Supplementary Material
GIQE: Generic Image Quality Enhancement via Nth Order Iterative Degradation

Pranjay Shyam[1], Kyung-Soo Kim[1*], and Kuk-Jin Yoon[2*]
[1]Mechatronics, Systems, and Control Laboratory, [2]Visual Intelligence Laboratory
Department of Mechanical Engineering, KAIST, Republic of Korea
{pranjayshyam, kyungsookim, kjyoon}@kaist.ac.kr

1. Degradation Samples

We first provide a qualitative overview of different training samples generated using the proposed Nth Order Iterative Degradation in Fig. 2. We additionally provide samples from the GOPRO-II (r,s,f,i,n) and GOPRO-III datasets in Fig. 1.

2. Architectural Overview

We provide an overview of the new transformer mechanism when proposed optimizations and feature enhancement mechanisms are integrated in Fig. 3. Furthermore we also provide the methodology used to stack multiple transformer blocks along with the long skip connection.

3. Ablation Studies

We extend the ablation study from the main paper and examine the effect of longer training time, number of channels, transformer blocks and multi-scale model configuration. First we examine the effect of number of channels in Fig. 4 (left) and number of transformer blocks on GOPRO dataset to determine the ideal combination for accurate deblurring. To examining the effect of number of channels in transformer blocks we vary the range from 24 to 180 with an interval of 12 channels and train the network following the training setting aforementioned. We observe that increasing the number of channels and number of transformer blocks results in increased performance. While the performance gain for 24 block single scale GIQE isnt substantial, we observe it to be higher for 36 block transformer. However as the channel size increases, the performance gains are not substantial. Thus we chose number of channels as 120 with 36 transformer, as it provides the best performance-parameter ratio. Owing to computational limitations, we restrict the experiment to 36 transformer blocks.

Second, we focus on whether increasing the feature quality from multiple scales affect the restoration quality. For this we use multiple GIQE branches with different channel and block sizes. Based on this ablation, we observe 3 stage network to result in peak restoration performance. Based on prior works that highlight increased training epochs to result in better performance, we examine that increasing training epochs by 100 times to improve performance by 0.27db. Hence the final GIQE model that achieves peak performance is obtained when window size is fixed to 8, channel size i.e. C_1, C_2, C_3 to 120, 96, 48 respectively with the number of transformer blocks (N_1, N_2, N_3) for each scale as 36, 24 and 12 respectively.

Table 1. Performance comparison with SoTA image restoration algorithms on different degradation-specific datasets. P, S, N, L refer to PSNR, SSIM, NIQE and LPIPS values, respectively.

Method	NTIRE-19 P (↑) / S (↑)	NTIRE-19 N (↓) / L (↓)	Method	Rain1400 P (↑) / S (↑)	Rain1400 N (↓) / L (↓)
Input	9.11 / 0.49	5.04 / 0.62	Input	23.64 / 0.75	5.17 / 0.28
DuRN-US [13]	13.63 / 0.57	5.61 / 0.61	EfficientDerain [4]	24.23 / 0.78	4.10 / 0.26
GridDehazenet [12]	12.96 / 0.50	5.80 / 0.75	MPRNet [24]	32.01 / **0.92**	3.12 / **0.11**
RefineDNet [27]	**19.54** / 0.54	4.71 / 0.67	DuRN [13]	30.64 / 0.91	**2.99** / 0.12
FFA-Net [16]	14.01 / 0.56	4.84 / 0.67	RESCAN [11]	31.28 / 0.90	3.65 / 0.58
DIDH [18]	19.47 / **0.75**	3.98 / 0.64	PreNet [17]	31.75 / 0.91	3.50 / 0.49
Uformer [20]	19.19 / 0.65	4.44 / 0.53	Uformer [20]	26.92 / 0.74	3.78 / 0.34
GIQE	19.29 / 0.67	4.37 / **0.52**	GIQE	**32.21** / 0.91	3.72 / 0.34
Raindrop			**GOPRO**		
Input	21.41 / 0.75	3.60 / 0.23	Input	25.64 / 0.79	4.69 / 0.31
A.GAN [15]	23.68 / 0.75	3.19 / 0.39	DMPHN [25]	31.20 / **0.94**	3.05 / 0.31
DuRN [13]	23.91 / 0.75	**3.14** / 0.38	DeblurGANv2 [10]	29.55 / 0.93	3.24 / 0.40
MPRNet [24]	24.19 / 0.79	3.34 / **0.29**	MPRNet	31.67 / 0.92	**2.58** / **0.29**
EfficientDerain [4]	23.72 / 0.75	3.08 / 0.42	DuRN	29.27 / 0.89	3.09 / 0.33
RESCAN	21.69 / 0.74	4.91 / 0.47	DPIR [26]	27.02 / 0.81	3.57 / 0.39
Uformer [20]	23.51 / 0.64	3.37 / 0.42	Uformer [20]	32.27 / 0.90	3.47 / 0.42
GIQE	**25.18** / **0.82**	3.19 / 0.34	GIQE	**33.48** / **0.94**	3.45 / 0.43

4. Quantitative Results

We now retrain the aforementioned network configuration with proposed degradation algorithm and enable multiple degradations and subsequently compare the performance of proposed image restoration algorithm with SoTA on Deblurring, Raindrop removal, deraining and dehazing. We thus utilize GOPRO, Raindrop, Rain1400 and NTIRE-19 datasets with results summarized in Fig. 1. From the quantitative results we can observe the proposed GIQE to outperform SoTA for deraining, raindrop removal and de-

*Co-corresponding authors. Listed in alphabetical order.

Figure 1. GOPRO-II (r,s,f,i,n) and GOPRO-III datasets.

blurring while achieving competitive performance for dehazing. We present qualitative results for dehazing in Fig. 5, deraining in Fig. 6, rain drop removal in Fig. 7 and deblurring in Fig. 8.

We now extend our evaluation to determine if current SoTA can be used to handle multiple degradations. For this we retrain them using the proposed degradation algorithm while following the training methodology mentioned in their respective works. We summarize the quantitative results in Tab. 3 and observe performance to be consistent across multiple degradation combinations. However all models undergo performance degradation when synthetic fog is present in the image, along with blurring. Despite this we observe the proposed GIQE to have consistent performance across all degradation types.

Table 2. Performance comparison with SoTA image deblurring algorithms.

Algorithm	GoPRO-II (r)	GoPRO-II (s)	GoPRO-II (f)	GoPRO-II (i)	GoPRO-II (n)
	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM
Input	25.64 / 0.79	26.55 / 0.80	15.24 / 0.53	11.33 / 0.49	16.71 / 0.52
DMPHN [25]	27.98 / 0.84	26.04 / 0.79	19.29 / 0.63	17.76 / 0.51	24.89 / 0.75
DeblurGANv2 [10]	28.92 / 0.89	26.68 / 0.81	16.87 / 0.57	17.18 / 0.63	23.46 / 0.77
MPRNet [24]	31.84 / 0.92	26.68 / 0.82	20.18 / 0.69	23.57 / 0.72	25.48 / 0.79
DuRN [13]	28.00 / 0.85	26.37 / 0.81	18.27 / 0.64	21.24 / 0.62	21.68 / 0.64
Uformer [20]	32.27 / 0.90	30.74 / 0.88	24.58 / 0.66	25.68 / 0.74	26.01 / 0.75
GIQE	33.48 / 0.94	30.18 / 0.84	26.19 / 0.71	27.98 / 0.77	28.12 / 0.81

5. Performance of SoTA Semantic Segmentation Algorithms in Diverse Conditions

We now extend the examination of the impact of different weather degradations on SoTA semantic segmentation algorithms available in mmsegmentation [2] library such as ANN [28], APCNet [5], CCNet [7], CGNet [21], Deeplabv3+ [1], DNL [22], PointRend [9], Non Local Net [19], PFPN [8], and OCRNet [23] to examine if observations drawn in Tab. 4 of the main paper translate across SoTA architectures while providing a qualitative summary of their performance in Fig. 9.

