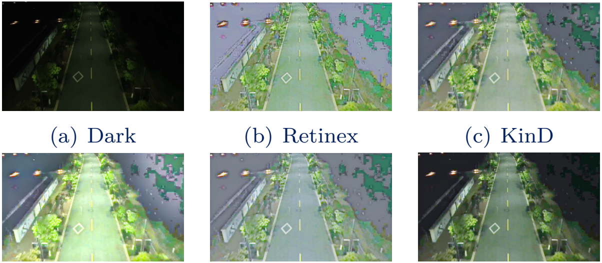
Semantic-Guided Zero-Shot Learning for Low-Light Image/Video Enhancement

Shen Zheng and Gaurav Gupta

College of Science and Technology, Wenzhou-Kean University

Outline

Challenges



(a) Dark (b) Retinex (c) KinD

(d) EnlightenGAN (e) Zero-DCE (f) Ours

Figure 1: Sample Enhancement Result on a nighttime aerial video frame

Challenges

- Degraded Feature and Contrast
- Poor Low-Level Perceptual Quality
- Substandard Performances for High-Level Tasks

HOW TO ADDRESS THOSE CHALLENGES?

Existing Solutions: Camera and Software

Camera

- Higher ISO
- Longer Exposure Time
- Noise ! Motion Blur !

Software

- Photoshop, Lightroom
- High Artist Skills !
- Inefficient at Diverse Illumination !

Existing Solutions: Conventional Methods

Histogram Equalization

- Adjusting Photo Intensities
- BPDHE [Ibrahim and Kong, 2007]
- WTHE [Wang and Ward, 2007]

Retinex Theory

- Reflectance and Illumination
- NPE [Wang et al., 2013]
- PIE [Fu et al., 2015], SRIE [Fu et al., 2016]
- LIME [Guo et al., 2016]

Existing Solutions: DL Methods

Supervised

- LLNet [Lore et al., 2017]
- RetinexNet [Wei et al., 2018], KinD [Zhang et al., 2019]
- DeepUPE [Wang et al., 2019]

Unsupervised

- EnlightenGAN [Jiang et al., 2021]

Zero-Shot

- Zero-DCE [Guo et al., 2020], Zero-DCE++ [Li et al., 2021]

Model Architecture

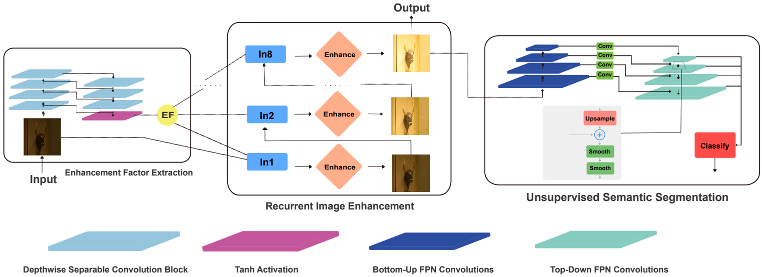


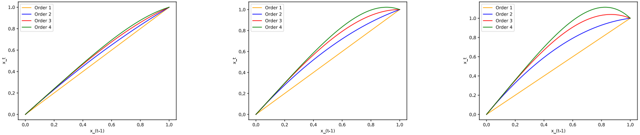
Figure 2: Proposed model architecture. Our model aims at enhancement factor extraction (EFE), recurrent image enhancement(RIE), and unsupervised semantic segmentation (USS)

Enhancement Factor Extraction

Figure 3: Enhancement factor visualization. Left: Low-Light Images. Right : Enhancement Factors. Darker region indicates lower values for the enhancement factor.

Recurrent Image Enhancement

$$x_t = x_{t-1} + x_r * (x_{t-1}^{\text{Order}} - x_{t-1}) \quad (1)$$



(a) $|x_r| = 0.2$ (b) $|x_r| = 0.5$ (c) $|x_r| = 0.8$

Figure 4: Recurrent image enhancement illustration with different enhancement factor x_r and different Order. Greater $|x_r|$ indicates a more intense enhancement.

