Low-Light Image and Video Enhancement: A Comprehensive Survey and Beyond

Shen Zheng, Yiling Ma*, Jinqian Pan*, Changjie Lu*, Gaurav Gupta

***Abstract*—This paper presents a comprehensive survey of low-light image and video enhancement. We first consider the challenging mixed over-/under-exposed images, which are underperformed by existing methods. To this end, we propose two variants of the SICE dataset named SICE_Grad and SICE_Mix. Because the lack of a low-light video dataset has impeded the extension of low-light image enhancement (LLIE) to videos, we introduce a large-scale, high-resolution video dataset named Night Wenzhou. Night Wenzhou is challenging since it is captured in fast motion with aerial scenes and street landscapes under diverse illuminations and degradation. Based on these newly proposed datasets and the existing benchmark datasets, we conduct extensive key technique analysis and experimental comparisons for representative LLIE methods. We then discuss the open challenges and suggest future research directions for the LLIE community.**

***Index Terms*—Low-Light Image and Video Enhancement, Low-Level Vision, Deep Learning, Computational Photography.**

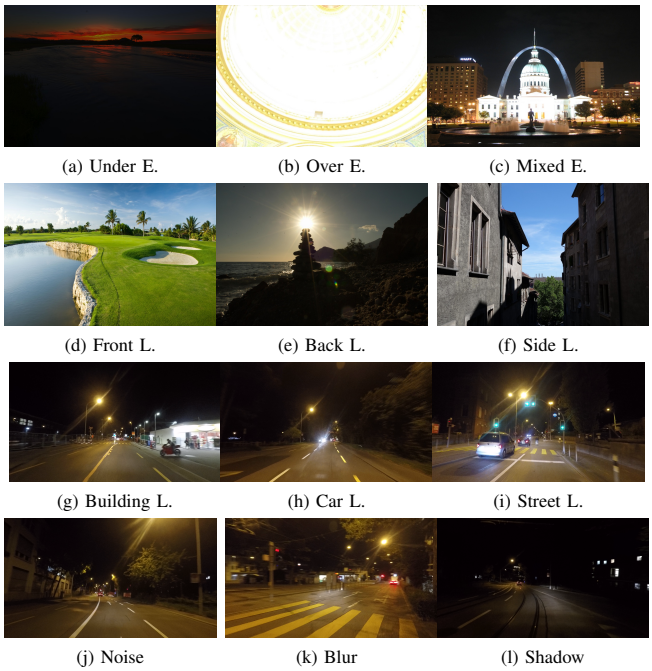
I. Introduction

Images are often captured under sub-optimal illumination conditions. Because of environmental factors (e.g., bad lightening, wrong beam angle) or technical constraints (e.g., small ISO, short exposure) [1], these images often have deteriorated features, and low contrast (See Fig. 1), which not only harm the low-level perceptual quality but also degrade the high-level vision tasks such as object detection [2], semantic segmentation [3], and depth estimation [4].

One plausible way to solve the issue above comes from the camera side. Admittedly, increasing ISO and exposure will improve the brightness of the images. However, increasing ISO leads to noise, whereas extending exposure results in motion blur [5], which makes the images look even worse. The other plausible way is to use image manipulation software like Photoshop or Lightroom to improve the visual quality of low-light images. However, both software requires artistic tastes and are time-consuming on large-scale datasets.

Unlike the camera and software approaches which demand manual efforts, low-light image enhancement (LLIE) aims to automatically improve the visibility of images taken in low-light conditions. It is an active research field that is related to various system-level applications, such as visual surveillance [6], autonomous driving [7], and Unmanned aerial vehicle [8].

Traditional methods are the only choice for LLIE in pre-deep learning eras. Most traditional LLIE methods utilize Histogram Equalization [9], [10], [11], Retinex theory [12], [13], [14], [15], [16], or Dehazing [17], [18]. These methods are theoretically sound but are empirically underperformed and inefficient. Therefore, in recent years, traditional LLIE methods tend to be a supplement (e.g., RetinexNet [19]) instead of competitors for deep learning LLIE methods, which has become popular because of their excellent effectiveness, efficiency, and generalizability.



(a) Under E. (b) Over E. (c) Mixed E.
(d) Front L. (e) Back L. (f) Side L.
(g) Building L. (h) Car L. (i) Street L.
(j) Noise (k) Blur (l) Shadow

Fig. 1: Example images at challenging illumination conditions. First row: different exposures (E.). Second and third row: different lightening (L.). Bottom row: different degradation.

Deep learning-based LLIE methods can be divided into supervised learning [20], [21], [22], [19], [5], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], unsupervised learning [41], [42], semi-supervised learning [43] and zero-shot learning [44], [45], [46], [47], [48], [49] methods. There has been a large amount of deep learning-based LLIE publication over the recent 5 years, and all learning strategies have their strengths and limitations. For example, supervised learning has state-of-the-art performance in benchmark datasets but generalizes poorly to unseen datasets, whereas unsupervised and zero-shot learning do the opposite. It is important to thoroughly

Shen Zheng is with the School of Computer Science, Carnegie Mellon University, Pittsburgh, USA. Jinqian Pan is with the Center for Data Science, New York University, NY, USA. Yiling Ma, Changjie Lu, and Gaurav Gupta is with the School of Science and Technology, Wenzhou-Kean University, Wenzhou, China.

* Indicates equal contribution.

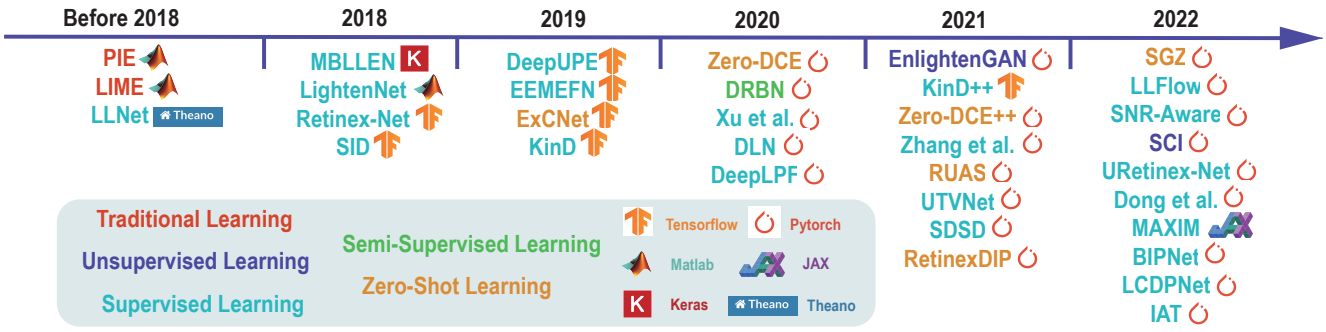


Fig. 2: Milestone for recent representative low-light image and video enhancement methods.

review the recent progress, which can provide an in-depth understanding, indicate current challenges, and suggest future works for the LLIE community.

There are three recent surveys about LLIE. Wang et al. [50] reviews traditional learning-based LLIE methods. Liu et al. [51] propose a new LLIE dataset named VE-LOL, reviews LLIE methods, and introduce a joint image enhancement and face detection network named EDTNet. Li et al. [1] introduce a new LLIE dataset named LLIV-Phone, reviews deep learning-based LLIE methods, and design an online demo platform for LLIE methods.

The existing surveys have several limitations. Firstly, their proposed dataset has either overexposure *or* underexposure for single images. That assumption contradicts real-world images, which often contain overexposure *and* underexposure in a single image. Secondly, their proposed dataset contains few videos, and even these videos are filmed in fixed shooting positions. That over-simplification is also contradictory to real-world videos often captured in motion. Besides, these surveys highlight the low-level perceptual quality and high-level vision tasks but ignore the system-level application, which is essential when LLIE methods are deployed in real-world products. Finally, they focused on methods by the first half of 2021. That is considered outdated from the end of 2022.

This paper makes the following contributions to the existing LLIE surveys:

- We present the latest comprehensive survey for low-light image and video enhancement. In particular, we make a modularized discussion focusing on structures and strategies and conduct an extensive qualitative and quantitative comparison with various full-reference and non-reference evaluation metrics.
- We introduce two image datasets named SICE_Grad and SICE_Mix. They are the first datasets that contain overexposure and underexposure in single images. This preliminary attempt points out the unresolved mixed over-/under-exposure challenge for the LLIE community.
- We propose Night Wenzhou, a large-scale high-resolution video dataset. Night Wenzhou is captured during fast motions and contains diverse illuminations, various landscapes, and miscellaneous degradation. It will facilitate the application of LLIE methods to real-world challenges like autonomous driving and UAV.
- We highlight the application of LLIE methods both at the algorithm level and at the system level. Based on that, we point out the open challenges and suggest directions for future work.

The rest of the paper is organized as follows. Section II provides a systematic review of existing LLIE methods. Section III introduces the benchmark datasets and the proposed datasets. Section IV makes empirical analysis and comparisons for representative LLIE methods. Section V discusses the open challenges and the corresponding future works. Section VI provides the concluding remarks.

II. Methods Review

A. Selection Criteria

We select the LLIE methods according to the following rubrics. Firstly, we focus on LLIE methods in the recent 5 years (2018-2022) and pay specific emphasis to deep learning-based LLIE methods in the recent 2 years (2021-2022) because of their rapid development. That being said, we also add three earlier works (2015-2017) to aid comparison. Secondly, we pick LLIE methods published in prestigious conferences (e.g., CVPR) and journals (e.g., TIP) and provided the official codes to ensure credibility and authenticity. Thirdly, for the paper published in the same year, we prefer works with more citations and Github stars. Finally, we include LLIE methods that significantly surpass previous state-of-the-art benchmark LLIE datasets.

