

A EFFECT OF DENSE SUPERVISION ON LOCALIZATION ACCURACY

To analyze the effect of our self-guided framework on localization, we evaluate detection performance across multiple IoU thresholds, ranging from 0.5 to 0.95. As illustrated in Figure 1, the performance gap between our method and the baseline increases consistently as the IoU threshold becomes more strict. This trend suggests that our framework not only improves overall detection robustness under low-light conditions but also leads to more precise object localization.

We believe this improvement is largely due to the dense pixel-level supervision introduced during training through the fused target image. By complementing sparse bounding box annotations with structure-preserving guidance, the backbone learns more discriminative features that enable tighter and more accurate object boundaries. These results support our hypothesis that dense supervision encourages spatially coherent and localization-aware feature representations.

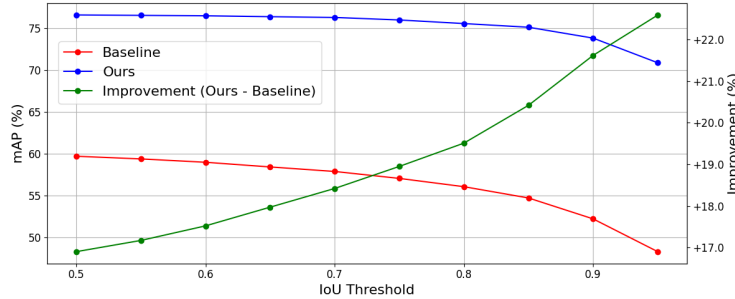


Figure 1: Detection mAP under different IoU thresholds on the DARK FACE dataset. As the IoU threshold increases from 0.5 to 0.95, the performance gap between our method and the baseline consistently widens, demonstrating improved localization accuracy enabled by our structure-aware supervision.

B PERFORMANCE VARIATION ACROSS LOW-LIGHT BENCHMARKS

Our framework consistently improves detection performance across all three low-light benchmarks: DARK FACE, ExDark, and nuImages. However, we observe that the magnitude of improvement varies across datasets, with the largest gain observed on DARK FACE. To better understand this discrepancy, we analyze the average relative brightness of each dataset, as shown in Figure 2. Among the three, DARK FACE contains the darkest scenes—nearly two times darker than ExDark and over five times darker than daytime nuImages. In such extreme low-light conditions, the pixel-level supervision provided by our framework becomes especially beneficial, as it effectively encourages the learning of structure-preserving features from low-contrast input image. In contrast, ExDark and nuImages (Night) contain relatively brighter images than DARK FACE, where standard detectors are already able to extract salient features. In these cases, while our framework still yields noticeable improvements, the relative gains are naturally smaller due to the less severe degradation in the input signal. This analysis highlights the strength of our approach under the most challenging illumination scenarios and explains the dataset-specific variation in performance improvements.

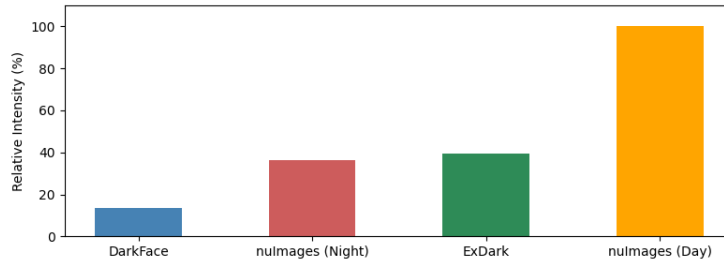


Figure 2: Relative brightness of each dataset, normalized to the mean intensity of nuImages (Day).

C ENSURING FAIR AND REPRODUCIBLE COMPARISON ON EXDARK

Splits and Pretrained Weights. For all experiments on the ExDark dataset, we strictly followed the evaluation protocol and dataset split introduced by MAET (Cui et al., 2021). To ensure a fair and reproducible comparison, we adopted the official weights released by the authors of each method. Specifically, the baseline YOLOv3 model as well as KinD (Zhang et al., 2019), Zero-DCE (Guo et al., 2020), and MAET (Cui et al., 2021) models were evaluated using the official pre-trained weights provided in the MAET GitHub repository. We independently verified all results by running inference using these weights, ensuring reproducibility under the same experimental conditions. For DAI-Net (Du et al., 2024), we used the weights publicly released by its authors, who trained the model using the same protocol as MAET. These weights are compatible with the same dataset split and detector architecture, ensuring a fair comparison.