For consistency, we used pretrained models with ResNet-50 [6] backbone or Hourglass-18 [6] (for OCRNet) and evaluated these algorithms on original cityscapes representing the baseline performance (Fig. 9, Tab. 3(I)). Subsequently, we extend the variations observe that while semantic segmentation algorithms are robust to an extent (avg performance drop of -0.06 mIOU) in foggy cityscapes (Fig. 10, Tab. 3(II)), the performance deteriorates as degradation intensity increases to rainy (Fig. 11, Tab. 3(III))(avg performance drop of -0.23 mIOU) as well as raindrops (Fig. 12, Tab. 3(IV))(avg performance drop of -0.06 mIOU) wherein the algorithms classify attributes within the water droplets as well. This results in a substantial decrease in performance when all of these variations occur together in mixed cityscapes (Fig. 13, Tab. 3(V))(avg performance drop of -0.19mIOU). Following [14] we examine if switching to deeper backbones (ResNet-101, Hourglass-48) aids in per-

Figure 2. Images generated by proposed N^{th} order degradation along with corresponding spatial distortion masks for a given clean input from GOPRO dataset.

formance under adverse weather conditions and conclude it to not improve performance (avg performance increase of +0.046 mIOU on mixed cityscapes).

Hence we retrain SoTA algorithms with R-101 (or H-W48) backbone SoTA from scratch following their original training methodology. We observe that despite tuning the hyperparameters, we were only able to achieve around 50%-60% of performance on mixed cityscapes (Tab. 4(I),

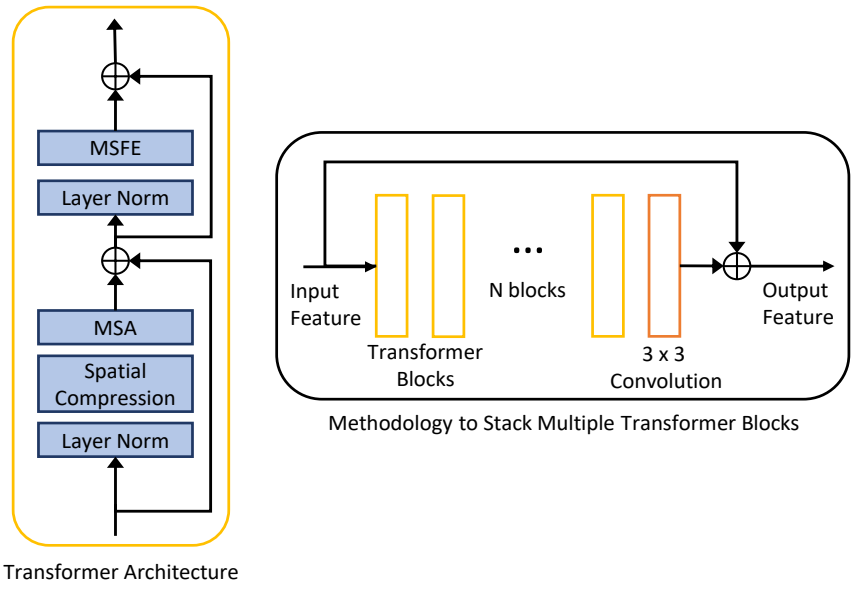

Figure 3. Overview of (left) proposed transformer module and (right) connection of multiple transformer modules with long skip connection.

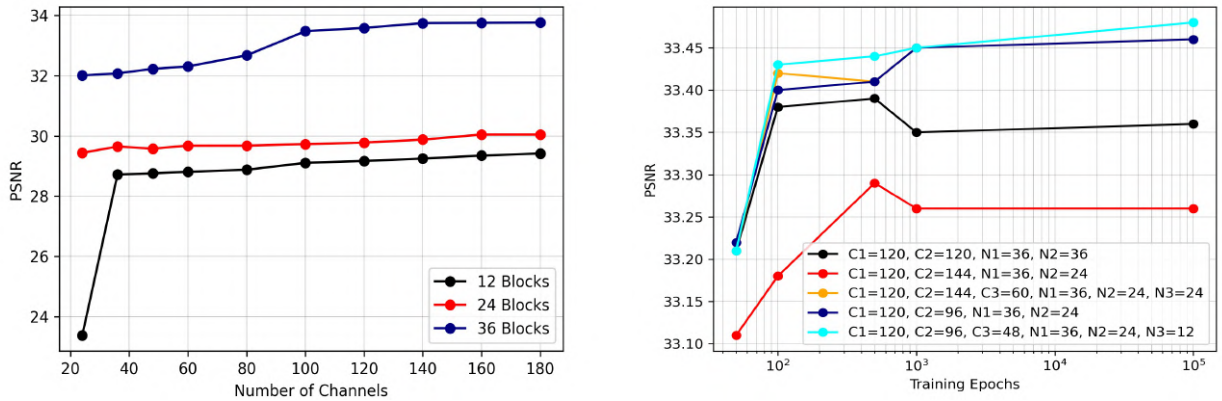

Figure 4. Ablation of Varying Channel Numbers, Number of Transformer Blocks, Extended Training Epochs and Multi-Scale Restoration algorithm.

(III)), hence we conclude that while retraining the algorithms on degraded dataset might ensure robustness for object detections, but such an approach is not possible for semantic segmentation. We then examine the effect of restoring degraded images using the proposed framework on pretrained semantic segmentation algorithms (Fig. 14, Tab. 4(II)) and observe that while all algorithms seem to benefit from restored images, the quantum varies wildly, ranging primarily due to over-saturation of images by the algorithm. Hence we have algorithms that benefit the most due to their robustness to color saturation, such as Deeplabv3+ and DNL, while on the other hand, algorithms such as APCNet and CCNet are extremely sensitive and fail in certain scenarios. Hence one can improve the performance of the combined framework of image restoration and semantic segmentation by ensuring the semantic segmentation network is robust to color variations while ensuring an image restoration algorithm is able to recover maximum areas.

6. Limitations

From the extensive studies, we can clearly identify some of the limitations of the proposed work and discuss certain directions that could improve the performance of blind im-

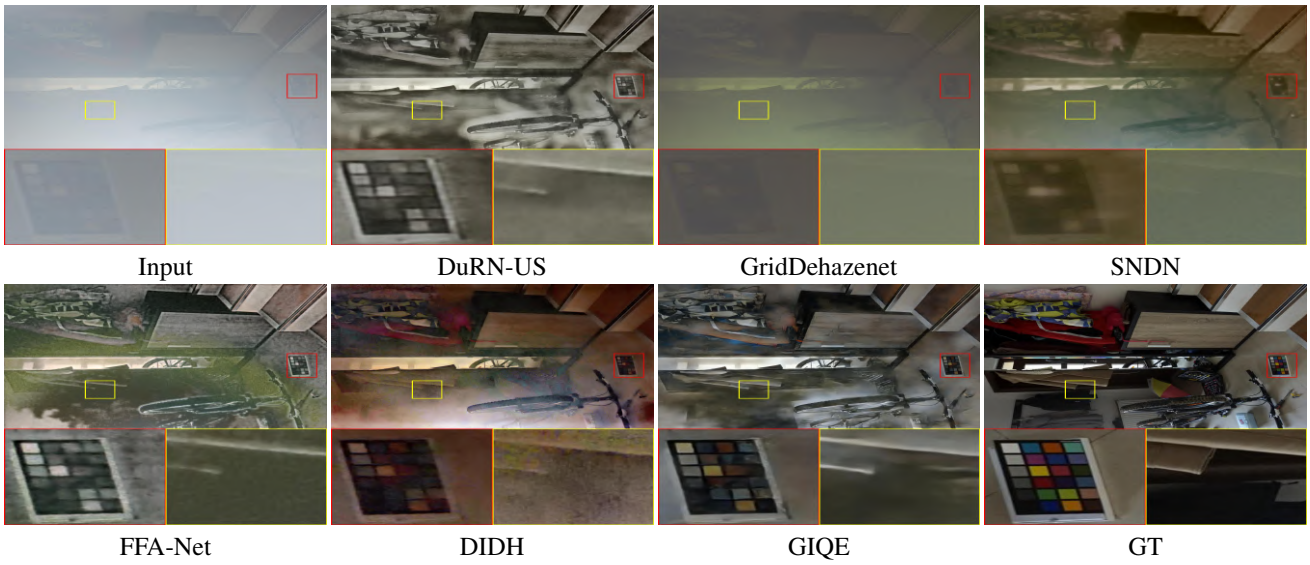

Figure 5. Qualitative comparison of SoTA dehazing models on hazy images.