B. Learning Strategies

Before 2017, traditional methods are the ad-hoc solution for LLIE. Since 2017, deep learning methods have started to dominate this field. Supervised learning (67.6 %) is so far the most popular strategy. From the year 2019, there are several methods using zero-shot learning (17.6 %), unsupervised learning (5.9 %), and semi-supervised (2.9 %). In this work, we categorize the existing LLIE methods' learning strategies into traditional, supervised, unsupervised, semi-supervised, and zero-shot learning.

Traditional Learning Traditional Learning methods in LLIE refer to learning methods that do not use neural networks. Mainstream traditional learning methods utilize Histogram Equalization, Retinex theory, or Dehazing.

Histogram Equalization-based methods spread out the frequent intensity values of an image to improve its global contrast. In this way, the low-contrast region of an image gains higher contrast, and the visibility improves. Retinex-based methods assume that an image can be decomposed into a reflectance map and an illumination map. The enhanced image can be obtained by fusing the enhanced illumination map and the reflectance map. Dehazing-based methods treat the inverted low-light images as haze images and apply dehazing algorithms to enhance the image.

Supervised Learning In LLIE, supervised learning refers to the learning strategy with paired images. The example will be one dataset with 1000 low-light images and another with 1000 normal-light images that are different in only illuminations. It is worth mentioning that the supervised learning method has achieved state-of-art-results in benchmark datasets.

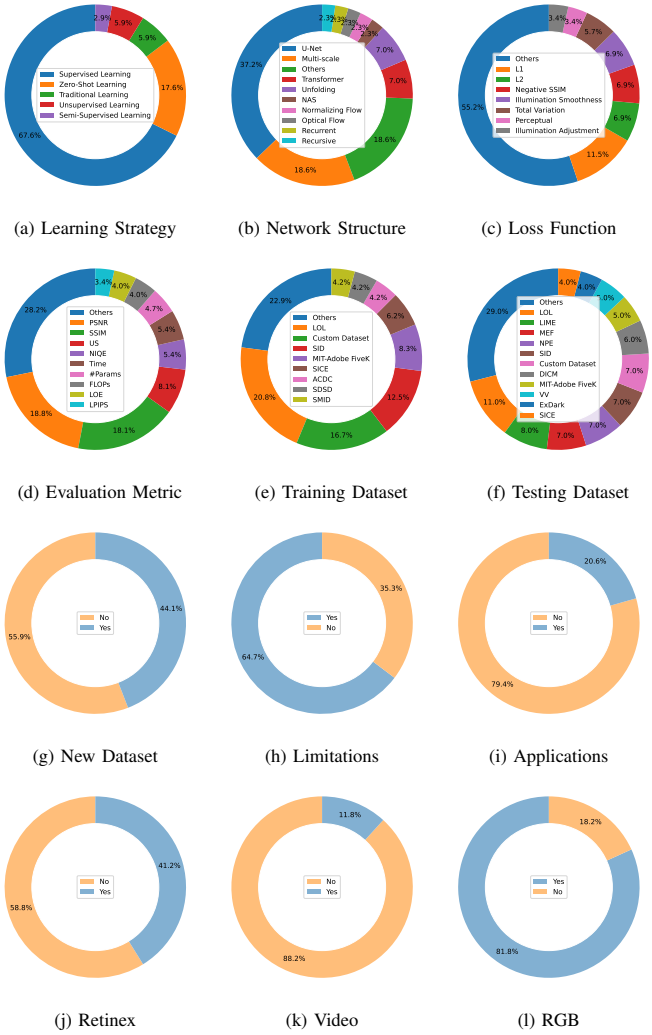
Unsupervised Learning In LLIE, unsupervised learning refers to the learning strategy without paired images. An example will be a dataset with low-light images and another with normal-light images that are different in more than illuminations. In this way, the unsupervised learning method avoids the tedious work of collecting paired images.

Semi-supervised learning In LLIE, semi-supervised learning is a learning strategy with a small number of paired images and a large number of unpaired images. An example will be a dataset with low-light images and another with normal-light images where most images are different in more than illuminations, and few images are different in only illuminations.

Zero-shot learning In LLIE, Zero-shot learning is a learning strategy which require neither paired data or unpaired training dataset. Instead, Zero-shot learning learns image enhancement at test time using data-free loss functions like exposure loss or color loss. Thanks to the data-free loss functions, zero-shot learning methods have good generalization ability, require a small number of parameters, and has fast inference speed.

Discussion The aforementioned learning strategies for LLIE have the following limitations.

- Traditional Learning methods' performances lag behind deep learning methods, even with their handcrafted priors and intricate optimization steps, which result in poor inference latency
- Supervised Learning methods rely heavily on the paired training dataset, but none of the approaches for obtaining such a dataset is feasible. Specifically, it is difficult to capture image pairs that are only different in illuminations; it is hard to synthesize images that fit the complex real-world scenes; it is expensive and time-consuming to retouch large-scale low-light images.
- Unsupervised Learning methods rely on the unpaired training dataset, which induces data bias. Because of the data bias, unsupervised learning methods like EnlightenGAN (EGAN) [41] and SCI [42] generalize poorly to the testing dataset that has significant domain gaps from the training dataset.
- Semi-supervised learning methods inherit the limitations of both supervised and unsupervised learning methods without fully utilizing their strengths. That's why semi-supervised learning has been used by only one representative LLIE method DRBN [43].
- Zero-shot learning methods require elaborate design about the data-free loss functions, which cannot cover the necessary properties of real-world low-light images. Besides, zero-shot learning methods' performance still lags behind supervised learning methods like LLFlow [33].



(a) Learning Strategy (b) Network Structure (c) Loss Function

(d) Evaluation Metric (e) Training Dataset (f) Testing Dataset

(g) New Dataset (h) Limitations (i) Applications

(j) Retinex (k) Video (l) RGB

Fig. 3: Donut Chart for methods summary.

C. Network Structures

Many LLIE methods utilize a U-Net-like (37.2 %) structure or multi-scale information (18.6 %). Some methods use transformers (7.0 %) or unfolding networks (7.0 %). A few methods (2.3 % for each) use Neural Architecture Search (NAS), Normalizing Flow, Optical Flow, Recurrent Network, or Recursive Network.

U-Net and multi-scale U-Net-like [52] structure is the most popular since it preserves high-resolution rich detail features and low-resolution rich semantic features, which are essential for LLIE. The same logic is applied to a non-U-Net-like structure that uses multi-scale information.

Transformers Recently, the transformers-based [53] method has surged in computer vision, especially high-level vision tasks, due to its ability to track long-range dependencies and capture global information in an image.
Unfolding and NAS The unfolding network (a.k.a. unrolling network) [54] has been used by several methods because it combines the wisdom of model-based and data-based approaches. NAS [55] is automating design of a neural network that generates the optimal result with a given dataset.
Normalizing and Optical Flow The normalizing flow-based [56] method transforms a simple probability distribution into a complex distribution with a sequence of invertible mappings, whereas optical flow-based [57] methods estimate the pixel-level motions of adjacent video frames.
Recurrent and Recursive Recurrent network [58] is a type of neural network that repeatedly process the input in chain structures, whereas recursive network [59] is a variant of the recurrent network that processes the input in hierarchical structures.
Discussion The aforementioned network structures for LLIE have the following limitations.

- Transformers is currently unpopular in low-level vision tasks like LLIE. Perhaps this is due to their impotence to integrate local and non-local attention and inefficiency at processing high-resolution images.
- The unfolding strategy requires elaborate network design, whereas NAS requires computationally expensive parameter learning.
- The normalizing flow and optical flow-based methods have poor computational efficiency and long inference latency.
- The recurrent network (and its LSTM variants [60]) have the vanishing gradient problem at large-scale data [61], whereas the recursive network additionally relies on the inductive bias of hierarchical distribution, which is unrealistic.

D. Loss Functions

The choice of loss functions is highly diverse among LLIE methods. 55.2 % of the LLIE methods use non-mainstream loss functions. Among mainstream loss functions, L_1 (11.5 %) is the most popular, whereas L_2 (6.9 %), Negative SSIM (6.9 %), and Illumination Smoothness (6.9 %) are also popular. Small amounts of methods use Total Variation (5.7 %), Perceptual (3.4 %), or Illumination Adjustment (3.4 %).
Full-Reference loss L_1 loss, L_2 loss, Negative SSIM loss, Perceptual loss [62] and Illumination adjustment loss [25] are full-reference loss functions (i.e., loss requiring paired images). L_1 loss targets the absolute difference between image pairs, whereas L_2 loss targets the squared difference. Therefore, L_2 loss penalizes large errors more and is more tolerant for small errors, whereas L_1 loss does the opposite. Like other low-level vision tasks [63], L_1 loss in LLIE is more popular than L_2 loss. Negative SSIM loss is based on the luminance, contrast, and structure between paired images. However, Negative SSIM loss is uncommon in LLIE. That is different from other low-level vision tasks like image deraining [64], where it gains tremendous popularity. Perceptual loss is the L_2 difference for paired images based on their extracted features from a network (usually VGG16 [65]). It is popular in low-level vision tasks like style transfer [62] but is less explored in LLIE. Illumination adjustment loss is the L_2 difference for illumination and illumination gradients between image pairs.
Non-Reference loss Total Variation (TV) loss [66] and illumination smoothness loss [25] are non-reference loss functions (i.e., losses that do not require paired images). TV loss measures the sum of the difference between adjacent pixels in vertical and horizontal directions for an image. Therefore, total variation loss suppresses irregular patterns like noise and blur and promotes smoothness in the image. Illumination smoothness loss is similar to total variation loss since it is written as the L_1 norm of illumination divided by the maximum variation. Despite their success at other low-level vision tasks like denoising and deblurring, [67], the variation-based methods have been less explored in LLIE.

