Computer Vision and Image Understanding 178 (2019) 30–42

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Getting to know low-light images with the Exclusively Dark dataset

Yuen Peng Loh, Chee Seng Chan *

Centre of Image and Signal Processing, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, 50603, Malaysia

ARTICLE INFO

Communicated by Nikos Paragios

MSC:
41A05
41A10
65D05
65D17

ABSTRACT

Low-light is an inescapable element of our daily surroundings that greatly affects the efficiency of our vision. Research works on low-light imagery have seen a steady growth, particularly in the field of image enhancement, but there is still a lack of a go-to database as a benchmark. Besides, research fields that may assist us in low-light environments, such as object detection, has glossed over this aspect even though breakthroughs-after-breakthroughs had been achieved in recent years, most noticeably from the lack of low-light data (less than 2% of the total images) in successful public benchmark datasets such as PASCAL VOC, ImageNet, and Microsoft COCO. Thus, we propose the Exclusively Dark dataset to elevate this data drought. It consists exclusively of low-light images captured in visible light only, with image and object level annotations. Moreover, we share insightful findings in regards to the effects of low-light on the object detection task by analyzing the visualizations of both hand-crafted and learned features. We found that the effects of low-light reach far deeper into the features than can be solved by simple "illumination invariance". It is our hope that this analysis and the Exclusively Dark dataset can encourage the growth in low-light domain researches on different fields. The dataset can be downloaded at https://github.com/cs-chan/Exclusively-Dark-Image-Dataset.

1. Introduction

Low-light environment is an integral part of our everyday activities. As day change to night, the amount of available light decreases, causing our surroundings to be increasingly dark, and subsequently affecting our abilities to perform even menial tasks due to the lack of visibility. Computer vision research and systems aimed at assisting people in daily activities, as well as improve safety and security could be especially helpful in such conditions (Leo et al., 2017). However, low-light research commonly focus on the image enhancement problem that hardly relates to assistive systems, or night vision surveillance that demands costly hardware, whereas more related domains like object detection are seldom given attention. Though significant breakthroughs have been achieved one after another in the object detection domain, they evidently deal with bright images while significantly lacking for low-light. We believe this is largely due to a lack of available dataset to facilitate and benchmark the research in this area.

Well known public object datasets, PASCAL VOC (Everingham et al., 2010), ImageNet (Russakovsky et al., 2015a), and Microsoft COCO (Lin et al., 2014), played an integral role in the advancements as they provide large scale data for many researchers to work on or as challenges that promote progress in object detection and recognition. The PASCAL VOC is one of the earliest object datasets with comparatively large amounts of images at that time, consisting many variations that represent realistic environments during a time where object datasets suffer from simplicity and bias (Torralba and Efros, 2011). Since the launch of the dataset in 2006, it has facilitated the development of many handcrafted approaches for object centric applications (Felzenszwalb et al., 2008; Wang et al., 2010). In 2011, the rise of internet data mining has led to the collection of even larger scale data, prominently the ImageNet, that led to the breakthrough of deep learning using Convolutional Neural Network (CNN) (Krizhevsky et al., 2012), and subsequently sparked a whole new generation of deep learning works in computer vision and machine learning domain. While datasets continue to grow in numbers, a new challenge arises in the form of data annotation because it is difficult for the human annotators to cope with the sheer numbers. Then enters Microsoft COCO in 2014, while not as large in numbers as the ImageNet, it brings to the table, comprehensive annotation covering a variety of tasks which includes recognition, segmentation, and captioning. While the progress brought by these datasets are remarkable, there is a glaringly obvious lapse, that is, less than 2% of the images in these influential datasets are captured in low-light conditions. Moreover, there are no publicly available datasets that specifically provide natural low-light images for object focused works to the best of our knowledge.

We believe this shortage of data has impeded both the understanding and development of computer vision in low-light environments. Thus, we are committed and hope to move the field forward in this direction through the Exclusively Dark (ExDARK) dataset. It contains 7363 low-light images from very low-light environments to twilight, and 12 object classes annotated on both image class level and local object

* Corresponding author.
E-mail address: cs.chan@um.edu.my (C.S. Chan).

https://doi.org/10.1016/j.cviu.2018.10.010
Received 24 August 2017; Received in revised form 29 October 2018; Accepted 30 October 2018
Available online 22 November 2018

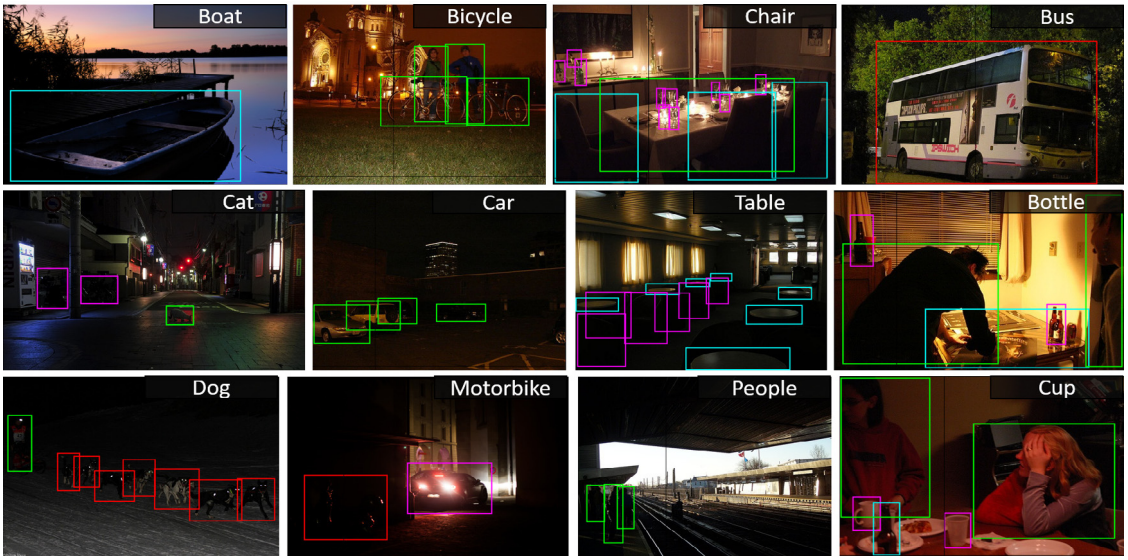

Fig. 1. Example images from the Exclusively Dark dataset with image and object level annotations. [Best viewed in color.]

bounding boxes, as shown in Fig. 1. We believe this database could facilitate a better understanding of the low-light phenomenon focusing on objects, unlike the current trend of low-light research works where limited samples were used for benchmarking enhancement algorithms, or camera dependent images like thermal imaging and near infra-red for surveillance that are costly and do not show realistic images.

This paper presents two contributions. First, we propose the Exclusively Dark (ExDARK) dataset which to the best of our knowledge, is the largest collection of natural low-light images taken in visible light to-date with object level annotation. Secondly, we provide an object-focused analysis of low-light images using the state-of-the-art algorithms in both hand-crafted and learned features for a better understanding of low-light vision and its difference from vision with sufficient illumination.

2. Related works

This section discusses the related work, particularly the common data used in low-light researches and renown public object datasets.

2.1. Low-light data

Works on low-light commonly address two different areas. The first is enhancement, where algorithms are proposed to improve the visibility of the contents in low-light images or videos. The second is in surveillance which can be categorized as detection tasks, but the data used differ greatly from typical object detection due to the use of different types of cameras.

Low-light image enhancement: The datasets used to benchmark enhancement works are commonly taken from those used for quality evaluation, but not necessarily a standard that is widely used in the low-light domain, such as the IVC database (Le Callet and Autrusseau, 2005) that were collected for general image enhancement works instead of low-light enhancement. As such, the images were synthetically darkened to simulate low-light so that the original image can be used as a groundtruth for comparison (Lim et al., 2015; Lore et al., 2017). There are also those who proposed datasets for enhancement but the amount is low, with less than 100 images (Wang et al., 2013) which is common for image quality works. Whereas some chose to combine datasets to obtain a larger variety for benchmarking (Fu et al., 2016b; Jung et al., 2017). It is also common to capture or download low-light images for qualitative assessments (Huang et al., 2013; Li et al., 2015; Fu et al., 2016a; Guo et al., 2017). Essentially, the datasets used are highly inconsistent. Nonetheless, recently two notable datasets for low-light image enhancement has been proposed, the See-in-the-Dark dataset (SID) (Chen et al., 2018) and LOw Light paired dataset (LOL) (Wei et al., 2018) that simulates low-light images by adjusting the camera exposure time and ISO. The SID provides 5094 short exposure images (low-light) corresponding to 424 long-exposure images (bright), whereas the LOL consists of 500 image pairs. While these datasets provide the much needed bright and low-light image pairs, the images do not represent real low-light environments such as nighttime. Moreover, the images are captured using specific cameras and do not contain dynamic objects because it is crucial for the image pairs to perfectly match. Contrarily, our proposed ExDARK dataset is made up of images captured in real low-light environments, contains dynamic objects such as cars, people, etc., and unconstrained by the imaging device.

