LEBANESE AMERICAN UNIVERSITY

Low-Light Image Enhancement for Object Classification using Deep Learning

By

Rayan Abdul Razzak Al Sobbahi

A thesis

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Engineering

School of Engineering

May 2021

LAU
الجامعة اللبنانية الأميركية
Lebanese American University

THESIS APPROVAL FORM

Student Name: Rayan El Sobbahi I.D.#: 201501537

Thesis Title: Low-Light Image Enhancement for Object Classification using Deep Learning

Program: Computer Engineering

Department: E.C.E

School: Engineering

The undersigned certify that they have examined the final electronic copy of this thesis and approved it in Partial Fulfillment of the requirements for the degree of:

M.S in the major of Computer Engineering

Thesis Advisor's Name: Joe Tekli

Signature: [redacted] Date: 11 / 05 / 21
Day Month Year

Committee Member's Name: Wissam Fawaz

Signature: [redacted] Date: 11 / 05 / 21
Day Month Year

Committee Member's Name: Zahi Nakad

Signature: [redacted] Date: 11 / 05 / 21
Day Month Year

LAU
الجامعة اللبنانية الأميركية
Lebanese American University

THESIS COPYRIGHT RELEASE FORM

LEBANESE AMERICAN UNIVERSITY NON-EXCLUSIVE DISTRIBUTION LICENSE

By signing and submitting this license, you (the author(s) or copyright owner) grants the Lebanese American University (LAU) the non-exclusive right to reproduce, translate (as defined below), and/or distribute your submission (including the abstract) worldwide in print and electronic formats and in any medium, including but not limited to audio or video. You agree that LAU may, without changing the content, translate the submission to any medium or format for the purpose of preservation. You also agree that LAU may keep more than one copy of this submission for purposes of security, backup and preservation. You represent that the submission is your original work, and that you have the right to grant the rights contained in this license. You also represent that your submission does not, to the best of your knowledge, infringe upon anyone's copyright. If the submission contains material for which you do not hold copyright, you represent that you have obtained the unrestricted permission of the copyright owner to grant LAU the rights required by this license, and that such third-party owned material is clearly identified and acknowledged within the text or content of the submission. IF THE SUBMISSION IS BASED UPON WORK THAT HAS BEEN SPONSORED OR SUPPORTED BY AN AGENCY OR ORGANIZATION OTHER THAN LAU, YOU REPRESENT THAT YOU HAVE FULFILLED ANY RIGHT OF REVIEW OR OTHER OBLIGATIONS REQUIRED BY SUCH CONTRACT OR

AGREEMENT. LAU will clearly identify your name(s) as the author(s) or owner(s) of the submission, and will not make any alteration, other than as allowed by this license, to your submission.

Name: Rayan Abdul Razzak Al Sobbahi – ID #: 201501537

Signature: [redacted] Date: 30 / 04 / 2021
Day Month Year

LAU
الجامعة اللبنانية الأميركية
Lebanese American University

PLAGIARISM POLICY COMPLIANCE STATEMENT

I certify that:

1. I have read and understood LAU's Plagiarism Policy.
2. I understand that failure to comply with this Policy can lead to academic and disciplinary actions against me.
3. This work is substantially my own, and to the extent that any part of this work is not my own I have indicated that by acknowledging its sources.

Name: Rayan Abdul Razzak Al Sobbahi – ID #: 201501537

Signature: [redacted]

Date: 30 / 04 / 2021
Day Month Year

ACKNOWLEDGMENT

This thesis would not have been possible without the support of many people. Big thanks to my advisor, Dr. Joe Tekli, who read and corrected my revisions and was very helpful, motivating and supporting through the research. Also, thanks to my committee members, Dr. Wissam Fawaz and Dr. Zahi Nakad, as well as the members of the Department of Electrical and Computer engineering who offered guidance and support.

I would like also to thank my friends and senior computer engineering students of our department who helped filling many of the surveys conducted in this project.

Finally, many thanks to my family and parents for their support, care, and love throughout two years of continuous work on this thesis.

Low-Light Image Enhancement for Object Classification using Deep Learning

Rayan Abdul Razzak Al Sobbahi

ABSTRACT

Low-light image (LLI) enhancement is an important image processing task that aims at improving the illumination of images taken under low-light conditions. Recently, a remarkable progress has been made in utilizing deep learning (DL) approaches for LLI enhancement. In this thesis, we perform a concise and comprehensive review and comparative study of the most recent DL models used for LLI enhancement. We address LLI enhancement in two ways: i) standalone, as a separate task, and ii) end-to-end, as a pre-processing stage embedded within another high-level computer vision task, namely object detection and classification. We also conduct a feature analysis of DL feature maps extracted from normal, low-light, and enhanced images, and perform the occlusion experiment to better understand the effect of enhancement on object detection and classification. We then address a common problem of these models depicted by their design as standalone solutions without focusing on the impact of enhancement on high-level computer vision tasks like object classification. Our review and empirical evaluations show that enhancing LLI visual quality does not necessarily correlate with improved object detection and classification performance, and may even deteriorate it, especially in cases where enhanced images include extreme artifacts. To solve the problem, we propose a new LLI enhancement model that performs image-to-frequency filter learning and is designed for seamless integration into classification models. Through this integration, the classification model is embedded with an internal enhancement capability and is jointly trained to optimize both enhancement and classification performance. We conduct a large battery of experiments involving 76 testers to evaluate our approach's LLI enhancement quality. When evaluated as a standalone enhancement model, our solution consistently ranks first or second among five state of the art enhancement techniques both quantitatively and qualitatively. When embedded with a classification model, our solution achieves an average of 5.5% improvement in classification accuracy, compared with the traditional pipeline of separate enhancement followed by classification. Results clearly produce robust classification performance on both low light and normal light images.

Keywords: Image Enhancement, Low-light Conditions, Deep Learning, Object Detection and Classification, Homomorphic Filtering, Empirical Comparison, Comparative Study

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

With the rapid spread of digital devices and photo-taking gadgets such as smart phones, pads, and tablets, capturing digital images has become an easy and common trend in our world. These images may be influenced by various poor visibility conditions like low-light, noise, haze, snow, and blur, among others. Low-light is a prominent element of our daily life that largely impacts the effectiveness of our vision and our ability to perceive meaningful content from badly illuminated objects, especially from Low-Light Images (LLIs). Low-light conditions are affected by the time of day (e.g., nighttime or twilight), the location (e.g., indoor or outdoor), and the availability of adequate light sources (e.g., natural and man-made lights) (Loh and Chan 2019). Modern artificial intelligence-based applications like autonomous spacecrafts, drones, autopilot car systems, robots, security surveillance systems, among others, essentially rely on visualizing and understanding outdoor environments. Such systems use cameras as their vision sensors to perform high-level computer vision tasks like classification, detection, semantic segmentation, and tracking. While these systems show good performance during normal and clear outdoor conditions, yet low-light conditions which account to a considerable time of our daily life can largely challenge the visual perception of such systems and significantly affect their robustness and hinder their market deployment (Yang et al. 2020; VidalMata et al. 2020). Hence, LLI enhancement has emerged: i) as a standalone image processing task that aims at illuminating LLIs and improving their visual quality, and ii) as a pre-processing step embedded within another high-level computer vision task to improve its performance.

Numerous traditional enhancement techniques have been proposed to tackle LLI enhancement. For instance, gamma correction methods, e.g., (Huang et al. 2013; Zhi et al. 2018) rely on a nonlinear transformation-based function where the different gray regions of the image are stretched or compressed by modifying the gamma correction parameter. Histogram equalization methods like CLAHE (Pisano et al. 1998) and DHE (Abdullah-Al-Wadud et al. 2007) stretch the histogram of the image to make it uniform and enhance its contrast. Other methods rely on the Retinex theory (Land and McCann

1971) which splits the image into reflectance and illumination components: reflectance describes the intrinsic properties of the image's objects and is assumed to be constant under varying light conditions, while illumination represents the varying lightness in the whole image. Typical Retinex-based approaches like MultiScale Retinex (MSR) (Rahman et al. 1996), Single Scale Retinex (SSR) (Jobson et al. 1997a), MultiScale Retinex with Color Restoration (MSRCR) (Jobson et al. 1997b), LIME (Li et al. 2015) and SRIE (Fu et al. 2016) adopt Retinex theory to perform LLI enhancement defined as an illumination estimation problem. However, deep learning techniques have demonstrated better performance and efficiency when compared with traditional methods (Tao et al. 2017; Guo et al. 2020).

Deep Learning (DL) approaches have been recently utilized to enhance LLIs and have shown great success. These approaches are data-driven as they require training datasets of LLIs and their corresponding Normal-Light Images (NLIs). Yet most DL approaches face two major challenges (Yang et al. 2020): i) the data aspect challenge – state of art enhancement models mainly rely on synthetic training datasets which might not be well representative of real world LLIs that incorporate nonlinear and complex degradations due to their visual quality; and ii) **the goal aspect challenge** - LLI enhancement is usually embedded as a pre-processing step in another high-level computer vision task, while the enhancement model itself is not initially designed for the target task. One major question is whether a LLI enhancement method – which performs well as a standalone component – can improve (or not) the performance of the high-level computer vision task as a whole. VidalMata et al. (2020) investigate the effect of image restoration and enhancement on object classification performance. The evaluation demonstrates that enhancing the image quality does not necessarily improve the classification performance but rather degrades it, especially when the enhanced images contain extreme artifacts.

In this thesis, we address **the goal aspect challenge** by designing a DL-based LLI enhancement model which is tailored for object classification. In the first part of the thesis report, we briefly describe and categorize the different models and techniques related to the task, while illustrating some of their main characteristics. Then, we empirically compare the models in two ways: i) standalone, as a separate task by analyzing the visual and perceptual performance of 10 publicly available enhancement models, and ii) end-to-end, as a pre-processing stage embedded within another high-level computer vision task: by comparing the performance of 4 object detection and classification models, applied on images enhanced by each of the 10 LLI enhancement models considered in the previous experiment. We also perform a DL feature analysis experiment to compare the feature maps extracted from LLIs, NLIs and enhanced images, and run the occlusion experiment (Zeiler and Fergus 2014) to better understand the effect of LLI enhancement on preserving the semantic features needed by the object detection and classification task. To our knowledge, this is the first comparative study dedicated to DL-based models for LLI enhancement, and we hope the obtained results will foster and guide further research on the subject.

In the second part of the thesis report, we introduce our approach which consists of two contributions: i) introducing a novel *LLI Enhancer* model based on image-to-frequency filter learning, and ii) introducing a *LLI Enhancer-Classifier* model, which integrates the enhancer model into a state of art object classification solution. On the one hand, the *LLI Enhancer* model performs image-to-frequency filter learning, inspired from homomorphic filtering traditionally used for LLI enhancement in which a frequency filter comprising only two parameters is devised to effectively filter the frequency-based LLI. On the other hand, the *LLI Enhancer-Classifier* model integrates the *LLI Enhancer* into a typical classification model, namely ResNet50 (He et al. 2016), to perform a joint training that optimizes both enhancement and classification performance simultaneously. Note that our solution is not tied to ResNet50, and is designed to use typical feature extractors utilized with existing classification models including VGG16 (Simonyan and Zisserman 2015), MobileNetv2 (Sandler et al. 2018), and SqueezeNet (Iandola et al. 2016), among others, thus making it unconstrained from any special architecture.

We perform a large battery of experiments to evaluate the performance of our approach. One the one hand, quantitative and qualitative evaluations on our *LLI Enhancer* model show competitive results compared with state of the art enhancement models like ZeroDCE (Guo et al. 2020), EnlightenGAN (Jiang et al. 2019) and DeepUPE (Wang et al. 2019). On the other hand, we compare our *LLI Enhancer-Classifier* model against the traditional pipeline commonly followed in the literature where separately preprocessed enhanced LLIs are evaluated on classification models pre-trained on benchmarks with abundant Normal Light Images (NLIs). *Enhancer-Classifier* results show a robust classification against both LLIs and NLIs, producing an average 5.5% improvement in classification accuracy on both synthetic LLIs form the Pascal VOC 2007 dataset (Everingham et al. 2012) and real-world images of ExDark dataset (Loh and Chan 2019).

The remainder of this report is organized as follows. Chapter 2 provides an overview of the LLI enhancement task. Chapter 3 presents our research motivation, aim, objectives, and contributions. Chapter 4 describes and categorizes the most prominent DL-based LLI enhancement models. Chapter 5 presents an empirical comparative study of 10 of the most recent LLI enhancement models. Chapter 6 describes the design and implementation of our *LLI Enhancer* model, and its empirical results. Chapter 7 describes the design and implementation of our *LLI Enhancer-Classifier* model, and its empirical results. Chapter 8 highlights the impact and limitations of our research and discusses future work. Finally, chapter 9 concludes the report.

Chapter 2

Overview of LLI Enhancement

The main objective of LLI enhancement is to improve the visual quality of an image by boosting its illumination and contrast while avoiding amplified noise or exposed artifacts. Formally, a low-light image $\mathrm{I}_{\mathrm{LLI}}$ is the output of a degradation function:

$$\mathrm{I}_{\mathrm{LLI}} = D\,(\mathrm{I}_{\mathrm{NLI}}, \delta), \tag{1}$$

where D denotes a degradation mapping function, $\mathrm{I}_{\mathrm{NLI}}$ the NLI, and δ the parameter of the degradation process (e.g. illumination level). Generally, the degradation process is complex as it may encompass – in addition to the illumination of the image – other factors like artifacts and noise. The enhancement task aims at recovering an approximation of $\mathrm{I}_{\mathrm{NLI}}$ denoted by $\mathrm{I}_{\mathrm{Enhanced}}$, generated from $\mathrm{I}_{\mathrm{LLI}}$ as follows:

$$\mathrm{I}_{\mathrm{Enhanced}} = F\,(\mathrm{I}_{\mathrm{LLI}}, \theta), \tag{2}$$

where F is the LLI enhancement model and θ its adjustment parameters. Here, we distinguish between two main categories of LLI enhancement models: i) traditional and ii) deep learning.

2.1 Traditional Approaches

Most traditional LLI enhancement techniques rely on mathematical or algorithmic models to perform the enhancement task. For instance, gamma correction methods, e.g., (Huang et al. 2013; Zhi et al. 2018) use a nonlinear transformation-based function in which a gamma correction parameter is adjusted to stretch or compress different gray regions of the image, aiming to enhance it. Also, histogram equalization methods, e.g., (Abdullah-Al-Wadud et al. 2007; Pisano et al. 1998; Wang et al. 1999) rely on a cumulative distribution function to change the image output gray levels such that they fit into a uniform distribution. The original LLI is mapped to its enhanced counterpart with an approximately uniform gray-level distribution. Yet the latter methods generally ignore spatially varying lightness and usually result in under or over brightened regions. In addition, Retinex theory –i.e., the theory of the human retinal cortex (Land and McCann 1971), has been utilized to perform LLI enhancement. Based on the nature of color perception by the human eye and the modeling of color constancy, methods in this category aim to remove the effects of illumination from the image leaving it with the reflective nature of its objects (Jobson et al. 1997a; Rahman et al. 1996; Jobson et al. 1997b). According to the theory, the Human Visual System (HVS) perceives the content and colors of the image constantly under varying or uneven lighting conditions, and thus only the major characteristics of the objects depicted in the reflection component are retained by the HVS (Lee et al. 2015). As a result, the reflectance component of the image is considered to be constant under varying light conditions and holds the inherent characteristics of visual objects. The Retinex model is thus used to estimate the illumination component of the image and retain its reflectance component, preserving the image's inherent features to allow more accurate image processing. Typical Retinex-based approaches like MultiScale Retinex (MSR) (Rahman et al. 1996), MultiScale Retinex with Color Restoration (MSRCR) (Jobson et al. 1997b) and Single Scale Retinex (SSR) (Jobson et al. 1997a) try to restore the illumination map and use it for enhancement. More recently, Wang et al. (2013) design an enhancement method which preserves the naturalness of images with non-uniform illumination. A bright pass filter is used to split the image into its reflectance and illumination components which respectively link to the details and naturalness of the image. Additionally, a bi-log transformation is applied to impose a balance between details and naturalness. Fu et al. (2016) propose a fusion based enhancement method. A simple illumination estimating algorithm based on morphological closing is used to decompose the image into its Retinex based components. The estimated illumination map is then adjusted and improved following an effective multi-scale fusion-based approach. Also, Fu et al. (2016) introduce SRIE, a weighted variational model that estimates both reflectance and illumination components of the input image. The model uses a better prior representation than the logarithmic transformation-based regularization and an alternating minimization scheme is utilized to solve the model. Li et al. (2015) propose a simple yet effective solution namely LIME. First, the illumination of each pixel is estimated by finding the maximum pixel of its RGB channels, then the illumination map is recovered by applying a structure prior. A joint denoising algorithm namely BM3D (Dabov et al. 2006) is applied as a post processing step for LIME. Li et al. (2018) design a robust Retinex model which takes noise into consideration. The method simultaneously estimates a structure revealing reflectance along with a smoothed illumination map and a noise map. The augmented Lagrange multiplier-based algorithm is utilized to solve the involved optimization problem. Yet most of the Retinex-based approaches assume that enhancement does not affect image reflectance, regardless of the color distortions or lost details that result from applying the Retinex model (Wang et al. 2019). In addition, Retinex-based enhancement quality is highly dependent on a set of carefully hand-crafted parameters allowing to estimate the resulting illumination map (Wei et al. 2018).

2.2 Deep Learning Approaches

In contrast to the traditional algorithmic or mathematical enhancement approaches, Deep Learning (DL) enhancement models are essentially data-driven, where training datasets of LLIs and NLIs are used to drive the learning process. DL models are a special kind of machine learning algorithms made of multilayered artificial neural networks, inspired by the structure and function of the human brain. They aim to find unknown structures or patterns in the input distribution so that they discover good representations of the data and learn its features through a hierarchical architecture (Deng 2014). DL techniques have gained great attention in the past few years as the most effective machine learning solutions to perform LLI enhancement, outperforming traditional methods based on histogram equalization e.g. (Pisano et al. 1998; Abdullah-Al-Wadud et al. 2007); and Retinex theory e.g., (Jobson et al. 1997a; Jobson et al. 1997b; Rahman et al. 1996). They accept LLIs as input, and propagate them through the DL model to learn a variety of features needed for the enhancement task. Paired labels of LLIs/NLIs are essentially needed to train the DL model under a supervised setting, allowing it to learn how to perform the enhancement task. A loss function is one of the main elements of a DL solution, allowing to evaluate how well a given model fits the training data. Through an iterative self-evaluation process, the loss function usually guides the DL model to reduce the error in its own predictions. In this context, commonly used DL loss functions like Mean Absolute Error (MAE, or L1 loss) and Mean Square Error (MSE, or L2 loss) might not always be suitable to accurately evaluate the visual quality of enhanced LLIs (Wang et al. 2004). Given the various elements that affect the quality of the image including illumination levels, color deviations, artifacts, noise, etc., recent studies have introduced more sophisticated loss functions to improve the quality of enhanced LLIs, including: perceptual loss (Lv et al. 2018), illumination smoothness loss (Wei et al. 2018), and adversarial loss (Wang et al. 2019), among others.

While they usually require expensive training time and effort (Abu-Khzam et al. 2019; Abu-Khzam et al. 2015), yet various reasons have contributed to the leap of DL algorithms and their applications, including (Deng 2014): i) the substantial increase in computational capabilities (e.g., GPUs), ii) the lower costs of computing hardware, iii) the significant advances of machine learning algorithms (Salem et al. 2018; Ebrahimi et al. 2020; Abu-Khzam et al. 2018), and iv) the increasing availability of training data. In chapter 4, we thoroughly describe and categorize the recent DL models for LLI enhancement.

Chapter 3

Proposal

In this chapter we describe the motivation of our work, then state its aim and list its main objectives. We finally highlight the major contributions of the research.

3.1 Motivations

The motivations behind this research can be summarized as follows:

Motivation 1:

The existing literature lacks a comprehensive survey and empirical study that is dedicated to reviewing and evaluating DL-based enhancement techniques for images taken under low-light conditions (i.e., LLIs).

Motivation 2:

Existing DL classification models show impressive accuracy results when processing images taken under normal and clear-light conditions (i.e., NLIs), yet their performance degrades significantly when challenged by low-light conditions (Yang et al. 2020; VidalMata et al. 2020). So, classification models do not provide a robust performance against both LLIs and NLIs. Therefore, a dedicated LLI enhancement model is needed to help boost the performance of classification models under low-light conditions.

3.2 Aim

Given the two motivations mentioned above, this research aims at designing a LLI enhancement model tailored for the object classification task.

3.3 Objectives

To fulfill our aim, the following five objectives are considered:

1. Survey and categorize existing DL-based LLI enhancement models.

2. Evaluate the models at three levels: enhancement performance, detection & classification performance, and feature preservation performance.
3. Design a new LLI enhancement solution which is learnable through a DL model, while allowing a seamless integration with typical architectures and loss functions used for classification models.
4. Integrate the designed enhancement model into one state of art classification model.
5. Validate the enhancement and classification performance of the proposed model and compare it with recent existing techniques.

3.4 Contributions

Our research offers four main contributions:

1. Surveying and comparing existing DL-based LLI enhancement models at three different levels: enhancement, detection & classification, and feature preservation.
2. Designing an enhancement model tailored for high-level computer vision tasks, particularly the object classification task.
3. Embedding classification models with an internal enhancement capability.
4. Producing a robust classification performance against varying light conditions.

Chapter 4

Deep Learning-based LLI Enhancement: Review

This chapter provides an in-depth review of most prominent and recent DL-based LLI enhancement models. We organize the models in five main categories: i) Encoder-decoder and Convolutional Neural Network (CNN)-based models, ii) Retinex theory-based models, iii) Fusion-based models, iv) Generative Adversarial Network (GAN)-based models, and more recent v) Zero Reference models.

4.1 Encoder-decoder and CNN-based Models

Various works have focused on utilizing encoder-decoder models, CNNs, or have integrated them together to perform LLI enhancement.

Encoder-decoder models: An encoder-decoder is a DL model designed to learn a mapping from an input domain to an output domain through a two-stage network comprising: i) an encoder which encodes the input into a latent feature representation, and ii) a decoder which decodes and reconstructs the original features to predict the output. While largely used in image-to-image translation applications (Minaee et al. 2020), encoder-decoder solutions have been recently developed for image enhancement, where the input is a LLI, and the output is its enhanced counterpart, e.g., (Lore et al. 2017; Jiang et al. 2018; Xu et al. 2018). An autoencoder is a special type of encoder-decoder which aims at learning a reduced encoding for the data, and to generate from the reduced encoding a representation as close as possible to its original input (Minaee et al. 2020). There are many variants of autoencoders such as: i) sparse autoencoders: extracting sparse features from the input data, by penalizing hidden unit biases (Ranzato et al. 2006) or unit activations (Le et al. 2011), ii) denoising autoencoders: recovering the correct input from a corrupted version of the input data, by forcing the network to learn the structure of the input distribution (Vincent et al. 2008), and iii) convolutional autoencoders: combining CNNs and autoencoders, where the encoder consists of a series of convolutional and pooling layers and the decoder consists of deconvolutional and unpooling layers.

LLNet (Lore et al. 2017) is one of the earliest DL approaches for LLI enhancement. It uses a stacked-sparse denoising autoencoder (SSDA) as its deep neural network architecture with three denoising autoencoder layers comprising hidden units with no use of convolutional layers. The model is trained on synthetic LLIs obtained from normal images through gamma correction and Gaussian noise induction, and uses the L2 loss function. Experimental results by Lore et al. (2017) highlight a tradeoff between the sharpness of the enhanced image and its noise levels. The model shows competitive results when compared with traditional approaches based on histogram equalization (Abdullah-Al-Wadud et al. 2007; Pisano et al. 1998) and gamma adjustment.

CNN models: A Convolutional Neural Network (CNN) is a DL network consisting of a regularized version of the multilayer perceptron that uses the linear convolution mathematical operation in place of general matrix multiplication in at least one of its layers. CNNs are highly effective and have been commonly used in various computer vision applications (Guo et al. 2016), allowing to extract, distinguish, and assemble complex visual features (patterns) from the images' visual properties and objects. A typical CNN consists of three types of consecutive layers: i) convolutional layers: using kernels to convolve the whole image as well as intermediate feature maps and generate new feature maps, ii) pooling layers: reducing the feature map dimensions and the number of network parameters, and iii) fully connected layers: mapping a 2D feature map into a 1D feature vector that either refers to a certain number of categories for image classification or is utilized for further processing. CNNs have been largely used for the image classification task, including famous architectures such as VGG16 and VGG19 (Simonyan and Zisserman 2015), AlexNet (Krizhevsky et al. 2012), and ResNet (He et al. 2016).

LLCNN (Tao et al. 2017) is one of the early CNN-based models for LLI enhancement. It is built using specially designed convolutional modules inspired from inception modules (convolving an input using different size convolutional layers and then combining their outputs to the next layer) and residual modules (employing shortcut connections). It uses a Structural Similarity Index (SSIM) (Wang et al. 2004) based loss function and relies on synthetic LLIs created through random gamma adjustment for training the network. The model demonstrates superior performance compared with LLNet (Lore et al. 2017) and many traditional approaches (Abdullah-Al-Wadud et al. 2007; Pisano et al. 1998; Rahman et al. 1996; Jobson et al. 1997a; Fu al et. 2016).

Gharbi et al. (2017) propose a deep bilateral CNN based model to perform fast real-time enhancement. The approach aims at processing a low-resolution version of the image in which a bilateral grid of affine coefficients is estimated. Then a slicing operation is used to up-sample the affine coefficients into the full image resolution. The model is designed to learn global and local features and preserve edges. L2 loss is used to train the network on the MIT FiveK dataset (Bychkovsk et al. 2011). The results demonstrate the effectiveness of the model in real time image enhancement. One limitation mentioned by the authors is the network's strong dependence on the modelling assumptions and constraints related to the affine transformations in the bilateral space.

Chen et al. (2018) introduce a learning to *See In the Dark* (SID) model for image enhancement and noise suppression designed to process images under extreme low-light conditions. The model relies on Fully Convolutional Networks (FCNs) (using convolutional layers only) and is trained using L1 loss on a newly collected dataset of raw LLIs taken by the imaging sensors of Sony 7SII and Fujifilm X-T2 cameras. Although the model is able to suppress noise and produce proper coloring, it is limited to raw data obtained using a specific camera sensor and the images of the SID dataset do not contain pictures of humans and dynamic objects (Chen et al. 2018).

Integrated models: Jiang et al. (2018) propose LL-RefineNet, a deep refinement network consisting of two symmetrical paths: forward and backward. In the forward path, high-level features with global content are extracted and then gradually fused with low-level features with local content and refined during the backward refinement path. The model relies on synthetic LLIs based on impulse and Gaussian noise and guided using a mixed loss function of L1 and L2 losses. Results show that the model outperforms LLCNN (Tao et al. 2017) and many traditional approaches both quantitatively and qualitatively.

Xu et al. (2018) introduce LRCNN: a Low-light Residual Connection based Convolutional Network, consisting of: i) a convolutional encoder-decoder structure in which the encoder is used for feature extraction and the decoder for denoising, connected with a ii) sequence of fully connected layers for brightness enhancement. Residual connections are used to better preserve the details of the original image. The network is guided by an L2 based loss function and is trained on a synthetic dataset of LLIs simulated from the CVG-UGR database. Results show that the model can remove noise and properly adjust light intensity.

Wang et al. (2018) introduce a Global Illumination Aware and Detail-preserving NETwork (GLADNET) comprising: i) a global illumination estimation step using an encoder-decoder structure where the encoder consists of convolution layers and the decoder consists of resize convolutional layers (Odena et al. 2016), followed by ii) a reconstruction step through a series of convolutions where the input image is concatenated with the predicted features from the encoder-decoder to better preserve the original image features. The network is trained on a synthesized dataset collected from RAISE (Dang-Nguyen et al. 2015) and guided by L1 loss. Results show that the model produces clear and natural enhanced images with preserved details.

4.2 Retinex Theory-based Models

Other DL approaches are inspired by the Retinex theory (Land and McCann 1971) that decomposes the image into a constant reflectance map and a light varying illumination map (cf. Section 2.1). Multi Scale Retinex Net (MSR-Net) (Shen et al. 2017) is one of the early models in this category. It performs LLI enhancement in three stages. The input LLI is first processed as a set of multi-scale logarithmic transformations. The transformed image is then fed into a CNN, and is finally processed through a dedicated color restoration function. The model is trained using the L2 loss function and a synthesized dataset obtained from the UCID dataset (Schaefer and Stich 2003), the BSD dataset (Arbelaez et al. 2011), and Google Images. While the model is effective in producing images with rich colors and clear textures, yet it sometimes fails to properly handle the image edge features as it tends to produce some darkness around the edges, especially in bright regions (referred to as the “halo effect”) (Shen et al. 2017).

Another approach is RetinexNet (Wei et al. 2018) which consists of two subnetworks: i) DecomNet that aims at learning the decomposition of the image into its reflectance and illumination components based on Retinex theory, and ii) EnhanceNet that performs illumination adjustment and enhancement through a dedicated encoder-decoder structure which uses multiscale concatenation to maintain the global and local illumination of the enhanced image. A joint denoising operation using 3D transform-domain filtering (BM3D) denoising algorithm (Dabov et al. 2006) is then applied on the reflectance component. Wei et al. (2018) introduce their own training dataset named LOw-Light (LOL), consisting of 500 pairs of real LLIs and NLIs. They also put forth a multi-term loss function combining reconstruction, invariable reflectance, and illumination losses. The resulting enhanced images are produced with a good image decomposition learning and are deemed visually pleasing by the authors.

Li et al. (2018) introduce LightenNet, a CNN model made of 4 convolutional layers for i) patch extraction and representation, ii) feature enhancement, iii) non-linear mapping, and iv) reconstruction. It is designed to predict the Retinex illumination map component from the original LLI, which is then used to produce the enhanced image. The network learns through a synthesized dataset obtained by the Retinex model and is guided by the L2 loss function. The enhanced images are visually pleasing with well restored content. Yet, the method shows a degraded performance while applied on low-quality images due to noise or JPG compression resulting in noise and artifacts amplification (Li et al. 2018).

