Joint image enhancement and localization framework for vehicle model recognition in the presence of non-uniform lighting conditions

Kezebou, Landry, Oludare, Victor, Panetta, Karen, Agaian, Sos

Landry Kezebou, Victor Oludare, Karen Panetta, Sos Agaian, "Joint image enhancement and localization framework for vehicle model recognition in the presence of non-uniform lighting conditions," Proc. SPIE 11734, Multimodal Image Exploitation and Learning 2021, 117340Q (12 April 2021); doi: 10.1117/12.2586036

SPIE.

Event: SPIE Defense + Commercial Sensing, 2021, Online Only

Joint Image Enhancement and Localization Framework for Vehicle Model Recognition in the Presence of Non-Uniform Lighting Conditions

Landry Kezebou[1], Victor Oludare[1], Karen Panetta[1], Sos Agaian[2]
[1]Tufts University, Medford, MA, 02155
[2]City University of New York, New York, NY 100161

ABSTRACT

Recognizing the model of a vehicle in natural scene images is an important and challenging task for real-life applications**.** Current methods perform well under controlled conditions, such as frontal and horizontal view-angles or under optimal lighting conditions. Nevertheless, their performance decreases significantly in an unconstrained environment, that may include extreme darkness or over illuminated conditions. Other challenges to recognition systems include input images displaying very low visual quality or considerably low exposure levels. This paper strives to improve vehicle model recognition accuracy in dark scenes by using a deep neural network model. To boost the recognition performance of vehicle models, the approach performs joint enhancement and localization of vehicles for non-uniform-lighting conditions. Experimental results on several public datasets demonstrate the generality and robustness of our framework. It improves vehicle detection rate under poor lighting conditions, localizes objects of interest, and yields better vehicle model recognition accuracy on low-quality input image data.

Grants: This work is supported by the US Department of Transportation, Federal Highway Administration (FHWA), grant contract: **693JJ320C000023**

***Keywords*—Image enhancement, vehicle model and make recognition, object detection, deep learning, object recognition in the dark, vehicle detection and recognition in the dark.**