Low-light denoising: Denoising is a notable subfield of low-light enhancement as noise is a significant problem in low-light visual data. Due to the nature of the environment where such data is captured, modern digital cameras rely on exposure timing and sensitivity settings to compensate for the lack of light which in turn bring about significant noise signals to the resultant images or videos. In related low-light enhancement researches, the noise problem is dealt with either as a post-processing of enhancement (Fu et al., 2016a; Guo et al., 2017; Shen et al., 2017) or incorporated into the enhancement process (Malm et al., 2007; Kim et al., 2015; Fu et al., 2016b; Su and Jung, 2017; Lore et al., 2017; Li et al., 2018). However, we note that the data, particularly images, used for these works are either synthetically generated, typically by adding Poisson noise and/or Gaussian noise into the synthetically darkened images without much consideration for accurate modeling of the noise content in contrast to real low-light data, or captured using specific cameras so as to have prior information for noise modeling. Hence, a specified low-light image database containing multitudes of images captured using unconstrained hardware will be the next challenge for such studies of noise in the low-light domain.

Low-light surveillance: Thermal and near infrared cameras are generally used to counter limited light for surveillance operations at night. Common object detection is not usually addressed in surveillance works, although the closest would be face recognition (Li et al., 2007; Kang et al., 2014) and pedestrian detection (Davis and Keck, 2005; Dong et al., 2007; Elguebaly and Bouguila, 2013; Qi et al., 2014; Zhao et al., 2015). Datasets such as the OTCVBS (Davis and Keck, 2005; Davis and Sharma, 2007; Li et al., 2007; Bilodeau et al., 2014), LSI (Olmeda et al., 2013), and LDHF (Kang et al., 2014) were acquired

with careful setup using sophisticated hardware that is much more difficult to obtain as compared to visible light images taken by general digital cameras. Moreover, the unrealistic images provided are not suitable for the practical understanding of low-light in common vision.

2.2. Popular object datasets

PASCAL VOC: The PASCAL VOC (Everingham et al., 2010) object dataset grew from 2005 till 2012, with annual challenges that encouraged researchers to develop ever improving algorithms to outdo one another in the spirit of progress. It began with only 4 object classes and 3787 images sourced from existing datasets. Initially containing simple object images, it has been continuously improved with more challenging images, and additional annotations. The last update to the dataset in 2012 puts the cumulative total at 26,305 images with 20 object classes, including annotations for object region of interest and segmentations.

ImageNet: ImageNet (Russakovsky et al., 2015a) was open to public in 2010 as the largest object image dataset, and gained great interest from the community especially in 2012 where its database of over 1 million images and 1000 image level object classes has allowed CNNs to be optimized and set a new benchmark in the object image classification task (Krizhevsky et al., 2012). The provided images are very challenging, where each of them is categorized into one of the object classes as long as there are instances of the object, regardless if the objects are either occluded or if the image contains other objects. Since then, ImageNet has become the de facto dataset for object image works, either as the main benchmark (Krizhevsky et al., 2012) or as the fundamental data for transfer learning (Donahue et al., 2014; Lee et al., 2017). In 2017, the dataset has reach new heights with more than 14 million images, and 1000 classes of which 200 of them has bounding box annotation for object detection tasks.

Microsoft COCO: The latest of notable object datasets is the Microsoft COCO (Lin et al., 2014), released in 2014. The quantity of images provided are not up to that of ImageNet, though its advantage is in the completeness of the image annotations. Specifically, 80 object classes annotated from bounding box for the detection task, to pixel level annotation for the segmentation task, as well as description of each image for the image captioning task. Similar to ImageNet, the content of the images are highly challenging where even a small instance of an object's part is annotated.

As a summary, despite that the three aforementioned datasets are challenging and large, the number of low-light images within them are considerably small, as shown in Table 1. In contrast, our proposed dataset, the ExDARK has 7363 images, inclusive of 223 images from our initial low-light pedestrian dataset (Loh and Chan, 2015), with 12 object classes annotated to the bounding box level. Although it is not massive, the low-light images would provide approximately 400% more than the low-light images found in the aforementioned datasets combined.

3. Exclusively Dark dataset

This section discusses (1) the motivation in establishing an object in low-light image dataset, (2) our observations on the handling of low-light images by past and present researches, and (3) the properties of the ExDARK.

Aspiration for low-light image data. A significant motivation in the effort to introduce a singular low-light image dataset is that there is none that is available to-date to set the standards for research in this domain. Even in low-light image enhancement works, real low-light images were mostly downloaded or captured on an ad hoc basis (Huang et al., 2013; Li et al., 2015; Fu et al., 2016a; Guo et al., 2017). On the other hand, large scale object datasets (Everingham et al., 2010; Russakovsky et al., 2015a; Lin et al., 2014) that claims data variations and generalization hardly provide enough low-light images, as shown in Table 1, to represent the true extend of environments and challenges faced in such conditions despite being an integral element in daily vision. Hence, with our proposed ExDARK, we hope to provide a staple collection of images for benchmarking low-light research works, and bring together different areas of expertise to focus on low-light conditions, for instance, image understanding, image enhancement, object detection, etc.

Table 1
Approximate number of low-light images in various public object datasets, and our proposed Exclusively Dark dataset.

Dataset		Low-light image [Amount (% from total)]
Microsoft COCO	Training	149 (0.18%)
	Validation	163 (0.4%)
	Testing 2014	138 (0.34%)
	Testing 2015	115 (0.14%)
	Total	565 (0.23%)
ImageNet	Training 2012	255 (0.02%)
	Validation 2012	38 (0.08%)
	Testing 2012	51 (0.05%)
	Validation 2013	12 (0.26%)
	Testing 2013	22 (0.23%)
	Training 2014	72 (0.12%)
	Total	450 (0.03%)
PASCAL VOC	2007	123 (1.24%)
	2008	72 (1.66%)
	2009	43 (1.58%)
	2010	50 (1.43%)
	2011	48 (1.32%)
	2012	17 (0.79%)
	Total	353 (1.34%)
Exclusively Dark	2018	7363 (100%)

Handling of low-light images. Based on our observations, we found that low-light is commonly glossed over in object dataset analyses (Everingham et al., 2015; Russakovsky et al., 2015b; Lin et al., 2014) with the preferred emphasis on object instances, scale, occlusion, and quantity. Therefore, it is not surprising that state-of-the-art object detectors, past and present (Felzenszwalb et al., 2008; Wang et al., 2010; Krizhevsky et al., 2012; Simonyan and Zisserman, 2014; He et al., 2016), were not designed nor were they analyzed, given the samples they had to work with. This has also indirectly led many researchers to oversimplify the diversity and challenges of low-light images. Considering very early computer vision works, such as well-known feature extractors (Lowe, 2004; Dalal and Triggs, 2005), had already strove for illumination invariance in their designs, it is understandable that many would consider illumination or low-light condition as just an auxiliary element to other challenges without going into a deeper understanding. Particularly, with the emergence of deep learning, machine learning is expected to be able to counteract this problem with ease. However, we show in our analyses in Section 4 that there is more to be studied than just relying on machine intelligence.

Knowing low-light images. We believe that the characterization of the low-light condition as just "illumination variation" is insufficient as the "variations" encompass much more. For example, low-light condition can emerge depending on the time of day (e.g. twilight, nighttime), location (e.g. indoor, outdoor), and the availability of light sources and their types (e.g. the sun, man-made lights). The combination of these three factors can create a great deal of disparity between image to image or even within an image itself. The impact of these variations has been left unexplained in most works, especially in object detection tasks, however, a grasp of their behavior can potentially advance the field. Though, rather than disregarding the milestones of researches so far, we simply believe that a gap has been overlooked in the common analysis, which we intend to fill in for a more thorough understanding of computer vision.

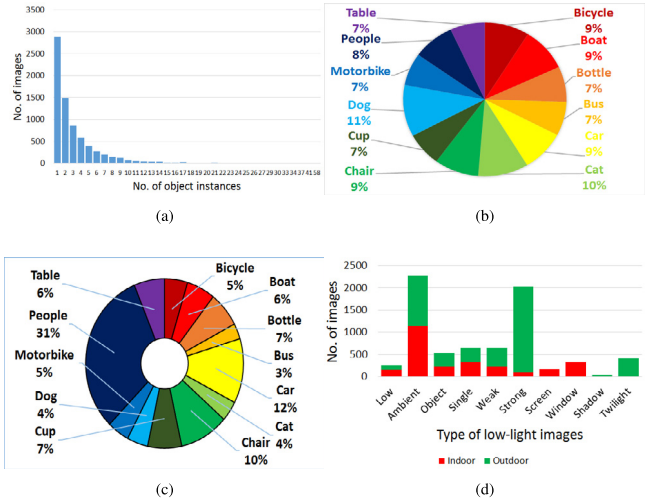

Fig. 2. Statistics of ExDARK dataset. (a) Object instances per image; (b) Fraction of image classes; (c) Object occurrence in dataset; (d) Image illumination types.[Best viewed in color.].

3.1. ExDARK dataset statistics

The ExDARK is a low-light object image dataset, where an image is categorized as low-light if it has either low or significant variations in illumination. The dataset currently has 7363 images with 12 object classes, namely *Bicycle, Boat, Bottle, Bus, Car, Cat, Chair, Cup, Dog, Motorbike, People,* and *Table*.

Data collection. We performed data collection from a variety of sources targeting the specified object classes. Most of the low-light images were downloaded from internet websites and search engines, namely *Flickr.com, Photobucket.com, Imgur.com, Deviantart.com, Gettyimages.com,* and *Google Search*. We used keywords related to low-light conditions, such as *dark, low-light, nighttime,* etc., to manually search and download the images.

We have also sub-sampled images from some public datasets, mainly PASCAL VOC, ImageNet, and Microsoft COCO, while there are also additional small amounts of images from other datasets (Russell et al., 2008; Philbin et al., 2008). Furthermore, we increased the variation of the images by extracting frames of low-light scenes from a collection of movies, as well as manually capture low-light images using different models of smart phones and digital cameras.