Wang et al. (2019) describe Retinex Decomposition based Generative Adversarial Network (RDGAN) which consists of two subnetworks: i) Retinex Decomposition Net (RDNet) that decomposes the LLI into its illumination and reflectance components, and ii) Fusion Enhancement Net (FENet) that fuses the decomposed parts into an enhanced image. The model is trained using the SICE dataset (Cai et al. 2018) and utilizes a novel adversarial loss function based on GANs to improve visual quality. While the model can properly recover the details and colors of the original LLI, yet it also tends to amplify noise and JPEG artifacts that are not obvious in the LLI, thus possibly degrading the quality of the enhanced image (Wang et al. 2019).

Zhang et al. (2019) introduce KinD (Kindling the Darkness) consisting of three networks: i) layer decomposition that decomposes the image into reflectance and illumination components, ii) reflectance restoration which aims at removing degradations that are concentrated in the dark regions of the reflectance, and iii) illumination adjustment which distributes the illumination across the image. The authors design an integrated loss function based on L1, L2, and SSIM (Wang et al. 2004) losses, and train the model on the LOL dataset (Wei et al. 2018). Results in (Zhang et al. 2019) show that the model produces enhanced images with properly adjusted lightness and suppressed noise.

Wang et al. (2019) introduce a Deep Underexposed Photo Enhancement (DeepUPE) model which performs image-to-illumination map learning. It consists of an encoder network (i.e., a pre-trained VGG16 (Simonyan and Zisserman 2015) that extracts the image's local and global features, followed by a bilateral grid based up-sampling allowing to produce the image's full resolution illumination map. The latter is then used to enhance the image based on the Retinex model. The authors use an integrated loss function combining reconstruction, smoothness, and color losses. A newly proposed dataset of underexposed images and expert retouched references is used for training and evaluation. Results by Wang et al. (2019) show a good recovery of the image details, contrast, and colors.

4.3 Fusion-based Models

Some LLI enhancement models consider fusing the derived images or feature maps by multiple traditional or DL techniques to combine their advantages into a final enhanced image. MBLLEN, a Multi-Branch Low-Light Enhancement Network (Lv et al. 2018) is one of the earliest models in this category. It uses a dedicated feature extraction module to extract the LLI features at each of its 10 convolutional layers, and then enhances the features at each layer using an encoder-decoder based enhancement module. It finally fuses the multi-branch enhanced features to form the enhanced image. The model uses a loss function composed of structure, context, and region losses. It learns through a synthesized dataset of LLIs from the Pascal VOC dataset (Everingham et al. 2012). The resulting enhanced images have good brightness and contrast with minimal artifacts.

Shin et al. (2018) propose ACA-net, an Adversarial Context Aggregation network consisting of a Context Aggregation Network (CAN) applied with an adversarial GAN-based loss function. First, image illumination is boosted using two gamma correction functions, then the corresponding feature maps are extracted using convolutional layers and passed through a CAN which uses dilated convolutions to perform an effective aggregation of the global contextual information in the image. The network is guided by an integrated loss function combining reconstruction and adversarial losses, and the training data is synthesized based on aesthetic visual analysis (AVA) dataset (Murray et al. 2012). The model shows superior performance compared with MSR-Net (Shen et al. 2017).

Another fusion-based approach is DFN (Deep Fusion Network) (Cheng et al. 2019), which combines three traditional enhancement techniques: CLAHE (Pisano et al. 1998), log correction and bright channel enhancement. It runs the three models on the same input LLI and produces the feature confidence maps from the three derived images using an encoder-decoder network. It then weights the derived images by the obtained confidence maps in an element-wise fusion to output the final enhanced image. The model aims at combining the significant features emphasized by the constituent enhancement methods. It utilizes an integrated loss function composed of L1 and L2 losses and is trained on a dataset synthesized from 600 NLIs using gamma correction. Although it shows good performance, yet the authors mention that DFN may add smoothing and artificial edges in the fused image as it lacks an edge preserving capability (Cheng et al. 2019).

Wang et al. (2019) utilize attention modules to selectively enhance useful features while suppressing features that are not so important for the network, and use multi-scale feature fusion to combine global features with strong semantic features at deeper layers. The model consists of feature extraction blocks (FEB), where each FEB is a convolutional block made up of an attention module and two convolutional layers, and a feature fusion block (FFB) which fuses multilevel features through pixel-wise addition and channel connection. The model uses a Peak Signal to Noise Ratio (PSNR) based loss function and is trained on real images from the SID (Chen et al. 2018) and S7ISP (Schwartz et al. 2019) datasets, as well as synthetic images produced based on the Pascal VOC dataset (Everingham et al. 2012). The model shows competitive results compared with many traditional enhancement approaches (Jobson et al. 1997b; Fu et al. 2016; Li et al. 2015).

Lv et al. (2020) introduce an attention-guided model that aims at handling image enhancement and denoising simultaneously by using Under Exposure (UE) attention maps and noise maps that guide the model attention in a region-aware adaptive manner. The model consists of four components: i) Attention Net: produces the UE maps used to avoid over-enhanced regions, ii) Noise Net: estimates the noise distribution map, iii) Enhancement Net: extracts and enhances features then fuses them through a multi-branch CNN concatenation and iv) Reinforce Net: uses dilated convolutions to improve the image contrast and details. The integrated loss function combines L1, L2, SSIM (Wang et al. 2004), and VGG19 (Simonyan and Zisserman 2015) based perceptual losses, among others. The network is trained on a synthesized dataset from publicly available datasets like Pascal VOC (Everingham et al. 2012) and Microsoft (MS) COCO (Lin et al. 2014). Extensive evaluation experiments show the superior performance of the proposed model compared with LLNet (Lore et al. 2017), MBLLEN (Lv et al. 2018), SRIE (Fu et al. 2016), LIME (Li et al. 2015), among others.

Ren et al. (2019) propose a deep hybrid network consisting of two streams that simultaneously learn: i) the global content and ii) the salient edge contents of the input image. The first stream uses a residual encoder-decoder and the second stream utilizes a novel spatially variant recurrent neural network (RNN) to model the edge details. The network is guided by an integrated loss function combining L2, perceptual, and adversarial losses, and is trained using MIT-Adobe FiveK dataset (Bychkovsky et al. 2011). The enhanced images are shown to be visually pleasing with minimal artifacts and color distortions (Ren et al. 2019).

Xiang et al. (2019) introduce a multi-branch encoder-decoder architecture combining: i) DCGAP: a Dilated Convolution and Global Average Pooling module used to better learn the image global features, and ii) ConvLSTM: a Convolutional Long Short-Term Memory that allows remembering and preserving the features learned at the different branches. The model is guided by L1 and SSIM (Wang et al. 2004) losses and is trained using the LOL dataset (Wei et al. 2018) and 1000 synthetic images based on RAISE (Dang-Nguyen 2015). The model successfully enhances LLI visual quality while minimizing noise and artifacts (Xiang et al. 2019).

4.4 GAN-based Models

Recently, Generative Adversarial Networks (GANs) have been attracting attention for image-to-image mapping applications (Goodfellow et al. 2014), and have been successfully employed for the LLI enhancement task. A typical GAN is made-up of two networks: a generator and a discriminator. The generative network is trained to generate realistic synthetic data samples from a data distribution of interest, while the discriminative network is trained to distinguish fake samples produced by the generator form the true data distribution. The generative network's training objective is to increase the error rate of the discriminative network, as it attempts to "fool" the discriminator network by producing novel candidates that the discriminator thinks are not synthesized. DL models like encoder-decoders and CNNs are used for the generator and discriminator networks. In the context of image enhancement, LLIs are used as real samples and enhanced images as fake samples to be generated.

Meng et al. (2019) introduce one of the earliest GAN-based models to perform LLI enhancement, consisting of an encoder-decoder based generator supplemented by a fusion network that combines features from the different layers of the encoder-decoder. Through adversarial learning, the discriminator is trained to differentiate a LLI from an enhanced image while the generator is trained to fool the discriminator. The model learns using a vehicle dataset of daytime and nighttime images that are not exactly taken at the same scenes, and is driven by an integrated loss function combining adversarial, perceptual, and total variation losses. A major problem highlighted by the authors is the tendency of the model to miss objects that are strongly illuminated in nighttime images.

Hua and Xia (2018) propose a GAN-based approach supported by Image Quality Assessment (IQA) techniques, in particular an image quality assessment network NIMA (Talebi and Milanfar 2018) which relies on the VGG16 (Simonyan and Zisserman 2015) feature extractor to minimize the model's dependence on the training dataset and boost its de-noising and de-blurring performance. The authors use an integrated loss function combining IQA, content, and total variation losses, and introduce a synthesized dataset based on General100 (Dong et al. 2016) and other image sources by applying Gaussian correction, Gaussian blur, and noise induction techniques on the normal images. Results in (Hua and Xia 2018) highlight a certain balance between noise suppression and the preservation of image details.

Kim et al. (2019) introduce Low-LightGAN which applies spectral normalization on the network to make the training more stable and accurate. It uses a combination of loss functions including adversarial, perceptual, color, and total variation losses, specifically tuned to produce visually pleasing images. The authors propose a task-driven training dataset based on local illumination synthesis rather than global low-light synthesis, so that over-saturated bright regions in the image are avoided. Results show good performance although the model may add artifacts in the background of the enhanced images.

Yangming at al. (2019) combine Retinex theory and GANs. Their generative network includes: i) a decomposition part that decomposes the image into its reflectance and illumination components, and ii) an enhancement part that enhances the lightness of images taken from the CSID dataset (Chen et al. 2018). The loss function combines regularization, reconstruction, and adversarial losses, among others. Results by Yangming at al. (2019) show that combining Retinex theory and GANs can effectively handle LLI enhancement.

Chen et al. (2018) propose a Deep Photo Enhancer (DPE) model using a GAN-based architecture for image enhancement, while considering paired and unpaired training settings (i.e., with and without LLI/NLI pairs[1]). A global feature U-Net (Ronneberger et al. 2015) is used to investigate the paired training setting. Two network architectures are used for unpaired training: 1-way GAN and 2-way GAN. In addition, two improvements are added to stabilize the training: adaptive WGAN (Arjovsky et al. 2017) and individual batch normalization for the generator. The loss function is based on L2 and adversarial losses. The authors produce a dataset extracted from MIT-Adobe 5K (Bychkovsky et al. 2011) and HDR images selected from Flickr images. Results mainly show good quality enhanced images with natural colors, yet the authors also highlight that the model might amplify noise in very dark and noisy images.

[1] Unpaired training is increasingly used with GANs and consists in training the model using unmatched training data, e.g., LLIs and NLIs which are produced separately, and which do not necessary match.

A recent approach by Jiang et al. (2019) introduces EnlightenGAN: a first successful attempt at generalizing well to various real-world scenes while using unsupervised learning for image enhancement based on GANs. The model undergoes unpaired training, uses an attention guided U-Net (Ronneberger et al. 2015) as the backbone for the generator, and includes a global relativistic discriminator (Jolicoeur-Martineau 2018) along with a local one to handle spatially varying light conditions in the image. Self-regularization is adopted for the loss function and the attention mechanism, since the model is independent from reference training labels. The loss function combines local and global discriminator adversarial losses and a self-feature preserving loss. The training dataset consists of unpaired LLIs and NLIs sampled from the LOL (Wei et al. 2018), RAISE (Dang-Nguyen et al. 2015) and HDR datasets (Gharbi et al. 2017; Kalantari and Ramamoorthi 2017). Results by Jiang et al. (2019) demonstrate a successful enhancement of dark areas while preserving the texture details and producing naturalistic images with no under- or over-exposed regions.

4.5 Zero Reference Models

A recent approach by Guo et al. (2020) opens the door for a new category of LLI enhancement techniques which does not require paired or unpaired training data (hence the name “Zero Reference”). The authors introduce Zero Reference Deep Curve Estimation (Zero-DCE) which entirely reformulates the LLI enhancement task: from an image-to-image mapping task into an image-to-light curves estimation task. Inspired by curve adjustment techniques used in digital photo editing solutions, the authors design light enhancement curves that are learned and estimated by a lightweight deep curve estimation network (DCE-Net), and are then iteratively applied on the input LLI to produce the final enhanced image. The model can be trained in the absence of paired or even unpaired training data by using non-reference loss functions such as spatial consistency, exposure control, color constancy, and illumination smoothness losses that can indirectly evaluate the quality of enhancement. The proposed method is computationally efficient and shows superior performance compared with DL enhancement models like EnlightenGAN (Jiang et al. 2019), RetinexNet (Wei et al. 2018), and LIME (Li et al. 2015), among others.

Another recent approach by Zhang et al. (2020) presents a self-supervised DL model for LLI enhancement, which can be trained using only LLIs with no need for paired or unpaired training data. Relying on the Retinex model and entropy theory, the authors devise a well-tuned loss function that includes a new method to compute reflectance loss. The method is based on the assumption that the enhanced image should have enough information and should comply with the original image. This is achieved by applying histogram equalization on the LLI to improve its information entropy. Driven by the newly designed referenceless loss function, a CNN model is then trained on the LOL dataset (Wei et al. 2018) to perform the enhancement task. Results by Zhang et al. (2020) show that the model produces visually pleasing images with short training time, and exhibits good real-time performance.

4.6 Discussion

Table 1 summarizes the main characteristics of recent DL-based LLI enhancement solutions. While many DL enhancement models have been shown to outperform their traditional counterparts, e.g., (Lore et al. 2017; Tao et al. 2017; Jiang et al. 2018), yet most of them share several challenges.

Challenge 1: Most approaches consider supervised learning where paired LLIs/NLIs are needed to train the models. Yet collecting large datasets of real-world LLIs and their corresponding daytime counterparts for the same scenes is difficult and challenging. To counter this problem, most techniques utilize synthetic LLIs produced from NLIs using light correction and noise induction techniques like gamma correction and Gaussian noise. However, synthetic LLIs do not always accurately represent real world low-light conditions, which usually encompass non-linear and spatially varying light conditions and noise levels, and are difficult to simulate mathematically. In an attempt to ease the restriction of paired or unpaired training labels and counter the synthetic LLIs performance problem, one recent approach by Guo et al. (2020) redefines the LLI enhancement task from an image-to-image learning task where the enhanced image is the final output of the network, to an image-to-curve estimation where light curves are learned and applied to enhance the image. The model achieves a good performance, thus opening new horizons for formulating the enhancement task. Another study by Jiang et al. (2019) describes a GAN-based unsupervised learning method (i.e., EnlightenGAN), performing enhancement without the need for training pairs or the LLIs' daytime counterparts. The latter achieves good performance levels and shows a lot of promise since unpaired image datasets are much easier to come by compared with paired ones.

Challenge 2: Most of the approaches tend to struggle whenever low-quality, noisy, or very dark images are considered during enhancement. This underlines the need for a proper understanding and modeling of the quality and noise elements in an image when conducting image enhancement or when designing a new LLI enhancement approach.

Challenge 3: Most existing techniques are developed as standalone solutions aiming to improve the illumination and the quality of LLIs. Yet, the latter's impact on high-level computer vision tasks like object detection and classification remains uncertain, where high-level image features might be distorted or lost during the enhancement task, thus leading to reduced or non-improving end-to-end performance.

Challenge 4: It is difficult to fairly compare most existing models for two main reasons: i) lack of a large standard dataset of paired LLIs/NLIs that are taken from real-world scenes and represent various low-light conditions, and ii) lack of a (set of) common and standard metric(s) that can accurately evaluate the visual perception of enhanced image quality. As can be seen in Table 1, different datasets and evaluation metrics are used to train and evaluate the visual performance of different enhancement models.

In the following empirical study, we further discuss the above challenges aiming to acquire a better understanding of the issues at stake and shed light on possible future directions.

Table 1: Characteristics of DL-based LLI enhancement models.

	Model	Description	Evaluation Metrics[2]	Loss function[2]	Training Datasets
Encoder-decoder and CNN based	LLNet (Lore et al 2017)	- First application of DL to enhance LLIs - Uses a sparse stacked denoising autoencoder for contrast enhancement and denoising	PSNR & SSIM (Wang et al 2004)	L2	Synthetic based on CVG-UGR database[3]
	LLCNN (Tao et al 2017)	- Uses CNN modules based on inception modules and residual connections	PSNR, SSIM, LOE (Wang et al 2013) & SNM (Hojatollah and Wang 2013)	SSIM	Synthetic based on CVG-UGR database[3]
	LL-RefineNet (Jiang et al 2018)	- Uses two symmetrical paths: forward to extract high level features and backward to fuse and refine with low level features	PSNR, SSIM, & Root-MSE	L1 & L2	Synthetic based CVG-UGR database[3]
	HDR-Net (Gharbi et al 2017)	- Performs real time enhancement using a deep bilateral CNN which processes images in their low-resolution version	PSNR	L2	MIT-Adobe FiveK (Bychkovsky et al 2011)
	SID (Chen et al 2018)	- Utilizes fully convolutional networks to enhance and denoise sensor raw data	PSNR & SSIM	L1	SID (Chen et al 2018)
	LRCNN (Xu et al 2018)	- Uses a deep residual convolutional encoder-decoder along with fully connected layers for contrast enhancement and denoising	PSNR & SSIM	L2	Synthetic based on CVG-UGR database[3]
	GladNet (Wang et al 2018)	- Uses an encoder-decoder to estimate illumination and a CNN for content reconstruction	-----------	L1	Synthetic based on RAISE (Dang-Nguyen et al 2015)
Retinex theory based	MSR-Net (Shen et al 2017)	- Uses Retinex theory to construct a CNN network that learns a mapping from dark to bright images	SSIM, NIQE (Mittal et al 2013), & DE (Amigó et al 2007)	L2	Synthetic based on UCID (Schaefer and Stich 2003), BSD (Arbelaez et al 2011), and Google images
	RetinexNet (Wei et al 2018)	- Decomposes the image into its reflectance and illumination components, then performs enhancement and denoising	-----------	Reconstruction, IR, & IS	Real from LOL & synthetic based on RAISE (Dang-Nguyen et al 2015)
	LightenNet (Li et al 2018)	- Learns an image to illumination map translation through a CNN	MSE, PSNR &SSIM	L2	Synthetic based on 600 pairs of normal & Retinex-darkened images
	RDGAN (Wang et al 2019)	- Learns to decompose the image into reflectance and illumination, then fuses them into a final enhanced image	PSNR & FSIMc (Zhang et al 2011)	Multi term decomposition, content & adversarial	SICE (Cai et al 2018)
	KinD (Zhang et al 2019)	- Decomposes the image into reflectance and illumination components, clears reflectance degradations, and adjusts illumination	PSNR, SSIM, LOE, & NIQE	Based on: L1, L2, & SSIM	LOL (Wei et al 2018)
	DeepUPE (Wang et al 2019)	- Learns an image-to-illumination mapping, then applies the Retinex model to enhance the image	PSNR & SSIM	Reconstruction, smoothness, & color	3000 pairs of underexposed & expert retouched images
Fusion based	MBLLEN (Lv et al 2018)	- Extracts features at every layer of the CNN, then enhances them via an encoder-decoder, and finally fuses them into an enhanced image	PSNR, SSIM, AB (Chen et al 2006), VIF (Sheikh and Bovik 2006), LOE, & TMQI (Hojatollah and Wang 2013)	Structure, context & regional	Synthetic based on Pascal VOC (Everingham et al 2012)
	ACA-Net (Shin et al 2018)	- Utilizes context aggregation networks to aggregate the global context of the image	PSNR & SSIM	Reconstruction & adversarial	Synthetic based on AVA (Murray et al 2012)
	DFN (Cheng et al 2019)	- Combines features extracted using an encoder decoder from images derived by traditional methods	MSE, PSNR, SSIM, & NIQE	L1 & L2	Synthetic based on datasets from (Gu et al 2016; Ma et al 2017)
	(Wang et al 2019)	- Utilizes attention-based modules to enhance important features and suppress non-vital features	PSNR	PSNR	SID (Chen et al 2018), S7ISP (Schwartz et al 2019), & synthetic based on Pascal VOC (Everingham et al 2012)
	(Lv et al 2020)	- Produces two attention maps to guide exposure enhancement and denoising, and performs enhancement using a multi-branch CNN	PSNR, SSIM, VIF, LOE, TMQI, AB, & LPIPS (Zhang et al 2018)	L1, L2, bright, structure, perceptual, & regional	Synthetic based on publicly available datasets: Pascal VOC (Everingham et al 2012), MS COCO (Lin et al 2014), (Grubinger et al 2006; Bileschi 2006)
	(Ren et al 2019)	- Utilizes an RNN to learn salient edge contents and an encoder-decoder to learn the global contents	PSNR & SSIM	L2, perceptual, & adversarial	MIT-Adobe FiveK (Bychkovsky et al 2011)
	(Xiang et al 2019)	- Uses a multi-branch encoder-decoder supplemented by ConvLSTM module	PSNR & SSIM	L1 & SSIM	LOL (Wei et al 2018) & synthetic based on RAISE (Dang-Nguyen et al 2015)
GAN based	(Meng et al 2019)	- Utilizes an encoder-decoder with a fusion network for the generator	Cosine similarity	Adversarial, perceptual, & TV	OAV dataset (Milford and Wyeth 2012)
	(Hua and Xia 2018)	- Uses GANs joint with an image quality assessment network to improve visual quality	PSNR, SSIM, MSSSIM (Wang et al 2003) & IWSSIM (Wang and Li 2011)	Adversarial, content, TV & IQA	Synthetic based on General100 (Dong et al 2016), Sun-Hayes80 (Sun and Hays 2012) & Urbanal100 (Huang et al 2015)
	LowLightGAN (Kim et al 2019)	- Uses spectral normalization to stabilize the training, and produces its dataset based on local illumination synthesis	NIQE	Adversarial, perceptual, color, & TV	Synthetic based on DIV2K (Agustsson and Timofte 2017)
	(Yangming et al 2019)	- Combines Retinex Theory and GANs for image decomposition then enhancement	MSE, PSNR & SSIM	Regularization, adversarial, smooth L1, reconstruction, decomposition, enhancement, & MSSSIM	CSID (Chen et al 2018)
	DPE (Chen et al 2018)	- Utilizes paired and unpaired training settings based on GANs for enhancement	PSNR & SSIM	L2 & adversarial	MIT-Adobe 5K (Bychkovsky et al 2011), & HDR images selected from Flickr images
	EnlightenGAN (Jiang et al 2019)	- Performs unsupervised learning, using a U-Net (Ronneberger et al 2015) as the backbone for the generator, and includes global and local relativistic discriminators to handle spatially varying light conditions	NIQE	Non-reference: self-feature preserving, & adversarial	Collected from LOL (Wei et al 2018), RAISE (Dang-Nguyen et al 2015), & HDR sources (Gharbi et al 2017; Kalantari and Ramamoorth 2017)
Zero Reference	Zero-DCE (Guo et al 2020)	- Learns image-to-light enhancement curve mappings through a lightweight CNN model - Does not require paired or unpaired training data	PSNR, SSIM, MAE & PI (Blau and Michaeli 2018)	Non-reference: SC, EC, CC & IS	360 multi-exposure sequences from part1 of SICE (Cai et al 2018)
	(Zhang et al 2020)	- Introduces a referenceless loss function designed based on the Retinex model and entropy theory to train a self-supervised CNN model - Does not require paired or unpaired training data	GE, CE, GMI, GMG, PSNR, SSIM, LOE, & NIQE	Non-reference: reconstruction, reflectance, & IS	LOL (Wei et al 2018)

[2] PSNR: Peak Signal to Noise Ratio, SSIM: Structural Similarity Index, LOE: Lightness Order Error, SNM: Structure Natural Measure, NIQE: Natural Image Quality Evaluator, DE: Discrete Entropy, IR: Invariable Reflectance, IS: Illumination Smoothness, AB: Average Brightness, VIF: Visual Information Fidelity, TV: Total Variation, MSSIM: Multiscale SSIM, IQA: Image Quality Assessment, OAV: Open Available Vehicle, IWSSIM: Information-Weighted SSIM, PI: Perpetual Index, SC: Spatial Consistency, EC: Exposure Control, CC: Color Constancy, GE: Gray Entropy, CE: Color Entropy, GMI: Gray Mean Illumination, and GMG: Gray Mean Gradient

[3] http://decsai ugr es/cvg/dbimagenes/

Chapter 5

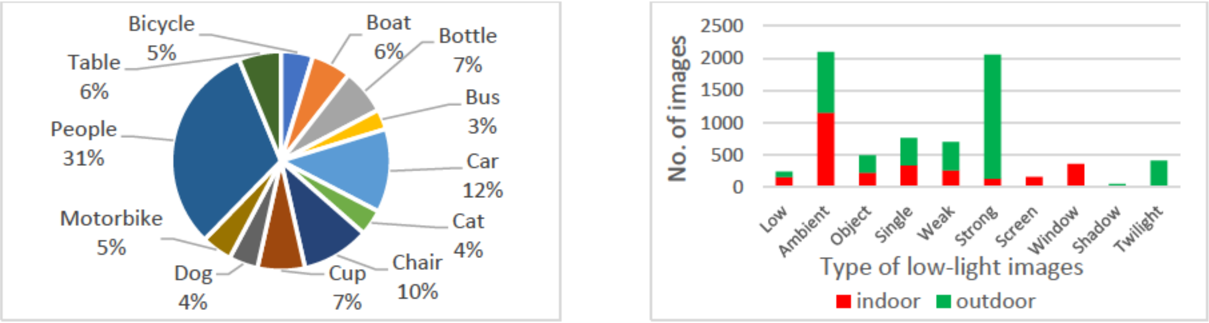
Comparative Study

In this chapter we describe the empirical experiments performed to compare the enhancement models and present the results along with analysis. Section 5.1 describes the test data and experimental setup. Section 5.2 presents the results of Experiment 1: comparing visual and perceptual LLI enhancement quality, Section 5.3 covers Experiment 2: comparing object classification and detection quality, and Sections 5.4 and 5.5 cover Experiment 3: comparing image feature maps and occlusion results. Section 5.6 presents a recap of the results and highlights interesting future directions.

5.1 Overview

As mentioned previously, one of the main challenges facing LLI enhancement is the lack of common benchmarks and metrics for empirical evaluation. Hence, in this section, we describe the test data, experiments, and evaluation metrics that we adopt in our study.

5.1.1 Test Data



(a) Object occurrences in dataset

(b) Distribution of images on illumination types

Figure 1: ExDark statistics reported from (Loh and Chan 2019)

We use two well-known datasets to conduct our empirical evaluation: ExDark (Loh and Chan 2019) and LOL (Wei et al. 2018). ExDark consists of 7,363 LLIs captured in real-world low-light environments, and contains 12 different object classes like people, cats, dogs, bicycles, etc. (Fig. 1a). Every instance of the 12 classes is associated with a bounding box annotation making the dataset applicable for training and evaluation on object detection and classification models. In addition, the dataset is split among 10 types of low-light conditions found in indoor and outdoor environments, varying from extremely dark images to images with spatially varying illumination depending on the location and the presence of light sources (Fig. 1b).

LOL (Low-light) (Wei et al. 2018) is made of 500 LLI/NLI pairs. The NLIs refer to a variety of real scenes taken in houses, campuses, clubs, etc. Yet, most of their LLI counterparts are created by changing the camera exposure and ISO sensitivity of the image sensor in order to simulate low-light conditions and thus they do not represent real low-light environments (Loh and Chan 2019). Hence, we refer to LOL as a quasi-synthetic dataset. Note that LOL images do not contain moving objects (such as people, animals, and vehicles) as the image pairs require exact position matching between LLI/NLI pairs (sample LOL image pairs are shown in Fig. 2).

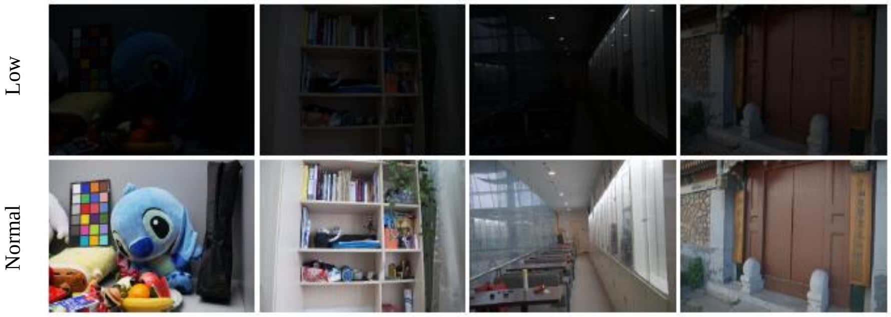


Figure 2: Sample pairs of LLIs/NLIs from the LOL dataset (Wei et al. 2018)

5.1.2 Experiments and Metrics

Our empirical evaluation consists of three main experiments: i) *visual and perceptual quality evaluation*, ii) *detection and classification quality evaluation*, and iii) *feature analysis*.

5.1.2.1 Experiment 1 – Perceptual and Visual Quality

In this experiment, we perform an image quality assessment (IQA) that aims at evaluating whether an image is visually pleasing and how it is visually perceived. Image quality refers to the different visual attributes of the image and focuses on the perceptual assessment of viewers. IQA methods are generally either i) quantitative: based on objective evaluation metrics, or ii) qualitative: based on the human perception of visual quality. In this study, we conduct both quantitative and qualitative evaluations, by comparing the visual quality achieved by 10 of the recent DL-based LLI enhancement models.

Quantitative comparison: We evaluate the enhancement models against four objective evaluation metrics commonly used in the literature: i) Natural Image Quality Evaluator (*NIQE*) (Mittal et al. 2013), ii) Blind/Reference-less Image Spatial Quality Evaluator (*BRISQUE*) (Mittal et al. 2012), iii) Structural Similarity Index (*SSIM*) (Wang et al. 2004) and iv) Peak Signal to Noise Ratio (*PSNR*).