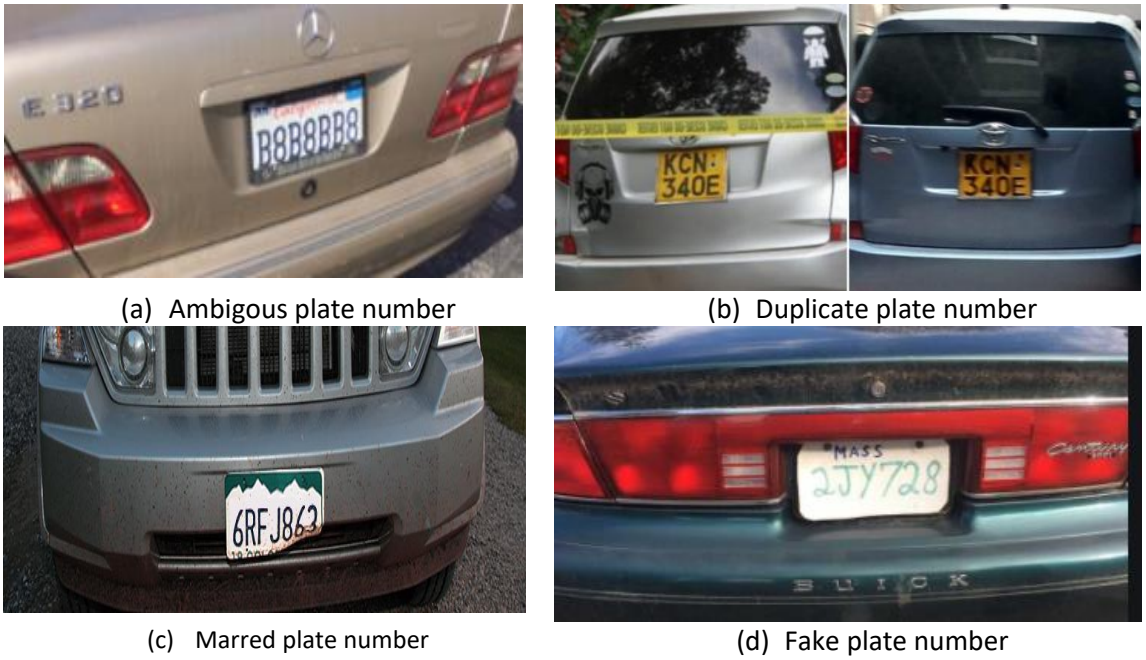
1. INTRODUCTION

The recent revolutionary advancement in the field of deep learning and computer vision in particular, has had a transformational impact on intelligent transportation systems. Vehicle plate number identification and recognition systems [1], [2] , vehicle detection and classification [3]–[5], vehicle reidentification [6], [7] , incidents detection systems [8], [9], and spatio-temporal tracking systems [10], are some of the numerous computer vision technologies that find direct application in smart transportation systems.

As the sheer volume of traffic surveillance cameras continues to grow, monitoring the vast amounts of live feeds and producing timely actionable responses to incidents becomes intractable. Thus, it has become imperative to develop assistive technologies to make transportation systems much smarter with capabilities of integrating a wider range of applications including advanced security.

The ability to recognize the make and model of a vehicle is vital in providing an extra layer of security in surveillance applications. Consider for example a plate number verification system that is capable of efficiently identifying and recognizing a plate number attached to a vehicle and effectively retrieving the residential address and other essential details of the owner to which the vehicle is registered. Such a system will fail to identify cases in which license plates are swapped from one vehicle to another to perpetuate crimes or may simply fail if images are of low-quality under non-uniform lighting conditions. There has been several reported cases of truck drivers using several tricks to instantly swap or partially occlude their plate number to avoid paying toll fees, and then swap back to the original plate once they've gone pass the toll gate [11], [12]. Automatically recognizing the make and model of a particular vehicle along with its associated license plate will help address such loopholes while altogether providing an additional layer of security.

Multimodal Image Exploitation and Learning 2021, edited by Sos S. Agaian, Vijayan K. Asari, Stephen P. DelMarco, Sabah A. Jassim, Proc. of SPIE Vol. 11734, 117340Q
 · doi: 10.1117/12.2586036



(a) Ambigous plate number (b) Duplicate plate number

(c) Marred plate number (d) Fake plate number

***Figure 1**: Samples of altered plate numbers which elude plate number identification and recognition systems.*

State-of-the-art object detection algorithms such as YOLOv3 [13], SSD [14], and Fast-RCNN [15], have demonstrated very high accuracy in object detection tasks, specifically vehicle detection and categorization (car, buses, trucks, motorcycle, bicycle). However, detection accuracy of these models is significantly impaired in presence of non-uniform illumination. This detection impairment in turn reduces the efficacity of vehicle recognition models. Furthermore, most existing vehicle recognition models are designed to work well under controlled environments and optimal lighting conditions, as such they suffer considerable performance degradation in presence of non-ideal conditions such as low-light, hazy, and foggy weather.

To address these challenges, a deep neural network-based framework for joint enhancement and localization for real time vehicle make and model recognition was developed. The proposed framework incorporates from end-to-end: **a)** a lightweight Non-Uniform Light Enhancement via Deep Curve Estimation (NULE-DCE) model for automatically transforming the input images/frames into their well illuminated version with fine-grain details, **b)** Yolov5 object detector for automatically detecting and extracting patches of objects of interest, and **c)** a trained Resnet model for recognizing the make and model of the detect vehicles.

The notable contributions of this research endeavor can be summarized as follows:

a) An efficient lightweight model for non-uniform light enhancement through well-defined and pretrained light enhancement curves technique is proposed. The proposed NULE-DCE model is a hybrid fusion of Zero-DCE [16] and DUAL [17] light enhancement models. Unlike existing low light enhancement algorithms, the model considers the case of underexposed (low-light) and overexposed image enhancement while not requiring paired data for training. The model is also very lightweight and can run at very high frame rate making it suitable for a framework with computationally intensive object detection and object recognition models.
b) We proposed a trained ResNet 50 model with high accuracy for vehicle make and model recognition.
c) Lastly, an end-to-end framework for automatically localizing and recognizing vehicles make and model in the presence of non-uniform lighting is developed.

Computer simulations demonstrate that the proposed lightweight automatic non-uniform light enhancement method outperforms current state-of-the-art. Additionally, we demonstrate that the proposed end-to-end framework substantially boosts vehicle detection and recognition accuracy.

