Light the Night: A Multi-Condition Diffusion Framework for Unpaired Low-Light Enhancement in Autonomous Driving –Supplementary Material–

Jinlong Li
Cleveland State University
Cleveland, Ohio, USA
lijinlong1117@gmail.com

Jinlong Li
Cleveland State University
Cleveland, Ohio, USA
lijinlong1117@gmail.com

Jinlong Li
Cleveland State University
Cleveland, Ohio, USA
lijinlong1117@gmail.com

Jinlong Li
Cleveland State University
Cleveland, Ohio, USA
lijinlong1117@gmail.com

Jinlong Li
Cleveland State University
Cleveland, Ohio, USA
lijinlong1117@gmail.com

Second Author
Institution2
First line of institution2 address
secondauthor@i2.org

Within this supplementary section, we present additional insights through extensive comparisons of our LightDiff with various state-of-the-art (SOTA) methods for 3D detection tasks and provide visualizations showcasing the outcomes.

1. More 3D perception comparisons

Here we compare our LightDiff with more SOTA methods mentioned in our paper on the nuScenes nighttime validation set, as shown in Table 1. The SOTA methods include all the supervised methods mentioned in our paper (Afifi et al., Zhao et al., URetinexNet, SNR-aware) with all of their released pre-trained models, all the unsupervised methods mentioned in our paper (ExCNet, SCI, Zero-DCE, ZeroDCE++, EnlightenGAN, CLIP_LIT) as well as the version of their released pre-trained models), and some selected methods (SCI, EnlightenGAN, CLIP_LIT) retrained on the nuScenes training set. Notably, applying most image enhancement methods directly to preprocess images for 3D perception tasks is ineffective. In other words, using image enhancement methods to pre-process images may cause target inconsistency (human vision v.s. machine vision), since image restoration goals, such as image quality (*i.e.*, PSNR, SSIM), may conflict with the accuracy objectives of machine vision in detection tasks.

2. More visual comparisons

Detailed visual comparisons, as depicted in Figures 1, 2, 3, 4, and 5, showcase the prowess of LightDiff against **other 18 compared methods**. LightDiff adeptly enhances nighttime images, preserving intricate texture details and luminance without succumbing to issues such as over/under-exposure. Specific methods, namely SCI, EnlightenGAN, and CLIP_LIT, when retrained on the nuScenes training set, exhibit suboptimal performance compared to their original models, possibly due to global enhancement disturbances triggered by excessive darkness, imbalanced training data (daytime images are 26K and nighttime images are 6K), and dynamic noise inherent in real driving scenarios.

3. Challenging dynamic driving scenarios

The nuScenes dataset, gathered through autonomous vehicles navigating real, complex environments, stands out as one of the most comprehensive autonomous driving datasets. Notably, it lacks paired daytime-nighttime data specifically tailored for low-light enhancement tasks. In contrast to static datasets for low-light enhancement, datasets derived from autonomous driving scenarios capture scenes that are not only more variable but also more challenging. For example, Figs 1, 2, 3, 4 and 5 show all comparisons are ineffective on nuScenes nighttime validation set intuitively.

4. Training data limitation

Throughout the training process, we exclusively utilize the nuScenes daytime training set, omitting the nighttime counterpart. This strategic decision, while posing challenges due to the unknown lighting distribution in genuine nighttime images, demonstrates the resilience of our model. Despite never encountering authentic nighttime scenes like some enhancement methods, our model consistently achieves commendable enhancement results. It is noteworthy that the generated images by our LightDiff model exhibit a style consistent with the real nuScenes daytime dataset. Conversely, other comparison methods, trained across diverse brightness levels on other different style enhancement datasets, exhibit a more varied color palette.

Table 1. 3D detection performance on the nuScenes nighttime validation set comparing with more SOTA methods. Both BEVDepth and BEVStereo are trained using the nuScenes daytime training set. The best and second performance are marked in **red** and **blue**. * indicates that it has been retrained on the nuScenes training set.