Updated Results of MAET. Importantly, we note that the scores reported in the original MAET and DAI-Net papers for several baseline and enhancement methods (e.g., KinD, Zero-DCE) have since been revised by the MAET authors. The updated values, along with corresponding weights, are now available on the official MAET GitHub repository¹. In this work, we adopt these corrected scores to provide the most up-to-date and consistent comparison across all methods. For more recent enhancement methods that were not included in MAET or DAI-Net (e.g. SCI, Retinexformer) we conducted the evaluations ourselves using their official codebases. Each method was trained and tested under exactly the same conditions as MAET, using the YOLOv3-608 detector, COCO pre-training, and the same dataset split and training schedule. This ensures that all results are directly comparable and evaluated under a unified protocol.

Method	Bicycle	Boat	Bottle	Bus	Car	Cat	Chair	Cup	Dog	Motorbike	People	Table	mAP
Baseline (Redmon, 2018)	79.8	75.3	78.1	92.3	83.0	68.0	69.0	79.0	78.0	77.3	81.5	55.5	76.4
KinD (Zhang et al., 2019)	80.1	77.7	77.2	93.8	83.9	66.9	68.7	77.4	79.3	75.3	80.9	53.8	76.3
RUAS (Liu et al., 2021)	80.4	75.9	74.8	89.8	79.3	69.8	67.4	70.2	74.6	70.4	78.0	54.4	73.7
Zero-DCE (Guo et al., 2020)	84.1	77.6	<u>78.3</u>	93.1	83.7	70.3	69.8	77.6	77.4	76.3	81.0	53.6	76.9
Zero-DCE++ (Li et al., 2021)	79.4	76.0	74.8	91.6	83.3	69.0	68.0	72.3	78.6	75.2	78.7	56.4	75.3
SCI (Ma et al., 2022)	82.0	77.9	78.1	92.9	83.9	72.9	<u>71.8</u>	78.0	<u>81.0</u>	75.9	82.7	57.2	77.9
Retinexformer (Cai et al., 2023)	80.9	72.8	75.6	91.0	83.4	71.7	69.7	72.1	76.5	76.9	80.6	54.7	75.5
GLARE (Zhou et al., 2024)	79.3	73.3	74.1	90.5	83.5	68.2	69.3	75.1	78.0	74.0	79.2	55.2	75.0
T2 (Cui et al., 2024)	82.1	76.5	77.4	89.6	82.8	67.1	67.8	75.3	77.4	74.7	80.3	54.5	75.5
MAET (Cui et al., 2021)	83.1	<u>78.5</u>	75.6	92.9	83.1	<u>73.4</u>	71.3	79.0	79.8	77.2	81.1	57.0	77.7
DAI-Net (Du et al., 2024)	<u>83.8</u>	75.8	75.1	94.2	<u>84.1</u>	74.9	73.1	79.2	82.2	76.4	80.7	59.8	78.3
Ours	82.0	79.9	80.6	93.6	85.1	71.2	71.7	79.2	80.0	79.8	<u>82.4</u>	<u>57.8</u>	78.6

Table 1: Per-class detection mAP (%) comparison on the ExDark dataset using YOLOv3. All models were evaluated under the MAET protocol. Baseline, KinD, Zero-DCE, and MAET results were reproduced using official weights provided by the MAET authors. DAI-Net results are based on the weights released by its authors under the same protocol. Note that some of the reported values (e.g., Baseline, KinD, Zero-DCE) were updated by the MAET authors after publication, and our evaluation reflects these revised results. Our method achieves the highest overall mAP among all compared approaches.

¹MAET GitHub: https://github.com/cuiziteng/ICCV_MAET.