NIQE (Mittal et al. 2013) is a *non-reference* metric or “blind” evaluation metric in which only the LLIs are available for assessment. It measures the deviations from statistical regularities seen in natural images without training on human rated distorted images or even exposure to distorted images. The quality of the test image represents the distance between a multivariate Gaussian (MVG) fit of the natural scene statistic (NSS) features derived from the test image, and a MVG model of the quality aware features extracted from a corpus of natural images.

BRISQUE (Mittal et al. 2012) is also a *non-reference* evaluation metric. It belongs to a class of opinion-aware metrics which evaluate the image based on models trained on databases of human rated distorted images and associated subjective opinion scores. In *BRISQUE*, the extracted features are derived based on a spatial natural scene statistical model. Then, a mapping is learned between the feature space and human based quality scores using a regression module, namely a support vector machine regressor (SVM-R) (Schölkopf et al. 2000).

SSIM (Wang et al. 2004) is a *full reference* metric in which a known reference image is needed for assessment. It measures the structural similarity between images based on independent comparisons of their luminance, contrast, and structure features. Given a ground truth image x with N pixels and maximum pixel value L, and given the corresponding enhanced image y, a simplified version of *SSIM* is defined as follows (Wang et al. 2004):

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C1)(2\sigma_{xy} + C2)}{(\mu_x{}^2 + \mu_y{}^2 + C1)(\sigma_x{}^2 + \sigma_y{}^2 + C2)} \tag{3}$$

,where $\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i$, $\sigma_x = (\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)^2)^{1/2}$, $\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)$,

$C1 = (k_1 L)^2$ and $C2 = (k_2 L)^2$ are constants for avoiding instability, with $k_1 << 1$ and $k_2 << 1$.

PSNR is another commonly used *full reference* evaluation metric. It is defined using the maximum pixel value (denoted as *L*) and the mean squared error (MSE or L2 loss) between images. Given a ground truth image x with *N* pixels and the corresponding enhanced image *y*, the *PSNR* between *x* and *y* is defined as follows:

$$PSNR(x, y) = 10 \times \log_{10}\left(\frac{L^2}{\frac{1}{N}\sum_{i=1}^{N}(x(i) - y(i))^2}\right) \tag{4}$$

In addition to the above objective metrics, we also evaluate the *noise level* in the enhanced images to acquire a complete viewpoint of the enhancement quality achieved by the models. We follow the approach proposed in (Liu et al. 2012) which relies on a patch-based noise level estimation algorithm. The algorithm selects weak textured patches from a single noisy image based on the gradients of the patches and their statistics. Then it estimates the noise level from the selected patches using principal component analysis.

Qualitative comparison: In addition to the quantitative study, we also perform a qualitative evaluation to assess the human visual perception of images enhanced by the 10 models used in this experiment. To do so, we randomly select 20 LLIs from our test data, i.e., 10 from each dataset (ExDark and LOL), and display them along with their enhanced counterparts in two dedicated surveys (for the LOL survey, we also display the corresponding NLIs)[4]. Responders are asked to rate each image considering six visual

[4] ExDark: https://cutt.ly/0fN4evQ, and LOL: https://cutt.ly/TfN4r6G

IQA criteria including: i) level of illumination, ii) level of exposure (over/under-exposed regions), iii) level of noise, iv) color deviations, v) clearness of contents and details, and vi) overall beauty. A total of 32 testers (senior computer engineering and master's students) were invited to contribute to the experiment, where 16 testers participated in each survey and independently rated every enhancement model on an integer scale from 1 to 10 (i.e., worst to best). A total of 1,600 responses were collected for each dataset, with every model receiving 160 rating scores. The ratings are aggregated for every enhancement model to evaluate its overall visual perceptual quality.

5.1.2.2 Experiment 2 – Detection and Classification Quality

In this experiment, we compare the performance achieved by 4 different object detection and classification models applied on the enhanced images from ExDark dataset using the 10 LLI enhancement methods considered in the previous experiment. We utilize mean Average Precision (*mAP*) as a commonly used metric to assess object detection and classification quality. For each object class, we generate the corresponding Precision-Recall (*P-R*) curve and compute the Average Precision (*AP*) per class from the area covered under the *P-R* curve. We then compute *mAP* for the object detection model as the average of the *AP* scores calculated for all the classes.

5.1.2.3 Experiment 3 – Feature Analysis

In this experiment, we compare the feature maps extracted from the LLIs, NLIs, and enhanced images from the LOL dataset using one of the object detection models from Experiment 2. A feature map is an $m \times n$ matrix which represents the output of a filter applied to a layer of the object detection model. A layer in a DL-based model usually consists of a sequence of feature maps. In this experiment, we consider the feature maps from three sample layers of the detection model: i) a sequence of large (e.g., 64×64 cell) maps from one of the layers belonging to the model's backbone, ii) a sequence of smaller (e.g., 16×16 cell) maps from an intermediary layer, and iii) a sequence of minimal size (e.g, 1×1 cell) maps from the model's last layer. To our knowledge, this is the first quantitative feature map evaluation study of its kind in the literature. We introduce two new metrics to compare feature maps: i) Feature Map Matrix Similarity (*FMMS*), and ii) top-*N* Active Feature Map Similarity (*topN-AFMS*). *FMMS* computes the cosine similarity measure between the feature maps of two (sets of) images at a given layer of the DL model, highlighting overall image feature similarity. More formally, given two images *x* and *y* whose feature maps are extracted at layer *n* of the DL model:

$$FFMS(x,y) = \frac{\sum_{i=1\ \ |n|} Sim_{Cosine}(F_i^x, F_i^y)}{|n|} \in [0,1], \tag{5}$$

where F_i^x and F_i^y are the i^{th} feature maps of images *x* and *y*, |*n*| the number of feature maps at layer *n*, and Sim_{cosine} the legacy cosine matrix similarity measure[5]:

$$Sim_{Cosine}(F_i, F_j) = \frac{\sum_q \sum_r w_i(q,r) \times w_j(q,r)}{\sqrt{\sum_q \sum_r w_i(q,r)^2 \times \sum_q \sum_r w_j(q,r)^2}} \in [0,1], \tag{6}$$

where $w_i(q, r)$ is the feature map (matrix) F_i position at coordinates *q* and *r*. Note that *FMMS* can be extended to compare two sets of pair-wise matching images (e.g., comparing sets of LLI/NLI, LLI/enhanced, or NLI/enhanced image pairs) by computing the similarity between every matching pair and then averaging over the total number of image pairs.

As for *topN-AFMS*, it compares the most active feature maps between two sets of pair-wise matching images, in order to help describe the behavior of a detection model and its response activity against the fed images. Identifying the most active feature maps gives insight into the features that might be most impactful on object detection and classification quality. Given two sets of pair-wise matching images $X=\{x_1, \ldots, x_t\}$ and $Y=\{y_1, \ldots, y_t\}$ where doublet (x_i, y_i) designates a matching pair (e.g., LLI/NLI, LLI/enhanced, or NLI/enhanced), and given the images' feature maps extracted at layer *n* of the DL model, we produce two vectors $V_X = < w_X(1), \ldots, w_X(|n|) >$ and $V_Y = < w_Y(1), \ldots, w_Y(|n|) >$ of size |*n*| each, where weights $w_X(i)$ and $w_Y(i)$ designate the number of times feature map *i* at layer *n* occurs among the top-active feature maps (based on their average) in image set *X* and *Y*

[5] We adopt the cosine measure due to its common usage in the literature (McGill 1983), yet other vector or matrix similarity measures could have been used such as Pearson Correlation Coefficient or Dice.

respectively. For instance, $w_X(i)$=10 means that feature map *i* has been identified 10 times (i.e., in 10 different images of set *X*) as one of the top active feature maps at layer *n*. Consequently, computing *topN-AFMS* between image sets *X* and *Y* comes down to computing the similarity between their vectors:

$$\begin{aligned} \text{top}N\text{-AFMS}(\text{X},\text{Y}) &= \text{Sim}_{\text{Cosine}}(\text{V}_\text{X},\text{V}_\text{Y}) \\ &= \frac{\sum_{\text{i}=1\ldots|n|} w_X(\text{i}) \times w_Y(\text{i})}{\sqrt{\sum_{r=1\ldots|n|} w_X(\text{i})^2 \times \sum_{r=1\ldots|n|} w_Y(\text{i})^2}} \in [0,1] \end{aligned} \tag{7}$$

where *N* is the number of most active feature maps at a certain layer of the DL model (e.g., we consider *N*=16 and compute the *top16-AFMS* in our empirical study, cf. Section 5.4.1).

In addition to computing *FFMS* and *topN-AFMS*, we utilize the occlusion experiment proposed in (Zeiler and Fergus 2014), where a black square is used to mask particular regions of an image while monitoring the output of the object detection model. The black square is slided over all the regions of the image allowing to produce a heatmap describing object detection confidence scores (in case of a detection – zero scores are produced otherwise). The significance of the experiment lays in the fact that the output of the object detection model should not change when the regions that are not so important for detection are occluded, and should vanish when the regions responsible of the detection are occluded. The occlusion experiment is applied on images containing a single object. If an image has a lot of regions that result in a misdetection if occluded, then we say the image holds weak features allowing to easily misdetect its object. Contrarily, if an image has no specific region that causes misdetection when occluded, then the image maintains strong features allowing to detect its object despite occlusion. While the authors in (Zeiler and Fergus 2014) describe the occlusion experiment, yet they do not define a quantitative approach to evaluate its results. Here, we introduce an objective metric: Occlusion based Average Misdetection Regions (*OAMR*) that quantifies the average number of regions contributing to misdetecting objects in a set of images. More formally, given a set of images $X=\{x_1, \ldots, x_n\}$ with *n* images of same size and a fixed size black box sliding over all the regions of the image then:

$$OAMR(X) = \frac{\sum_{i=1...n} C_{x_i}}{n}, \tag{8}$$

where $C_{x_i}$is the count of regions contributing to a misdetection in image x_i. A low *OAMR* indicates that a small number of regions causes misdetections, meaning that the images mostly contain strong features contributing to high object detection quality. A high *OAMR* indicates that many regions cause misdetections, and thus the images hold weak features leading to low object detection quality. In short, high quality LLI enhancement models would minimize *OAMR*.

5.2 Experiment 1: Perceptual and Visual Quality

In this section, we present quantitative and qualitative evaluations of the performance achieved by 10 recent DL-based LLI enhancement models, namely: RetinexNet[6] (Wei et al. 2018), GladNet[7] (Wang et al. 2018), LLNet[8] (Lore et al. 2017), LightenNet[9] (Li et al. 2018), DPE[10] (Chen et al. 2018), EnlightenGAN[11] (Jiang et al. 2019), MBLLEN[12] (Lv et al. 2018), DeepUPE[13] (Wang et al. 2019), RDGAN[14] (Wang et al. 2019), and Zero-DCE[15] (Guo et al. 2020). We run the latter on both ExDark and LOL datasets using the models' pre-trained weights and author-recommended configurations which are publicly available online.

5.2.1 Quantitative Comparison

To perform a quantitative evaluation, we process the results produced by each of the mentioned DL models through four commonly used metrics in the literature: i) Natural Image Quality Evaluator (*NIQE*) (Mittal et al. 2013), ii) Blind/Reference-less Image Spatial Quality Evaluator (*BRISQUE*) (Mittal et al. 2012), iii) Structural Similarity Index (*SSIM*) (Wang et al. 2004) and iv) Peak Signal to Noise Ratio (*PSNR*) (cf. Section 5.1.2.1).

[6] https://github.com/weichen582/RetinexNet
[7] https://github.com/weichen582/GLADNet
[8] https://github.com/kglore/llnet_color
[9] https://li-chongyi.github.io/sub_projects html
[10] https://github.com/UtopiaHu/Deep-Photo-Enhancer
[11] https://github.com/TAMU-VITA/EnlightenGAN
[12] https://github.com/Lvfeifan/MBLLEN
[13] https://github.com/wangruixing/DeepUPE
[14] https://github.com/WangJY06/RDGAN
[15] https://github.com/Li-Chongyi/Zero-DCE

We use the first two metrics, i.e., *NIQE* and *BRISQUE*, to evaluate the enhanced images from the ExDark and LOL datasets as stand-alone images without referring to their NLI counterparts. We use the third and fourth metrics to evaluate the enhanced images from the LOL dataset against their reference NLI counterparts. We also evaluate noise levels in the enhanced images[16] as an added indicator of the enhancement quality achieved by the models. In addition to the 10 models being evaluated, we also provide the scores obtained for the original LLIs, which we use as a reference to compare with the latter. Models producing *NIQE* and *BRISQUE* scores that are lower/higher than the original LLI scores are considered to be better/worse in improving the visual quality of the LLIs, and models producing *PSNR* and *SSIM* scores that are higher/lower than the original LLI scores are considered to be better/worse in improving the visual quality of the LLIs. Results for the ExDark and LOL datasets are provided in Table 2 and Table 3 and samples are visualized in Fig. 4 and Fig. 5.

Based on the results in Table 2 and Table 3, we highlight the following observations:

- **Results of the *NIQE* and *BRISQUE* metrics do not seem correlated**: Models that perform well following the first metric might perform poorly with the second. Considering the ExDark dataset for instance, MBLLEN has the best *NIQE* score while showing the third best *BRISQUE* score. Also, RDGAN achieves the best *BRISQUE* score while showing the fourth worst *NIQE* score. Similarly, for the LOL dataset, LLNet which is the best following the *NIQE* metric produces the worst *BRISQUE* score.

Table 2: Results for the ExDark dataset, ranked from best (#1) to worst (#10) following each of the metrics (red color refers to the best score and green to the second best for every metric).

(a) Results ranked following *NIQE* ↓

Approach	**Rank**	***NIQE***	***BRISQUE***	**Noise**
MBLLEN	1	**3 26**	25.75	**19.95**
EnlightenGAN	2	**3.41**	**25.34**	20.06
GladNet	3	3 58	27.07	20.148
DPE	4	3.64	29.05	20.108
LightenNet	5	3.657	27.53	20.16
DeepUPE	6	3.658	28.09	20.106
RDGAN	7	3.66	**24.75**	20.150
Original LLIs	---	3.69	30.52	20.00
LLNet	8	3 91	32.56	**19.77**
Zero-DCE	9	3 98	29.89	20.29
RetinexNet	10	4 12	30.84	20.09

(b) Results ranked following *BRISQUE* ↓

Approach	**Rank**	***NIQE***	***BRISQUE***	**Noise**
RDGAN	1	3.66	**24.75**	20 150
EnlightenGAN	2	**3.41**	**25.34**	20.06
MBLLEN	3	**3.26**	25.75	**19 95**
GladNet	4	3.58	27.07	20 148
LightenNet	5	3.657	27.53	20 16
DeepUPE	6	3.658	28.09	20 106
DPE	7	3.64	29.05	20 108
Zero-DCE	8	3.98	29.89	20 29
Original LLIs	---	3.69	30.52	20.00
RetinexNet	9	4.12	30.84	20.09
LLNet	10	3.91	32.56	**19.77**

(c) Results ranked following *Noise Level* ↓

Approach	**Rank**	***NIQE***	***BRISQUE***	***Noise***
LLNet	1	3.91	32.56	**19.77**
MBLLEN	2	**3.26**	25.75	**19 95**
Original LLIs	---	3.69	30.52	20.00
EnlightenGAN	3	**3.41**	**25.34**	20.06
RetinexNet	4	4.12	30.84	20.09
DeepUPE	5	3.658	28.09	20.106
DPE	6	3.64	29.05	20.108
GladNet	7	3.58	27.07	20.148
RDGAN	8	3.66	**24.75**	20.150
LightenNet	9	3.657	27.53	20 16
Zero-DCE	10	3.98	29.89	20 29

[16] For noise level evaluation, we use: i) a random subset of 500 images from ExDark including 50 images from each of the 10 different lighting conditions, as well as ii) the whole LOL dataset.

Table 3: Results for the LOL dataset, ordered from best (#1) to worst (#10) following each of the metrics (red color refers to the best score and green to the second best for every metric).

(a) Results ranked following *NIQE*↓

Approach	Rank	*NIQE*	*BRISQUE*	*SSIM*	*PSNR*	Noise
LLNet	1	**4.17**	33.03	**0.66**	**17.50**	**19.75**
MBLLEN	2	**4.22**	**20.38**	0.59	17.30	**19.99**
EnlightenGAN	3	4.97	24.41	0.60	16.25	20.15
RetinexNet	4	5.30	**24.19**	0.57	16.23	20.55
Original LLIs	---	6.03	24.87	0.16	7.74	19.99
DPE	5	6.64	24.51	0.371	9.41	20.09
RDGAN	6	7.04	27.85	0.62	14.97	20.49
GladNet	7	7.23	28.67	**0.67**	**19.26**	20.86
LightenNet	8	7.68	28.81	0.370	10.13	20.25
DeepUPE	9	7.81	27.91	0.39	10.57	20.11
Zero-DCE	10	8.54	31.80	0.54	14.16	21.35

(b) Results ranked following *BRISQUE*↓

Approach	Rank	*NIQE*	*BRISQUE*	*SSIM*	*PSNR*	Noise
MBLLEN	1	**4.22**	**20.38**	0.59	17.30	**19.99**
RetinexNet	2	5.30	**24.19**	0.57	16.23	20.55
EnlightenGAN	3	4.97	24.41	0.60	16.25	20.15
DPE	4	6.64	24.51	0.371	9.41	20.09
Original LLIs	---	6.03	24.87	0.16	7.74	19.99
RDGAN	5	7.04	27.85	0.62	14.97	20.49
DeepUPE	6	7.81	27.91	0.39	10.57	20.11
GladNet	7	7.23	28.67	**0.67**	**19.26**	20.86
LightenNet	8	7.68	28.81	0.370	10.13	20.25
Zero-DCE	9	8.54	31.80	0.54	14.16	21.35
LLNet	10	**4.17**	33.03	**0.66**	**17.50**	**19.75**

(c) Results ranked following *SSIM* ↑

Approach	Rank	*NIQE*	*BRISQUE*	*SSIM*	*PSNR*	Noise
GladNet	1	7.23	28.67	**0.67**	**19.26**	20.86
LLNet	2	**4.17**	33.03	**0.66**	**17.50**	**19.75**
RDGAN	3	7.04	27.85	0.62	14.97	20.49
EnlightenGAN	4	4.97	24.41	0.60	16.25	20.15
MBLLEN	5	**4.22**	**20.38**	0.59	17.30	**19.99**
RetinexNet	6	5.30	**24.19**	0.57	16.23	20.55
Zero-DCE	7	8.54	31.80	0.54	14.16	21.35
DeepUPE	8	7.81	27.91	0.39	10.57	20.11
DPE	9	6.64	24.51	0.371	9.41	20.09
LightenNet	10	7.68	28.81	0.370	10.13	20.25
Original LLIs	---	6.03	24.87	0.16	7.74	19.99

(d) Results ranked following *PSNR*↑

Approach	Rank	*NIQE*	*BRISQUE*	*SSIM*	*PSNR*	Noise
GladNet	1	7.23	28.67	**0.67**	**19.26**	20.86
LLNet	2	**4.17**	33.03	**0.66**	**17.50**	**19.75**
MBLLEN	3	**4.22**	**20.38**	0.59	17.30	**19.99**
EnlightenGAN	4	4.97	24.41	0.60	16.25	20.15
RetinexNet	5	5.30	**24.19**	0.57	16.23	20.55
RDGAN	6	7.04	27.85	0.62	14.97	20.49
Zero-DCE	7	8.54	31.80	0.54	14.16	21.35
DeepUPE	8	7.81	27.91	0.39	10.57	20.11
LightenNet	9	7.68	28.81	0.371	10.13	20.25
DPE	10	6.64	24.51	0.370	9.41	20.09
Original LLIs	---	6.03	24.87	0.16	7.74	19.99

(e) Results ranked following *Noise Level* ↓

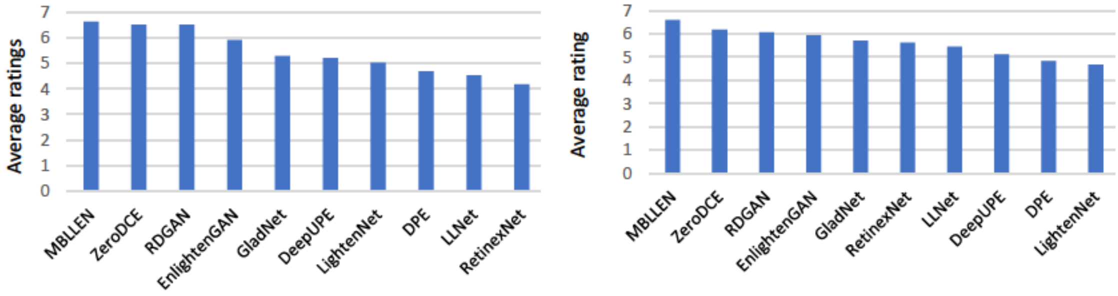
Approach	Rank	*NIQE*	*BRISQUE*	*SSIM*	*PSNR*	Noise
LLNet	1	**4.17**	33.03	**0.66**	**17.50**	**19.75**
Original LLIs	---	6.03	24.87	0.16	7.74	19.99
MBLLEN	2	4.22	**20.38**	0.59	17.30	**19.99**
DPE	3	6.64	24.51	0.37	9.41	20.09
DeepUPE	4	7.81	27.91	0.39	10.57	20.11
EnlightenGAN	5	4.97	24.41	0.60	16.25	20.15
LightenNet	6	7.68	28.81	0.37	10.13	20.25
RDGAN	7	7.04	27.85	0.62	14.97	20.49
RetinexNet	8	5.30	24.19	0.57	16.23	20.55
GladNet	9	7.23	28.67	**0.67**	**19.26**	20.86
Zero-DCE	10	8.54	31.80	0.54	14.16	21.35

- **LOL dataset results show discrepancies among *NIQE*, *BRISQUE*, *SSIM*, and *PSNR* metrics**: For instance, GladNet is ranked as the fourth worst model following *NIQE* and *BRISQUE,* yet it shows the best *SSIM* and *PSNR* scores. Also, MBLLEN which achieves the best score for *BRISQUE* and the second best for *NIQE* comes only at the fifth place following *SSIM*.

- **Most of the models produce noise levels higher than the original LLIs, with a few exceptions:** Most enhancement models tend to amplify or integrate noise into the enhanced images, except for LLNet and MBLLEN with both ExDark and LOL. LLNet achieves the minimum noise levels as it tends to over-smooth image details. MBLLEN produces the second lowest noise levels and is consistent with the good scores achieved by both *NIQE* and *BRISQUE* metrics with both ExDark and LOL datasets thus highlighting its good enhancement performance. EnlightenGAN produces some of the best scores on the ExDark dataset and adds a minimal amount of noise. In contrast, Zero-DCE consistently shows the highest noise levels and produces some of the worst *NIQE* and *BRISQUE* scores, indicating that a high noise level tends to distort enhancement quality.

5.2.2 Qualitative Comparison

In addition to the quantitative study, we also perform a qualitative evaluation, where 32 participants were asked to rate samples of enhanced images from each of the two datasets used in our experiments (Fig. 4 and Fig. 5), providing their perception of the images' visual quality. Every sample image was independently rated on an integer scale from 1 to 10 (i.e., worst to best). Results for each enhancement model are compiled in Fig. 3.



(a) ExDark dataset scores **(b) LOL dataset scores**

Figure 3: Average tester rating scores compiled for every enhancement model, and ranked from best to worst

Based on the results in Fig. 3, we highlight the following observations:

- **Regarding the ExDark dataset**: MBLLEN tends to entirely illuminate the images to look visually pleasing and beautiful (e.g. bicycle and cat in Fig. 4a, c) and is ranked as the best enhancement model. Images enhanced by Zero-DCE and RDGAN show good illumination levels and well-preserved contents, and are ranked as second and third best models. EnlightenGAN tends to produce visually pleasing images with no over or under exposed regions, and is ranked as the fourth best model. GladNet tends to increase image illumination but shows some color deviation and noise, and is ranked as the fifth best model. DeepUPE, LightenNet and DPE add minimal touches on the images and tend to show low illumination levels. They are ranked at the sixth, seventh, and eighth positions, respectively. LLNet increases image illumination, yet it also tends to over-smooth certain image details (e.g., pedestrian street in Fig. 4a). It is ranked as the ninth and second last model. Finally, RetinexNet is ranked as the tenth and last model as it produces significant noise and tends to over-expose certain artifacts in the enhanced images (e.g., bicycle in Fig. 4a).

- **Regarding the LOL dataset**: MBLLEN produces visually pleasing images with vivid and natural colors and is ranked as the best model thus demonstrating its good enhancement quality. RDGAN, Zero-DCE, and EnlightenGAN show naturalistic colors with preserved contents and texture. They are ranked at the second, third, and fourth positions, respectively. GladNet sufficiently boosts image illumination but usually shows pale colors and tends to add noise. It is ranked at the fifth position. RetinexNet boosts image illumination while showing exposed artifacts. It is ranked at the sixth position. LLNet tends to highly smoothen image details while showing pale lighting, and is ranked at the seventh position. DeepUPE, DPE and LightenNet minimally enhance image illumination and tend to incorporate noise into the images. Together as a group, they are ranked as the three worst models.

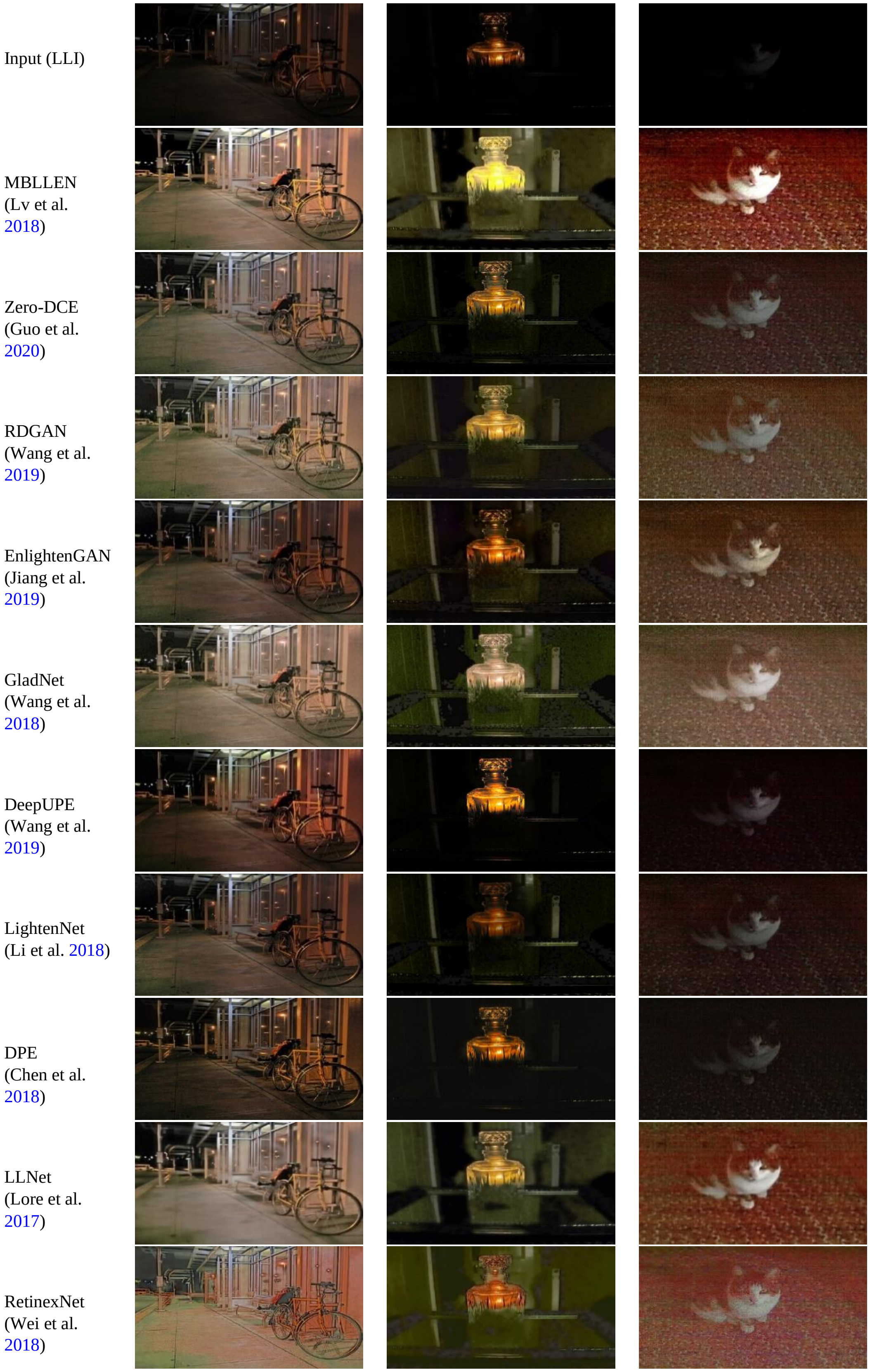
5.2.3 Discussion

To sum up, we review and summarize the results of both quantitative and qualitative tests.

First, concerning the *quantitative evaluation metrics*: most metrics used in this study fail to produce model rankings which closely match the qualitative (human) evaluation rankings. For instance, *NIQE* results are compatible with the qualitative scores in certain aspects, by i) showing that MBLLEN achieves the best/second best enhancement quality for ExDark/LOL datasets, ii) showing that RetinexNet and LLNet are amongst the worst performing models when applied on the ExDark dataset, and iii) producing very close scores for DPE, LightenNet, and DeepUPE which only perform slight enhancement to the images of ExDark, in accordance with their sequential human rankings. However, *NIQE* does not show consistent results when it comes to capturing the illumination and noise components in the enhanced images. For instance, in the case of the ExDark dataset, Zero-DCE and RDGAN are ranked among the worst models following *NIQE* as they produce high noise levels (Table 2c). Yet, they are relatively better ranked by human testers, producing higher scores than DPE, LightenNet, and DeepUPE which have better *NIQE* scores and lower illumination levels. This might be due to the fact that the noise produced by Zero-DCE and RDGAN is not clearly apparent in the images and maybe visually overlooked by the users in favor of good illumination. Moreover, LLNet shows the best *NIQE* score while maintaining the lowest noise level for the LOL dataset, yet it exhibits the fourth worst human scores due to the over-smoothing and the exposed artifacts it produces. *BRISQUE* shows similar inconsistencies while quantifying image illumination. For example, Zero-DCE shows higher *BRISQUE* scores compared with DeepUPE, DPE, and LightenNet, and yet surpasses the latter models in terms of human tester ratings. This is probably due to the seemingly better illumination as perceived by most testers. In addition, all considered metrics fail to produce consistent rankings among themselves, suggesting the need to design more accurate objective metrics that behave in accordance with human visual perception.