Methods	BEVDepth				BEVStereo			
	AP↑	ATE↓	ASE↓	AOE↓	AP↑	ATE↓	ASE↓	AOE↓
Night Images	0.134	0.787	0.195	**0.957**	0.124	0.746	**0.205**	0.714
Afifi et al.	0.082	0.836	**0.172**	1.203	0.083	0.801	0.214	0.794
URetinex-Net	0.053	0.831	0.184	1.114	0.035	0.782	0.243	0.803
SNR-Aware-LOLv1	0.053	0.821	0.186	1.130	0.035	0.772	0.239	0.801
SNR-Aware-LOLv2real	0.053	0.821	0.186	1.130	0.035	0.771	0.239	0.800
SNR-Aware-LOLv2synthetic	0.084	0.800	0.190	1.036	0.061	0.787	0.227	0.875
Zero-DCE	0.077	0.832	0.190	1.087	0.053	0.761	0.225	0.864
Zero-DCE++	0.089	0.826	0.197	1.029	0.077	0.780	0.224	0.787
RUAS-LOL	0.039	0.817	0.179	1.101	0.027	0.757	0.239	0.752
RUAS-MIT5K	0.044	0.812	0.179	1.119	0.027	0.772	0.241	0.744
RUAS-DarkFace	0.039	0.816	**0.178**	1.102	0.027	0.758	0.239	0.757
SCI-easy	0.085	0.807	0.193	1.110	0.055	0.778	0.232	0.947
SCI-medium	0.042	0.812	0.180	1.107	0.027	0.765	0.241	0.741
SCI-difficult	0.067	0.828	0.187	1.071	0.032	0.764	0.239	0.774
EnlightenGAN	0.067	0.838	0.188	1.146	0.040	0.770	0.239	0.799
LESNet	0.111	0.823	0.183	1.209	0.111	0.798	**0.204**	0.710
CLIP_LIT	0.078	0.822	0.185	1.060	0.049	0.785	0.227	0.821
SCI*	0.062	0.829	0.181	1.133	0.034	0.783	0.235	0.801
EnlightenGAN*	**0.138**	**0.786**	0.193	**0.948**	**0.128**	**0.743**	**0.204**	**0.680**
CLIP_LIT*	0.131	0.791	0.199	0.972	0.121	0.753	0.211	0.739
Ours	**0.176**	**0.774**	0.180	1.108	**0.170**	**0.690**	0.210	**0.620**

(a) Night Image **(b)** Afifi et al. **(c)** URetinex-Net **(d)** SNR-Aware-LOLv1

(e) SNR-Aware-LOLv2real **(f)** SNR-Aware-LOLv2synthetic **(g)** Zero-DCE **(h)** Zero-DCE++

(i) RUAS-LOL **(j)** RUAS-MIT5K **(k)** RUAS-DarkFace **(l)** SCI-easy

(m) SCI-medium **(n)** SCI-difficult **(o)** EnlightenGAN **(p)** LESNet

(q) SCI-retrained **(r)** EnlightenGAN-retrained **(s)** CLIP-LIT-retrained **(t) Ours**

Figure 1. Complete comparisons with all methods on an example nighttime image of the nuScenes validation set. Our LightDiff excels in handling challenging dark regions, restoring clear texture details and satisfactory luminance.

(a) Night Image (b) Afifi et al. (c) URetinex-Net (d) SNR-Aware-LOLv1 (e) SNR-Aware-LOLv2real (f) SNR-Aware-LOLv2synthetic (g) Zero-DCE (h) Zero-DCE++ (i) RUAS-LOL (j) RUAS-MIT5K (k) RUAS-DarkFace (l) SCI-easy (m) SCI-medium (n) SCI-difficult (o) EnlightenGAN (p) LESNet (q) SCI-retrained (r) EnlightenGAN-retrained (s) CLIP-LIT-retrained **(t) Ours**

Figure 2. Complete comparisons with all methods on an example nighttime image of the nuScenes validation set. Our LightDiff excels in handling challenging dark regions, restoring clear texture details and satisfactory luminance.

