Spatial Feedback Learning to Improve Semantic Segmentation in Hot Weather

by

Shyam Nandan Rai, Vineeth Balasubramanian, Anbumani Subramanian, C V Jawahar

in

BMVC 2020

Report No: IIIT/TR/2020/-1

Centre for Visual Information Technology
International Institute of Information Technology
Hyderabad - 500 032, INDIA
September 2020

Spatial Feedback Learning to Improve Semantic Segmentation in Hot Weather

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

[1] IIIT Hyderabad
[2] IIT Hyderabad
[3] Intel Corporation

Abstract

High-temperature weather conditions induce geometrical distortions in images which can adversely affect the performance of a computer vision model performing downstream tasks such as semantic segmentation. The performance of such models has been shown to improve by adding a restoration network before a semantic segmentation network. The restoration network removes the geometrical distortions from the images and shows improved segmentation results. However, this approach suffers from a major architectural drawback that is the restoration network does not learn directly from the errors of the segmentation network. In other words, the restoration network is not task aware. In this work, we propose a semantic feedback learning approach, which improves the task of semantic segmentation giving a feedback response into the restoration network. This response works as an attend and fix mechanism by focusing on those areas of an image where restoration needs improvement. Also, we proposed loss functions: Iterative Focal Loss (iFL) and Class-Balanced Iterative Focal Loss (CB-iFL), which are specifically designed to improve the performance of the feedback network. These losses focus more on those samples that are continuously miss-classified over successive iterations. Our approach gives a gain of 17.41 mIoU over the standard segmentation model, including the additional gain of 1.9 mIoU with CB-iFL on the Cityscapes dataset.

1 Introduction

Standard image datasets such as ImageNet [6], Cityscapes [4], and IDD [28] are often taken in a clear and well-illuminated environment. However, real-world images often suffer from variations in weather conditions such as rain, fog, snow, and temperature. Computer vision models trained on standard datasets to perform tasks such as segmentation, detection, and classification often struggle to overcome performance degradation when tested on such real-world images with weather variations. Hence, to overcome such problems, restoration

Code available at: https://github.com/shyam671/Spatial-Feedback-Learning-to-ImproveSemantic-Segmentation-in-Hot-Weather.

networks [9, 19, 23, 33] can be added before the vision models, to minimize the domain shift caused by different weather conditions to improve performance.

In this work, we focus on improving the performance of the semantic segmentation model in hot weather conditions, such as in certain areas of the Middle East, Africa, Asia or even other places. We choose semantic segmentation because it is highly sensitive to a slight domain shift due to dense labeling and its wide application in autonomous driving systems. Hot weather conditions, also termed as atmospheric turbulence [34], introduce geometrical distortions into images, leading to incorrect perceptions of concepts and erroneous semantic segmentation. We address this problem by adding a restoration network before the semantic segmentation model that removes the geometrical distortions. The problem with such a two-stage framework is that the restoration network does not directly learn from the errors of the semantic segmentation network. In other words, the restoration model is not task-aware. To overcome this problem, we introduce a feedback module that uses the information obtained from image regions with incorrect segmentation prediction, while training the restoration network. The additional information through the feedback module helps the restoration network focus on image regions where restoration needs to be improved, thus resulting in better segmentation. We repeat this process for several iterations to refine the final result. In summary, the feedback module provides the restoration network with an attend-and-fix mechanism through which it progressively improves the segmentation results. Previously used loss functions [16] in feedback frameworks gave constant weights to all samples in an iteration and did not focus on those samples that are consistently misclassified across training iterations. This can however be achieved by using the idea of focal loss [18], which gives more focus on highly misclassified samples and less focus on well-classified samples. In this work, we hence propose Iterative Focal Loss (iFL) that progressively focuses on those samples that are consistently misclassified over the iterations. Also, to handle class imbalance, we propose Class-Balanced Iterative Focal Loss (CB-iFL). We perform extensive experiments on the Cityscapes dataset showing the efficacy of our feedback module and loss functions. Our ablation study shows that our feedback can remove noisy predictions and improve the semantic segmentation in atmospheric turbulence. We also demonstrate that our method can correctly segment classes occupying small area such as 'rider' and 'poles', which are important classes for autonomous driving systems. In summary, our key contributions are:

- We introduce the notion of a semantic feedback module in an end-to-end framework that improves semantic segmentation in hot weather conditions/atmospheric turbulence. Our feedback module provides an attend-and-fix mechanism for better restoration of images in atmospheric turbulence.
- We propose two general loss functions: Iterative Focal Loss (iFL) and Class-Balanced Iterative Focal loss (CB-iFL) designed to improve the feedback framework by increasing focus on consistently misclassified samples and handle imbalance in a dataset.
- We conduct a comprehensive suite of experiments to study the proposed methodology and loss functions on the Cityscapes dataset, and show the promise of this method across these empirical studies.

