CWNet: Causal Wavelet Network for Low-Light Image Enhancement

Supplementary Material

A. Network Structure

A.1. Detailed Structure of Feature Extraction (FE)

In the FE module, we utilize WTConv [6], H-Conv, V-Conv, and D-Conv. Each component is described in detail below:

WTConv [6] employs a convolution kernel of size 5 and is configured with 3 levels of wavelet downsampling to progressively reduce spatial resolution while preserving frequency information.

The **H-Conv, V-Conv, and D-Conv** modules utilize specifically designed convolution kernels to capture directional features in the input data. These kernels are as follows:

- Horizontal Convolution (H-Conv):

$$\text{Horizontal Kernel:} \quad \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$$

  This kernel is designed to emphasize horizontal edges in the input data.

- Vertical Convolution (V-Conv):

$$\text{Vertical Kernel:} \quad \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

  This kernel is designed to capture vertical edge features.

- Diagonal Convolution (D-Conv):

$$\text{Diagonal Kernel:} \quad \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

  This kernel is designed to detect diagonal structures.

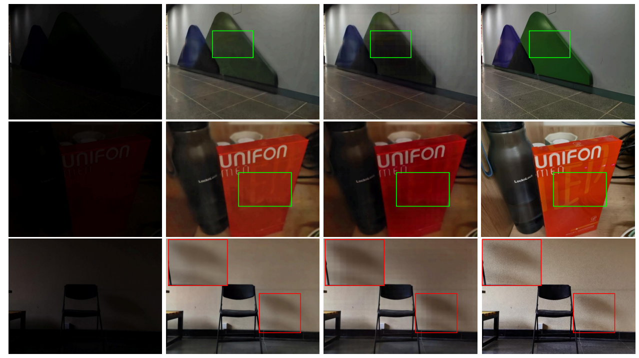

Figure 9. Visual comparison of results with and without the causal inference. Focus on the highlighted regions to observe differences in color and brightness, demonstrating the module's effectiveness.

Methods	DICM	LIME	MEF	NPE	VV	AVG
Kind [57]	**3.61**	4.77	4.82	4.18	3.84	4.24
MIRNet [52]	4.04	6.45	5.50	5.24	4.74	5.19
SGM [51]	4.73	5.45	5.75	5.21	4.88	5.21
FECNet [14]	4.14	6.04	4.71	4.50	3.75	4.55
HDMNet [24]	4.77	6.40	5.99	5.11	4.46	5.35
Bread [10]	4.18	4.72	5.37	4.16	**3.30**	4.35
Retinexformer [3]	4.01	**3.44**	<u>3.73</u>	<u>3.89</u>	<u>3.71</u>	<u>3.76</u>
UHDFormer [39]	4.42	4.35	4.74	4.40	4.28	4.44
Wave-Mamba [61]	4.56	4.45	4.76	4.54	4.71	4.60
CWNet	<u>3.92</u>	<u>3.58</u>	**3.66**	**3.61**	3.74	**3.70**

Table 4. NIQE scores on DICM, LIME, MEF, NPE, and VV datasets. Lower NIQE scores indicate better perceptual quality. The **best** and <u>second-best</u> results in each column are shown in bold and underlined respectively. "AVG" denotes the average NIQE scores across the five datasets.

B. Experiments

B.1. Visualization Comparison on LOL-v2-Synthesis Dataset.

Fig.10 presents a comparison on the LOL-v2-Synthesis dataset between our CWNet and current state-of-the-art methods, including FECNet [14], FourLLIE [38], Retinexformer [3], UHDFour [19], UHDFormer [39], and Wave-Mamba [61].

In the first row, other methods exhibit stripe noise in the sky region. Specifically, FECNet shows severe hue distortion, while other methods lack image sharpness. In the second row, similar artifacts are observed, with stripe-like noise appearing in the background areas.

In the third and fifth rows, other methods exhibit color distortion when compared to the reference ground truth. In contrast, our results appear more natural and exhibit better visual quality. This demonstrates the effectiveness of our causal inference component in mitigating color distortion and preserving semantic structures by disentangling causal relationships in the feature space.

In the fourth and last rows, upon magnification, our results exhibit the most consistent and visually pleasing brightening effect. The comparison on the LOL-v2-Synthesis dataset further demonstrates that our CWNet achieves natural brightness restoration and excels in preserving semantic and color consistency.