Figure 3: Additional Results on DARK FACE (Yang et al., 2020) dataset. *Zoom in for best view.*

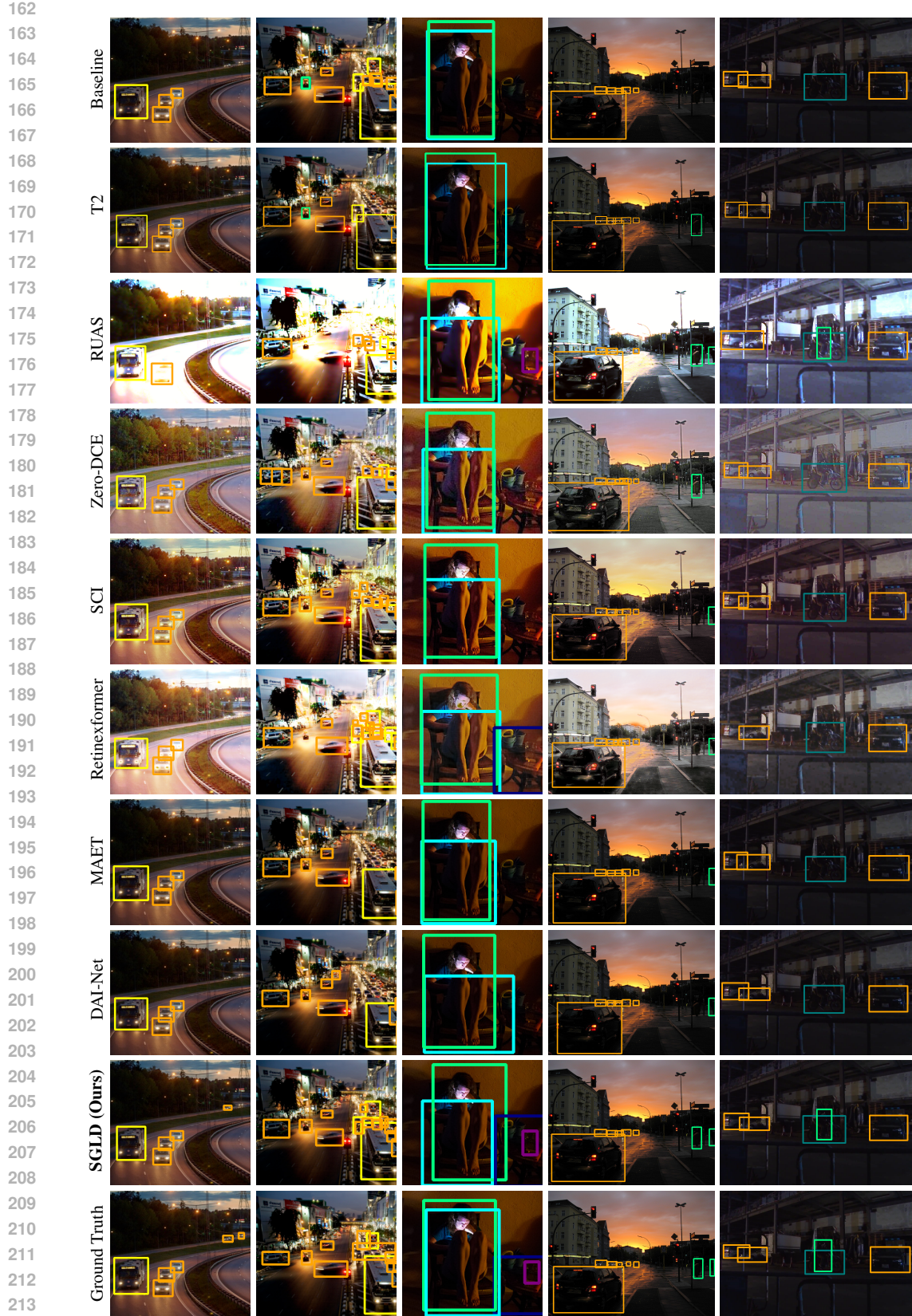


Figure 4: Additional Results on ExDark (Loh & Chan, 2019) dataset.