2 Related Work

Computer Vision across Weather Conditions: It is well-known that dynamic changes in weather due to rain, haze, and atmospheric turbulence adversely affect [13, 20] the perfor-

mance of computer vision algorithms. Several physics-based methods [10, 11] have been proposed to circumvent the tedious task of data collection for individual weather conditions. Recently, meta-learning based methods [29] have become popular for generating such datasets. These dataset generation efforts have helped in building deep learning models [15, 17] that attempt to remove distortions caused due to various weather conditions. In particular, for the case of atmospheric turbulence, most efforts are based on classical methods such as adaptive optics [22], lucky imaging [8], and Fourier analysis [31] to remove such distortion present in an image. Apart from this, machine learning-based approaches have also been proposed [30, 34] to remove atmospheric turbulence. As mentioned earlier, all of these efforts are not task-aware, and focus only on removal of distortion, independent of the task at hand. Very recently, Rai *et al*. [23] proposed a deep learning methodology to remove atmospheric turbulence in the context of semantic segmentation. However, their framework is also a two-stage approach and is hence not end-to-end trainable, with no explicit feedback from the semantic segmentation module. We overcome these limitations in our work, while using lighter and efficient learnable models.

Feedback Mechanisms in Deep Learning: Initial approaches of integrating feedback into deep learning models can be traced to [1, 2, 21], where a feedback loop was used to improve the performance of hand pose estimation and human pose estimation tasks. Later, Li et al [32] proposed a general feedback framework that used recurrent networks for improving the performance of vision tasks. The feedback connection, however, had no learnable network. Recently, Shama *et al*. [27] proposed a learnable feedback mechanism that has been used to improve the generation quality in Generative Adversarial Networks. There have also been recent feedback approaches [16] used to improve the performance of models in image super-resolution tasks. However, none of these existing efforts focus on the problem we are addressing in this work.

Loss Functions for Feedback Networks: Feedback networks involve iteratively training a network, such that samples that are misclassified in a previous iteration are given more penalty in the current iteration. This provides the seed for our idea, which involves changing the loss function over the iterations to reflect this characteristic of feedback networks. Earlier feedback methods [21] did not exploit this idea, rather they used the same loss function across all iterations. Recently, Zamir *et al*. [32] introduced episodic curriculum learning, where they adopted an iteration-varying loss to enforce a curriculum. Later, Li *et al*. [16] used a similar idea with a weighted iterative L_1 loss function, to improve the performance of image super-resolution models. However, neither of these efforts explicitly increased the model's focus on previously misclassified samples. We overcome this problem by proposing a loss function for this purpose in our feedback framework which also shares similarities with the AdaBoost [7] algorithm.

3 Semantic Feedback Learning: Methodology

3.1 Problem Formulation

We begin our formulation with a set of turbulent images: $I_T = \{I_T^i : i = 1,2,\ldots,n\}$ and a corresponding non-turbulent image set: $I_{NT} = \{I_{NT}^i : i = 1,2,\ldots,n\}$ and semantic segmentation annotation set: $I_S = \{I_S^i : i = 1,2,\ldots,n\}$, where n is the total number of images. Our network architecture consists of three modules: a restoration network R, a semantic segmentation network S, and a feedback network F. The restoration network R follows an encoder-decoder

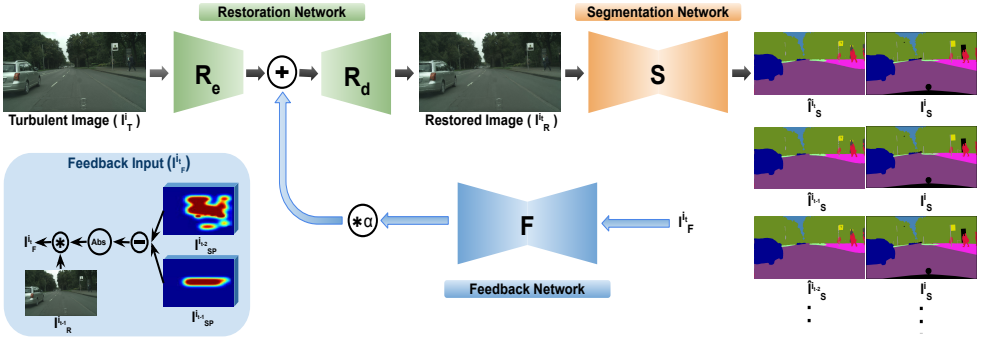

Figure 1: **Semantic feedback learning framework:** Our architecture consists of 3 networks: a restoration network R, a segmentation network S, and a feedback network F. An input image I^i_T is passed through encoder R_e of R, whose output is modified by F for better restoration in areas where outputs of S in a previous iteration were incorrect. The modified output of R_e is then passed through R_d to give $I^{i_t}_R$, at iteration t. $I^{i_t}_R$ is further passed into S to give $\hat{I}^{i_t}_S$ and $I^{i_t}_{SP}$. The feedback input $I^{i_t}_F$ given to F is the absolute difference of $I^{i_t-1}_{SP}$ and $I^{i_t-2}_{SP}$, which is multiplied to $I^{i_t-1}_R$ to focus on regions where restoration needs to be improved.

