Gradient-Free Adversarial Training Against Image Corruption for Learning-based Steering

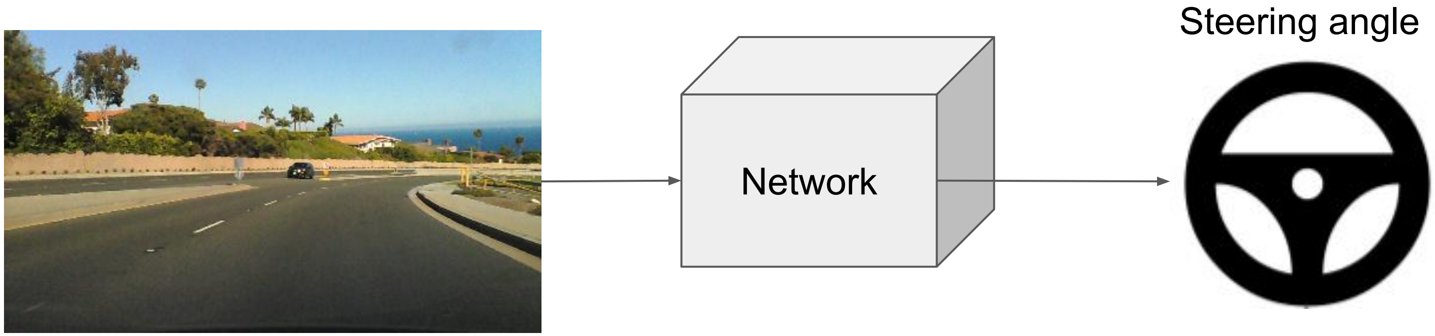
Paper ID: 3594

Motivation

- For safety, autonomous vehicles (AVs) need to drive under varying lighting, weather, and visibility conditions in different environments
- Internal (e.g. sensors) and external (e.g. environments) factors can pose significant challenges to perceptual data processing and affect control and decision-making of AVs
- To harden neural network systems against image degradation and improve robustness of learning algorithms

Target Task

- Learn to steer in end-to-end autonomous driving
  - *perception and control*



Key Contributions

A scalable training scheme to improve robustness of autonomous driving against image corruption, while increasing overall accuracy of learning-based steering with adversarial data augmentation

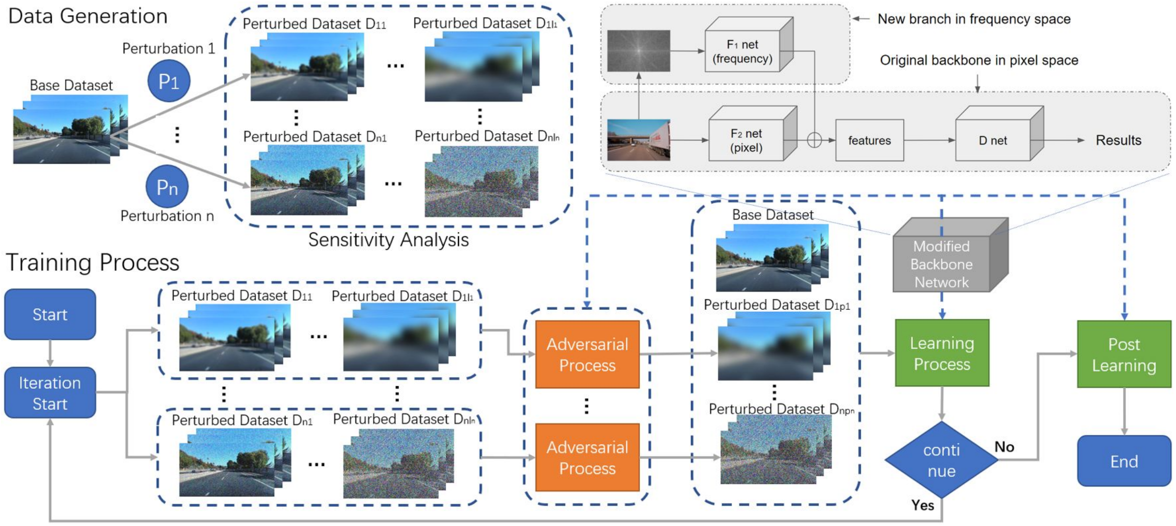
- Among the first to train on single image perturbations, while improving overall performance on functional combination of perturbations or previously unseen perturbations