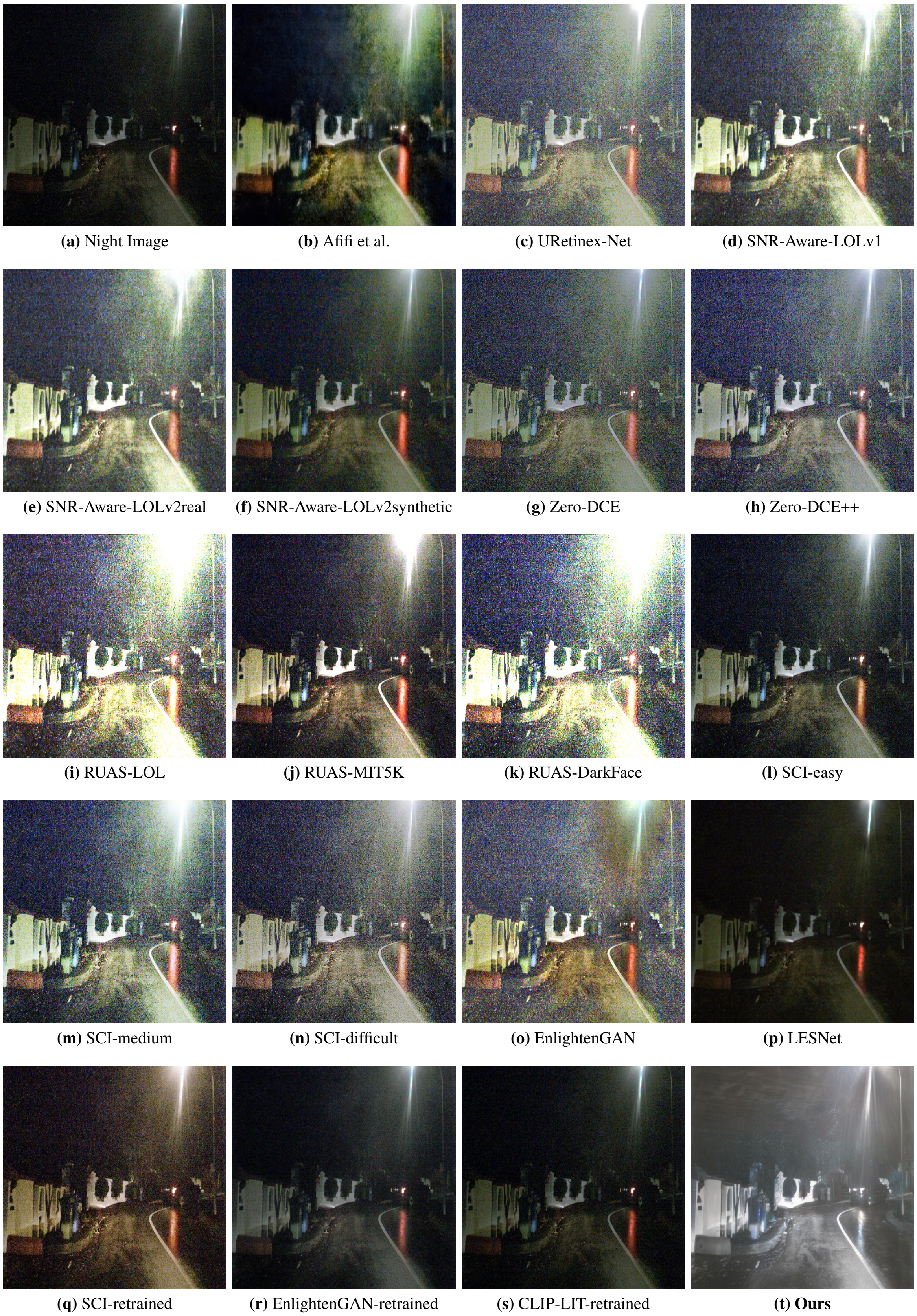
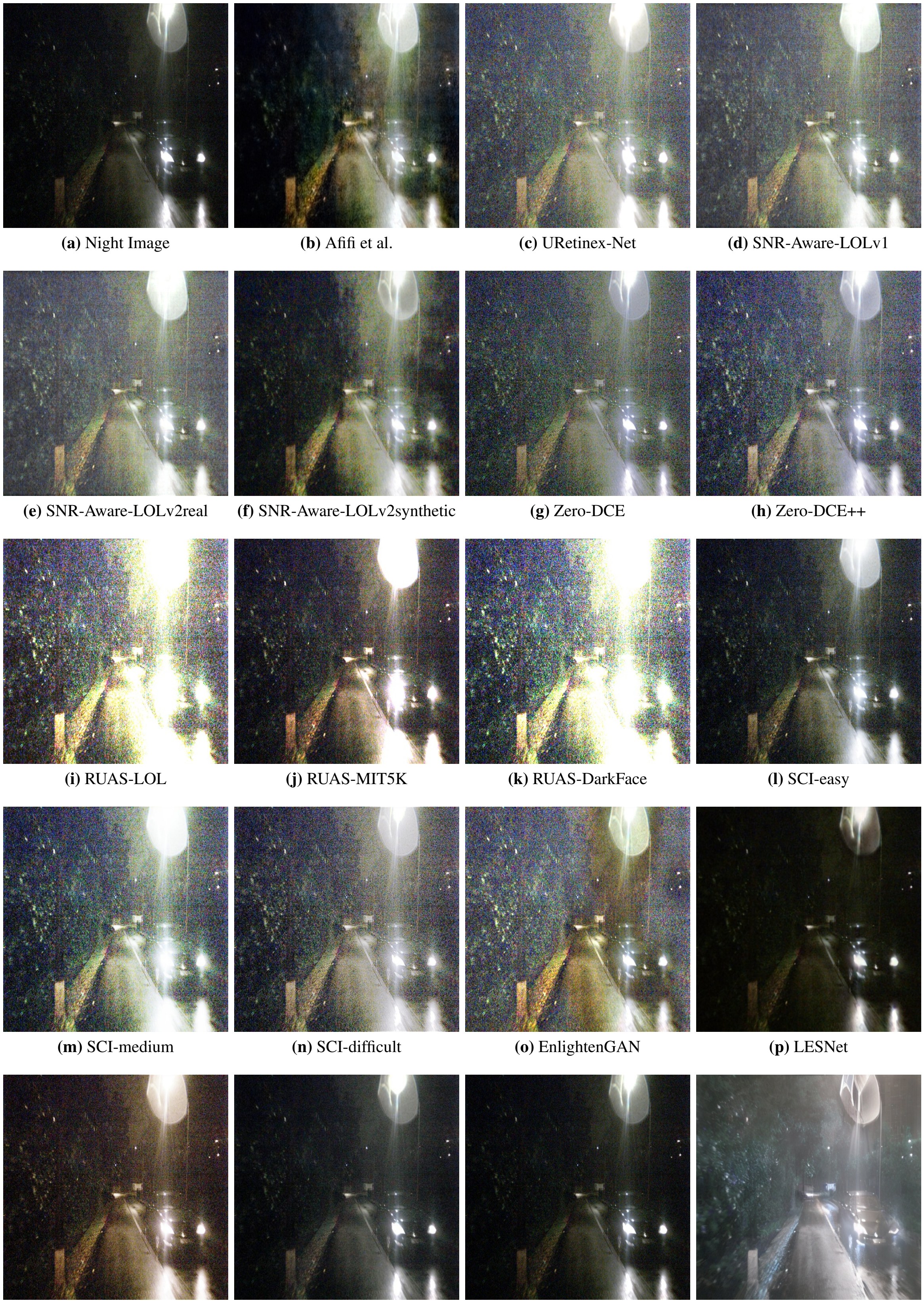


