

Supplementary Material - Lightweight HDR Camera ISP for Robust Perception in Dynamic Illumination Conditions via Fourier Adversarial Networks

Pranjay Shyam¹

pranjayshyam@kaist.ac.kr

Sandeep Singh Sengar²

sengar@di.ku.dk

Kuk-Jin Yoon¹

kjyoon@kaist.ac.kr

Kyung-Soo Kim¹

kyungsookim@kaist.ac.kr

¹ Korea Advanced Institute of Science and Technology (KAIST)

Daejeon, Republic of Korea

² University of Copenhagen

Copenhagen, Denmark

1 Extended Analysis of Ablation Studies

Extending the analysis presented in the main paper, we examine the impact of using fourier discriminator network during training by analyzing the fourier spectrum of enhanced images. For this, we choose five different training configurations,

- (I) Two stage network with 5 RDB blocks and proposed cMSFE-A, referred as Baseline.
- (II) Baseline trained with fourier discriminator using only magnitude from gray scale version of enhanced image.
- (III) Baseline trained with complete fourier spectrum from gray scale version of enhanced image.
- (IV) Baseline trained with gray scale based fourier discriminator and RGB based PatchGAN [1] discriminator.
- (V) Replacing gray scale based fourier discriminator with RGB based, wherein per channel fourier transform is calculated and used for training the baseline.

From qualitative results summarized in Fig. 1, we can observe images captured in low illumination content to have a perceivable difference in magnitude spectrum, specifically along the higher frequencies that represent edge information. Subsequently, we observe baseline algorithm to improve image quality as observable from magnitude spectrum (Fig. 1 (I)), which is improved if magnitude spectrum is utilized as adversarial training process (Fig. 1 (II)). While it improves structural consistency of enhanced image with respect to ground truth, the quantum is irregular in high frequency regions resulting in sharp lines visible in

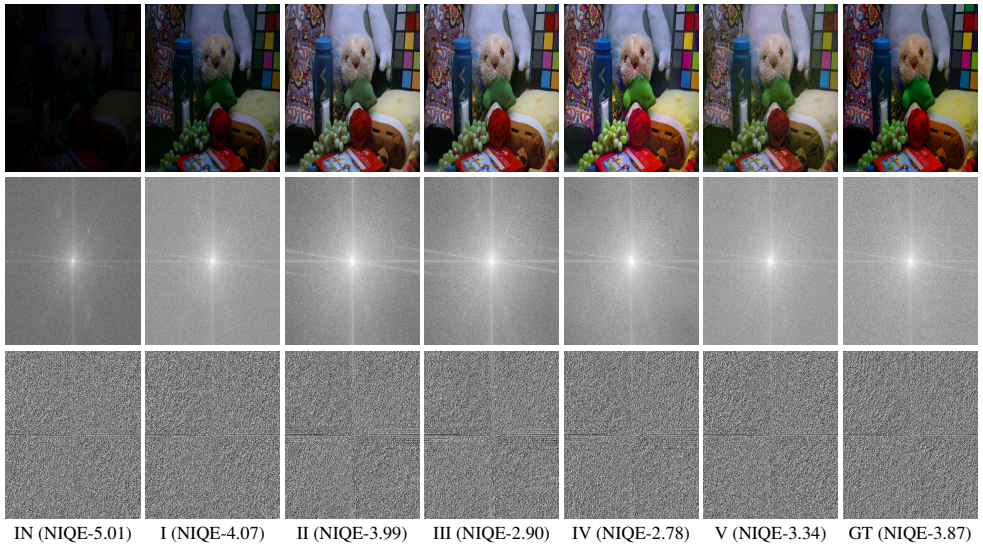


Figure 1: Magnitude and Phase information of Images enhanced by frameworks following different training methodologies.

magnitude spectrum. This irregular increase can be attributed to the absence of phase information in the optimization process, that can be overcome by utilizing the complete frequency information during optimization (Fig. 1 (III)). Using phase spectrum ensures the magnitude plot of enhanced image is balanced while ensuring higher naturalness of images, as measured using NIQE. Following prior works involving GANs, we are motivated to examine if using PatchGAN improves image quality further or not. From the magnitude spectrum and NIQE scores (Fig. 1 (IV)) we observe that magnitude spectrum that is sharper and balanced, while NIQE is lower, demonstrating the images to be more natural. We believe this to arise from the method to classify patches within images as real or fake, thus ensuring more natural enhanced images. Compelled by this, we explore if calculating FFT per channel of RGB images to result in more natural images, however, we observe a higher NIQE score and a dim image (compared to enhanced images generated by other approaches), convincing us to conclude that Fourier spectrum of grayscale images is more beneficial. Hence we demonstrate that Fourier discriminator ensures higher structural details whereas PatchGAN ensures a more natural image. We proceed with this configuration to compare performance with SoTA algorithms on sRGB and RAW images while examining its practicality in real-world applications.

2 Extended Performance Comparison with SoTA

2.1 Image Enhancement on sRGB Images

While we evaluated the performance of SoTA LLIE algorithms on commonly used LOL dataset, we witnessed a lack of diversity within evaluation scenarios that are more likely to be encountered in deployment settings (e.g., multiple illumination sources, effects of clouds and shadows) as well as illumination conditions (overcast, night). Hence we additionally evaluated the performance of SoTA algorithms on SICE dataset that contains both moderate

(-1 ev) to extreme (-3 ev) low illumination conditions while also capturing moderate (+1 ev) to high (+3 ev) high exposure conditions, providing a diverse range of evaluation scenarios. We summarize qualitative results of different SoTA algorithms in moderately illuminated conditions (-1 ev¹, Fig. 2, Fig. 5 along with extremely dark conditions (-3 ev, Fig. 3, Fig. 4) from LOL and SICE datasets and summarize the quantitative performance in Tab. 1. While all algorithms suffer consistent drop when evaluated on images captured at -1ev setting (Tab. 1), the performance drop (LPIPS and SSIM) was more significant in images captured at -3ev. We believe this arises from extreme illumination variation wherein certain regions are under-exposed, and local illumination sources illuminate certain regions unevenly. These scenarios require region-specific enhancement, resulting in enhancement inconsistency observed in SoTA algorithms (such as RetinexNet, URIE) constructed using the assumption of a single global illumination model.