- A systematic approach for analyzing, predicting, and quantifying impact of an image degradation on network using sensitivity analysis and Fréchet Inception Distance (FID)

- A benchmark that contains autonomous driving datasets with different perturbations and comparison on performance of recent methods on different datasets and backbones

Pipeline of Our System

- Data generation: perturbed datasets of each factor at multiple levels based on the FID-parameterized sensitivity analysis results.
- Training process: (1) augment the training dataset with “adversarial images” given each perturbation, then combine the base and perturbed datasets to maximize overall performance iteratively; (2) fine-tune the model in post learning.



Experiments: Training Data

- ***Clean Driving Data***

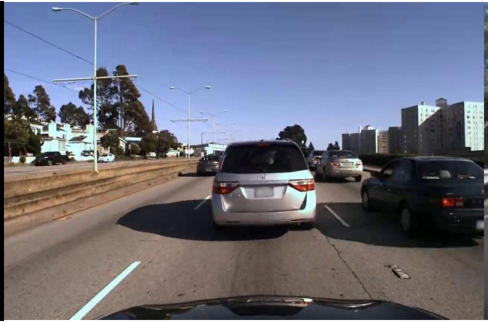
  -- Audi 2020, Honda 2018, SullyChen 2018, and Waymo 2020

- ***Single Perturbation:***

  -- blur, noise, distortion, R, G, B, H, S, V channels

Examples of Clean Driving Video

Audi 2020



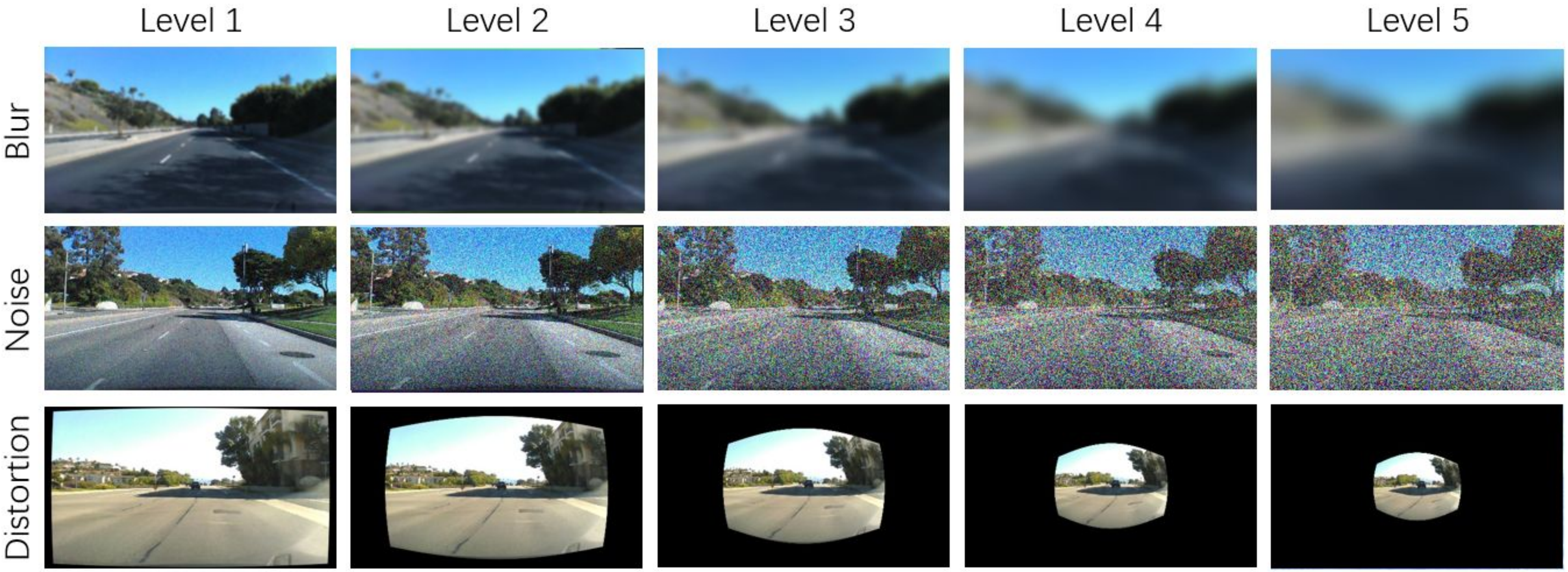
Honda 2018

SullyChen 2018

Waymo 2020

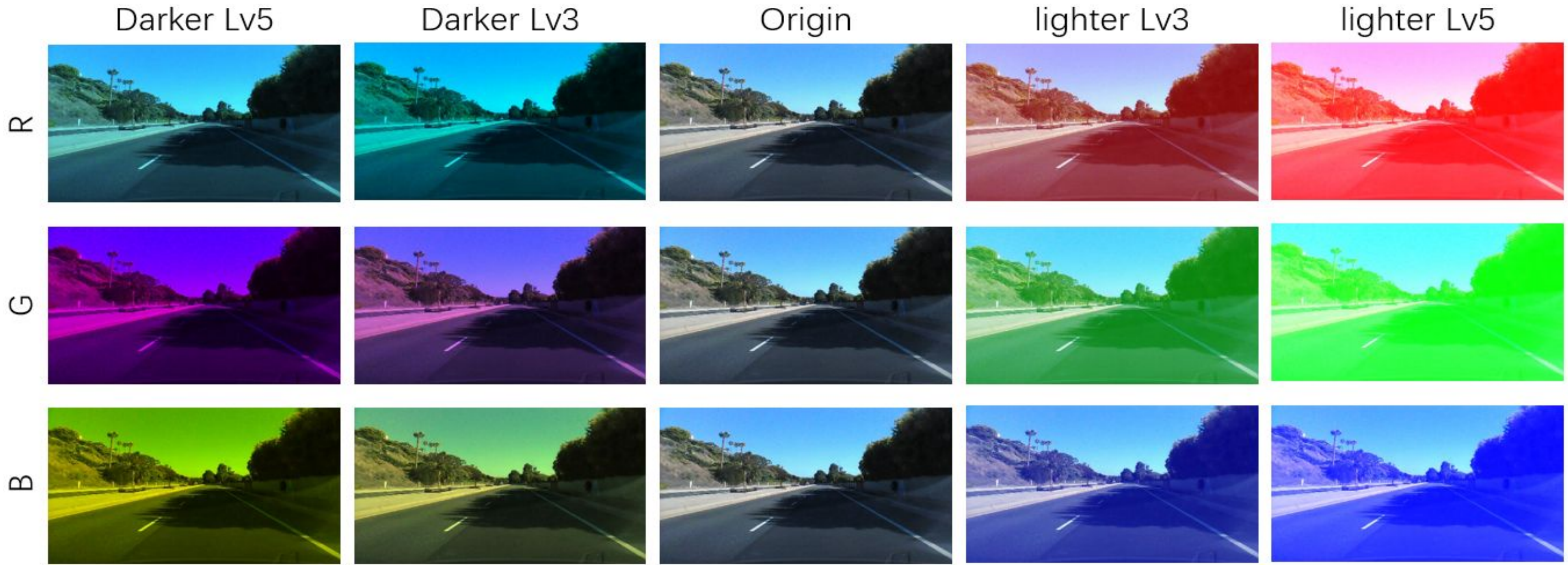
Images with Single Perturbation

- Blur, noise, and distortion: most commonly used and directly affect image quality



