attacks (STA) [2].

Conclusion: Our method also improves robustness against other distributional shifts.

TEXAS The University of Texas at Austin

NVIDIA

[1] Recht, B. et al. Do CIFAR-10 Classifiers Generalize to CIFAR-10?
[2] Engstrom, L. et al. Exploring the Landscape of Spatial Robustness. *ICML*, 2019.

NEURAL INFORMATION PROCESSING SYSTEMS

Thank You!

Our code and pretrained models are available on GitHub:

https://github.com/VITA-Group/AugMax

TEXAS The University of Texas at Austin

NVIDIA