B.2. Quantitative and Qualitative Comparison on the DICM, LIME, MEF, NPE, and VV Datasets

Quantitative Comparison. We evaluated CWNet on five independent datasets: DICM [18] (69 images), LIME [11] (10 images), MEF [27] (17 images), NPE [41] (85 images), and VV (24 images). The evaluation was conducted using the no-reference image quality assessment metric NIQE [28], with lower scores indicating better perceptual quality and naturalness. Tab. 4 presents the NIQE results. As shown, CWNet outperforms several state-of-the-art (SOTA) methods, including SKF-SNR, UHDFormer, and Wave-Mamba. To ensure a fair comparison, all methods were pretrained on the LSRW-Huawei dataset.

Qualitative Comparison. We provide qualitative comparisons on the DICM, LIME, MEF, NPE, and VV datasets to visually demonstrate the effectiveness of CWNet.

1) Fig.11 shows the qualitative comparison on the DICM dataset. In the first row, SKF-SNR produces distorted and overexposed results, which negatively impact the visual quality. UHDFormer generates relatively natural results but suffers from overexposure, particularly in the highlighted regions, where the flower colors are overly brightened. Wave-Mamba exhibits blurred edge details and insufficient exposure control. In contrast, CWNet produces clearer and more natural enhancement results. In the second row, CWNet produces sharper and more natural results, while SKF-SNR suffers from underexposure and noise artifacts. UHDFormer and Wave-Mamba produce relatively blurry results. In the third row, SKF-SNR introduces unavoidable noise, and both UHDFormer and Wave-Mamba generate unnatural sky colors. CWNet, however, produces results with consistent and natural sky colors, demonstrating its robustness in handling color distortions.

2) Fig.12 presents the qualitative comparison on the LIME dataset. SKF-SNR fails to achieve sufficient brightness enhancement, while UHDFormer and Wave-Mamba produce reasonable brightness but suffer from blurriness and lack of detail. In contrast, CWNet generates sharper textures and richer details, further validating the robustness of our high- and low-frequency modeling in capturing fine-grained details and global structures.

3) Fig.13 illustrates the qualitative comparison on the MEF dataset. SKF-SNR produces unnatural flame colors and insufficient brightness enhancement. UHDFormer and Wave-Mamba exhibit varying degrees of blurriness, while CWNet achieves the clearest and most visually pleasing enhancement results.

4) Fig.14 shows the qualitative comparison on the NPE dataset. Similar to the MEF dataset, SKF-SNR fails to provide sufficient brightness enhancement, and both UHDFormer and Wave-Mamba suffer from blurriness. CWNet, on the other hand, produces visually superior results with richer detail information.

Metrics	UHDFour [19]	UHDFormer [39]	Wave-Mamba [61]	CWNet
UICM ↑	0.7464	0.9571	0.9082	**0.9663**
NIQE ↓	5.385	5.564	5.689	**4.494**

Table 5. Visualization quality comparison on DarkFace. The **best** results in each column are shown in bold.

5) Fig.15 demonstrates the qualitative comparison on the VV dataset. SKF-SNR generates distorted enhancement results with significant noise artifacts. UHDFormer suffers from overexposure, as observed in the highlighted facial regions, while Wave-Mamba produces blurry results. In contrast, CWNet produces natural enhancements with well-preserved details.

B.3. Quantitative and Qualitative Comparison on the DarkFace Dataset

We conducted quantitative and qualitative comparisons against current state-of-the-art (SOTA) methods on the DarkFace dataset [50]. The DarkFace dataset, with 6,000 real-world low-light images, serves as a challenging benchmark for low-light image enhancement.

For quantitative evaluation, we employed two metrics: NIQE and the Underwater Image Colorfulness Measure (UICM)[34], which is commonly used to evaluate colorfulness and naturalness in enhanced images. Unlike NIQE, where lower scores indicate better perceptual quality, higher UICM scores reflect greater colorfulness and naturalness. As shown in Tab.5, our method achieves the best performance across both metrics, highlighting its effectiveness in low-light image enhancement.

