# Munich to Dubai: How far is it for Semantic Segmentation? (Supplementary Material)

Shyam Nandan Rai[1], Vineeth N Balasubramanian[2], Anbumani Subramanian[3], C. V. Jawahar[4]
[1,4]IIIT Hyderabad, [2]IIT Hyderabad, [3]Intel Corporation
[1]shyam.nandan@research.iiit.ac.in, [2]vineethnb@iith.ac.in,
[3]anbumani.subramanian@intel.com, [4]jawahar@iiit.ac.in

***Summary:*** We describe the parameters used to simulate atmospheric turbulence images and analyze the effect of change in turbulence parameters on segmentation results. Then, we perform an ablation study on the loss functions used in our restoration and segmentation network. Finally, we group all the small object classes of Cityscapes dataset according to their importance for autonomous applications and show class-wise improvement in semantic segmentation by using our proposed framework.

## 1. Parameters of Turbulent Images

We use an efficient method proposed by [4], which was computationally inexpensive as compared to computer graphics methods. The imaging parameters we use for the simulation are: focal distance of the camera = $300mm$ with lens diameter $\approx 5.357cm$ having a pixel size $4e-3mm$. The virtual imaging system is placed at an elevation of $4m$. Structure constant $C_n^2$ is an important variable for measuring atmospheric turbulence. Higher the value of $C_n^2$ stronger the turbulence. $C_n^2$ = 0 indicates a medium free of turbulence. The value for structure constant $C_n^2$ here is $9e-14m^{-2/3}$. Light traveling from the object is assumed to be having a spherical wavefront with a wavelength of $550nm$. Figure 1 shows an example of the simulated image on these parameters.

**Figure 1:** Shows the simulated turbulent images of Cityscapes dataset. The green boxes shows the enlarged patches of the turbulent image. It is evident that the linear structure of objects such as poles is lost due to the geometrical distortion caused by atmospheric turbulence.

## 2. Additional Ablation Studies

**Loss Function:** We perform an ablative study on the loss function $l_{gen}$ (equation 1), which is used to train our restoration network. We train all our models for 15 epochs with a learning rate of $1e-4$ on the Cityscapes dataset. Table 1 shows the performance of our restoration model when it is trained on different loss components of $l_{gen}$. We find that removing adversarial loss from $l_{gen}$ drastically reduces the performance of the restoration network. Whereas, the addition of perceptual loss in $l_{gen}$ gives a marginal improvement in the restoration. We also show the effectiveness of using a natural logarithm on the CORAL loss, which is loss component of the joint segmentation loss (equation 3). $l_{coral}$ (CORAL loss component in equation 3) is the natural logarithm of CORAL loss. The idea for applying logarithm on top of the CORAL loss was to handle the overshooting values of CORAL loss, which made the training of the segmentation network more stable. From Table 2, we

| Loss | PSNR | SSIM | MS-SSIM | MSE |
|---|---|---|---|---|
| $l_{con}$ | 18.438 | 0.4226 | 0.7312 | 299.024 |
| $l_{con}+l_{adv}$ | 24.245 | 0.7725 | 0.9404 | 153.970 |
| $l_{con}+l_{adv}+l_{per}$ | 25.914 | 0.8042 | 0.9611 | 134.682 |

**Table 1:** Shows the performance of our restoration network, when trained on different components of $l_{gen}$. We can see by removing perceptual loss and adversarial loss from $l_{gen}$, there is a drastic decrease in the restoration performance observed in the all image quality metric, whereas, the addition of perceptual loss to $l_{gen}$ improves the restoration marginally. $l_{con}$, $l_{adv}$, and $l_{per}$ are the content loss, adversarial loss, and perceptual loss respectively.