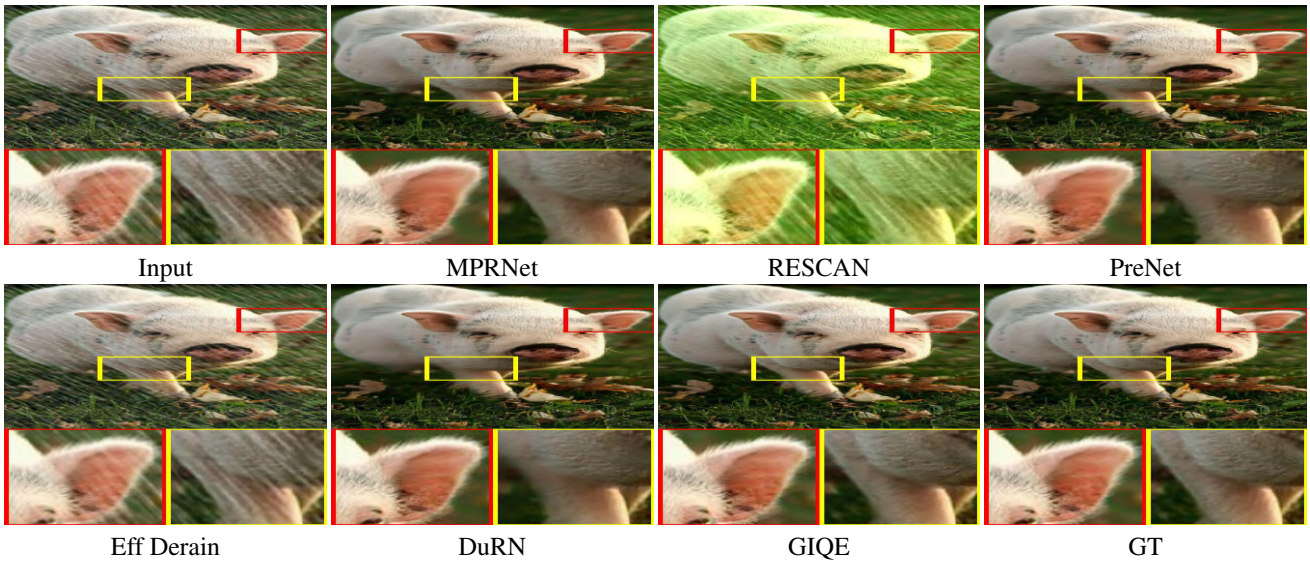

Figure 6. Qualitative comparison of SoTA deraining models on synthetic rainy images.

age restoration algorithm, (1) Over-saturation of Images - We observe the recovered images to be over-saturated and thus believe that longer training time within an adversarial framework should ensure improved color saturation and perceptual quality. (2) We observe weak restoration performance in recovering haze-affected regions at a distance. This can be improved by making architectural modifications within the underlying architecture so as to increase the receptive field while improving feature interpolation. While we could extend this work and explore these scenarios, we would tackle these issues in potential future works to limit the scope of this paper.

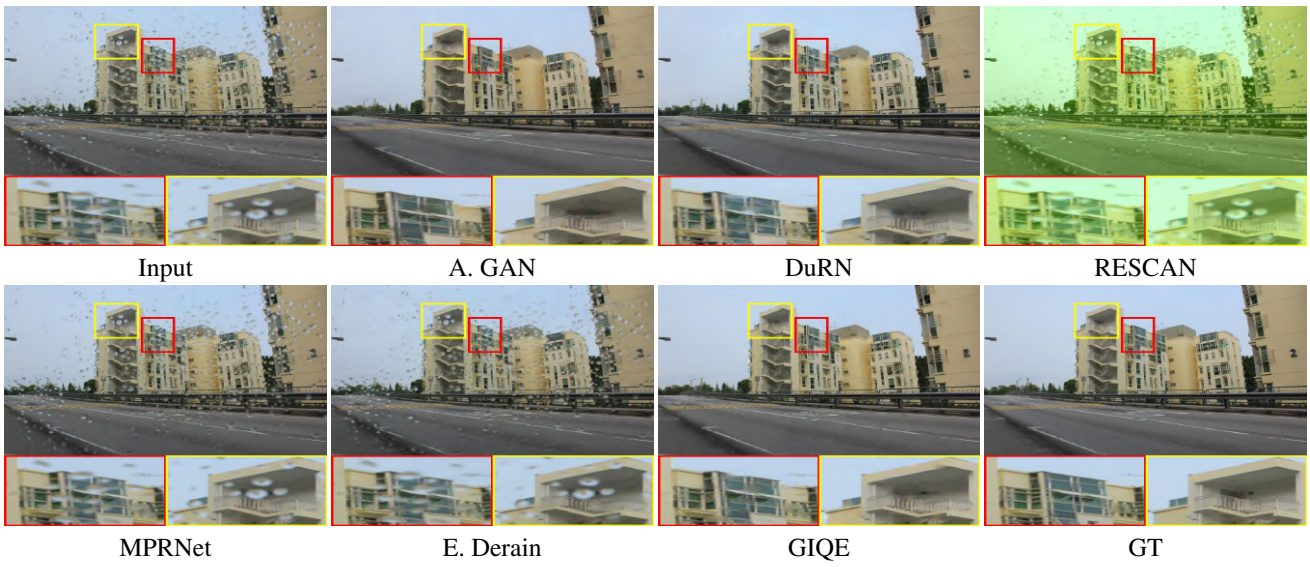

Figure 7. Qualitative comparison of SoTA raindrop removal models on rainy images.

Figure 8. Qualitative comparison of SoTA deblurring models on images.