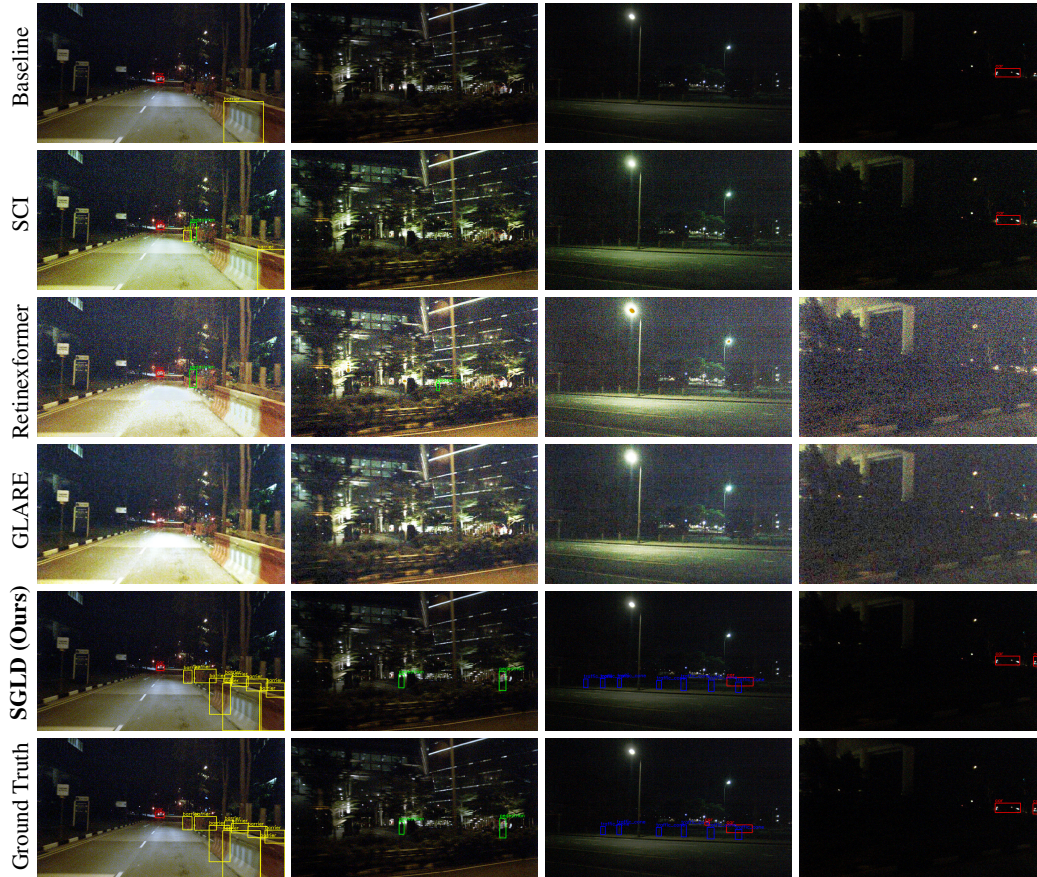


Figure 5: Additional qualitative comparison with enhancement-based method on nuImages (Caesar et al., 2020).