Loss Function

- Spatial consistency loss

$$L_{spa} = \frac{1}{A}\sum_{i=1}^{A}[\sum_{j\in\phi(i)} \left(\left|(Y_i - Y_j)\right| - \left|(I_i - I_j)\right|\right)^2 + \alpha * \sum_{k\in\psi(i)} (|(Y_i - Y_k)| - |(I_i - I_k)|)^2] \tag{2}$$

- RGB loss

$$L_{rgb} = \sum_{\forall(i,j)\in\zeta} \sqrt{\left(\left(Y^i\right) - \left(Y^j\right)\right)^2 + \varepsilon^2}, \quad \zeta = \{(R,G),(R,B),(G,B)\} \tag{3}$$

Loss Function

- Brightness loss

$$L_{bri} = \frac{1}{A} \sum_{a=1}^{A} |Y_a - E| \tag{4}$$

- TV loss

$$L_{tv} = \frac{1}{CHW} \sum_{c=1}^{C} \sum_{h=1}^{H} \sum_{w=1}^{W} \left[(\nabla_x Y_{c,h,w})^2 + (\nabla_y Y_{c,h,w})^2 \right] \tag{5}$$

- Semantic loss

$$L_{sem} = \frac{1}{HW} \sum_{1 \leq i \leq H, 1 \leq j \leq W} -\beta \left(1 - p_{i,j}\right)^{\gamma} \log p_{i,j} \tag{6}$$

Implementation Details

Training details

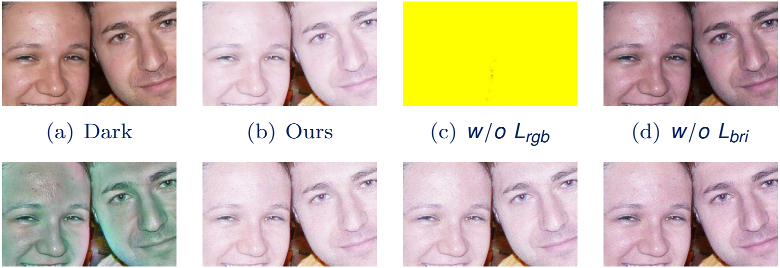
- Training Datasets: 2002 images, 512 X 512 resolutions
- Optimizer: Adam [Kingma and Ba, 2014]
- Framework: Pytorch [Paszke et al., 2019]
- Weight Initialization: N(0, 0.02)
- Batch Size: 6
- Epoch: 100
- Gradient Clipping: 0.1

Evaluation Dataset

Name	Number	Format	Type	Metric
NPE [Wang et al., 2013]	10	RGB	Real	N, B
LIME [Guo et al., 2016]	84	RGB	Real	N, B
MEF [Ma et al., 2015]	17	RGB	Real	N, B
DICM [Lee et al., 2012]	64	RGB	Real	N, B
VV	24	RGB	Real	N, B
LOL [Wei et al., 2018]	15	RGB	Real	P, S, M
DarkBDD	100	RGB	Real	N, B
DarkCityScape	150	RGB	Synthesis	P, S, M

Table 1: Dataset description. Where U, B stands for UNIQUE and BRISQUE, and P, S, M stands for PSNR, SSIM, MSE, respectively.

Ablation Study



(a) Dark
(b) Ours
(c) $w/o\ L_{rgb}$
(d) $w/o\ L_{bri}$
(e) $w/o\ L_{tv}$
(f) $w/o\ L_{spa}$
(g) $w/o\ L_{sem}$
(h) $w/o\ L_{spa}\ L_{sem}$

Figure 5: Visual comparison on loss function ablations

Quantitative Comparison: Efficiency

Method	RT↓	Params↓	FLOPs↓	Score↑
Retinex[Wei et al., 2018]	0.121	0.555	587.5	2.30
MBLLEN[Lv et al., 2018]	0.526	0.450	301.1	3.05
KinD[Zhang et al., 2019]	0.147	8.160	575.0	3.36
EnlightenGAN[Jiang et al., 2021]	0.008	8.637	273.2	2.94
Zero-DCE[Guo et al., 2020]	0.003	0.079	84.99	2.60
Ours	0.001	0.011	0.120	4.04

Table 2: Model Efficiency and User Study Score Comparison. We select an image of size 1200×900 for experiments. 'RT' is the inference time in seconds per image. 'Params' are the numbers of trainable parameters in millions per image, and 'FLOPs' are the numbers of floating-point operations in billions per image.