architecture, and hence is further divided into an encoder R_e and a decoder R_d, which outputs the restored image I^i_R. The restored image is provided as input to the segmentation module S, whose output is in turn input to the feedback module F. This is summarized in Figure 1. Consider an input turbulent image I^i_T, which is passed through R_e giving the latent representation h^i of the input image. h^i is further decoded by R_d to give the corresponding restored image I^i_R. The restoration process can be formalized as:

$$h^i = R_e(I^i_T) \qquad \text{and} \qquad I^i_R = R_d(h^i) \tag{1}$$

I^i_R is then passed into the semantic segmentation network S to give semantic segmentation output $\hat{I}^i_S$.

3.2 Semantic Feedback Learning

We now explain the semantic feedback learning framework, where the semantic feedback information from S is passed on to the restoration network. The restoration task can be formulated as a recurrent process, where it learns to fix the mistakes of its previous output by leveraging the difference of output probability response map of S. This probability response map acts as a spatial attention mechanism for the restoration network enabling it to focus on regions that need to be restored. Now, we introduce the notion of feedback into our framework. We denote the current training iteration as t. The hidden output of R_e is then given by h^i_t, when I^i_T is given as the input turbulent image at the tth iteration. Similarly, $I^{i_t}_R$ and $\hat{I}^{i_t}_S$ denote the restored image and segmentation output, respectively, at iteration t. Now, to leverage the previous restored image $I^{i_t-1}_R$ and its previous consecutive probability response maps obtained from S: $I^{i_t-1}_{SP}$ and $I^{i_t-2}_{SP}$, we propose a feedback network F to introduce feedback information into the hidden representation h^{i_t}. We now explain how the above inputs are combined to provide a feedback input to the network F. Firstly, we take the absolute difference between $I^{i_t-1}_{SP}$ and $I^{i_t-2}_{SP}$, which gives us a weighted region, where regions with higher weights need to have better restoration. This information is subsequently merged

with the previous restored image via element-wise multiplication. Our formulations at time t can hence be written as:

$$\textit{Feedback Input} : I_F^{i_t} = I_R^{i_{t-1}} \odot (\text{abs}(I_{SP}^{i_{t-1}} - I_{SP}^{i_{t-2}})) \tag{2}$$

$$\textit{Restoration Network} : h^{i_t} = R_e(I_T^i) + \alpha(F(I_F^{i_t})) \tag{3}$$

$$I_R^{i_t} = R_d(h^{i_t}) \tag{4}$$

$$\textit{Segmentation Network} : I_{SP}^{i_t} = S(I_R^{i_t}) \tag{5}$$

$$\hat{I}_S^{i_t} = \arg\max(I_{SP}^{i_t}) \tag{6}$$

where α controls the amount of feedback given to the network. Figure 1 also shows the visual pipeline of our overall framework described above.

3.3 Iterative Focal Loss

Since it was proposed, Focal loss (FL) [18] is widely used as a loss function in object detection. This loss reduces the relative loss for well-classified examples and puts more focus on hard examples which helps improve learning. Formally, focal loss is defined as:

$$FL(p_j) = -\delta(1-p_j)^{\gamma}\log(p_j) \tag{7}$$

where p_j is model's estimated probability of being in class j, γ is the focusing parameter and δ is a scalar for balancing the loss. Now, in Equation 7, we change the γ to an iteration-dependent monotonically increasing function, $\gamma(t)$. This shows that as the iteration t increases, the value of $\gamma(t)$ also increases. This will result in more focus on those examples that are misclassified across consecutive iterations. We include this iterative Focal Loss (iFL) in our feedback framework, and this is formally defined as:

$$iFL(p_j) = -\delta(1-p_j)^{\gamma(t)}\log(p_j) \tag{8}$$

In order to further address class imbalance issues (which are common in semantic segmentation datasets), we use a recent idea proposed by Cui *et al.* [6], which suggests normalizing the loss in a manner based on the sample density of each class. We hence propose a Class-Balanced Iterative Focal Loss (CB-iFL), which is given as below:

$$CB\text{ - }iFL(p_j) = -\frac{1-\beta}{1-\beta^{m_j}}(1-p_j)^{\gamma(t)}log(p_j) \tag{9}$$

where β is a smoothing factor and m_j is the class frequency of class j. The additional multiplicative term in CB-iFL forces the network focus as much on small classes such as 'rider' as on large classes such as 'sky'. This is useful in scenarios especially when an important class (rider in this case) occupies less area in an image than others. Now, in the next subsequent sections, we perform extensive experiments to show the effectiveness of our proposed feedback framework and losses. The total loss of our feedback framework is:

$$L_{total} = L_{restoration} + L_{segmentation} \tag{10}$$

where $L_{restoration}$ is the L1 loss and $L_{segmentation}$ is the CB-iFL loss that is used for training the restoration and segmentation networks, respectively.

