# Semantic-Guided Zero-Shot Learning for Low-Light Image/Video Enhancement
## *Supplementary Material*

Shen Zheng
Wenzhou-Kean University
Wenzhou, China
zhengsh@kean.edu

Gaurav Gupta
Wenzhou-Kean University
Wenzhou, China
ggupta@kean.edu

## 1. Introduction

In this supplementary material, we first show the specific model architecture for our proposed model. Next, we display more images from our proposed dataset. After that, we conduct more visual comparisons on low-light images and low-light video frames. Competing models are previous state-of-the-arts including PIE [8], LIME [13], RetinexNet [48], MBLLEN [34], KinD [52], ZeroDCE [11] and EnlightenGAN [20]. For pure perceptual comparison, we select images from NPE [47], LIME [13], MEF [35], DICM [25], VV [1] and LOL [2]. For task-driven visual comparison we choose images from our proposed DarkBDD, which is selected from BDD10k [49], and DarkCityScape, which is synthesized from CityScape [4].

## 2. Specific model architecture

We present the specific model architecture of our enhancement factor extraction network (EFE) in Table 1 and the depthwise separable convolution layer in Table 2.

| **Require**: Input Image $x$ | |
|---|---|
| **Name** | **Details** |
| Conv1 | Input; DS(3,32); ReLU; |
| Conv2 | DS(32,32); ReLU; |
| Conv3 | DS(32,32);ReLU; |
| Conv4 | DS(32,32);ReLU; |
| Concat1 | Concat(Conv3, Conv4); |
| Conv5 | DS(64,32);ReLU; |
| Concat2 | Concat(Conv2, Conv5); |
| Conv6 | DS(64,32);ReLU; |
| Concat3 | Concat(Conv1, Conv6); |
| Conv7 | DS(64,3);Tanh; Output; |

Table 1. The architecture of EFE. Where "DS" is depthwise separable convolution layer with (input channel, output channel). "Concat" represents tensor concatenation. "ReLU" and "Tanh" are activation functions.

## 3. Samples images from proposed dataset

### 3.1. DarkBDD samples

Following are sample photos from our proposed DarkBDD dataset.

[1] https://sites.google.com/site/vonikakis/datasets

| **Require**: Input Channel $in$, Output Channel $out$ | |
|---|---|
| **Name** | **Details** |
| DConv | Conv($in$, $in$, 3, 1, 1, $in$) |
| PConv | Conv($in$, $out$, 1, 1, 0, 1) |

Table 2. The architecture of depthwise separable convolution layers. Where “Conv” stands for convolution operation with (input channel, output channel, kernel size, stride, padding, groups)