Furthermore a visual inspection of enhanced images (Fig. 2, Fig. 3, Fig. 4, Fig. 5) from SoTA algorithms across different settings reveal several limitations, specifically in terms of changing image properties and features within enhanced images. RetinexNet, while enhancing images, also stylizes them, resulting in large NIQE (indicative of deteriorated perceptual quality) and low SSIM (reduced structural consistency) value. Apart from image stylization issues, we also observe presence of noise in results generated by RetinexNet, GLAD, and DSLR. On a closer inspection of results from KinD and MBLLN, we observe the loss in texture and color information. While results from DALE are affected by contrast variation, DLN and URIE suffer from reduced image sharpness arising due to poor edge recovery. These observations highlight the limitations of current SoTA algorithms towards noise amplification, inaccurate recovery of texture, color, and edge information. However, we observe a majority of algorithms achieving visually pleasing images compared to the ground truth, which is corroborated by a lower NIQE (compared to ground truth) and LPIPS score, indicating these images are more natural.

The proposed approach achieves a higher SSIM score and lower NIQE and LPIPS scores compared to prior algorithms, thereby representing better texture and color information while ensuring accurate edge consistency that is visually corroborated by enhanced images. We attribute this performance to be driven by using an adversarial Fourier network during the training cycle that aided the underlying network to enhance images while suppressing noise. In terms of PSNR score comparison, while it is commonly used for low level image processing tasks such as dehazing, deblurring, etc., it cannot be solely relied upon in LLIE as in some cases (DLN, KinD and ours), the enhanced images are more visually pleasing than corresponding ground truth.

2.2 Image Enhancement on RAW Images

We now examine the performance of SoTA algorithms on RAW images to evaluate if they could incorporate the functionality of camera-ISP and benefit from raw signals instead of sRGB images. To ensure that SoTA LLIE algorithms work seamlessly on demosaiced images and generate sRGB images that are subsequently up-scaled using bicubic interpolation, we use Sony-SID dataset and choose KinD, GLAD, EnlightenGAN as they are comparatively shallow networks, thus requiring lower computational resources to be retrained on demosaiced images. These algorithms are retrained following their respective optimization processes and evaluated with SoTA learning based ISP such RAW2RGB-GAN [44], TENet

¹ev represents Exposure Value

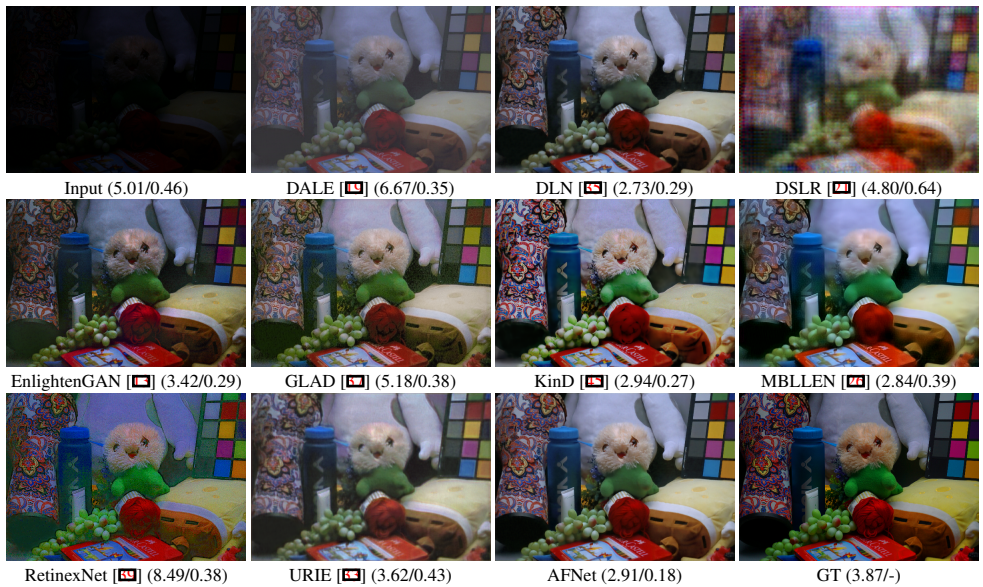


Figure 2: Performance of SoTA algorithms on image from LOL dataset. Numbers in brackets represent NIQE and LPIPS score respectively.

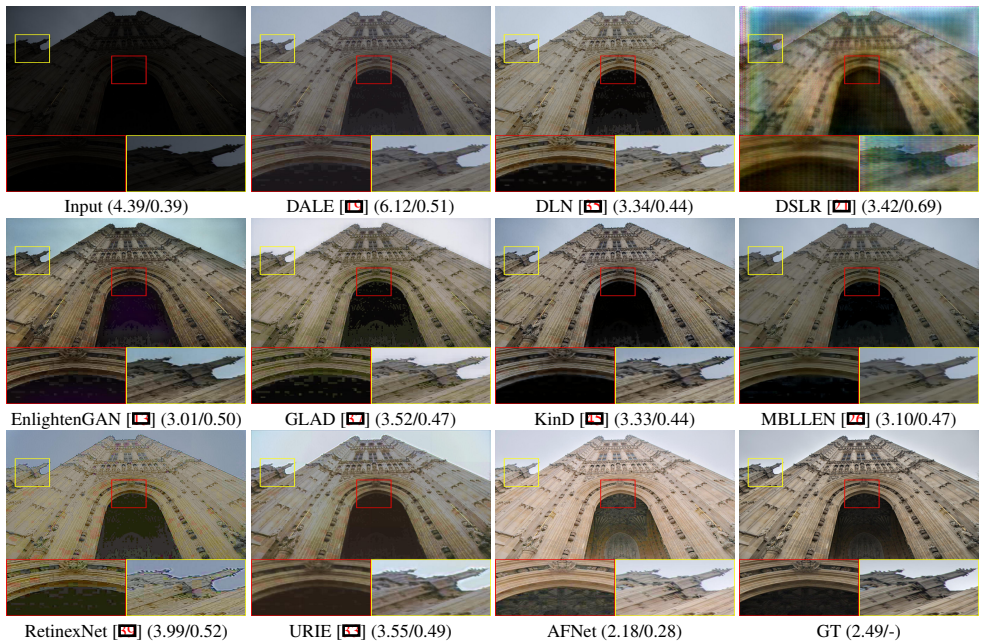


Figure 3: Performance of SoTA algorithms on extremely dark image from SICE dataset captured at -3ev setting. Numbers in brackets represent NIQE and LPIPS score respectively.