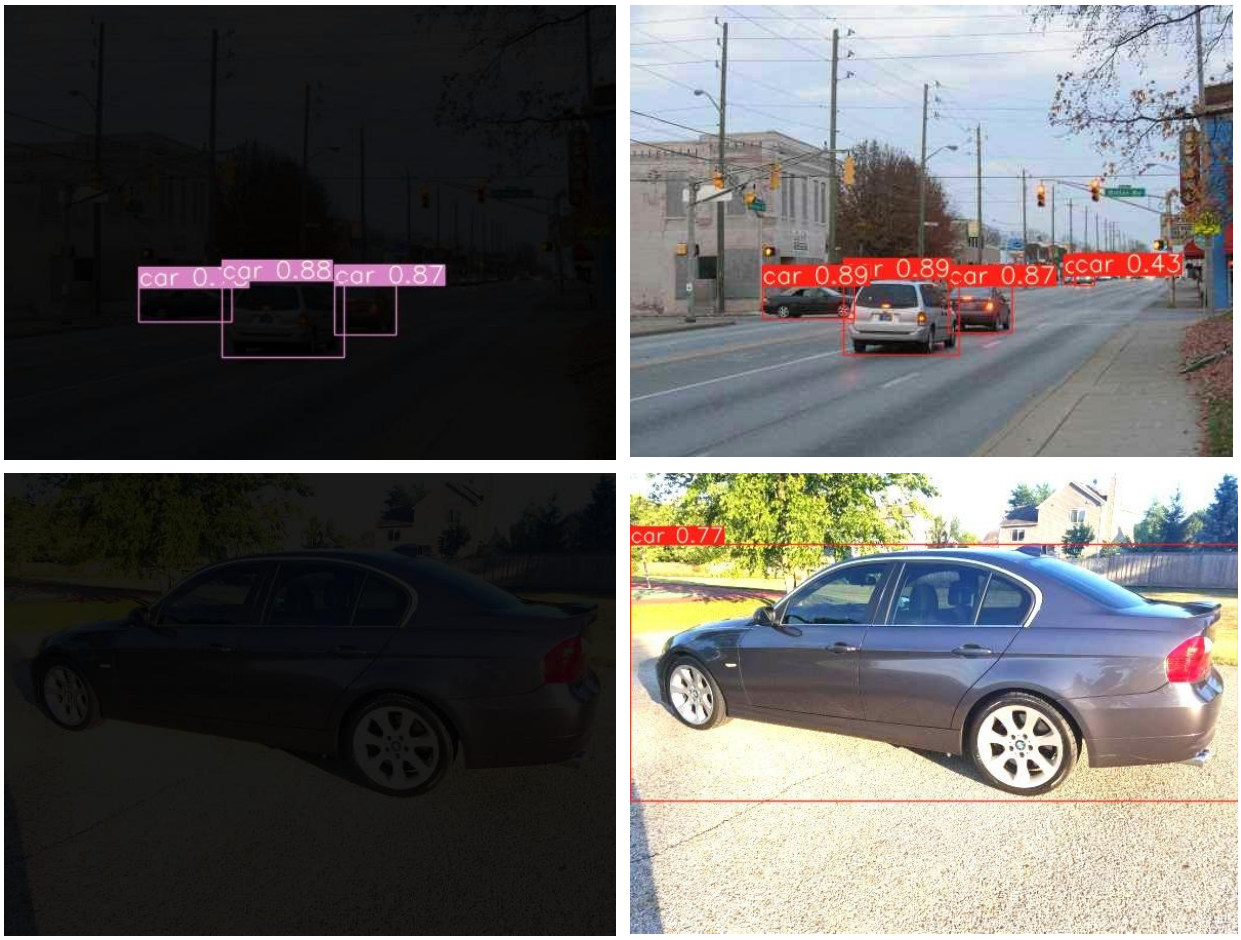


***Figure 2**: Vehicle detection in presence of lowlight using YOLOv5 model. Images on the right are enhanced versions of the lowlight images on the left using the proposed NULE-DCE model. It can be seen that the detection model fails to detect some of the vehicles of interest in the dark.*