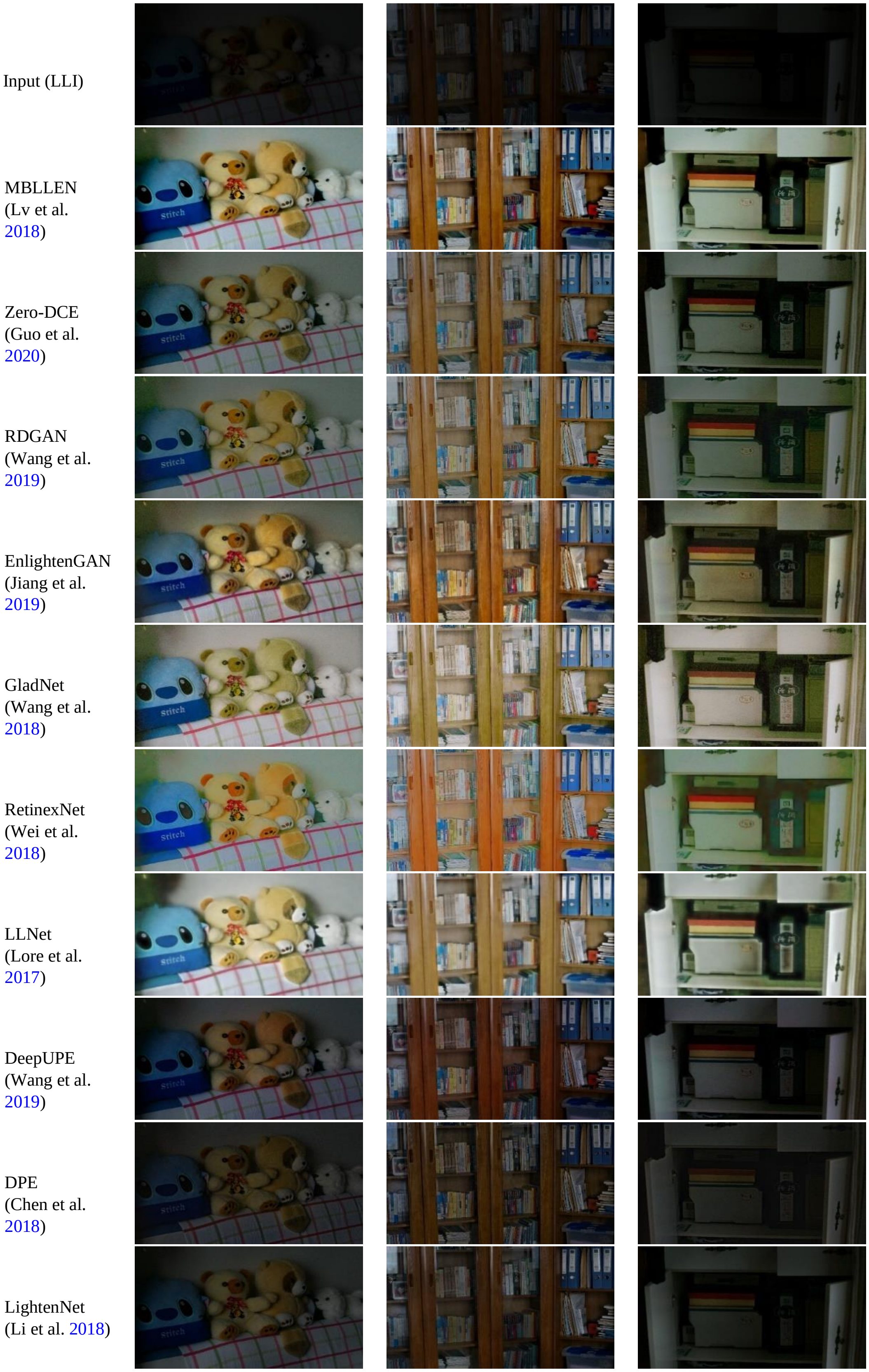
Second, concerning the *best performing models*: EnlightenGAN is consistently ranked among the best enhancement models on both ExDark and LOL datasets. Although its good performance on LOL can be due to using it as part of the model’s training dataset, yet its performance on ExDark proves its capability of generalizing to real world scenes. The results of this model highlight the potential of unsupervised GAN-based solutions in performing LLI enhancement. ZeroDCE is ranked as one of the best models following human ratings. It shows good illumination levels and preserves image contents, but tends to incorporate noise into the enhanced images (producing the highest noise levels for both datasets). The latter highlights the potential of ZeroDCE which reformulates the enhancement task using image-to-light curve estimation mapping, while eliminating the need for paired and unpaired training data. Also, the supervised MBLLEN model achieves some of the best quantitatively and qualitatively enhancement results on both ExDark and LOL datasets. This may be due to its large training dataset of synthetic LLIs (16,925 images) generated based on the Pascal VOC (Everingham et al. 2012) object detection and classification benchmark, allowing it to better generalize and handle real-world LLIs (namely those in ExDark and LOL). In addition, MBLLEN extracts and enhances the features at every layer of the used CNN model thus allowing global and local level feature enhancement.

Third, concerning the *noise element*: most enhancement models tend to incorporate significant noise into the enhanced images, thus distorting their quality. Notably, LLNet achieves minimal noise levels on both datasets, while sufficiently boosting image illumination. Its underlying Stacked Sparse Denoise Autoencoder (SSDA) (Lore et al. 2017) seems promising and could be effective if properly tuned and designed to maintain a good balance between noise suppression and over-smoothing. Nonetheless, we note that the noise factor and de-noising techniques need to be given special attention, especially that the present evaluation metrics do not simultaneously quantify illumination and noise levels. This might suggest the need for new and more robust metrics that are consistent with the humans' visual perception of enhanced image quality.


Input (LLI)
MBLLEN (Lv et al. 2018)
Zero-DCE (Guo et al. 2020)
RDGAN (Wang et al. 2019)
EnlightenGAN (Jiang et al. 2019)
GladNet (Wang et al. 2018)
DeepUPE (Wang et al. 2019)
LightenNet (Li et al. 2018)
DPE (Chen et al. 2018)
LLNet (Lore et al. 2017)
RetinexNet (Wei et al. 2018)

(a) Sample 1 **(b)** Sample 2 **(c)** Sample 3

Figure 4: Visual human comparison of enhanced LLIs from ExDark dataset, ordered from best to worst


Input (LLI)
MBLLEN (Lv et al. 2018)
Zero-DCE (Guo et al. 2020)
RDGAN (Wang et al. 2019)
EnlightenGAN (Jiang et al. 2019)
GladNet (Wang et al. 2018)
RetinexNet (Wei et al. 2018)
LLNet (Lore et al. 2017)
DeepUPE (Wang et al. 2019)
DPE (Chen et al. 2018)
LightenNet (Li et al. 2018)
Stitch

(a) Sample 1 **(b)** Sample 2 **(c)** Sample 3

Figure 5: Visual human comparison of enhanced LLIs from the LOL dataset, ordered from best to worst

5.3 Experiment 2: Detection & Classification Quality

High-level computer vision tasks like object detection and classification usually suffer from a degraded performance when processing LLIs (Yang et al. 2020; VidalMata et al. 2020). In this experiment, we aim to verify whether enhancing LLI illumination and quality would improve the performance of the object detection task. To do so, we perform a comparative analysis using 4 object detection models: YOLOv3 (You Only Look Once version 3)[17] (Redmon and Farhadi 2018), RetinaNet[18] (Lin et al. 2017), SSD (Single Shot MultiBox Detector)[19] (Wei et al. 2016), and Mask RCNN (Region based CNN)[20] (He et al. 2017). We apply the models on the entire original ExDark dataset as well as its enhanced versions produced by the 10 enhancement models considered in our previous experiment.

5.3.1 Experimental Setup

We use the detection models' recommended weights, pre-trained on Microsoft COCO (Lin et al. 2014): a large-scale object detection, segmentation, and captioning dataset. This allows a generic evaluation, rather than training and fine-tuning the detection models using ExDark's LLIs or their enhanced counterparts, which would defeat the purpose of the experiment. After all, we aim to enhance LLIs to make them usable by existing detection models trained on large benchmark datasets of real world NLIs that are abundantly available. To do so, we leverage the ground truth bounding box annotations provided in the ExDark dataset to perform our experiments. We post-process the predictions provided by the detection models trained on COCO and fit them to ExDark. The COCO dataset consists of 80 different classes of objects, and ExDark consists only of 12 classes all of which are included in COCO. Few ExDark classes are more generic than their COCO counterparts, e.g., *couch* and *bench* classes in COCO are annotated as *chair* in ExDark, *wine glass* in COCO is annotated as *cup* in ExDark, and *truck* in COCO is annotated as *car* in ExDark. Hence, we match the classes from both datasets by converting COCO's

[17] https://github.com/ultralytics/yolov3
[18] https://github.com/fizyr/keras-retinanet
[19] https://github.com/pierluigiferrari/ssd_keras
[20] https://github.com/matterport/Mask_RCNN

predictions to the equivalent class annotations in ExDark and ignoring all the classes predicted by COCO that do not exist in ExDark, thus bounding the detections to ExDark's 12 classes. As for the comparison task, we utilize the mean Average Precision *(mAP*) metric commonly used to evaluate the performance of object detection and classification models in the literature (cf. Section 5.1.2.2).

5.3.2 Experimental Results

Table 4 presents the *mAP* results achieved by the 4 detection models, applied on the original ExDark dataset and its enhanced versions produced by the 10 enhancement models considered in our study. Results highlight the following observations:

- **The enhancement models produce consistent results:** They rank almost the same across the 4 considered detection models, with a few fluctuations between the top-ranked and bottom-ranked models. Namely, DeepUPE ranks as the top enhancement model with 3 out of the 4 detection models (and comes only at 2nd place with YOLOv3), whereas MBLLEN ranks at 1st place with YOLOv3 and comes 2nd with the other 3 models. At the other end of the spectrum, RetinexNet and EnlightenGAN consistently share the last two positions among the four detection models. The remaining enhancement models mostly rank the same across all detection models.

- **Minimal enhancement models produce good detection results:** DeepUPE, LightenNet, and DPE which perform minimal enhancement and show low illumination levels in Experiment 1 (cf. Section 5.2.2), are consistently ranked among the best models in terms of detection performance in this experiment. This can be reasoned to their minimalistic enhancement, which keep the enhanced images attached to their original LLIs, and thereby preserve their original features and semantics.

- **Object detection quality does not always correlate with visual enhancement quality**: A striking example is EnlightenGAN which holds some of the worst object detection results in this experiment, despite consistently producing some of the best LLI enhancement results in Experiment 1. EnlightenGAN uses a reference-less self-feature preserving loss based on a pre-trained VGG16 model (Simonyan and Zisserman 2015), which may not be able to effectively preserve the image features to itself, thus showing a degraded detection performance. On the opposite side of the spectrum, DeepUPE produces some of the best object detection results in this experiment, despite showing non-remarkable LLI enhancement results in Experiment 1. DeepUPE utilizes a pre-trained VGG16 encoder network to extract the image features before enhancing it. The powerful feature extraction capabilities of VGG16 might be a reason behind its good detection performance. Hence, an improved LLI visual enhancement quality does not seem to directly translate into improved detection and classification quality.

- **Few exceptions to the previous observations:** Results produced by MBLLEN and RetinexNet tend to contradict some of the previous observations. On the one hand, MBLLEN boasts some of the best enhancement quality levels on ExDark from Experiment 1, and consistently exhibits the second-best *mAP* levels across most detection models in the present experiment. The good *mAP* results can be attributed to MBLLEN's large-scale training dataset: PASCAL VOC (Everingham et al. 2012) consisting of 16,925 images containing dynamic objects and classes similar to those in the ExDark dataset, thus probably allowing for a better preservation of the image visual contents and semantics. On the other hand, RetinexNet bears some of the worst enhancement quality levels and produces the worst object detection quality levels. This means that image enhancement quality is not completely disassociated from detection quality and can affect the object detection task.

5.3.3 Discussion

To summarize the above observations: i) most enhancement models produce consistent results and behave similarly across the object detection models evaluated in our study, ii) object detection quality does not always correlate with visual enhancement quality, where good enhancement models seem to perform badly when used for object detection, and vice versa, and iii) a few exceptions to the previous observation show that image enhancement quality is not completely disassociated from object detection quality, and can improve the object detection task.

Interestingly, most detection models tend to perform better on the original LLIs, compared with the enhanced images. While this seems counter-intuitive, yet a similar observation was made in a recent study in (VidalMata et al. 2020), where the authors summarize the results of the UG^2 Challenge workshop held at IEEE CVPR 2018[21], and which aims at assessing the influence of image restoration and enhancement techniques in improving the performance of classification models like VGG16 and VGG19 (Simonyan and Zisserman 2015), InceptionV3 (Szegedy et al. 2016), and ResNet50 (He et al. 2016). Extensive experimentation on a new video benchmark dataset representing both ideal conditions and common aerial image artifacts, demonstrate that improving image quality does not necessarily lead to an improved classification performance, and may even degrade it in certain cases where images include extreme artifacts. However, the improvement in detection quality that is consistently produced by few enhancement models like MBLLEN and DeepUPE suggests that LLI enhancement can help improve object detection performance if designed in a special way to highlight and preserve the features of interest to the object detection task.

Also, one aspect that might affect object detection quality is the level of noise added in the enhanced images. By comparing with the results of Experiment 1, we realize that MBLLEN produces some of the lowest noise levels compared with the other enhancement models (Table 2c) while producing some of best detection results in this experiment. ZeroDCE and RDGAN which are ranked among the best enhancement models by human testers in Experiment 1 (Fig. 3), produce some of the worst noise levels (Table 2c) and show a degraded detection performance in this experiment. This suggests that a proper balancing between visual features and noise levels should be maintained to improve the detection task

[21] 2018 *IEEE* Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, June 18-22, 2018

Table 4: *mAP* results for the ExDark dataset, ranked from best (#1) to worst (#10) following each detection model (red color refers to the best score and green to the second best for every detection model).

(a) Results ranked following YOLOv3 (Redmon and Farhadi 2018)

Approach	Rank	YOLOv3	RetinaNet	SSD	Mask RCNN
MBLLEN	1	63.35	58.90	57.61	56.18
DeepUPE	2	63.24	59.14	57.74	56.64
LightenNet	3	62.30	57.77	56.51	54.66
Original LLIs	---	61.79	60.14	57.83	56.43
DPE	4	61.79	57.10	55.93	54.07
Zero-DCE	5	60.24	52.80	47.82	50.03
GladNet	6	58.37	50.98	51.51	49.35
RDGAN	7	58.22	50.53	51.14	48.98
LLNet	8	54.80	50.23	49.81	47.70
RetinexNet	9	48.14	38.13	40.87	32.94
EnlightenGAN	10	44.19	39.52	40.16	38.28

(b) Results ranked following RetinaNet (Lin et al. 2017)

Approach	Rank	YOLOv3	RetinaNet	SSD	Mask RCNN
Original LLIs	---	61.79	60.14	57.83	56.43
DeepUPE	1	63 24	59.14	57.74	56.64
MBLLEN	2	63 35	58.90	57.61	56.18
LightenNet	3	62 30	57.77	56.51	54.66
DPE	4	61.79	57.10	55.93	54.07
Zero-DCE	5	60 24	52.80	47.82	50.03
GladNet	6	58 37	50.98	51.51	49.35
RDGAN	7	58 22	50.53	51.14	48.98
LLNet	8	54.80	50.23	49.81	47.70
EnlightenGAN	9	44 19	39.52	40.16	38.28
RetinexNet	10	48 14	38.13	40.87	32.94

(c) Results ranked following SSD (Wei et al. 2016)

Approach	Rank	YOLOv3	RetinaNet	SSD	Mask RCNN
Original LLIs	---	61.79	60.14	57.83	56.43
DeepUPE	1	63.24	59.14	57.74	56.64
MBLLEN	2	63.35	58.90	57.61	56.18
LightenNet	3	62.30	57.77	56.51	54.66
DPE	4	61.79	57.10	55.93	54.07
GladNet	5	58.37	50.98	51.51	49.35
RDGAN	6	58.22	50.53	51.14	48.98
LLNet	7	54.80	50.23	49.81	47.70
Zero-DCE	8	60.24	52.80	47.82	50.03
RetinexNet	9	48.14	38.13	40.87	32.94
EnlightenGAN	10	44.19	39.52	40.16	38.28

(d) Results ranked following Mask RCNN (He et al. 2017)

Approach	Rank	YOLOv3	RetinaNet	SSD	Mask RCNN
DeepUPE	1	63 24	59.14	57.74	56.64
Original LLIs	---	61.79	60.14	57.83	56.43
MBLLEN	2	63 35	58.90	57.61	56.18
LightenNet	3	62 30	57.77	56.51	54.66
DPE	4	61.79	57.10	55.93	54.07
Zero-DCE	5	60 24	52.80	47.82	50.03
GladNet	6	58 37	50.98	51.51	49.35
RDGAN	7	58 22	50.53	51.14	48.98
LLNet	8	54.80	50.23	49.81	47.70
EnlightenGAN	9	44 19	39.52	40.16	38.28
RetinexNet	10	48 14	38.13	40.87	32.94

5.4 Experiment 3: Feature Analysis

In this experiment, we attempt to better understand the impact of LLI enhancement on high-level computer vision tasks by comparing the feature maps extracted from LLIs, NLIs, and their enhanced counterparts. To our knowledge, this is the first quantitative feature map evaluation study of its kind in the literature.

5.4.1 Experimental Setup

We present and discuss the results obtained by the SSD detection model (Wei et al. 2016) applied on images from the LOL dataset. We particularly utilize SSD512, a variant of SSD using 512 input image size pre-trained using the Microsoft COCO dataset (Lin et al. 2014). We chose this detection model for our analysis because of its convenient architecture (VGG16+6 extra sequential feature layers), comprising a sequence of stacked convolutional and pooling layers, and allowing to easily extract the feature maps of interest. In this experiment, we consider the feature maps from three sample layers of the detection model: i) a sequence of large maps from conv4_3 (64*64*512) – one of the layers belonging to the model's backbone, ii) a sequence of smaller maps from conv8_2 (16*16*512) – an intermediary layer, and iii) a sequence of minimal size maps from conv11_2 (1*1*256) – the model's last layer (Fig. 6).

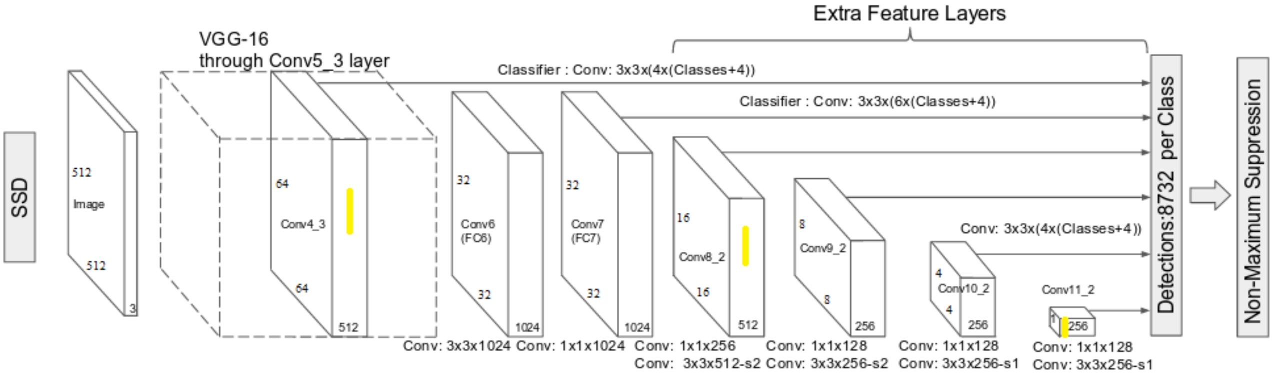
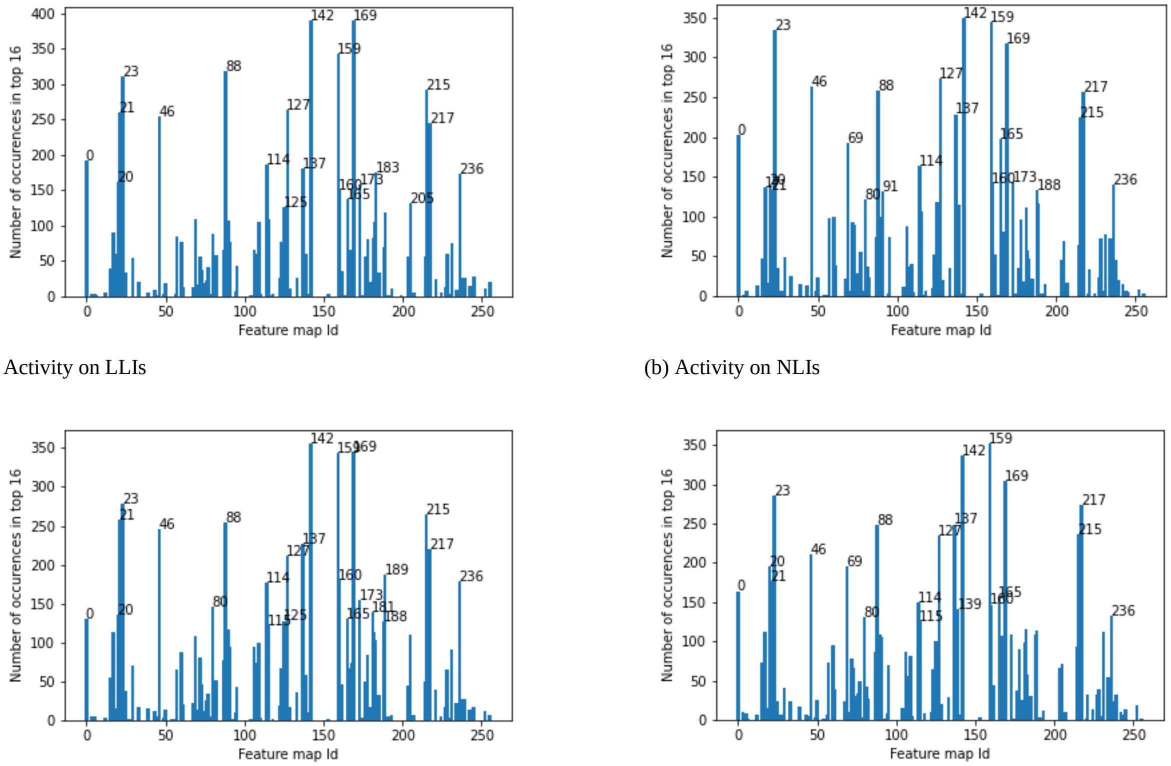


Figure 6: SSD512 architecture: layers marked in yellow are used in the analysis (modified based on (Wei et al. 2016))



(a) Activity on LLIs

(b) Activity on NLIs

(c) Activity on enhanced images using DPE (Chen et al. 2018)

(d) Activity on enhanced images using MBLLEN (Lv et al. 2018)

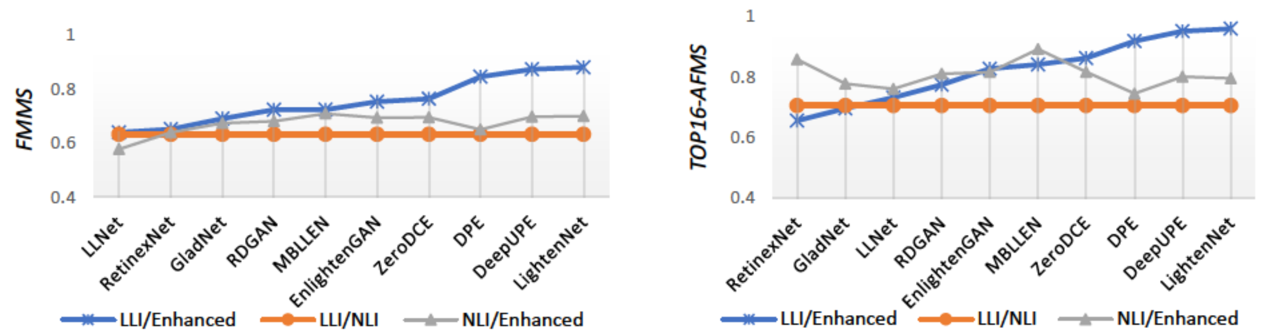
Figure 7: Visualizing top-16 active feature maps for layer conv11_2 for the LOL dataset. The labels on top of the bars refer to the feature map id within the layer

We introduce two new metrics to compare the feature maps at a given layer of the DL model: i) Feature Map Matrix Similarity (*FMMS*), and ii) top N Active Feature Map Similarity (*topN-AFMS*). *FMMS* computes the cosine similarity measure between the feature maps of two (sets of) images at given layer of the DL model, highlighting overall image feature similarity (Equation (5) in Section 5.1.2.3). Here, we expect that the enhanced images share maximum feature similarity (producing maximum *FMMS* scores) with NLIs, compared with their LLI counterparts. In other words, we expect a high-quality enhanced image to share more similar features with a NLI, compared with a LLI. *TopN-AFMS* compares the most active feature maps between two sets of pair-wise matching images, to help describe the behavior of the detection model and its response activity against the input images (Equation (7) in Section 5.1.2.3). The most active and responsive feature maps are those having the highest average activation at a certain layer of the DL model, while feature maps having zero average activations refer to dead or inactive maps. Identifying and comparing the most active feature maps gives insight into the features that might be most impactful on object detection and classification quality.

For instance, Fig. 7 shows the number of occurrences of the top-16 active feature maps extracted using SSD at layer conv11_2, considering all the images from the LOL dataset. One can see that feature maps 142 and 169 are the most active with about 400 occurrences among the LLIs (Fig. 7a), while feature maps 142 and 159 are the most active with about 350 occurrences among the top-16 active maps for NLIs (Fig. 7b). Note that images enhanced using the DPE model (which produced the worst *SSIM* and *PSNR* scores for LOL and minimal illumination levels in Experiment 1 – Table 3c, d) produce activity maps which are similar to those of LLIs, whereas images enhanced using MBLLEN (which produced the second best *NIQE* and the best *BRISQUE* scores – Table 3a, b – as well as the best subjective scores for LOL) produce activity maps which are similar to those of NLIs. Here, we expect that enhanced images share more similar active feature maps (producing higher *top16-AFMS* scores) with NLIs, compared with their LLI counterparts. In other words, we expect an enhanced image to preserve the most important (active) features that would be present in a NLI, compared with a LLI which tends to loosen certain image features.