270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323

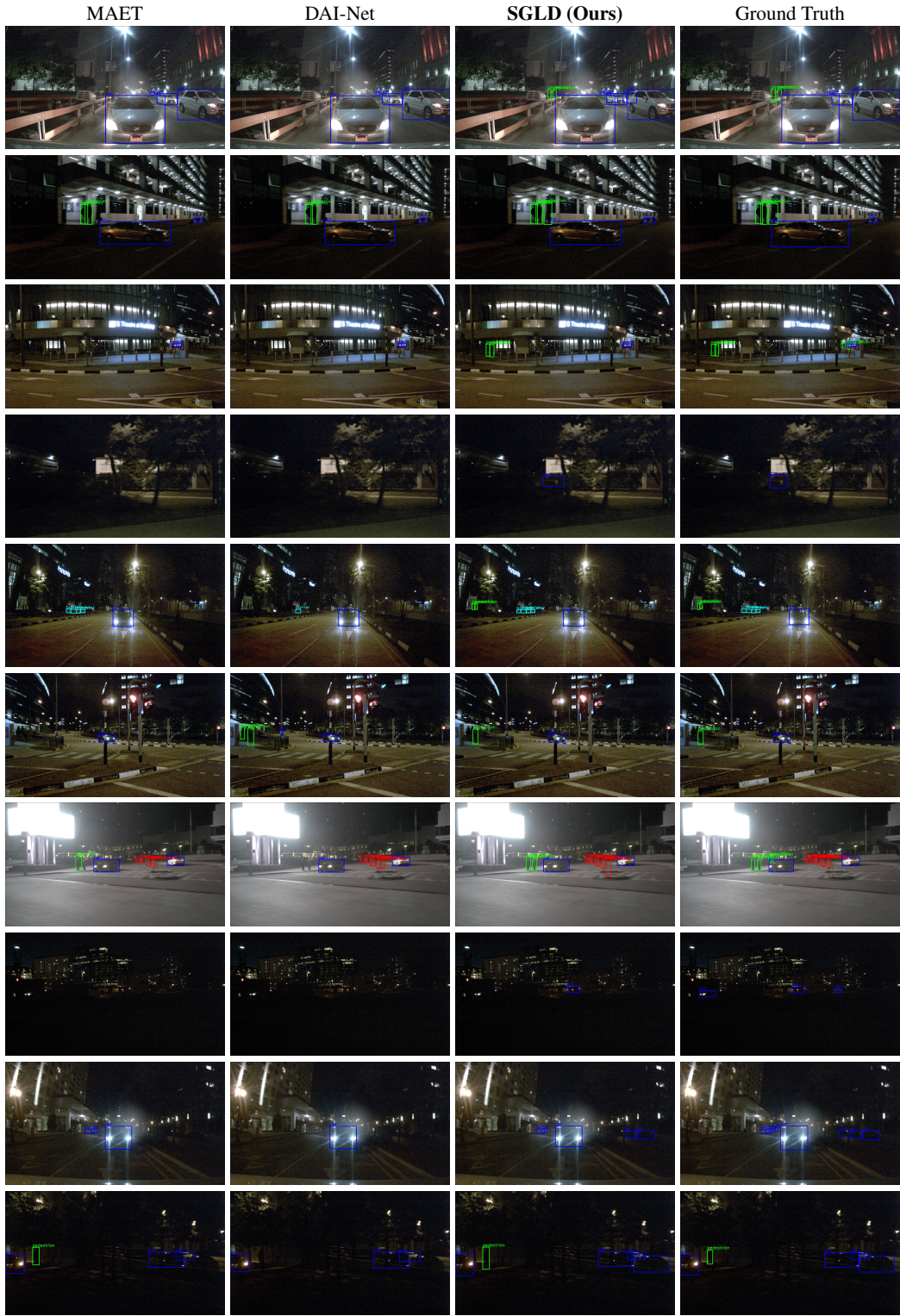


Figure 6: Additional qualitative comparison with domain adaptation method on nuImages (Caesar et al., 2020).

REFERENCES

- Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11621–11631, 2020.
- Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *ICCV*, 2023.
- Xiaohan Cui, Long Ma, Tengyu Ma, Jinyuan Liu, Xin Fan, and Risheng Liu. Trash to treasure: Low-light object detection via decomposition-and-aggregation. In *AAAI*, 2024.
- Ziteng Cui, Guo-Jun Qi, Lin Gu, Shaodi You, Zenghui Zhang, and Tatsuya Harada. Multitask act with orthogonal tangent regularity for dark object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 2553–2562, 2021.
- Zhipeng Du, Miaoqing Shi, and Jiankang Deng. Boosting object detection with zero-shot day-night domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12666–12676, 2024.
- 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*, pp. 1780–1789, 2020.
- Chongyi Li, Chunle Guo, and Chen Change Loy. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE transactions on pattern analysis and machine intelligence*, 44(8):4225–4238, 2021.
- Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, and Zhongxuan Luo. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10561–10570, 2021.
- 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.
- Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *CVPR*, 2022.
- Joseph Redmon. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.
- Wenhan Yang, Ye Yuan, Wenqi Ren, Jiaying Liu, Walter J Scheirer, Zhangyang Wang, Taiheng Zhang, Qiaoyong Zhong, Di Xie, Shiliang Pu, et al. Advancing image understanding in poor visibility environments: A collective benchmark study. *IEEE Transactions on Image Processing*, 29:5737–5752, 2020.
- 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*, pp. 1632–1640, 2019.
- Han Zhou, Wei Dong, Xiaohong Liu, Shuaicheng Liu, Xiongkuo Min, Guangtao Zhai, and Jun Chen. Glare: Low light image enhancement via generative latent feature based codebook retrieval. In *European Conference on Computer Vision*, pp. 36–54. Springer, 2024.