[15], PyNet [15], AWWNet [15], as well as RAW image based enhancement algorithms such as Rawpy², SID [15]. From quantitative results in Tab. 2, we observe comparable performance

²<https://letmaik.github.io/rawpy/api/index.html>



Figure 4: Performance of SoTA algorithms on extremely dark image from SICE dataset captured at -3ev setting. Numbers in brackets represent NIQE and LPIPS score respectively.



Figure 5: Performance of SoTA algorithms on moderately dark image from SICE dataset captured at -1ev setting. Numbers in brackets represent NIQE and LPIPS score respectively.

Table 1: Performance Evaluation of different illumination conditions on sRGB images from SICE dataset.

Algorithm	SICE -3ev			SICE -1ev			SICE 1ev			SICE 3ev		
	PSNR / SSIM	NIQE / LPIPS		PSNR / SSIM	NIQE / LPIPS		PSNR / SSIM	NIQE / LPIPS		PSNR / SSIM	NIQE / LPIPS	
Input	7.32 / 0.21	4.34 / 0.46		9.88 / 0.46	3.49 / 0.32		7.18 / 0.51	5.38 / 0.49		5.81 / 0.26	7.13 / 0.73	
DALE	16.39 / 0.53	4.61 / 0.45		17.82 / 0.59	2.77 / 0.35		12.60 / 0.57	2.41 / 0.33		8.90 / 0.53	2.34 / 0.39	
DLN	17.93 / 0.57	2.72 / 0.45		16.44 / 0.60	2.32 / 0.36		14.75 / 0.60	2.25 / 0.32		10.84 / 0.58	2.35 / 0.36	
DSLR	15.43 / 0.41	3.62 / 0.65		14.53 / 0.48	2.95 / 0.57		11.29 / 0.51	2.75 / 0.53		8.60 / 0.49	2.67 / 0.54	
EnlightenGAN	16.17 / 0.55	2.65 / 0.47		15.49 / 0.59	2.37 / 0.38		12.63 / 0.59	2.37 / 0.36		9.71 / 0.55	2.60 / 0.38	
GLAD	17.74 / 0.56	2.87 / 0.47		17.78 / 0.61	2.49 / 0.36		17.64 / 0.64	2.42 / 0.33		15.31 / 0.63	2.36 / 0.37	
KinD	16.72 / 0.54	2.92 / 0.46		17.64 / 0.60	2.44 / 0.35		14.78 / 0.60	2.45 / 0.32		10.62 / 0.57	2.66 / 0.35	
MBLLEN	15.36 / 0.50	2.77 / 0.47		15.48 / 0.53	2.64 / 0.39		14.31 / 0.56	2.72 / 0.38		12.75 / 0.56	3.07 / 0.44	
RetinexNet	17.18 / 0.51	3.85 / 0.52		14.78 / 0.52	3.46 / 0.41		12.97 / 0.55	3.13 / 0.35		10.14 / 0.54	2.93 / 0.38	
URIE	17.91 / 0.57	2.78 / 0.49		17.03 / 0.59	2.21 / 0.40		14.91 / 0.60	2.30 / 0.36		11.75 / 0.57	2.48 / 0.39	
Ours	19.48 / 0.67	2.04 / 0.21		20.79 / 0.71	1.57 / 0.18		18.12 / 0.75	3.40 / 0.31		16.63 / 0.84	3.96 / 0.23	

Table 2: Performance Evaluation of different illumination conditions on RAW images from SID-Sony and ELD datasets.

Algorithm	SID-Sony			ELD-Sony (RAW)			ELD-Sony (sRGB)		
	PSNR / SSIM	NIQE / LPIPS		PSNR / SSIM	NIQE / LPIPS		PSNR / SSIM	NIQE / LPIPS	
Rawpy	20.17 / 0.53	7.07 / 0.58		21.02 / 0.61	5.54 / 0.47		-	-	-
EnlightenGAN	24.27 / 0.64	4.68 / 0.53		25.98 / 0.87	3.71 / 0.14		18.98 / 0.69	3.62 / 0.38	
RAW2RGB-GAN	23.55 / 0.78	4.00 / 0.71		24.08 / 0.71	4.20 / 0.34		-	-	-
KinD	26.91 / 0.73	4.10 / 0.39		25.00 / 0.81	3.43 / 0.28		18.40 / 0.76	3.43 / 0.28	
GLAD	27.11 / 0.82	3.86 / 0.39		25.25 / 0.87	3.48 / 0.18		20.25 / 0.78	3.48 / 0.24	
SID	28.88 / 0.78	4.39 / 0.43		24.89 / 0.71	4.96 / 0.33		-	-	-
TENet	24.17 / 0.73	3.18 / 0.31		23.63 / 0.74	5.76 / 0.31		-	-	-
PyNet	25.01 / 0.69	3.79 / 0.34		24.08 / 0.77	4.04 / 0.20		-	-	-
PyNet-CA	24.24 / 0.64	4.02 / 0.41		21.98 / 0.72	4.02 / 0.29		-	-	-
AWNNet	25.09 / 0.66	3.98 / 0.39		23.49 / 0.77	5.76 / 0.38		-	-	-
Ours	27.67 / 0.84	3.94 / 0.37		25.75 / 0.83	3.98 / 0.24		20.72 / 0.83	3.19 / 0.23	

between sRGB and RAW image enhancement algorithms (Fig. 6). Furthermore we observe learning based algorithms to generate visually pleasing images with lower noise and lower NIQE scores, compared to traditional ISP mechanism represented by Rawpy that generates noisy image. This is representative of issues faced by traditional camera-ISP in low illumination conditions. We compare the enhanced and processed images with ground truth that is captured in well illuminated condition.