Fig.17 illustrates the qualitative comparison on the DarkFace dataset. UHDFormer produces severe color distortions, particularly in bright regions, indicating its limited generalization ability on this dataset. In the first and last rows, the zoomed-in regions show that our method outperforms UHDFormer and Wave-Mamba in detail preservation and sharpness. Furthermore, in the second row, our method demonstrates superior exposure control, yielding natural and visually pleasing results. These observations validate the effectiveness of CWNet in low-frequency brightness control for natural exposure and high-frequency detail enhancement for sharper textures.

C. Ablation Study and Downstream Application

C.1. Ablation Study: Validating the Effectiveness of Core Modules

Fig.9 demonstrates the impact of causal inference on CWNet's performance. By focusing on the highlighted regions, it is evident that removing causal inference leads to

Number	PSNR↑	SSIM↑	LPIPS↓
L=1, C=1	21.06	0.6381	0.1772
L=2, C=2	**21.50**	0.6397	**0.1562**
L=3, C=3	21.34	**0.6401**	0.1587

Table 6. Ablation Study on Negative Sample Quantity in the Causal Inference Module. The first column represents the number of I_l and I_c. The **best** results in each column are shown in bold.

inconsistencies in brightness and semantic coherence. In contrast, incorporating causal inference ensures both brightness restoration and semantic consistency, effectively preserving color fidelity and structural details. This validates the module's ability to model causal relationships, enhancing overall enhancement quality.

C.2. Ablation Study on Negative Sample Quantity in the Causal Inference

We conducted an ablation study within the Causal Inference module to determine the optimal number of negative samples for brightness degradation I_l (simulating underexposed regions) and color anomaly I_c (introducing color distortions), as summarized in Tab.6. The results indicate that the optimal performance is achieved when the number of samples for both brightness degradation and color anomaly is set to 3.

C.3. Object Detection on DarkFace Dataset

To evaluate how enhanced images affect downstream tasks, we conducted object detection experiments on the DarkFace dataset [50]. We evaluated our method on 200 randomly selected images from the dataset using the official YOLOv5 model pretrained on the COCO dataset [25], as YOLOv5 is a widely used object detection framework and COCO provides a diverse set of object categories. Fig.18 presents the visual comparison results, showing that our method achieves superior detection accuracy compared to others.

In the first experiment, our method achieves higher confidence scores for detected pedestrians compared to other methods and uniquely detects the motorcycle on the right, likely due to its ability to enhance fine-grained details in low-light conditions. In the second experiment, our method detects more pedestrians with higher confidence and correctly identifies the traffic light, a task that is particularly challenging due to the small size and low contrast of the traffic light in the original image. In contrast, Wave-Mamba misclassifies building lights as traffic lights.

These results demonstrate that our enhancement method generates clearer images while preserving semantic structures effectively. By leveraging causal inference, CWNet emphasizes intrinsic image content, enhancing downstream task performance such as object detection.

C.4. Superior Edge Detection Performance

Fig. 16 compares edge detection performance across state-of-the-art methods, highlighting CWNet's ability to restore details more precisely, particularly in highlighted regions. This superior performance validates the effectiveness of CWNet's high-frequency module in preserving fine details and structural fidelity, further demonstrating the robustness of our architecture.

Figure 10. Visual comparison on LOL-v2-Synthesis dataset.

