WED-AM-214

Improving Robustness of Semantic Segmentation to Motion-Blur Using Class-Centric Augmentation

Aakanksha , Rajagopalan A. N.

Image Processing and Computer Vision (IPCV) Lab

Indian Institute of Technology Madras

JUNE 18-22, 2023

CVPR

VANCOUVER, CANADA

Overview:

Input Blurred Image

Standard Semantic Segmentation Network

Predicted Segmentation Map

Input Blurred Image

Semantic Segmentation Network trained with our augmentation

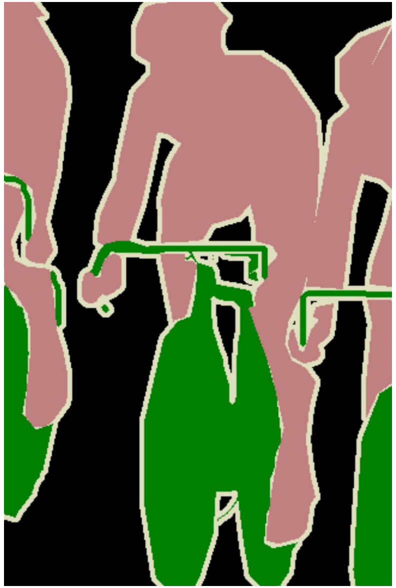
Predicted Segmentation Map

Semantic Segmentation:

- Classify each pixel to a class from a set of object/stuff classes

Input Image

Predicted Segmentation Map

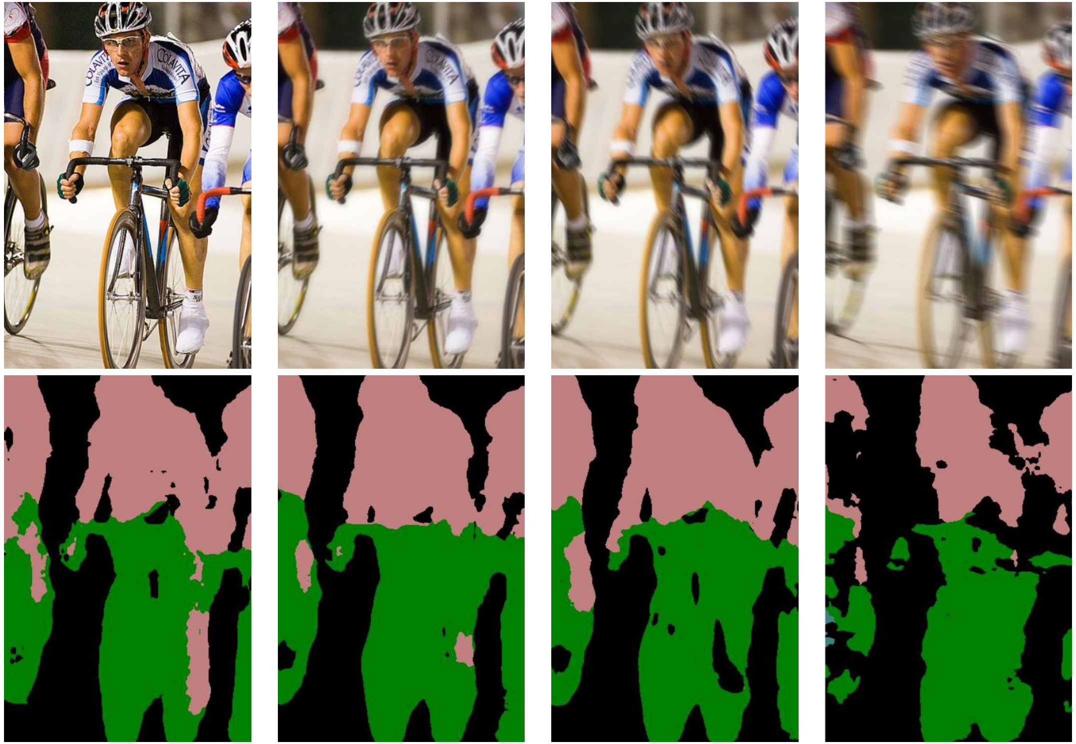


Ground Truth
Segmentation Map

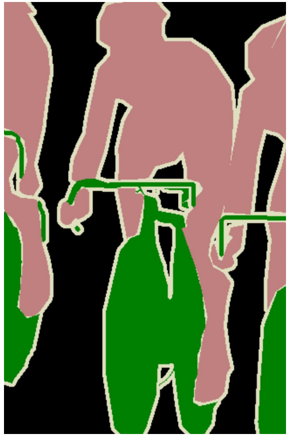
Robustness of Semantic Segmentation:

Input Image

Predicted Segmentation Map



Sharp | L1 | L2 | L3



Ground Truth
Segmentation Map

** Note that blur severity increases from L1 to L3

Related Works:

- Investigating the impact of blur on performance [1]

- Benchmarking robustness to common corruptions and perturbations [2,3]

- Increasing robustness to generic common corruptions and perturbations [4]

[1] Igor Vasiljevic, Ayan Chakrabarti, and Gregory Shakhnarovich. Examining the impact of blur on recognition by convolutional networks. arXiv preprint arXiv:1611.05760, 2016. 3
[2] Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. Proceedings of the International Conference on Learning Representations, 2019.
[3] Christoph Kamann and Carsten Rother. Benchmarking the robustness of semantic segmentation models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 8828–8838, 2020.
[4] Christoph Kamann and Carsten Rother. Increasing the robustness of semantic segmentation models with painting-by-numbers. In European Conference on Computer Vision, pages 369–387. Springer, 2020.

Key Idea:

- Motion blur is common and unavoidable and causes performance drop

- Motion blur is diverse and challenging

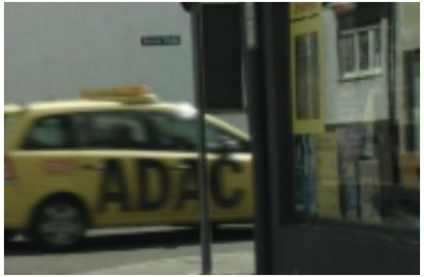
- Augment training images with synthetic motion-blurred images (comprising the entire spectrum from dynamic scenes to camera-shake blur) to improve robustness to motion-blur in semantic segmentation

Generating synthetic motion-blur:

- GANs - more realistic but offer no control or interpretability
- Motion-blur kernels can be used along with ground truth segmentation maps to generate synthetic dynamic-scene motion-blur

Sharp Image

GT Segmentation Map

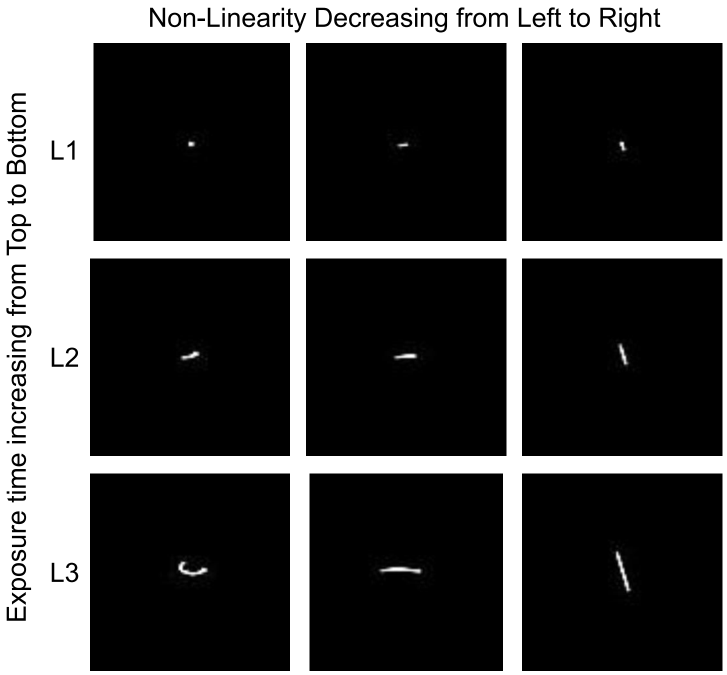


Synthetic dynamic-scene blur

Blur Kernel Generation:

- Diverse blur kernels are generated [5] by varying -
  (a) non-linearity of the camera trajectory, and
  (b) exposure time of the camera

- Increased exposure time leads to more severe blur. So, L1 corresponds to lowest blur and lowest exposure time while L3 corresponds to the highest.