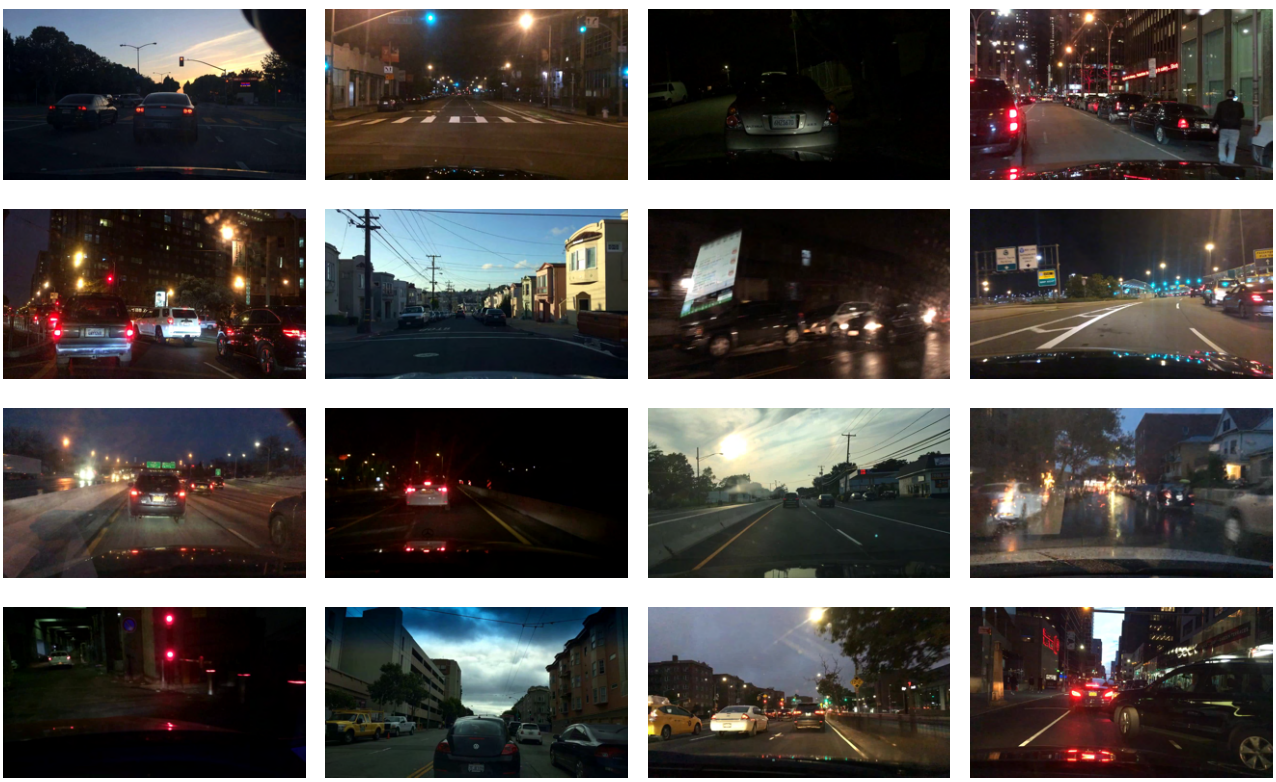


Figure 1. Example images from our DarkBDD dataset. Our selected low-light images have diverse brightness, contrast, and exposure. Due to its challenging nature, it is suitable as a low-light image enhancement benchmark.

### 3.2. DarkCityScape samples

Following are sample photos from our proposed DarkCityScape dataset.

Figure 2. Example images from our DarkCityScape dataset. Upper two rows: groundtruth normal-light images. Lower two rows: our synthesis low-light images. The low-light photos are synthesized using gamma correction. The synthesized dataset is challenging because it hides most image details from human perceptions.

## 4. More Visual Comparison on Low-Light Images

Fig. 3 shows that our model effectively enhances the low-light regions and makes superior edge preservation. In contrast, other models either fail to improve the image contrast or generate unpleasant artifacts.

(a) Dark

(b) MBLLEN [34]

(c) PIE [8]

(d) Retinex [48]

(e) ZeroDCE [11]

(f) Ours

Figure 3. Visual Comparison on DICM [25]

Fig. 4 shows that our model generates the most natural and balanced enhancement result. In comparison, other models either fail to enhance the dark regions or produce significant color deviation.

(a) Dark

(b) MBLLEN [34]

(c) PIE [8]

(d) Retinex [48]

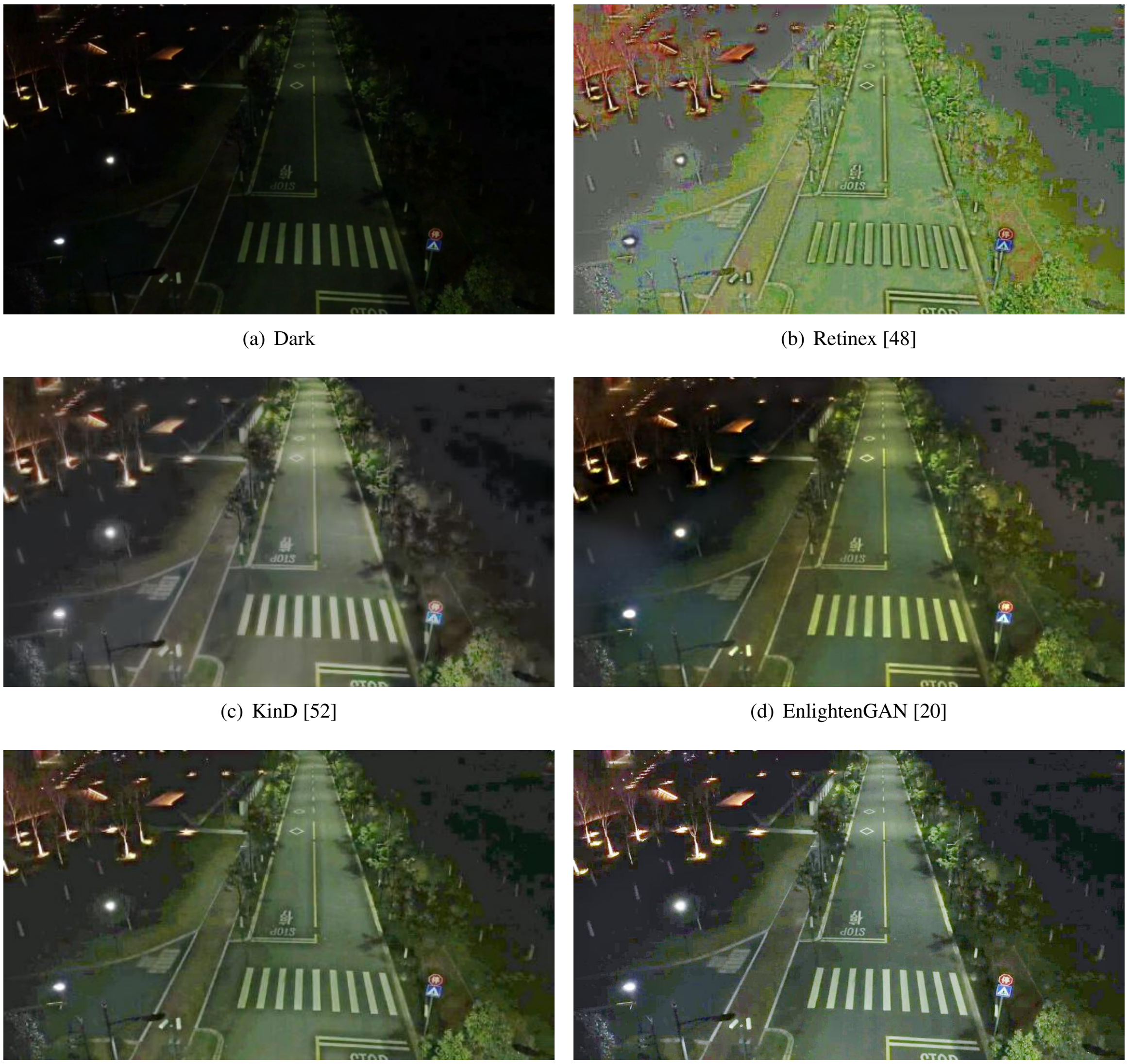
(e) ZeroDCE [11]