One noticeable observation was that the performance of these algorithms surpassed their corresponding sRGB models, hence to examine if it is caused by camera-ISP or change in the distribution of images, we use models trained on sRGB images from LOL and evaluate the performance on the ELD dataset (ELD-Sony). From the quantitative results in Tab. 2, we summarize that nonlinearities caused by camera-ISP indeed affects the performance of LLIE algorithms. As we are comparing performance of LLIE algorithms on sRGB and RAW images, we donot require to observe performance of learning based ISP networks on sRGB images.

3 Practical Implications of IE algorithms in Real-World Scenarios

We now explore practical applications of current image enhancement algorithms in real world scenarios such as object detection and semantic segmentation. As there lacks a paired

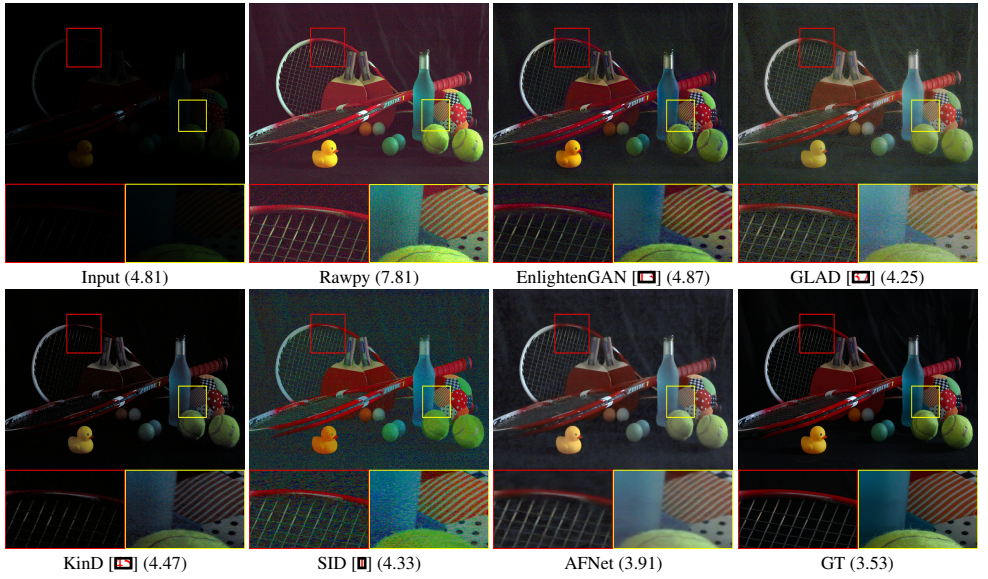


Figure 6: Performance of SoTA algorithms on RAW image from ELD dataset. Numbers in brackets represent NIQE scores, wherein the ground truth is represented by the JPEG image generated by onboard camera-ISP.

dataset diverse enough to capture and represent all illumination conditions, while providing different label attributes, we customize cycle-gan algorithm to translate images present in benchmark datasets to capture different extreme illumination conditions while sharing the same labels. Our modifications are inspired by recent developments in efficiently training GAN based algorithms such as multi-scale discriminators proposed in Pix-to-Pix HD [56], Augmentation [44], Self Attention [44] and Contrastive Learning [15] with the complete pipeline summarized in Fig. 7, with qualitative results presented in Fig. 8. To evaluate the quality of translated images, we compare the performance of with original CycleGAN and CoMoGAN [27] that represents the best algorithm for performing image translation by modeling the complete illumination range and providing a control parameter for varying the illumination in an image.

We train the proposed cyclegan for 10 epochs, using similar settings as original cyclegan framework and utilizing JOL [52] and BDD100K dataset [43] for unpaired training. We donot train the modified network extensively as we require different light sources that are present in generated images. Since these images are representative of real conditions, we leverage this imperfect network to analyze the quality of enhanced images and its subsequent effect on underlying algorithms in presence of multiple local illumination sources.

Subsequently we use this modified CycleGAN to generate low illumination version of cityscapes [9] and COCO-val [22] datasets and analyze the performance of SoTA semantic segmentation and object detection algorithms available in mmseg [8] and mmdetection [8] software packages. Furthermore we also utilize the Exdark dataset [25] to evaluate performance of object detectors in varying illumination conditions using models pretrained on COCO dataset. We donot retrain these algorithms as the objective is to quantify the impact of varying illumination conditions on pretrained models and examine if improving image quality improves performance or not.

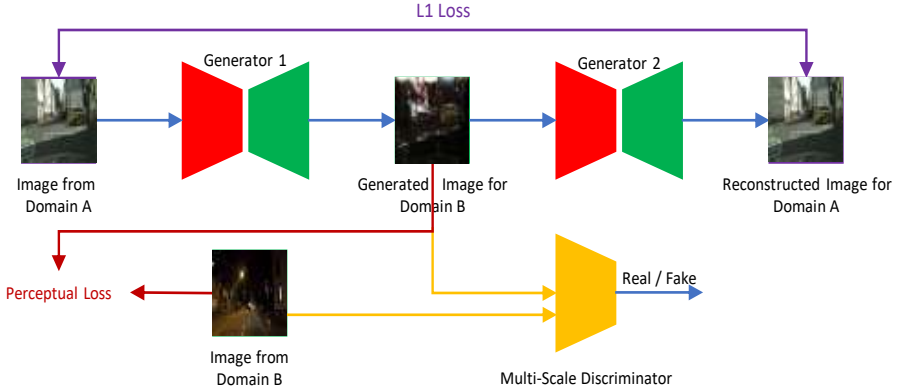


Figure 7: Overview of the modified CycleGAN

Table 3: Performance of SoTA object detection algorithms in low-light and enhanced images using ExDark Dataset.