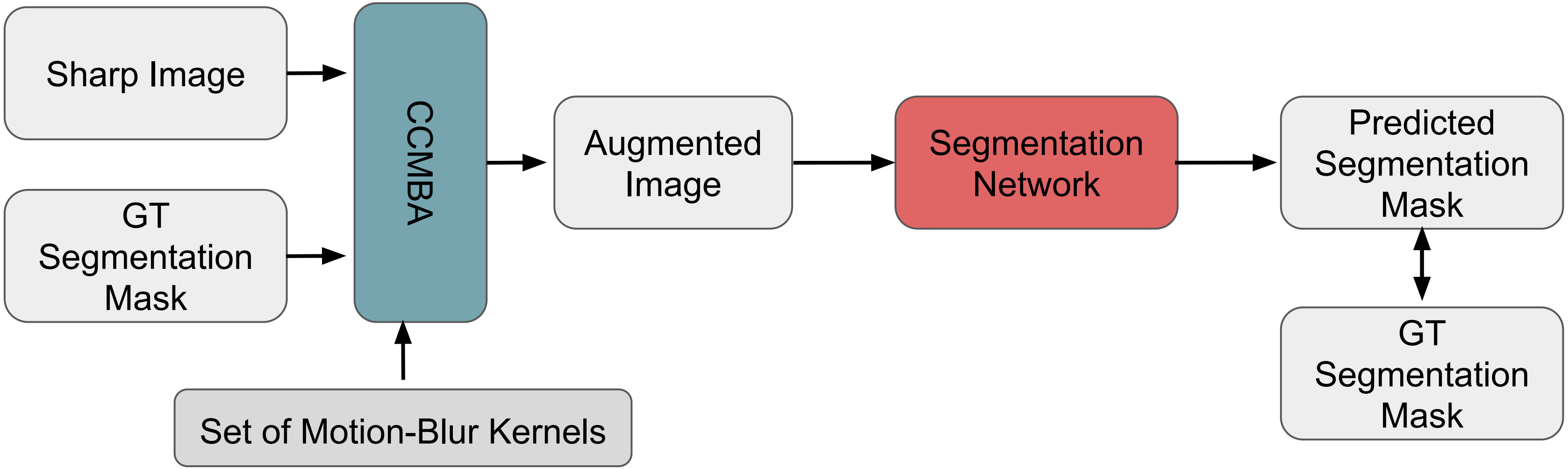


** Note that the shown blur kernels are enlarged for better visibility.

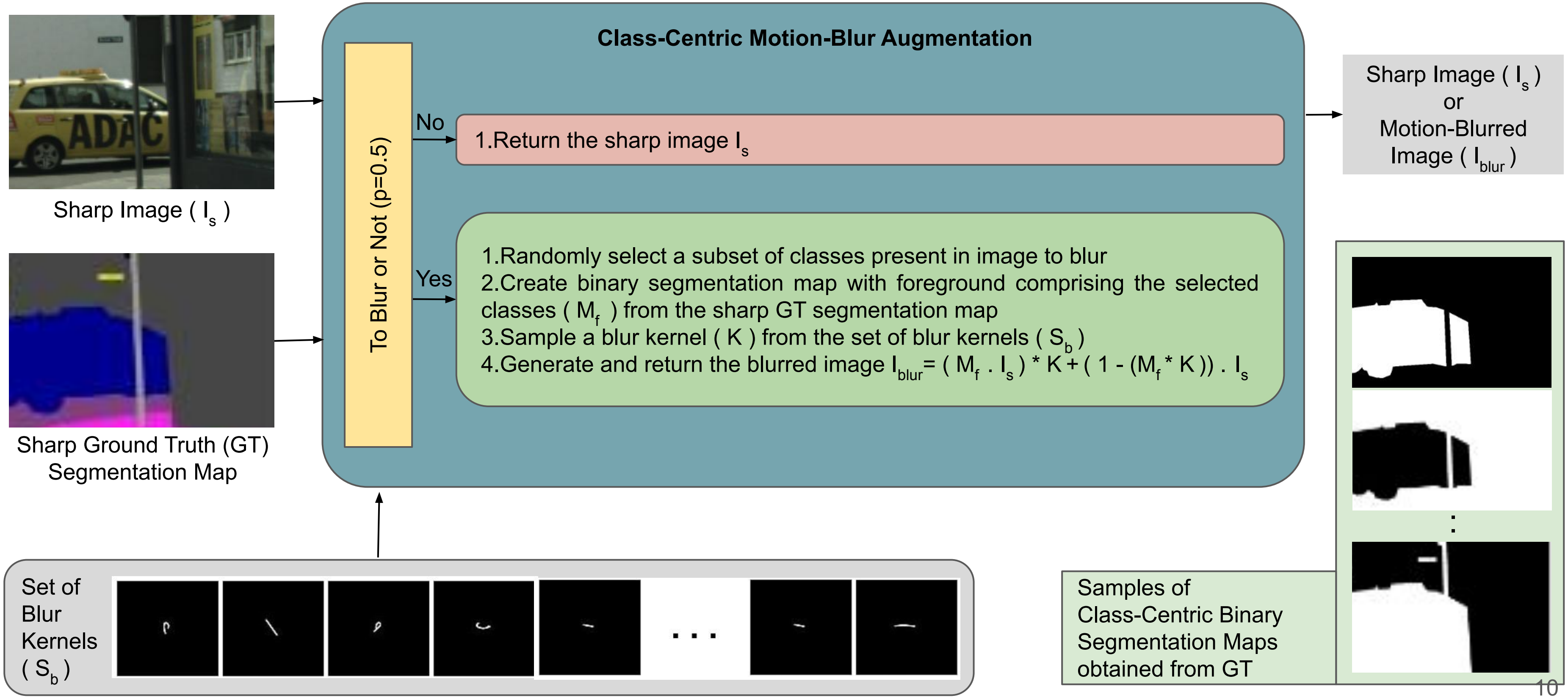
[5] Giacomo Boracchi and Alessandro Foi. Modeling the performance of image restoration from motion blur. IEEE Transactions on Image Processing, 21(8):3502–3517, 2012.