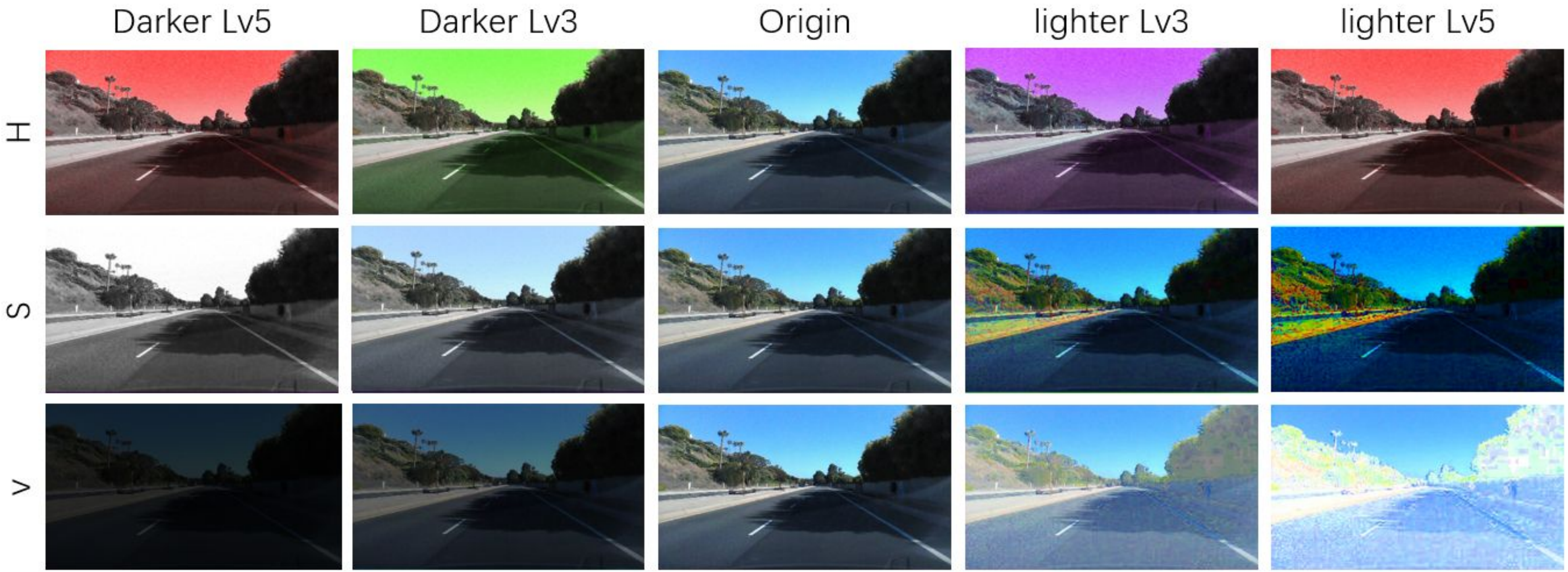
Images with Single Perturbation

- Red, Green, and Blue (RGB): 3 basic color representation of an image



Images with Single Perturbation

- Hue, Saturation and Brightness Value (HSV): 3 common metrics of an image



Single-Perturbation Videos: Blur, Noise, Distortion

Clean

Blur

Noise

Distortion

Single-Perturbation Videos: RGB

R-

R+

Clean

G-

G+

B-

B+