2. RELATED WORK

Fine-grained vehicle recognition still remains a challenge despite continuous effort to push the accuracy boundary of existing models on such tasks. The public release of fine-grain vehicle classification datasets such as VMMR [18], Stanford Dataset [19], and CompCars [20], has spurred increasing interest in the computer vision community to develop sophisticated models for performing fine-grain vehicle classification. However, the nature of these training datasets, which only consider specific views of vehicles (frontal view, rear view, side view) and under controlled lighting conditions, has skewed the strength of existing vehicle fine-grained recognition models on the type of data trained on. As a result of these biases, such recognition models do not account for some of the real-world scenarios such as nighttime/darkness, hazy or foggy weather conditions. Hence, there is noticeable decline in performance when these models are tested on input data that simulate other real-world conditions such as non-uniform lighting.

A. Fine-Grained Vehicle Classification/Recognition

Earlier classical systems favored using popular feature extraction methods such as SIFT, SURF and HOG, coupled with a simple classifier such as SVM, Nearest Neighbor and Random Forest [21]–[23]. However, these methods rely on the strength of the feature extractor used and fail to consider many real-world scenarios.

Wang et al. [24] proposed a vehicle recognition algorithm based on a multiple feature subspace and transfer learning. The proposed model is divided into a) an offline training pipeline based on deep belief network (DBN) which uses

multiple restricted Boltzmann machines to extract relevant features; and b) an online transfer learning algorithm which generates labels for new samples. The pipeline is then retrained end-to-end for improved efficacity. Manzoor et al. [25] introduced a Vehicle Make and Model recognition system based on engineered local and global feature vectors used to train a Random Forest classifier. However, performance of these models is suboptimal on large-scale datasets.

Other methods have leveraged more advanced deep neural network techniques such as ResNet, VGG and MobileNet models for training more accurate vehicle recognition models. Ma et al. [26] recently proposed an AI based visual attention model for vehicle make and model recognition which is based on Recurrent Attention Unit (RAU) concatenated with the CNN layers of the Resnet101 architecture. The model proposed by Ma et al achieves state-of-the-art accuracy on both the Stanford and CompCars datasets. Fang et al. [27] proposed a deep neural network based fine-grained vehicle recognition model by automatically extracting local and global features and cues to distinctively recognize vehicle from 281 classes with high degree of accuracy. The proposed method is able to first identify what subtle parts of the image contain the most descriptive cues from which hierarchical feature maps are extracted via multi-layer feed-forward CNN architecture network. Lee et al. [28] proposed an incremental improvement over Vanilla SqueezeNet using bypass connections to extract high dimensional features from input images. Principal Component Analysis is then applied to reduce the features dimensionality upon which k-means clustering is used to cluster vehicles of similar classes.

Although some of the aforementioned systems for vehicle make and model recognition have achieved good accuracy on benchmark datasets, these methods fail to consider certain non-ideal real-world conditions. In this paper, we focus our attention on addressing the problem of vehicle recognition in presence of non-uniform lighting conditions to help overcome some of the drawbacks of existing models. To this end, we develop a real-time joint enhancement and localization framework upon which we build a robust vehicle make and model recognition system suitable for both optimal and suboptimal lighting conditions.

B. Lowlight Image Enhancement

Numerous research work has been done to address the problem of lowlight image enhancement. From classical low-level image processing methods [17], [29] to neural network models [16], [30]–[33], to more recent generative adversarial network models [34], [35].

Guo et al. [29] proposed LIME, a very effective and considerably simple model for lowlight image enhancement using well coined low-level image processing techniques to articulate a generalizable mathematical formula for automatically estimating illumination maps for optimal image enhancement. Guo leverages on the Retinex modeling theory to extract a first level illumination map by finding the maximum intensities of pixels across the RGB channels. This step is followed by an Augmented Lagrangian Multiplier (ALM) algorithm which exploits the structure of the illumination map to derive a more refined map without color saturation. Zhang et al. [30] proposed an alternative approach to LIME for practical lowlight enhancement called KinD. Similar to the LIME method, KinD also uses Retinex theory to formulate the auto-enhancement objective. But unlike LIME, KinD enhances the image from two decoupled subspaces: an illumination component used for auto-adjusting the lighting exposure in the image; and the reflectance component responsible for degradation removal. However, these methods sometimes result in color distortions in output images and perform sub-optimally on overexposed images.