Object annotations. The collected data is annotated on two levels, the first is image class annotation where the images are sorted into the 12 classes based on the object instances only, regardless if the object is the majority in the image. Second is the bounding box annotation of the objects, where every instance of any of the 12 classes are annotated in all images using Piotr's Computer Vision Matlab toolbox (Dollár, 0000).

Fig. 2 shows the statistics of the number of images and their fraction with respect to the annotations. Most of the images provide a single instance of an object, but a considerable amount of the images have more instances. The highest number of bounding box annotations found in an image is 58, as shown Fig. 2a. Images that contain multiple instances can be a mixture of different objects, as shown in Fig. 1. While we kept a relatively balanced number of images in the image level annotation as shown in Fig. 2b, most of the bounding box annotations are from the *People* class, as seen in Fig. 2c. Among the total of 23,710 object instances annotated, 7460 are *People*, from single person to a crowd. We believe this would be useful for pedestrian detection work as well.

Types of low-light. From our collection of data, we have also identified 10 types of low-light conditions, in indoor and outdoor environments, that are commonly captured in images. Examples of the types are shown in Fig. 3 and explained as follows:

- **Low**: Images with very low illumination and hardly visible details.
- **Ambient**: Images with weak illumination and the light source is not captured within.
- **Object**: Images where there is/are brightly illuminated object[1](s) but surroundings are dark and the light source is not captured within.
- **Single**: Images where a single light source is visible.
- **Weak**: Images with multiple visible but weak light sources.
- **Strong**: Images with multiple visible and relatively bright light sources.
- **Screen**: *Indoor* images with visible bright screens (i.e. computer monitors, televisions).
- **Window**: *Indoor* images with bright windows as light sources.
- **Shadow**: *Outdoor* images captured in daylight but the objects are shrouded in shadows.
- **Twilight**: *Outdoor* images captured in twilight (i.e. time of day between dawn and sunrise, or between dusk and sunset).

We hope this categorization of low-light images will be valuable for future research, particularly for the low-light image enhancement domain, as identifying different illumination types could assist in the design of enhancement algorithms to handle the over and under enhancement problem accordingly. Fig. 2d shows the statistics of the different illumination types found in the ExDARK dataset.

Table 2
Number of images per object class used for analyses.

Dataset Class	Exclusively Dark Number of Image	Microsoft COCO Number of image
Bicycle	652	603
Boat	679	650
Bottle	547	650
Bus	527	564
Car	638	650
Cat	735	650
Chair	648	651
Cup	519	650
Dog	801	650
Motorbike	503	644
People	609	650
Table	505	650
Total	7363	7662

4. Analyzing image features in low-light

In this section, we look into the effectiveness of image features, commonly used in object tasks, on the ExDARK. In particular, we employ object proposal algorithms that make use of hand-crafted features (Zitnick and Dollár, 2014; Cheng et al., 2014; Fang et al., 2016), and object classification CNN (He et al., 2016) that learns features, to study their behavior in low-light images in comparison to bright images, as well as to gain new insights on this domain.

In our study, Microsoft COCO (MS-COCO) is used as the baseline dataset in our analysis. However, since the ExDARK has considerably less images compared to the MS-COCO, we sub-sampled bright images from MS-COCO for a fair comparison. Table 2 shows the number of images for each class of the ExDARK and the subset from MS-COCO that we have randomly extracted based on the classes of interest. For the MS-COCO images, only the annotations of the 12 chosen object classes are kept for the analysis while the rest are discarded. As a result, there are a total of 23,710 object instances in the ExDARK and 34,370 in the MS-COCO subset.

[1] The illuminated object is not necessarily from the 12 specified classes.

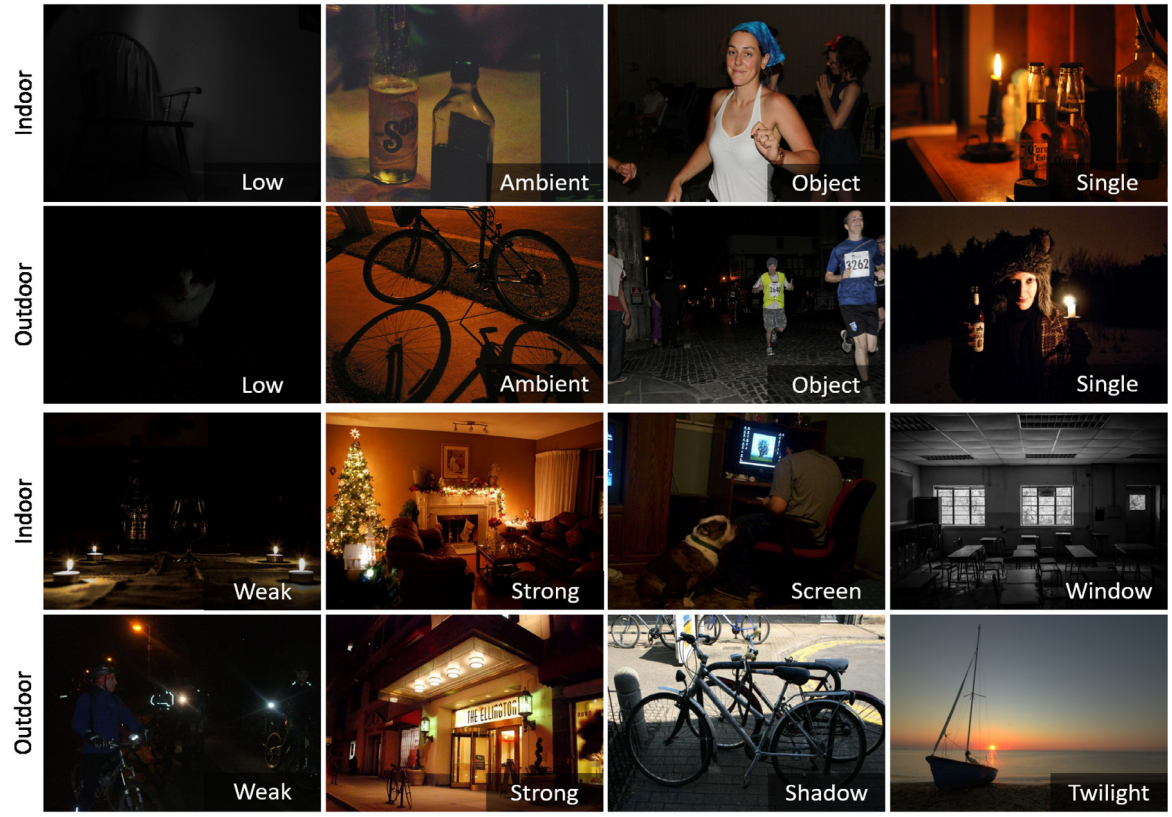

Fig. 3. Example of low-light image types in the ExDARK dataset.

4.1. Performance of hand-crafted features

Hand-crafted features are designed computations to extract meaningful information, based on established insights on the behaviors of the image contents, as opposed to learned features where computational models are trained to discover the meaningful information by itself. While the progress of deep learning in these few years has seen a shift in preference towards learned features, hand-crafted features are still employed, particularly for the object proposal task due to their high speed and low complexity nature. In this analysis, we intend to look into the abilities of classically hand engineered features when handling low-light images, thus we engage algorithms that use different types of features for our comparison, namely Edge Boxes[2] (Zitnick and Dollár, 2014), BING[3] (Cheng et al., 2014), and Adobe Boxes[4] (Fang et al., 2016), instead of deep learning based proposers (Ren et al., 2015; Redmon et al., 2016). A brief description of these methods are as follows:

- *Edge Boxes*, as stated in the name, proposes object bounding boxes by grouping *edges*, and uses the edge inside the bounding box to compute a score indicating the likelihood of object (objectness).
- *BING* is based on correlation between object boundaries and norm of image *gradients*. To this end, they implement SVM classification on the binarized norm gradients of bounding boxes to determine which box likely bounds a full object. Another SVM is then used on the SVM output scores to calibrate a final objectness score.
- *Adobe Boxes* uses groups of *superpixels* with high contrast from the background as the representation of object parts, named adobes, to propose object bounding boxes. The spatial concentration of adobes are used to calculate the objectness score. This method can also be used to refine proposals produced by other methods, which the paper shows works well when combined with BING (AdobeBING).

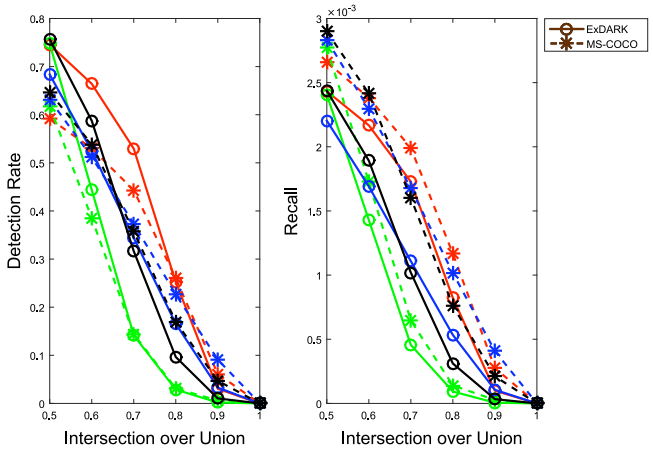

Fig. 4. Detection rate and recall of Edge boxes (red), BING (green), Adobe Boxes (blue), AdobeBING (black), at maximum proposal of 1000 boxes tested on ExDARK and MS-COCO.