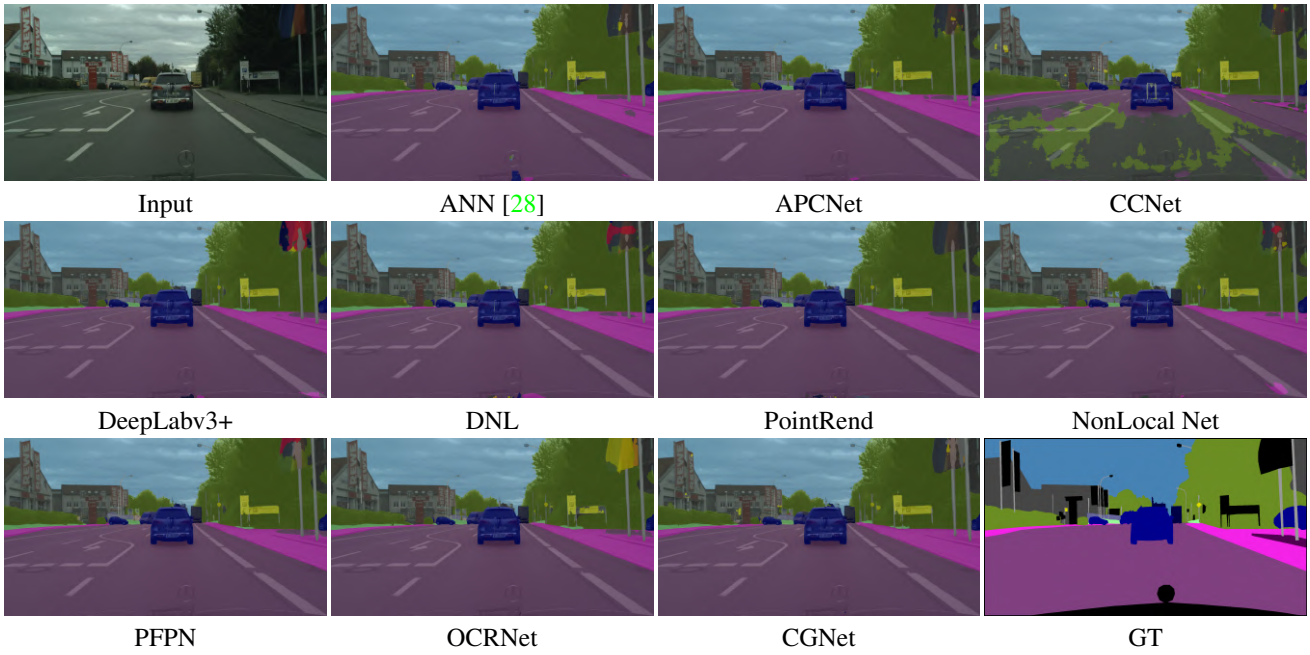

Figure 9. Baseline Qualitative performance of SoTA models on Cityscapes valset.

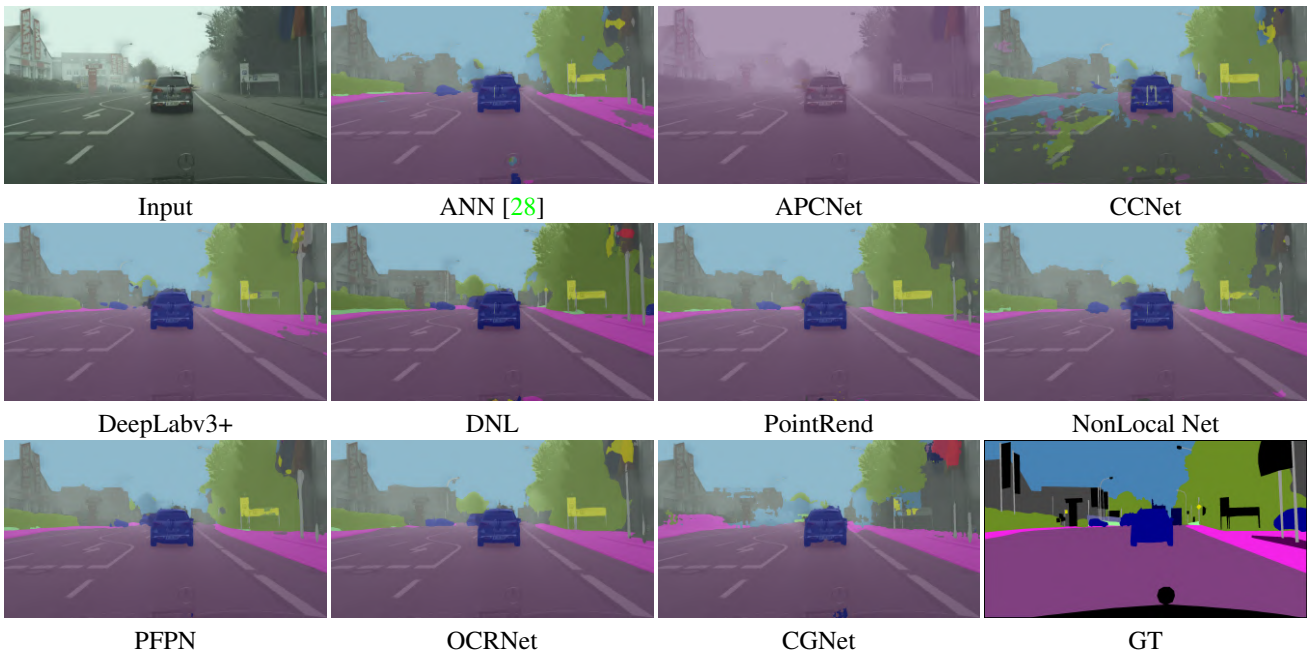

Figure 10. Qualitative performance of SoTA models on Foggy Cityscapes valset.

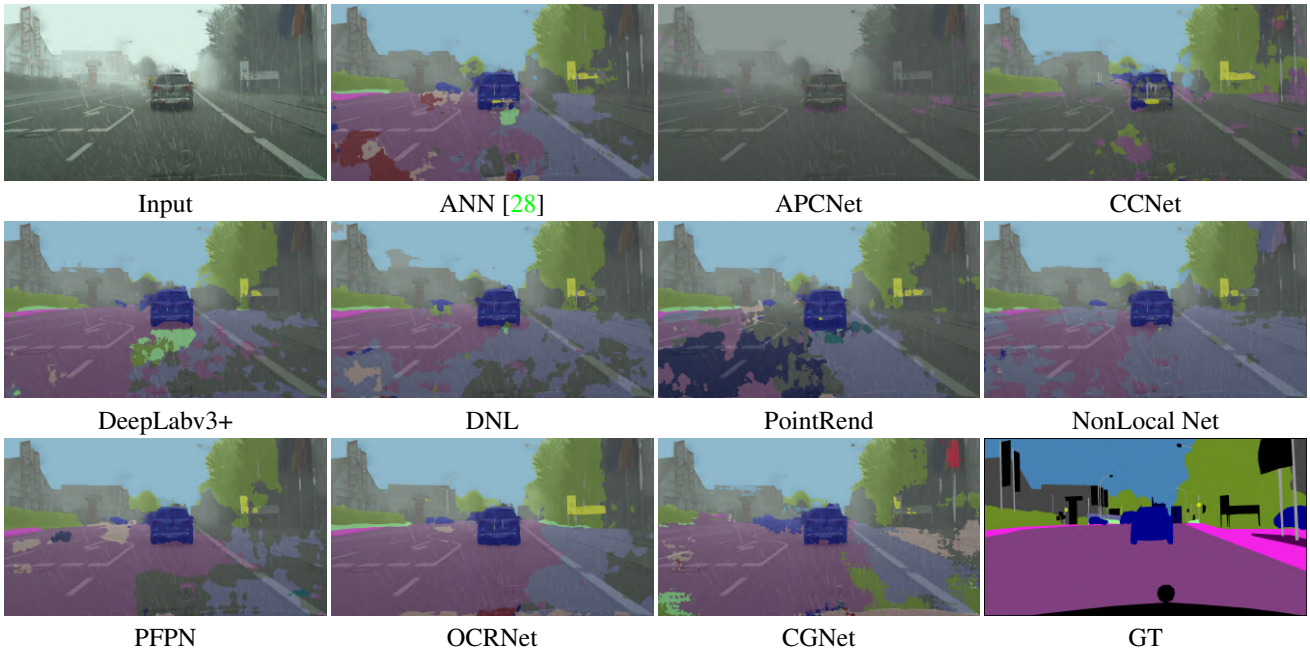

Figure 11. Qualitative performance of SoTA models on Rainy Cityscapes valset.