Liang et al. [33] introduced a Deep Bilateral Retinex model which is a deep neural network model that learns to predict the pixel-wise illumination and noise maps in a bilateral space to produce an equivalent enhanced image output by exploiting the inherent connections between the spatially-varying noise and the illumination layers. Liang accentuates the focus on handling measurement of noise in the formulation of the training objective of the network. Wei et al. [31] proposed a Deep-Retinex decomposition model for lowlight enhancement. Deep-Retinex consists of three steps trainable end-to-end. The first step decomposes the input lowlight image into illumination and reflectance map via a Decom-Net model. This is followed by an encoder-decoder network for adjusting the illumination map. Lastly, a multi-scale concatenation is used to enforce local and global consistency of pixels color and contextual information. The methods are computationally expensive and require paired training data. Furthermore, some patches in the image appear under or over enhanced.

The EnlightenGAN model proposed by Jiang et al. [34] and the AGCRN model proposed by Oludare et al. [35] leverage on the concept of Generative Adversarial Network to train efficient models that learn to automatically map input images from the lowlight domain to its equivalent well illuminated domain. EnlightenGAN is trained in an unsupervised manner such that lowlight data does not require a direct pair. Instead, another pool of well-enhanced images is used to formulate the target domain from which the network models an equivalent hyperspace representation. This adequately translates the image from the lowlight domain to the illuminated domain. AGCRN on the other hand is trained on paired lowlight data which allows the model to learn a more optimal model for translating images from the lowlight domain to the well-illuminated domain. AGCRN provides several advantages over EnlightenGAN in terms of network architecture and loss function which helps achieve superior results on benchmark datasets. However, these GAN-Based methods can be more computationally expensive than the classical approach and also demands vast amounts of training data. Additionally, it overlooks the reverse problem which consists of toning down overexposed/over-illuminated images.

Zhang et al. [17] also introduced a robust correction exposure model via dual illumination estimation (DUAL). This method starts by extracting the forward and reverse illumination map from which intermediate enhancement results of the input image and its inverted version are generated. The final enhanced image is generated by fusing the input image with the intermediate underexposed and overexposed corrected images from the input and its inverse, respectively. Hence achieving dual illumination correction. Guo et al. [16] proposed Zero-DCE, a deep curve estimation network model for auto enhancing both lowlight and overexposure images in a single network by modeling light enhancement curves without requiring reference or paired data. The major advantage of Zero-DCE it that it is very lightweight and high speed (capable of operating at up to 500 Fps) and also very efficient for both underexposed and overexposed image enhancement, while yielding better or comparable results to other state-of-the-art models. Zero-DCE feeds the input image to a shallow network of 6 convolutional layers to extract pixel wise enhancement maps which are then used to iteratively enhance the images using fine-tuned light enhancement curves. Although these methods are capable of auto enhancing both underexposed (lowlight) and overexposed images, they seem to underperform on extremely dark images, whereas the overexposure correction leaves a semblance of haze.

The proposed NULE-DCE model addresses the shortcomings of these methods by capitalizing on the strengths of Retinex modeling theory used in the DUAL (dual illumination net) [17] and the LIME [29] methods; and the learnable fast speed light enhancement curves of Zero-DCE model [16], to produce high quality visual enhancement of both underexposed and overexposed images. The fast speed and efficacy of the proposed model makes it very suitable for the joint enhancement and localization framework for vehicle model and make recognition.