Quantitative Comparison: Others

Dataset	Dark	PIE	Retinex	MBLLEN	KinD	Zero-DCE	Ours
LOL	13.20/0.48	20.18/0.77	17.59/0.54	21.21/0.84	19.29/0.76	20.38/0.78	20.60/0.79
DCS	16.22/0.77	17.49/0.83	10.54/0.65	22.52/0.88	12.28/0.73	22.59/0.94	25.97/0.97

Table 3: PSNR ↑ / SSIM ↑ Comparison on Synthesized Low-Light Images

Method	NPE	LIME	MEF	DICM	VV	DarkBDD	Average
Dark	0.793 / 19.81	0.826 / 21.81	0.738 / 23.56	0.795 / 21.57	0.826 / 23.62	0.799 / 61.62	0.796 / 28.67
PIE	0.801 / 21.72	0.791 / 22.72	0.752 / 11.02	0.791 / 21.72	0.832 / 26.54	0.796 / 53.22	0.794 / 26.16
LIME	0.786 / 18.24	0.774 / 20.44	0.722 / 15.25	0.758 / 23.48	0.820 / 27.14	OME / OME	OME / OME
Retinex	0.828 / 16.04	0.794 / 31.47	0.755 / 20.08	0.770 / 29.53	0.824 / 29.58	0.792 / 50.77	0.794 / 29.57
MBLLEN	0.793 / 34.46	0.768 / 30.26	0.717 / 37.44	0.787 / 32.44	0.719 / 26.13	0.772 / 51.40	0.759 / 35.35
KinD	0.792 / 19.65	0.766 / 39.29	0.747 / 31.36	0.776 / 32.71	0.814 / 29.34	0.778 / 49.38	0.779 / 33.62
Zero-DCE	0.814 / 17.06	0.811 / 21.40	0.762 / 16.84	0.777 / 27.35	0.835 / 24.26	0.800 / 59.37	0.800 / 27.71
Ours	0.786 / 13.25	0.807 / 19.99	0.785 / 13.92	0.801 / 26.12	0.836 / 31.72	0.815 / 57.06	0.805 / 27.01

Table 4: UNIQUE ↑ / BRISQUE ↓ Comparison on Real-World Low-Light Images.