4 Experiments

4.1 Experimental Setup

Dataset: We use a physics-based method proposed by Schwartzman *et al*. [26] to generate atmospheric turbulent images. This method efficiently injects atmospheric turbulence into images by a series of 2D image transformations. Using the above method, we synthesized an image dataset consisting of 2975 training image pairs and 500 validation image pairs. Each image pair consists of a turbulent image and a corresponding non-turbulent image from the Cityscapes [4] dataset. Each non-turbulent image has a semantic segmentation label map that divides the image into 19 semantic labels, excluding the void labels.

Network Details: Our restoration framework consists of a UNet [25] which predicts the warping field. The warping field is then bi-linearly applied on the input turbulent image to remove the geometrical distortions. We use ERFNet [24] as our semantic segmentation network because of its small size and efficacy, which makes the entire framework end-to-end trainable. The feedback network consists of 2 average pooling layers, 2 convolutional layers, 2 batch normalization [12] layers and a ReLU layer. The average pooling layers are at the start and end of the network. The middle layers consist of a convolutional layer which is followed by a batch normalization layer and a ReLU layer as an activation layer.

Training Details: The learning rates for R, F and S are $2e-4$, $2e-4$ and $5e-4$ respectively, with Adam [14] as the optimizer. The learning rate of R and F decays by a factor of 0.5 at every 30 epochs and for S, the learning rate decays by a factor of 0.99 at every epoch. For feedback inputs, at iteration $t=1$, $I_R^{i_1}$ and $I_{SP}^{i_1}$ are computed by feeding zero tensors into F. At $t=2$, $I_{SP}^{i_{t-2}} = I_{SP}^{i_0}$, we feed $I_{SP}^{i_{t-2}}$ as $I_{SP}^{i_1}$ into F. The focusing parameter $\gamma(t)$ for iFL was chosen to be: $\gamma(t) = \{0,\ if\ t=1;\ 0.1t,\ if\ t>1\}$. During inference, we use the same feedback network as in training, and the final output is taken in the last iteration.

Baselines: We used UNet and ERFNet for the restoration and segmentation networks respectively, each of which is simpler architectures than those used in Rai *et al*. [23]. These chosen architectures for this work occupy only 16.9% of the parameters in Rai *et al*. [23], thus making it easier for end-to-end training, and improve the performance on their work, despite this reduction in parameters. The work of Rai *et al*. [23] with these architectures was chosen as the baseline for a fair comparison. Now, we use the feedback method proposed by Zamir *et al*. [32] and Shama *et al*. [27] in our framework, to compare the performance with our proposed feedback method. We train all the feedback framework for 3 iterations(t) for fair comparison. We ran three trials of each experiment, and report the mean of the mIoUs across the trials.

4.2 Results

Sanity Check: We perform two experiments to check whether our feedback framework provides useful feedback information into the restoration network. In the first experiment, we pass as input into F an image whose pixel values are randomly sampled from a normal distribution having mean 0 and standard deviation 1. In the second experiment, the input image into F is the image obtained from the multiplication of $I_R^{i_{t-1}}$ and normally distributed image having the same statistics of experiment 1. The results obtained from the first and second experiments have mIoUs of 45.03 and 51.29, respectively - which is far less than our

