Learning Semantic Degradation-Aware Guidance for Recognition-Driven Unsupervised Low-Light Image Enhancement

Naishan Zheng*, **Jie Huang***, **Man Zhou**†, **Zizheng Yang**, **Qi Zhu**, **Feng Zhao**†

University of Science and Technology of China, Hefei, China

{nszheng, hj0117, manman, yzz6000, zqcrafts}@mail.ustc.edu.cn, fzhao956@ustc.edu.cn

Abstract

Low-light images suffer severe degradation of low lightness and noise corruption, causing unsatisfactory visual quality and visual recognition performance. To solve this problem while meeting the unavailability of paired datasets in wide-range scenarios, unsupervised low-light image enhancement (ULLIE) techniques have been developed. However, these methods are primarily guided to alleviate the degradation effect on visual quality rather than semantic levels, hence limiting their performance in visual recognition tasks. To this end, we propose to learn a Semantic Degradation-Aware Guidance (SDAG) that perceives the low-light degradation effect on semantic levels in a self-supervised manner, which is further utilized to guide the ULLIE methods. The proposed SDAG utilizes the low-light degradation factors as augmented signals to degrade the low-light images, and then capture their degradation effect on semantic levels. Specifically, our SDAG employs the subsequent pre-trained recognition model extractor to extract semantic representations, and then learns to self-reconstruct the enhanced low-light image and its augmented degraded images. By constraining the relative reconstruction effect between the original enhanced image and the augmented formats, our SDAG learns to be aware of the degradation effect on semantic levels in a relative comparison manner. Moreover, our SDAG is general and can be plugged into the training paradigm of the existing ULLIE methods. Extensive experiments demonstrate its effectiveness for improving the ULLIE approaches on the downstream recognition tasks while maintaining a competitive visual quality. Code will be available at https://github.com/zheng980629/SDAG.

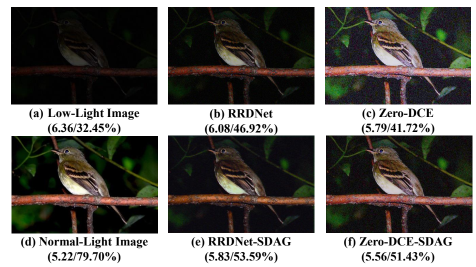


Figure 1: The illustration of visual quality and visual recognition performance on the testing set of the simulated low-light image CUB (Wah et al. 2011) dataset. The numbers in the parentheses represent the average NIQE values and top-1 accuracy evaluated on VGG16 classifier. As can be seen, the ULLIE approaches (Zero-DCE and RRDNet) improve the visual quality of low-light images but are insufficient for visual recognition. In contrast, by integrating our SDAG into these ULLIE methods, the visual recognition performance is significantly improved while retaining the visual quality.

Introduction

Shooting images in a low-light environment often suffer low visibility due to low lightness and noise corruption. Degraded by these factors, the low-light images not only experience the problem of poor human visual perception, but also pose challenges for downstream visual recognition tasks, such as object detection (Ren et al. 2015), face recognition (Deng et al. 2019), and image segmentation (Chen et al. 2017). To address this issue, various low-light image enhancement methods have been developed (Zhang, Zhang, and Guo 2019; Wang et al. 2021; Wu et al. 2022). However, most of them are designed in a supervised manner, limiting their applications due to the lack of paired datasets in wide-range scenarios.

In recent years, we have witnessed the progress of unsupervised low-light image enhancement (ULLIE) algorithms (Guo et al. 2020; Zhu et al. 2020; Liu et al. 2021) that do not rely on paired datasets. Specifically, these approaches enhance the quality of low-light images with lightness enhancement and noise removal, thus alleviating the degradation effect of the low-light condition. However, they are primarily devised to improve human visual perception, which could be inadequate for downstream visual recognition tasks (see Fig. 1).

To explain this phenomenon, it has been proved that the degradation effect on visual recognition is often on semantic levels (Wang et al. 2020), while previous ULLIE methods cannot guarantee to reduce this effect, leading to uncovered and incomplete feature representation for recognition tasks (see Fig. 2). Previous works based on supervised approaches solve this problem by guiding the enhancement

*These authors contributed equally.

†Corresponding authors: F. Zhao and M. Zhou.