Single-Perturbation Videos: HSV

H-

H+

Clean

S-

S+

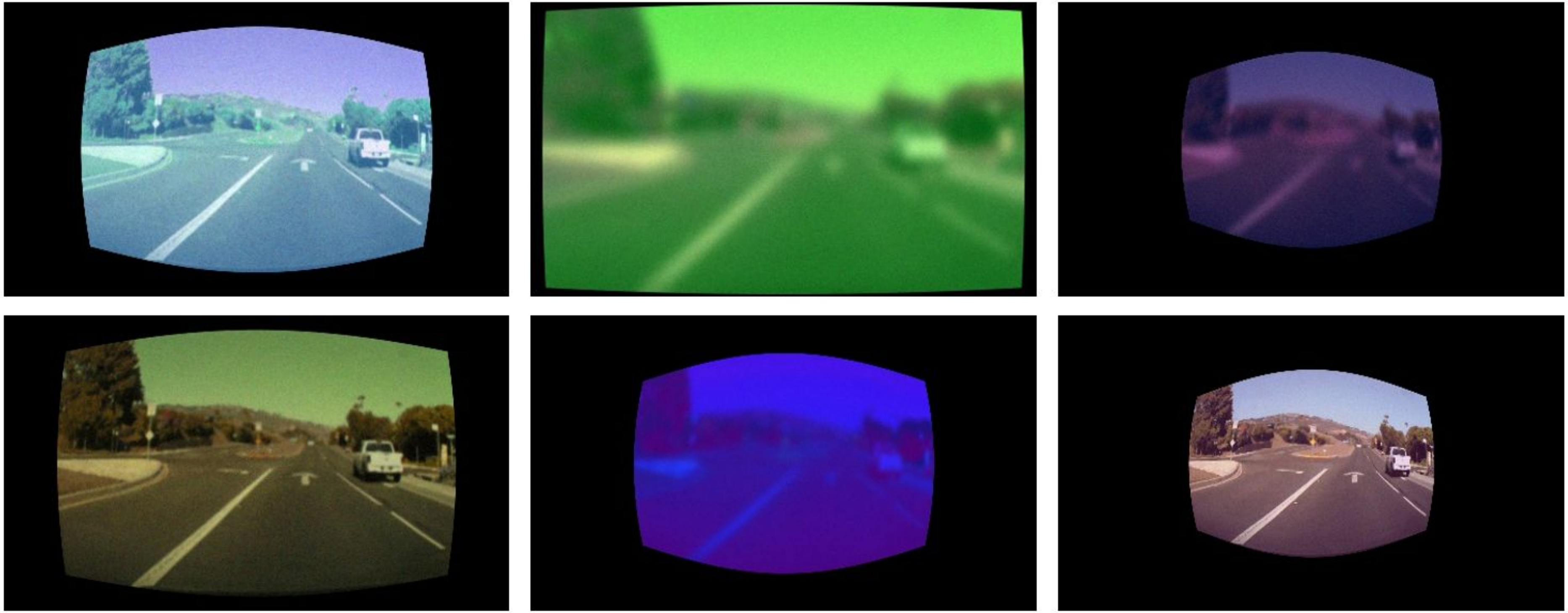
V-

V+

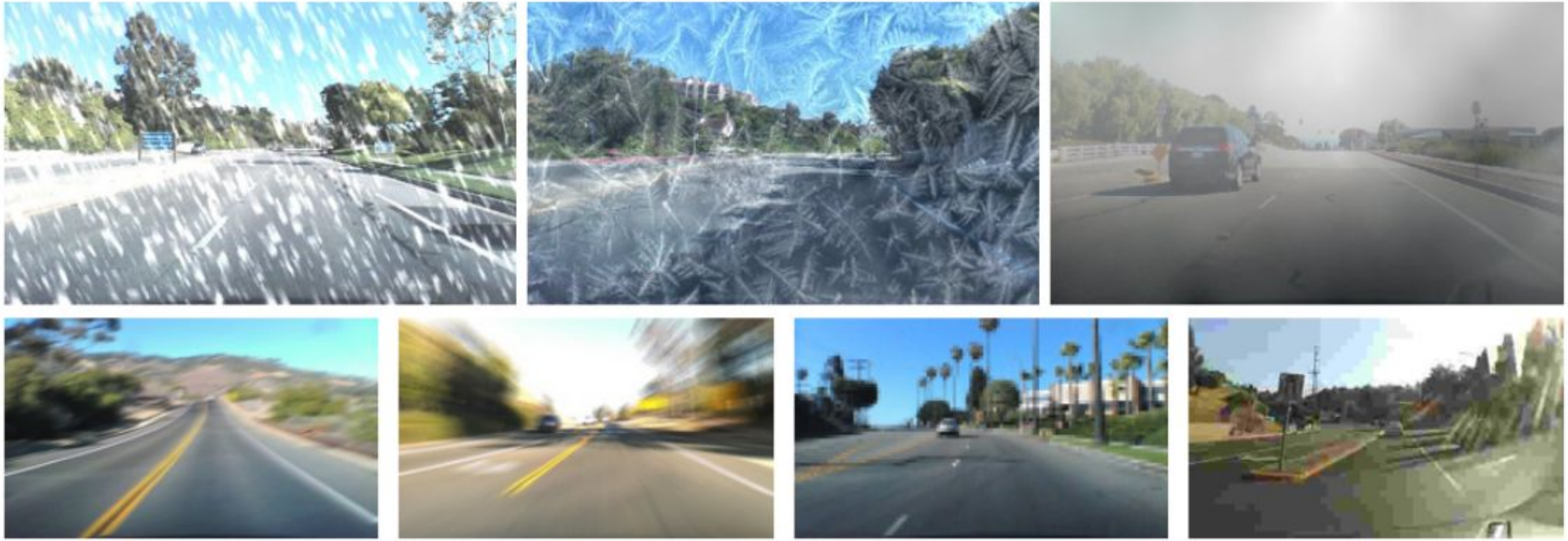
Experiments: Test Scenarios

- ***Clean Driving Data***
- ***Single Perturbation***
- ***Combined Perturbation:*** combinations of all seen factors w/ random severity
- ***Unseen Perturbation:*** motion blur, zoom blur, pixelate, jpeg compression, snow, frost, fog, etc. from ImageNet-C