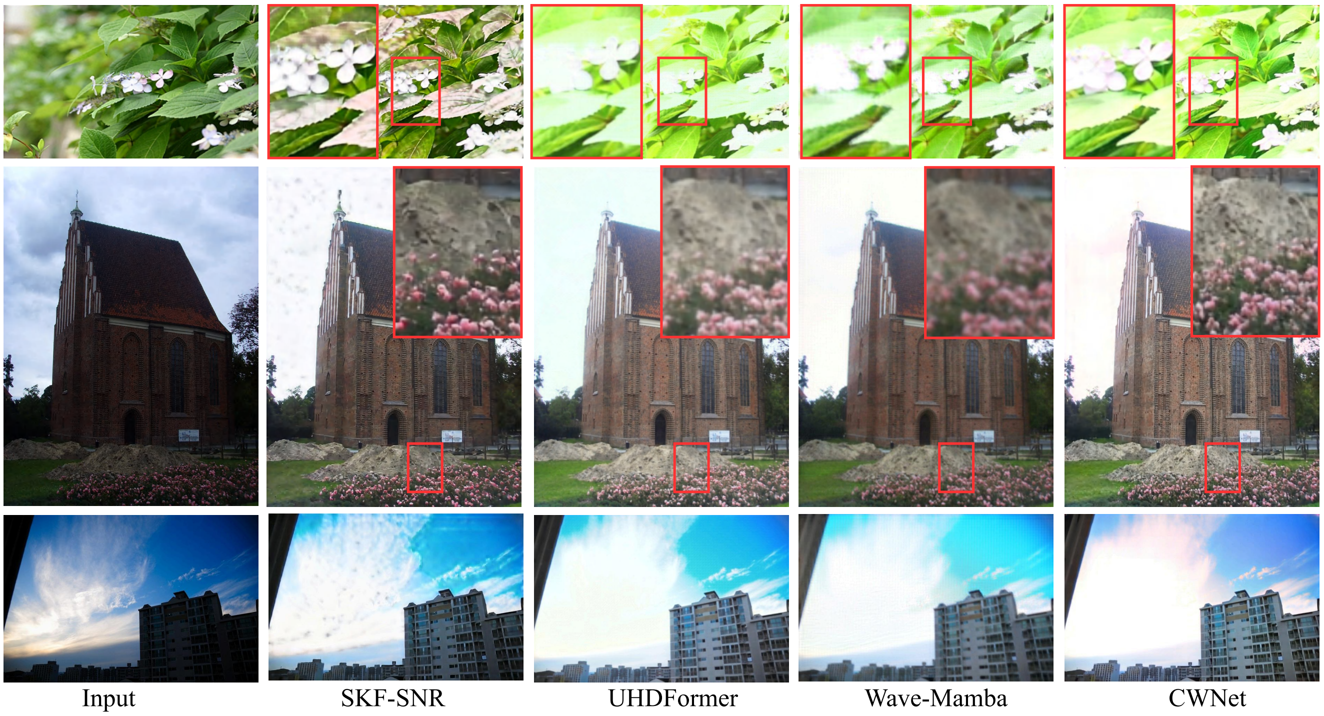

Figure 11. Visual comparison on DICM dataset.

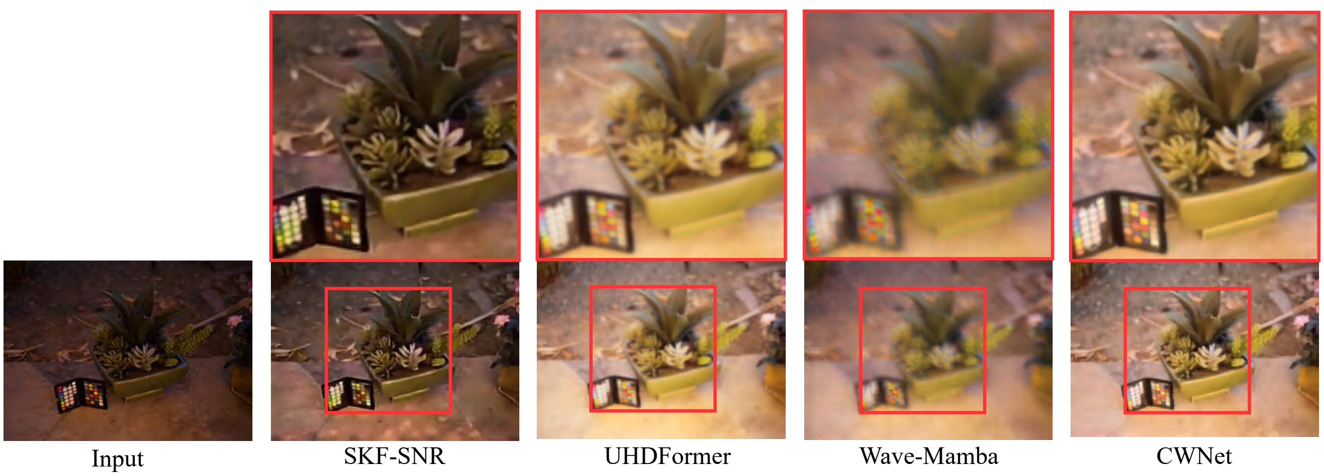

Figure 12. Visual comparison on LIME dataset.