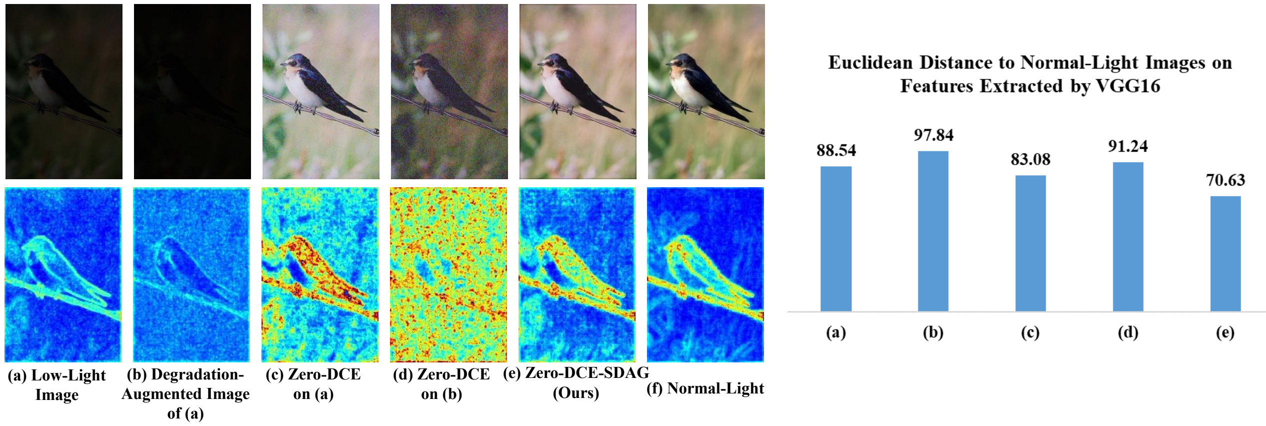


Figure 2: To explore the impact of low-light degradation factors and ULLIE algorithms on deep representations, in the left part, we first augment the low-light image (a) by the low-light degradation factors (low-lightness and noise pollution) to generate the degradation-augmented image (b). Then, we process (a) and (b) by a ULLIE algorithm: Zero-DCE (Guo et al. 2020) to generate the corresponding enhanced images (c) and (d). Their features extracted by the pre-trained VGG16 classifier are presented in the bottom row images. In the right part, we further present the mean Euclidean distance between the deep features of the degraded/enhanced images and corresponding normal-light ones on the CUB dataset. In summary, we can observe that: i) the more an image is degraded by low-lightness and noise corruption, the greater the effect of degradation on semantic levels will be; and ii) existing ULLIE methods are inadequate to reduce the degradation effect on the semantic level. By integrating our SDAG into with ULLIE methods, the degradation effect on the semantic level is mitigated with lower mean Euclidean distance.

models to learn either semantic information of high-quality ground truth or semantic labels (Wang et al. 2020; Kim et al. 2021; Yang et al. 2022), hence alleviating the degradation effect on semantic levels. However, for existing ULLIE techniques, it is challenging to meet this goal due to the lack of guidance from clean ground truth or semantic labels.

To tackle the shortcomings, we propose to learn a Semantic Degradation-Aware Guidance (SDAG) that perceives the low-light degradation effect on the semantic level in a self-supervised manner, and then utilize it to guide the learning of ULLIE methods by minimizing this degradation effect. To construct such a guidance, we notice a phenomenon that the more an image is degraded by low-light degradation factors (e.g., low-lightness and noises corruption), the greater degradation effect on semantic levels arises (see Fig. 2). Based on this observation, we design the SDAG to capture the relative degradation effect on the semantic level between the original low-light image and its augmented degraded counterparts. Similar to the ranker strategy (Zhang et al. 2019), such a comparative mechanism endows the proposed SDAG with the awareness of the degradation effect on the semantic level by measuring the reconstruction effect.

Specifically, we formulate the SDAG in a self-reconstruction manner (see Fig. 4), which is trained to relatively reconstruct the enhanced low-light image and its augmented degraded formats. To facilitate downstream recognition tasks, our SDAG employs its feature extractor to extract the semantic representation of input enhanced images, followed by a reconstruction part that processes the features to reconstruct the input image. During training, by constraining the reconstruction effect of the original enhanced image to overwhelm that of its augmented degraded counterpart, our SDAG is able to perceive the degradation effect on semantic levels, where more degraded semantic-level features perform worse on reconstructing its input image. After training, the SDAG can be easily plugged into the training paradigm of existing ULLIE methods [see Fig. 3(b)], which guides them to alleviate the degradation effect on the semantic level (see Fig. 2) by optimizing the reconstruction effect of SDAG on enhanced results. Extensive experiments show the effectiveness of our SDAG in instructing existing ULLIE methods to produce improved enhancement results that elevate the visual recognition performance significantly while preserving the visual quality. Integrating our SDAG into existing ULLIE frameworks achieves the top-1 accuracy improvement of VGG16 on the CUB dataset from 41.72% to 51.43% and 46.92% to 53.59% on the improved enhancement images.

Our contributions can be summarized as follows:

- We propose to learn a Semantic Degradation-Aware Guidance (SDAG) that improves the performance of ULLIE methods on the downstream visual recognition tasks. To achieve this goal, our SDAG is trained to perceive the low-light degradation effect on the semantic level in a self-supervised manner, which is generic and can be integrated into the standard training paradigm of ULLIE algorithms to alleviate the degradation effect.
- We design the SDAG in a self-reconstruction manner. It consists of a fixed extractor that extracts the semantic representation, and a reconstruction part that is constrained to reconstruct the original enhanced input image and its augmented degraded versions of different degrees.
- We evaluate the performance by applying our method to

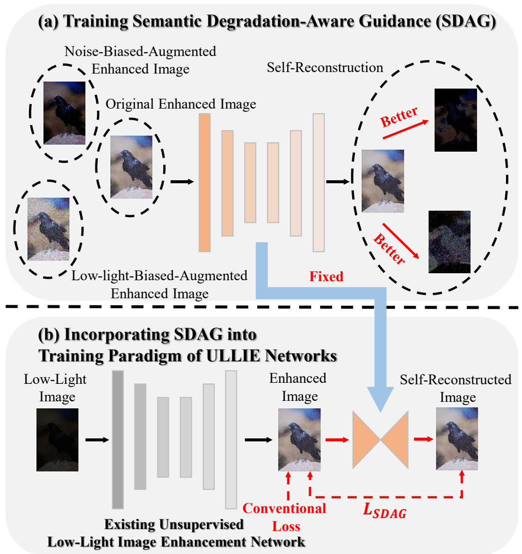


Figure 3: The overview of our proposed framework. Firstly, SDAG is trained to learn to self-reconstruct the input enhanced image, where the reconstruction effect of the original enhanced image is constrained to be better than the enhanced result of augmented degraded versions. Then, the trained SDAG is integrated into the training paradigm of ULLIE methods, which guides them to alleviate the degradation on semantic levels by optimizing the reconstruction effect of enhanced images on the SDAG.

some existing ULLIE frameworks. Experiments demonstrate that the improved enhancement results can significantly elevate the visual recognition performance, as well as retaining a competitive visual effect.