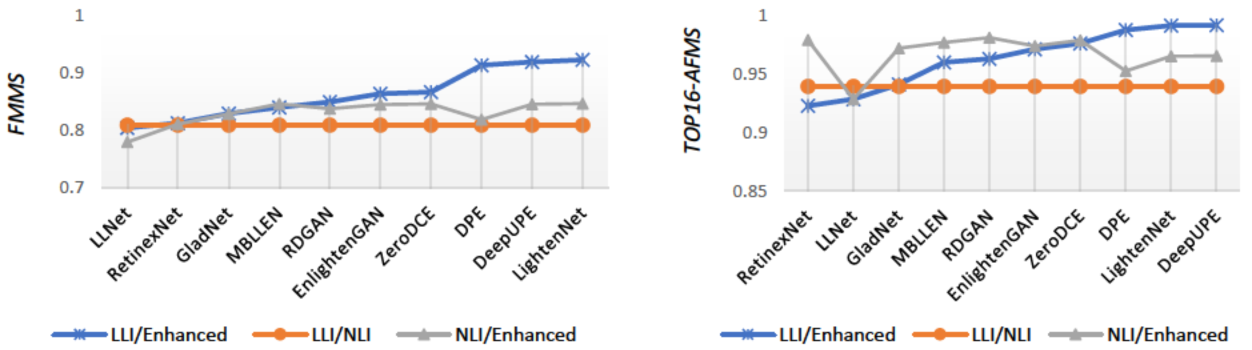
5.4.2 Experimental Results

Fig. 8, Fig. 9, and Fig. 10 show the results of *FMMS* and *top16-AFMS* metrics respectively comparing LLIs/NLIs, LLIs/enhanced, and NLIs/enhanced images from the LOL dataset. Based on the obtained results, we highlight the following observations:



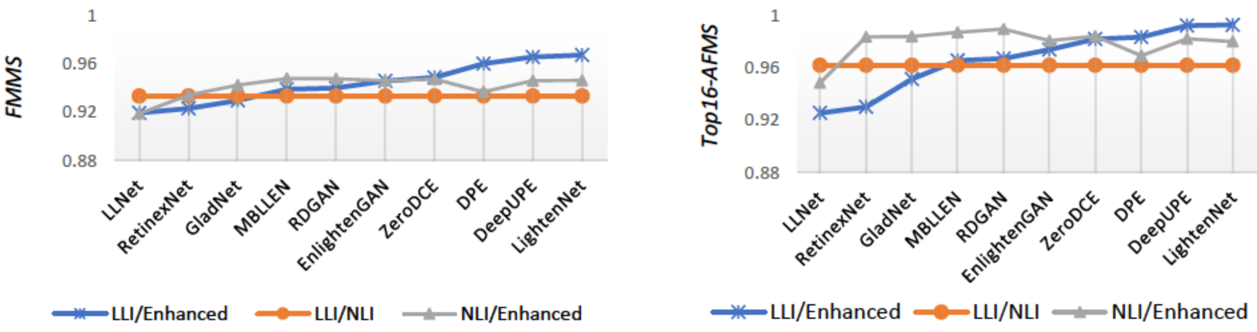
(a) ***FMMS*** **results** **(b)** ***top16-AFMS*** **results**

Figure 8: Similarity measures for layer conv4_3 of SSD512 (Wei et al. 2016) applied on the LOL dataset, ordered following LLI/Enhanced



(a) FMMS results **(b) top*16*-AFMS results**

Figure 9: Similarity measures for layer conv8_2 of SSD512 (Wei et al. 2016) applied on the LOL dataset, ordered following LLI/Enhanced



(a) FMMS results **(b) top*16*-AFMS results**

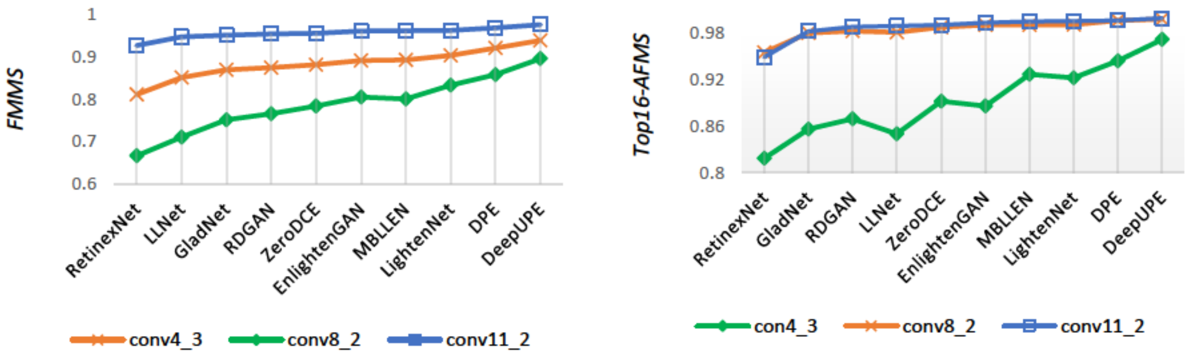
Figure 10: Similarity measures for layer conv11_2 of SSD512 (Wei et al. 2016) applied on the LOL dataset, ordered following LLI/Enhanced

- ***Feature Map Matrix Similarity*** (*FMMS*) results for conv4_3 in Fig. 8a show that *FMMS*(LLI, enhanced) > *FMMS*(NLI, enhanced) for most enhancement models. This means that the enhanced images tend to share more features at conv4_3 with their LLI counterparts, compared with the corresponding NLIs, and thus remain attached to their initial LLIs. Almost the same pattern holds for conv8_2 in Fig. 9a (with the exception of MBLLEN). However, results at conv11_2, i.e., the deepest layer of SSD, show that most models produce maps which are closer to those of NLIs versus LLIs, except for DPE, DeepUPE, and LightenNet which remain largely attached to the initial LLIs. Results for the latter three models might be due to their minimal enhancement (Experiment 1 in Section 5.2.2) which makes them more faithful to the original LLIs. Additionally, LLNet shows the lowest *FMMS*(NLI, enhanced) and *FMMS*(LLI, enhanced) levels for all layers and more prominently for conv11_2 in Fig. 10a. This can be due to the over smoothing applied by LLNet on the enhanced images (Experiment 1) making them loose their fine details, especially in the deepest layers of the detection model (i.e., conv11_2) where the high-level features incorporated in the fine details are out of interest. Finally, MBLLEN shows some of the best results with *FMMS*(LLI, enhanced) approaching *FMMS*(LLI, NLI) and simultaneously producing approximately the highest *FMMS*(NLI, enhanced) scores compared with all other models and in all three layers. Other models do not share similar measures, for example LLNet produces very close *FMMS*(LLI, enhanced) and *FMMS*(LLI, NLI) scores in both conv4_3 and conv8_2 layers, and yet it shows the lowest *FMMS*(NLI, enhanced) score. Moreover DeepUPE has very close *FMMS* (NLI, enhanced) to that of MBLLEN, yet it shows much higher *FMMS*(LLI, enhanced).

- ***Top 16 Active Feature Map Similarity*** (*top16-AFMS*) results show that *top16-AFMS*(NLI, enhanced) > *top16-AFMS*(LLI, enhanced) in all layers and for most enhancement models except for DPE, DeepUPE, and LightenNet. This means that most enhancement models tend to activate the same feature maps in the detection model (e.g., SSD in this case) compared with NLIs, and succeed to diverge from the most active feature maps of the LLIs towards those of the NLIs. We also notice from Fig. 8b, 9b and 10b, that MBLLEN's *top16-AFMS*(LLI, enhanced) is approaching *top16-AFMS*(LLI, NLI) in all layers such that its *top16-AFMS*(NLI, enhanced) produces high scores compared with all other models. This shows that MBLLEN's enhanced images share very similar active features with their NLI counterparts, compared with their LLIs. MBLLEN's *FMMS* and *top16-AFMS* scores seem to corroborate the results of Experiment 1 (cf. Section 5.2.1 and 5.2.2) where the model produces one of the best enhancement quality results compared with the other models on the LOL dataset. To sum up, most enhancement models fall short of simultaneously producing high *FMMS*(NLI, enhanced) and high *top16-AFMS*(NLI, enhanced) along with *FMMS*(LLI, enhanced) ≈ *FMMS*(LLI, NLI) and *top16-AFMS*(LLI, enhanced) ≈ *top16-AFMS*(LLI, NLI), which suggest that the enhanced images remain attached to the original LLIs and do not diverge towards the actual NLIs.

We further apply our feature analysis to 500 sample images from the ExDark dataset. Here, we only compute *FMMS* and *top16-AFMS* for LLI/enhanced images since the dataset does not include NLIs. Based on the results in Fig. 11, we highlight the following observations:

- The behavior of SSD seems to be different between LOL and ExDark datasets. For instance, results in Fig. 11b show that the *top16-AFMS*(LLI, enhanced) for LLNet at conv11_2 is higher than those of RetinexNet, GladNet and RDGAN when applied on ExDark, while LLNet produces the lowest scores when applied on LOL (Fig. 10b). Similar fluctuations occur with the other enhancement models, producing different activity responses when applied on the quasi-synthetic LOL dataset, compared with the real-world ExDark.



(a) *FMMS* similarity results

(b) *top16-AFMS* similarity results

Figure 11: Similarity measures between LLIs and enhanced images for SSD512 (Wei et al. 2016) applied on the ExDark subset, ordered following conv11_2 results

- Most enhancement models (except for RetinexNet) produce high *FMMS*(LLI, enhanced) and *top16-AFMS*(LLI, enhanced) scores and show minimal variations compared with the results produced with the LOL dataset (Fig. 10a, b). This means that the enhanced images produced by most models remain mostly faithful to their original LLIs and do not diverge enough from the LLIs to promote better features that can be more useful for the object detection task.

5.5 Occlusion Experiment

To better interpret and understand the effect of enhancement on the preservation of image features and what may be lost during enhancement, we utilize the occlusion experiment proposed in (Zeiler and Fergus 2014) where: i) a black square is used to mask particular sections of the image, ii) the black square is slided over all the possible sections of the image, allowing to iii) perform object detection for every slided mask, producing a heatmap highlighting the object detection confidence scores (in case of a detection, and zero otherwise). The occlusion experiment is performed on images containing a single object each, so that the detection model focuses solely on them. Its rationale is two-fold: i) if an image contains many regions which may cause a misdetection if occluded, then the image is assumed to hold weak features allowing to easily misdetect its object, and ii) if an image contains no specific region that might cause a misdetection if occluded – given all the masks slided over the entire image, then the image is assumed to hold strong features that allow to correctly detect its object.

In the following, we present both quantitative and qualitative evaluations of the occlusion experiment applied on enhanced images produced by the 10 DL-based LLI enhancement models used in the previous experiments.

5.5.1 Quantitative Evaluation

We perform the occlusion experiment on 100 sample images from the ExDark dataset, considering only images containing single objects. All the images are resized equally, and the same size of the moving black box is used with all of them. We consider the original LLIs and their enhanced counterparts produced by the 10 enhancement models considered in our study, and process each of them through the 4 object detection models used in Experiment 2 (cf. Section 5.3). A confidence score of 0.15 is used to limit the number of detections produced by all models. We make use of the Occlusion based Average Misdetection Regions (*OAMR*) metric (cf. Section 5.1.2.3), which highlights the ability of an enhancement model to include stronger features in the enhanced images by producing lower scores (i.e., lower number of regions contributing to misdetections) compared with the original LLIs. Results are reported in Table 5 and highlight the following observations:

- DeepUPE, MBLLEN, and DPE produce some of the best (low) *OAMR* results, which is consistent with their high *mAP* scores obtained in Experiment 2 (cf. Section 5.3.2), despite DeepUPE and DPE's minimal enhancement quality in Experiment 1 (cf. Section 5.2.1 and 5.2.2). This corroborates with our observations from Experiment 2, where a good enhancement quality does not necessarily translate into better feature preservation and improved object detection quality.

- None of the remaining enhancement models (with the exception of DeepUPE, MBLLEN, and DPE) produce an *OAMR* score lower than that of the original LLIs, indicating that the models are adding more regions which contribute to misdetections, and are thereby loosing significant object features upon image enhancement.

- MBLLEN produces one of the best (lowest) average *OAMR* scores, reflecting good feature preservation in the enhanced images. This seems consistent with its top enhancement quality achieved in Experiment 1. On the other end of the spectrum, RetinexNet shows the worst (highest) average *OAMR* scores, which is consistent with its bad enhancement quality achieved in Experiment 1. This seems to indicate that visual enhancement quality and feature preservation performance might not be completely unrelated, and that good visual enhancement balanced with proper feature handling could strengthen the object features upon image enhancement.

Table 5: OAMR ↓ results for 100 images from ExDark dataset, ranked from best (#1) to worst (#10) following each detection model (red color refers to the best score and green to the second best for every detection model).

(a) Results ranked following YOLOv3 (Redmon and Farhadi 2018)

Approach	Rank	YOLOv3	RetinaNet	SSD	Mask RCNN
Original LLIs	---	5.70	6.31	5.81	8.17
DeepUPE	1	**5.82**	**6.53**	**5 38**	**7.51**
MBLLEN	2	**6.42**	7.31	**5.76**	9.93
DPE	3	6.61	**5.58**	5.78	11.16
LightenNet	4	6.75	8.34	6 37	9.48
EnlightenGAN	5	7.40	9.56	6.68	9.66
GladNet	6	7.49	9.51	8.05	**9.40**
RDGAN	7	7.70	9.06	7.73	10.74
ZeroDCE	8	7.99	9.69	8 28	11.02
LLNet	9	8.83	11.23	7.46	11.89
RetinexNet	10	18.57	22.35	18.25	26.04

(b) Results ranked following RetinaNet (Lin et al. 2017)

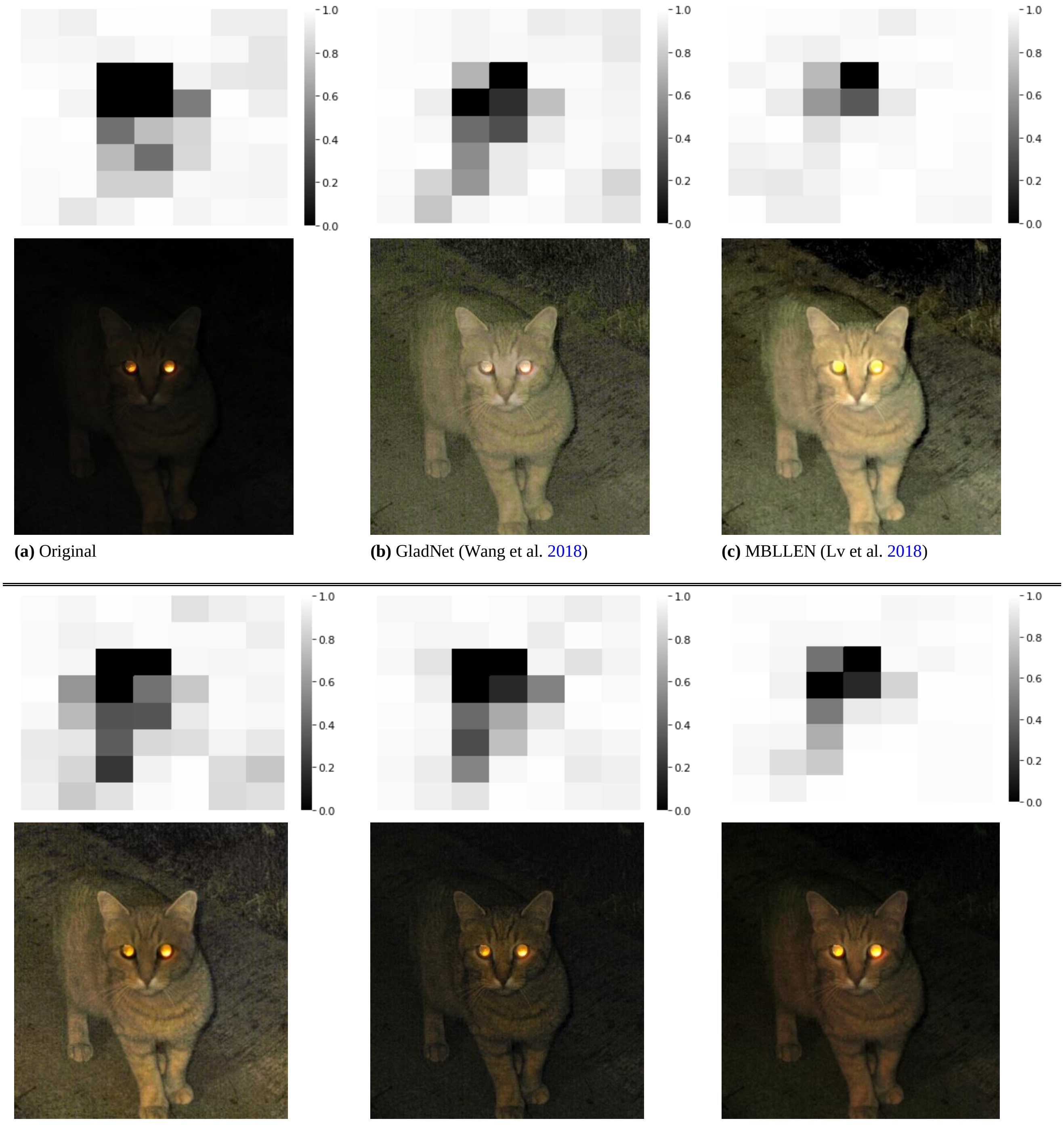
Approach	Rank	YOLOv3	RetinaNet	SSD	Mask RCNN
DPE	1	6.61	**5 58**	5.78	11.16
Original LLIs	---	5.70	6 31	5.81	8.17
DeepUPE	2	**5.82**	**6 53**	**5.38**	**7.51**
MBLLEN	3	**6.42**	7 31	**5.76**	9.93
LightenNet	4	6.75	8 34	6.37	9.48
RDGAN	5	7.70	9.06	7.73	10.74
GladNet	6	7.49	9 51	8.05	**9.40**
EnlightenGAN	7	7.40	9 56	6.68	9.66
ZeroDCE	8	7.99	9.69	8.28	11.02
LLNet	9	8.83	11.23	7.46	11.89
RetinexNet	10	18 57	22.35	18 25	26.04

(c) Results ranked following SSD (Wei et al. 2016)

Approach	Rank	YOLOv3	RetinaNet	SSD	Mask RCNN
DeepUPE	1	**5.82**	**6.53**	**5 38**	**7.51**
MBLLEN	2	**6.42**	7.31	**5.76**	9.93
DPE	3	6.61	**5.58**	5.78	11.16
Original LLIs	---	5.70	6.31	5.81	8.17
LightenNet	4	6.75	8.34	6 37	9.48
EnlightenGAN	5	7.40	9.56	6.68	9.66
LLNet	6	8.83	11.23	7.46	11.89
RDGAN	7	7.70	9.06	7.73	10.74
GladNet	8	7.49	9.51	8.05	**9.40**
ZeroDCE	9	7.99	9.69	8 28	11.02
RetinexNet	10	18.57	22.35	18.25	26.04

(d) Results ranked following Mask RCNN (He et al. 2017)

Approach	Rank	YOLOv3	RetinaNet	SSD	Mask RCNN
DeepUPE	1	**5.82**	**6 53**	**5.38**	**7.51**
Original LLIs	---	5.70	6 31	5.81	8.17
GladNet	2	7.49	9 51	8.05	**9.40**
LightenNet	3	6.75	8 34	6.37	9.48
EnlightenGAN	4	7.40	9 56	6.68	9.66
MBLLEN	5	**6.42**	7 31	**5.76**	9.93
RDGAN	6	7.70	9.06	7.73	10.74
ZeroDCE	7	7.99	9.69	8.28	11.02
DPE	8	6.61	**5 58**	5.78	11.16
LLNet	9	8.83	11.23	7.46	11.89
RetinexNet	10	18 57	22.35	18 25	26.04



(a) Original

(b) GladNet (Wang et al. 2018)

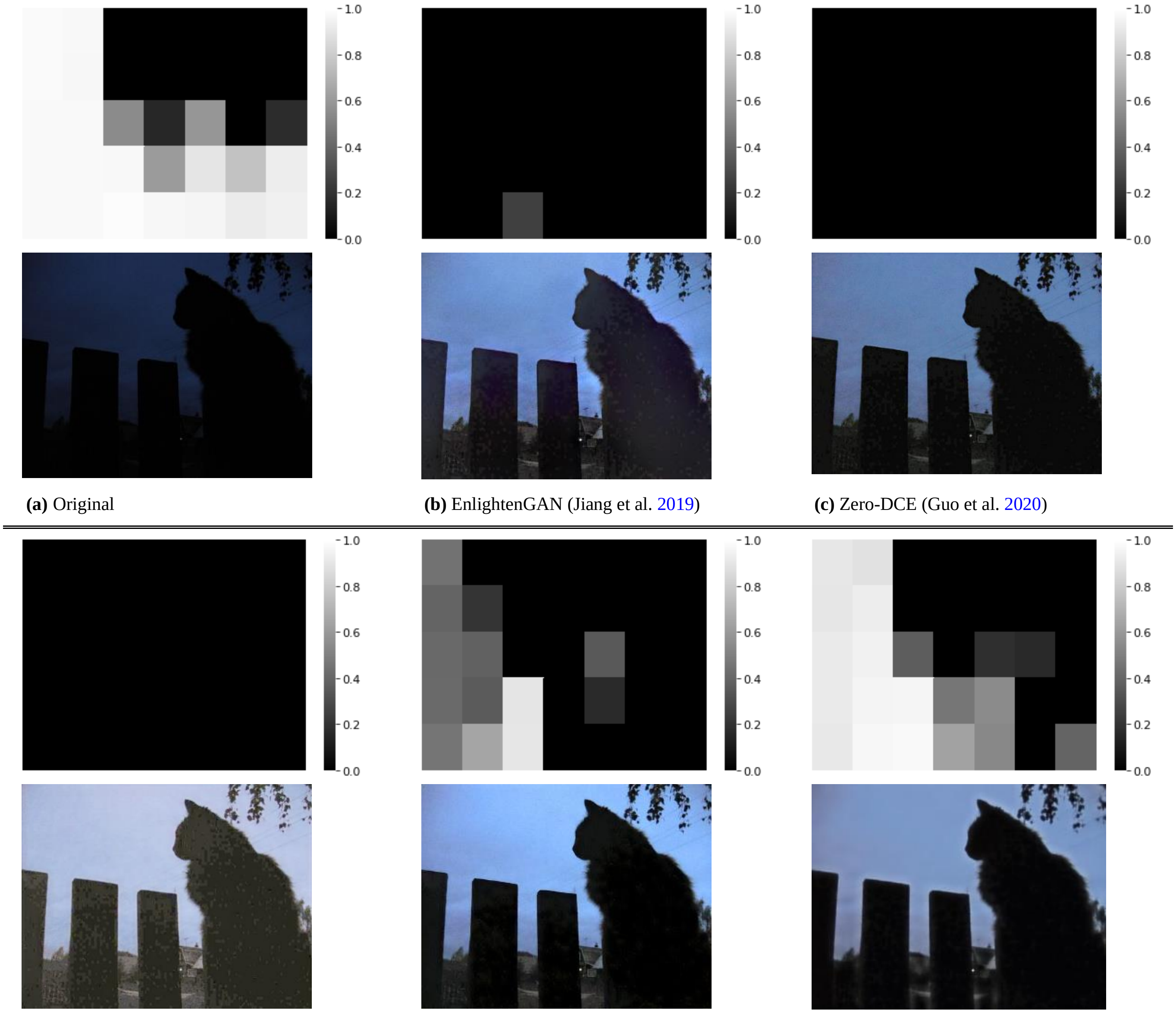
(c) MBLLEN (Lv et al. 2018)

(d) EnlightenGAN (Jiang et al. 2019)

(e) DPE (Chen et al. 2018)

(f) DeepUPE (Wang et al. 2019)

Figure 12: Occlusion experiment on a sample LLI from ExDark and its enhanced counterparts: case 1



(a) Original

(b) EnlightenGAN (Jiang et al. 2019)

(c) Zero-DCE (Guo et al. 2020)

(d) GladNet (Wang et al. 2018)

(e) MBLLEN (Lv et al. 2018)

(f) LLNet (Lore et al. 2017)

Figure 13: Occlusion experiment on a sample LLI from ExDark and its enhanced counterparts: case 2

5.5.2 Qualitative Evaluation

In addition to the quantitative evaluation, and in order to shed further light on the results of the occlusion experiment, we qualitatively evaluate and discuss two typical cases using the YOLOv3 detection model: i) successful detection in both the LLI and the enhanced image, and ii) successful detection in the LLI and misdetection in the enhanced image. The cases where we have a misdetection in the LLI are not beneficial for this experiment since they do not reflect any information about the initial LLI features**.**

Case 1 – Successful detection in the LLI and the enhanced image*:*

Fig. 12 shows the occlusion heatmaps obtained on a sample LLI and its enhanced counterparts.

Results in Fig. 12 highlight the following observations:

- The heatmap of the original LLI shows 4 zero-confidence score (dark) regions concentrated around the face of the *cat* object, which seem to contribute to its misdetection. In contrast, the heatmaps of the enhanced images show a lesser number of zero-score (dark) regions contributing to the misdetection of the *cat* object. This means that the enhancement models seem to integrate better features into the enhanced images, allowing to improve their detection confidence scores.

- MBLLEN shows the best features with only one region resulting in a misdetection. In other words, regions which were initially responsible for misdetecting the *cat* object in the original image are no longer causing a misdetection after MBLLEN' s enhancement.

- Although EnlightenGAN shows one of the best visual quality results among all enhancement models in Experiment 1 (Section 5.2.1), yet it produces 3 zero-confidence (dark) regions resulting in misdetections (Fig. 12d). In contrast, DeepUPE which shows a minimal enhancement quality in Experiment 1 produces only 2 regions resulting in misdetections (Fig. 12f). Similarly, DPE shows low illumination while producing only 3 misdetection regions (Fig. 12e), identically to EnlightenGAN which seemingly shows better illumination and enhancement quality (Fig. 12d). The latter observations show that a good visual enhancement quality does not necessarily translate into better object detection features. This corroborates the results from Experiments 1 and 2, where EnlightenGAN on the one hand, and DeepUPE and DPE on the other hand, respectively show high/low visual enhancement quality versus low/high object detection performance.

Case 2 – Successful detection in the LLI and misdetection in the enhanced image.

Fig. 13 shows the occlusion heatmaps obtained on another sample LLI from the ExDark dataset, and its enhanced counterparts. Results highlight the following observations:

- Although object *cat* is successfully detected in the original LLI, yet it contains 11 regions contributing to a misdetection. This shows that the LLI initially holds weak features that poorly contribute to the object detection task.

- The enhanced images produced by EnlightenGAN, ZeroDCE, and GladNet, result in a complete misdetection of object *cat*, showing that the enhancement models have loosened the few features that were contributing to the object detection task in the original LLI.

- The enhanced image produced by MBLLEN allows a successful object detection. However, it includes 22 regions contributing to a misdetection, which is double the number of misdetection regions present in the original LLI (=11). This shows that MBLLEN loosened some of the features while preserving others that were most important to the detection task.

- LLNet allows a successful object detection and shows better feature preservation compared with its counterparts. This can be due to the minimal noise incorporated by LLNet in comparison with its counterparts (Experiment 1, Table 2c). This shows that the amplification or integration of noise into the enhanced image seems to loosen the features that are useful for object detection.

5.5.3 Discussion

To sum up, we review and discuss the results of our feature analysis experiment.

First, the enhanced images produced by most enhancement models tend to activate the same detection model feature maps compared with LLIs. In other words, enhanced images produced by most models tend to share more features with their LLI counterparts, compared with the corresponding NLIs. They fail to diverge from the features of the LLIs towards those of the NLIs and remain attached to their initial LLIs.

Second, results of the occlusion experiment show that successful object detection in enhanced images seems to be related to the number of (mis)detection regions in the occlusion heatmap, which in turn highlights the number of (loosened and) preserved features in the resulting enhanced image, compared with its original LLI. *OAMR* results show that most of the enhancement models tend to produce enhanced images with more regions contributing to misdetections and thus showing weakly embodied semantic features.

Third, an important aspect to be considered here is the level of noise added in the enhanced images. Referring to the results of Experiments 1 and 2, we realize that MBLLEN produces some of the lowest noise levels (cf. Section 5.2.1) and some of the best *mAP* results (cf. Section 5.3.2) compared with the other enhancement models, and accordingly produces good *OAMR* scores in this experiment. On the other side of the spectrum, ZeroDCE produces the highest noise level amongst the enhancement models (cf. Section 5.2.1) with uncompetitive *mAP* results (cf. Section 5.3.2), and accordingly produces some of the worst *OAMR* scores. This suggests that preserving the image features that are useful for object detection, coupled with a reduction in noise levels, can help improve detection performance.

5.6 Recap and Directions

5.6.1 Recap of Empirical Results

To sum up, we recap our observations and findings as follows.

From *Experiment 1 - Visual and Perceptual Quality*:

- Most of the LLI enhancement models evaluated in our study still fall short of producing properly illuminated enhanced images with good visual quality. They fail to strike a good balance between image illumination level, noise level, exposure level, and color deviation. Some models successfully improve one aspect while ignoring others and tend to incorporate significant noise into the enhanced images, thus distorting their quality.

- Results for the IQA (Image Quality Assessment) objective metrics used in this study do not closely match human evaluation ratings. The metrics also fail to produce consistent rankings among themselves.

From *Experiment 2 – Detection and Classification quality*:

- Improving LLI visual quality does not necessarily boost object detection and classification quality. Many models evaluated in our study tend to produce enhanced images which deteriorate object detection performance rather than improving it. This can be attributed to the fact that most existing enhancement models were developed as standalone solutions, and were not designed to be embedded as a pre-processing step for high-level computer vision tasks like object detection.

- The level of noise added in the enhanced images seems to affect detection quality. By comparing with the results of Experiment 1, we realize that many models producing low noise levels tend to produce some of the best detection results in Experiment 2. This suggests that a proper balancing between noise level and visual features could improve the detection task.

From *Experiment 3 – Feature Analysis*:

- Enhanced images produced by most models tend to share more features with their LLI counterparts, compared with the corresponding NLIs. They fail to diverge from the features of the LLIs towards those of the NLIs, and remain attached to their initial LLIs. This contributes to a drop in detection performance, which is usually further exacerbated by the added artifacts and noise resulting from the enhancement process.

- Most enhancement models tend to produce enhanced images with more regions contributing to misdetections and thus show weakly embodied semantic features. The enhancement task should consider enriching enhanced images with strong features that make detection models more robust and confident in their predictions.

5.6.2 Potential Directions

Based on our literature review and empirical observations, we highlight a few potential directions:

- There is a need to design more accurate IQA objective metrics that simultaneously quantify illumination and noise levels and behave in accordance with the human visual perception of image quality.

- There is a need to produce LLI enhancement models that can be used as a pre-processing step for other high-level computer vision task such as object detection and classification. In this context, LLI enhancement should be formulated while considering the preservation of the semantic features necessary for the high-level task at hand.

- The preservation of semantic features should consider decoupling the LLI from the enhanced image such that it does not diverge beyond the actual similarity between the NLI and LLI, while maintaining at the same time high similarity with the NLI. While most supervised learning models tend to use a perceptual loss between the enhanced image and the NLI, they should also consider limiting the loss between the LLI and the enhanced image to that between the LLI and NLI.

- The noise factor and de-noising techniques need to be given special attention when designing new LLI enhancement models, especially that noise seems to consistently affect visual enhancement quality as well as object detection quality.

- One of the best enhancement models in our empirical evaluation: EnlightenGAN (Jiang et al. 2019), follows the unsupervised learning paradigm, and thus highlights the potential of unsupervised LLI enhancement techniques. This would eliminate the need for paired training images, and would allow the use of real-world datasets which are increasingly available, rather than relying on synthetic datasets which are scarce and fail to mimic real LLIs.

- Another promising direction is presented by Zero-DCE (Guo et al. 2020), which entirely reformulates the LLI enhancement task to learn a mapping between LLIs and estimated light curves, thus releasing the need of paired and unpaired training data. The model achieved good object detection quality compared with many other DL enhancement models and was qualitatively favored by human testers as it sufficiently boosted image illumination albeit adding more noise. Such an approach could be revolutionary if properly extended or fine-tuned to maintain a good balance between illumination level, noise level, and semantic feature preservation.