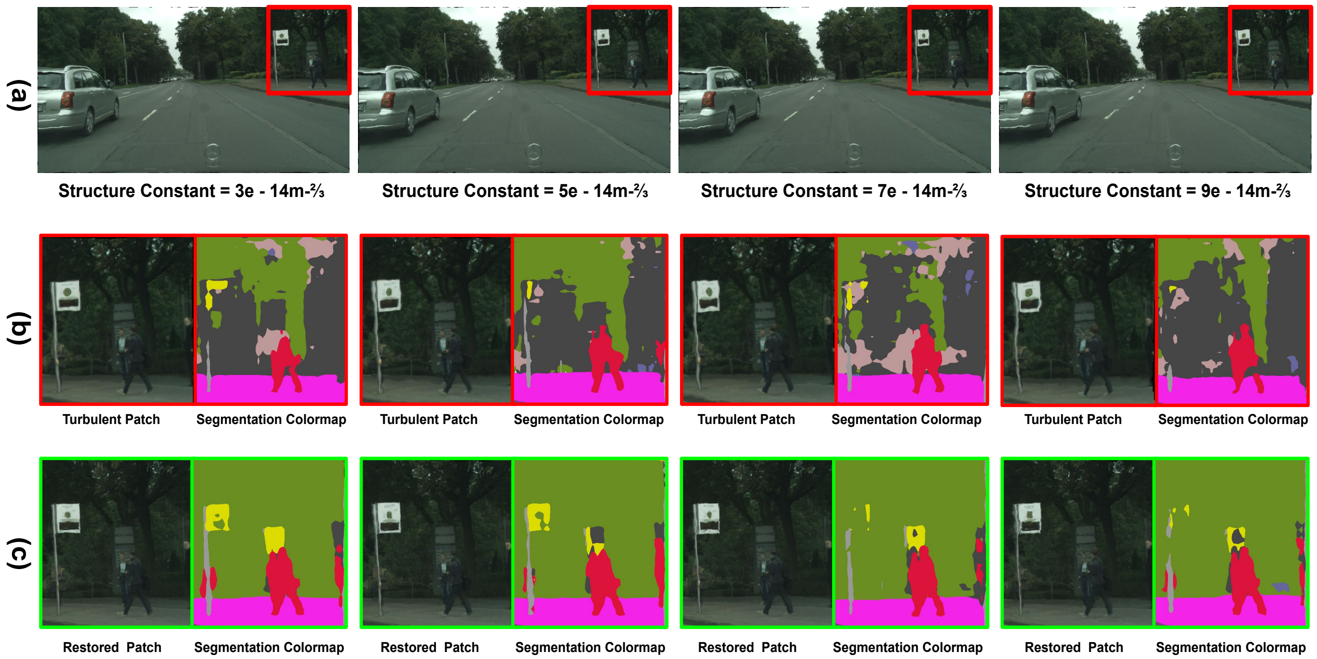


**Figure 2:** The figure shows the segmentation results on the restored and turbulent image at different values of structure constant $C_n^2$. We can notice as the value of $C_n^2$ increases, the segmentation results start to degrade (observe the segmentation result of the white signboard in the magnified patch). (a) Turbulent images at different values of $C_n^2$. (b) Magnified turbulent image patch (red box) and its corresponding segmentation colormap by DeepLabV3. (c) Restored image patch and corresponding segmentation colormap obtained from our framework.

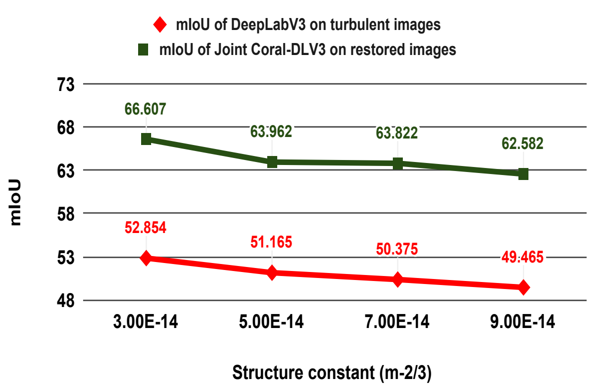


**Figure 3: mIoU vs structure constant($C_n^2$).** Comparison of the performance of the semantic segmentation model on the set of turbulent and restored images of Cityscapes. The plot shows that the performance of DeepLabV3 on turbulent images and Joint Coral-DLV3 on restored images decrease as the turbulence in the environment increases, which is measured by $C_n^2$.

see a marginal improvement in segmentation results on the Cityscapes dataset by using $l_{coral}$.

| Loss | mIoU |
|---|---|
| Coral Loss | 56.831 |
| $l_{\text{coral}}$ | 57.011 |

**Table 2:** Improvement in segmentation results by applying logarithm to the CORAL loss component of joint segmentation loss (equation 3). $l_{\text{coral}}$ is the natural logarithm of CORAL loss.