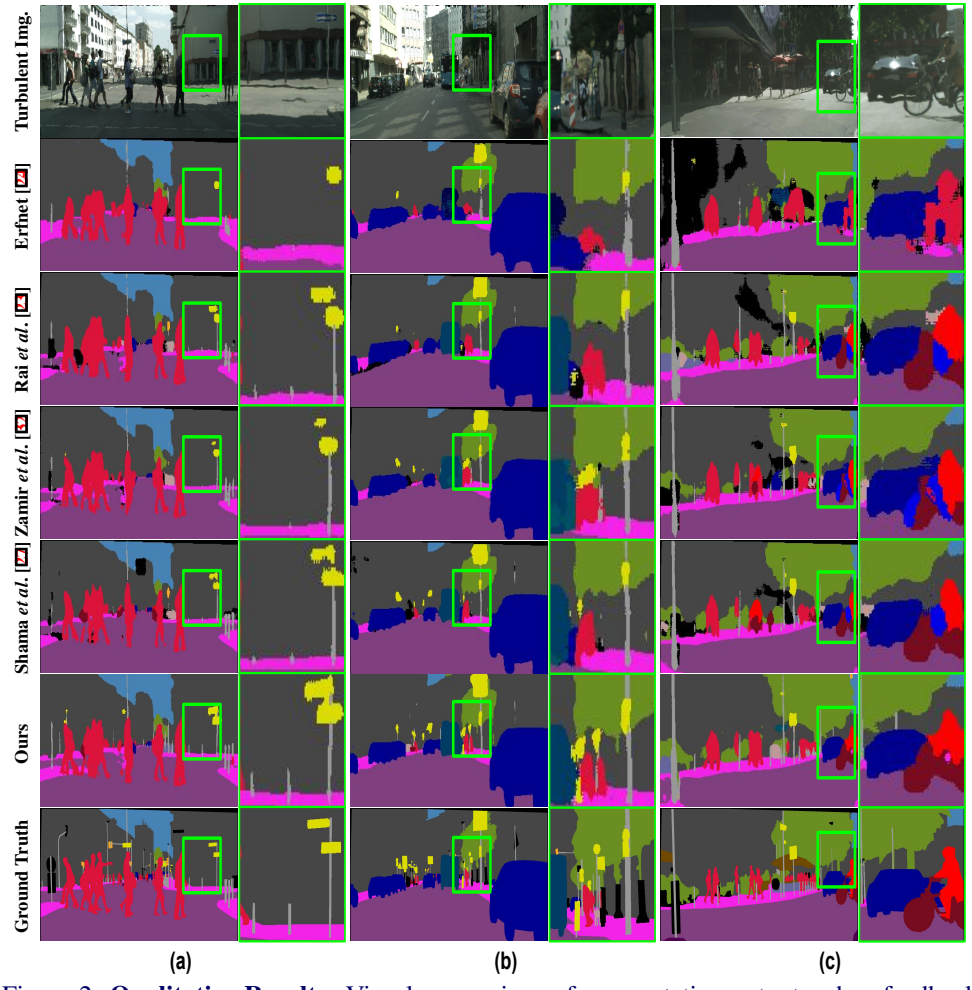

Figure 2: **Qualitative Results:** Visual comparison of segmentation outputs when feedback and non-feedback methods are applied to an input turbulent image. We can observe that our proposed feedback framework correctly segments finer structures such as poles and signs.

method. This helped us infer that our restoration network indeed benefited from the semantic feedback given into the network, which was better than randomly feeding input.

Result Discussion: We compare our proposed feedback method with previously proposed feedback methods Zamir *et al*. [32] and Shama *et al*. [27] and non-feedback methods Rai *et al*. [23], which is also the current state-of-the-art for semantic segmentation in atmospheric turbulence. Table 1 shows the results. Using the feedback model of Zamir *et al*. [32] and Shama *et al*. [27] into our framework did not provide adequate semantic feedback information into the restoration network, which resulted in reduced performance when compared with our method. Our method can correctly segment even small classes such as 'poles', 'traffic signs' shown in Figure 2. This encouraged us to do further analysis of the improvement in those classes that are most important for an autonomous driving system that belongs to Group 4 according to Chen *et al*. [3]. Table 2 shows that our method gives a large improvement of 23.56 mIoU for Group 4 classes over ERFNet itself. Our proposed loss function

Method	road	swalk	build.	wall	fence	pole	tlight	sign	veg.	terrain	sky	person	rider	car	truck	bus	train	mbike	bicycle	**mIoU**
ERFNet [23] (NT.)	97.55	81.52	91.34	54.56	54.21	60.15	63.53	72.68	91.49	63.97	93.24	77.14	56.33	92.98	68.73	77.43	60.10	43.45	68.87	72.067
ERFNet [23]	94.12	65.17	81.84	14.43	20.13	27.01	10.66	31.67	84.28	50.13	87.80	44.44	13.21	82.39	24.85	14.93	12.29	7.42	36.77	42.291
Rai *et al.* [23]	94.81	68.72	85.37	35.49	32.07	35.34	33.17	44.46	86.71	52.37	89.89	57.30	26.98	86.71	43.33	46.70	29.26	13.56	50.59	53.306
Zamir *et al.* [23]	95.43	69.79	85.24	36.01	33.04	34.98	33.47	46.24	86.59	53.56	89.91	57.34	31.30	85.53	52.97	55.01	31.78	17.35	51.10	55.086
Shama *et al.* [23]	95.19	69.43	85.49	37.11	**37.12**	34.33	35.51	45.77	86.68	54.63	90.04	57.45	32.64	86.94	52.36	56.90	39.40	18.80	51.02	56.147
Ours	95.99	71.24	85.69	42.76	35.34	39.54	35.74	47.20	86.95	54.97	**90.22**	57.82	30.34	87.44	53.83	57.72	42.58	23.33	52.74	57.446
Ours (iFL)	95.85	71.24	**85.99**	44.14	32.95	36.34	35.41	48.15	**86.97**	53.57	**90.22**	58.94	**35.28**	87.30	52.55	60.79	**51.26**	19.57	51.69	57.801
Ours (CB-iFL)	**96.27**	**72.59**	85.80	**44.30**	35.03	**40.99**	**38.31**	**50.86**	86.87	**55.97**	90.12	**60.04**	35.14	**88.01**	**59.46**	**67.50**	48.43	**24.16**	**54.62**	**59.709**

Table 1: Classwise semantic segmentation results of various methods on the Cityscapes dataset. Our proposed feedback method outperforms other methods with and without feedback modules. We train all the feedback models for 3 iterations(t). NT. shows the method is trained and validated on non-turbulent dataset whereas all other methods are trained and validated on turbulent dataset. Best results are in bold.