Combined Perturbation Test Data Samples



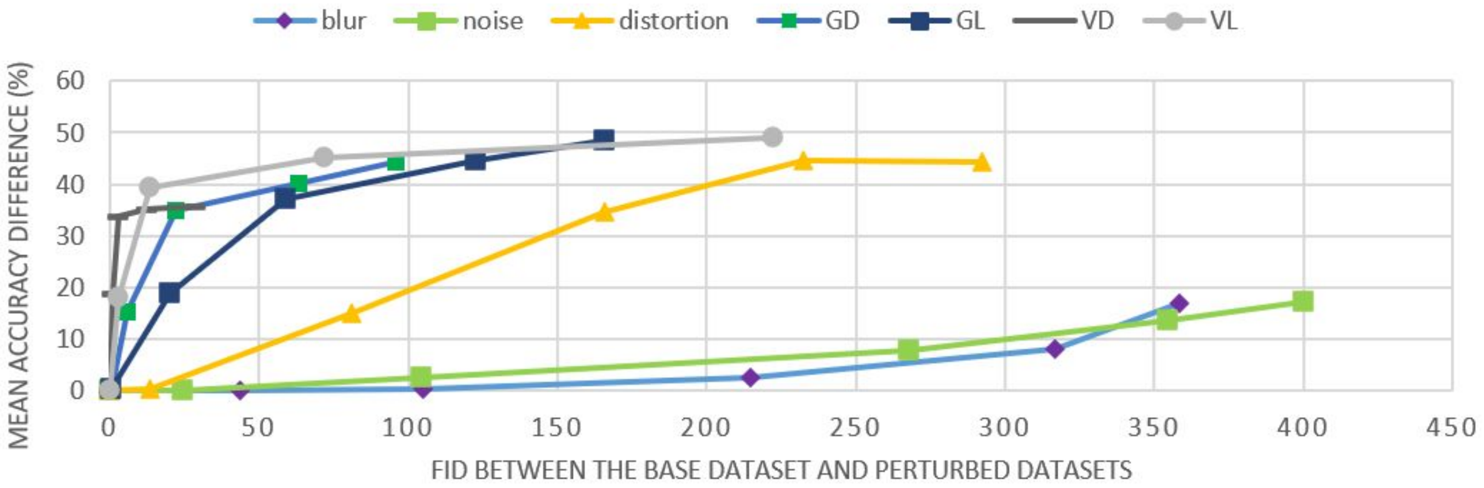
Unseen Perturbation Test Data Samples



left to right (top; bottom): snow, frost, fog;
motion blur, zoom blur, pixelate, jpg compression

Analysis: FID-based Parameterization

- *Fréchet Inception Distance* (FID): a distance metric considering image pixels & features, mean and covariance
- *FID vs. Mean Accuracy Difference*
  - More sensitive to the channel-level perturbations (i.e., R, G, B, H, S, V channels) than the image-level perturbations (i.e., from blur, noise, distortion).
  - least sensitive to blur and noise but most sensitive to the *intensity value* in V channel.



Experiments – Comparison with 5 different methods

Our method outperforms other SOTA methods

- Highest maximum & average MA improvements (MMAI & AMMI in %)
- Lowest mean corruption errors (mCE in %)

	Scenarios									
	Clean	Single Perturbation			Combined Perturbation			Unseen Perturbation		
Method	AMAI↑	MMAI↑	AMAI↑	mCE↓	MMAI↑	AMAI↑	mCE↓	MMAI↑	AMAI↑	mCE↓
Data Augmentation	-0.44	46.88	19.97	51.34	**36.1**	11.97	75.84	27.5	7.92	81.51
Adversarial Training	-0.65	30.06	10.61	74.42	17.89	6.99	86.82	16.9	8.17	89.91
MaxUp	-7.79	38.30	12.83	66.56	26.94	16.01	72.60	23.43	5.54	81.75
AugMix	-5.23	40.27	15.01	67.49	26.81	15.45	68.38	28.70	8.85	87.79
Ours w/o FS	0.93	48.57	20.74	49.47	33.24	17.74	63.81	29.32	9.06	76.20
Ours	**2.12**	**49.97**	**23.92**	**37.30**	33.15	**22.12**	**54.38**	**33.16**	**13.81**	**61.61**