Outline of Our Approach:

- Our approach is class-centric motion blur augmentation (CCMBA)



Class-Centric Motion Blur Augmentation:



Samples of Augmented Image

Sharp Image

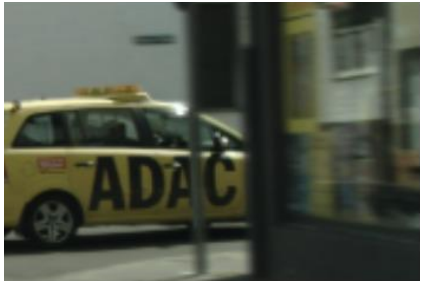
Only Airplane Blurred

Only Airplane Sharp

All Blurred

Sharp Image

Only Car Blurred



Only Car Sharp

All Blurred

Baseline Methods :

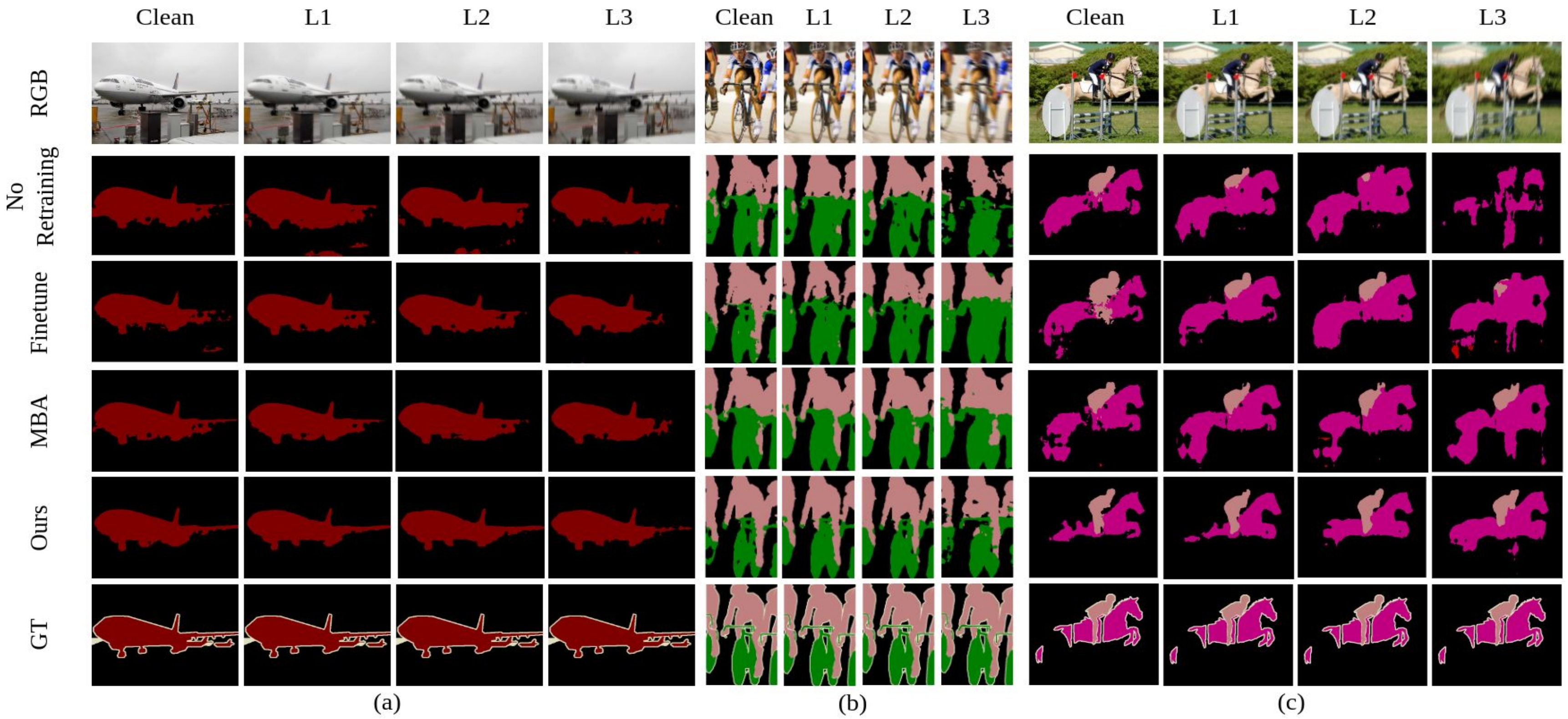
- *No Retraining* - to check pre-trained models's robustness to blur
- *Deblurring* - to check the effectiveness of deblurring as a pre-processing step
- *Finetuning* - to enable the network to learn the slight domain shift
- *Motion Blur Augmentation (MBA)* - to compare space invariant blur augmentation with our class-centric space-variant blur augmentation