3. METHODOLOGY

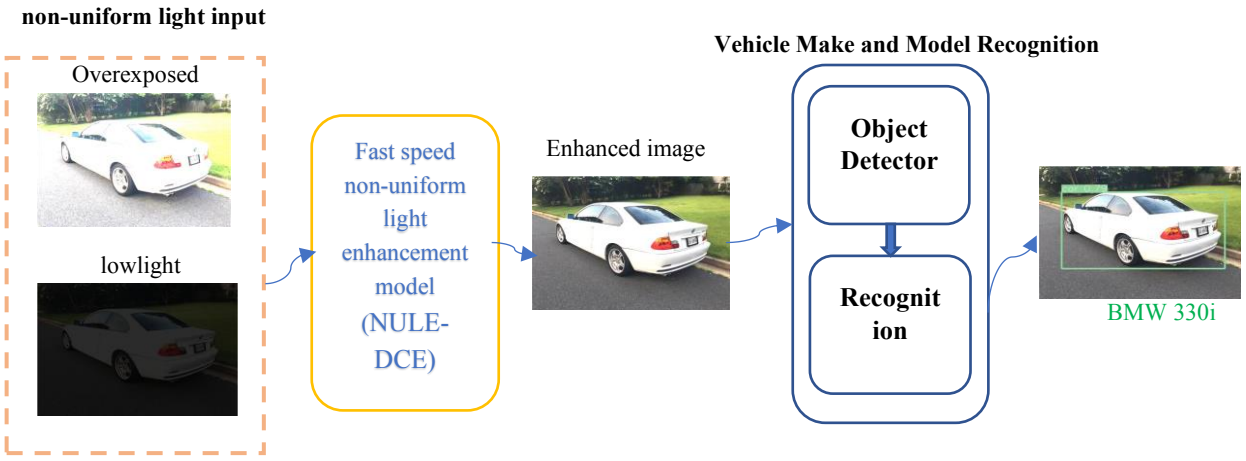


***Figure 3**: Framework for Vehicle Make and Model Recognition in presence of non-uniform lighting. The Light enhancement module is capable of performing high speed input preprocessing to reveal objects of interest and boost performance of object detection and subsequent recognition models. The overall framework achieves high accuracy for vehicle recognition in low-light.*

The proposed joint-enhancement and localization framework for vehicle model and make recognition is a fluid end-to-end integration of three sub-modules working collaboratively to achieve desired results., including: a) non-uniform light auto-correction and auto-enhancement model (NULE-DCE), b) object detection model, and c) vehicle make and model recognition model. Figure 3 shows the structural architecture of the proposed framework. The individual modules are trained independently and consolidated into a pipeline where they work hand in hand to accurately detect and recognize vehicles by their make and model in presence of non-uniform lighting including lowlight (underexposure) and overexposure. The light enhancement module preprocesses the image by correcting for non-uniform lighting in the input image/video feed to boost the performance of the object detector which demonstrates reduced detection efficiency in presence of bad lighting conditions. Detected objects are subsequently passed through the recognition model for accurate vehicle make and model recognition. The Vehicle make and model recognition is a difficult task given limited availability data per class. Much less data is available with desired poor lighting conditions. As such it is more effective to train the recognition model on well illuminated images, making the enhancement module even more crucial to help the system achieve desired performance on distorted test input.

3.1. No-Reference Non-Uniform Light Enhancement via Deep Curve Estimation (NULE-DCE)

Because the object detection and recognition models are computationally intensive, it is imperative to develop a very lightweight enhancement model which could preprocess the input feed at a very high speed to achieve real-time operations. Figure 4 shows the network model for the proposed lightweight non-uniform light enhancement model.