Chapter 6
LLI Enhancer Model

While the reviewed and evaluated enhancement models show success when used as standalone for only the purpose of enhancement, yet they all share a common limitation: they are not tailored for high-level computer vision tasks like object classification while they are expected to produce enhanced images which boost the tasks performance. In our in-depth evaluation study for the performance of state of art classification and detection models on LLI datasets preprocessed by recent enhancement models, the results in section 5.3.2 show that the involved enhancement adds slight improvement or even does not improve the detection and classification performance. The study shows that a good enhancement quality is not necessarily correlated with an improved detection and classification quality. Therefore, the enhancement task should essentially consider into account high-level computer vision tasks like object classification to make them more robust against low light and normal light conditions. In what follows, we elaborate on the detailed steps involved in the design of our enhancement model which is feasible to integration into classification models and evaluate its enhancement performance quantitatively and qualitatively.

6.1 Methodology

We design an enhancement model which performs an image to frequency filter learning instead of an image-to-image learning. The enhancement is based on homomorphic filtering where a special filter of only two parameters is devised to filter the image frequency components in the Fourier transform domain. The two parameters are estimated using any of the feature extractors usually used in classification models. We call our model: Low Light Homomorphic Filtering Network (LLHFNet). We next detail the key elements of the enhancement model namely: homomorphic filtering, enhancement filter design, network architecture and loss function.

6.1.1 Homomorphic Filtering (HF)

HF based enhancement methods use the Retinex model representation of the image to convert the illumination and reflectance components which combine multiplicatively to an additive form in the logarithmic domain. The additive components are separated linearly in the Fourier transform frequency domain in which high frequency components are associated with reflectance while low frequency components correspond to illumination. A high pass filter is used to suppress low frequencies and amplify high frequencies.

The steps for HF are as follows:

1. The logarithm of both sides of the Retinex model is taken to convert from multiplicative form to additive form as follows:

   Retinex Model:

   $$M(x, y) = I(x, y) \times R(x, y), \tag{9}$$

   where $M(x, y)$ is the original image, $I(x, y)$ is the illumination component, and $R(x, y)$ is the reflectance component.

   Logarithm:

   $$\ln M(x, y) = \ln I(x, y) + \ln R(x, y) \tag{10}$$

2. The Fourier transform is applied to convert the image from the spatial domain to the frequency domain:

   $$F[\ln M(x, y)] = F[\ln I(x, y) + \ln R(x, y)] \tag{11}$$

   And more concisely equation (11) can be written as:

   $$M(u, v) = I(u, v) + R(u, v), \tag{12}$$

where $M(u,v)$, $I(u,v)$ and $R(u,v)$ are the Fourier transforms of $M(x,y)$, $I(x,y)$ and $R(x,y)$. $I(u,v)$ is mainly concentrated in the low frequency range while $R(u,v)$ is concentrated in the high frequency range.

3. For enhancement, an appropriate high pass filter with transfer function $H(u,v)$ is applied:

$$S(u,v) = H(u,v) \times M(u,v) = H(u,v)I(u,v) + H(u,v)R(u,v) \tag{13}$$

4. The inverse Fourier transform is employed to transform the image from the frequency domain to the spatial domain. Let $s(x,y)$ be the inverse Fourier transform of $S(u,v)$, then the inverse Fourier transform of equation (13) is:

$$\begin{aligned} s(x,y) &= F^{-1}\big(H(u,v)I(u,v)\big) + F^{-1}\big(H(u,v)R(u,v)\big) \\ &= h_I(x,y) + \ h_R(x,y) \end{aligned} \tag{14}$$

5. Finally, the exponential or logarithmic inverse is applied on equation (14) to obtain the final enhanced image denoted by $E(x,y)$ as below:

$$E(x,y) = \exp[s(x,y) \ \ = \exp[h_I(x,y)]\exp[h_R(x,y)] \tag{15}$$

The HF algorithm flow is shown in Fig. 14. In this figure, Log is the logarithmic transform, FFT and IFFT are the fast Fourier transform and its inverse respectively, $H(u,v)$ is the frequency filtering function and Exp is the exponential operation.

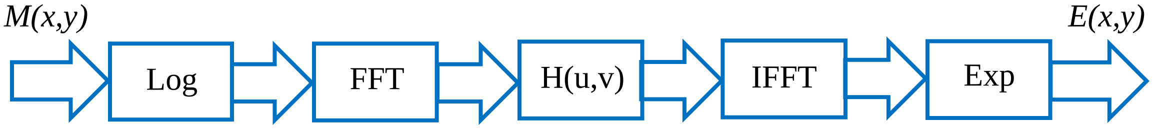


Figure 14: HF algorithm flow (modified based on (Wang et al. 2020))

6.1.2 Enhancement Filter Design

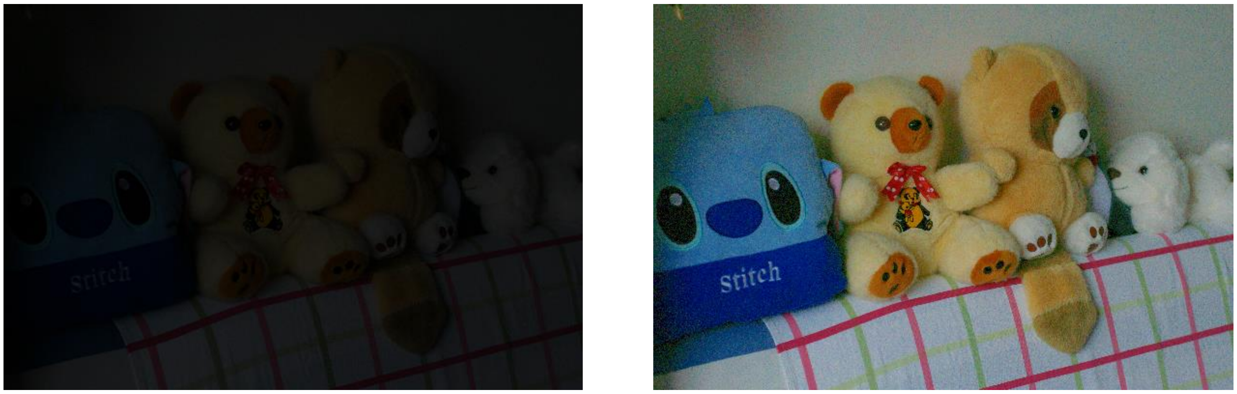
A core part of the HF algorithm is $H(u,v)$, the frequency filtering transform. We want our filter to be simple and effective in performing the enhancement. The Fourier transform of the original image i.e., $M(u,v)$ at (0,0) represents its DC-term which corresponds to its average brightness in the spatial domain (Gonzalez and Woods, 2018). It is noticeable that for LLIs, $M(0{,}0)$ is a large negative value which reflects the low brightness of these images and so the brightness can be enhanced by increasing $M(0{,}0)$. This leads to the first parameter of the enhancement filter denoted by $\gamma_L \in [0{,}1]$ placed at $H(0{,}0)$. The smaller the value of γ_L, the higher is the brightness level. The remaining frequency components of $M(u,v)$ correspond to the image variations and are filtered using the second parameter of the enhancement filter denoted by $\gamma_H \in [0{,}1]$. The larger the value of γ_H, the sharper are the contents of the image. Finally, the enhancement filter is as follows:

$$H(u,v) = \begin{cases} \gamma_L & (0{,}0) \\ \gamma_H & otherwise \end{cases} \tag{16}$$

We run the HF algorithm on the Value channel of the HSV (Hue-Saturation-Value) color domain instead of using the RGB domain. We follow this step for four reasons: i) it is more efficient to apply FFT and its inverse on only one channel instead of three, ii) the Value channel corresponds to the lightness of the image which we aim to improve and this channel will be only affected while the Hue and Saturation will remain preserved, iii) the obtained enhancement quality on the Value channel is better than that on the RGB channels as noticed experimentally and iv) more simplicity is maintained for the enhancement filter in HSV domain compared to the RGB domain which may need two parameters per each of its channels to achieve a good enhancement quality.

We show in Figures 15 and 16 the correspondence of the designed enhancement filter for different exposure levels. In Fig.15, the LLI has a very low exposure level, so by using an enhancement filter of parameters $(\gamma_L = 0.25,\ \gamma_H = 0.45)$, the HF algorithm produces a visually pleasing image with minimal artifacts. While in Fig.16 the image has a medium exposure level so larger values of γ_L and γ_H are used to perform a minimal enhancement and avoid overexposure. Therefore, a learner is needed to estimate the values of γ_L and γ_H

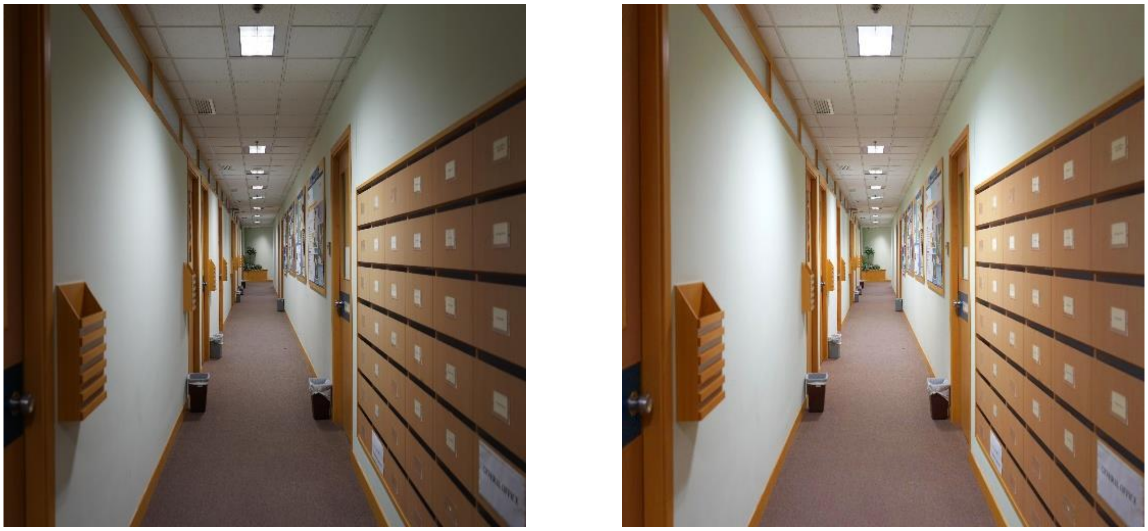
which produce the best enhancement quality and handle the different exposure levels of input images.



(a) LLI

(b) Enhanced Image ($\gamma_L = 0.25, \gamma_H = 0.45$)

Figure 15: LLI from LOL (Wei et al. 2018) dataset and its enhanced counterpart using HF algorithm



(a) LLI

(b) Enhanced Image ($\gamma_L = 0.60, \gamma_H = 0.70$)

Figure 16: LLI from SICE (Cai et al. 2018) dataset and its enhanced counterpart using HF algorithm

6.1.3 Network Architecture

Since the enhancement filter is formed of two parameters γ_L and γ_H which are to be estimated via a deep learner, then we can benefit from existing feature extractors to perform the task. Our HF based enhancement algorithm eases the restriction of using custom architectures like in the case of image-to-image learning and rather allows using any feature extractor to perform an image to only 2 frequency filter parameters mapping.

Our enhancement model architecture is formed of two major parts:

i) A feature extractor which is responsible of extracting high-level features from input images. The extractor can be any of the commonly used feature extractors for object classification task like VGG16 (Simonyan and Zisserman 2015), ResNet50 (He et al. 2016), MobileNetv2 (Sandler et al. 2018), SqueezeNet (Iandola et al. 2016), among others. The first layer of the extractor is modified to accept an input image of one channel corresponding to the Value channel in HSV domain.

ii) An enhancement head which is formed of four convolutional layers followed by ReLU activation and max pooling layers used to downsize the feature maps obtained from the extractor. The last convolutional layer is of size 1x2x1 and followed by Sigmoid activation function to limit the 2 output values representing γ_L and γ_H to the range [0,1]. The detailed architecture of LLHFNet is shown in Fig.17.

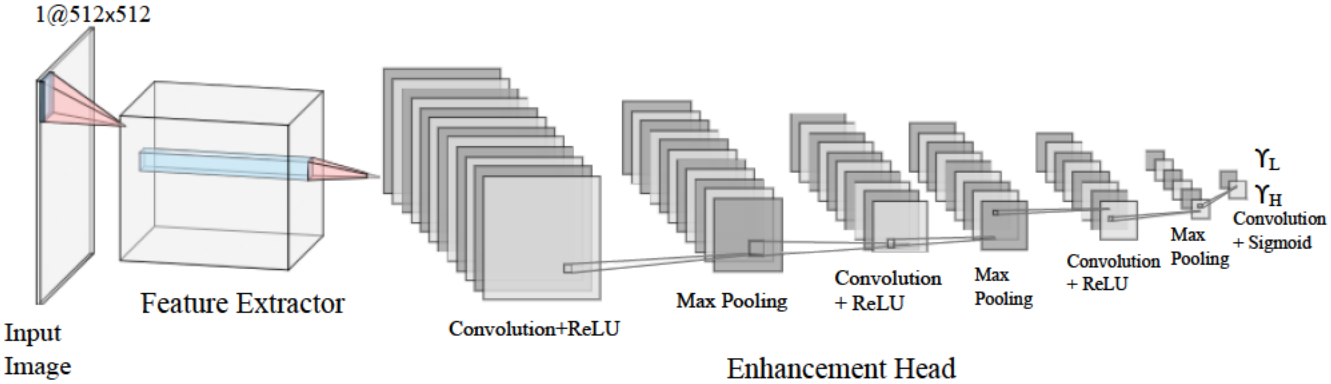


Figure 17: Enhancement model network architecture

6.1.4 Loss Function

The loss function is a major element of the LLI enhancement model that drives the entire learning process. We follow a supervised training setting in which reference-based loss functions are needed. The model uses Multi-Scale Structural Similarity index (MS-SSIM) (Wang et al. 2003) as its loss function. It is noticed in section 5.2.3 that image quality assessment (IQA) metrics do not always show proper correlation with our visual perception and so MS-SSIM (Wang et al. 2003) may produce wrong measures for the enhancement quality during training. Since our enhancement model optimizes only two parameters using MS-SSIM loss, then the glitches in IQA can be easily monitored. It is noticed that the model has sometimes a tendency to predict values of γ_L greater than that of γ_H, which in turn produce an enhanced image that is over smoothed and has color deviations making it not perceptually pleasing, yet at the same time this tendency is encouraged by lower MS-SSIM loss values. To minimize the impact of this miscorrelation between the quantitative measure and the qualitative perception, a regularize term $l = \gamma_L - \gamma_H$ is added to the loss function. This term encourages the model to maintain γ_H values greater than γ_L allowing to produce better enhanced images. Finally, our loss function equation is as follows:

$$\begin{aligned} loss\ (enhanced, NLI) &= 1 - MSSSIM\ (enhanced, NLI) + \alpha\ l \\ &= 1 - MSSSIM\ (enhanced, NLI) + \alpha\ (\gamma_L - \gamma_H) \end{aligned} \tag{17}$$

The α term is used to weight the impact of the regularize term on the overall loss. It is found empirically that values of α in the range $[0.05 - 0.1]$ produce good results in overall.

We show in Fig. 18 the overall framework of our enhancement model.

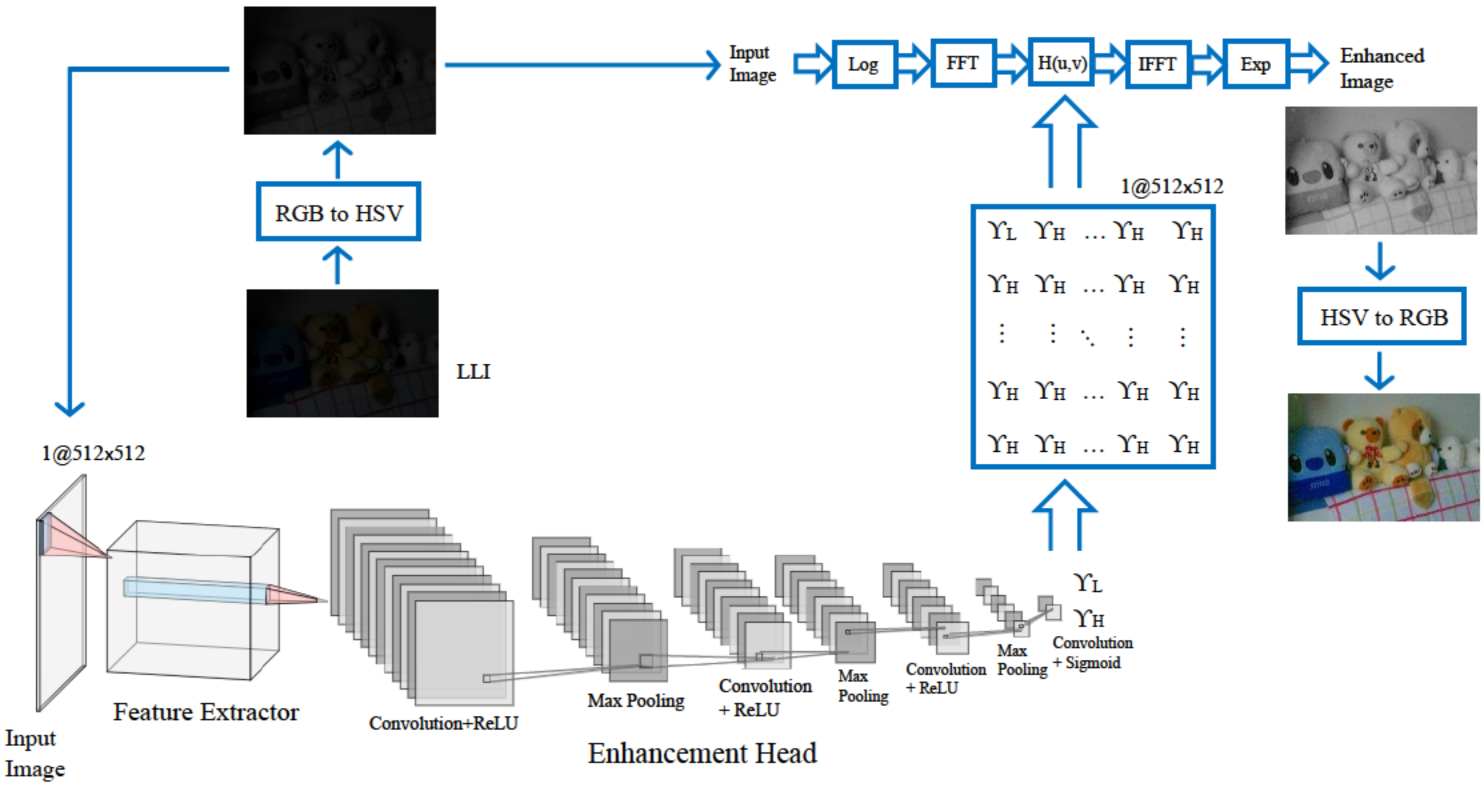


Figure 18: LLHFNet framework

6.2 Experimental Results

6.2.1 Implementation Details

Our enhancement model uses a supervised training setting where paired LLIs/NLIs are required. We therefore employ the Part1 of SICE (Cai et al. 2018) dataset to train our proposed model. The dataset is made of 360 multi-exposure sequences allowing the model to be trained on variety of exposure conditions ranging from underexposed to overexposed images. We excluded the extremely underexposed and overexposed images as our model will struggle handling them. Yet we only kept 30 extremely underexposed images but changed their ground truth to images corresponding to a low exposure level rather than a normal exposure level. We aim by this trick to teach our model to avoid inducing a tough enhancement on extremely underexposed images to make them look like normal and instead perform a slight enhancement which will minimize exposed artifacts and distortions. Our dataset is made up of 1700 images for training and 450 for validation. Although the dataset is small, yet our enhancement model does not require a lot of training data as it relies on powerful pre-trained feature extractors for its backbone. All training images are resized to 512x512. We utilize five pre-trained feature extractors namely: VGG16 (Simonyan and Zisserman 2015), ResNet50 (He et al. 2016), MobileNetv2 (Sandler et al. 2018), SqueezeNet (Iandola et al. 2016) and DenseNet (Huang et al. 2017), and train all the architectures on the formed dataset.

LLHFNet is implemented using PyTorch on a P100 Tesla Nvidia GPU. A batch size of 8 is used. Adam optimizer with default parameters and a reduce on plateau based decaying learning rate with initial value of 1e-4 are used for network optimization.

For our empirical evaluation we use 767 paired LLIs/NLIs from Part2 of SICE (Cai et al. 2018) dataset collected similar to the approach followed by Guo et al. (2020). The images are resized to size 1200x900x3. We compare our enhancement model to two state of art traditional approaches: SRIE (Fu et al. 2016) and LIME (Li et al. 2015), and three most recent DL based approaches: ZeroDCE (Guo et al. 2020), EnlightenGAN (Jiang et al. 2019) and DeepUPE (Wang et al. 2019).

6.2.2 Quantitative Evaluations

To perform a quantitative image quality assessment, we employ the commonly used full reference metrics: Peak Signal to Noise Ratio (PSNR), Structural Similarity index (SSIM) (Wang et al. 2004) and Mean Absolute Error (MAE) on the Part2 SICE subset (Cai et al. 2018). In table 6 our enhancement model using MobileNetv2 (Sandler et al. 2018) as its feature extractor ranks second best following SSIM (Wang et al. 2004) metric and best following PSNR and MAE. The results show that our model is competitive compared to the recent state of art enhancement solutions despite following an image to only 2 frequency filter parameters learning.

Table 6: Quantitative comparison for the enhancement quality of different models. The best result is in red and second best is in green. LLHFNet uses MobileNetv2 (Sandler et al. 2018) as its feature extractor.

Model	SSIM ↑	PSNR ↑	MAE ↓
SRIE	0.54	14.41	127.08
LIME	0.57	16.17	108.12
DeepUPE	0.49	13.52	142.01
EnlightenGAN	**0.59**	16.21	102.78
ZeroDCE	**0.59**	**16.57**	**98.78**
LLHFNet	**0.58**	**16.89**	**94.99**

We show in table 7 the quantitative assessment results for our enhancement model while using different feature extractors. It is noticeable that all the architectures achieve good and competitive results when compared to state of art enhancement solutions in table 6. VGG16 (Simonyan and Zisserman 2015) which has a very dense architecture provides some of the best measures and similarly MobileNetv2 (Sandler et al. 2018) is ranked among the best architectures while providing a good compromise between model size and efficiency. SqueezeNet (Iandola et al. 2016) is ranked as the worst model compared to the given feature extractors possibly due to its lightweight architecture, yet it still shows competitive results when compared to the enhancement solutions in table 6. Thus, our enhancement approach can be used with different feature extractors making it independent of a custom architecture and feasible to integration with object classification models.

Table 7: Quantitative comparison of different feature extractors used for the enhancement model. The best result is in red and second best is in green.

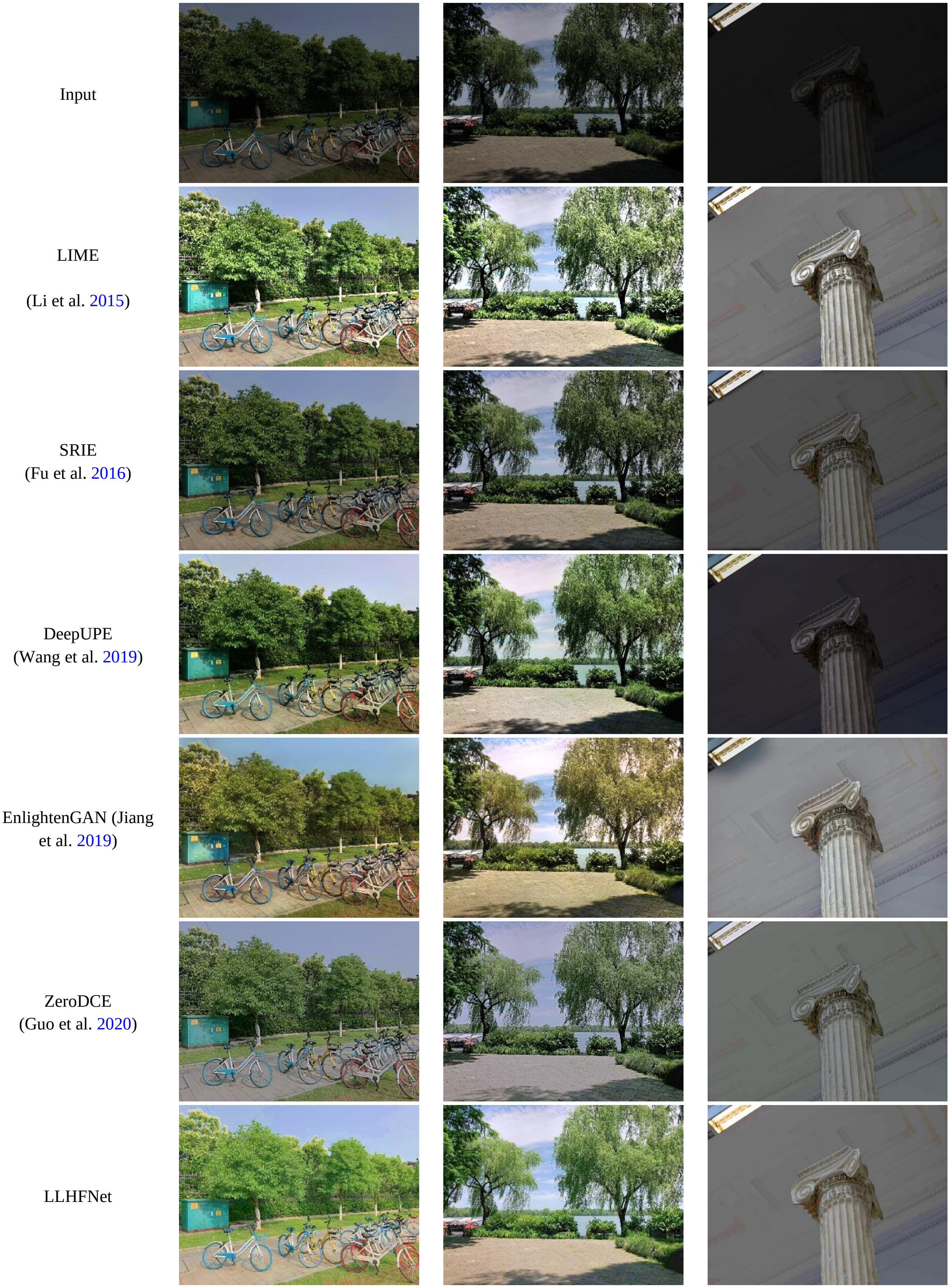
Feature Extractor	SSIM ↑	PSNR ↑	MAE ↓
MobileNetv2	**0.583**	**16.896**	**94.992**
VGG16	**0.582**	**16.897**	**94.064**
ResNet50	0.577	16.686	96.152
DenseNet	0.576	16.716	97.253
SqueezeNet	0.575	16.593	99.129

Table 8 shows the runtimes of different models averaged on 100 images of size 1200×900×3 using Tesla P100 GPU. As can be seen, LLHFNet runtime is associated with the feature extractor architecture it uses. For instance, it ranks as third best model while using SqueezeNet (Iandola et al. 2016) which has a lightweight architecture, and it drops in rank with denser architectures like VGG16 (Simonyan and Zisserman 2015). Our model is also much more efficient in all its architectures when compared to traditional enhancement approaches like SRIE (Fu et al. 2016) and LIME (Li et al. 2015). While recent DL based enhancement models like ZeroDCE (Guo et al. 2020) and EnlightenGAN (Jiang et al. 2019) show faster runtimes than LLHFNet, yet our approach is targeted for the object classification task rather than only the enhancement task, and so it relies on feature extractors used by classification models thus affecting its inference performance.

Table 8: Runtime comparisons of different enhancement models. LLHFNet (M) refers to our enhancement model with feature extractor M. Models are ranked from best to worst.