(f) Ours

Figure 4. Visual Comparison on VV

## 5. More Visual Comparison on Low-Light Video Frames

Fig. 5 shows that our model generates a natural appearance for the low-light video frame. As opposed to our result, other models either have severe color distortion or generate sub-optimal regional contrast.



(a) Dark
(b) Retinex [48]
(c) KinD [52]
(d) EnlightenGAN [20]
(e) ZeroDCE [11]
(f) Ours

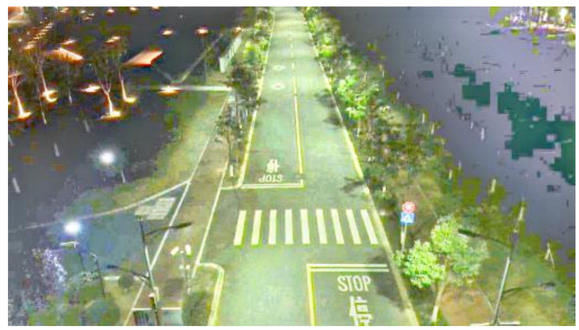
Figure 5. Visual Comparison on Low-Light Video Frames (1)

Fig. 6 shows that our model produces pleasing texture for the low-light video frame. Compared with our result, other models have significant over/under exposure with large background artifacts.


(a) Dark


(b) Retinex [48]


(c) KinD [52]




(d) EnlightenGAN [20]