E. Evaluation Metrics

Many LLIE methods choose PSNR (18.8 %) or SSIM (18.1 %) as the evaluation metrics. Apart from PSNR and SSIM, the User Study (US) (8.1 %) is a popular method. Small amounts of methods use NIQE (5.4 %), Inference Time (5.4 %), #Params (4.7 %), FLOPs (4.0 %), LOE (4.0 %), or LPIPS (3.4 %) as the evaluation metrics.
Full-Reference Metrics Peak Signal-to-Noise Ratio (PSNR), Structure Similarity Index (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS) [68] are full-reference image quality evaluation metrics. The PSNR measures similarity at the pixel level between image pairs, whereas SSIM measures the similarity according to luminance, contrast, and structure. LPIPS measures the difference between two images at patch level at a pretrained network. Therefore, higher PSNR and SSIM and lower LPIPS indicate better visual quality.
Non-Reference Metrics Natural Image Quality Evaluator (NIQE) [69] and Lightness order Error (LOE) [13] are non-reference image quality evaluation metrics. Both NIQE and LOE target the naturalness of an image. Specifically, NIQE compares the gap between an image to a model trained from images of natural scenes, where LOE analyze that image using lightness-order errors. Therefore, a lower NIQE and LOE indicate better visual quality.
Subjective Metrics A user study is the only subjective metric used for representative LLIE methods. Typically, the user study score is the mean opinion score from a group of participants. A high user study score means better perceptual quality from human perspectives.
Efficiency Metrics Efficiency metrics include inference time, Numbers of Parameters (#Params), and Floating Point Operations (FLOPs). A shorter inference time and a smaller #Params and FLOPs indicates better efficiency.

F. Training and Testing Data

Popular benchmark training data for LLIE include LOL (20.8 %), SID (12.5 %), and MIT-Adobe FiveK (8,3 %). A small amount of method uses SICE (6.2 %), ACDC (4.2 %),

TABLE I: Table Summary for existing benchmark dataset and the proposed SICE_Grad and SICE_Mix.

Dataset	Number	Resolutions	Type	Train	Paired	App.
NPE [13]	8	Varies	Real	n	n	n
LIME [16]	10	Varies	Real	n	n	n
MEF [70]	17	Varies	Real	n	n	n
DICM [71]	64	Varies	Real	n	n	n
VV	24	Varies	Real	n	n	n
LOL [19]	500	400×600	Real	y	y	n
VE-LOL [51]	13,440	Various	Both	b	y	y
ACDC [72]	4,006	1,080×1,920	Real	y	n	y
DCS [49]	150	1,024×2,048	Syn	n	y	y
DarkFace [73]	10,000	720×1,080	Real	y	n	y
ExDark [2]	7,363	Varies	Real	y	n	y
SICE [74]	4,800	Varies	Both	y	y	n
SICE_Grad	589	600×900	Both	y	y	n
SICE_Mix	589	600×900	Both	y	y	n

SDSD (4.2 %), or SMID (4.2 %). However, many methods utilize their custom dataset (16.7 %) or other datasets (22.9 %) that are less popular for training. Popular benchmark testing data for LLIE include LOL (11.0 %) and LIME (8.0 %). Some methods utilize MEF (7.0 %), NPE (7.0 %), and SID (7.0 %). A small number of methods use DICM (6.0 %), MIT-Adobe FiveK (5.0 %), VV (5.0 %), ExDark (4.0 %), or SICE (4.0 %). Similar to the case of training data, many methods utilize their custom dataset (7.0 %) or other datasets (29.0 %) that are less popular for testing datasets. The detailed training and testing datasets discussion will be in Section IV.

G. Others

New Dataset The number of LLIE methods that introduce new datasets (55.9 %) surpasses the number of LLIE methods that only use existing datasets (44.1 %). This reflects the importance of data for LLIE.

Limitations Most LLIE methods (64.7 %) do not mention their limitations and future works. This makes it hard for future researchers to improve upon their work.

Applications Most LLIE methods (79.4 %) do not relate low-level image enhancement to high-level applications like detection or segmentation. Therefore, the practical values of these methods remain a question.

Retinex There are a number of methods (41.2 %) that utilize Retinex theory for LLIE enhancement. However, most LLIE methods (58.8 %) do not utilize Retinex theory. Hence, the Retinex theory remains a popular but non-dominant choice for LLIE.

Video Most LLIE methods (88.2 %) do not consider Low-Light Video Enhancement (LLVE) tasks. This isn't good since many real-world low-light visual data are in video format.

RGB Most LLIE methods (81.8 %) uses RGB data for training. This is great since RGB is much more popular than RAW for modern digital devices like laptops or smartphones.

III. Datasets

A. Benchmark Datasets

NPE / LIME / MEF / DICM [13], [16], [70], [71] carries 8/10/17/64 real low-light images of various resolutions. They contain indoor items and decorations, outdoor buildings, streetscapes, and natural landscapes, and they are all for testing.

VV[1] contains 24 real multi-exposure images of various resolutions. It contains personal traveling photos with indoor and outdoor persons and natural landscapes for testing.

LOL [19] contains 500 pairs of real low-light images of 400 × 600 resolutions. It only contains indoor items and is split into 485 training images and 15 testing images.

VE-LOL [51] contains 13,440 real and synthetic low-light images and image pairs of various resolutions. It has diversified scenes, including natural landscapes, streetscapes, buildings, human faces, etc. The paired portion VE-LOL-L has 2100 pairs for training and 400 pairs for testing, whereas the unpaired portion VE-LOL-H has 6940 images for training and 4000 for testing. Besides, the VE-LOL-H portion contains detection labels for high-level object detection tasks.

ACDC [72] contains 4,006 real low-light images of resolution 1,080 × 1,920. It includes autonomous driving scenes with adverse conditions (1000 foggy, 1000 snowy, 1000 rainy, and 1006 nighttime) and has 19 classes. In particular, the ACDC nighttime contains 400 training images, 106 validation images, and 500 test images. Besides, ACDC contains semantic segmentation labels which allow high-level semantic segmentation tasks.

DCS [49] contains 150 synthetic low-light images of resolution 1,024 × 2,048. Specifically, it is synthesized with gamma correction upon the original CityScape [75] dataset, and it contains urban scenes with fine segmentation labels (30 classes), which allow high-level instance segmentation, semantic segmentation, and panoptic segmentation tasks. The Dark CityScape (DCS) dataset is intended for testing only.

DarkFace [73] contains 10,000 real low-light images of resolution 720 × 1,080. It contains nighttime streetscapes with many human faces in each image. It consists of 6,000 training and validation images and 4,000 testing images. Since it contains object detection labels, it can be applied to high-level object detection tasks.

ExDark [2] contains 7,363 real low-light images of varies resolutions. It contains images with diversified indoor and outdoor scenes under 10 illumination conditions with 12 object classes. It is split into 4,800 training images and 2,563 testing images. It contains object detection labels and can be applied to high-level object detection tasks.

SICE [74] contains real and synthetic multi-exposure images of various resolutions. It contains images with diversified indoor and outdoor scenes synthesized with different exposure levels. The train/val/test follows a 7:1:2 ratio. In particular, SICE contains both normal-exposed and ill-exposed images. Therefore, it can be used for supervised, unsupervised, and zero-shot learning.

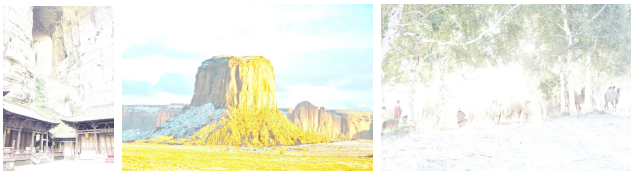
Discussion The current benchmark datasets for LLIE have the following limitations.

- Many datasets use synthetic images to meet the paired image requirement for supervised learning methods. These image synthesis techniques usually follows simple gamma correction or exposure adjustment, which does

[1] https://sites.google.com/site/vonikakis/datasets

(a) SICE (low) [74]

(b) SICE (normal) [74]



(c) SICE (high) [74]

(d) SICE_Grad

(e) SICE_Mix

Fig. 4: Examples Images for SICE (low-light, normal-light, high-light) and the proposed SICE_Grad and SICE_Mix.

not fit the diverse illuminations in the real-world. Consequently, methods trained with these synthetic images generalize poorly to real-world images.

- The existing datasets consider single images rather than both images and videos. That is because that high-quality low-light videos are hard to capture and that many methods cannot process high-resolution video frames in real-time. However, the real-world applications (e.g., visual surveillance, autonomous driving, and UAV) are heavily dependent on videos rather than single images. The lack of low light video dataset greatly undermines the benefits of LLIE for these fields.
- The existing datasets either consider underexposure or overexposure only, or consider underexposure and overexposure in separate images in a dataset. There is no dataset that contains mixed under-/overexposure in single images. The detailed discussions is in III-C.

B. New Image Dataset

We synthesize two new datasets, dubbed **SICE_Grad** and **SICE_Mix**, based on the SICE [74] dataset. To obtain these two datasets, we first reshape the original SICE dataset to a resolution of 600×900. After that, we obtain panels from images in SICE that have the same background but different exposures. The next step is different for SICE_Grad and SICE_Mix. For SICE_Grad, we arrange the panels from low exposure to high exposure. To make it more challenging, we randomly placed some normally exposed panels at the end instead of at the mid. For SICE_Mix, we permute all panels at random. Example images for the original SICE and the proposal SICE_Grad and SICE_Mix are in Fig. 4. The table summary for SICE_Grad and SICE_Mix is in Tab. I.