**mIoU vs Structure Constant:** Structure Constant $C_n^2$ [3] is an important parameter to measure the atmospheric turbulence as it is directly proportional to the atmospheric temperature. Hot cities like Dubai normally have $C_n^2$ values which are in the order of $10^{-14}m^{-2/3}$ which increases to $10^{-13}m^{-2/3}$ in extremely hot days. Hence, to show the effect of increasing $C_n^2$ on our segmentation model, we vary the value of $C_n^2$ from $3e - 14m^{-2/3}$ to $9e - 14m^{-2/3}$ and show segmentation result at different $C_n^2$. Figure 3 shows the mIoU for turbulent and restored images on Munster city frames present in the Cityscapes dataset. We can see, as the value of $C_n^2$ increases the mIoU for turbulent and restored images decreases. Figure 2 shows the

| Method | Group 2 | Group 3 | | Group 4 | | | | mIoU |
|---|---|---|---|---|---|---|---|---|
| | pole | tlight | tsign | person | rider | mbike | bicycle | |
| DeepLabV3 [2] | 23.14 | 30.39 | 37.51 | 44.06 | 15.43 | 15.65 | 29.41 | 27.941 |
| Joint Coral-DLV3 | 33.28 | 39.29 | 43.59 | 60.73 | 39.21 | 36.41 | 52.79 | 43.614 |
| *IoU Gain* | *10.15* | *8.90* | *6.08* | *16.67* | *23.78* | *20.76* | *23.39* | *15.675* |

**Table 3:** Shows the segmentation improvement for the small object classes, which are grouped according to their importance [1] in the autonomous navigation system. The classes become more valuable with the increase in the group number. We can also observe from the table, the segmentation improvement from atmospheric turbulence is highest in group 4, which consists of the most important and valuable classes for autonomous navigation. In the table, tlight, tsign, and mbike denote the traffic light, traffic sign, and motorcycle classes of Cityscapes dataset.

| Channel Attention | Inter-connection | Multi-scale | PSNR | SSIM | MS-SSIM | MSE |
|---|---|---|---|---|---|---|
| ✗ | ✗ | ✗ | 20.9341 | 0.6013 | 0.8477 | 244.04 |
| ✗ | ✗ | ✓ | 21.6617 | 0.6249 | 0.8615 | 235.33 |
| ✗ | ✓ | ✓ | 22.0432 | 0.6495 | 0.8889 | 217.12 |
| ✓ | ✓ | ✓ | 22.1411 | 0.6517 | 0.8910 | 214.68 |

**Table 4:** Ablation of our proposed CA-MSRB block. We empirically find that combining all the component gave the best performance.

qualitative analysis on different values of $C_n^2$.

**Channel Attentive Multi-scale Residual Block:** Next, we perform ablation study on our proposed CA-MSRB block. The CA-MSRB block consist of channel-attention, inter connection between the convolution layers and has multi-scale convolutional layers. Table 4 shows the performance of each component of CA-MSRB.

## 3. Performance on Small and Important Objects

Small classes of the Cityscapes dataset such as poles, traffic lights, or person are severely affected by atmospheric turbulence, as shown in Figure 1. The reason behind such cause is that small objects like poles lose their linear shape and become geometrically distorted, whereas, for large classes like sky and road, the impact of turbulence is not so prominent. But, these small class are much more importance [1] when trained semantic segmentation models are applied to the autonomous driving system. For instance, rider or person class as these classes are more vulnerable and valuable than the sky in case of self-driving cars. Hence, by using our proposed framework, we analyze the improvement in segmentation accuracy for the small objects, which are grouped according to their importance. Table 3 shows the improvement in semantic segmentation accuracy for small object categories of Joint Coral-DLV3, compared against DeepLabV3.

## References


[1] B.-k. Chen, C. Gong, and J. Yang. Importance-aware semantic segmentation for autonomous driving system. In *IJCAI*, 2017.

[2] L. Chen, G. Papandreou, F. Schroff, and H. Adam. Rethinking atrous convolution for semantic image segmentation. *CoRR*, abs/1706.05587, 2017.

[3] D. L. Fried. Limiting resolution looking down through the atmosphere. *JOSA*, 56(10):1380–1384, 1966.

[4] A. Schwartzman, M. Alterman, R. Zamir, and Y. Y. Schechner. Turbulence-induced 2d correlated image distortion. In *ICCP*, 2017.