Related Work

Low-Light Image Enhancement

Conventional low-light image enhancement methods (Guo, Li, and Ling 2016; Fu et al. 2016b; Wang et al. 2013) often employ the Retinex theory (Land 1977), which decomposes an image into illumination and reflectance components, and attributes illumination adjustment to the illumination component. Recently, deep learning-based algorithms (Zhang, Zhang, and Guo 2019; Fan et al. 2020; Wang et al. 2021; Zhang et al. 2022; Wu et al. 2022) have been developed to improve the visual quality of low-light images. However, most of them are based on supervised learning and rely on paired dataset, limiting their applications in wide-range scenarios. To this end, some ULLIE algorithms are designed to improve the visual quality of the low-light images in an unsupervised manner. Guo et al. (Guo et al. 2020) proposed to estimate the pixel-wise and high-order curves for adjusting the dynamic range of low-light images without any paired or unpaired data. RRDNet (Zhu et al. 2020) implements the illumination adjustment and noise removal by decomposing the low-light image into illumination, reflectance, and noise components, and an illumination-guided noise estimation loss. RUAS (Liu et al. 2021) proposes a cooperative bilevel search strategy with low-light priors to explore the network architectures from a compact search space in an unsupervised way. SCI (Ma et al. 2022) estimates a cascaded illumination learning process and constructs the self-calibrated module to realize the convergence between results of each stage for fast, flexible, and robust enhancement. Although these approaches enhance the visual quality of low-light images by alleviating the low-light degradation effect, they are inadequate for improving the performance of downstream visual recognition tasks due to the lack of restoring such an effect on semantic levels.

Image Enhancement for Visual Recognition

In recent years, we have witnessed progress in applying enhancement techniques for improving visual recognition. Following the classic image enhancement pipeline, a line of methods attempt to produce enhancement results constrained to the semantic representation of clean ground truth (Son et al. 2020; Lee et al. 2021). Another line of works address this issue by alleviating the degradation effect on the recognition models. For instance, (Wang et al. 2020; Kim et al. 2021; Yang et al. 2022) intend to produce high-quality-like features of degraded images on classification models with the assistance of semantic representation from the clean ground truth. With the supervision of semantic labels and degradation parameters, (Cui et al. 2021, 2022) introduce the degradation transformation into the object detection framework to improve its performance. However, all of these solutions necessitate semantic labels or clean ground truth, thus limiting their applications in wide-range scenarios.

Method

Overview

In this work, to improve the recognition performance on the enhanced images produced by ULLIE algorithms without any normal-light images and semantic labels, we propose a Semantic Degradation-Aware Guidance (SDAG) to capture the relative degradation effect on semantic levels in a self-supervised manner. By integrating our SDAG into the training paradigm of existing ULLIE algorithms, they will be optimized to generate visually pleasing results with a mitigated degradation effect on semantic levels (see Fig. 2), resulting in improved recognition performance. The overall framework of our SDAG is illustrated in Fig. 3. The pipeline involves the following two stages:

Stage 1: Training Semantic Degradation-Aware Guidance. The proposed SDAG is trained to self-reconstruct the enhanced low-light image and its augmented degraded formats in a self-supervised manner, as described in Fig. 4. In the constraint of the self-reconstruction effect of the original enhanced images overwhelms that of its augmented degraded counterpart, our SDAG is capable of perceiving the low-light degradation effect on the semantic level in a self-supervised manner based on the self-reconstruction effect.

Stage 2: Incorporating SDAG into Training Paradigm of ULLIE Networks. The trained SDAG is integrated into the training paradigm of ULLIE methods, which guides