4.1.1. Quantitative evaluation

This evaluation is to assess the ability of hand-crafted features to detect objects in both bright and low-light images, disregarding the identity of the objects. Experiments were performed to compare the detection (detections/groundtruths) and recall (detections/proposals) rates between the datasets using each proposal method. In the tests, all the methods were set to produce a maximum of 1000 bounding boxes, however the total could be less depending on the algorithms' ability to

[2] https://github.com/pdollar/edges.
[3] using implementation provided by Fang et al. (2016) of Adobe Boxes
[4] https://github.com/fzw310/AdobeBoxes-v1.0-/tree/master/AdobeBoxes(v1.0).

Table 3
Average proposals, average detections, detection rate, and recall of different proposal methods at maximum proposal of 1000 and IoU of 0.7.

Methods	Dataset	Avg. Prop./im	Avg. Det./im	Det. Rate	Recall
Edge boxes	MS-COCO	998	**1.9871**	0.4430	**0.0020**
	ExDARK	987	1.7050	**0.5295**	0.0017
BING	MS-COCO	1000	0.6457	0.1439	0.0006
	ExDARK	1000	0.4483	0.1392	0.0004
Adobe boxes	MS-COCO	1000	1.6753	0.3735	0.0017
	ExDARK	999	1.1039	0.3428	0.0011
Adobe BING	MS-COCO	1000	1.6010	0.3569	0.0016
	ExDARK	1000	1.0209	0.3170	0.0010

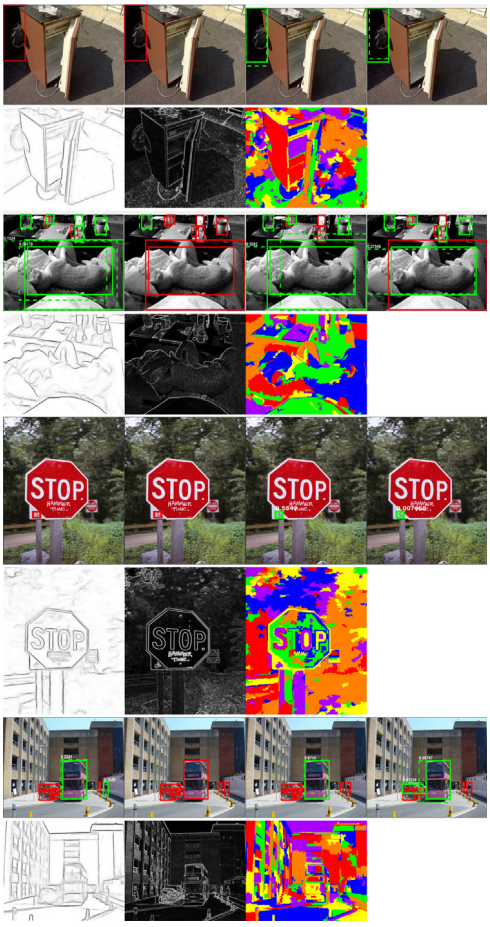

Fig. 5. MS-COCO images: Examples of proposals (top) and visualizations of their respective features (bottom). (Red: undetected groundtruth; Green: detected groundtruth, Green dotted: proposed box) From left: Edge Boxes, BING, Adobe Boxes, and AdobeBING. (Maximum proposals = 1000; IoU = 0.7.) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

confidently propose the boxes. As for the evaluation, the Intersection over Union (IoU) metric is used, where varying thresholds, from 0.5 to 1.0, were tested.

Implicitly, as the IoU increases, the detection rate and recall will reduce as the criteria to constitute a detection becomes stricter, as seen in Fig. 4. At lower IoU, the detection rate is higher for images from the ExDARK but the condition gradually inverts as the IoU increases. From the onset, the higher detection rate on the ExDARK seems to indicate more object detections, however, the results in Table 3 shows that the average detection in the low-light images are less than MS-COCO for all methods. Hence, we postulate that the reason for the observed higher detection rate is caused by the number of groundtruth where the images in MS-COCO contain more objects that remain undetected.

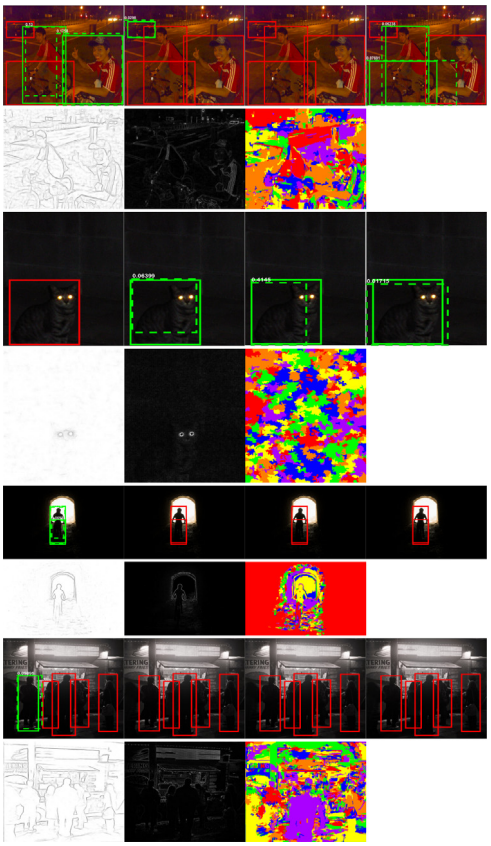

Fig. 6. ExDARK images: Examples of proposals (top) and visualizations of their respective features (bottom). (Red: undetected groundtruth; Green: detected groundtruth, Green dotted: proposed box) From left: Edge Boxes, BING, Adobe Boxes, and AdobeBING. (Max. proposals = 1000; IoU = 0.7.) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

These undetected objects can be attributed to the complexity of the MS-COCO images where many of the objects are too small, occluded, or only partially shown in the image, a common trait in challenging bright datasets. Whereas the images from ExDARK mostly contain the full objects where the main challenge comes from the illumination. Nonetheless, the low detection rate showed by ExDARK at higher IoU is also an indication that it is more challenging to get an accurate localization in low-light images as compared to bright images.

On the other hand, the recall rate on the ExDARK is obviously lower than the MS-COCO data using any of the studied methods. This result infers that most of the proposals in the low-light images are not valuable, even though the average proposal per image may be lower than that in MS-COCO, such as for the Edge Boxes and Adobe Boxes in Table 3.

4.1.2. Qualitative evaluation

We further study the results of different features by examining qualitative examples of both bright and low-light images in Figs. 5 and 6 respectively, as well as visualizations of the features used by the proposers.

In Fig. 5, we notice that the MS-COCO images have objects that are very small compared to the image size, which cause the studied methods, particularly the Edge Boxes and BING to fail. This is evident in their respective edge and gradient images, where the features are unable to capture the details of really small objects. On the other hand, Adobe Boxes and AdobeBING are better as superpixels are more precise in segmenting the objects from the background, but it still could not solve the problem.

On the contrary, the failures in the ExDARK are not due to object scale, but from factors related to low-light, as shown in Fig. 6. The first

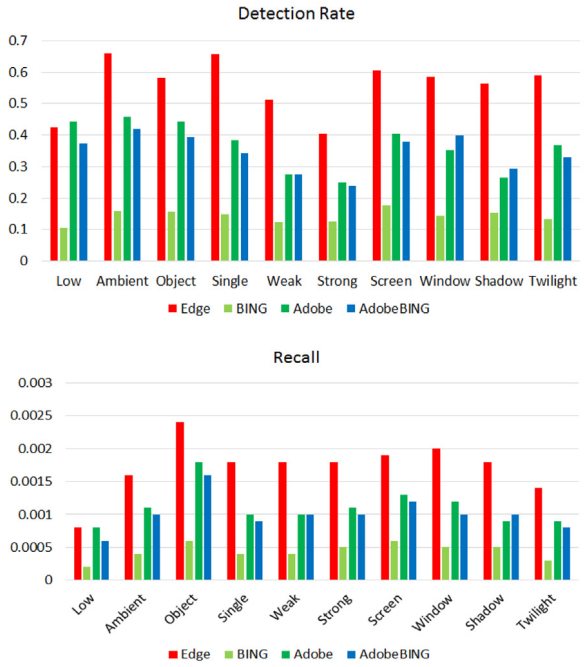

Fig. 7. Detection rate and recall of Edge Boxes, BING, Adobe Boxes, AdobeBING on ExDark dataset, sorted into different low-light image types. (Maximum proposals = 1000; IoU = 0.7.)

is due to the additional noise in low-light images that causes the failure due to interference from extra features, as seen in the first two row of images in Fig. 6. Even if there are successful proposals, we can see that the alignment is rather far from the groundtruth. These noises are usually caused by the high camera ISO setting used to compensate the low light level but at the same time it makes the camera oversensitive to the surrounding light. The other cause is the blending of the objects either to the background or to other objects, as seen in the last two rows of examples in Fig. 6. The methods are especially weak for these types of conditions because the gradient boundaries are unclear and the superpixels were unable to distinguish the difference between the low valued pixels of objects and backgrounds.

4.1.3. Further look into low-light

We take a further look into the detection and recall rates of each method separated into the 10 types of low-light images that we have established in Section 3.1. Fig. 7 shows the detection and recall rates, where Edge Boxes performs the best for all types of low-light conditions. Images with *Ambient* and *Single* lighting have the best detection rates, while *Low* and surprisingly, *Strong* lighting are the weakest. Whereas for the recall, the *Object* lighting type is the best while *Low* is the weakest. Fig. 8 shows examples of Edge Boxes detections in the different types of lighting.