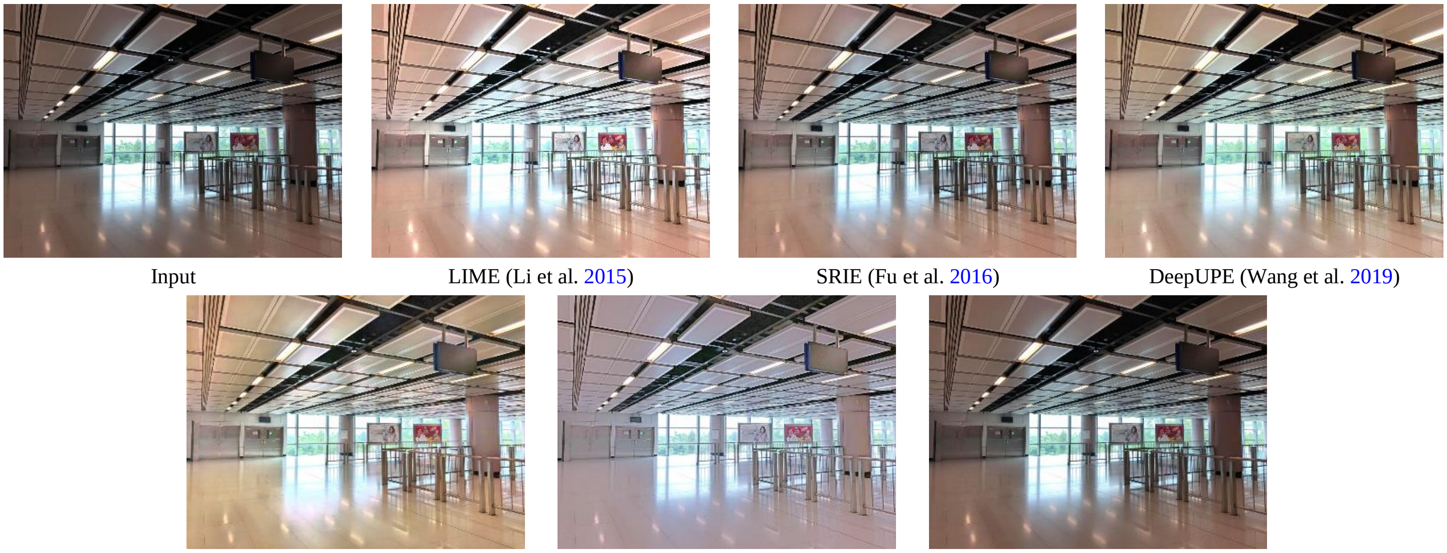
Model	**Runtime (in seconds)**	**Platform**
ZeroDCE	**0.0014**	PyTorch (GPU)
EnlightenGAN	0.0055	PyTorch (GPU)
LLHFNet (SqueezeNet)	0.0117	PyTorch (GPU)
DeepUPE	0.0183	TensorFlow (GPU)
LLHFNet (MobileNetv2)	0.0213	PyTorch (GPU)
LLHFNet (ResNet50)	0.0507	PyTorch (GPU)
LLHFNet (DenseNet)	0.0606	PyTorch (GPU)
LLHFNet (VGG16)	0.0763	PyTorch (GPU)
LIME	0.4914	MATLAB (CPU)
SRIE	12.1865	MATLAB (CPU)

6.2.3 Qualitative and Perceptual Evaluations



(a) Sample 1 (b) Sample 2 (c) Sample 3

Figure 19*:* Visual comparison of sample LLIs from SICE part2 subset (Cai et al. 2018) and their enhanced versions. LLHFNet is based on MobileNetv2 (Sandler et al. 2018) feature extractor


Input

LIME (Li et al. 2015)

SRIE (Fu et al. 2016)

DeepUPE (Wang et al. 2019)

EnlightenGAN (Jiang et al. 2019)

ZeroDCE (Guo et al. 2020)

LLHFNet

Figure 20: Visual comparison of an input image with normal exposure level and its enhanced versions. LLHFNet is based on MobileNetv2 (Sandler et al. 2018) feature extractor

We show in Fig. 19 a visual comparison of sample LLIs enhanced by the different models used in our evaluation. As can be seen LLHFNet produces visually pleasing images with minimal artifacts. In the first and second images (Fig.19.a and b) our model shows the best green color restoration for the trees and grass. Moreover, in the first image our model is able to uncover the dark regions of the fence and in the second image it properly restores the white cloud without overexposing it like in the case with EnlightenGAN (Jiang et al. 2019) enhancement or deviating its color to look blue like in ZeroDCE (Guo et al. 2020) and SRIE (Fu et al. 2016) enhancements. In the third image (Fig.19.c) our model shows a good illumination level and produces results similar to ZeroDCE (Guo et al. 2020) and SRIE (Fu et al. 2016). In Fig.20, we show the enhancements obtained while using an input image with almost a normal exposure level. While models like LIME (Li et al. 2015) and EnlightenGAN (Jiang et al. 2019) tend to overexpose the image especially the light from the windows, our model performs a slight and minimal enhancement. It is worth noting that our enhancement approach properly handles normal exposure levels in input images, and it can pass the image without any enhancement by producing filter parameters $\gamma_L = \gamma_H = 1$.

We additionally perform a qualitative user study to evaluate the human's perception of the results produced by our model and the models considered in our evaluation. A total of 20 images from Part2 of SICE dataset (Cai et al. 2018) are used in our study in which the reference input LLI and the enhanced image are placed side by side. A total of 76 users (senior master and computer engineering students) are invited to independently assign scores for the visual quality of the enhanced images. The scores range from 0 to 10 (worst to best) and users are asked to give scores based on three criteria: i) level of exposure (over- or under- exposed), ii) color deviations and iii) overall beauty of the image. Each model received at least 60 scores as we filtered certain outliers identified in the responses. These outliers correspond to scores which are either extremely low for images which have a good quality or extremely high for images which are not very visually pleasing. We show in Fig 21. the scores obtained for the different enhancement models. It is notable that our model is ranked second best, and its results are favored by the human users thus indicating its capability of producing perceptually pleasing enhanced images.

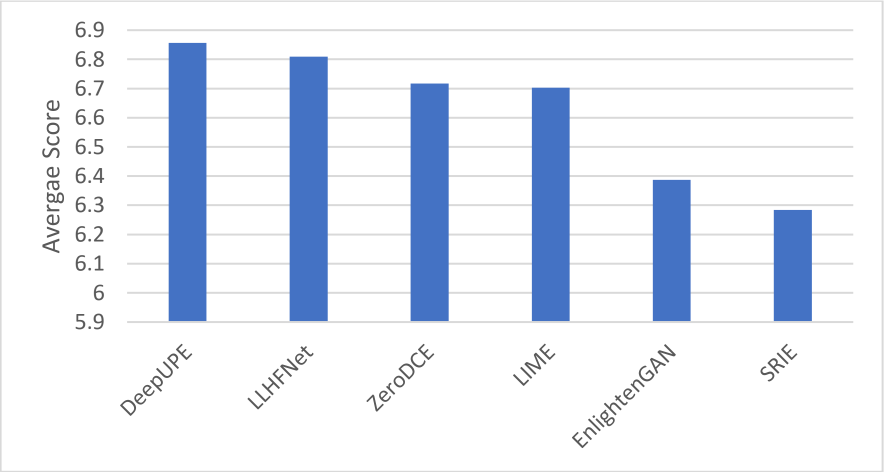
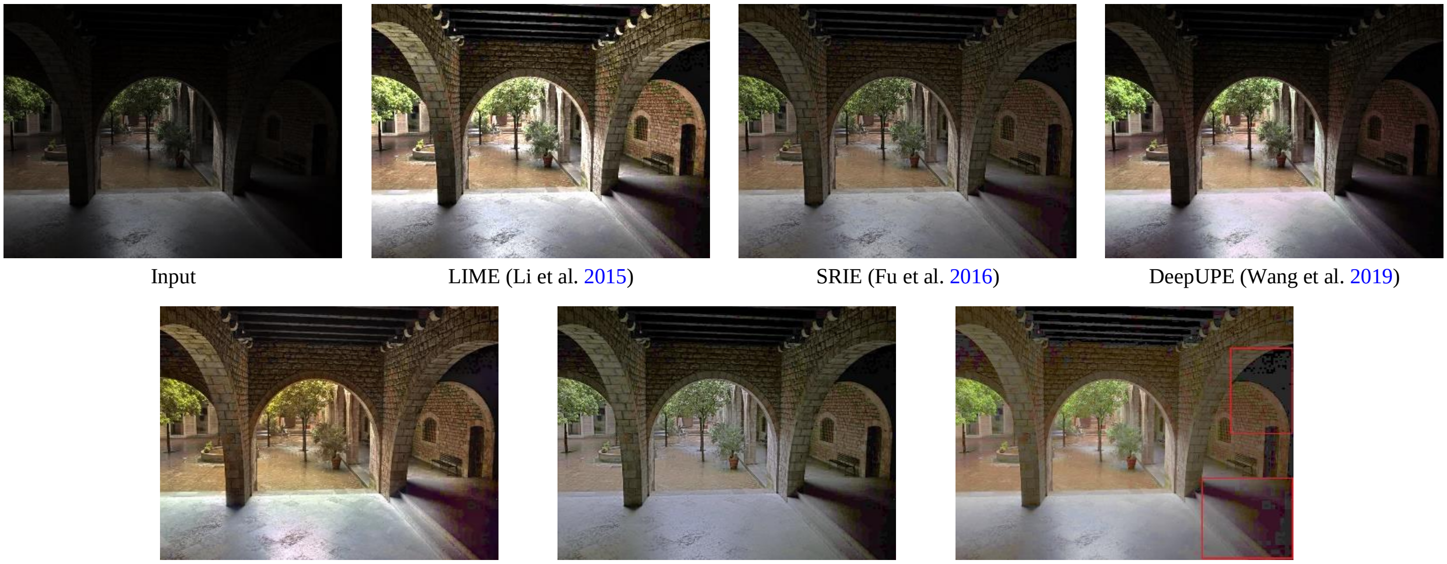


Figure 21: Average user scores for the enhancement models ranked from best to worst

6.2.4 Limitations

Similar to many enhancement models, our approach fails to handle extremely LLIs and tends to produce artifacts. It also shows exposed artifacts in certain very dark regions in the image as can be seen in the samples in Fig. 22 where the highlighted red boxes in our model enhancement show some of these artifacts. Yet, our model properly restores images with low to medium exposure levels and can perfectly handle normal exposed images making it a good enhancement approach whose final goal is to be integrated into classification models.



Input

LIME (Li et al. 2015)

SRIE (Fu et al. 2016)

DeepUPE (Wang et al. 2019)

EnlightenGAN (Jiang et al. 2019)

ZeroDCE (Guo et al. 2020)

LLHFNet

Figure 22: Visual comparison on a challenging LLI where artifacts may appear

Chapter 7
LLI Enhancer-Classifier Model

In this chapter we describe the design of the integration of our enhancement model into one classification model specifically ResNet50 (He et al. 2016) and call it Enhancer-Classifier model. ResNet50 (He et al. 2016) is based on residual blocks which have shortcut connections and is made up of 50 layers for feature extraction and one fully connected (FC) layer for classification. We choose ResNet50 (He et al. 2016) as it is very effective for the classification task and has smaller model size and lower training time compared to VGG16 (Simonyan and Zisserman 2015) for example. This allows faster fine tuning and evaluation for the combined models. We aim by integrating our enhancement model into ResNet50 (He et al. 2016) to perform a joint learning and optimization for both enhancement and classification performance simultaneously. Such an approach will embed an internal enhancement capability to the classifier allowing it to handle LLIs and NLIs and at the same time it will adapt to the enhanced images thus resulting in a robust classification.

7.1 Methodology

7.1.1 Designs Tried

In our first approach to perform the integration, the network architecture is designed as follows: we feed the input image into the ResNet50 (He et al. 2016) feature extractor then pass its output feature maps to the enhancement head which estimates the frequency filter parameters and then use the HF algorithm to produce the enhanced image which will be optimized through the enhancement loss. We then use another extractor network made up of six convolutional layers followed by ReLU activation layers and maxpooling layers to downsize the feature maps of the enhanced image to the same size of those obtained at the output of ResNet50 (He et al. 2016). Then we merge both feature maps and pass them to the FC layer used by ResNet50 (He et al. 2016) classifier to predict the classification scores for the given image and optimize them through a dedicated classification loss function. The architecture is shown in Fig. 23.

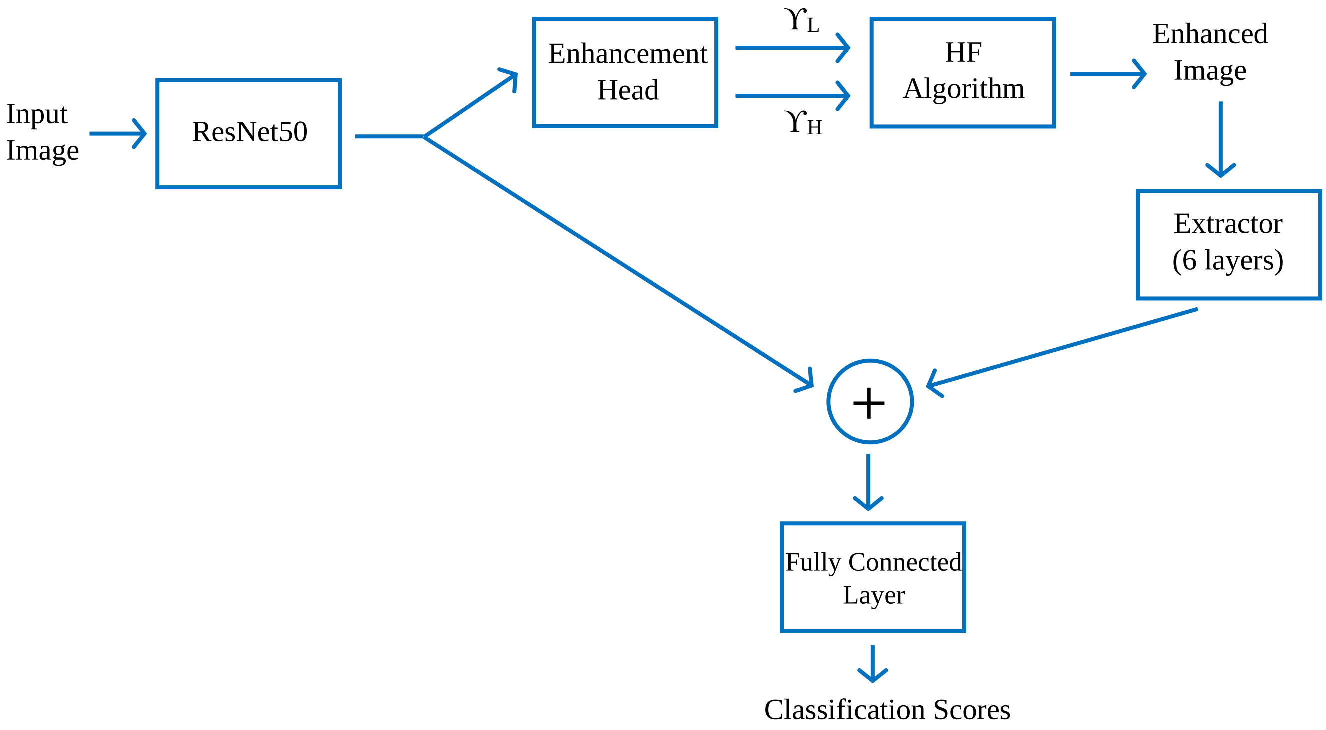


Figure 23: Architecture of the first design of the Enhancer-Classifier model

After training the network, the obtained classification accuracy values were not good compared to ResNet50 (He et al. 2016) classification model with no integrated enhancement capability. We attribute this to the fact that the feature maps obtained through the extractor using only six convolutional layers are not at the same feature level of those extracted through the deeper 50 layers of ResNet50 (He et al. 2016). So, the performed merging operation may be distorting the high-level features at the output of the feature extractor and thus degrading the classification results.

In the second approach, we overcome the issue of different feature levels by feeding the enhanced image back to ResNet50 (He et al. 2016) in a second stage of feature extraction. We then pass the output feature maps to a FC layer for classification. The architecture is shown in Fig. 24. After training the network, the classification accuracy improved and exceeded that of the classification model alone indicating a better performance. Yet, the issue with this architecture is the stability over each experiment as we noticed that the classification accuracy was fluctuating largely with each experiment. We attribute this to the fact that the feature extractor is performing dual feature extraction tasks as it first processes the input image then at a second stage handles the enhanced image and is optimized to improve both enhancement and classification. This may overload the extractor and alienate the feature maps from what they are expected to be.

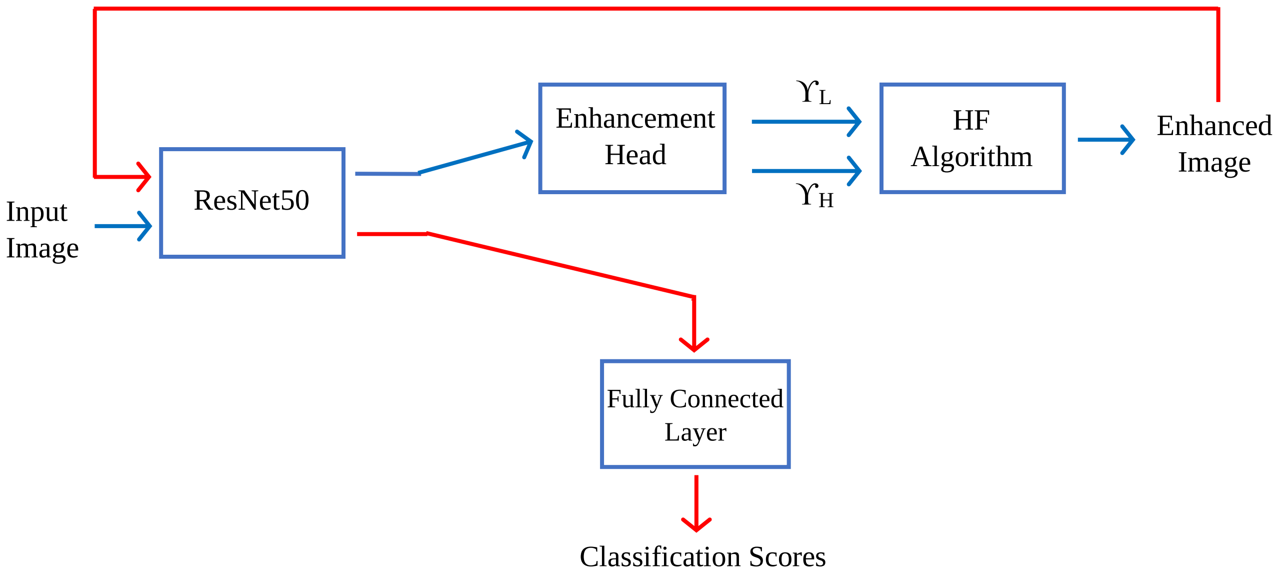


Figure 24: Architecture of the second design of the Enhancer-Classifier model

7.1.2 Final Design

In our final approach, we design the Enhancer-Classifier model as follows: the input image is fed into a first ResNet50 (He et al. 2016) feature extractor whose output feature maps are then passed to the enhancement head to estimate the frequency filter parameters and finally process them through the HF algorithm to produce the enhanced image which will be optimized by the enhancement loss. We then feed the enhanced image into a second ResNet50 (He et al. 2016) feature extractor to extract features which are at the same level of those obtained from the first extractor. Then both feature maps are merged in an element wise addition and passed to the FC layer of the second ResNet50 (He et al. 2016) to produce the classification scores which represent the output of the Enhancer-Classifier model along with the enhanced image. Moreover, we add the FC layer of the first

ResNet50 (He et al. 2016) at its output branch and use it only for the classification loss to further optimize the classification performance of the first feature extractor. We found experimentally that this layer improves the results, and we attribute this to the fact that the first extractor is performing a dual task of optimizing both enhancement and classification performance. This may overload the extractor and thus it is better to use the classification loss of the added FC layer to improve its optimization. The final design of our Enhancer-Classifier model is shown in Fig. 25.

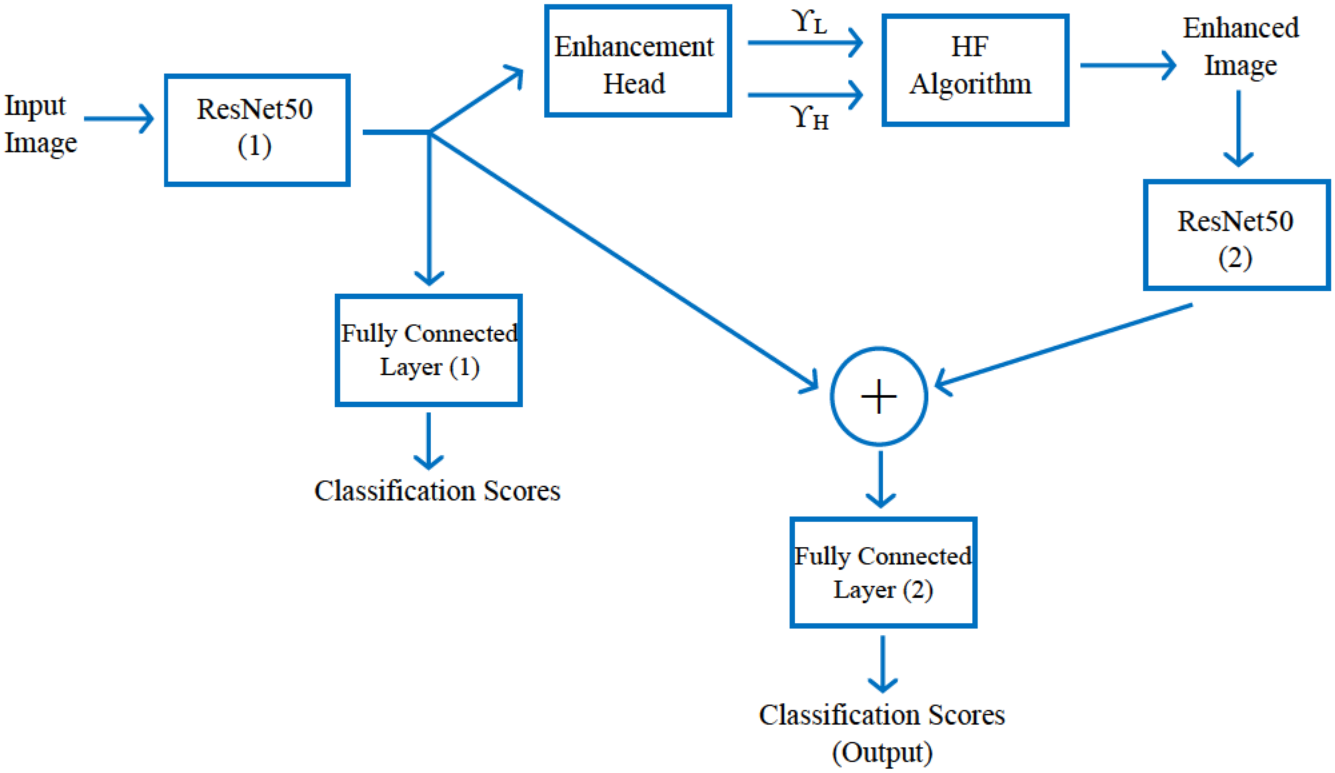


Figure 25: Final design of the Enhancer-Classifier model

7.1.3 Loss Function

The loss function is used to perform a joint optimization for both classification and enhancement. We use the cross-entropy loss for the classification loss ($clsLoss$) for both FC layers ($clsLoss1$ for the first layer and $clsLoss2$ for the second layer). The enhancement loss ($enhLoss$) is same to equation 17 with $\alpha = 0.08$. The overall loss is as follows:

$$\begin{aligned} loss &= clsLoss + enhLoss \\ &= clsLoss1 + clsLoss2 + enhLoss \end{aligned} \tag{18}$$

7.2 Experimental Results

7.2.1 Datasets

Since two tasks are joint together in a supervised training setting, our Enhancer-Classifier model requires paired LLIs/NLIs for enhancement and class labels for classification. So, we use Pascal VOC dataset (2012+2007) (Everingham et al. 2012) for training the model. First, we synthetically generate five different exposure levels using gamma correction with gamma values {4.5, 3.5, 2.5} correspond to low exposure levels and gamma values {0.5, 0.8} correspond to over exposure levels. The reason for using a mixture of five levels ranging from underexposed to overexposed is to allow the enhancement algorithm to learn handling various input exposure levels. In total the training dataset consists of 8500 images and the validation dataset of 1125 images equally divided among the used exposure levels along with their reference NLIs and class labels. The images are converted to HSV domain where only the Value channel-based image is used in training.

For our evaluation, we form a test only subset from Pascal VOC2007 (Everingham et al. 2012) dataset made up of 3000 images divided equally among the used exposure levels. In addition to the synthetic images, we use 3000 real world LLIs from ExDark dataset (Loh and Chan 2019) to further evaluate the performance of the Enhancer-Classifier model. All the images for training and evaluation are resized to 512x512.

7.2.2 Implementation Details

Our Enhancer-Classifier is implemented using PyTorch on a P100 Tesla Nvidia GPU. A batch size of 8 is used. Adam optimizer with default parameters and a reduce on plateau based decaying learning rate with initial value of 1e-5 are used for network optimization. For the first epoch, we multiply the classification loss by 0.1 to give it less weight to the advantage of stabilizing enhancement and warming up the joint models. Furthermore, a pre-trained ResNet50 (He et al. 2016) on ImageNet (Deng et al. 2009) database is used.

We train three different models as follows: i) the Enhancer-Classifier on the 8500 gamma corrected images (five exposure levels) of the training dataset (referred to as **EnhCls (ϒ corrected)**), ii) ResNet50 (He et al. 2016) classifier with no enhancement capability on the same dataset (referred to as **Cls (ϒ corrected)**) and iii) ResNet50 (He et al. 2016) classifier with no enhancement capability on the 8500 NLIs of the training dataset (referred to as **Cls (NLIs)**).

7.2.3 Results

We show in table 9 the classification accuracy results obtained by the three trained models evaluated on the ExDark (Loh and Chan 2019) subset, the mixed exposure levels (ϒ corrected) based Pascal VOC2007 (Everingham et al. 2012) test only subset and their reference NLIs subset. The following observations can be highlighted from the results:

1. The **EnhCls (ϒ corrected)** model which has an internal enhancement capability shows an improvement of 3.86% compared to the **Cls (ϒ corrected)** model alone trained on the same dataset of mixed exposure levels (3/5 have low exposure) and evaluated on the Pascal VOC2007 test only subset of similar exposure levels distribution.
2. The **EnhCls (ϒ corrected)** model evaluated on the gamma corrected images of Pascal VOC2007 test only subset achieves approximately a similar accuracy to the **Cls (NLIs)** trained and evaluated on the NLIs of the same subset. This indicates that the embedded enhancement contributes to an improvement equivalent to training the classifier on NLIs only.
3. The **EnhCls (ϒ corrected)** model achieves the best accuracy on the NLIs of Pascal VOC2007 test only subset even better than the **Cls (NLIs)** model which is trained and evaluated on the NLIs. This is due to the ability of our enhancement algorithm to handle NLIs by adding only a minor enhancement and avoiding over exposure as indicated in section 6.2.3.
4. The **Cls (NLIs)** model trained on NLIs shows a highly degraded performance when evaluated on gamma corrected images of Pascal VOC2007 test only subset, while the **Cls (ϒ corrected)** model trained on gamma corrected images

shows a better performance on gamma corrected images and approximately similar performance on NLIs. This indicates that the **Cls (ϒ corrected)** model has a better robustness against varying light conditions. Therefore, training a classifier model on varying exposure levels is better than limiting the training to normal light conditions.

5. The **EnhCls (ϒ corrected)** model depicts the best classification accuracy on the 3000 images of ExDark test subset confirming the model good performance on real-world LLIs.
6. The Enhancer-Classifier model shows robust classification performance on both LLIs and NLIs.

Table 9: Classification accuracy (in %) of the three trained models evaluated using Pascal VOC2007 (Everingham et al. 2012) test only subset and ExDark (Loh and Chan 2019) subset.

Dataset	**EnhCls (ϒ corrected)**	**Cls (ϒ corrected)**	**Cls (NLIs)**
VOC2007 test only (ϒ corrected)	**86.22**	82.36	78.09
VOC2007 test only (NLIs)	**88.42**	86.26	86.54
ExDark	**71.73**	68.48	61.17

We show in tables 10 and 11 the classification results obtained by ResNet50 (He et al. 2016) classifier trained on NLIs of our training dataset and evaluated on enhanced images of Pascal VOC2007 (Everingham et al. 2012) test only (ϒ corrected) subset and ExDark (Loh and Chan 2019) subset. The original images are enhanced using EnlightenGAN (Jiang et al. 2019), DeepUPE (Wang et al. 2019), ZeroDCE (Guo et al. 2020) and our proposed enhancement model based on ResNet50 (He et al. 2016) feature extractor. We aim from this evaluation to compare our joint models (Enhancer-Classifier model) against the normal pipeline usually followed in the literature and based on using the LLI enhancement model separately as a preprocessing step after which classification is performed on the enhanced images using a classifier pre-trained on NLIs. This pipeline may result in a slight or even no improvement on the classification performance as mentioned by VidalMata et al. (2020) and we want to validate our Enhancer-Classifier against it. The results highlight the following observations:

1. Our enhancement model (LLHFNet) shows the best classification accuracy values on both Pascal VOC2007 (Everingham et al. 2012) test only subset and ExDark (Loh and Chan 2019) subset (80.81% and 66.18% respectively) compared with other enhancement solutions. Thus, our enhancement model as standalone model is able to significantly improve the target classification task.
2. The **EnhCls (ϒ corrected)** model (table 9) shows an improvement of almost 5.5% compared to the best performing enhancement solution on both Pascal VOC2007 (Everingham et al. 2012) test only subset and ExDark (Loh and Chan 2019) subset following the normal pipeline in tables 10 and 11. Thus, the joint optimization and training of enhancement and classification models proves to be efficient and better than the normal pipeline depicted by separate enhancement followed by classification.

Table 10: Classification accuracy (in %) of the classifier trained on NLIs and evaluated on original and enhanced Pascal VOC2007 (Everingham et al. 2012) test only (ϒ corrected**)** subset by different enhancement models. LLHFNet uses ResNet50 (He et al. 2016) feature extractor.

Model	**Cls (NLIs)**
Original	78.09
LIME	75.90
EnlightenGAN	78.40
DeepUPE	79.53
SRIE	79.71
ZeroDCE	79.87
LLHFNet	**80.81**

Table 11: Classification accuracy (in %) of the classifier trained on NLIs and evaluated on original and enhanced ExDark (Loh and Chan 2019) subset by different enhancement model. LLHFNet uses ResNet50 (He et al. 2016) feature extractor.

Model	**Cls (NLIs)**
Original	61.17
LIME	61.51
EnlightenGAN	62.17
DeepUPE	64.34
SRIE	63.27
ZeroDCE	65.30
LLHFNet	**66.18**

In our third evaluation, we compare the three trained models using enhanced Pascal VOC2007 (Everingham et al. 2012) test only subset and enhanced ExDark (Loh and Chan 2019) subset by the different enhancement models. We show the results in tables 12 and 13 and the following observations can be drawn:

1. The **EnhCls (ϒ corrected)** model depicts the best classification results although it is internally performing a second enhancement in addition to the separate enhancement performed by the considered enhancement models. This indicates that the Enhancer-Classifier can perfectly adapt to the data domain of enhanced

images which may contain artifacts and amplified noise and can additionally add a minor enhancement when needed and thereby imposing further improvements.

2. The **Cls (NLIs)** model trained only on NLIs shows the worst classification results compared to the other two models. This may indicate that the data domain of NLIs does not correlate with that of enhanced images which contain artifacts, noise, and varying light conditions. Therefore, processing enhanced images using classifiers pre-trained on NLIs may not be the perfect strategy to benefit from the enhancement task as preprocessing step for classification.