Figure 3. Complete comparisons with all methods on an example nighttime image of the nuScenes validation set. Our LightDiff excels in handling challenging dark regions, restoring clear texture details and satisfactory luminance.



(a) Night Image **(b)** Afifi et al. **(c)** URetinex-Net **(d)** SNR-Aware-LOLv1

(e) SNR-Aware-LOLv2real **(f)** SNR-Aware-LOLv2synthetic **(g)** Zero-DCE **(h)** Zero-DCE++

(i) RUAS-LOL **(j)** RUAS-MIT5K **(k)** RUAS-DarkFace **(l)** SCI-easy

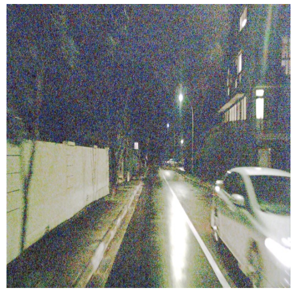
(m) SCI-medium **(n)** SCI-difficult **(o)** EnlightenGAN **(p)** LESNet

(q) SCI-retrained **(r)** EnlightenGAN-retrained **(s)** CLIP-LIT-retrained **(t) Ours**

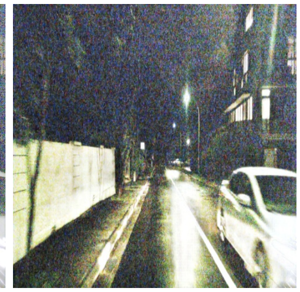
Figure 4. Complete comparisons with all methods on an example nighttime image of the nuScenes validation set. Our LightDiff excels in handling challenging dark regions, restoring clear texture details and satisfactory luminance.