Experiments – Comparison on 3 backbone networks

Our method outperforms AugMix on 3 networks

- Highest maximum & average MA improvements (MMAI & AMMI in %)
- Lowest mean corruption errors (mCE in %)

	Scenarios									
	Clean	Single Perturbation			Combined Perturbation			Unseen Perturbation		
Method	AMAI↑	MMAI↑	AMAI↑	mCE↓	MMAI↑	AMAI↑	mCE↓	MMAI↑	AMAI↑	mCE↓
AugMix+Nvidia	-0.12	40.64	10.94	76.48	25.97	16.79	64.41	**22.23**	5.99	84.95
Ours+Nvidia	**2.48**	**43.51**	**13.51**	**67.78**	**28.13**	**17.98**	**61.12**	16.93	**6.70**	**80.92**
AugMix+Comma.ai	-5.25	55.59	9.56	86.31	31.32	**0.77**	**106.1**	37.91	7.97	89.99
Ours+Comma.ai	**0.36**	**62.07**	**15.68**	**70.84**	**38.01**	0.74	108.32	**42.54**	**12.15**	**77.08**
AugMix+ResNet152	-4.23	20.84	1.45	96.24	12.21	6.71	80.19	15.40	2.87	97.62
Ours+ResNet152	**-0.96**	**24.29**	**5.19**	**79.76**	**16.05**	**8.02**	**75.16**	**16.58**	**5.33**	**85.68**

Experiments – Comparison on 4 driving datasets

Our method outperforms AugMix on 4(+1) datasets

- Highest maximum & average MA improvements (MMAI & AMMI in %)
- Lowest mean corruption errors (mCE in %)

	Scenarios									
	Clean	Single Perturbation			Combined Perturbation			Unseen Perturbation		
Method	AMAI↑	MMAI↑	AMAI↑	mCE↓	MMAI↑	AMAI↑	mCE↓	MMAI↑	AMAI↑	mCE↓
AugMix on SullyChen	-5.23	40.27	15.01	67.49	26.81	15.45	68.38	28.70	8.85	87.79
Ours on SullyChen	**1.46**	**49.76**	**22.87**	**40.84**	**33.15**	**22.12**	**54.38**	**33.87**	**13.51**	**62.50**
AugMix on Audi	-8.24	81.89	32.22	55.27	75.49	50.23	41.98	73.06	27.39	77.51
Ours on Audi	**5.98**	**97.57**	**48.50**	**10.27**	**87.56**	**62.38**	**25.80**	**77.22**	**32.71**	**39.14**
AugMix on Honda10k	-0.12	40.64	10.94	76.48	25.97	16.79	64.41	**22.23**	5.99	84.95
Ours on Honda10k	**2.48**	**43.51**	**13.51**	**67.78**	**28.13**	**17.98**	**61.12**	16.93	**6.70**	**80.92**
AugMix on Honda100k	-11.41	63.85	14.08	70.64	**68.95**	47.69	40.12	**61.68**	16.32	88.36
Ours on Honda100k	**-2.55**	**67.35**	**19.88**	**53.26**	65.10	**48.60**	**36.94**	51.90	**18.29**	**72.84**
AugMix on Waymo	18.27	45.40	23.30	59.31	**22.95**	16.92	66.36	**57.65**	29.10	55.63
Ours on Waymo	**20.34**	**46.85**	**26.76**	**52.84**	21.34	**18.24**	**64.58**	56.98	**31.12**	**53.18**

Experiments – Comparison on CIFAR-100 (Classification)