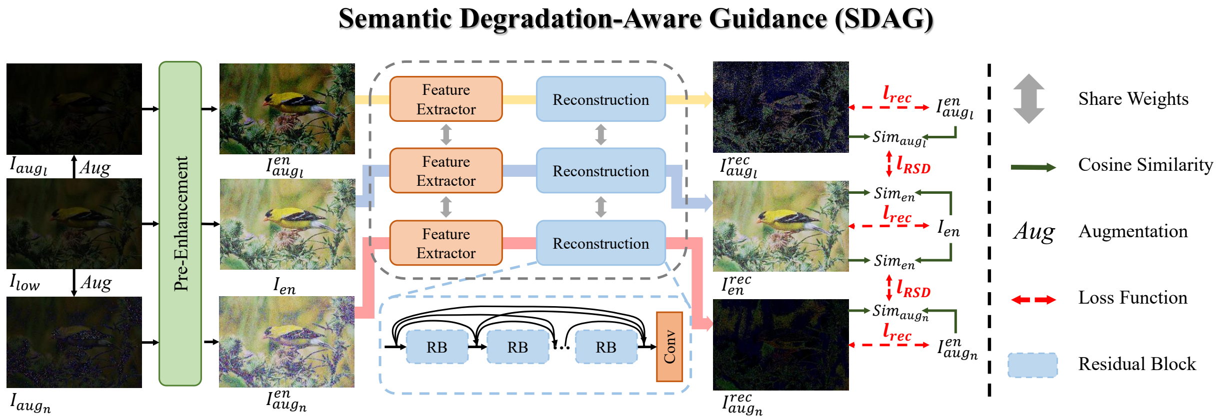


Figure 4: The illustration of the proposed Semantic Degradation-Aware Guidance (SDAG). Firstly, the low-light image, L_{low} is augmented by the low-light degradation factors into the low-light-biased-augmented image, I_{aug_l}, and the noise-biased-augmented image, I_{aug_n}, which are further enhanced by a pre-enhancement operation that derives I_{en}, $I^{en}_{aug_l}$, and $I^{en}_{aug_n}$, respectively. Then, SDAG employs the feature extractor of the downstream recognition model to extract semantic representation of them, and a reconstruction part to self-reconstruct the input enhanced images guided by L_{rec}, L_{RSD}, and L_{AS} (omitted in figure). In particular, the reconstruction loss L_{rec} constrains the reconstruction result to be closed with input, while the relative similarity distance loss L_{RSD} constrains the reconstruction effect of I_{en} to overwhelm $I^{en}_{aug_l}$ and $I^{en}_{aug_n}$. After training, SDAG can perceive the low-light degradation effect on the semantic level by measuring reconstruction effect (see Fig. 5).

them to alleviate the degradation on the semantic level for better visual recognition performance (see Fig. 3(b)).

Semantic Degradation-Aware Guidance

We observe that the more an image is degraded by low-lightness and noise corruption, the greater the effect of degradation on semantic levels (see Fig. 2). Therefore, considering the low-light degradation includes poor visibility, low contrast, color deviation, and noise pollution simultaneously, we augment an original low-light image, I_{low}, to a low-light-biased-augmented image, I_{aug_l} that is degraded more on lightness, and a noise-biased-augmented image, I_{aug_n} that is degraded more on the noise level. The first augmentation is implemented by gamma correction, which introduces low-lightness, color deviation, and low contrast while the second is achieved by adding Gaussian noise. To capture the degradation effect on the semantic-level features of enhanced degraded images, we first apply a pre-enhancement operation to obtain the corresponding enhanced images, I_{en}, $I^{en}_{aug_l}$, and $I^{en}_{aug_n}$, where this pre-enhancement operation is better to set as the target improved ULLIE method, and the final result is robust to other enhancement methods (see *Ablation Studies*). Then, we employ a pre-trained recognition model as the feature extractor to extract the semantic representations of the above enhanced images, respectively. These representations are further processed by a reconstruction part to self-reconstruct the input images, I^{rec}_{en}, $I^{rec}_{aug_l}$, and $I^{rec}_{aug_n}$, respectively, where this part is simple and includes four residual blocks, deconvolution, and LeakyReLU layers (see Fig. 4).

To measure the self-reconstruction effect, we calculate the cosine similarity between the enhanced low-light image and its self-reconstructed image to represent their similarity:

$$Sim_{en} = \frac{I_{en} \cdot I^{rec}_{en}}{\|I_{en}\|_2 \|I^{rec}_{en}\|_2}, \tag{1}$$

where $\cdot$ means dot product, and $\|\cdot\|_2$ denotes the L_2 normalization. Sim_{aug_l} and Sim_{aug_n} are obtained by the same operation on the augmented enhanced formats $I^{en}_{aug_l}$, and $I^{en}_{aug_n}$, respectively.

To enable the proposed SDAG to be aware of semantic degradation, we design the following loss function to constrain the self-reconstructed effect of the original enhanced image and its augmented degraded counterparts.

Reconstruction Loss. The reconstruction loss guarantees the self-constructed effect for the input images:

$$\begin{aligned} L_{rec} = &\|I_{en} - I^{rec}_{en}\|_1 \\ &+ \alpha * \|I^{en}_{aug_l} - I^{rec}_{aug_l}\|_1 + \alpha * \|I^{en}_{aug_n} - I^{rec}_{aug_n}\|_1, \end{aligned} \tag{2}$$

where $\|\cdot\|_1$ denotes L_1 normalization. α is set to 0.1 to enable the reconstruction effect of the original enhanced image is stronger than that of augmented degraded images.

Relative Similarity Distance Loss. To establish the awareness of our SDAG on the semantic degradation caused by low-lightness and noise corruption, we apply the relative similarity distance loss to constraint the reconstruction effect of the original enhanced image to overwhelm that of its augmented degraded counterparts by:

$$\begin{aligned} L_{RSD} = &margin(Sim_{en}, Sim_{aug_l}, \sigma) \\ &+ margin(Sim_{en}, Sim_{aug_n}, \sigma), \end{aligned} \tag{3}$$

where $margin(a, b, \sigma) = max(0, -(a+b) + \sigma)$ and σ denotes the relative similarity distance, which is set to 0.5 in our experiments. The effect of σ is discussed in the *Ablation Studies* section.