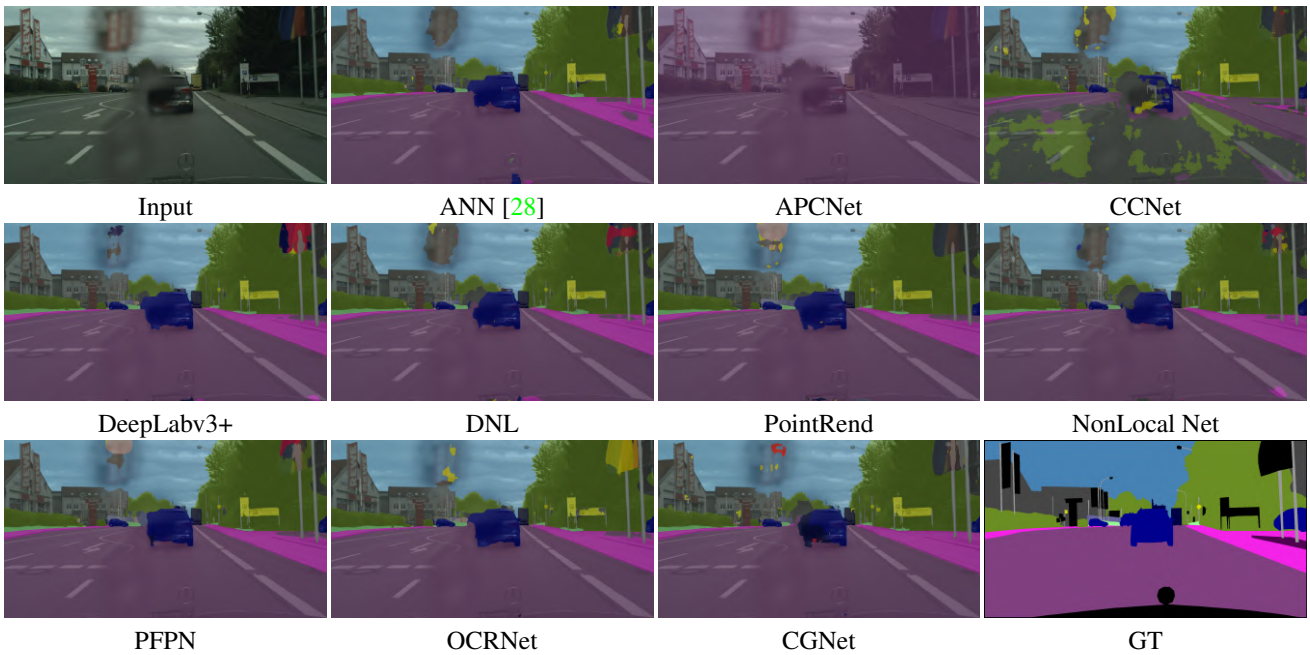

Figure 12. Qualitative performance of SoTA models on Raindrop Cityscapes valset.

Table 3. Semantic segmentation results for Cityscapes Dataset under Diverse Conditions