Results On Synthetic Generated Blur:

Method	VOC				Cityscapes							
	DeepLabv3+				DeepLabv3+				Segformer			
	Clean	L1	L2	L3	Clean	L1	L2	L3	Clean	L1	L2	L3
No-Retraining	**77.2**	69.6	53.1	36.5	<u>75.6</u>	70.4	58.1	41.4	<u>81.0</u>	78.2	73.2	62.5
Deblurring	-	69.3	65.9	58.2	-	72.2	70.9	66.5	-	<u>78.5</u>	77.5	<u>75.3</u>
Finetuning	67.4	71.9	<u>69.6</u>	<u>63.9</u>	70.6	<u>74.2</u>	70.6	<u>68.3</u>	79.8	**80.2**	**79.1**	**76.0**
MBA	74.6	<u>72.9</u>	69.2	60.3	60.4	73.3	<u>71.2</u>	66.9	79.6	<u>78.5</u>	77.0	74.1
CCMBA (Ours)	<u>76.5</u>	**74.6**	**72.1**	**66.0**	**76.2**	**75.6**	**73.6**	**70.4**	**81.1**	**80.2**	<u>78.7</u>	**76.0**

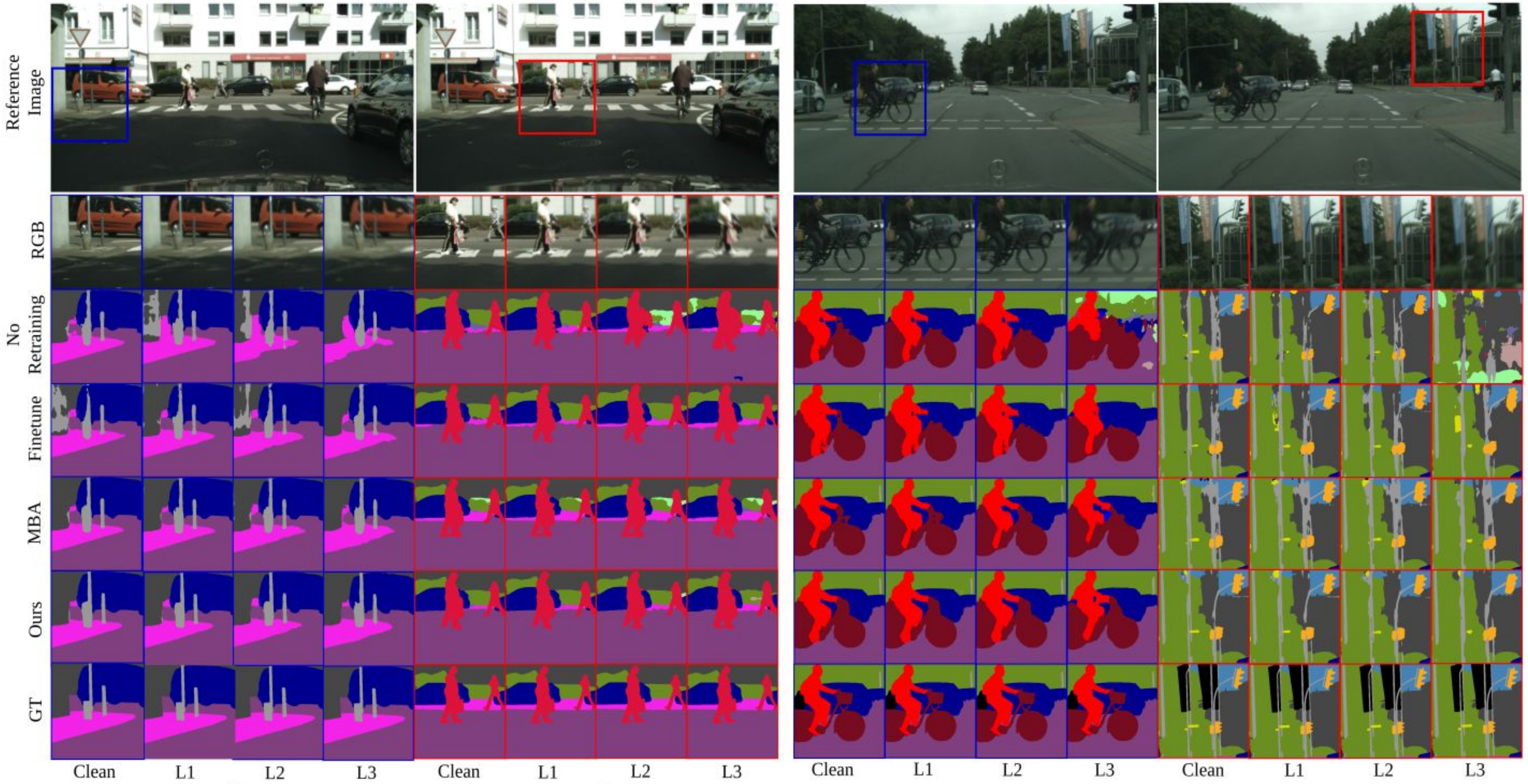
- Our augmentation improves the performance in the presence of blur while retaining the performance for clean/sharp images.

Results On Synthetically Generated Blur (VOC):



- More consistent results are obtained across clean and blurred data (with different severity levels) using our approach.

Results On Synthetically Generated Blur (Cityscapes):



- More consistent results are obtained across clean and blurred data (with different severity levels) using our approach.

Results On Cityscapes-C:

Method	DeepLabv3+				Segformer			
	Clean	S1	S2	S3	Clean	S1	S2	S3