Classes	ERFNet [23]	Ours	IoU Gain
Person	44.44	60.04	*15.60*
Rider	13.21	35.14	*21.93*
Car	82.39	88.01	*05.62*
Truck	24.85	59.46	*34.61*
Bus	14.93	67.50	*52.57*
Motorcycle	07.42	24.16	*16.74*
Bicycle	36.77	54.62	*17.85*
mIoU	32.001	55.561	***23.560***

Table 2: **mIoU gain in important classes:** Segmentation performance improvement in Group 4 [3] classes, which are the most important for the autonomous driving system using our feedback restoration method.

Loss	**mIoU**
Cross-Entropy Loss (CE Loss)	57.45
CE Loss + Li *et al.* [16]	57.53
Ours (iFL)	**57.80**
Weighted CE Loss	59.11
Ours (CB-iFL)	**59.71**

Table 3: **Performance comparison of loss functions for feedback networks:** We train all losses on our feedback framework. Our proposed iFL and CB-iFL outperforms prior losses for the feedback framework.

iFL and CB-iFL further improve the performance of our framework shown in Table 3 and Figure 5. The key advantage we get from using our loss function is that it progressively improves the segmentation results via feedback network by increasing focus on those pixels that are continuously misclassified, unlike giving a constant weight to all the pixels at every iteration as proposed in Li *et al.* [16]. We train our feedback framework over a range of iterations(1-7) and validate the mIoU at each iteration as shown in Figure 3. We find the optimal performance at 3^{rd} iteration, after which performance saturates for higher iterations with slightly lower mIoU. Similarly, we tune α over a range of values, shown in Figure 4. We empirically find the optimal performance at $\alpha = 0.001$. Now, to visually analyze the performance of our feedback framework across the iterations, we show the semantic segmentation results for 3 iterations in Figure 6. The improvement in semantic segmentation across iterations by removing the false segmentation output shows its efficacy.

Ablation Study: To show the effectiveness of our semantic feedback learning, we choose Rai *et al.* [23] as our baseline, and run multiple studies. *Method 1:* Train the baseline for twice the number of epochs. *Method 2:* Doubled the size of the hidden representation of the restoration network in the baseline. *Method 3:* Double the size of the segmentation network in the baseline. *Method 4:* Combine Methods 2 and 3. *Method 5:* Combined Methods 1-3. Our method is adding only semantic feedback with a single iteration into the baseline in these experiments. Table 7 shows the resultant mIoU of all setups, among which our feedback

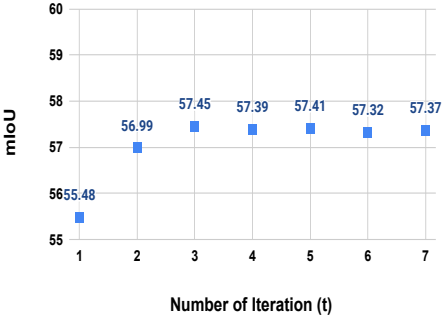

Figure 3: **Analysis on number of iterations:** We train our feedback network trained over a range of iterations on Cityscapes and find its peak performance in the 3*rd* iteration that saturates at higher iterations with slightly lower mIoU.

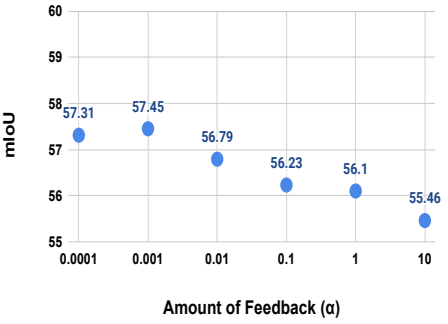

Figure 4: **Amount of feedback vs mIoU:** mIoU on Cityscapes over multiple values of α, a hyper-parameter that controls the fraction of feedback required in the network. Optimal performance was obtained when $\alpha = 0.001$.