Method	B.B.	Baseline		EnlightenGAN		DLN		GLAD		RetinexNet		AFNet	
		AP	AP ₃₀	AP	AP ₃₀	AP	AP ₃₀	AP	AP ₃₀	AP	AP ₃₀	AP	AP ₃₀
HTC [8]	R-50	0.308	0.619	0.284	0.573	0.285	0.574	0.277	0.558	0.210	0.440	0.303	0.507
GFL [10]	R-50	0.298	0.599	0.276	0.556	0.273	0.552	0.262	0.530	0.196	0.408	0.315	0.590
PAA [10]	R-50	0.309	0.618	0.290	0.580	0.288	0.576	0.280	0.561	0.213	0.439	0.299	0.471
Auto Assign [10]	R-50	0.290	0.592	0.271	0.554	0.271	0.552	0.262	0.533	0.196	0.407	0.290	0.431
D. DETR [10]	R-50	0.300	0.609	0.280	0.569	0.281	0.568	0.272	0.548	0.211	0.432	0.288	0.438
Faster RCNN [10]	R-50	0.278	0.584	0.254	0.541	0.252	0.536	0.244	0.519	0.179	0.394	0.273	0.501
RetinexNet [10]	R-50	0.290	0.593	0.266	0.552	0.264	0.546	0.252	0.524	0.188	0.403	0.296	0.492
SCNet [10]	R-50	0.311	0.627	0.290	0.586	0.287	0.581	0.277	0.559	0.217	0.452	0.289	0.663
SSD-300 [10]	VGG16	0.254	0.526	0.246	0.505	0.249	0.509	0.240	0.493	0.170	0.358	0.253	0.395
YOLOF [8]	R-50	0.300	0.596	0.277	0.553	0.274	0.548	0.262	0.526	0.170	0.358	0.293	0.454
Yolov3-320 [10]	DarkNet-53	0.246	0.518	0.242	0.508	0.246	0.517	0.237	0.498	0.167	0.358	0.273	0.417
DANet [10]	R-50	0.319	0.631	0.303	0.599	0.299	0.595	0.289	0.576	0.221	0.451	0.307	0.514

We summarize the quantitative results of SoTA object detection algorithms in Tab. 3 and Tab. 4 on ExDark and COCO-val datasets respectively, while demonstrating the qualitative performance in Fig. 9. Furthermore we utilize the JOL [8] dataset that captures object and road information captured in dynamic illumination conditions such as driving across cities and examine performance of object detection algorithms in different conditions, with qualitative results summarized in Fig. 10. Similarly we perform the same experiment on low illuminated translated images from cityscapes datasets with quantitative results summarized in Tab. 5 and qualitative results on low illuminated images and enhanced images in Fig. 11.

From these results we can summarize that low illumination conditions indeed affects the performance of SoTA algorithms whereas enhanced images can improve performance of these algorithms as summarized from results on ExDark dataset. While we observe performance of SoTA Object Detection and Semantic Segmentation algorithms to improve on enhanced images on COCO and Cityscapes datasets, the performance doesn't reach the baseline performance. We believe this to arise from image stylization within original images as a consequence of image translation, that is consistent across all image translation networks. However we do observe performance improvement when images are enhanced and distortions by lightening conditions minimized. Hence we can conclude that image enhancement algorithm can improve performance of SoTA perception algorithms.

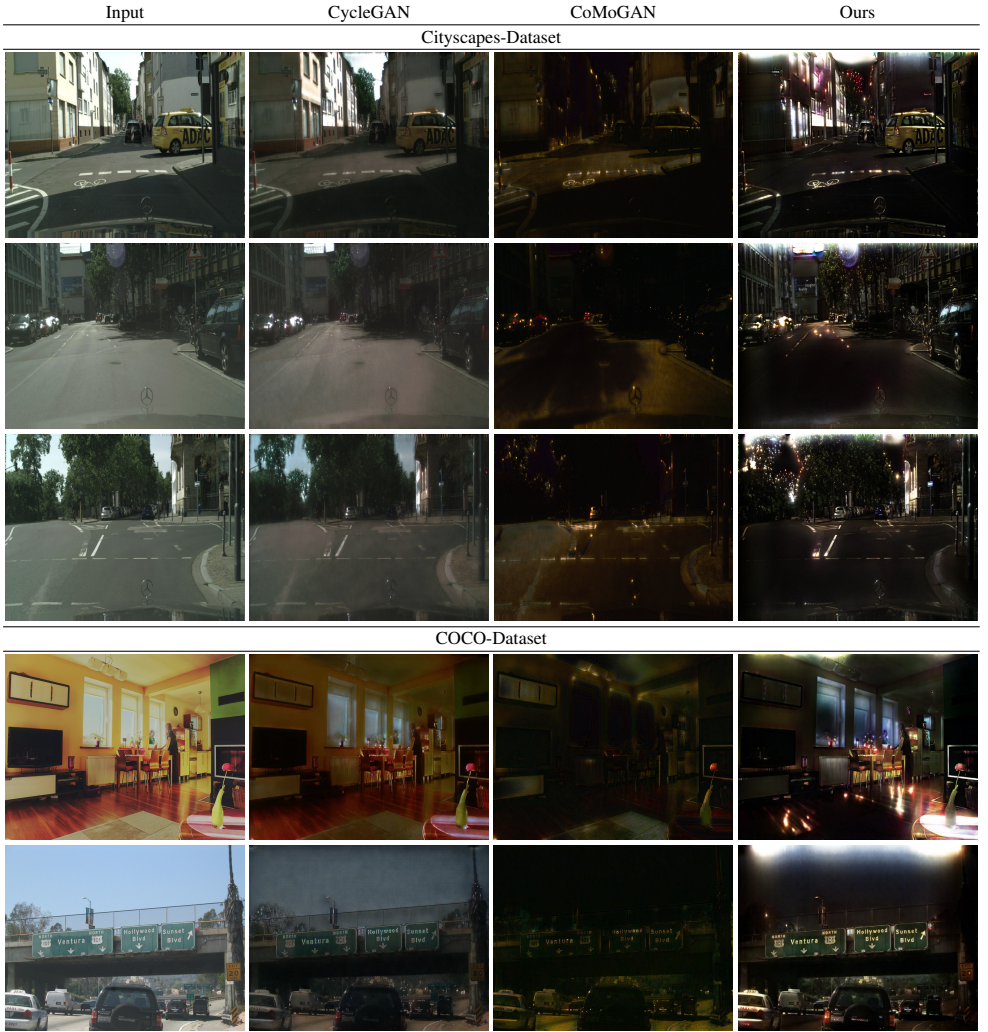


Figure 8: Visual Analysis of Images generated using different Image Translation Algorithms

Table 4: Performance of SoTA object detection algorithms in low-light and enhanced images using COCO Dataset.