(a) Night Image

(b) Afifi et al.



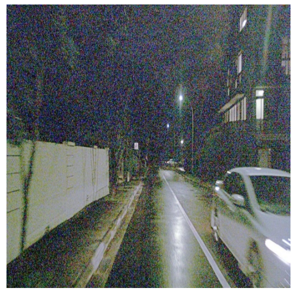
(c) URetinex-Net



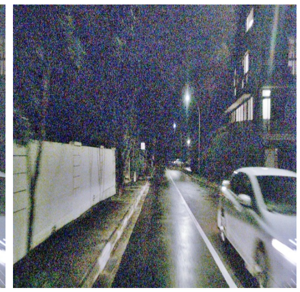
(d) SNR-Aware-LOLv1

(e) SNR-Aware-LOLv2real

(f) SNR-Aware-LOLv2synthetic

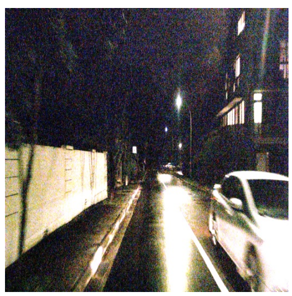


(g) Zero-DCE



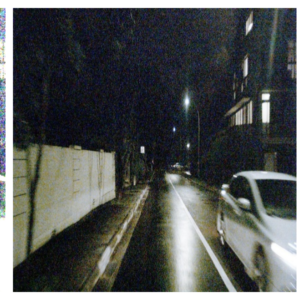
(h) Zero-DCE++

(i) RUAS-LOL

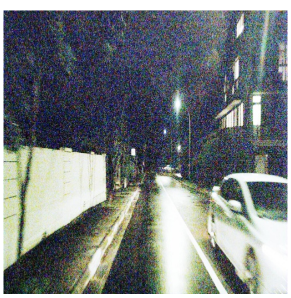


(j) RUAS-MIT5K

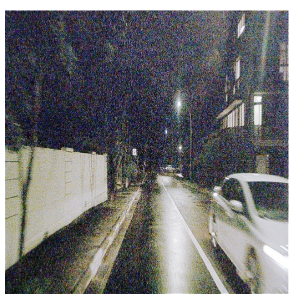
(k) RUAS-DarkFace



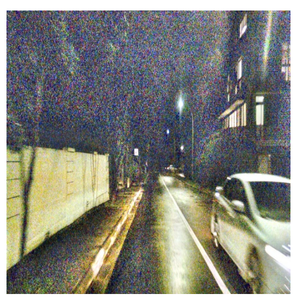
(l) SCI-easy



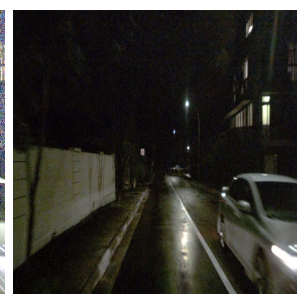
(m) SCI-medium



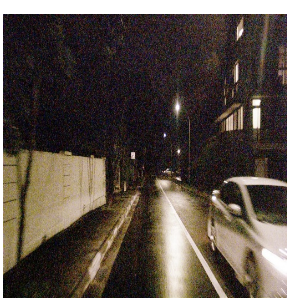
(n) SCI-difficult



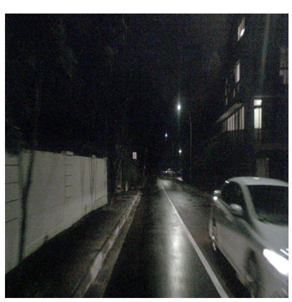
(o) EnlightenGAN



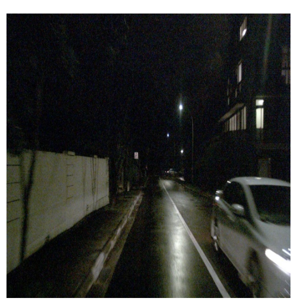
(p) LESNet



(q) SCI-retrained



(r) EnlightenGAN-retrained



(s) CLIP-LIT-retrained

(t) Ours

Figure 5. Complete comparisons with all methods on an example nighttime image of the nuScenes validation set. Our LightDiff excels in handling challenging dark regions, restoring clear texture details and satisfactory luminance.