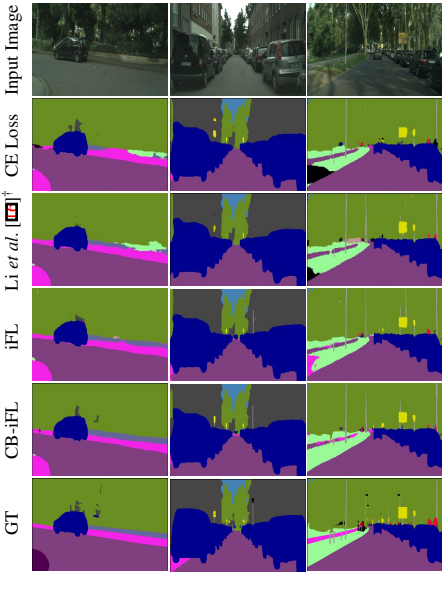

Figure 5: **Feedback losses comparison:** Shows the efficacy of iFl and CB-iFL over previous losses that were used to train the feedback network. († + CE Loss)

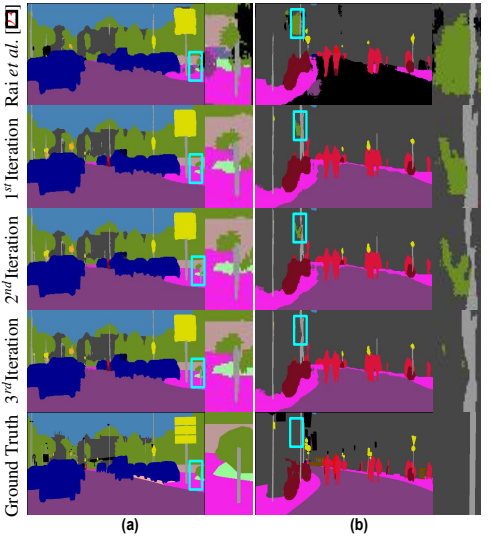

Figure 6: **Contribution of semantic feedback:** We can observe that in example (a), our feedback module progressively improves the segmentation results, whereas, in (b), it tries to remove the false segmentation output.

method performs the best. Now to provide further insight into our feedback module and demonstrate how the feedback response improves semantic segmentation in atmospheric turbulence, we visualize the feedback response in form of a response map and show the improvement in semantic segmentation over the iteration. Figure 8 shows that as the number of iterations increases, some of the yellow regions (representing high error response) changes into a red area that reflects a low error, reflecting improvement in semantic segmentation in

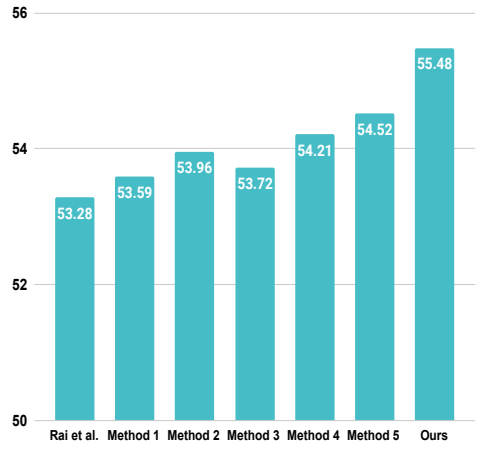

Figure 7: Plot shows the effectiveness of our proposed feedback module, which is trained on a single iteration (t) over various methods that are trained over a higher number of epochs and parameters.

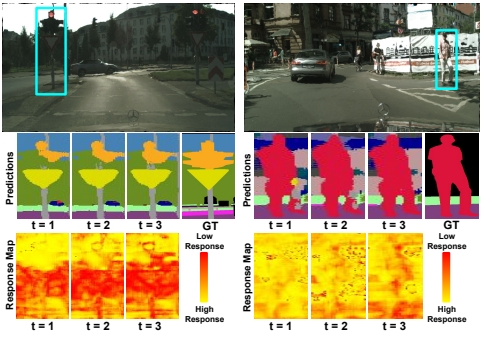

Figure 8: **Feedback response visualization:** It shows image areas where restoration needs to be improved for better segmentation. Yellow color indicates the high response map representing an incorrect image region, and red represents the correct image area with low response. As the number of iterations increases, some of the yellow areas are changed into red areas, reflecting improvement in segmentation results.

those areas. We also perform an analysis between the feedback input and the mIoU of the semantic segmentation network.

5 Conclusion

In this work, we demonstrated the ability of semantic feedback learning to improve semantic segmentation models in hot weather conditions. We propose a feedback framework that consists of a restoration network, a segmentation network, and a feedback network. The feedback network gives a feedback response to the restoration network, which attends those areas and fixes regions that need to be restored. We further boosted the performance of our model by new loss functions: iFL and CB-iFL. We proved the effectiveness of our proposed feedback model and losses through extensive experiments and ablation studies. Our work unlocks doors for further potential application of the feedback mechanism in other weather conditions such as snow, rain, and fog.

Acknowledgement: This work is partly supported by DST, Govt of India, through the IMPRINT program.

References

[1] Vasileios Belagiannis and Andrew Zisserman. Recurrent human pose estimation. In *FG*. IEEE, 2017.

[2] Joao Carreira, Pulkit Agrawal, Katerina Fragkiadaki, and Jitendra Malik. Human pose estimation with iterative error feedback. In *CVPR*, 2016.