No-Retraining	75.6	71.2	65.7	56.9	<u>81.0</u>	<u>77.8</u>	74.6	68.8
Finetuning	70.6	<u>72.4</u>	<u>70.4</u>	<u>68.1</u>	79.8	**78.0**	**76.0**	**73.1**
MBA	60.4	57.9	54.5	50.8	79.6	77.4	75.3	71.9
PbN [6]	<u>76.1</u>	72.3	68.7	63.2	-	-	-	-
CCMBA (Ours)	**76.2**	**74.0**	**72.3**	**68.9**	**81.1**	77.7	<u>75.9</u>	<u>72.3</u>

- Our augmentation improves the performance in the presence of blur while retaining the performance for clean/sharp images even on Cityscapes-C benchmark.

[6] Christoph Kamann and Carsten Rother. Increasing the robustness of semantic segmentation models with paintingby-numbers. In European Conference on Computer Vision, pages 369–387. Springer, 2020.

Results on Real Blur (GoPro and REDS):

- Our approach generalizes better to real blur as well and maintains better consistency across sharp and blur images when compared with baselines.

Conclusion

- An effective data augmentation scheme using ground truth segmentation maps and synthetic blur kernels was proposed for improving semantic segmentation robustness to motion blur.

- The class-centric nature of our augmentation enables it to perform well on real blur datasets like GoPro and REDS, especially for common classes like humans.

Thank You!