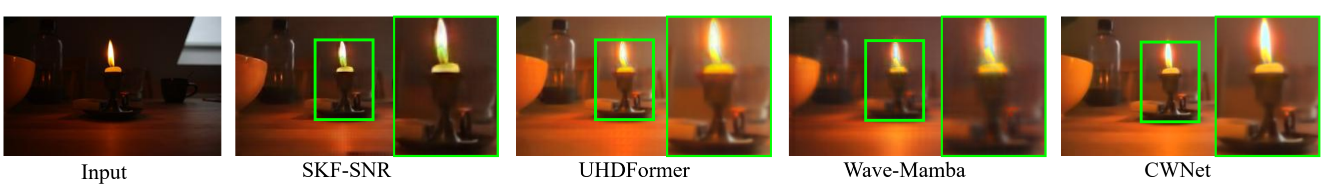

Figure 13. Visual comparison on MEF dataset.

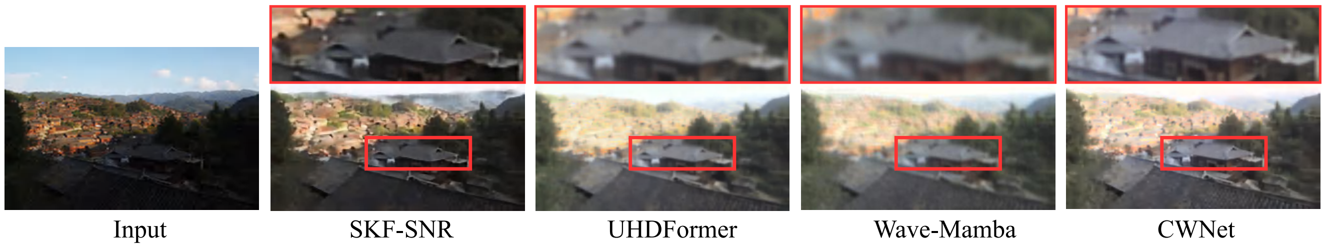

Figure 14. Visual comparison on NPE dataset.

Figure 15. Visual comparison on VV dataset.

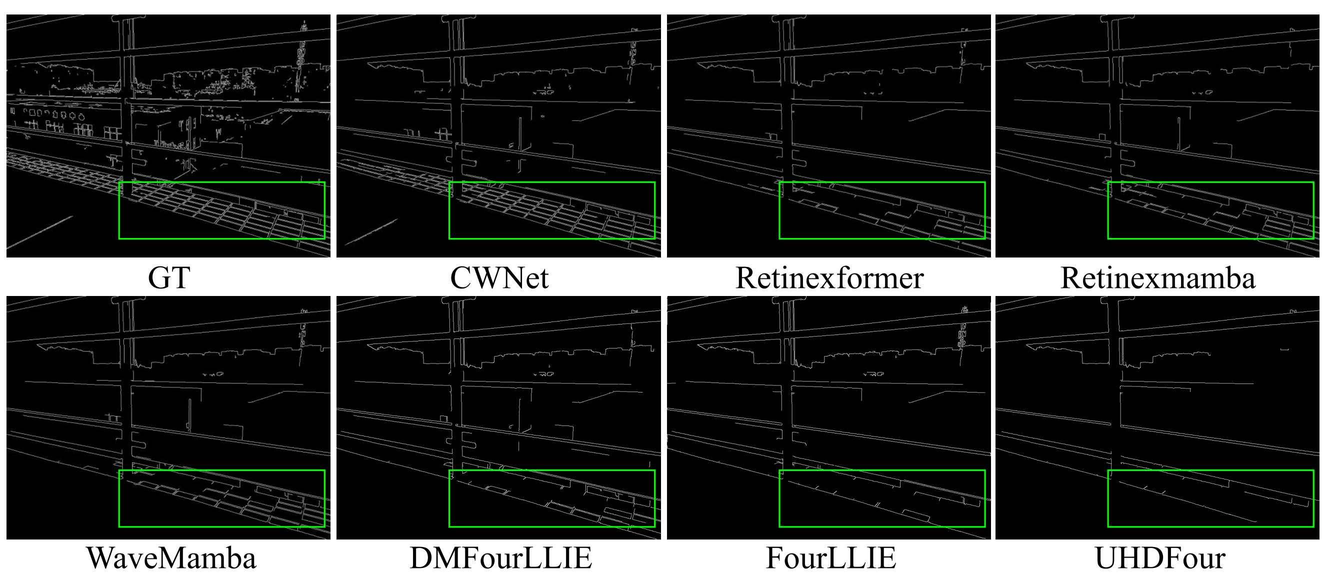

Figure 16. Edge detection comparison shows our method restores details more precisely, especially in highlighted regions, validating the high-frequency module.