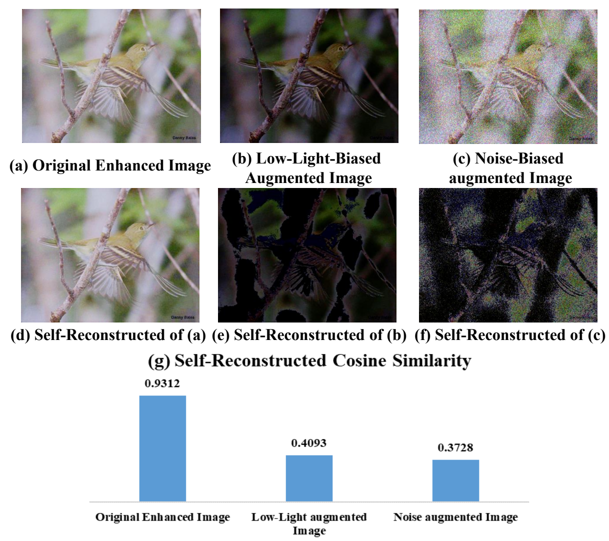


Figure 5: Examples of self-reconstruction effect. (a) The original enhanced image, (b) low-light-biased-augmented enhanced image, (c) noise-biased-augmented enhanced image, and (d)-(f) corresponding self-reconstructed images of (a)-(c), and (g) average cosine similarity between the reconstructed image and the corresponding input enhanced image on the CUB testing dataset. It demonstrates that our SDAG is able to reconstruct the original enhanced images with greater effect than its degraded augmented images.

Absolute Similarity Loss. The absolute similarity loss encourages the SDAG to increase the reconstruction effect of the original enhanced image and decreases that of its augmented degraded images independently:

$$L_{AS} = (1 - Sim_{en}) + Sim_{aug_l} + Sim_{aug_n} \quad (4)$$

Overall, the total loss of the SDAG can be expressed as :

$$L = L_{rec} + 0.1 * L_{RSD} + 0.1 * L_{AS} \quad (5)$$

The self-reconstruction effect is verified on the CUB dataset's testing set, demonstrating that our SDAG reconstructs the original enhanced image better than its degradation-augmented images (see Fig. 5).

Incorporating the SDAG into ULLIE Methods

As guidance to alleviate the degradation effect on the semantic level of enhanced results, our SDAG could be further plugged into the training paradigm of most existing ULLIE networks (see Fig. 3(b)). We choose two representative baseline methods, Zero-DCE (Guo et al. 2020) and RRDNet (Zhu et al. 2020) as backbones, and denote them with our SDAG as Zero-DCE-SDAG and RRDNet-SDAG, respectively. The former baseline approaches mainly to improve the lightness while the latter can further reduce the noise effect. The trained SDAG guides these ULLIE models to mitigate the degradation effect on semantic level by maximizing the reconstruction effect of SDAG on the enhanced results:

$$L_{SDAG} = \|I_{res} - I_{res}^{rec}\|_1, \quad (6)$$

where I_{res} is the enhanced images produced by ULLIE methods, and I_{res}^{rec} denotes the reconstructed images of the SDAG on I_{res}.

By plugging this additional guidance procedure with the conventional loss of Zero-DCE and RRDNet, the total loss functions of them is expressed by:

$$L_{ULLIE} = L_{conven} + 2L_{SDAG}, \quad (7)$$

where L_{conven} is the conventional loss functions in original Zero-DCE and RRDNet networks.

Experiments

Experimental Setup

Low-Light Image Classification. In the low-light image classification task, due to the lack of a dedicated low-light image classification dataset, we choose the CUB (Wah et al. 2011) dataset to simulate the low-light images for evaluation. Following the simulation process of (Lv, Li, and Lu 2021), we degrade the normal-light image to a low-light image with lightness reduced and noise corrupted. The training and testing sets in our experiments are identical to the official selection (Wah et al. 2011), which include 5994 images for training and 5794 images for testing.

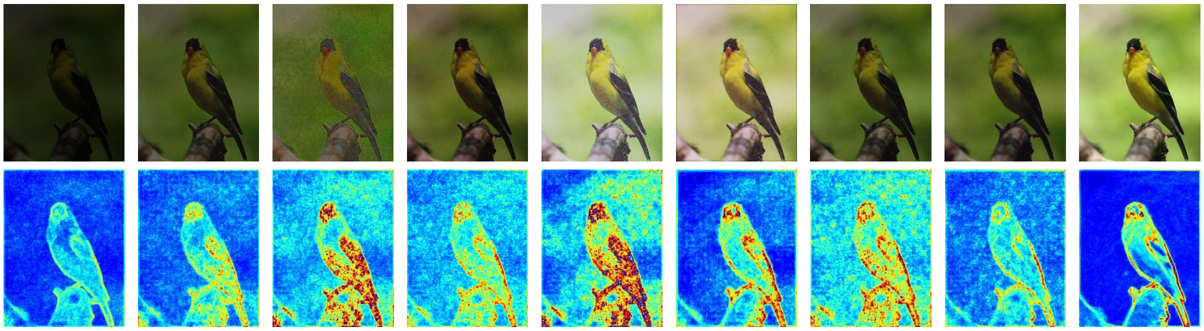
Dark Face Detection. In dark face detection tasks, we evaluate the effectiveness of our SDAG on the DARK FACK (Yuan et al. 2019) dataset. It consists of 10,000 dark images, with 6000 images for training/validation and 4000 for testing. Due to the unavailable bounding boxes of the testing set, we randomly select 500 images from the training and validation sets for performance comparison while the remaining 5,500 for training.