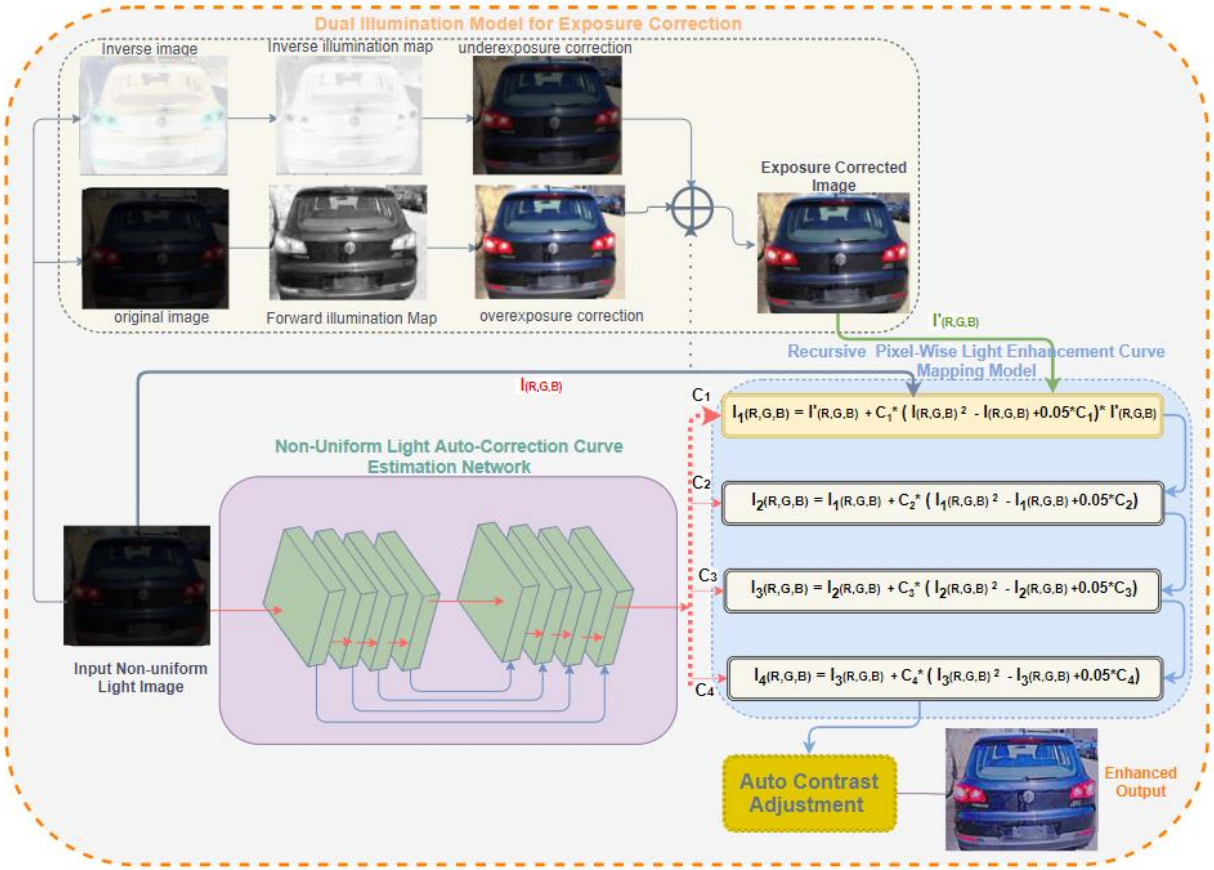


***Figure 4**: Non-Uniform Light Enhancement via Deep Curve Estimation (NULE-DCE) model for improved vehicle localization and recognition. The network is divided into 3 modules, a) a shallow network for predicting light enhancement curve inspired by Guo et al.* [16]*; b) a dual illumination network for simultaneous underexposure and overexposure correction inspired by Zhang et al.* [17]*; and c) a recursive pixel-wise light enhancement curve mapping.*

There are numerous state-of-the-art algorithms for performing low-light enhancement as discussed in section 2. However, these methods present several disadvantages if used in the proposed pipeline, the most important of which being "high inference time" which would constitute a system bottleneck and defeat the intended real-time speed. Furthermore, most algorithms only consider the case of low-light enhancement and neglect overexposure correction.

The model proposed in this work is mainly inspired by the Zero-DCE deep curve estimation for low-light enhancement proposed by Guo et al. [16], and the dual illumination estimation (DUAL) for robust exposure correction proposed by Zhang et al. [17]. Both these methods are lightweight and consider both underexposure and overexposure correction in one network. We combine the strength of both these algorithms to propose a more robust model capable of rendering better enhancement outputs. Essentially, we leverage the fast speed light enhancement curve estimation via shallow network as championed by Guo, and we make our recursive pixel-wise mapping network to benefit from the first level enhancement output of the dual illumination exposure correction network which helps stabilize the training and improve enhancement curve via backpropagation. Evaluation on numerous datasets and on vehicle recognition dataset as presented in the result section show the strength of the proposed methods as compared to other state-of-the-art.

Non-Uniform Light Auto-Correction Curve Estimation Network

The curve estimation network is essentially just a very shallow network of 9 blocks of 2D convolutional each followed by a **ReLU** activation function. Layers have kernel sizes of 32 and stride of 3x3. The output of the network is passed through a **tanh** activation function to generate final pixel-wise light enhancement curve parameters used for automatically correcting the input images. Feature maps from lower convolutional layers are concatenated to the outputs of upper layers to ensure local and global pixel color consistency in the enhanced outputs using predicted light enhancement curves. The proposed method is also a zero-reference based method which does not require paired data for training. As such, the training set incorporates a mixture of images of same scene at multiple exposure level. Sample such images are shown in Figure 5 which are drawn from original multi-exposure images used in the LVZ-HDR dataset [36]. Training is very fast and can be completed in 30 to 45 min, and only about 50 epochs are need.