The method performs quite well for the *Ambient* and *Single* light types because there are still enough light in the image to highlight the object features, particularly when the objects are nearer to the source of light. Whereas for very low light images, the objects are more likely to blend into the background. On the other hand, images taken in strongly lit low-light environments are expected to show more features, however, such environments are also more cluttered with objects and irregular light sources that result in complex images, subsequently deteriorating the detection performance.

Considering the recall, very low light images has the lowest value because either the contrast of the objects are too low for the object

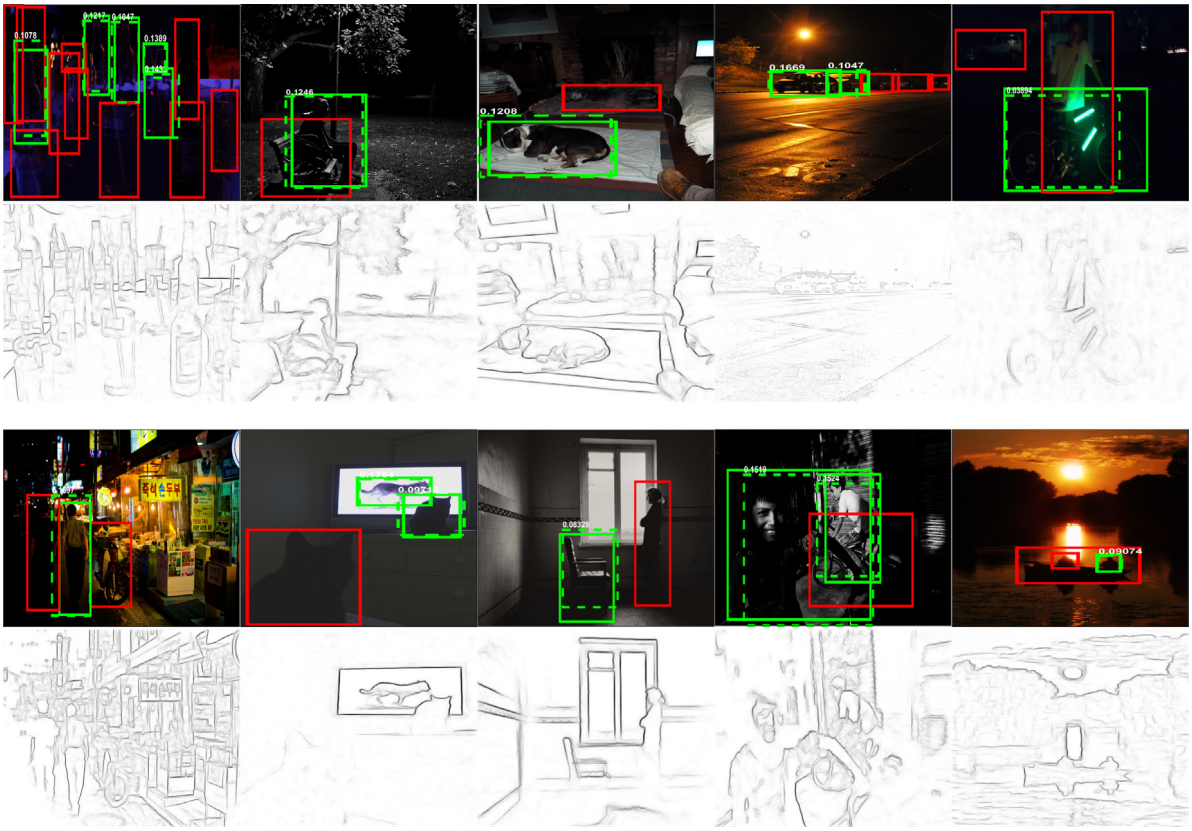

Fig. 8. Examples of Edge Boxes proposals (Max. proposals = 1000; IoU = 0.7) on different types of low-light images (top) and visualizations of their respective edge features (bottom). (Red: undetected groundtruth; and Green: detected groundtruth, Green dotted: proposed box) From left, first row: *Low, Ambient, Object, Single, Weak*; and second row: *Strong, Screen, Window, Shadow, Twilight*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

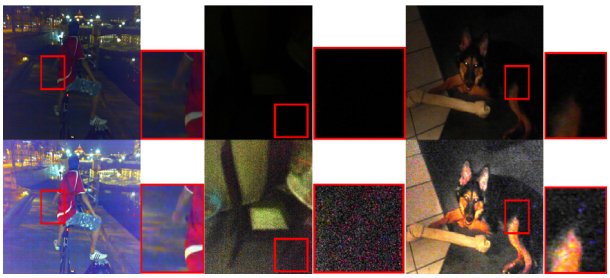

Fig. 9. Examples of noisy low-light images (top) and respective enhanced counterparts using LIME (Guo et al., 2017) (bottom) to show the severity of the signals. The red bounding boxes show the zoom-in areas of significant noisy signals . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

features to be extracted, or the image is saturated with noise due to the camera's high ISO setting. Images with a well illuminated object but low-light surroundings, give the best recall because the well lit object will mostly be detected even if the other objects in the low-light background are missed, hence aiding in the recall evaluation. For the most part, the detection rates using these hand-crafted approaches are below 70% for any type of low-light conditions, which leaves room for improvement towards a good low-light object detection system.

4.1.4. The noise problem

As found from the analyses in Sections 4.1.2 and 4.1.3, noise is a notable component in the low-light images. Moreover, various enhancement works distinctly discuss and address the problem by either employing existing off-the-shelf denoising algorithms as a post-processing step (Fu et al., 2016a; Guo et al., 2017; Shen et al., 2017) or incorporating mechanisms into the proposed enhancements (Lee et al., 2005; Malm et al., 2007; Kim et al., 2015; Fu et al., 2016b; Su and Jung, 2017; Lore et al., 2017; Li et al., 2018). These works brought forward various noise types that exist in low-light images which need to be addressed and the following three are the most noteworthy.

Poisson noise: Due to the nature of imaging devices that is discrete, in environments that have extremely low light it is necessary to increase the signal acceptance level but even so, the number of photons captured by the device's sensors fluctuate randomly which causes the resultant image to have noise. This noise signals conform to the Poisson distribution model, hence called Poisson noise.

Gaussian noise: Many works tend to address image noise using Gaussian-based denoisers due to the strategy using white Gaussian noise with unit variance to approximate Poisson noise (Remez et al., 2017). Moreover, Lee et al. (2005) note that portions of the Poisson noise behave similarly to Gaussian noise.

False color noise (FCN): FCN is especially noticeable to human observation as they appear as random pixels of varying colors that do not belong to the natural appearance of an image. This noise can be attributed to the clipping of color filters in the analog-to-digital conversion (ADC) of the signals.

Fig. 9 shows examples of low-light images from the ExDARK containing significant amounts of the aforementioned noise signals. As the signals are hardly noticeable on the original low-light images due to low intensity, we brightened them using the LIME algorithm (Guo et al., 2017) to emphasize the signals for easier observation. It is clearly noticeable that the noise significantly affects the quality of the images and possibly degrade the performance of features. Therefore, a test and analysis is conducted to ascertain their impact on hand-crafted object features. Specifically, the experiments in Section 4.1 were repeated on denoised ExDARK data. The BM3D (Dabov et al., 2007) was chosen as the denoiser for its performance and also due to its common application

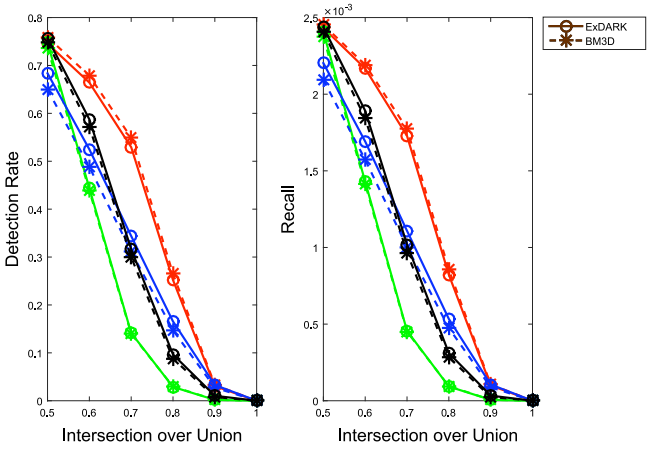

Fig. 10. Detection rate and recall of Edge boxes (red), BING (green), Adobe Boxes (blue), and AdobeBING (black), at maximum proposal of 1000 boxes tested on ExDARK and ExDARK denoised by BM3D (BM3D).

Table 4
Average proposals, average detections, detection rate, and recall of proposal methods tested on ExDARK and ExDARK denoised by BM3D (BM3D) at maximum proposal of 1000 and IoU of 0.7.

Methods	Dataset	Avg. Prop./im	Avg. Det./im	Det. Rate	Recall
Edge Boxes	ExDARK	987	1.7050	0.5295	0.0017
	BM3D	997	**1.7686**	**0.5492**	**0.0018**
BING	ExDARK	1000	0.4483	0.1392	0.0004
	BM3D	1000	0.4504	0.1399	0.0005
Adobe Boxes	ExDARK	999	1.1039	0.3428	0.0011
	BM3D	999	1.0024	0.3113	0.0010
Adobe BING	ExDARK	1000	1.0209	0.3170	0.0010
	BM3D	1000	0.9648	0.2996	0.0010

in low-light enhancement post-processing (Fu et al., 2016a; Guo et al., 2017; Shen et al., 2017).