Method	BB	road	s.walk	build.	wall	fence	pole	t.light	t.sign	veget.	terrain	sky	person	rider	car	truck	bus	train	moto.	bike	mIoU
(I) CityScapes [3]																					
ANN [28]	R-50	0.980	0.841	0.921	0.445	0.562	0.646	0.720	0.796	0.922	0.623	0.942	0.821	0.616	0.950	0.789	0.865	0.789	0.667	0.782	0.773
APCNet [5]	R-50	0.981	0.851	0.925	0.495	0.600	0.658	0.732	0.806	0.926	0.650	0.947	0.830	0.618	0.954	0.835	0.889	0.842	0.652	0.787	0.788
CCNet [7]	R-50	0.981	0.853	0.924	0.468	0.614	0.647	0.721	0.801	0.926	0.663	0.946	0.827	0.629	0.954	0.814	0.860	0.653	0.687	0.783	0.776
DeepLabV3+ [1]	R-50	0.984	0.866	0.930	0.512	0.625	0.695	0.743	0.822	0.927	0.635	0.949	0.842	0.658	0.956	0.811	0.888	0.847	0.686	0.795	0.798
DNL [22]	R-50	0.982	0.858	0.927	0.542	0.610	0.658	0.730	0.806	0.926	0.644	0.946	0.830	0.629	0.954	0.831	0.894	0.829	0.660	0.788	0.792
PointRend [9]	R-50	0.979	0.842	0.923	0.466	0.592	0.668	0.725	0.796	0.924	0.629	0.946	0.823	0.615	0.949	0.673	0.844	0.695	0.634	0.782	0.763
NonLocal Net [19]	R-50	0.980	0.846	0.925	0.553	0.588	0.644	0.720	0.797	0.923	0.637	0.943	0.822	0.606	0.950	0.792	0.861	0.813	0.669	0.779	0.781
PFPN [8]	R-50	0.980	0.844	0.922	0.394	0.567	0.656	0.707	0.794	0.920	0.618	0.944	0.818	0.608	0.943	0.586	0.721	0.729	0.607	0.776	0.744
OCRNet [23]	H-W18	0.983	0.862	0.929	0.519	0.653	0.691	0.728	0.798	0.928	0.654	0.949	0.829	0.631	0.953	0.822	0.903	0.810	0.645	0.775	0.793
CGNet [21]	R-50	0.970	0.791	0.899	0.459	0.526	0.569	0.589	0.688	0.907	0.588	0.919	0.734	0.520	0.920	0.516	0.678	0.477	0.500	0.710	0.682
(II) Foggy CityScapes																					
ANN [28]	R-50	0.974	0.817	0.817	0.354	0.492	0.589	0.661	0.756	0.787	0.583	0.700	0.795	0.600	0.932	0.719	0.802	0.680	0.633	0.762	0.708
APCNet	R-50	0.976	0.827	0.829	0.421	0.544	0.606	0.670	0.768	0.815	0.602	0.753	0.800	0.600	0.938	0.753	0.826	0.738	0.625	0.770	0.730
CCNet	R-50	0.977	0.830	0.828	0.395	0.567	0.592	0.659	0.760	0.811	0.618	0.771	0.796	0.612	0.933	0.700	0.772	0.505	0.654	0.767	0.713
DeepLabV3+	R-50	0.981	0.847	0.853	0.461	0.575	0.645	0.686	0.784	0.834	0.577	0.793	0.812	0.642	0.937	0.719	0.839	0.768	0.656	0.777	0.747
DNL	R-50	0.978	0.837	0.847	0.473	0.558	0.603	0.670	0.767	0.822	0.606	0.751	0.802	0.609	0.935	0.756	0.833	0.743	0.622	0.772	0.736
PointRend	R-50	0.975	0.822	0.847	0.406	0.545	0.612	0.667	0.760	0.831	0.588	0.797	0.799	0.595	0.935	0.618	0.808	0.655	0.603	0.765	0.717
NonLocal Net	R-50	0.976	0.821	0.822	0.484	0.521	0.582	0.654	0.752	0.791	0.588	0.755	0.793	0.592	0.921	0.719	0.777	0.686	0.638	0.759	0.717
PFPN	R-50	0.975	0.823	0.843	0.311	0.516	0.595	0.647	0.753	0.819	0.573	0.760	0.790	0.592	0.928	0.535	0.656	0.651	0.573	0.757	0.689
OCRNet	H-W18	0.978	0.842	0.845	0.459	0.608	0.634	0.663	0.755	0.839	0.612	0.775	0.799	0.609	0.938	0.796	0.844	0.747	0.612	0.752	0.742
CGNet	R-50	0.943	0.729	0.765	0.310	0.469	0.491	0.453	0.615	0.671	0.512	0.569	0.667	0.467	0.845	0.329	0.545	0.420	0.379	0.658	0.570
(III) Rainy CityScapes																					
ANN [28]	R-50	0.523	0.492	0.593	0.023	0.277	0.484	0.678	0.789	0.792	0.553	0.824	0.696	0.620	0.845	0.352	0.799	0.000	0.421	0.700	0.551
APCNet	R-50	0.387	0.404	0.546	0.015	0.468	0.497	0.664	0.791	0.766	0.504	0.820	0.666	0.631	0.759	0.431	0.153	0.000	0.467	0.771	0.513
CCNet	R-50	0.447	0.431	0.544	0.019	0.407	0.468	0.651	0.775	0.771	0.591	0.879	0.699	0.625	0.838	0.350	0.769	0.000	0.395	0.787	0.550
DeepLabV3+	R-50	0.568	0.472	0.558	0.033	0.525	0.549	0.692	0.819	0.822	0.574	0.893	0.704	0.666	0.835	0.165	0.849	0.000	0.421	0.788	0.575
DNL	R-50	0.358	0.411	0.395	0.032	0.444	0.485	0.671	0.790	0.754	0.504	0.822	0.669	0.621	0.783	0.344	0.754	0.000	0.515	0.776	0.533
PointRend	R-50	0.469	0.492	0.642	0.021	0.533	0.496	0.668	0.809	0.781	0.587	0.890	0.711	0.642	0.862	0.022	0.449	0.000	0.396	0.776	0.539
NonLocal Net	R-50	0.501	0.450	0.691	0.018	0.375	0.473	0.642	0.776	0.791	0.510	0.849	0.674	0.618	0.820	0.358	0.397	0.000	0.396	0.769	0.532
PFPN	R-50	0.697	0.620	0.663	0.039	0.378	0.500	0.667	0.803	0.778	0.578	0.878	0.708	0.634	0.881	0.086	0.839	0.000	0.420	0.791	0.577
OCRNet	H-W18	0.692	0.699	0.663	0.088	0.080	0.496	0.650	0.789	0.858	0.607	0.919	0.754	0.647	0.911	0.268	0.728	0.000	0.094	0.783	0.565
CGNet	R-50	0.426	0.397	0.647	0.025	0.105	0.360	0.397	0.643	0.644	0.351	0.776	0.538	0.493	0.551	0.022	0.482	0.000	0.001	0.592	0.392
(IV) Rain Drop CityScapes																					
ANN [28]	R-50	0.967	0.792	0.867	0.370	0.512	0.590	0.606	0.711	0.881	0.600	0.881	0.757	0.559	0.914	0.702	0.795	0.726	0.605	0.737	0.714
APCNet	R-50	0.968	0.807	0.864	0.411	0.559	0.595	0.639	0.701	0.880	0.613	0.867	0.759	0.557	0.920	0.706	0.795	0.769	0.585	0.736	0.723
CCNet	R-50	0.970	0.803	0.868	0.365	0.565	0.582	0.612	0.646	0.875	0.612	0.866	0.751	0.567	0.922	0.681	0.745	0.498	0.605	0.735	0.698
DeepLabV3+	R-50	0.977	0.830	0.885	0.439	0.596	0.642	0.625	0.725	0.887	0.614	0.881	0.752	0.585	0.926	0.713	0.848	0.798	0.613	0.749	0.741
DNL	R-50	0.972	0.815	0.876	0.466	0.562	0.593	0.635	0.712	0.884	0.621	0.878	0.761	0.562	0.923	0.736	0.826	0.787	0.580	0.739	0.733
PointRend	R-50	0.965	0.799	0.879	0.396	0.547	0.611	0.641	0.713	0.887	0.595	0.884	0.764	0.560	0.913	0.565	0.803	0.602	0.587	0.739	0.708
NonLocal Net	R-50	0.966	0.802	0.871	0.480	0.552	0.586	0.617	0.703	0.885	0.606	0.866	0.756	0.546	0.911	0.665	0.744	0.757	0.608	0.736	0.719
PFPN	R-50	0.969	0.803	0.874	0.328	0.527	0.597	0.622	0.695	0.883	0.593	0.873	0.751	0.550	0.912	0.490	0.640	0.681	0.536	0.730	0.687
OCRNet	H-W18	0.975	0.828	0.895	0.470	0.602	0.635	0.640	0.720	0.895	0.641	0.907	0.763	0.568	0.925	0.748	0.853	0.734	0.570	0.733	0.742
CGNet	R-50	0.959	0.745	0.844	0.352	0.491	0.518	0.482	0.614	0.848	0.550	0.845	0.560	0.416	0.878	0.332	0.585	0.386	0.403	0.646	0.603
(V) Mixed CityScapes																					
ANN [28]	R-50	0.775	0.653	0.640	0.072	0.394	0.512	0.585	0.695	0.754	0.513	0.686	0.711	0.537	0.849	0.574	0.722	0.403	0.503	0.677	0.592
APCNet	R-50	0.748	0.639	0.652	0.052	0.491	0.521	0.593	0.683	0.756	0.504	0.639	0.708	0.537	0.818	0.626	0.232	0.648	0.505	0.709	0.582
CCNet	R-50	0.759	0.644	0.629	0.061	0.479	0.498	0.572	0.659	0.752	0.539	0.711	0.706	0.549	0.849	0.561	0.688	0.378	0.512	0.717	0.593
DeepLabV3+	R-50	0.813	0.689	0.679	0.084	0.529	0.569	0.596	0.712	0.793	0.541	0.760	0.710	0.568	0.856	0.381	0.771	0.705	0.543	0.725	0.633
DNL	R-50	0.736	0.650	0.572	0.132	0.485	0.514	0.591	0.692	0.753	0.518	0.659	0.709	0.537	0.838	0.573	0.718	0.673	0.529	0.712	0.610
PointRend	R-50	0.771	0.676	0.717	0.062	0.491	0.530	0.600	0.702	0.786	0.545	0.744	0.723	0.548	0.858	0.121	0.579	0.219	0.466	0.724	0.572
NonLocal Net	R-50	0.774	0.655	0.726	0.046	0.462	0.500	0.581	0.684	0.757	0.509	0.707	0.706	0.531	0.819	0.523	0.465	0.590	0.504	0.709	0.592
PFPN	R-50	0.850	0.706	0.719	0.093	0.438	0.518	0.581	0.687	0.774	0.534	0.714	0.709	0.541	0.871	0.256	0.640	0.384	0.462	0.717	0.589
OCRNet	H-W18	0.852	0.745	0.739	0.157	0.271	0.551	0.586	0.696	0.822	0.584	0.780	0.722	0.550	0.891	0.491	0.680	0.661	0.444	0.714	0.628
CGNet	R-50	0.732	0.586	0.689	0.073	0.276	0.423	0.388	0.564	0.651	0.387	0.624	0.303	0.325	0.680	0.084	0.464	0.267	0.212	0.562	0.436
(VI) Heavy Backbone - Mixed CityScapes																					
ANN [28]	R-101	0.806	0.690	0.611	0.167	0.488	0.541	0.601	0.707	0.763	0.496	0.642	0.713	0.558	0.859	0.311	0.748	0.422	0.490	0.717	0.596
APCNet	R-101	0.825	0.691	0.724	0.070	0.494	0.557	0.600	0.701	0.804	0.572	0.731	0.737	0.588	0.855	0.676	0.816	0.600	0.532	0.722	0.647
CCNet	R-101	0.783	0.696	0.613	0.221	0.520	0.550	0.602	0.706	0.796	0.560	0.739	0.705	0.573	0.859	0.536	0.791	0.676	0.557	0.727	0.643
DeepLabV3+	R-101	0.818	0.725	0.712	0.072	0.559	0.581	0.605	0.718	0.799	0.556	0.676	0.737	0.579	0.884	0.612	0.825	0.701	0.591	0.745	0.658
DNL	R-101	0.799	0.669	0.640	0.158	0.521	0.548	0.617	0.711	0.799	0.565	0.739	0.735	0.580	0.874	0.700	0.823	0.698	0.534	0.724	0.654
PointRend	R-101	0.777	0.682	0.683	0.085	0.513	0.541	0.599	0.708	0.801	0.542	0.756	0.719	0.556	0.876	0.117	0.744	0.484	0.466	0.722	0.599
NonLocal Net	R-101	0.791	0.680	0.650	0.080	0.513	0.542	0.575	0.698	0.788	0.544	0.707	0.712	0.571	0.845	0.590	0.776	0.650	0.509	0.713	0.628
PFPN	R-101	0.845	0.699	0.667	0.197	0.452	0.539	0.589	0.702	0.799	0.550	0.733	0.726	0.567	0.877	0.281	0.723	0.509	0.500	0.729	0.615
OCRNet	H-W48	0.871	0.771	0.789	0.142	0.271	0.593	0.624	0.724	0.831	0.624	0.797	0.738	0.588	0.895	0.677	0.754	0.766	0.502	0.741	0.668