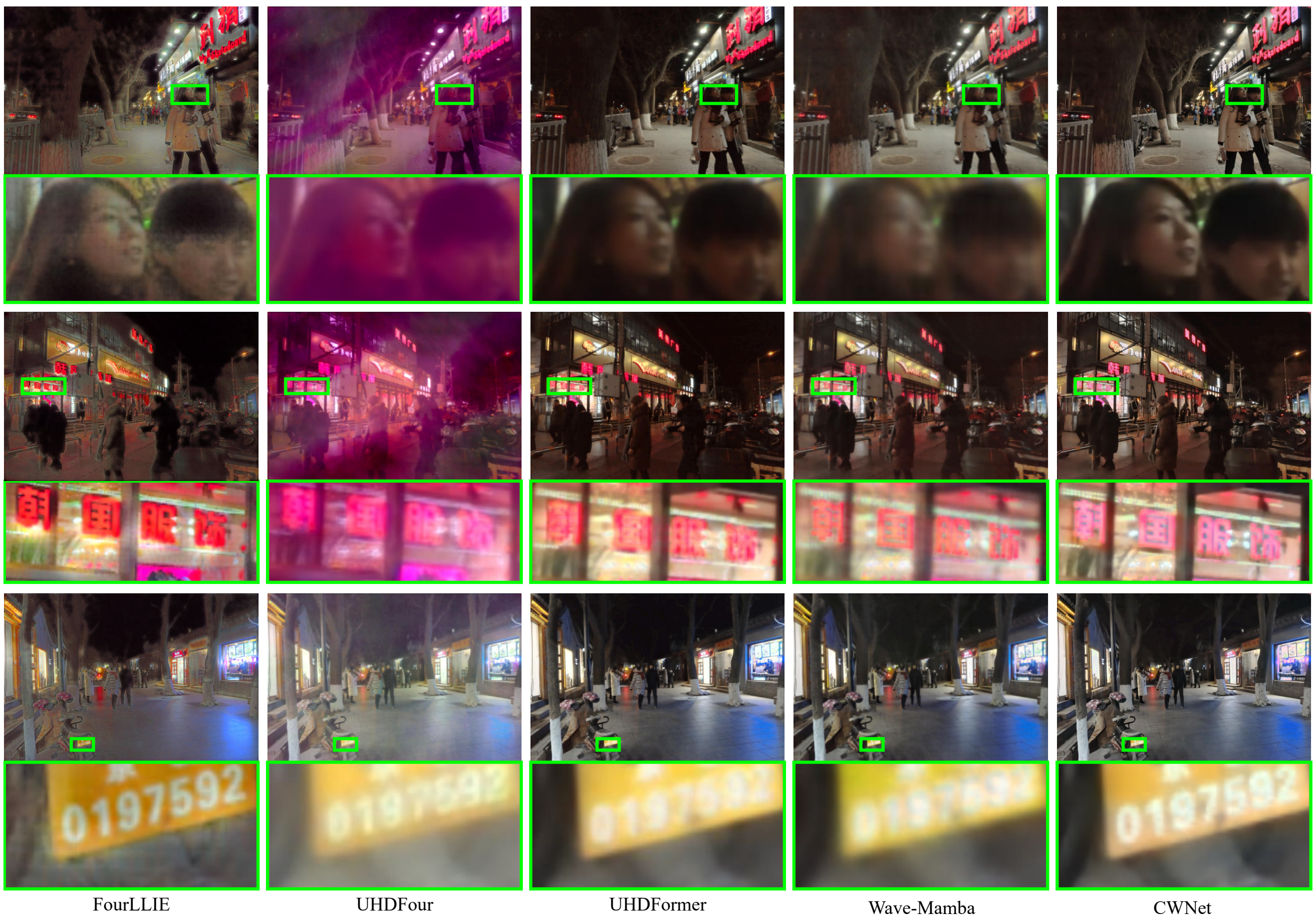

Figure 17. Visual comparison on DarkFace dataset.

Figure 18. Detection comparison results on DarkFace dataset.