Our SICE_Grad and SICE_Mix dataset has two significant advantages over existing datasets. Firstly, they are synthesized by permuting the panels of the SICE dataset. Therefore, we can use SICE_Grad or SICE_Mix paired with reshaped SICE dataset for training the supervised learning methods. On the other hand, we can also use SICE_Grad and SICE_Mix as testing datasets. Secondly, SICE_Grad and SICE_Mix contain extremely uneven exposure within a single image, often seen in real-world applications like UAVs.

C. Exposure Analysis

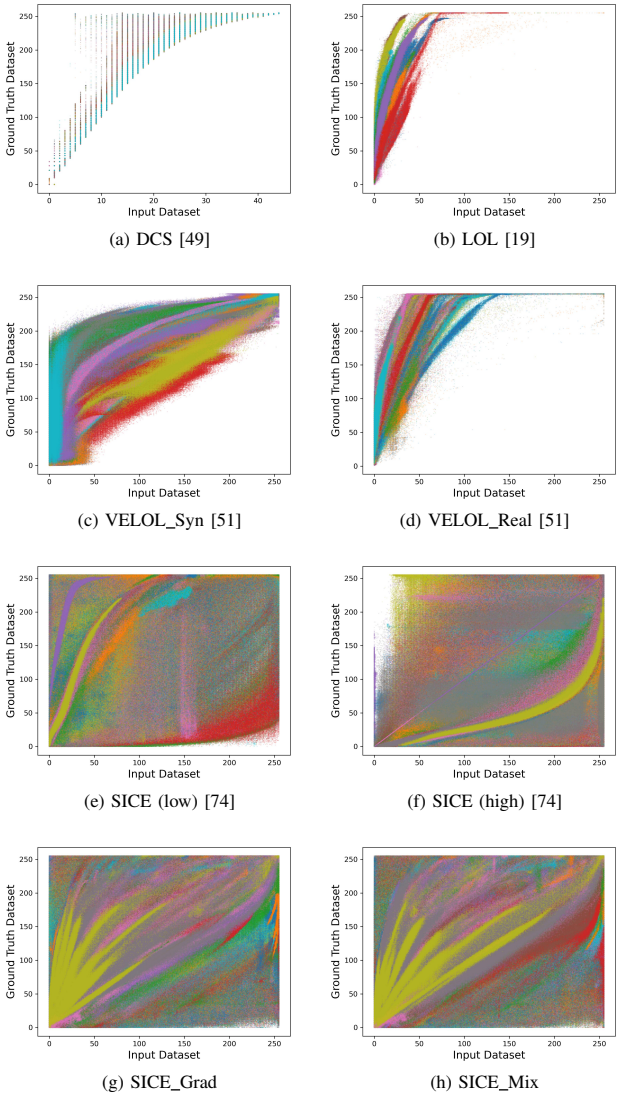
We conduct an exposure analysis to study the over-/underexposure for benchmark datasets with paired images. Specifically, we pick DCS [49], LOL [19], VE-LOL (Syn) [51], VE-LOL (Real) [51], and SICE [74] to compare with the proposed SICE_Grad and SICE_Mix.

Our exposure analysis result is shown in Fig. 5. The horizontal axis represents the pixel value for input (e.g., under-exposed, over-exposed) images. In contrast, the vertical axis represents the pixel value for the ground truth (i.e., normal-exposed) images. An individual curve describes an image pair; a plot with only concave/convex curves means that the input dataset contains only under/over-exposed images; a plot with both concave and convex curves contains both over-exposed and under-exposed images; a curve in the plot that is both concave and convex means that there is mixed overexposure and underexposure in a single image.

From the plots, we know that DCS [49], LOL [19], VE-LOL (Syn) [51], and VE-LOL (Real) [51] contain under-exposed images only. SICE (low-light) has a majority of under-exposed images and a minority of over-exposed images, whereas SICE (high-light) has a majority of over-exposed images and a minority of under-exposed images. SICE_Grad and SICE_Mix are unique because they not only have over-exposed and under-exposed images across the whole dataset but also mixed overexposure and underexposure in single images. This feature also makes SICE_Grad and SICE_Mix particularly challenging for image enhancement. Our experiments in IV show no representative LLIE method shows the satisfactory result on SICE_Grad or SICE_Mix.

D. New Video Dataset

We collect a large-scale dataset named **Night Wenzhou** to comprehensively analyze the performance of existing methods in real-world low-light conditions. In particular, the dataset contains aerial videos captured with DJI Mini 2 and



(a) DCS [49] (b) LOL [19]

(c) VELOL_Syn [51] (d) VELOL_Real [51]

(e) SICE (low) [74] (f) SICE (high) [74]

(g) SICE_Grad (h) SICE_Mix

Fig. 5: Over-/underexposure analysis with different datasets that has paired images. For each subfigure, the horizontal axis refers to the input dataset, whereas the vertical axis refers to the ground truth dataset.

streetscapes captured with GoPro HERO7 Silver. All videos are taken nighttime in Wenzhou, China, and have an FPS of 30. The table summary for Night Wenzhou is in Tab. II. As we can see, Night Wenzhou contains videos of 2 hours and 3 minutes and has a size of 26.144 GB.

The Night Wenzhou dataset is challenging since it contains large-scale high-resolution videos of diverse illumination conditions (e.g., extremely dark, underexposure, moonlight, uneven illumination, etc.). Besides, it contains various degradation (e.g., noise, blur, shadows, artifacts) commonly seen in real-world applications like autonomous driving. The Night Wenzhou dataset can be used to train unsupervised and zero-shot learning methods and to test LLIE methods with any learning strategy. Example images for Night Wenzhou are in Fig. 6.

TABLE II: Table Summary for Night Wenzhou dataset.

Device	Resolution	Duration (h:m:s)	Size (GB)
GoPro	1440 × 1920	0:09:08	1.871
GoPro	1440 × 1920	0:13:04	2.675
GoPro	1440 × 1920	0:17:41	3.727
GoPro	1440 × 1920	0:17:40	3.727
GoPro	1440 × 1920	0:17:43	3.727
GoPro	1440 × 1920	0:11:15	2.316
GoPro	1440 × 1920	0:05:14	1.029
GoPro	1440 × 1920	0:17:57	3.727
GoPro	1440 × 1920	0:02:25	0.526
GoPro Total		**1:52:07**	**23.325**
DJI	1530 × 2720	0:01:47	0.500
DJI	1530 × 2720	0:00:27	0.126
DJI	1530 × 2720	0:00:42	0.198
DJI	1530 × 2720	0:00:42	0.177
DJI	1530 × 2720	0:00:27	0.127
DJI	1530 × 2720	0:06:21	1.604
DJI	1530 × 2720	0:00:27	0.087
DJI Total		**0:10:53**	**2.819**
Total		**2:03:00**	**26.144**

IV. Evaluations

A. Quantitative Comparisons

PSNR, SSIM, and LPIPS [76] are full-reference metrics, whereas UNIQUE [77], BRISQUE [78], and SPAQ [79] are non-reference metrics.

Table III is the first table for quantitative comparison. We can see that LLFlow [33] has the best performance: it achieves the best PSNR, SSIM, and LPIPS on LOL [19] and VE-LOL (Real) [51], and the best PSNR and SSIM on DCS [49]. KinD++ [29] has the second-best performance: it achieves the best PSNR, SSIM, and LPIPS on VE-LOL (Syn) and SICE_Mix, and the best PSNR and LPIPS on SICE_Grad. Besides, Zero-DCE [45] has the best LPIPS on DCS [49], whereas KinD [25] has the best SSIM on SICE_Grad. No other methods achieve the best score at any metrics.

Table IV is the second table for quantitative comparison. The competition for this table is much more intense than the previous one. The only method that has 4 best scores is Zero-DCE [45]; the only method that has 3 best scores is KinD++ [29]; the method with 2 best scores include RetinexNet [19] and SGZ [49]; the method with 1 best score include KinD [25], LLFlow [33], URetinexNet [35], and SCI [42]. RUAS [47] is the only method that does not score best in any metrics for any dataset.

Table V is the third table for the quantitative comparison. In this table, RUAS [47] has the best UNIQUE and BRISQUE on DarkFace [73]; RetinexNet [19] has the best SPAQ for DarkFace [73], and ExDark [2]. LLFlow [33] has the best UNIQUE for ExDark [2].

Table VI shows the Quantitative Comparison of model efficiency. We choose ACDC [72] as the benchmark for efficiency comparison since it contains images of 2K resolution (i.e., 1080 × 1920), which is closer to real-world applications such as autonomous driving, UAV, and photography. We can see that SGZ [49] has the best FLOPs and Inference Time, whereas SCI [42] has the best #Params. Besides, it is worth mentioning Zero-DCE [45], RUAS [47], SGZ [49], and SCI [42] achieve real-time processing on a single GPU.

TABLE III: Quantitative Comparison on LOL, DCS, VE-LOL (Syn & Real), SICE_Mix, SICE_Grad.