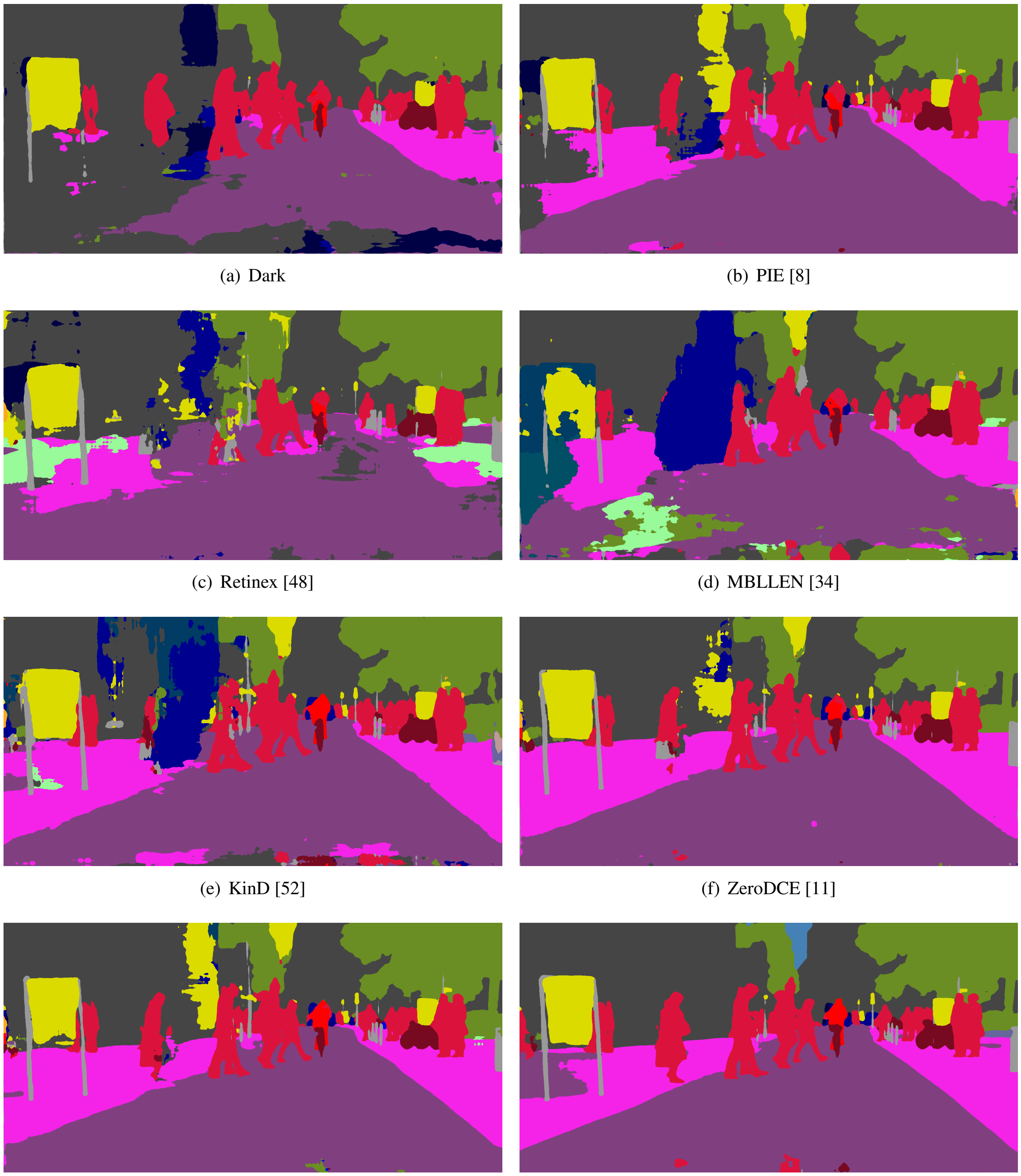

(e) ZeroDCE [11]


(f) Ours

Figure 6. Visual Comparison on Low-Light Video Frames (2)

## 6. More on Semantic Segmentation

Fig. 7 shows our model helps accurate semantic segmentation of most objects and is closest to the groundtruth. In comparison, other models result in large areas of incorrect segmentation.



(a) Dark (b) PIE [8]

(c) Retinex [48] (d) MBLLEN [34]

(e) KinD [52] (f) ZeroDCE [11]

(g) Ours (h) GT

Figure 7. Visual Comparison on Semantic Segmentation

## References


[1] Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. Learning photographic global tonal adjustment with a database of input/output image pairs. In *CVPR 2011*, pages 97–104. IEEE, 2011.

[2] 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.

[3] Qiang Chen, Philippe Montesinos, Quan Sen Sun, Peng Ann Heng, et al. Adaptive total variation denoising based on difference curvature. *Image and vision computing*, 28(3):298–306, 2010.

[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 *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016.

[5] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[6] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010.

[7] Zhiwen Fan, Liyan Sun, Xinghao Ding, Yue Huang, Congbo Cai, and John Paisley. A segmentation-aware deep fusion network for compressed sensing mri. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 55–70, 2018.

[8] Xueyang Fu, Yinghao Liao, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation. *IEEE Transactions on Image Processing*, 24(12):4965–4977, 2015.

[9] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[10] Abel Gonzalez-Garcia, Davide Modolo, and Vittorio Ferrari. Objects as context for detecting their semantic parts. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6907–6916, 2018.

[11] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1780–1789, 2020.

[12] Mengxi Guo, Mingtao Chen, Cong Ma, Yuan Li, Xianfeng Li, and Xiaodong Xie. High-level task-driven single image deraining: Segmentation in rainy days. In *International Conference on Neural Information Processing*, pages 350–362. Springer, 2020.

[13] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2):982–993, 2016.

[14] 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.

[15] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[16] Zhe Hu, Sunghyun Cho, Jue Wang, and Ming-Hsuan Yang. Deblurring low-light images with light streaks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3382–3389, 2014.