Method	Baseline		Dark		EnlightenGAN		DLN		GLAD		RetinexNet		AFNet	
	AP	AP ₅₀	AP	AP ₅₀	AP	AP ₅₀	AP	AP ₅₀	AP	AP ₅₀	AP	AP ₅₀	AP	AP ₅₀
HTC [9]	0.433	0.622	0.273	0.417	0.279	0.423	0.273	0.417	0.276	0.422	0.244	0.378	0.279	0.426
GFL [10]	0.429	0.612	0.257	0.393	0.266	0.405	0.257	0.393	0.262	0.400	0.232	0.357	0.264	0.402
PAA [11]	0.433	0.610	0.269	0.403	0.277	0.412	0.269	0.403	0.273	0.410	0.238	0.360	0.274	0.409
Auto Assign [12]	0.404	0.596	0.253	0.395	0.261	0.406	0.253	0.395	0.258	0.404	0.224	0.355	0.260	0.404
D. DETR [13]	0.468	0.663	0.290	0.431	0.292	0.431	0.290	0.431	0.293	0.437	0.256	0.385	0.298	0.442
Faster RCNN [14]	0.384	0.590	0.228	0.376	0.234	0.386	0.228	0.376	0.234	0.384	0.201	0.334	0.232	0.382
RetinaNet [15]	0.374	0.567	0.222	0.359	0.231	0.369	0.222	0.359	0.227	0.366	0.194	0.317	0.228	0.367
SCNet [16]	0.445	0.641	0.293	0.454	0.300	0.460	0.293	0.454	0.295	0.455	0.262	0.411	0.298	0.460
SSD-300 [17]	0.256	0.438	0.172	0.307	0.179	0.315	0.172	0.307	0.177	0.313	0.149	0.267	0.174	0.309
YOLOv3 [18]	0.375	0.570	0.240	0.385	0.246	0.394	0.240	0.385	0.245	0.393	0.205	0.335	0.244	0.393
Yolov3-320 [19]	0.279	0.491	0.195	0.354	0.197	0.358	0.195	0.354	0.197	0.358	0.163	0.301	0.194	0.354
DANet [20]	0.423	0.612	0.288	0.438	0.291	0.443	0.288	0.438	0.291	0.444	0.253	0.390	0.293	0.445



Figure 9: Qualitative Performance of SoTA object detectors on Low Light and Enhanced Images.

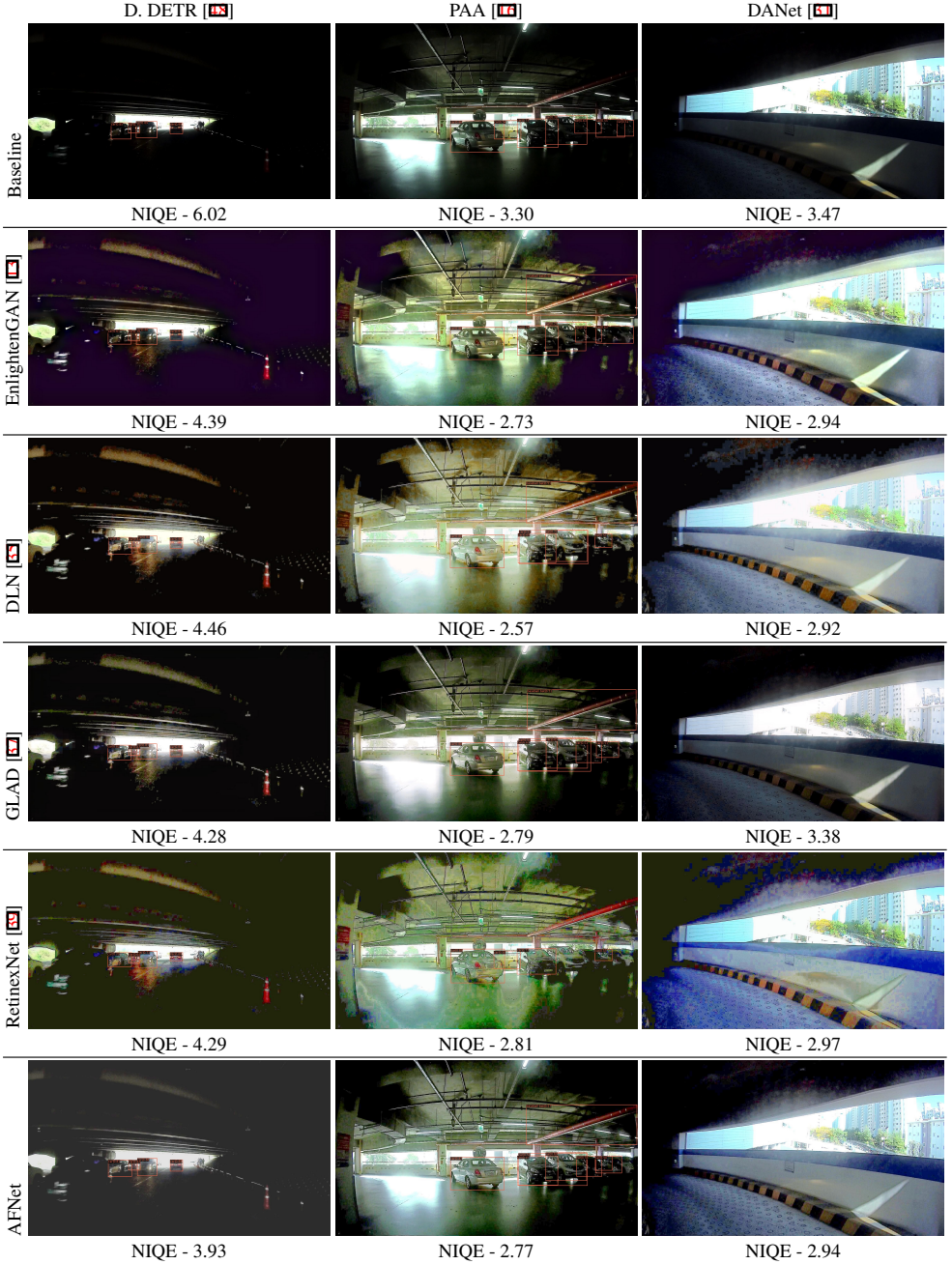


Figure 10: Qualitative Performance of SoTA image enhancement algorithms on images captured in wild and subsequent effect on SoTA object detection algorithms.