Based on the quantitative results shown in Fig. 10 and Table 4, there is only a minor improvement for the Edge Boxes, and worse, degrades the results of BING, Adobe Boxes and AdobeBING. Fig. 11 shows some examples of their features before and after denoising. For the Edge Boxes, it can be seen that the denoising improves the edge features of the objects in the image which contributed to the better performance, however, there seems to be an increase of artifacts. A similar behavior is observed from the superpixel features used by Adobe Boxes where there are clear box-like artifacts that have occluded the object and degraded the performance. As for BING, the BM3D managed to reduce some noise but the effect is insignificant, which is in agreement with the quantitative observation.

In summary, we can deduce a few inferences from these findings. First, denoising is only able to assist some features, such as edges, where it brings out some features but at the same time it may increase artifacts. This is mainly due to the nature of the BM3D algorithm that uses "blocks" filtering that is not designed for low-light conditions. Secondly, the detection rate only improved by a small margin after denoising. This indicates that the challenge for computer vision tasks in low-light is not only due to noise, but also the lack of signals in low-light conditions. Thus, these two paths: (a) denoising for low-light data; and (b) low-light enhancement that retrieves informative signals, are potential directions for research growth.

4.2. Insights from learned features

As we have explored hand-crafted features in Section 4.1, here we explore the capabilities of learned features in low-light. In contrast to hand-crafted features, learned features rely on the computation of

Fig. 11. Comparison between low-light images (Top) and their BM3D denoised counterpart (Bottom) with the visualization of their respective features used for object proposal. From left: Edge Boxes, BING, and Adobe Boxes (Red: undetected groundtruth; Green: detected groundtruth, and Green dotted: proposal box).

machine learning algorithms to uncover the best representations for a given task. At first, the features learned remain largely unknown as we could not fully comprehend the high dimensional representations generated by machines. Nevertheless, many works had since visualized high dimensional data and features (Donahue et al., 2014; Zeiler and Fergus, 2014; Mahendran and Vedaldi, 2015; Yosinski et al., 2015; Lee et al., 2017) to understand and find out what the machines "see".

In this section, we attempt to uncover the features in low-light images by visualizing a straight forward object image classification CNN, as opposed to the more intricate object detection networks. Specifically, we fine-tuned the pre-trained Resnet-50 model (He et al., 2016), on the Microsoft COCO and ExDARK data, and evaluated their performance based on different ratios of bright and dark data used in the fine-tuning. Then, we look into the behavior of the learned representations in two ways. First, the t-SNE (Maaten and Hinton, 2008) is used to visualize a 2D mapping on the clustering behavior achieved by the learned feature vectors. Second, the visualization of the activations in convolution maps corresponding to the spatial location on the images (Yosinski et al., 2015) in order to find out which part of an image "triggers" the classification outcome, i.e. the attention of the network.

4.2.1. Classification performance

It is commonly agreed that CNN performs better when trained with more general data, i.e. very large numbers of images with complex variations. However, on account that the amount of images in the ExDARK is still too small to train a full CNN model from scratch, we approach the task by fine-tuning the existing Resnet-50 model that is pre-trained using ImageNet. The Resnet model is chosen for this task because it is currently one of the top performing architectures in both the ILSVRC and Microsoft COCO challenges.

The training setup of the experiments include replacing the last classification layer of the pre-trained Resnet-50 model which has 1000 object classes for the ImageNet into the 12 object classes of the experimented dataset. The learning rate of this new layer is set as 0.001, while the pre-trained layers have a lower learning rate of 0.0001, and they are kept constant throughout the training. The optimization scheme used is the Stochastic Gradient Descent with batch size of 32. The pre-processing of the training data includes augmentation by cropping and jittering for better model generalization, as well as subtracting with the training dataset's mean RGB image as normalization. All of the models used were trained for 50 epochs.

The data stated in Table 2 are used for the experiments. We set aside 400 images per object class for the training, where 250 of them were used to fine-tune the model and 150 were used for validation. Hence, both the Microsoft COCO and ExDARK provide 4800 training images each, while the remaining 2862 and 2563 respectively make up the test set. Table 5 shows baseline results of models trained using the subset and all relevant data of the Microsoft COCO[5] training and validation set, with and without fine-tuning. It can be seen that the performance is clearly lacking when classifying low-light data.

Table 5
Accuracy of Resnet-50 models trained using all relevant data from the Microsoft COCO and the extracted subset detailed in Table 2, with and without fine-tuning. MS-COCO: performance on Microsoft COCO test images only, ExDARK: performance on ExDARK test images only, Overall: performance on test images of both sets.

Training data	Fine-tuning	Test accuracy		
		MS-COCO	ExDARK	Overall
All	No	54.82%	40.27%	47.94%
All	Yes	61.60%	50.84%	56.52%
Subset	No	54.16%	34.57%	44.90%
Subset	Yes	62.75%	43.15%	53.49%

Table 6
Accuracy of Resnet-50 models fine-tuned using different ratios of bright images (Microsoft COCO) and low-light images (ExDARK). MS-COCO: performance on Microsoft COCO test images only, ExDARK: performance on ExDARK test images only, Overall: performance on test images of both sets.

Model	Training ratio	Test accuracy		
	MS-COCO:ExDARK	MS-COCO	ExDARK	Overall
1	10:0	62.75%	43.15%	53.49%
2	9:1	**63.31%**	48.89%	56.50%
3	8:2	62.16%	52.75%	57.71%
4	7:3	61.25%	55.05%	58.32%
5	6:4	61.50%	55.64%	58.73%
6	5:5	61.18%	58.45%	**59.89%**
7	4:6	59.89%	58.99%	59.47%
8	3:7	58.00%	59.54%	58.73%
9	2:8	57.27%	61.45%	59.24%
10	1:9	55.38%	62.27%	58.64%
11	0:10	46.30%	**62.58%**	53.99%

Our main experiments use different ratios of bright to low-light images, from 10:0 (only bright images) to 0:10 (only low-light images) maintaining the same overall number of training images, to fine-tune different models and observe the classification outcomes on the same independent testing data and the test results are shown in Table 6.

A few inferences can be drawn from these results. First, the notion that the illumination variation of low-light can be addressed in the same manner as noise (as training data augmentation) is improper. As we can see in the results, the models that were fine-tuned with less amount of low-light images are weaker at classifying them, and gradually increases with the ratio, indicating dataset dependency. On the other hand, we had a presumption that balanced or generalized training data would enable the model to learn features that are mutually useful for both types of images and subsequently achieve best classification performance, but

[5] Images that contain the 12 objects as in ExDARK.

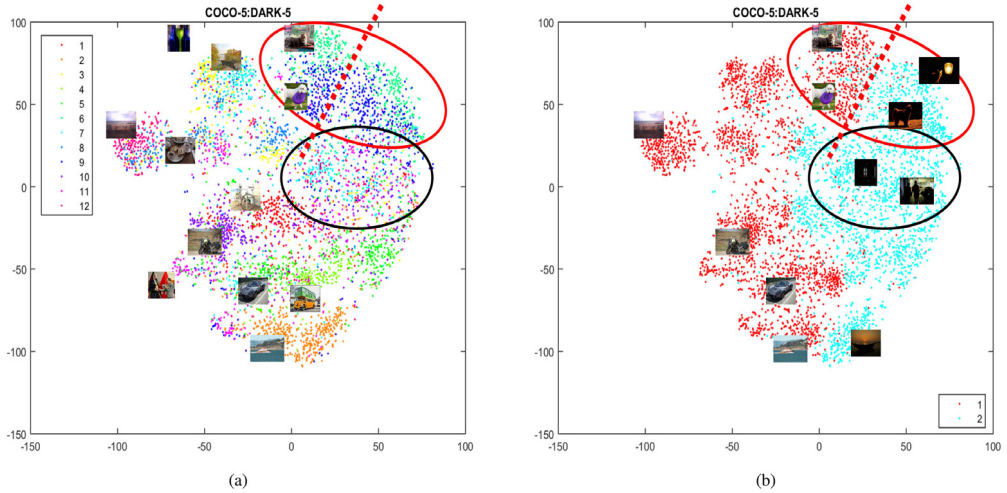

Fig. 12. t-SNE embedding of features vectors from Resnet-50 fine-tuned on 5:5 ratio of bright and low-light images. (a) Class 1–12: Bicycle, Boat, Bottle, Bus, Car, Cat, Chair, Cup, Dog, Motorbike, People, and Table; (b) Type 1–2: Bright (MS-COCO), and Low-light (ExDARK) images. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