Qualitative Comparison on VV Dataset


(a) Dark


(b) PIE


(c) LIME


(d) Retinex


(e) MBLLEN

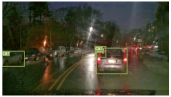

(f) KinD


(g) Zero-DCE


(h) Ours

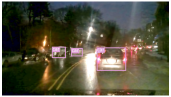
Low-Light Detection

(i) Dark

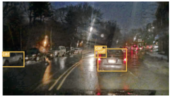


(j) PIE

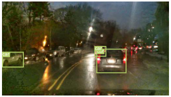
(k) Retinex



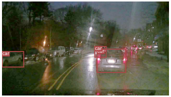
(l) MBLLEN



(m) KinD



(n) Enlighten-GAN

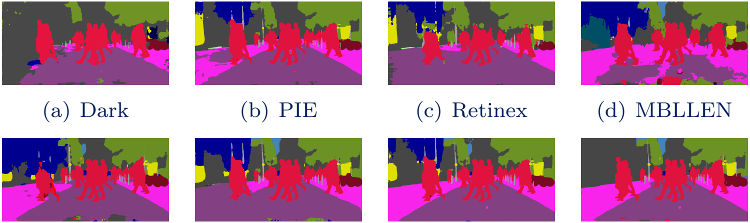


(o) Zero-DCE

(p) Ours

Figure 6: Object Detection Results on DarkBDD

Low-Light Segmentation



(a) Dark (b) PIE (c) Retinex (d) MBLLEN

(e) KinD (f) Zero-DCE (g) Ours (h) GT

Figure 7: Semantic Segmentation Results on DarkCityScape

Conclusion and Future Works

- Semantic-Guided, Zero-Shot LLIE Network
- EFE, RIE, USS with five non-reference loss functions
- No Paired Images, Unpaired Datasets, or Segmentation Labels
- Low-Level Enhancement and High-Level Semantics
- Efficient and Effective
- Future: Motion blur, Mirror Reflection

References I

Fu, X., Liao, Y., Zeng, D., Huang, Y., Zhang, X.-P., and Ding, X. (2015).
A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation.
IEEE Transactions on Image Processing, 24(12):4965–4977.

Fu, X., Zeng, D., Huang, Y., Zhang, X.-P., and Ding, X. (2016).
A weighted variational model for simultaneous reflectance and illumination estimation.
In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Guo, C., Li, C., Guo, J., Loy, C. C., Hou, J., Kwong, S., and Cong, R. (2020).
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.

References II

Guo, X., Li, Y., and Ling, H. (2016).
Lime: Low-light image enhancement via illumination map estimation.
IEEE Transactions on image processing, 26(2):982–993.

Ibrahim, H. and Kong, N. S. P. (2007).
Brightness preserving dynamic histogram equalization for image contrast enhancement.
IEEE Transactions on Consumer Electronics, 53(4):1752–1758.

Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., and Wang, Z. (2021).
Enlightengan: Deep light enhancement without paired supervision.
IEEE Transactions on Image Processing, 30:2340–2349.

Kingma, D. P. and Ba, J. (2014).
Adam: A method for stochastic optimization.
arXiv preprint arXiv:1412.6980.

References III

Lee, C., Lee, C., and Kim, C.-S. (2012).
Contrast enhancement based on layered difference representation.
In 2012 19th IEEE International Conference on Image Processing, pages 965–968. IEEE.

Li, C., Guo, C., and Loy, C. C. (2021).
Learning to enhance low-light image via zero-reference deep curve estimation.
arXiv preprint arXiv:2103.00860.

Lore, K. G., Akintayo, A., and Sarkar, S. (2017).
Llnet: A deep autoencoder approach to natural low-light image enhancement.
Pattern Recognition, 61:650–662.

Lv, F., Lu, F., Wu, J., and Lim, C. (2018).
Mbllen: Low-light image/video enhancement using cnns.
In BMVC, page 220.

References IV

Ma, K., Zeng, K., and Wang, Z. (2015).
Perceptual quality assessment for multi-exposure image fusion.
IEEE Transactions on Image Processing, 24(11):3345–3356.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019).
Pytorch: An imperative style, high-performance deep learning library.
arXiv preprint arXiv:1912.01703.

Wang, Q. and Ward, R. K. (2007).
Fast image/video contrast enhancement based on weighted thresholded histogram equalization.
IEEE transactions on Consumer Electronics, 53(2):757–764.

References V

Wang, R., Zhang, Q., Fu, C.-W., Shen, X., Zheng, W.-S., and Jia, J. (2019).
Underexposed photo enhancement using deep illumination estimation.
In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 6849–6857.

Wang, S., Zheng, J., Hu, H.-M., and Li, B. (2013).
Naturalness preserved enhancement algorithm for non-uniform illumination images.
IEEE Transactions on Image Processing, 22(9):3538–3548.

Wei, C., Wang, W., Yang, W., and Liu, J. (2018).
Deep retinex decomposition for low-light enhancement.
arXiv preprint arXiv:1808.04560.

Zhang, Y., Zhang, J., and Guo, X. (2019).
Kindling the darkness: A practical low-light image enhancer.
In Proceedings of the 27th ACM International Conference on Multimedia, pages 1632–1640.