Methods	LOL [19]			DCS [49]			VE-LOL (Syn) [51]			VE-LOL (Real) [51]			SICE_Mix			SICE_Grad		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
RetinexNet[19]	17.559	0.645	0.381	N/A	N/A	N/A	15.606	0.449	0.769	17.676	0.642	0.441	12.397	0.606	0.407	12.450	0.619	0.364
KinD[25]	15.867	0.637	0.341	13.145	0.720	0.304	16.259	0.591	0.432	20.588	0.818	0.143	12.986	0.656	0.346	13.144	**0.668**	0.302
Zero-DCE[45]	14.861	0.562	0.330	16.224	0.849	**0.172**	14.071	0.369	0.652	18.059	0.580	0.308	12.428	0.633	0.362	12.475	0.644	0.314
KinD++[29]	15.724	0.621	0.363	N/A	N/A	NONE	**16.523**	**0.613**	**0.411**	17.660	0.761	0.218	**13.196**	**0.657**	**0.334**	**13.235**	0.666	**0.295**
RUAS[47]	11.309	0.435	0.377	11.601	0.412	0.449	12.386	0.357	0.642	13.975	0.469	0.329	8.684	0.493	0.525	8.628	0.494	0.499
SGZ[49]	15.345	0.573	0.334	16.369	0.854	0.204	13.830	0.385	0.664	18.582	0.584	0.309	10.866	0.607	0.415	10.987	0.621	0.364
LLFlow[33]	**19.341**	**0.839**	**0.142**	**20.385**	**0.897**	0.240	15.440	0.476	0.517	**24.152**	**0.895**	**0.098**	12.737	0.617	0.388	12.737	0.617	0.388
URetinexNet[35]	17.278	0.688	0.302	16.009	0.755	0.369	15.273	0.466	0.591	21.093	0.858	0.103	10.903	0.600	0.402	10.894	0.610	0.356
SCI[42]	14.784	0.525	0.333	14.264	0.689	0.249	12.542	0.373	0.681	17.304	0.540	0.307	8.644	0.529	0.511	8.559	0.532	0.484

TABLE IV: Quantitative Comparison on NPE [13], LIME [16], MEF [70], DICM [71], and VV.

Methods	NPE [13]			LIME [16]			MEF [70]			DICM [71]			VV		
	UNI	BRI	SPAQ	UNI	BRI	SPAQ	UNI	BRI	SPAQ	UNI	BRI	SPAQ	UNI	BRI	SPAQ
RetinexNet [19]	0.801	16.533	71.264	0.787	24.31	70.468	0.742	14.583	**69.333**	0.778	**22.877**	62.550	N/A	N/A	N/A
KinD [25]	0.792	20.239	70.444	0.766	39.783	67.180	0.747	32.019	63.266	0.776	33.092	59.946	0.814	29.439	**61.453**
Zero-DCE [45]	**0.814**	17.456	**72.945**	0.811	20.437	67.736	**0.762**	17.321	66.864	0.777	27.560	57.402	**0.836**	34.656	60.716
KinD++ [29]	0.801	19.507	71.742	0.748	**19.954**	**73.414**	0.732	27.781	67.831	0.774	27.573	**62.744**	N/A	N/A	N/A
RUAS [47]	0.706	47.852	61.598	0.783	27.589	62.076	0.713	23.677	60.701	0.710	38.747	47.781	0.770	38.370	47.443
SGZ [49]	0.783	**14.615**	72.367	0.789	20.046	67.735	0.755	**14.463**	66.134	0.777	25.646	55.934	0.824	31.402	58.789
LLFlow [33]	0.791	28.861	67.926	0.805	27.060	66.816	0.710	30.267	67.019	**0.807**	26.361	61.132	0.800	31.673	61.252
URetinexNet [35]	0.737	25.570	70.066	**0.816**	24.222	67.423	0.715	22.346	66.310	0.765	26.453	59.856	0.801	30.085	55.399
SCI	0.702	28.948	64.054	0.747	23.344	64.574	0.733	15.335	64.616	0.720	31.263	48.506	0.779	**26.132**	48.667

TABLE V: Quantitative Comparison on DarkFace [73] and ExDark [2].

Methods	DarkFace [73]			ExDark [2]	
	UNI	BRI	SPAQ	UNI	SPAQ
RetinexNet [19]	0.737	18.574	**54.966**	0.708	**66.330**
KinD [25]	0.737	48.311	41.070	0.728	55.690
Zero-DCE [45]	0.720	26.194	47.868	0.729	52.700
KinD++ [29]	0.719	32.492	52.905	0.723	61.036
RUAS [47]	**0.740**	**13.770**	42.329	0.712	47.785
SGZ [49]	0.713	24.647	47.392	0.729	51.236
LLFlow [33]	0.708	22.284	51.544	**0.735**	56.116
URetinexNet [35]	0.739	15.148	51.290	0.722	57.291
SCI [42]	0.719	19.511	46.046	0.709	50.618

TABLE VI: Efficiency Comparison on the ACDC [72] (resolution of 1080 $\times$ 1920) using a single NVIDIA GeForce RTX 3090 GPU. Blue indicates real-time capability.

Methods	FLOPs	#Params	Time
RetinexNet [19]	N/A	0.5550	N/A
KinD [25]	1103.9117	8.1600	3.5288
Zero-DCE [45]	164.2291	0.0794	0.0281
KinD++ [29]	N/A	8.2750	N/A
RUAS [47]	6.7745	0.0034	0.0280
SGZ [49]	**0.2135**	0.0106	**0.0026**
LLFlow [33]	892.7097	1.7014	0.3926
URetinexNet [35]	1801.4110	0.3401	0.2934
SCI [42]	0.7465	**0.0003**	0.0058

TABLE VII: Semantic Segmentation Result Comparison on ACDC [72] and DCS [49].

ACDC [72]			DCS [49]		
Methods	mPA	mIoU	Methods	mPA	mIoU
KinD [25]	60.79	49.18	PIE [14]	68.89	61.97
Zero-DCE [45]	59.00	49.51	RetinexNet [19]	66.76	57.96
RUAS [47]	50.42	44.48	MBLLEN [21]	59.06	51.98
SGZ [49]	61.65	49.50	KinD [25]	71.69	63.42
LLFlow [33]	**62.68**	49.30	Zero-DCE [45]	74.20	64.36
URetinexNet [35]	62.32	48.71	Zero-DCE++ [46]	74.43	65.51
SCI [42]	57.52	**49.66**	SGZ [49]	**74.50**	**65.87**

Table VII shows the Quantitative Comparison of semantic segmentation. For both ACDC [72] and DCS [49], we feed the enhanced image into a semantic segmentation model named PSPNet [80] and calculate the mPA and mIoU score with the default thresholds. On ACDC[72], LLFlow [33] has the best mPA, whereas SCI [42] has the best mIoU. On DCS [49], SGZ [49] has the best mPA and mIoU.

Table VIII shows the Quantitative Comparison of object detection. In particular, we feed the enhanced image into a face detection model named DSFD [81] and calculate the IoU score with different IoU thresholds (0.5, 0.6, 0.7). We find that LLFlow [33] achieves the best IoU with all given thresholds.

B. User Studies

Since there are few effective metrics to evaluate the visual quality of low-light video enhancement, we conducted a user study to assess the performances of different methods on the proposed Night Wenzhou dataset. Specifically, we ask 100 adults participants to watch the enhancement results of 7 models, including EGAN [41], KinD [25], KinD+ [29], MBLLEN [21], RetinexNet [19], SGZ [49], and Zero-DCE [45]. They are asked to vote '1' to '5' for each method, where '1' indicates the worst performance, and '5' indicates the best.

We make a stacked bar graph in Fig. 7 to show the category-wise information for different methods. It can be seen that RetinexNet [19] has the most (37 %) of '1's; Zero-DCE [45] has the most (33 %) of '2's; EGAN [41] has the most (40 %) of '3's; MBLLEN has the most (45 %) of '4's; SGZ [49] has the most (39 %) of '5's. Therefore, RetinexNet [19] is voted to have the worst performance, whereas SGZ [49] is voted to have the best performance.

C. Qualitative Comparisons

Fig. 8 shows the qualitative comparison for an image from the VV dataset. Our finding are as follows: 1) RUAS [47]

TABLE VIII: Object Detection Result Comparison on DarkFace with different IoU thresholds.

Learning	Methods	IoU@0.5	IoU@0.6	IoU@0.7
TL	LIME [16]	0.244	0.083	0.010
SL	LLNet [20]	0.228	0.063	0.006
	LightenNet [22]	0.270	0.085	0.011
	MBLLEN [21]	0.269	0.092	0.012
	KinD [25]	0.255	0.081	0.010
	KinD++ [29]	0.271	0.090	0.011
	URetinexNet [35]	0.283	0.101	0.015
	LLFlow [33]	**0.290**	**0.103**	**0.016**
UL	EGAN [41]	0.261	0.088	0.012
ZSL	ExCNet [44]	0.276	0.092	0.010
	Zero-DCE [45]	0.281	0.092	0.013
	Zero-DCE++ [46]	0.278	0.090	0.012
	SGZ [49]	0.279	0.092	0.012

Fig. 6: Video frames from our Night Wenzhou dataset.

produces under-exposed trees 2) RUAS [47], SCI [42], and URetinexNet [35] produce over-exposed skies. 3) LIME [16], Zero-DCE [45], and LLFlow [33] yield artifacts. 4) MBLLEN [21] oversmooths the image. 5) LLFlow [33] yield color distortion. 6) PIE [14], KinD [25], and SGZ [49] have good perceptual quality.

Fig. 9 and Fig. 10 show the qualitative comparison for an image from the SICE_Grad dataset and the SICE_Mix dataset, respectively. We find that no method generates a good result on SICE_Grad or SICE_Mix. In particular, most methods successfully enhanced the under-exposed regions but made the over-exposed region even brighter. The lack of contrast from the homogeneous over-exposure makes it hard to distinguish any detail in these enhanced regions.