Implementation Details. We implemented our SDAG with PyTorch on a single NVIDIA GTX 2080Ti. We use the Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.99$ for a total of $300K$ iterations. The filter weights are randomly initialized with the standard Gaussian function. The initial learning rate is set to 2×10^{-4} and reduced by 0.5 every $50k$ iteration. The batch and patch sizes are set to 16 and 64×64, respectively.

Comparison of Methods. For performance evaluation, we compare our method with the baseline networks. Besides, four conventional algorithms (CRM (Ying et al. 2017), JieP (Cai et al. 2017), SRIE (Fu et al. 2016b), and Fusion (Fu et al. 2016a)), two unsupervised algorithms (RUAS (Liu et al. 2021) and SCI (Ma et al. 2022)), and two supervised algorithms (RetinexNet (Wei et al. 2018) and URetinexNet (Wu et al. 2022)) are chosen for comparison. For unsupervised techniques, we train their models on the low-light datasets to generate the best possible results. For supervised methods, due to the unavailable paired normal-light images, we employ publicly trained models to generate all the results.

Visual Recognition Performance

Low-Light Image Classification. We investigate the effectiveness and consistency of the proposed SDAG in performance improvement across two recognition models, *i.e.*, VGG16 (Simonyan and Zisserman 2014) and AlexNet (Krizhevsky, Sutskever, and Hinton 2012). Note



(a) Low-Light Image (b) JieP (c) RetinexNet (d) RUAS (e) Zero-DCE (f) Zero-DCE-SDAG (g) RRDNet (h) RRDNet-SDAG (i) Normal-Light Image

Figure 6: Visual comparison with state-of-the-art low-light image enhancement methods on the testing set of CUB dataset. The corresponding feature maps in the bottom row are extracted by pre-trained VGG16 classifier.

Recognition Model	Method					
AlexNet	CRM	SRIE	JieP	Fusion	RUAS	SCI
	32.95/6.19	22.26/6.34	29.08/6.24	31.84/6.08	29.08/6.16	34.39/5.84
	RetinexNet	URetinexNet	Zero-DCE	ZERO-DCE-SDAG	RRDNet	RRDNet-SDAG
	11.98/8.25	32.61/6.21	27.39/**5.76**	**33.82**/5.87	28.99/6.17	**36.05/6.03**
VGG16	CRM	SRIE	JieP	Fusion	RUAS	SCI
	46.84/5.89	45.79/6.18	47.08/5.90	46.98/6.04	48.05/5.86	51.14/5.64
	RetinexNet	URetinexNet	Zero-DCE	ZERO-DCE-SDAG	RRDNet	RRDNet-SDAG
	22.19/7.67	50.57/5.61	41.72/5.79	**51.43/5.56**	46.92/6.08	**53.59/5.83**

Table 1: Comparison results on the CUB (Wah et al. 2011) testing set in terms of the top-1 accuracy (%) $\uparrow$ and NIQE $\downarrow$.

that the VGG16 and AlexNet are pre-trained with the ImageNet (Deng et al. 2009) dataset, and we fine-tune their fully-connected layer on the normal-light CUB dataset. In the experiments, following DDP (Wang et al. 2020), we utilize the first three layers of a pre-trained VGG16 network as the feature extractor in our SDAG for the VGG16 recognition model, and the first three layers of a pre-trained AlexNet network as the feature extractor in our SDAG for the AlexNet recognition model. We measure the top-1 accuracy on the testing set of the CUB dataset to perform the evaluation. As illustrated in Table 1, for the two recognition models, the classification performance on the enhanced images produced by Zero-DCE and RRDNet is lower than most baseline methods. By integrating our SDAG into their training paradigm, Zero-DCE-SDAG and RRDNet-SDAG both achieve better performance and obtain superior accuracy than the baseline methods. In particular, Zero-DCE-SDAG and RRDNet-SDAG achieve Top-1 accuracy ranging from 41.72% to $51.43\%$and 46.92% to 53.59% for the VGG16 recognition model and from 27.39% to 32.82% and 28.99% to 36.05% for the AlexNet recognition model.

Dark Face Detection. We employ the DSFD (Li et al. 2019), a well-known face detector trained on the WIDER FACE dataset (Yang et al. 2016), as the face detector to evaluate the dark face detection performance. Fig. 7 depicts the precision-recall (P-R) curves and the average precision (AP) of each low-light image enhancement algorithm. Integrating the proposed SDAG into the training paradigm of the ULLIE method improves the AP from 27.8 (Zero-DCE) to 43.5 (Zero-DCE-SDAG), and 21.5 (RRDNet) to 34.4 (RRDNet-DSFD). It further demonstrates the effectiveness of our SDAG in guiding ULLIE algorithms to generate enhanced results with superior recognition performance.

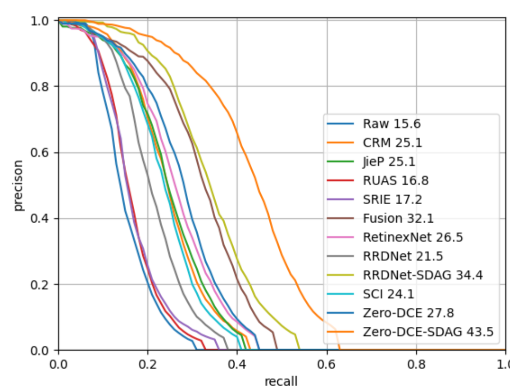


Figure 7: The performance of dark face detection.