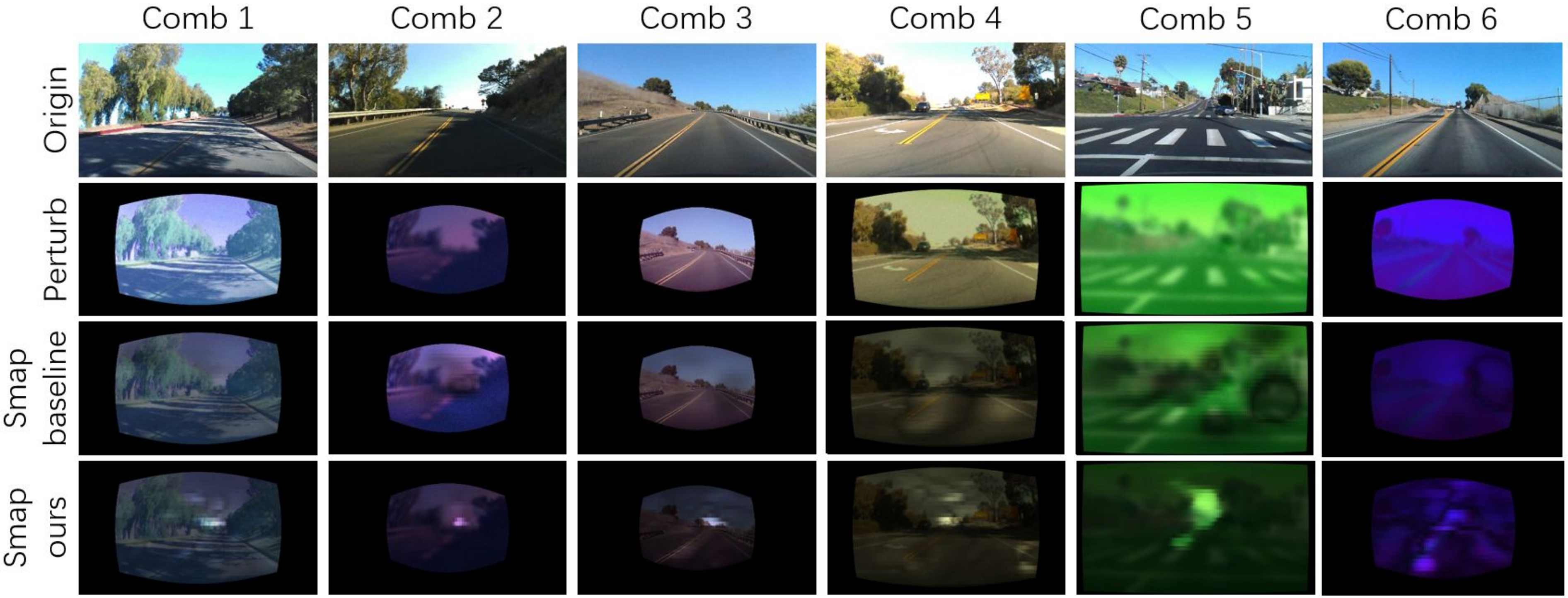
Our method outperforms 7 other methods on 4 different backbones

- Achieving ***lowest mean corruption errors*** (mCE in %)

	Standard	Cutout	Mixup	CutMix	AutoAugment*	Adv Training	AUGMIX	Ours
AllConvNet	56.4	56.8	53.4	56.0	55.1	56.0	42.7	**25.6**
DenseNet	59.3	59.6	55.4	59.2	53.9	55.2	39.6	**26.3**
WideResNet	53.3	53.5	50.4	52.9	49.6	55.1	35.9	**20.5**
ResNeXt	53.4	54.6	51.4	54.1	51.3	54.4	34.9	**15.4**
Mean	55.6	56.1	52.6	55.5	52.5	55.2	38.3	**22.0**

Experiments – Visualization

Saliency map for baseline vs. ours. With ours, the network can better focus on important areas (e.g., road in front) instead of random areas in baseline



Summary

- Leverage sensitivity analysis & FID-based parameterization
- Use adversarial data augmentation on single basis perturbations for training, improve performance on complex (combination or previously unseen) perturbations
- Improve accuracy & robustness for autonomous steering
- Achieve significant performance improvement
  - Up to ***97%*** on a *single* source of image degradation
  - Up to ***87%*** on combinations of *multiple* perturbations
  - Up to ***77%*** on previously *unseen* image corruption

Thank you!