Fig. 11 shows the qualitative comparison (w/ object detection) for an image from the DarkFace dataset [73]. In particular, the bounding box in the figure is annotated with the predicted class and probability. Our findings are as follows: 1) KinD [25], Zero-DCE [45], RUAS [47], SGZ [49], and SCI [42] produces under-exposure images, especially for the right half. Therefore, many objects in their enhanced image are not detected. 2) KinD [25] produces oversmoothed result. That is why the object detector is way off target in its enhanced image. 3) KinD++ [29], URetinexNet [35] and LLFlow [33] are good in terms of image enhancement. However, both KinD++ [29] and URetinexNet [35] yield artifacts. That's why LLFlow's

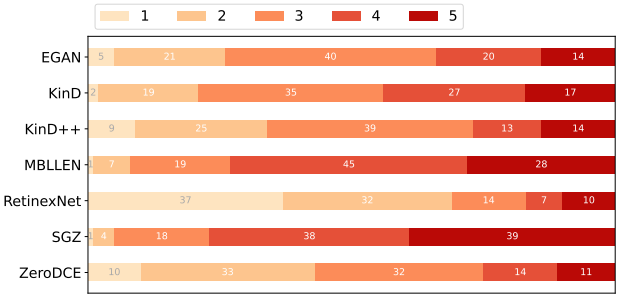
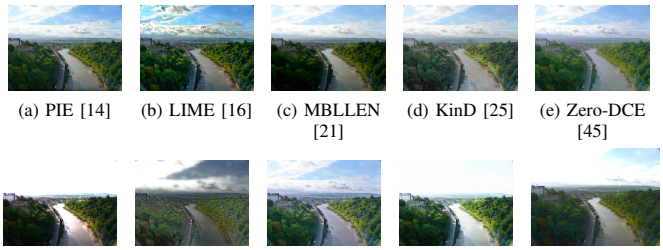


Fig. 7: Stacked Bar Plot from the results of the user study. Since the numbers of participants are exactly 100, the count shown in the plot is equivalent to the percentage.



(a) PIE [14] (b) LIME [16] (c) MBLLEN [21] (d) KinD [25] (e) Zero-DCE [45]

(f) RUAS [47] (g) LLFlow [33] (h) SGZ [49] (i) SCI [42] (j) URetinexNet [35]

Fig. 8: Visual comparison on the VV dataset.

[33] enhancement yields better object detection results.

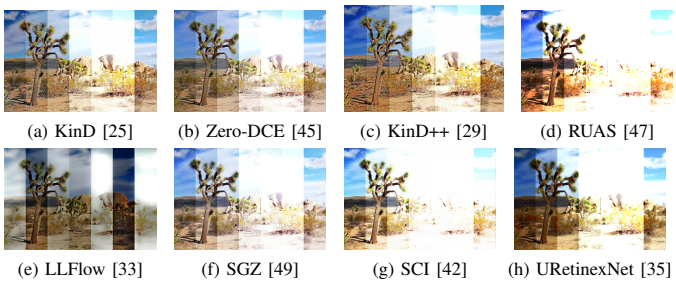
Fig. 12 shows the qualitative comparison (w/ semantic segmentation) for an image from the DCS dataset [49]. In particular, pixels with different predicted classes are visualized with different colors. Our findings are as follows 1) RetinexNet [19], MBLLEN [21], and KinD [25] leads to large areas of incorrect segmentation. Most incorrect segmentation occurs on pedestrians and sidewalks. 2) Zero-DCE [45] and SGZ [49] are close to the GT. However, SGZ [49] leads to better segmentation results for the objects in the distance.

Fig. 13 shows the qualitative comparison for a video frame from the Night Wenzhou dataset. Our findings are as below. 1) RetinexNet [19], KinD [25], and Zero-DCE [45] produces images with poor contrast, extreme color deviation, over-smoothed details, and significant noises, blurs, and artifacts. 2) EGAN [41] produces over-exposed images. 3) MBLLEN [21], KinD++ [29], and SGZ [49] produces images with good exposure. However, MBLLEN [21] oversmooths the detail, whereas KinD++ [29] generates artifacts and has more color deviation than SGZ [49].

V. Future Prospects

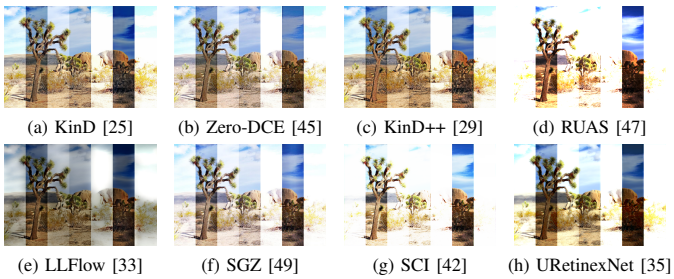
A. Uneven Exposure

Given images or videos, existing LLIE methods correct the under-exposed region but either ignore the over-exposed region or make the over-exposed region even brighter. Real-world images and videos often exhibit uneven exposure for different regions. Ideally, the LLIE methods should brighten the under-exposed region while darkening the over-exposed region.



(a) KinD [25] (b) Zero-DCE [45] (c) KinD++ [29] (d) RUAS [47]
(e) LLFlow [33] (f) SGZ [49] (g) SCI [42] (h) URetinexNet [35]

Fig. 9: Visual comparison on our SICE_Grad dataset.



(a) KinD [25] (b) Zero-DCE [45] (c) KinD++ [29] (d) RUAS [47]
(e) LLFlow [33] (f) SGZ [49] (g) SCI [42] (h) URetinexNet [35]

Fig. 10: Visual comparison on our SICE_Mix dataset.

Future works could solve the uneven exposure challenge in the following ways. Firstly, we can collect a large-scale, real-world dataset that contains mixed over-/under-exposure in a single image. Our SICE_Grad and SICE_Mix is a good starting point, but more work needs to be done. Secondly, we can adopt Laplacian Pyramid (e.g., DSLR [82]) or multi-branch fusion (e.g., TBEFN [83]) network to capture the multi-scale exposure information. Finally, use vision transformers [53], which can better capture global exposure information than CNNs. While IAT [40] has made the preliminary attempt in this field, its network structure could be improved to model more complex exposures.

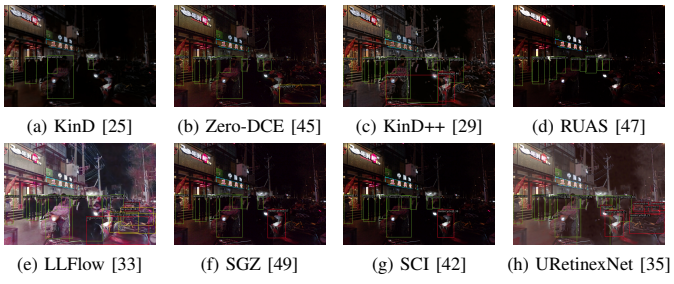
B. Integrating Low-Level and High-Level Tasks

The existing LLIE methods treat LLIE as a post-processing step before high-level vision algorithms. Whether this post-processing step is beneficial to the high-level vision algorithms remains dubious.

Future works could integrate low-level and high-level tasks to ensure the former contributes to the latter. The first way is to introduce a large-scale labeled low-light dataset for both low-level and high-level tasks. The second way is to perform joint training for low-level and high-level networks to benefit each other mutually (e.g., Deblur-YOLO [84]). The third way is to embed low-level image processing into high-level vision tasks using domain adaptation (e.g., YOLO-in-the-Dark [85]).

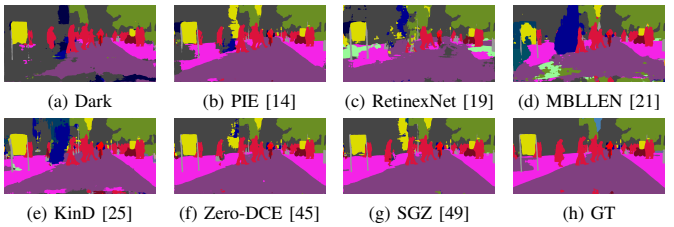
C. Preserving and Utilizing Semantic Information

While enhancing the low-light regions, the existing LLIE methods may also remove the semantic information, disturbing human understanding and degrading high-level vision algorithms. Ideally, the LLIE methods should enhance the low-light regions by preserving *and* utilizing the semantic information.



(a) KinD [25] (b) Zero-DCE [45] (c) KinD++ [29] (d) RUAS [47]
(e) LLFlow [33] (f) SGZ [49] (g) SCI [42] (h) URetinexNet [35]

Fig. 11: Object Detection Comparison on the DarkFace [73] dataset.



(a) Dark (b) PIE [14] (c) RetinexNet [19] (d) MBLLEN [21]
(e) KinD [25] (f) Zero-DCE [45] (g) SGZ [49] (h) GT

Fig. 12: Semantic Segmentation Comparison for the DCS [49] dataset. Note that the GT here refers to the segmentation result on the ground truth image.

Future works should attend to the semantic information in the following ways. First, we could introduce a large-scale dataset with fine annotated semantic labels to ensure that the semantic information is preserved during image enhancement. Besides, we could incorporate semantic priors [86] using the unfolding strategy or integrate semantic-aware modules using domain adaptation [87].

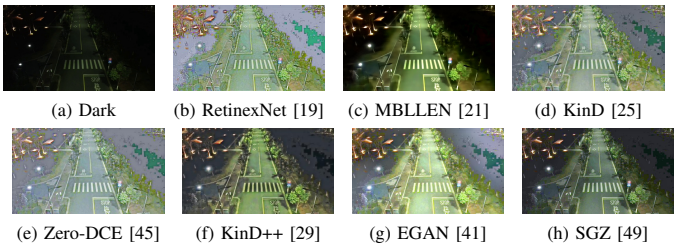
D. Low-Light Video Enhancement

Many existing LLIE methods cannot meet the real-time processing requirements of LLVE. Even for methods that meet the time constraint, a direct application to videos leads to flickering artifacts [1].