[17] Haidi Ibrahim and Nicholas Sia Pik Kong. Brightness preserving dynamic histogram equalization for image contrast enhancement. *IEEE Transactions on Consumer Electronics*, 53(4):1752–1758, 2007.

[18] 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.

[19] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. Multi-scale progressive fusion network for single image deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8346–8355, 2020.

[20] 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. *IEEE Transactions on Image Processing*, 30:2340–2349, 2021.

[21] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016.

[22] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[23] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Fast and accurate image super-resolution with deep laplacian pyramid networks. *IEEE transactions on pattern analysis and machine intelligence*, 41(11):2599–2613, 2018.

[24] Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977.

[25] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation. In *2012 19th IEEE International Conference on Image Processing*, pages 965–968. IEEE, 2012.

[26] Chongyi Li, Chunle Guo, Linghao Han, Jun Jiang, Ming-Ming Cheng, Jinwei Gu, and Chen Change Loy. Lighting the darkness in the deep learning era. *arXiv preprint arXiv:2104.10729*, 2021.

[27] Chongyi Li, Chunle Guo, and Chen Change Loy. Learning to enhance low-light image via zero-reference deep curve estimation. *arXiv preprint arXiv:2103.00860*, 2021.

[28] Jinjiang Li, Xiaomei Feng, and Zhen Hua. Low-light image enhancement via progressive-recursive network. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.

[29] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017.

[30] 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*, pages 2980–2988, 2017.

[31] Daqi Liu, Miroslaw Bober, and Josef Kittler. Visual semantic information pursuit: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 2019.

[32] Ding Liu, Bihan Wen, Xianming Liu, Zhangyang Wang, and Thomas S Huang. When image denoising meets high-level vision tasks: A deep learning approach. *arXiv preprint arXiv:1706.04284*, 2017.

[33] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2017.

[34] Feifan Lv, Feng Lu, Jianhua Wu, and Chongsoon Lim. Mbllen: Low-light image/video enhancement using cnns. In *BMVC*, page 220, 2018.

[35] Kede Ma, Kai Zeng, and Zhou Wang. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 24(11):3345–3356, 2015.

[36] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing*, 21(12):4695–4708, 2012.

[37] Roozbeh Mottaghi, Xianjie Chen, Xiaobai Liu, Nam-Gyu Cho, Seong-Whan Lee, Sanja Fidler, Raquel Urtasun, and Alan Yuille. The role of context for object detection and semantic segmentation in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 891–898, 2014.

[38] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Icml*, 2010.

[39] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *arXiv preprint arXiv:1912.01703*, 2019.

[40] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

[41] Wenqi Ren, Sifei Liu, Lin Ma, Qianqian Xu, Xiangyu Xu, Xiaochun Cao, Junping Du, and Ming-Hsuan Yang. Low-light image enhancement via a deep hybrid network. *IEEE Transactions on Image Processing*, 28(9):4364–4375, 2019.

[42] 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*, pages 234–241. Springer, 2015.

[43] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016.

[44] Qing Wang and Rabab K Ward. Fast image/video contrast enhancement based on weighted thresholded histogram equalization. *IEEE transactions on Consumer Electronics*, 53(2):757–764, 2007.

[45] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6849–6857, 2019.

[46] Sicheng Wang, Bihan Wen, Junru Wu, Dacheng Tao, and Zhangyang Wang. Segmentation-aware image denoising without knowing true segmentation. *arXiv preprint arXiv:1905.08965*, 2019.

[47] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, 22(9):3538–3548, 2013.

[48] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018.

[49] Fisher Yu, Wenqi Xian, Yingying Chen, Fangchen Liu, Mike Liao, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving video database with scalable annotation tooling. *arXiv preprint arXiv:1805.04687*, 2(5):6, 2018.

[50] Lu Yuan and Jian Sun. Automatic exposure correction of consumer photographs. In *European Conference on Computer Vision*, pages 771–785. Springer, 2012.

[51] Weixia Zhang, Kede Ma, Guangtao Zhai, and Xiaokang Yang. Uncertainty-aware blind image quality assessment in the laboratory and wild. *IEEE Transactions on Image Processing*, 30:3474–3486, 2021.

[52] 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*, pages 1632–1640, 2019.

[53] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017.

[54] Shen Zheng, Yuxiong Wu, Shiyu Jiang, Changjie Lu, and Gaurav Gupta. Deblur-yolo: Real-time object detection with efficient blind motion deblurring. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2021.

[55] Minfeng Zhu, Pingbo Pan, Wei Chen, and Yi Yang. Eemefn: Low-light image enhancement via edge-enhanced multi-exposure fusion network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13106–13113, 2020.