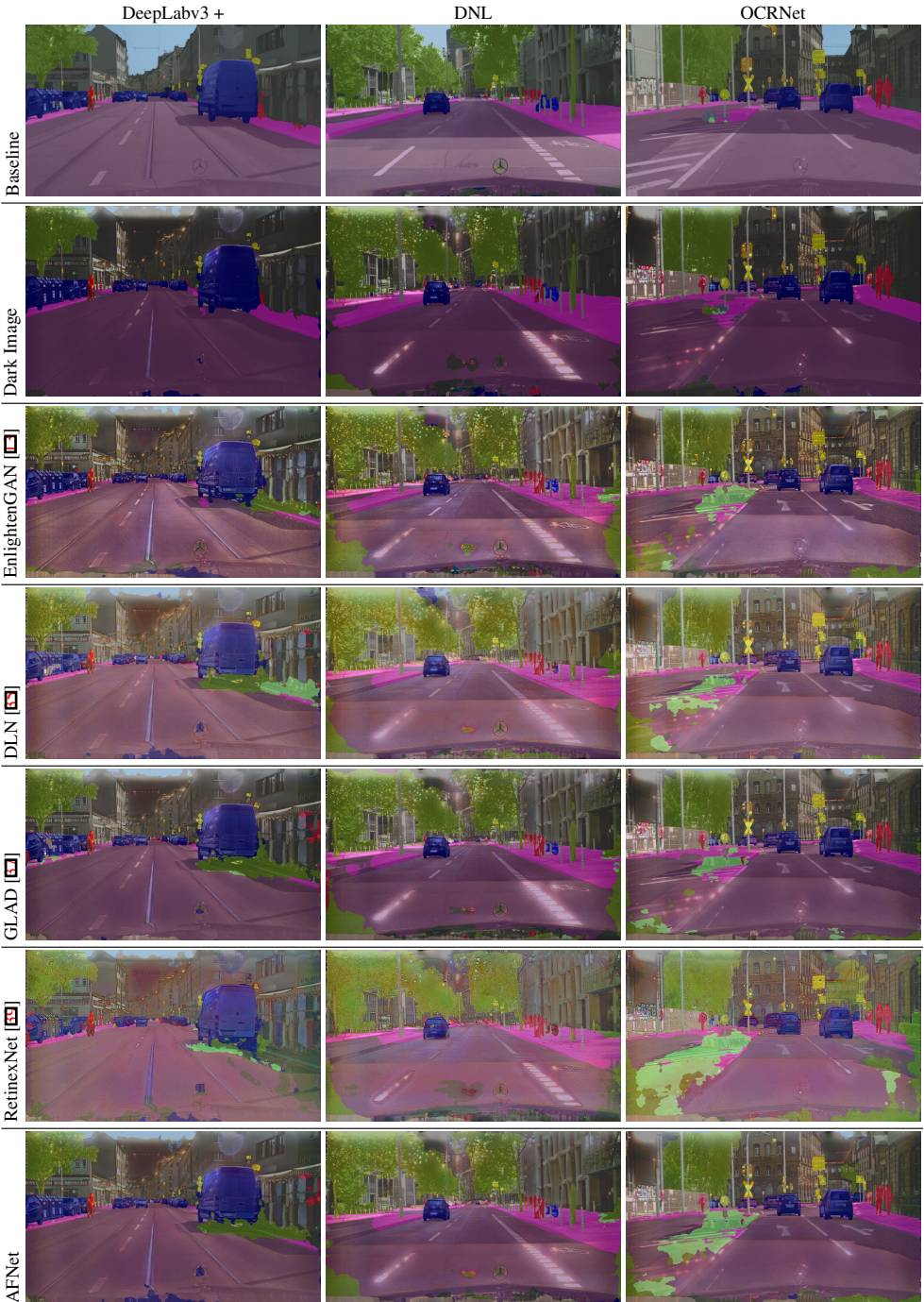


Figure 11: Qualitative Performance of SoTA Semantic Segmentation Algorithms on Low Light and Enhanced Images.

Table 5: Semantic segmentation results for Cityscapes Dataset under Diverse Conditions

Method	B.B.	Baseline	Low Light	EnlightenGAN	RetinexNet	DLN	GLAD	AFNet
ANN [10]	R-50	0.773	0.573	0.470	0.331	0.418	0.457	0.590
APCNet [9]	R-50	0.788	0.590	0.477	0.359	0.433	0.471	0.554
CCNet [10]	R-50	0.776	0.580	0.497	0.356	0.447	0.477	0.542
DeepLabV3+ [9]	R-50	0.798	0.629	0.544	0.415	0.503	0.526	0.669
DNL [10]	R-50	0.792	0.608	0.480	0.373	0.454	0.471	0.559
PointRend [10]	R-50	0.763	0.576	0.493	0.347	0.453	0.484	0.558
NonLocal Net [10]	R-50	0.781	0.567	0.487	0.338	0.418	0.469	0.530
PFPN [10]	R-50	0.744	0.561	0.458	0.330	0.438	0.469	0.632
OCRNet [10]	H-W18	0.793	0.653	0.565	0.395	0.525	0.574	0.660
CGNet [10]	R-50	0.682	0.499	0.409	0.297	0.374	0.371	0.601

References

- [1] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3291–3300, 2018.
- [2] Kai Chen, Jiangmiao Pang, Jiaqi Wang, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jianping Shi, Wanli Ouyang, Chen Change Loy, and Dahua Lin. Hybrid task cascade for instance segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
- [3] Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jiarui Xu, Zheng Zhang, Dazhi Cheng, Chenchen Zhu, Tianheng Cheng, Qijie Zhao, Buyu Li, Xin Lu, Rui Zhu, Yue Wu, Jifeng Dai, Jingdong Wang, Jianping Shi, Wanli Ouyang, Chen Change Loy, and Dahua Lin. MMDetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*, 2019.
- [4] 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.
- [5] Qiang Chen, Yingming Wang, Tong Yang, Xiangyu Zhang, Jian Cheng, and Jian Sun. You only look one-level feature. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2021.
- [6] MMSegmentation Contributors. MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark. <https://github.com/open-mmlab/mms Segmentation>, 2020.
- [7] 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.
- [8] Linhui Dai, Xiaohong Liu, Chengqi Li, and Jun Chen. Awnet: Attentive wavelet network for image isp. *arXiv preprint arXiv:2008.09228*, 2020.
- [9] 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.
- [10] Zilong Huang, Xinggang Wang, Lichao Huang, Chang Huang, Yunchao Wei, and Wenyu Liu. Ccnet: Criss-cross attention for semantic segmentation. 2019.
- [11] Andrey Ignatov, Luc Van Gool, and Radu Timofte. Replacing mobile camera isp with a single deep learning model. *arXiv preprint arXiv:2002.05509*, 2020.
- [12] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