Visual Quality Evaluation

We present the visual comparison on the testing set of the CUB and DARK FACE datasets in Fig 6 and Fig. **??**.

Loss Function	Top-1 accuracy↑ (%)
Baseline	41.72
w/o L_{Rec}	24.67
w/o L_{RSD}	45.16
w/o L_{AS}	47.32.
Full Loss	**51.43**

Table 2: Ablation studies of the contribution of loss functions in the proposed SDAG. "Baseline" and "Full loss" mean the classification performance on the enhanced images produced by Zero-DCE and Zero-DCE-SDAG, respectively.

JieP and RUAS yield the results with underexposure, while ZERO-DCE yields overexposed enhanced results. RetinexNet generates enhanced images with severe artifacts and color distortion. Although RRDNet considers noise removal, its enhanced images suffer from serious noise pollution. In contrast, our Zero-DCE-SDAG and RRDNet-SDAG produce enhanced results with improved brightness and reduced noise. Further, we visualize the deep features of the enhanced results by the pre-trained VGG16, as illustrated in the second row of Fig. 6. It indicates that integrating our SDAG into the standard training paradigm of ULLIE algorithms can guide them to generate improved enhancement results with an alleviated semantic degradation effect.

Ablation Studies

Based on the Zero-DCE network, we conduct experiments to demonstrate the effectiveness of the proposed SDAG and evaluate the classification performance on the CUB dataset by the VGG recognition model.

Contribution of Each Loss. Table 2 describes the top-1 accuracy performance on the enhanced results produced by Zero-DCE integrated with our SDAG trained by various combinations of the loss functions. Removing the reconstruction loss L_{Rec} will lead to the collapse of the SDAG, which cannot achieve self-reconstruction. Thus, the performance is worse than the backbone. The SDAG cannot exactly perceive the degradation effect on semantic levels when the relative similarity distance loss L_{RSD} is discarded, thus the classification performance on the enhanced images will not be improved significantly. Removing the absolute similarity loss L_{AS} cannot guarantee the self-reconstructed quality of our SDAG for the input images, resulting in our SDAG being ambiguous to the semantic degradation effect. The proposed SDAG with full loss functions could perceive the semantic degradation effect, hence improving the recognition performance on the improved enhancement results.

Effect of Relative Similarity Distance. To implement the degradation effect awareness on semantic levels, we limit the minimum relative distance between the reconstructed similarity between the enhanced image and its augmented formats to σ. Here, we investigate the effect of the minimum relative distance on the classification performance in Table 3. The model performs best when $\sigma = 0.5$ and the semantic degradation awareness is perturbed when lower or higher σ. In our experiments, we set σ to 0.5 by default.

σ	0.3	0.4	0.5	0.6	0.7
Acc ↑ (%)	48.79	50.27	**51.43**	51.06	49.31

Table 3: The effect of minimum relative similarity distance σ. "Acc" means the top-1 classification accuracy. In our experiments, $\sigma = 0.5$ is the default parameter.

Operation	GC	HE	Zero-DCE
Top-1 accuracy↑ (%)	45.56	46.82	**51.43**
Operation	SRIE	JieP	Fusion
Top-1 accuracy↑ (%)	50.17	50.54	51.15

Table 4: The effect of the pre-enhancement operation. "GC" and "HE" mean the gamma correction and histogram equalization operations, respectively. Zero-DCE is the default pre-enhancement operation in this section.

Ablation of Pre-Enhancement. This section explores whether the selection of the pre-enhancement method influences the effectiveness of the proposed SDAG. Aiming to improve the recognition performance of Zero-DCE, besides setting it as the default pre-enhancement operation, we also investigate the other enhancement algorithms: gamma correction, histogram equalization, SRIE, JieP, and Fusion. As shown in Table 4, due to the insufficient enhancement effect of the first two algorithms, the corresponding classification performance is unsatisfactory, however, it is stable for the last four. Therefore, our SDAG is robust to the pre-enhancement operation.

Conclusion

In this paper, we develop a novel framework for improving the performance of unsupervised low-light image enhancement (ULLIE) methods on the downstream visual recognition tasks. Without the guidance of ground truth or semantic labels, we devise a Semantic Degradation-Aware Guidance (SDAG) that guides the ULLIE approaches to alleviate the low-light degradation effect on semantic levels. In particular, the SDAG extracts semantic representation and is trained in a self-reconstruction manner, where the reconstruction effect between varied degraded enhanced images is constrained differently. Extensive experiments demonstrate that our SDAG is flexible and can perceive degradation effect on the semantic level, which can guide the ULLIE methods to improve the performance of visual recognition tasks while retaining competitive visual quality.

Acknowledgements

This work was supported by the Anhui Provincial Natural Science Foundation under Grant 2108085UD12, and the JKW Research Funds under Grant 20-163-14-LZ-001-004-01. We acknowledge the support of GPU cluster built by MCC Lab of Information Science and Technology Institution, USTC.

References

Cai, B.; Xu, X.; Guo, K.; Jia, K.; Hu, B.; and Tao, D. 2017. A joint intrinsic-extrinsic prior model for Retinex. In *Proceedings of the IEEE International Conference on Computer Vision*, 4000–4009.

Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; and Yuille, A. L. 2017. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 40(4): 834–848.