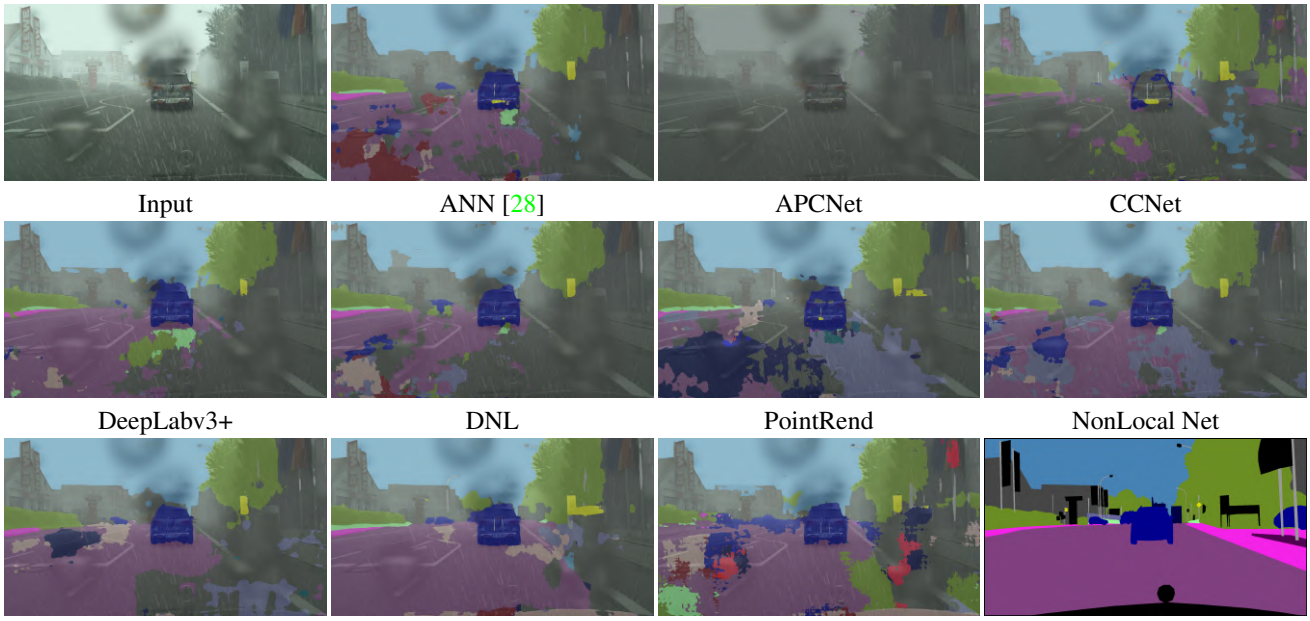

Figure 13. Qualitative performance of SoTA models on Mixed Cityscapes valset.

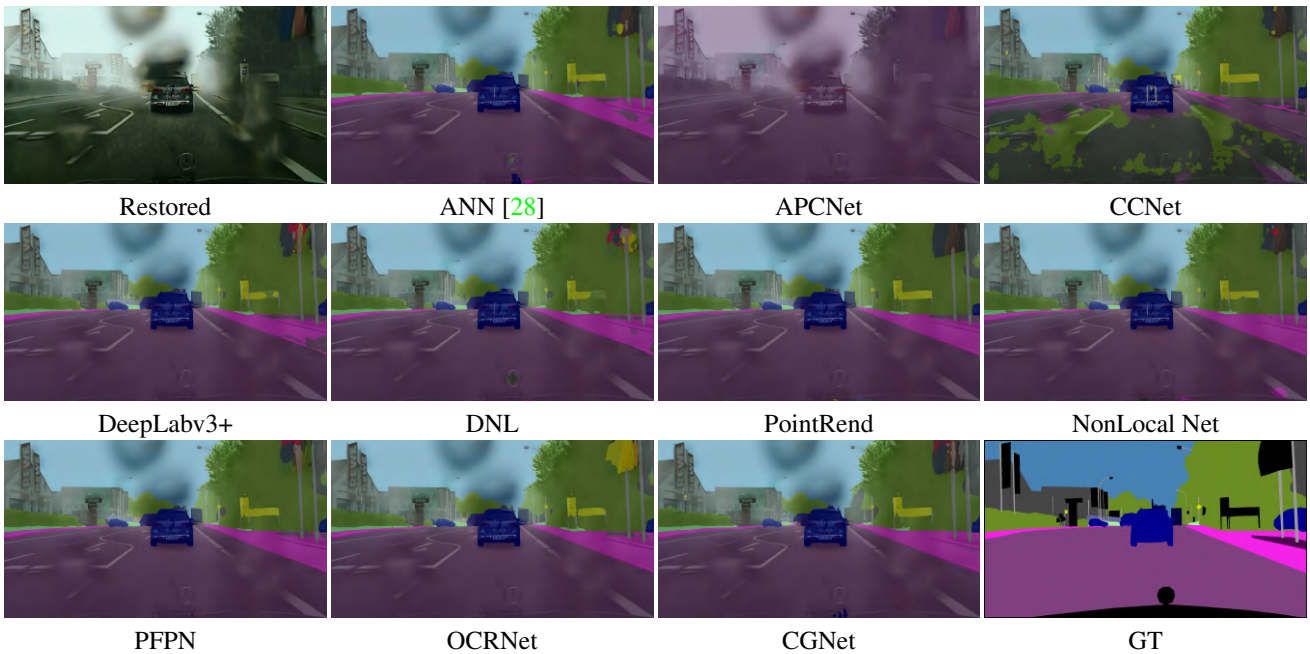

Figure 14. Qualitative performance of SoTA models on Restored Mixed Cityscapes valset.

Table 4. Semantic segmentation results for Cityscapes Dataset under Diverse Conditions