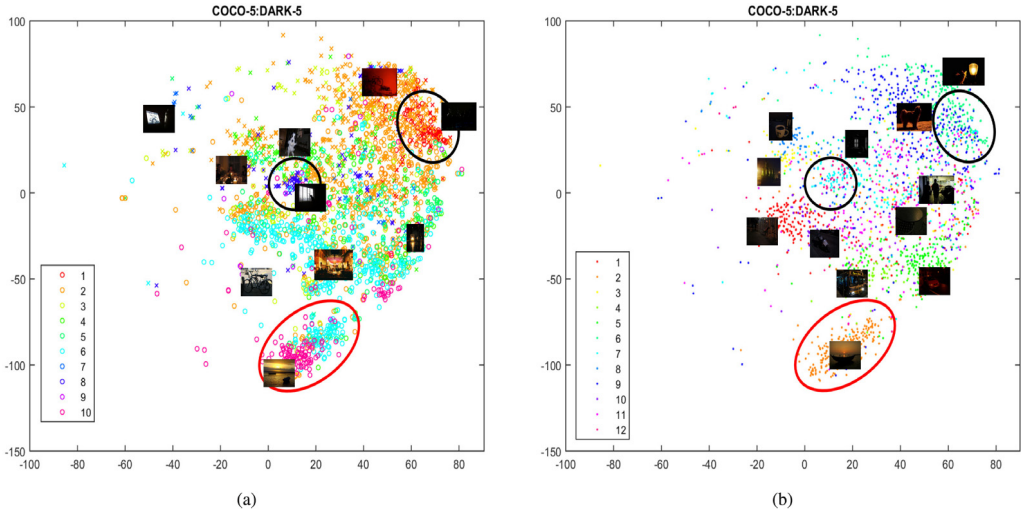

Fig. 13. t-SNE embedding of feature vectors from Resnet 50, fine-tuned on 5:5 ratio low-light images. (a) Separated by indoor ('x') and outdoor ('o') and color coded by the type of light conditions, 1–10: Low, Ambient, Object, Single, Weak, Strong, Screen (indoor only), Window (indoor only), Shadow (Outdoor only), and Twilight (outdoor only); (b) Color coded by classes, Class 1–12: Bicycle, Boat, Bottle, Bus, Car, Cat, Chair, Cup, Dog, Motorbike, People, and Table. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the results indicate otherwise. While the overall classification accuracy of Model 6 is the best, it appears to be a trade-off result as its performance is no better than a model specifically trained and tested on either bright (Model 1) or low-light (Model 11) images, even though they are addressing the same classification task. Hence, we bring forth the following two deductions: (1) the dataset dependent performance concurs the necessity of a low-light only dataset, and (2) the observation that a balanced training data did not raise the overall performance suggests bright and low-light data belong to different clusters that requires separate modeling. We are keen to explore further into the features to understand and verify this behavior.

Training amount influence. Additionally, we inspected the influence of data amount on the performance, specifically by training two additional models, A and B, by varying the image amount as shown in Table 7. Model A was trained using the same ratio as Model 6, but with all available training images of the subset, i.e. doubling the total training images used for Model 6. On the other hand, Model B is trained using only the low-light images but half the amount of those used to train Model 11. As seen in Table 7, the performance improves with more training data as shown by Model A and deteriorates when the data is reduced as in Model B. This is in line with the notion that CNNs require more data to improve its performance. However, the important message from this finding is that more low-light data is indeed needed to boost the performance of such systems.

Table 7
Accuracy of Resnet-50 models fine-tuned using different ratios of bright images (Microsoft COCO) and low-light images (ExDARK). MS-COCO: performance on Microsoft COCO test images only, ExDARK: performance on ExDARK test images only, Overall: performance on test images of both sets.

Model	Training ratio	Test accuracy		
	MS-COCO:ExDARK	MS-COCO	ExDARK	Overall
6	5:5	61.18%	58.45%	59.89%
A	10:10	**63.31%**	**63.71%**	**63.50%**
11	0:10	46.30%	62.58%	53.99%
B	0:5	40.46%	55.52%	47.58%

(a) Bicycle (b) Bottle (c) Car (d) Boat (e) Bus

(f) Bottle (g) Cup (h) Bottle (i) Car (j) People

(k) Cat (l) Dog (m) Bus (n) Chair (o) Cat

Fig. 14. Test images (top) and the visualization of activation maps (bottom). (a)–(e) Correctly classified low-light images; (f)–(j) Misclassified low-light images; (k)–(o) Misclassified bright images. (Classification results in sub-caption; and the correct (groundtruth) class labels are: (f) *Cat*, (g) *Chair*, (h) *Cup*, (i) *Dog*, (j) *Motorbike*, (k) *Motorbike*, (l) *People*, (m) *Dog*, (n) *Table*, (o) *Bicycle*). [Best viewed in color.].

4.2.2. Feature analysis with t-SNE

We look into the features learned by the Resnet-50 model fine-tuned on 5:5 data ratio (Model 6) using the t-SNE algorithm[6] (Maaten and Hinton, 2008). In a classification CNN, the output produced by the last convolution layer is the high level representation that is used by the subsequent fully connected layers that act as the classifier. Hence, to study the behavior of the high level features, we extracted the feature vectors of the last pooling layer of Model 6 when classifying the testing images. The t-SNE is then used to reduce these $1 \times 1 \times 2048$ dimension feature vectors into a 2-dimension embedding which shows the relationship between the features.

Fig. 12 shows the embedding of the test images generated by t-SNE and color coordinated by the object classes, and image types. Noticeable grouping of the object classes can be seen in Fig. 12a, and classes that are relatively similar, such as Cat (5-green) and Dog (9-dark blue) are grouped closely, as circled in red. We deduced that the learned features are able to capture high level abstraction of objects, though considerable amounts of confusion are still present, as seen by the mixture of colors circled in black.

We further look into the feature embeddings from another perspective by marking the scatter points with two colors to show the bright and low-light images as in Fig. 12b. Surprisingly, it shows a clear separation between bright images from the MS-COCO dataset (red) with the low-light images of ExDARK (blue). This observation is interesting because the feature vectors visualized are high level representations used to achieve object classification, whereas the brightness or intensity difference in images should be a low level feature. In our initial intuition, we believed that the training data normalization, and the data progression through the layers of the CNN towards high level abstraction should have normalized and disregarded the brightness between bright and low-light data as it is not a crucial feature for the classification of objects. However, the t-SNE embedding shows otherwise, which is a clear indication that even though the model is trained on both types of image for the same task, the features learned are inherently different. For example, the region for Cat and Dog classes (circled in red) has a distinct split (red dotted line). Moreover, the region that do not have a distinct clustering of classes (circled in black) are found within the low-light image cluster, thus pointing out that the features learned for low-light images may not be as robust as those for bright images.

Furthermore, we examined the t-SNE embedding by color coordinating the scatter plot based on the types of low-light, as well as differentiating them by indoor and outdoor environments, as illustrated in Fig. 13. Firstly, the CNN features seem to be able to distinguish indoor and outdoor by a small degree of confidence. We can see that the indoor images seem to cluster to the upper half of the embedding while the outdoor images are scattered throughout. On the other hand, the CNN features appear to have the ability to distinguish certain types of low-light images, such as *Low* (1-red), *Strong* (6-light blue), and *Twilight* (10-pink), though this ability may interfere with its robustness for the object classification task. As we show in the comparison between Figs. 13a and 13b, the clustering of *Low* (1-red) and *Window* (8-dark blue) illumination type features (circled in black) have caused confusion to *Cat*, *Chair*, *Dog*, and *People* object classes. However, the clustering of the features may be stronger for the classification task, such as the *Boat* class cluster (circled in red) grouping both *Strong* and *Twilight* images together, though a separation can still be seen. Hence, we surmise that CNN model unwittingly learns low-light properties which can be a hindrance to the object classification task.

4.2.3. Attention analysis with activation maps

In this section, we delve into the activation maps of the trained model to find out its attention when performing the classification, and verify if low-light elements are an influence to it. Specifically, we chose to visualize the activation maps before the last pooling layer of Model 6 (last convolution output before the fully connected layer), so that the spatial location of the activations are preserved.

The visualization is done by first extracting the $7 \times 7 \times 2048$ dimension activation maps of the model when classifying an image. These maps are then aggregated into a single map by selecting the maximum value among the maps for every spatial location. Thus, the resultant aggregated map will have high values for locations that are either highly activated or gives high contribution to the classifier. This map is then resized to the original image's dimensions and superimposed onto the image, whereby we will be able to visualize the model's attention on the image that led to the classification result.

Fig. 14 shows a few examples of the classified test images and their respective activation regions. Our analysis found that in low-light images, the attention of the model are often drawn to the bright sources of light, either partially or entirely. For example, the activation maps of the correctly classified images in Figs. 14a–14e shows that while the main attention is on the object of interest, the light sources are either within the attention (Figs. 14a–14c) or directly shine on the objects (Figs. 14d–14e). While the model can "overlook" the light sources, like in Fig. 14e, there are many cases, such as Figs. 14f–14j, where the attention of the model is overtaken by the brightest areas and causes misclassification. Yet this is not an issue for bright images, where the misclassification is commonly due to the attention being on another object instead of the labeled class, as shown in Figs. 14k–14o.

5. Summary and conclusion

In this paper, the Exclusively Dark dataset is introduced in hopes of providing a go-to database for low-light research works and also to encourage the community to look into the challenges of low-light environments that has long been glossed over, especially in application based researches such as object detection. Unlike common object datasets, the Exclusively Dark consists fully of low-light images captured in visible light with image and object level annotations of up to 12 classes, as well as a distinction of up to 10 types of low-light conditions.

Using this dataset, we performed an extensive analysis of low-light images from the perspective of object detection by digging deep into the behavior of common features, both hand-crafted and learned, in which we found interesting insights. We found that the design of hand-crafted features are mainly for bright conditions, thus unable to adequately address cases of noise and lack of details that frequently exist in low-light images. Similarly, a state-of-the-art denoising algorithm is also insufficient to handle the noise that frequently occurs alongside low-light data.

Conversely, our investigation into learned features by training CNNs using both bright and low-light data indicated that, indeed the number of low-light data should be increased for better performance in low-light conditions. Furthermore, by visualizing the feature vectors and activation maps of a CNN, we have come to understand that low-light "alters" object features, i.e. the same object in bright and low-light yields amply different features. Moreover, the irregularity of illumination greatly challenges the attention of features that is not found in bright environments. Therefore, object detection in low-light is not to be trifled with lightly, but instead requires careful consideration and a dedicated dataset is needed to push progress forward.

While our study has been focused on object detection based feature analysis, we believe there are more to be unraveled in the low-light domain. For this reason, we hope the Exclusively Dark to be a valuable database for future ventures, either to further understand the vision behavior or improve the performance of practical tasks in low-light.