Cui, Z.; Qi, G.-J.; Gu, L.; You, S.; Zhang, Z.; and Harada, T. 2021. Multitask AET with orthogonal tangent regularity for dark object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2553–2562.

Cui, Z.; Zhu, Y.; Gu, L.; Qi, G.-J.; Li, X.; Zhang, R.; Zhang, Z.; and Harada, T. 2022. Exploring resolution and degradation clues as self-supervised signal for low quality object detection. *arXiv preprint arXiv:2208.03062*.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Li, F.-F. 2009. ImageNet: A large-scale hierarchical image database. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 248–255. Ieee.

Deng, J.; Guo, J.; Xue, N.; and Zafeiriou, S. 2019. ArcFace: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4690–4699.

Fan, M.; Wang, W.; Yang, W.; and Liu, J. 2020. Integrating semantic segmentation and Retinex model for low-light image enhancement. In *Proceedings of the 28th ACM International Conference on Multimedia*, 2317–2325.

Fu, X.; Zeng, D.; Huang, Y.; Liao, Y.; Ding, X.; and Paisley, J. 2016a. A fusion-based enhancing method for weakly illuminated images. *Signal Processing*, 129: 82–96.

Fu, X.; Zeng, D.; Huang, Y.; Zhang, X.-P.; and Ding, X. 2016b. A weighted variational model for simultaneous reflectance and illumination estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2782–2790.

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 Conference on Computer Vision and Pattern Recognition*, 1780–1789.

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.

Kim, I.; Han, S.; Baek, J.-w.; Park, S.-J.; Han, J.-J.; and Shin, J. 2021. Quality-agnostic image recognition via invertible decoder. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12257–12266.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25.

Land, E. H. 1977. The Retinex theory of color vision. *Scientific American*, 237(6): 108–129.

Lee, Y.; Jeon, J.; Ko, Y.; Jeon, B.; and Jeon, M. 2021. Task-driven deep image enhancement network for autonomous driving in bad weather. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 13746–13753. IEEE.

Li, J.; Wang, Y.; Wang, C.; Tai, Y.; Qian, J.; Yang, J.; Wang, C.; Li, J.; and Huang, F. 2019. DSFD: Dual shot face detector. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5060–5069.

Liu, R.; Ma, L.; Zhang, J.; Fan, X.; and Luo, Z. 2021. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 10561–10570.

Lv, F.; Li, Y.; and Lu, F. 2021. Attention guided low-light image enhancement with a large scale low-light simulation dataset. *International Journal of Computer Vision*, 129(7): 2175–2193.

Ma, L.; Ma, T.; Liu, R.; Fan, X.; and Luo, Z. 2022. Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5637–5646.

Ren, S.; He, K.; Girshick, R.; and Sun, J. 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, 28.

Simonyan, K.; and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Son, T.; Kang, J.; Kim, N.; Cho, S.; and Kwak, S. 2020. URIE: Universal image enhancement for visual recognition in the wild. In *European Conference on Computer Vision*, 749–765. Springer.

Wah, C.; Branson, S.; Welinder, P.; Perona, P.; and Belongie, S. 2011. The caltech-ucsd birds-200-2011 dataset.

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.

Wang, Y.; Cao, Y.; Zha, Z.-J.; Zhang, J.; and Xiong, Z. 2020. Deep degradation prior for low-quality image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11049–11058.

Wang, Y.; Wan, R.; Yang, W.; Li, H.; Chau, L.-P.; and Kot, A. C. 2021. Low-light image enhancement with normalizing flow. *arXiv preprint arXiv:2109.05923*.

Wei, C.; Wang, W.; Yang, W.; and Liu, J. 2018. Deep Retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*.

Wu, W.; Weng, J.; Zhang, P.; Wang, X.; Yang, W.; and Jiang, J. 2022. URetinex-Net: Retinex-based deep unfolding network for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5901–5910.

Yang, S.; Luo, P.; Loy, C.-C.; and Tang, X. 2016. WIDER FACE: A face detection benchmark. In *Proceedings of the*

IEEE Conference on Computer Vision and Pattern Recognition, 5525–5533.

Yang, Z.; Dong, W.; Li, X.; Wu, J.; Li, L.; and Shi, G. 2022. Self-Feature Distillation with Uncertainty Modeling for Degraded Image Recognition. In *European Conference on Computer Vision*, 552–569. Springer.

Ying, Z.; Li, G.; Ren, Y.; Wang, R.; and Wang, W. 2017. A new low-light image enhancement algorithm using camera response model. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 3015–3022.

Yuan, Y.; Yang, W.; Ren, W.; Liu, J.; Scheirer, W. J.; and Wang, Z. 2019. UG2+ Track 2: A collective benchmark effort for evaluating and advancing image understanding in poor visibility environments. *arXiv preprint arXiv:1904.04474*.

Zhang, W.; Liu, Y.; Dong, C.; and Qiao, Y. 2019. RankSRGAN: Generative adversarial networks with ranker for image super-resolution. In *Proceedings of the IEEE International Conference on Computer Vision*, 3096–3105.

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

Zhang, Z.; Zheng, H.; Hong, R.; Xu, M.; Yan, S.; and Wang, M. 2022. Deep color consistent network for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1899–1908.

Zhu, A.; Zhang, L.; Shen, Y.; Ma, Y.; Zhao, S.; and Zhou, Y. 2020. Zero-shot restoration of underexposed images via robust Retinex decomposition. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, 1–6.