Method	BB	road	s.walk	build.	wall	fence	pole	t.light	t.sign	veget.	terrain	sky	person	rider	car	truck	bus	train	moto.	bike	mIoU
(I) Restored and Enhanced Images																					
ANN [28]	R-50	0.857	0.789	0.643	0.211	0.487	0.655	0.73	0.715	0.819	0.597	0.813	0.757	0.587	0.995	0.69	0.841	0.469	0.622	0.761	0.688
APCNet	R-50	0.769	0.731	0.664	0.074	0.618	0.611	0.703	0.698	0.894	0.621	0.651	0.73	0.601	0.83	0.652	0.233	0.742	0.545	0.729	0.618
CCNet	R-50	0.778	0.772	0.662	0.069	0.615	0.544	0.611	0.693	0.803	0.546	0.727	0.761	0.665	0.865	0.681	0.803	0.485	0.600	0.727	0.711
DeepLabV3+	R-50	0.991	0.852	0.854	0.471	0.579	0.655	0.696	0.788	0.837	0.580	0.800	0.817	0.648	0.943	0.722	0.840	0.773	0.657	0.786	0.749
DNL	R-50	0.980	0.839	0.852	0.478	0.567	0.608	0.672	0.771	0.822	0.616	0.759	0.809	0.611	0.936	0.763	0.839	0.748	0.629	0.780	0.739
PointRend	R-50	0.790	0.777	0.811	0.105	0.508	0.584	0.611	0.719	0.924	0.664	0.759	0.760	0.577	0.949	0.226	0.661	0.278	0.505	0.807	0.594
NonLocal Net	R-50	0.843	0.715	0.731	0.05	0.507	0.544	0.584	0.784	0.771	0.594	0.791	0.832	0.587	0.872	0.623	0.467	0.631	0.615	0.763	0.644
PFPN	R-50	0.888	0.785	0.837	0.110	0.478	0.586	0.595	0.720	0.922	0.588	0.832	0.772	0.621	1.012	0.316	0.695	0.458	0.470	0.730	0.619
OCRNet	H-W18	0.931	0.79	0.856	0.204	0.381	0.676	0.691	0.717	0.952	0.604	0.841	0.853	0.587	0.940	0.612	0.730	0.720	0.445	0.856	0.708
CGNet	R-50	0.789	0.621	0.818	0.148	0.307	0.449	0.394	0.656	0.733	0.392	0.746	0.346	0.413	0.706	0.106	0.504	0.362	0.359	0.623	0.461
(II) Retrained Deeper Model																					
ANN [28]	R-101	0.472	0.323	0.509	0.096	0.053	0.072	0.174	0.444	0.410	0.369	0.188	0.248	0.252	0.392	0.129	0.223	0.169	0.310	0.426	0.244
APCNet	R-101	0.482	0.428	0.521	0.257	0.048	0.189	0.547	0.552	0.282	0.281	0.428	0.392	0.367	0.540	0.229	0.235	0.090	0.432	0.394	0.386
CCNet	R-101	0.419	0.327	0.416	0.105	0.083	0.137	0.425	0.558	0.311	0.407	0.482	0.268	0.333	0.471	0.214	0.492	0.084	0.195	0.572	0.343
DeepLabV3+	R-101	0.488	0.603	0.412	0.056	0.278	0.318	0.282	0.381	0.459	0.084	0.334	0.294	0.309	0.319	0.374	0.284	0.033	0.036	0.247	0.305
DNL	R-101	0.316	0.428	0.243	0.206	0.025	0.196	0.127	0.429	0.357	0.228	0.392	0.285	0.203	0.395	0.429	0.626	0.304	0.466	0.245	0.204
PointRend	R-101	0.281	0.367	0.204	0.219	0.197	0.220	0.263	0.391	0.263	0.234	0.278	0.146	0.278	0.279	0.050	0.348	0.203	0.013	0.404	0.207
NonLocal Net	R-101	0.276	0.290	0.379	0.349	0.037	0.211	0.272	0.309	0.261	0.305	0.392	0.289	0.174	0.072	0.124	0.338	0.027	0.116	0.229	0.287
PFPN	R-101	0.396	0.235	0.273	0.353	0.315	0.375	0.370	0.516	0.242	0.211	0.197	0.356	0.341	0.076	0.204	0.341	0.308	0.073	0.222	0.325
OCRNet	H-W48	0.326	0.215	0.281	0.054	0.227	0.059	0.207	0.244	0.296	0.110	0.068	0.253	0.118	0.274	0.372	0.282	0.109	0.183	0.265	0.295
CGNet	R-101	0.286	0.215	0.306	0.311	0.195	0.214	0.167	0.263	0.309	0.251	0.101	0.247	0.244	0.165	0.485	0.521	0.324	0.176	0.359	0.107

References

[1] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *ECCV*, 2018. 2, 9

[2] MMSegmentation Contributors. MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark. https://github.com/open-mmlab/mmsegmentation, 2020. 2

[3] 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 *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 9

[4] Qing Guo, Jingyang Sun, Felix Juefei-Xu, Lei Ma, Xiaofei Xie, Wei Feng, and Yang Liu. Efficientderain: Learning pixel-wise dilation filtering for high-efficiency single-image deraining. In *AAAI*, 2021. 1

[5] Junjun He, Zhongying Deng, Lei Zhou, Yali Wang, and Yu Qiao. Adaptive pyramid context network for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 2, 9

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2

[7] Zilong Huang, Xinggang Wang, Lichao Huang, Chang Huang, Yunchao Wei, and Wenyu Liu. Ccnet: Criss-cross attention for semantic segmentation. 2019. 2, 9

[8] Alexander Kirillov, Ross Girshick, Kaiming He, and Piotr Dollar. Panoptic feature pyramid networks. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2019. 2, 9

[9] Alexander Kirillov, Yuxin Wu, Kaiming He, and Ross Girshick. Pointrend: Image segmentation as rendering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9799–9808, 2020. 2, 9

[10] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2019. 1, 2

[11] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *European Conference on Computer Vision*, pages 262–277. Springer, 2018. 1

[12] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In *ICCV*, 2019. 1

[13] Xing Liu, Masanori Suganuma, Zhun Sun, and Takayuki Okatani. Dual residual networks leveraging the potential of paired operations for image restoration. In *Proc. Conference on Computer Vision and Pattern Recognition*, pages 7007–7016, 2019. 1, 2

[14] Claudio Michaelis, Benjamin Mitzkus, Robert Geirhos, Evgenia Rusak, Oliver Bringmann, Alexander S Ecker, Matthias Bethge, and Wieland Brendel. Benchmarking robustness in object detection: Autonomous driving when winter is coming. *arXiv preprint arXiv:1907.07484*, 2019. 2

[15] Rui Qian, Robby T. Tan, Wenhan Yang, Jiajun Su, and Jiaying Liu. Attentive generative adversarial network for raindrop removal from a single image. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 1

[16] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 11908–11915, 2020. 1

[17] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 1

[18] Pranjay Shyam, Kuk-Jin Yoon, and Kyung-Soo Kim. Towards domain invariant single image dehazing. *arXiv preprint arXiv:2101.10449*, 2021. 1

[19] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7794–7803, 2018. 2, 9

[20] Zhendong Wang, Xiaodong Cun, Jianmin Bao, and Jianzhuang Liu. Uformer: A general u-shaped transformer for image restoration. *arXiv preprint arXiv:2106.03106*, 2021. 1, 2

[21] Tianyi Wu, Sheng Tang, Rui Zhang, Juan Cao, and Yongdong Zhang. Cgnet: A light-weight context guided network for semantic segmentation. *IEEE Transactions on Image Processing*, 30:1169–1179, 2020. 2, 9

[22] Minghao Yin, Zhuliang Yao, Yue Cao, Xiu Li, Zheng Zhang, Stephen Lin, and Han Hu. Disentangled non-local neural networks, 2020. 2, 9

[23] Yuhui Yuan, Xilin Chen, and Jingdong Wang. Object-contextual representations for semantic segmentation. 2020. 2, 9

[24] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *CVPR*, 2021. 1, 2

[25] Hongguang Zhang, Yuchao Dai, Hongdong Li, and Piotr Koniusz. Deep stacked hierarchical multi-patch network for image deblurring. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 1, 2

[26] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. Plug-and-play image restoration with deep denoiser prior. *arXiv preprint*, 2020. 1

[27] Shiyu Zhao, Lin Zhang, Ying Shen, and Yicong Zhou. Refinednet: A weakly supervised refinement framework for single image dehazing. *IEEE Transactions on Image Processing*, 30:3391–3404, 2021. 1

[28] Zhen Zhu, Mengde Xu, Song Bai, Tengteng Huang, and Xiang Bai. Asymmetric non-local neural networks for semantic segmentation. In *International Conference on Computer Vision*, 2019. 2, 7, 8, 9, 10, 11