- [13] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *arXiv preprint arXiv:1906.06972*, 2019.
- [14] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. *arXiv preprint arXiv:2006.06676*, 2020.
- [15] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *arXiv preprint arXiv:2004.11362*, 2020.
- [16] Kang Kim and Hee Seok Lee. Probabilistic anchor assignment with iou prediction for object detection. In *ECCV*, 2020.
- [17] 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. doi: 10.1109/cvpr.2019.00656. URL <http://dx.doi.org/10.1109/CVPR.2019.00656>.
- [18] 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.
- [19] Dokyeong Kwon, Guisik Kim, and Junseok Kwon. Dale: Dark region-aware low-light image enhancement. *arXiv preprint arXiv:2008.12493*, 2020.
- [20] Xiang Li, Wenhai Wang, Lijun Wu, Shuo Chen, Xiaolin Hu, Jun Li, Jinhui Tang, and Jian Yang. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. *arXiv preprint arXiv:2006.04388*, 2020.
- [21] Seokjae Lim and Wonjun Kim. Dslr: Deep stacked laplacian restorer for low-light image enhancement. *IEEE Transactions on Multimedia*, 2020.
- [22] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.
- [23] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, 2017.
- [24] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. Ssd: Single shot multibox detector. *ECCV*, 2016.
- [25] Yuen Peng Loh and Chee Seng Chan. Getting to know low-light images with the exclusively dark dataset. *Computer Vision and Image Understanding*, 178:30–42, 2019. doi: <https://doi.org/10.1016/j.cviu.2018.10.010>.
- [26] Feifan Lv, Feng Lu, Jianhua Wu, and Chongsoon Lim. Mblen: Low-light image/video enhancement using cnns. In *BMVC*, page 220, 2018.

- [27] Fabio Pizzati, Pietro Cerri, and Raoul de Charette. CoMoGAN: continuous model-guided image-to-image translation. In *CVPR*, 2021.
- [28] Guocheng Qian, Jinjin Gu, Jimmy S Ren, Chao Dong, Furong Zhao, and Juan Lin. Trinity of pixel enhancement: a joint solution for demosaicking, denoising and super-resolution. *arXiv preprint arXiv:1905.02538*, 2019.
- [29] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement, 2018.
- [30] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Jun 2017.
- [31] Pranjay Shyam, Kuk-Jin Yoon, and Kyung-Soo Kim. Dynamic anchor selection for improving object localization. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9477–9483, 2020. doi: 10.1109/ICRA40945.2020.9197076.
- [32] Pranjay Shyam, Kuk-Jin Yoon, and Kyung-Soo Kim. Weakly supervised approach for joint object and lane marking detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 2885–2895, October 2021.
- [33] Taeyoung Son, Juwon Kang, Namyup Kim, Sunghyun Cho, and Suha Kwak. Urie: Universal image enhancement for visual recognition in the wild. In *ECCV*, 2020.
- [34] Thang Vu, Kang Haeyong, and Chang D Yoo. Scnet: Training inference sample consistency for instance segmentation. In *AAAI*, 2021.
- [35] Li-Wen Wang, Zhi-Song Liu, Wan-Chi Siu, and Daniel P.K. Lun. Lightening network for low-light image enhancement. *IEEE Transactions on Image Processing*, 2020. doi: 10.1109/TIP.2020.3008396.
- [36] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [37] Wenjing Wang, Chen Wei, Wenhan Yang, and Jiaying Liu. Gladnet: Low-light enhancement network with global awareness. In *Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference*, pages 751–755. IEEE, 2018.
- [38] 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.
- [39] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018.
- [40] Tianyi Wu, Sheng Tang, Rui Zhang, Juan Cao, and Yongdong Zhang. Cgnet: A lightweight context guided network for semantic segmentation. *IEEE Transactions on Image Processing*, 30:1169–1179, 2020.

- [41] Minghao Yin, Zhulian Yao, Yue Cao, Xiu Li, Zheng Zhang, Stephen Lin, and Han Hu. Disentangled non-local neural networks, 2020.
- [42] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2636–2645, 2020.
- [43] Yuhui Yuan, Xilin Chen, and Jingdong Wang. Object-contextual representations for semantic segmentation. 2020.
- [44] Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena. Self-attention generative adversarial networks. In *International Conference on Machine Learning*, pages 7354–7363, 2019.
- [45] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, pages 1632–1640, New York, NY, USA, 2019. ACM. ISBN 978-1-4503-6889-6. doi: 10.1145/3343031.3350926. URL <http://doi.acm.org/10.1145/3343031.3350926>.
- [46] Yuzhi Zhao, Lai-Man Po, Tiantian Zhang, Zongbang Liao, Xiang Shi, et al. Saliency map-aided generative adversarial network for raw to rgb mapping. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 3449–3457. IEEE, 2019.
- [47] Benjin Zhu, Jianfeng Wang, Zhengkai Jiang, Fuhang Zong, Songtao Liu, Zeming Li, and Jian Sun. Autoassign: Differentiable label assignment for dense object detection. *arXiv preprint arXiv:2007.03496*, 2020.
- [48] Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. Deformable detr: Deformable transformers for end-to-end object detection. In *International Conference on Learning Representations*, 2021. URL <https://openreview.net/forum?id=gZ9hCDWe6ke>.
- [49] Zhen Zhu, Mengde Xu, Song Bai, Tengting Huang, and Xiang Bai. Asymmetric non-local neural networks for semantic segmentation. In *International Conference on Computer Vision*, 2019. URL <http://arxiv.org/abs/1908.07678>.