[3] Bi-ke Chen, Chen Gong, and Jian Yang. Importance-aware semantic segmentation for autonomous driving system. In *IJCAI*, 2017.

[4] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *CVPR*, 2016.

[5] Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. Class-balanced loss based on effective number of samples. In *CVPR*, 2019.

[6] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*. IEEE, 2009.

[7] Yoav Freund, Robert Schapire, and Naoki Abe. A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, 14(771-780):1612, 1999.

[8] David L Fried. Probability of getting a lucky short-exposure image through turbulence. *JOSA*, 1978.

[9] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *CVPR*, 2017.

[10] Chen Gong, Wang Tang, and Zhou He-qin. A novel physics-based method for restoration of foggy day images. *Journal of Image and Graphics*, 2008.

[11] Shirsendu Sukanta Halder, Jean-François Lalonde, and Raoul de Charette. Physics-based rendering for improving robustness to rain. In *ICCV*, 2019.

[12] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv*, 2015.

[13] Sookyung Kim, Sunghyun Park, Sunghyo Chung, Joonseok Lee, Yunsung Lee, Hyojin Kim, Mr Prabhat, and Jaegul Choo. Learning to focus and track extreme climate events. 2019.

[14] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv*, 2014.

[15] Ruoteng Li, Robby T Tan, Loong-Fah Cheong, Angelica I Aviles-Rivero, Qingnan Fan, and Carola-Bibiane Schonlieb. Rainflow: Optical flow under rain streaks and rain veiling effect. In *ICCV*, 2019.

[16] Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwanggil Jeon, and Wei Wu. Feedback network for image super-resolution. In *CVPR*, 2019.

[17] Zhuwen Li, Ping Tan, Robby T Tan, Danping Zou, Steven Zhiying Zhou, and Loong-Fah Cheong. Simultaneous video defogging and stereo reconstruction. In *CVPR*, 2015.

[18] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *ICCV*, 2017.

[19] Yun-Fu Liu, Da-Wei Jaw, Shih-Chia Huang, and Jenq-Neng Hwang. Desnownet: Context-aware deep network for snow removal. *IEEE TIP*, 2018.

[20] Shree K Nayar and Srinivasa G Narasimhan. Vision in bad weather. In *ICCV*. IEEE, 1999.

[21] Markus Oberweger, Paul Wohlhart, and Vincent Lepetit. Training a feedback loop for hand pose estimation. In *ICCV*, 2015.

[22] Roberto Ragazzoni, Enrico Marchetti, and Gianpaolo Valente. Adaptive-optics corrections available for the whole sky. *Nature*, 2000.

[23] Shyam Nandan Rai, Vineeth N Balasubramanian, Anbumani Subramanian, and CV Jawahar. Munich to dubai: How far is it for semantic segmentation? In *WACV*, 2020.

[24] Eduardo Romera, José M Alvarez, Luis M Bergasa, and Roberto Arroyo. Erfnet: Efficient residual factorized convnet for real-time semantic segmentation. *Transactions on Intelligent Transportation Systems*, 2017.

[25] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015.

[26] Armin Schwartzman, Marina Alterman, Rotem Zamir, and Yoav Y Schechner. Turbulence-induced 2d correlated image distortion. In *ICCP*. IEEE, 2017.

[27] Firas Shama, Roey Mechrez, Alon Shoshan, and Lihi Zelnik-Manor. Adversarial feedback loop. In *ICCV*, 2019.

[28] Girish Varma, Anbumani Subramanian, Anoop Namboodiri, Manmohan Chandraker, and CV Jawahar. IDD: A dataset for exploring problems of autonomous navigation in unconstrained environments. In *WACV*. IEEE, 2019.

[29] Zuxuan Wu, Xin Wang, Joseph E Gonzalez, Tom Goldstein, and Larry S Davis. Ace: Adapting to changing environments for semantic segmentation. In *ICCV*, 2019.

[30] Yuan Xie, Wensheng Zhang, Dacheng Tao, Wenrui Hu, Yanyun Qu, and Hanzi Wang. Distortion-driven turbulence effect removal using variational model. *arXiv*, 2014.

[31] Yitzhak Yitzhaky, Itai Dror, and Norman S Kopeika. Restoration of atmospherically blurred images according to weather-predicted atmospheric modulation transfer function. *OptEn*, 1997.

[32] Amir R Zamir, Te-Lin Wu, Lin Sun, William B Shen, Bertram E Shi, Jitendra Malik, and Silvio Savarese. Feedback networks. In *CVPR*, 2017.

[33] Yupei Zheng, Xin Yu, Miaomiao Liu, and Shunli Zhang. Residual multiscale based single image deraining. In *BMVC*, 2019.

[34] Xiang Zhu and Peyman Milanfar. Removing atmospheric turbulence via space-invariant deconvolution. *IEEE PAMI*, 2012.