Future works should improve the efficiency and effectiveness of the LLIE methods so they can generalize well to LLVE. For efficiency, we could adopt lightweight architecture using manual design (e.g., MobileNet [88]) or NAS (e.g., NasNet [89]) to improve the inference latency. For effectiveness, we could exploit temporal information from adjacent or neighbor frames to suppress flickering artifacts. Besides, we should propose more large-scale high-resolution low-light video datasets for static and motion scenes. The introduction of Night Wenzhou is a good starting point for low-light video datasets in fast motions, but more work needs to be done.

E. Benchmark datasets

So far, there is no well-accepted benchmark dataset for LLIE. On the one hand, many supervised LLIE methods train on their dataset and generalize poorly to a dataset that has significant domain gaps from their training dataset. On the other hand, many LLIE methods test on their custom dataset, which is biased towards their method and unfair to others.



(a) Dark (b) RetinexNet [19] (c) MBLLEN [21] (d) KinD [25]

(e) Zero-DCE [45] (f) KinD++ [29] (g) EGAN [41] (h) SGZ [49]

Fig. 13: Visual Comparison for a video frame from the Night Wenzhou dataset.

Future works could learn from the success of CityScape [75] in semantic segmentation for benchmark training and testing. A well-accepted benchmark like CityScape includes large-scale real-world images and videos with diverse objects, has an appropriate train test split ratio, and contains ground truth annotations for high-level tasks. In the context of LLIE, we additionally require diverse illuminations and complex exposures.

F. Better Evaluations Metrics

The existing evaluation metrics for LLIE (e.g., PSNR, SSIM) are borrowed from other low-level vision tasks rather than designed explicitly for LLIE. In particular, methods with good scores in these evaluation metrics may not look good from the human perspective. Due to that limitation, LLIE methods heavily rely on user studies for evaluation. However, a user study is expensive and time-consuming for a large-scale dataset. Therefore, A better evaluation metric is desired.

Future works should propose an evaluation metric that fits the characteristic of low-light images and matches the user study scores. The current IQA metrics MUSIQ [90] utilize an image quality Transformer to attend the multi-scale multi-granularity information using original-resolution images and is well-match with subjective test scores. Adapting IQA metrics like MUSIQ to the task of LLIE remains a challenge.

G. Addressing Noises, Blurs and Artifacts

Most existing LLIE methods fail to suppress degradation like noise, blurs, and artifacts, which significantly degrade perceptual quality. In reality, these degradations can either exist before enhancement or appear during enhancement [91].

Future works can solve this issue in three ways. Firstly, we can design a low-light dataset with various degradation so LLIE methods can be trained to enhance the image and tackle degradation. Secondly, we can adopt loss functions that can suppress various degradation (e.g., Adaptive Tv loss [92]). Finally, we could join the task of denoising, deblurring, and LLIE in a single network (e.g., D2HNet [93]).

H. Low-Light and Bad Weather

The existing LLIE methods assume the weather conditions are good for low-light images and videos, whereas the real-world weather is usually bad (e.g., rain, snow, haze). The combination of low-light and bad weather significantly deteriorates the visual quality, which is a catastrophe for many high-level vision tasks [94].

Future works can address this challenge in two ways. One way is to collect, synthesize, or generate low-light datasets in specific bad weather so LLIE methods can be trained to address bad weather and low-light in one go. The other way is to leverage disentangled domain adaptation to tackle LLIE in bad weather (e.g., ForkGAN [95]). It is also interesting to suppress noise and blurs within the Retinex framework (e.g., Hao et al.[96])

VI. Conclusion

This paper presents a comprehensive survey of recent representative low-light image and video enhancement methods. We first introduce two SICE variants, SICE_Grad and SICE_Mix, to simulate the challenging over-/under-exposure scenes underperformed by the current LLIE methods. We then introduce a large-scale, high-resolution video dataset, Night Wenzhou, which has various illuminations and degradation. Next, we analyze the critical components of LLIE methods, including learning strategy, network structure, loss function, evaluation metric, training dataset, testing dataset, etc. After that, we conduct qualitative and quantitative comparisons of LLIE methods on the benchmark dataset and the proposed dataset. Based on the comparative analysis, we discuss the open challenges and suggest the future prospects.

References

[1] C. Li, C. Guo, L.-H. Han, J. Jiang, M.-M. Cheng, J. Gu, and C. C. Loy, "Low-light image and video enhancement using deep learning: A survey," *IEEE transactions on pattern analysis and machine intelligence*, 2021.

[2] Y. P. Loh and C. S. Chan, "Getting to know low-light images with the exclusively dark dataset," *Computer Vision and Image Understanding*, vol. 178, pp. 30–42, 2019.

[3] H. Wang, Y. Chen, Y. Cai, L. Chen, Y. Li, M. A. Sotelo, and Z. Li, "Sfnet-n: An improved sfnet algorithm for semantic segmentation of low-light autonomous driving road scenes," *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[4] M. Lamba, K. K. Rachavarapu, and K. Mitra, "Harnessing multi-view perspective of light fields for low-light imaging," *IEEE Transactions on Image Processing*, vol. 30, pp. 1501–1513, 2020.

[5] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in vhr optical remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 12, pp. 7405–7415, 2016.

[6] M. Yang, X. Nie, and R. W. Liu, "Coarse-to-fine luminance estimation for low-light image enhancement in maritime video surveillance," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2019, pp. 299–304.

[7] G. Li, Y. Yang, X. Qu, D. Cao, and K. Li, "A deep learning based image enhancement approach for autonomous driving at night," *Knowledge-Based Systems*, vol. 213, p. 106617, 2021.

[8] S. Samanta, A. Mukherjee, A. S. Ashour, N. Dey, J. M. R. Tavares, W. B. Abdessalem Karâa, R. Taiar, A. T. Azar, and A. E. Hassanien, "Log transform based optimal image enhancement using firefly algorithm for autonomous mini unmanned aerial vehicle: An application of aerial photography," *International Journal of Image and Graphics*, vol. 18, no. 04, p. 1850019, 2018.

[9] H. Ibrahim and N. S. P. Kong, "Brightness preserving dynamic histogram equalization for image contrast enhancement," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 4, pp. 1752–1758, 2007.

[10] Q. Wang and R. K. Ward, "Fast image/video contrast enhancement based on weighted thresholded histogram equalization," *IEEE transactions on Consumer Electronics*, vol. 53, no. 2, pp. 757–764, 2007.

[11] K. Nakai, Y. Hoshi, and A. Taguchi, "Color image contrast enhacement method based on differential intensity/saturation gray-levels histograms," in *2013 International Symposium on Intelligent Signal Processing and Communication Systems*. IEEE, 2013, pp. 445–449.

[12] E. H. Land, "The retinex theory of color vision," *Scientific american*, vol. 237, no. 6, pp. 108–129, 1977.

[13] S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE transactions on image processing*, vol. 22, no. 9, pp. 3538–3548, 2013.

[14] X. Fu, Y. Liao, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4965–4977, 2015.

[15] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2782–2790.

[16] X. Guo, Y. Li, and H. Ling, "Lime: Low-light image enhancement via illumination map estimation," *IEEE Transactions on image processing*, vol. 26, no. 2, pp. 982–993, 2016.

[17] X. Dong, Y. Pang, and J. Wen, "Fast efficient algorithm for enhancement of low lighting video," in *ACM SIGGRAPH 2010 Posters*, 2010, pp. 1–1.

[18] X. Zhang, P. Shen, L. Luo, L. Zhang, and J. Song, "Enhancement and noise reduction of very low light level images," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*. Ieee, 2012, pp. 2034–2037.

[19] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," *arXiv preprint arXiv:1808.04560*, 2018.

[20] K. G. Lore, A. Akintayo, and S. Sarkar, "Llnet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognition*, vol. 61, pp. 650–662, 2017.

[21] F. Lv, F. Lu, J. Wu, and C. Lim, "Mbllen: Low-light image/video enhancement using cnns." in *BMVC*, vol. 220, no. 1, 2018, p. 4.

[22] C. Li, J. Guo, F. Porikli, and Y. Pang, "Lightennet: A convolutional neural network for weakly illuminated image enhancement," *Pattern recognition letters*, vol. 104, pp. 15–22, 2018.

[23] R. Wang, Q. Zhang, C.-W. Fu, X. Shen, W.-S. Zheng, and J. Jia, "Underexposed photo enhancement using deep illumination estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6849–6857.

[24] M. Zhu, P. Pan, W. Chen, and Y. Yang, "Eemefn: Low-light image enhancement via edge-enhanced multi-exposure fusion network," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 13 106–13 113.

[25] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proceedings of the 27th ACM international conference on multimedia*, 2019, pp. 1632–1640.

[26] K. Xu, X. Yang, B. Yin, and R. W. Lau, "Learning to restore low-light images via decomposition-and-enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2281–2290.

[27] L.-W. Wang, Z.-S. Liu, W.-C. Siu, and D. P. Lun, "Lightening network for low-light image enhancement," *IEEE Transactions on Image Processing*, vol. 29, pp. 7984–7996, 2020.

[28] S. Moran, P. Marza, S. McDonagh, S. Parisot, and G. Slabaugh, "Deeplpf: Deep local parametric filters for image enhancement," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 12 826–12 835.

[29] Y. Zhang, X. Guo, J. Ma, W. Liu, and J. Zhang, "Beyond brightening low-light images," *International Journal of Computer Vision*, vol. 129, no. 4, pp. 1013–1037, 2021.