Table 12: Classification accuracy (in %) obtained by the three trained models evaluated on the original and enhanced images from Pascal VOC2007 (Everingham et al. 2012) test only subset. LLHFNet uses ResNet50 (He et al. 2016) feature extractor.

Model	**EnhCls (ϒ corrected)**	**Cls (ϒ corrected)**	**Cls (NLIs)**
Original	**86.22**	82.35	78.09
LIME	**81.76**	77.47	75.90
EnlightenGAN	**82.83**	80.32	78.40
DeepUPE	**85.25**	81.96	79.53
SRIE	**84.92**	81.28	79.71
ZeroDCE	**83.77**	81.33	79.87
LLHFNet	**82.44**	80.45	80.81

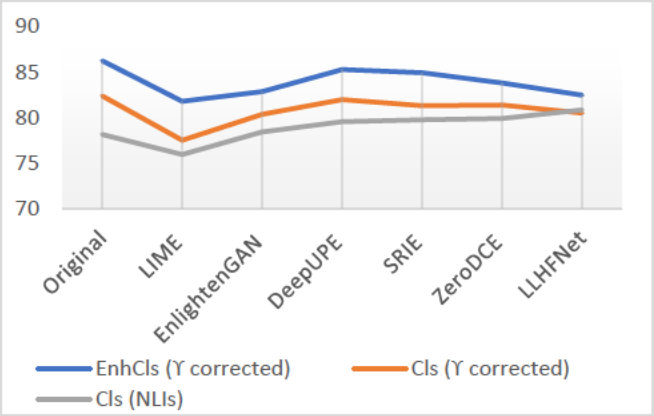
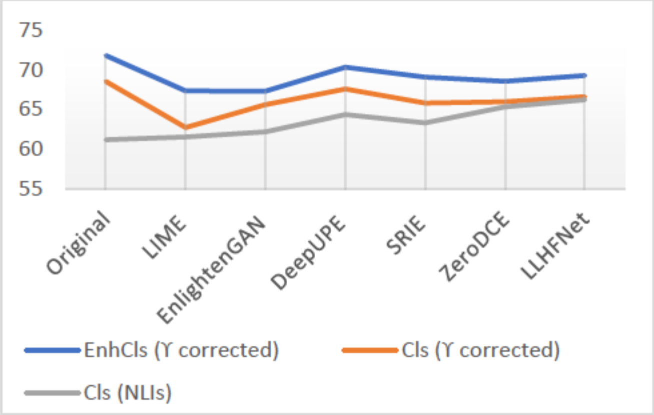


Table 13: Classification accuracy (in %) obtained by the three trained models evaluated on the original and enhanced images from ExDark (Loh and Chan 2019) subset. LLHFNet uses ResNet50 (He et al. 2016) feature extractor.

Model	**EnhCls (ϒ corrected)**	**Cls (ϒ corrected)**	**Cls (NLIs)**
Original	**71.73**	68.48	61.17
LIME	**67.28**	62.70	61.51
EnlightenGAN	**67.26**	65.55	62.16
DeepUPE	**70.27**	67.54	64.34
SRIE	**69.03**	65.79	63.27
ZeroDCE	**68.49**	65.92	65.29
LLHFNet	**69.23**	66.58	66.18



We finally evaluate the enhancement performance achieved by the **EnhCls (ϒ corrected)** model having the maximum classification accuracy to better understand if there is a relation between the best classification performance and the involved enhancement. For the Pascal VOC2007 (Everingham et al. 2012) test only subset we use full reference metrics: SSIM (Wang et al. 2004), PSNR and MAE, but for ExDark (Loh and Chan 2019) subset which does not have reference NLIs we use non reference metrics: NIQE (Mittal et al. 2013) and BRISQUE (Mittal et al. 2012). The results are shown in tables 14 and 15 and the following points can be noted:

1. The **EnhCls (ϒ corrected)** model achieves the best performance following all three metrics on the Pascal VOC2007 (Everingham et al. 2012) test only subset thus indicating the effectiveness of the internally embedded enhancement. Moreover, the **EnhCls (ϒ corrected)** has the best classification performance compared to the other models (tables 9 and 10) thus reflecting that the joint optimization used in the **EnhCls (ϒ corrected)** has created a possible correlation between the best achieved classification and the involved best enhancement quality.
2. The **EnhCls (ϒ corrected)** model shows the third worst performance on the ExDark (Loh and Chan 2019) subset following NIQE (Mittal et al. 2013) metric, although it has the best classification performance (tables 9 and 11). In contrast, EnlightenGAN (Jiang et al. 2019) is ranked as the best enhancement model following BRISQUE (Mittal et al. 2012) metric and second best following NIQE (Mittal et al. 2013) metric, yet it possesses some of the worst classification accuracy results (table 11). Thus, the best classification performance is not associated with the best enhancement quality.
3. Although we performed a joint optimization for both enhancement and classification tasks, yet it remains uncertain whether a good classification performance is consistent with a good enhancement quality.
4. It is notable that our proposed enhancement model alone achieves second best performance on Pascal VOC2007 (Everingham et al. 2012) test only subset following all three metrics and fourth best on ExDark (Loh and Chan 2019) subset following NIQE (Mittal et al. 2013) metric. This shows that the designed

enhancement algorithm is competitive compared to existing state of art enhancement models.

Table 14: Quantitative evaluation of enhancement performance on Pascal VOC2007 (Everingham et al. 2012) test only subset using the **EnhCls (ϒ corrected)** and other enhancement models. LLHFNet uses ResNet50 (He et al. 2016) feature extractor. The best result is in red and second best is in green. Models are ranked from best to worst following SSIM.

Model	**SSIM ↑**	**PSNR ↑**	**MAE ↓**
EnhCls (ϒ corrected)	**0.76**	**16.96**	**105.64**
LLHFNet	**0.731**	**15.69**	**119.92**
DeepUPE	0.730	14.30	143.89
ZeroDCE	0.67	14.96	139.05
SRIE	0.629	13.50	154.69
LIME	0.6286	13.33	159.68
EnlightenGAN	0.6284	13.63	152.32

Table 15: Quantitative evaluation of enhancement performance on ExDark (Loh and Chan 2019) subset using the **EnhCls (ϒ corrected)** and other enhancement models. LLHFNet uses ResNet50 (He et al. 2016) feature extractor. The best result is in red and second best is in green. Model are ranked from best to worst following NIQE.

Model	**NIQE ↓**	**BRISQUE ↓**
SRIE	**3.54**	29.33
EnlightenGAN	**3.71**	**27.29**
DeepUPE	3.87	**28.92**
LLHFNet	3.95	29.95
EnhCls (ϒ corrected)	3.96	29.31
LIME	4.01	30.50
ZeroDCE	4.14	30.64

Chapter 8

Limitations, Future Works, and Impact

This chapter presents the limitations of our research, future works that can be investigated, and impact of our designed model.

8.1 Limitations

We point out two major limitations of our enhancement approach:

1. Our LLI Enhancer, LLHFNet, used as a standalone enhancement solution, does not produce top enhancement performance. While it properly restores low and medium exposure levels, and properly handles normal exposure levels, yet it tends to produce exposed artifacts on extremely dark images, as discussed in section 6.2.4. Note that this is a common limitation with most existing enhancement solutions, and remains a major challenge in the literature.
2. The involved homomorphic filtering-based enhancement algorithm uses Fourier transform and its inverse, thus imposing additional computational overhead, which might hinder real-time LLI enhancement.

8.2 Future work

In the future, we plan to further investigate the below possible improvements:

1. Boosting the computational performance of the enhancement algorithm by using spatial domain-based homomorphic filtering instead of relying on the frequency domain-based approach.
2. Tackling the issue of exposed artifacts produced by the enhancement algorithm on LLIs with extremely low exposure levels.
3. Transforming the enhancement model into an unsupervised or semi-supervised solution, as an attempt to ease the dependence on paired LLIs/NLIs.
4. Integrating the enhancement model into other high-level computer vision tasks like object detection, semantic segmentation, and tracking.

8.3 Impact

Since nighttime accounts for a considerable time of our daily life, deploying any computer vision model that performs classification remains limited as such a system will struggle with low-light conditions. Our designed LLI Enhancer-Classifier model can perfectly tackle this critical issue as it performs a robust classification under both normal and low light conditions. This model can be integrated as a computer vision system into modern artificial intelligence-based applications like autopilot car systems, robot visual systems, security surveillance cameras, among others, in order to improve their operation in nighttime.

Chapter 9

Conclusion

In this report, we address the problem of LLI enhancement in two ways: i) standalone, as a separate task, and ii) end-to-end, as a pre-processing stage embedded within another high-level computer vision task, namely object classification.

First, we gave an overview of current DL-based LLI enhancement models, which we organized in five main categories: encoder-decoder and CNN-based, Retinex-based, Fusion-based, GAN-based, and more recent Zero Reference models. Then, we described the experimental evaluation and results comparing 10 of the most recent DL-based LLI enhancement models. We conducted three main experiments evaluating: i) visual and perceptual quality, where LLI enhancement models were evaluated as standalone applications, ii) detection and classification quality, achieved by 4 different object detection models applied on LLIs and their enhanced counterparts, where LLI enhancement models were embedded as a pre-processing step in the overall pipeline, and iii) feature analysis, considering the effect of LLI enhancement on the resulting image features and its impact on object detection performance. We then summarized our empirical observations and highlighted various potential research directions hoping that the unified presentation of DL-based LLI enhancement will contribute to strengthen further research on the subject.

Inspired by the results of the comparative study we proposed a DL-based *LLI Enhancer* model which is tailored for the object classification task. The model performs an image to a special designed frequency filter learning. The filter parameters are estimated via any of the feature extractors commonly deployed in the classification task and are then fed to a HF algorithm to enhance the original LLI. We then designed a *LLI Enhancer-Classifier* model which integrates our enhancement model into ResNet50 (He et al. 2016) to perform a joint optimization for both image enhancement and classification tasks. Experimental results show that the enhancement model possesses a competitive performance compared to state of art enhancement solutions. Moreover, the Enhancer-

Classifier model shows a robust classification performance against varying light conditions. It also significantly boosts the classification accuracy when compared with the traditional pipeline followed in the literature consisting of separate enhancement followed by classification. Furthermore, our results show that NLIs may have a different data domain from enhanced images, and processing enhanced images on classification models pre-trained on NLIs may not be a successful and effective approach to improve the classification task.

References

Abdullah-Al-Wadud, M., Kabir, M. H., Dewan, M. A., & Chae, O. (2007). A Dynamic Histogram Equalization for Image Contrast Enhancement. *IEEE Transactions on Consumer Electronics, 53*(2), 593-600.

Abu-Khzam, F. N., Daudjee, K., Mouawad, A. E., & Nishimura, N. (2015). On Scalable Parallel Recursive Backtracking. *Journal of Parallel and Distributed Computing*, 84:65-75.

Abu-Khzam, F., Li, S., Markarian, C., Heide, F., & Podlipyan, P. (2019). Efficient Parallel Algorithms for Parameterized Problems. *Theoretical Computer Science, 786*, 2-12.

Abu-Khzam, F. N., Markarian, C., Heide, F. M., & Schubert, M. (2018). Approximation and Heuristic Algorithms for Computing Backbones in Asymmetric Ad-hoc Networks. *Theory of Computing Systems*, *62*(8):1673-1689

Agustsson, E., & Timofte, R. (2017). NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 1122-1131).

Amigó, J. M., Kocarev, L., & Tomovski, I. (2007). Discrete Entropy. *Physica D: Nonlinear Phenomena, 228*(1), 77-85.

Arbelaez, P., Maire, M., Fowlkes, C., & Malik, J. (2011). Contour Detection and Hierarchical Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 33*(5), 898–916.

Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein Generative Adversarial Networks. In *International Conference on Machine Learning (ICML)* (pp. 214-223).

Bileschi, S. M. (2006). Streetscenes: Towards Scene Understanding in Still Images. *Massachusetts Inst. of Tech. Cambridge, Tech. Rep.*

Blau, Y., & Michaeli, T. (2018). The Perception-Distortion Tradeoff. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 6228-6237).

Bychkovsky, V., Paris, S., Chan, E., & Durand, F. (2011). Learning Photographic Global Tonal Adjustment with a Database of Input / Output Image Pairs. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 97-104).

Cai, J., Gu, S., & Zhang, L. (2018). Learning a Deep Single Image Contrast Enhancer from Multi-Exposure Images. *IEEE Transactions on Image Processing, 27*(4), 2049–2062.

Chen, C., Chen, Q., Xu, J., & Koltun, V. (2018). Learning to See in the Dark. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 3291-3300).

Chen, Y.-S., Wang, Y.-C., Kao, M.-H., & Chuang, Y.-Y. (2018). Deep Photo Enhancer: Unpaired Learning for Image Enhancement from Photographs with GANs. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 6306-6314).

Chen, Z., Abidi, B. R., Page, D. L., & Abidi, M. A. (2006). Gray-level grouping (GLG): An Automatic Method for Optimized Image Contrast Enhancement-part I: The Basic Method. *IEEE Transactions on Image Processing, 15*(8), 2290-2302.

Cheng, Y., Yan, J., & Wangy, Z. (2019). Enhancement of Weakly Illuminated Images by Deep Fusion Networks. In *IEEE International Conference on Image Processing (ICIP)* (pp. 924-928).

Dabov, K., Foi, A., & Egiazarian, K. (2006). Image Denoising with Block Matching and 3D Filtering. In *Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning* (pp. 354-365).

Dang-Nguyen, D., Pasquini, C., Conotter, V., & Boato, G. (2015). RAISE: A Raw Images Dataset for Digital Image Forensics. In *ACM Multimedia Systems Conference (MMSys)* (pp. 219-224).

Deng, L. (2014). A Tutorial Survey of Architectures, Algorithms, and Applications for Deep Learning. *APSIPA Transactions on Signal and Information Processing, 3*(2).

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database. In *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 248-255).

Dong, C., Loy, C., & Tang, X. (2016). Accelerating the Super-Resolution Convolutional Neural Network. In *European Conference on Computer Vision (ECCV)* (pp. 391-407).

Ebrahimi, D., Sharafeddine, S., Ho, P., & Assi, C. (2020). Autonomous UAV Trajectory for Localizing Ground Objects: A Reinforcement Learning Approach. *IEEE Transactions on Mobile Computing*, 1-1.

Everingham, M., Gool, L. V., Williams, C. K., Winn, J., & Zisserman, A. (2012). The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision, 88*(2), 303-338.

Fu, X., Zeng, D., Huang, Y., Liao, Y., Ding, X., & Paisley, J. (2016). A Fusion-Based Enhancing Method for Weakly Illuminated Images. *Signal Processing, 129*, 82-96.

Fu, X., Zeng, D., Huang, Y., Zhang, X.-P., & Ding, X. (2016). A Weighted Variational Model for Simultaneous Reflectance and Illumination Estimation. In *IEEE Conference on Computer Vision & Pattern Recognition (CVPR)* (pp. 2782-2790).

Gharbi, M., Chen, J., Barron, J. T., Hasinoff, S. W., & Durand, F. (2017). Deep Bilateral Learning for Real-time Image Enhancement. *ACM Transactions on Graphics, 36*(4), 118:1-118:12.

Gonzalez, R. C., & Woods, R. E. (2018). *Digital Image Processing (4th edition).* New York, NY: Pearson.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative Adversarial Networks. In *International Conference on Neural Information Processing Systems (NIPS)* (pp. 2672–2680).

Grubinger, M., Clough, P., Müller, H., & Deselaers, T. (2006). The IAPR TC12 Benchmark: A New Evaluation Resource for Visual Information Systems. *International Workshop OntoImage, 5*(10).

Gu, K., Zhai, G., Lin, W., & Liu, M. (2016). The Analysis of Image Contrast: From Quality Assessment to Automatic Enhancement. *IEEE Transactions on Cybernetics, 46*(1), 284–297.

Guo, C., Li, C., Guo, J., Loy, C. C., Hou, J., Kwong, S., & Cong, R. (2020). Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. In *IEEE Confernence on Computer Vision and Pattern Recognition (CVPR)* (pp. 1777-1786).

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep Learning for Visual Understanding: A Review. *Neurocomouting, 187*, 27- 48.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. In *IEEE International Conference on Computer Vision (ICCV)* (pp. 2980-2988).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770-778).

Hojatollah, Y., & Wang, Z. (2013). Objective Quality Assessment of Tone-mapped Images. *IEEE Transactions on Image Processing, 22*(2), 657-667.

Hua, W., & Xia, Y. (2018). Low-Light Image Enhancement Based on Joint Generative Adversarial Network and Image Quality Assessment. In *International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)* (pp. 1-6).

Huang, G., Liu, Z., Maaten, L. V., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2261-2269).

Huang, J.-B., Singh, A., & Ahuja, N. (2015). Single Image Super-Resolution from Transformed Self-Exemplars. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5197-5206).

Huang, S.-C., Cheng, F.-C., & Chiu, Y.-S. (2013). Efficient Contrast Enhancement Using Adaptive Gamma Correction with Weighting Distribution. *IEEE Transactions on Image Processing, 22*(3), 1032-1041.

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). SqueezeNet: AlexNet-level Accuracy with 50x Fewer Parameters and <0.5MB Model Size. *arXiv:1602.07360*.

Jiang, L., Jing, Y., Hu, S., Ge, B., & Xiao, W. (2018). Deep Refinement Network for Natural Low-Light Image Enhancement in Symmetric Pathways. *Symmetry 10*(10) 491.

Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., et al. (2019). EnlightenGAN: Deep Light Enhancement without Paired Supervision. *arXiv:1906.06972*.

Jobson, D. J., Rahman, Z., & Woodel, G. A. (1997a). Properties and Performance of a Center/Surround Retinex. *IEEE Transactions on Image Processing, 6*(3), 451-462.

Jobson, D. J., Rahman, Z., & Woodell, G. A. (1997b). A Multiscale Retinex for Bridging the Gap Between Color Images and the Human Observation of Scenes. *IEEE Transactions on Image processing, 6*(7), 965-976.

Jolicoeur-Martineau, A. (2018). The Relativistic Discriminator: A Key Element Missing from Standard GAN. *arXiv:1807.00734*.

Kalantari, N. K., & Ramamoorthi, R. (2017). Deep High Dynamic Range Imaging of Dynamic Scenes. *ACM Transactions on Graphics, 36*(4), 144:1–144:12.

Kim, G., Kwon, D., & Kwon, J. (2019). Low-Lightgan: Low-Light Enhancement via Advanced Generative Adversarial Network with Task-Driven Training. In *IEEE International Conference on Image Processing (ICIP)* (pp. 2811-2815).

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In *Neural Information Processing Systems Conference (NIPS)* (pp. 1106-1114).

Land, E., & McCann, J. (1971). Lightness and Retinex Theory. *Journal of the Optical Society of America, 61*(1), 1-11.

Le, Q. V., Ngiam, J., Coates, A., Lahiri, A., Prochnow, B., & Ng, A. Y. (2011). On Optimization Methods for Deep Learning. In *International Conference on Machine Learning (ICML)* (pp. 265–272).

Lee, H.-G., Yang, S., & Sim, J.-Y. (2015). Color Preserving Contrast Enhancement for Low-light Level Images based on Retinex. In *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)* (pp. 884-887).

Li, C., Guo, J., Porikli, F., & Pang, Y. (2018). LightenNet: A Convolutional Neural Network for Weakly Illuminated Image Enhancement. *Pattern Recognition Letters, 104*, 15-22.

Li, L., Wang, R., Wang, W., & Gao, W. (2015). A Low-light Image Enhancement Method for Both Denoising and Contrast Enlarging. *IEEE International Conference on Image Processing (ICIP)* (pp. 3730-3734).

Li, M., Liu, J., Yang, W., Sun, X., & Guo, Z. (2018). Structure-Revealing Low-Light Image Enhancement Via Robust Retinex Model. *IEEE Transactions on Image Processing, 27*(6), 2828-2841.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal Loss for Dense Object Detection. In *IEEE International Conference on Computer Vision (ICCV)* (pp. 2999-3007).

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., et al. (2014). Microsoft COCO: Common Objects in Context. In *European Conference on Computer Vision (ECCV)* (pp. 740-755).

Liu, X., Tanaka, M., & Okutomi, M. (2012). Noise Level Estimation Using Weak Textured Patches of a Single Noisy Image. In *IEEE International Conference on Image Processing (ICIP)* (pp. 665-668).

Loh, Y. P., & Chan, C. S. (2019). Getting to Know Low-light Images with the Exclusively Dark Dataset. *Computer Vision and Image Understanding, 178*, 30-42.

Lore, K. G., Akintayo, A., & Sarkar, S. (2017). LLNet: A Deep Autoencoder Approach to Natural Low-light Image Enhancement . *Pattern Recogntion, 61*, 650-662.

Lv, F., Li, Y., & Lu, F. (2020). Attention Guided Low-light Image Enhancement with a Large Scale Low-light Simulation Dataset. *arXiv:1908.00682*.

Lv, F., Lu, F., Wu, J., & Lim, C. (2018). MBLLEN: Low-light Image/Video Enhancement Using CNNs. In *British Machine Vision Conference (BMVC)* (pp. 220).

McGill, M. (1983). *Introduction to Modern Information Retrieval.* NewYork: McGrawHill.

Ma, K., Duanmu, Z., Wu, Q., Wang, Z., Yong, H., Li, H., & Zhang, L. (2017). Waterloo Exploration Database: New Challenges for Image Quality Assessment Models. *IEEE Transactions on Image Processing, 26*(2), 1004-1016.

Meng, Y., Kong, D., Zhu, Z., & Zhao, Y. (2019). From Night to Day: GANs Based Low Quality Image Enhancement. *Neural Processing Letters*, *50*(1), 799-814.

Milford, M. J., & Wyeth, G. F. (2012). SeqSLAM: Visual Route-based Navigation for Sunny Summer Days and Stormy Winter Nights. In *IEEE International Conference on Robotics and Automation (ICRA)* (pp. 1643-1649).

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., & Terzopoulos, D. (2020). Image Segmentation Using Deep Learning: A Survey. *arXiv:2001.05566*.

Mittal, A., Moorthy, A. K., & Bovik, A. C. (2012). No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Transactions on Image Processing, 21*(12), 4695-4708.

Mittal, A., Soundararajan, R., & Alan C. Bovik. (2013). Making a 'Completely Blind' Image Quality Analyzer. *IEEE Signal Processing Letters, 20*(3), 209-212.

Murray, N., Marchesotti, L., & Perronnin, F. (2012). AVA: A Large-scale Database for Aesthetic Visual Analysis. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2408-2415).

Odena, A., Dumoulin, V., & Olah, C. (2016). Deconvolution and Checkerboard Artifacts. *Distill*, 1(10).

Pisano, E. D., Zong, S., Hemminger, B. M., DeLuca, M., Johnston, R. E., Muller, K., et al. (1998). Contrast Limited Adaptive Histogram Equalization Image Processing to Improve the Detection of Simulated Spiculations in Dense Mammograms. *Journal of Digital Imaging, 11*(4), 193-200.

Rahman, Z., Jobson, D. J., & Woodell, G. A. (1996). Multi-scale Retinex for Color Image Enhancement. In *IEEE International Conference on Image Processing (ICIP)* (pp. 1003-1006).

Ranzato, M., Poultney, C., Chopra, S., & LeCun, Y. (2006). Efficient Learning of Sparse Representations with an Energy-based Model. In *Neural Information Processing Systems (NIPS)* (pp. 1137-1144).

Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. *arXiv:1804.02767*.

Ren, W., Liu, S., Ma, L., Xu, Q., & Xu, X. (2019). Low-Light Image Enhancement via a Deep Hybrid Network. *IEEE Transactions on Image Processing, 28*(9), 4364-4375.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv:1505.04597*.

Salem, C., Azar, D., & Tokajian, S. (2018). An Image Processing and Genetic Algorithm-Based Approach for the Detection of Melanoma in Patients. *Methods of Information in Medicine*, *57*(1), 74-80 .

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4510-4520).

Schaefer, G., & Stich, M. (2003). UCID: An Uncompressed Color Image Database. In *Storage and Retrieval Methods and Applications for Multimedia* (pp. 472-480).

Schölkopf, B., Smola, A. J., Williamson, R. C., & Bartlett, P. L. (2000). New Support Vector Algorithms. *Neural Computation* , *12*(5), 1207-1245.

Schwartz, E., Giryes, R., & Bronstein, A. M. (2019). DeepISP: Toward Learning an End-to-End Image Processing Pipeline. *IEEE Transactions on Image Processing*, *28*(2), 912-923.

Sheikh, H. R., & Bovik, A. C. (2006). Image Information and Visual Quality. *IEEE Transactions on Image Processing*, *15*(2), 430-444.

Shen, L., Yue, Z., Feng, F., Chen, Q., Liu, S., & Ma, J. (2017). MSR-net: Low-light Image Enhancement Using Deep Convolutional Network. *arXiv:1711.02488*.

Shin, Y.-G., Sagong, M.-C., Yeo, Y.-J., & Ko, S.-J. (2018). Adversarial Context Aggregation Network for Low-Light Image Enhancement. In *Digital Image Computing: Techniques and Applications (DICTA)* (pp. 1-5).

Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556*.

Sun, L., & Hays, J. (2012). Super-resolution from Internet-scale Scene Matching. In *IEEE International Conference on Computational Photography (ICCP)* (pp. 1-12).

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2818-2826).

Talebi, H., & Milanfar, P. (2018). Nima: Neural Image Assessment. *IEEE Transactions on Image Processing*, *27*(8), 3998–4011.

Tao, L., Zhu, C., Xiang, G., Li, Y., Jia, H., & Xie, X. (2017). LLCNN: A Convolutional Neural Network for Low-light Image Enhancement. In *IEEE Visual Communications and Image Processing (VCIP)* (pp. 1-4).

VidalMata, R. G., Banerjee, S., RichardWebster, B., Albright, M., Davalos, P., McCloskey, S., et al. (2020). Bridging the Gap Between Computational

Photography and Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1-1.

Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P.-A. (2008). Extracting and Composing Robust Features with Denoising Autoencoders. In *International Conference on Machine Learning (ICML)* (pp. 1096–1103).

Wang, J., Tan, W., Niu, X., & Yan, B. (2019). RDGAN: Retinex Decomposition Based Adversarial Learning for Low-Light Enhancement. In *IEEE International Conference on* Multimedia *and Expo* (*ICME)* (pp. 1186-1191).

Wang, L., Fu, G., Jiang, Z., Ju, G., & Men, A. (2019). Low-Light Image Enhancement with Attention and Multi-level Feature Fusion. In *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* (pp. 276-281).

Wang, R., Zhang, Q., Fu, C.-W., Shen, X., Zheng, W.-S., & Jia, J. (2019). Underexposed Photo Enhancement Using Deep Illumination Estimation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 6842-6850).

Wang, S., Zheng, J., Hu, H.-M., & Li, B. (2013). Naturalness Preserved Enhancement Algorithm for Non-Uniform Illumination Images. *IEEE Transactions on Image Processing, 22*(9)*, 3538-3548.*

Wang, W., Wei, C., Yang, W., & Liu, J. (2018). GLADNet: Low-Light Enhancement Network with Global Awareness. In *IEEE International Conference on Automatic Face & Gesture Recognition* (*FG)* (pp. 751-755).

Wang, Y., Chen, Q., & Zhang, B. (1999). Image Enhancement Based on Equal Area Dualistic Sub-image Histogram Equalization Method. *IEEE Transactions on Consumer Electronics, 45*(1), 68-75.

Wang, Z., & Li, Q. (2011). Information Content Weighting for Perceptual Image Quality Assessment. *IEEE Transactions on Image Processing, 20*(5), 1185-1198.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. (2004). Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing, 13*(4), 600–612.

Wang, Z., Simoncelli, E. P., Bovik, A.C. (2003). Multi-Scale Structural Similarity for Image Quality Assessment. In *IEEE Asilomar Conference on Signals, Systems, and Computers* (pp. 1398-1402).

Wei Liu, D. A., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). SSD: Single Shot MultiBox Detector. In *European Conference on Computer Vision (ECCV)* (pp. 21-37).

Wei, C., Wang, W., Yang, W., & Liu, J. (2018). Deep Retinex Decomposition for Low-Light Enhancement. In *British Machine Vision Conference (BMVC)* (pp. 155).

Xiang, Y., Fu, Y., Zhang, L., & Huang, H. (2019). An Effective Network with ConvLSTM for Low-Light Image Enhancement. In *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)* (pp. 221-233).

Xu, W., Lee, M., Zhang, Y., You, J., Suk, S., & Choi, J.-y. (2018). Deep Residual Convolutional Network for Natural Image Denoising and Brightness Enhancement. In *International Conference on Platform Technology and Service (PlatCon)* (pp. 1-6).

Yang, W., Yuan, Y., Ren, W., Liu, J., Scheirer, W. J., Wang, Z., & Zhang, T. (2020). Advancing Image Understanding in Poor Visibility Environments: A Collective Benchmark Study. *IEEE Transactions on Image Processing, 29,* 5737-5752.

Yangming, S., Xiaopo, W., & Ming, Z. (2019). Low-light Image Enhancement Algorithm Based on Retinex and Generative Adversarial Network. *arXiv:1906.06027*.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. In *European Conference on Computer Vision (ECCV)* (pp. 818-833).

Zhang, L., Zhang, L., Mou, X., & Zhang, D. (2011). Fsim: A Feature Similarity Index for Image Quality Assessment. *IEEE Transactions on Image Processing, 20*(8), 2378-2386.

Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 586-595).

Zhang, Y., Di, X., Zhang, B., & Wang, C. (2020). Self-supervised Image Enhancement Network: Training with Low-light Images Only. *arXiv:2002.11300*.

Zhang, Y., Zhang, I., & Guo, X. (2019). Kindling the Darkness: A Practical Low-light Image Enhancer. In *ACM International Conference on Multimedia (ACM MM)* (pp. 1632–1640).

Zhi, N., Mao, S., & Li, M. (2018). An Enhancement Algorithm for Coal Mine Low Illumination Images Based on Bi-Gamma Function. *Journal of Liaoning Technical University , 37*(1), 191-197.