Acknowledgments

This research is supported in part by the Fundamental Research Grant Scheme through the Ministry of Education Malaysia under Grant FRGS/1/2018/ICT02/UM/02/2 and in part by the Postgraduate Research Fund through the University of Malaya under Grant PG002-2016A. Also, we gratefully acknowledge the support of NVIDIA Corporation (United States) with the donation of the GPU used for this research.

[6] https://lvdmaaten.github.io/tsne/.

References

Bilodeau, G.A., Torabi, A., St-Charles, P.L., Riahi, D., 2014. Thermal–visible registration of human silhouettes: A similarity measure performance evaluation. Infrared Phys. Technol. 64, 79–86.

Chen, C., Chen, Q., Xu, J., Koltun, V., 2018. Learning to see in the dark. In: Computer Vision and Pattern Recognition (CVPR), 2018 IEEE Conference on.

Cheng, M.M., Zhang, Z., Lin, W.Y., Torr, P., 2014. Bing: Binarized normed gradients for objectness estimation at 300fps. In: Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. pp. 3286–3293.

Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K., 2007. Image denoising by sparse 3-d transform-domain collaborative filtering. IEEE Trans. Image Process. 16, 2080–2095.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. In: Computer Vision and Pattern Recognition (CVPR), 2005 IEEE Conference on. IEEE, pp. 886–893.

Davis, J.W., Keck, M.A., 2005. A two-stage template approach to person detection in thermal imagery. In: Application of Computer Vision, 2005. WACV/MOTIONS'05 Volume 1. Seventh IEEE Workshops on. IEEE, pp. 364–369.

Davis, J.W., Sharma, V., 2007. Background-subtraction using contour-based fusion of thermal and visible imagery. Comput. Vision Image Understanding 106, 162–182.

Dollár, P., Piotr's Computer Vision Matlab Toolbox (PMT). https://github.com/pdollar/toolbox.

Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T., 2014. Decaf: A deep convolutional activation feature for generic visual recognition. In: International Conference on Machine Learning. pp. 647–655.

Dong, J., Ge, J., Luo, Y., 2007. Nighttime pedestrian detection with near infrared using cascaded classifiers. In: Image Processing (ICIP), 2007 IEEE International Conference on. IEEE, pp. VI–185.

Elguebaly, T., Bouguila, N., 2013. Finite asymmetric generalized gaussian mixture models learning for infrared object detection. Comput. Vision Image Understanding 117, 1659–1671.

Everingham, M., Eslami, S.A., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A., 2015. The pascal visual object classes challenge: A retrospective. Int. J. Comput. Vis. 111, 98–136.

Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A., 2010. The pascal visual object classes (voc) challenge. Int. J. Comput. Vis. 88, 303–338.

Fang, Z., Cao, Z., Xiao, Y., Zhu, L., Yuan, J., 2016. Adobe boxes: Locating object proposals using object adobes. IEEE Trans. Image Process. 25, 4116–4128.

Felzenszwalb, P., McAllester, D., Ramanan, D., 2008. A discriminatively trained, multi-scale, deformable part model. In: Computer Vision and Pattern Recognition (CVPR), 2008 IEEE Conference on. IEEE, pp. 1–8.

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

Fu, X., Zeng, D., Huang, Y., Zhang, X.P., Ding, X., 2016b. A weighted variational model for simultaneous reflectance and illumination estimation. In: Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on. pp. 2782–2790.

Guo, X., Li, Y., Ling, H., 2017. Lime: Low-light image enhancement via illumination map estimation. IEEE Trans. Image Process. 26, 982–993.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on. pp. 770–778.

Huang, S.C., Cheng, F.C., Chiu, Y.S., 2013. Efficient contrast enhancement using adaptive gamma correction with weighting distribution. IEEE Trans. Image Process. 22, 1032–1041.

Jung, C., Yang, Q., Sun, T., Fu, Q., Song, H., 2017. Low light image enhancement with dual-tree complex wavelet transform. J. Vis. Commun. Image Represent. 42, 28–36.

Kang, D., Han, H., Jain, A.K., Lee, S.W., 2014. Nighttime face recognition at large standoff: Cross-distance and cross-spectral matching. Pattern Recognit. 47, 3750–3766.

Kim, M., Park, D., Han, D.K., Ko, H., 2015. A novel approach for denoising and enhancement of extremely low-light video. IEEE Trans. Consum. Electron. 61, 72–80.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems. pp. 1097–1105.

Le Callet, P., Autrusseau, F., 2005. Subjective quality assessment irccyn/ivc database. http://www.irccyn.ec-nantes.fr/ivcdb/.

Lee, S.H., Chan, C.S., Mayo, S.J., Remagnino, P., 2017. How deep learning extracts and learns leaf features for plant classification. Pattern Recognit. 71, 1–13.

Lee, S.W., Maik, V., Jang, J., Shin, J., Paik, J., 2005. Noise-adaptive spatio-temporal filter for real-time noise removal in low light level images. IEEE Trans. Consum. Electron. 51, 648–653.

Leo, M., Medioni, G., Trivedi, M., Kanade, T., Farinella, G.M., 2017. Computer vision for assistive technologies. Comput. Vision Image Understanding 154, 1–15.

Li, S.Z., Chu, R., Liao, S., Zhang, L., 2007. Illumination invariant face recognition using near-infrared images. IEEE Trans. Pattern Anal. Mach. Intell. 29, 627–639.

Li, M., Liu, J., Yang, W., Sun, X., Guo, Z., 2018. Structure-revealing low-light image enhancement via robust retinex model. IEEE Trans. Image Process. 27, 2828–2841.

Li, L., Wang, R., Wang, W., Gao, W., 2015. A low-light image enhancement method for both denoising and contrast enlarging. In: Image Processing (ICIP), 2015 IEEE International Conference on. IEEE, pp. 3730–3734.

Lim, J., Kim, J.H., Sim, J.Y., Kim, C.S., 2015. Robust contrast enhancement of noisy low-light images: Denoising-enhancement-completion. In: Image Processing (ICIP), 2015 IEEE International Conference on. IEEE, pp. 4131–4135.

Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft coco: Common objects in context. In: European Conference on Computer Vision. Springer, pp. 740–755.

Loh, Y.P., Chan, C.S., 2015. Unveiling contrast in darkness. In: Pattern Recognition (ACPR), 2015 3rd IAPR Asian Conference on. IEEE, pp. 266–270.

Lore, K.G., Akintayo, A., Sarkar, S., 2017. Llnet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recognit. 61, 650–662.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60, 91–110.

Maaten, L.v.d., Hinton, G., 2008. Visualizing data using t-sne. J. Mach. Learn. Res. 9, 2579–2605.

Mahendran, A., Vedaldi, A., 2015. Understanding deep image representations by inverting them. In: Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on. pp. 5188–5196.

Malm, H., Oskarsson, M., Warrant, E., Clarberg, P., Hasselgren, J., Lejdfors, C., 2007. Adaptive enhancement and noise reduction in very low light-level video. In: ICCV.

Olmeda, D., Premebida, C., Nunes, U., Armingol, J.M., Escalera, A.d.l., 2013. Lsi far infrared pedestrian dataset.

Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A., 2008. Lost in quantization: Improving particular object retrieval in large scale image databases. In: Computer Vision and Pattern Recognition (CVPR), 2008 IEEE Conference on. pp. 1–8.

Qi, B., John, V., Liu, Z., Mita, S., 2014. Use of sparse representation for pedestrian detection in thermal images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 274–280.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. In: Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on. pp. 779–788.

Remez, T., Litany, O., Giryes, R., Bronstein, A.M., 2017. Deep convolutional denoising of low-light images. arXiv preprint arXiv:1701.01687.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems. pp. 91–99.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015a. ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. 115, 211–252. http://dx.doi.org/10.1007/s11263-015-0816-y.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al., 2015b. Imagenet large scale visual recognition challenge. Int. J. Comput. Vis. 115, 211–252.

Russell, B.C., Torralba, A., Murphy, K.P., Freeman, W.T., 2008. Labelme: a database and web-based tool for image annotation. Int. J. Comput. Vis. 77, 157–173.

Shen, L., Yue, Z., Feng, F., Chen, Q., Liu, S., Ma, J., 2017. Msr-net: Low-light image enhancement using deep convolutional network. arXiv preprint arXiv:1711.02488.

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

Su, H., Jung, C., 2017. Low light image enhancement based on two-step noise suppression. In: Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on. IEEE, pp. 1977–1981.

Torralba, A., Efros, A.A., 2011. Unbiased look at dataset bias. In: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, pp. 1521–1528.

Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y., 2010. Locality-constrained linear coding for image classification. In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, pp. 3360–3367.

Wang, S., Zheng, J., Hu, H.M., Li, B., 2013. Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE Trans. Image Process. 22, 3538–3548.

Wei, C., Wang, W., Yang, W., Liu, J., 2018. Deep retinex decomposition for low-light enhancement. In: British Machine Vision Conference.

Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., Lipson, H., 2015. Understanding neural networks through deep visualization. arXiv preprint arXiv:1506.06579.

Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks. In: European Conference on Computer Vision. Springer, pp. 818–833.

Zhao, X., He, Z., Zhang, S., Liang, D., 2015. Robust pedestrian detection in thermal infrared imagery using a shape distribution histogram feature and modified sparse representation classification. Pattern Recognit. 48, 1947–1960.

Zitnick, C.L., Dollár, P., 2014. Edge boxes: Locating object proposals from edges. In: European Conference on Computer Vision. Springer, pp. 391–405.