[30] F. Zhang, Y. Li, S. You, and Y. Fu, "Learning temporal consistency for low light video enhancement from single images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 4967–4976.

[31] C. Zheng, D. Shi, and W. Shi, "Adaptive unfolding total variation network for low-light image enhancement," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 4439–4448.

[32] R. Wang, X. Xu, C.-W. Fu, J. Lu, B. Yu, and J. Jia, "Seeing dynamic scene in the dark: A high-quality video dataset with mechatronic alignment," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 9700–9709.

[33] Y. Wang, R. Wan, W. Yang, H. Li, L.-P. Chau, and A. Kot, "Low-light image enhancement with normalizing flow," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 3, 2022, pp. 2604–2612.

[34] X. Xu, R. Wang, C.-W. Fu, and J. Jia, "Snr-aware low-light image enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 714–17 724.

[35] W. Wu, J. Weng, P. Zhang, X. Wang, W. Yang, and J. Jiang, "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*, 2022, pp. 5901–5910.

[36] X. Dong, W. Xu, Z. Miao, L. Ma, C. Zhang, J. Yang, Z. Jin, A. B. J. Teoh, and J. Shen, "Abandoning the bayer-filter to see in the dark," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 431–17 440.

[37] Z. Tu, H. Talebi, H. Zhang, F. Yang, P. Milanfar, A. Bovik, and Y. Li, "Maxim: Multi-axis mlp for image processing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5769–5780.

[38] A. Dudhane, S. W. Zamir, S. Khan, F. S. Khan, and M.-H. Yang, "Burst image restoration and enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5759–5768.

[39] H. Wang, K. Xu, and R. W. Lau, "Local color distributions prior for image enhancement," in *European Conference on Computer Vision*. Springer, 2022, pp. 343–359.

[40] Z. Cui, K. Li, L. Gu, S. Su, P. Gao, Z. Jiang, Y. Qiao, and T. Harada, "Illumination adaptive transformer," *arXiv preprint arXiv:2205.14871*, 2022.

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

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

[43] W. Yang, S. Wang, Y. Fang, Y. Wang, and J. Liu, "From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 3063–3072.

[44] L. Zhang, L. Zhang, X. Liu, Y. Shen, S. Zhang, and S. Zhao, "Zero-shot restoration of back-lit images using deep internal learning," in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 1623–1631.

[45] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1780–1789.

[46] C. Li, C. Guo, and C. C. Loy, "Learning to enhance low-light image via zero-reference deep curve estimation," *arXiv preprint arXiv:2103.00860*, 2021.

[47] R. Liu, L. Ma, J. Zhang, X. Fan, and Z. 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*, 2021, pp. 10 561–10 570.

[48] Z. Zhao, B. Xiong, L. Wang, Q. Ou, L. Yu, and F. Kuang, "Retinexdip: A unified deep framework for low-light image enhancement," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 3, pp. 1076–1088, 2021.

[49] S. Zheng and G. Gupta, "Semantic-guided zero-shot learning for low-light image/video enhancement," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 581–590.

[50] W. Wang, X. Wu, X. Yuan, and Z. Gao, "An experiment-based review of low-light image enhancement methods," *Ieee Access*, vol. 8, pp. 87 884–87 917, 2020.

[51] J. Liu, D. Xu, W. Yang, M. Fan, and H. Huang, "Benchmarking low-light image enhancement and beyond," *International Journal of Computer Vision*, vol. 129, no. 4, pp. 1153–1184, 2021.

[52] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[53] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[54] K. Zhang, L. V. Gool, and R. Timofte, "Deep unfolding network for image super-resolution," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 3217–3226.

[55] B. Zoph and Q. V. Le, "Neural architecture search with reinforcement learning," *arXiv preprint arXiv:1611.01578*, 2016.

[56] D. Rezende and S. Mohamed, "Variational inference with normalizing flows," in *International conference on machine learning*. PMLR, 2015, pp. 1530–1538.

[57] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.

[58] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," California Univ San Diego La Jolla Inst for Cognitive Science, Tech. Rep., 1985.

[59] C. Goller and A. Kuchler, "Learning task-dependent distributed representations by backpropagation through structure," in *Proceedings of International Conference on Neural Networks (ICNN'96)*, vol. 1. IEEE, 1996, pp. 347–352.

[60] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[61] C. Lu, S. Zheng, Z. Wang, O. Dib, and G. Gupta, "As-introvae: Adversarial similarity distance makes robust introvae," *arXiv preprint arXiv:2206.13903*, 2022.

[62] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision*. Springer, 2016, pp. 694–711.

[63] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on computational imaging*, vol. 3, no. 1, pp. 47–57, 2016.

[64] W. Yang, R. T. Tan, S. Wang, Y. Fang, and J. Liu, "Single image deraining: From model-based to data-driven and beyond," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 43, no. 11, pp. 4059–4077, 2020.

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

[66] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM Journal on Scientific Computing*, vol. 17, no. 1, pp. 227–238, 1996.

[67] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE transactions on image processing*, vol. 18, no. 11, pp. 2419–2434, 2009.

[68] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 586–595.

[69] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal processing letters*, vol. 20, no. 3, pp. 209–212, 2012.

[70] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3345–3356, 2015.

[71] C. Lee, C. Lee, and C.-S. Kim, "Contrast enhancement based on layered difference representation of 2d histograms," *IEEE transactions on image processing*, vol. 22, no. 12, pp. 5372–5384, 2013.

[72] C. Sakaridis, D. Dai, and L. Van Gool, "Acdc: The adverse conditions dataset with correspondences for semantic driving scene understanding," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10 765–10 775.

[73] W. Yang, Y. Yuan, W. Ren, J. Liu, W. J. Scheirer, Z. Wang, T. Zhang, Q. Zhong, D. Xie, S. Pu *et al.*, "Advancing image understanding in poor visibility environments: A collective benchmark study," *IEEE Transactions on Image Processing*, vol. 29, pp. 5737–5752, 2020.

[74] J. Cai, S. Gu, and L. Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 2049–2062, 2018.

[75] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3213–3223.

[76] T. S. Dee, "Teachers and the gender gaps in student achievement," *Journal of Human resources*, vol. 42, no. 3, pp. 528–554, 2007.

[77] W. Zhang, K. Ma, G. Zhai, and X. Yang, "Uncertainty-aware blind image quality assessment in the laboratory and wild," *IEEE Transactions on Image Processing*, vol. 30, pp. 3474–3486, 2021.

[78] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on image processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[79] Y. Fang, H. Zhu, Y. Zeng, K. Ma, and Z. Wang, "Perceptual quality assessment of smartphone photography," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3677–3686.

[80] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.

[81] J. Li, Y. Wang, C. Wang, Y. Tai, J. Qian, J. Yang, C. Wang, J. Li, and F. Huang, "Dsfd: dual shot face detector," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5060–5069.

[82] S. Lim and W. Kim, "Dslr: deep stacked laplacian restorer for low-light image enhancement," *IEEE Transactions on Multimedia*, vol. 23, pp. 4272–4284, 2020.

[83] K. Lu and L. Zhang, "Tbefn: A two-branch exposure-fusion network for low-light image enhancement," *IEEE Transactions on Multimedia*, vol. 23, pp. 4093–4105, 2020.

[84] S. Zheng, Y. Wu, S. Jiang, C. Lu, and G. Gupta, "Deblur-yolo: Real-time object detection with efficient blind motion deblurring," in *2021 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2021, pp. 1–8.

[85] Y. Sasagawa and H. Nagahara, "Yolo in the dark-domain adaptation method for merging multiple models," in *European Conference on Computer Vision*. Springer, 2020, pp. 345–359.

[86] W. Yang, X. Wang, A. Farhadi, A. Gupta, and R. Mottaghi, "Visual semantic navigation using scene priors," *arXiv preprint arXiv:1810.06543*, 2018.

[87] S. Xie, Z. Zheng, L. Chen, and C. Chen, "Learning semantic representations for unsupervised domain adaptation," in *International conference on machine learning*. PMLR, 2018, pp. 5423–5432.

[88] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[89] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8697–8710.

[90] J. Ke, Q. Wang, Y. Wang, P. Milanfar, and F. Yang, "Musiq: Multi-scale image quality transformer," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5148–5157.

[91] F. Lv, Y. Li, and F. Lu, "Attention guided low-light image enhancement with a large scale low-light simulation dataset," *International Journal of Computer Vision*, vol. 129, no. 7, pp. 2175–2193, 2021.

[92] Q. Chen, P. Montesinos, Q. S. Sun, P. A. Heng *et al.*, "Adaptive total variation denoising based on difference curvature," *Image and vision computing*, vol. 28, no. 3, pp. 298–306, 2010.

[93] Y. Zhao, Y. Xu, Q. Yan, D. Yang, X. Wang, and L.-M. Po, "D2hnet: Joint denoising and deblurring with hierarchical network for robust night image restoration," in *European Conference on Computer Vision*. Springer, 2022, pp. 91–110.

[94] W. Liu, G. Ren, R. Yu, S. Guo, J. Zhu, and L. Zhang, "Image-adaptive yolo for object detection in adverse weather conditions," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 2, 2022, pp. 1792–1800.

[95] Z. Zheng, Y. Wu, X. Han, and J. Shi, "Forkgan: Seeing into the rainy night," in *European conference on computer vision*. Springer, 2020, pp. 155–170.

[96] S. Hao, X. Han, Y. Guo, X. Xu, and M. Wang, "Low-light image enhancement with semi-decoupled decomposition," *IEEE transactions on multimedia*, vol. 22, no. 12, pp. 3025–3038, 2020.