***Figure 5**: Sample data from the training set showing multiple exposure level for same scene from which the network learns to estimate the light enhancement curves to correct for both lowlight and overexposure.*

Dual Illumination Model for Exposure Correction

Zhang et al. proposes a simple dual illumination exposure correction by formulating the joint under and over exposure correction as a dual illumination estimation with adequate fusion of the input image and its inverted version. Leveraging on this observation we theorize that such single step enhancement could help boost the performance of the recursive pixel-wise light enhancement curve mapping which maps the light enhancement curves parameters generated by the convolutional blocks to the input image.

Given and input image $I_{(R,GB)}$, the inverse image is first generated as follows:

$$I_{inv_{(R,G,B)}} = 1 - I_{(R,GB)} \qquad (1)$$

Next, the forward and backward illumination maps are extracted from the original input image and its inverted version, respectively. Guo et al. in the LIME paper [29] suggests that the illumination map for an image $I_{(R,GB)}$ can be obtained by extracting the maximum of the R, G, B channel for each pixel. The illumination for a pixel x can be defined as:

$$L(x) = \max_{c \in \{R,G,B\}} I^{c}_{(R,GB)} \qquad (2)$$

Likewise, the inverse illumination map can be computed as $L_{inv}(x) = \max_{c \in \{R,G,B\}} I_{inv}{}^{c}_{(R,GB)}$

The intermediate under/over exposure images are generated using the forward and backward illumination maps, respectively. The final exposure corrected image is the result of fusing the original input image with the intermediate under and over exposure corrected images.

Figure 6: closer look at how the dual illumination module helps correct exposure in the input image.

Recursive Pixel-Wise Light Enhancement Curves Mapping

Through experimentation, we observe that fusing the exposure corrected output of the dual illumination enhancement module to the original input and then mapping the corresponding curves with Light Enhancement Equations helps the learning network to learn better enhancement curves for generating output with finer details and more consistent structural and color information both on local and global patches.

The output feature maps from the convolutional network module are used as parameters of the enhancement curves which are mapped to the input image in a pixel-wise fashion using the Light Enhancement Equations as follows:

$$I_{1_{(R,G,B)}} = I'_{(R,GB)} + C_1 * \left(I^2_{(R,G,B)} - I_{(R,G,B)} + 0.05 * C_1\right) * I'_{(R,G,B)} \qquad (3)$$

Where C_1, $I_{(R,GB)}$, $and\ I'_{(R,GB)}$ represent the output pixel-wise illumination curve parameters from the convolutional network block, the original non-uniform lighting input, and the intermediate enhancement from the dual illumination module.

The subsequent Light Enhancement curves are applied through the following equation:

$$I_{k_{(R,G,B)}} = I_{k-1_{(R,G,B)}} + C_k * \left(I_{k-1}{}^{2}_{(R,G,B)} - I_{k-1_{(R,G,B)}} + 0.05 * C_k\right) * I_{k-1_{(R,G,B)}} \qquad (4)$$

Where $I_{k_{(R,G,B)}}$, $I_{k-1_{(R,G,B)}}$ represent the enhancement output at the k^{th} and $(k-1)^{th}$ step, while C_k represents the curve parameters used for pixel-wise light enhancement at the k^{th} step.

Auto Contrast Correction

To further aid the training process of the curve estimation network, we introduce an automatic contrast correction component which helps control the contrast in the final output image and ensure global and local contrast consistency through backpropagation.

The contrast correction is applied to the R, G, B channels of an image $I_{(R,G,B)}$ using the following formula where R, G, B pixels values of the image to contrast-correct have been scaled to [0, 1] range:

$$I_{con_{(R,G,B)}} = F * (I_{(R,G,B)} - 0.502) + 0.502 \qquad (5)$$

Where F represents the contrast correction factor calculated for a given desired contrast level $C \in [0, 1]$ as follows:

$$F = \frac{1.0156(C + 1)}{(1.0156 - C)} \qquad (6)$$

After contrast correction using Equation (5), output pixels of the resultant contrast-corrected image are truncated to the [0, 1] range according to the following formula:

$$I(x) = \begin{cases} I(x) \ if \ 0 \le I(x) \le 1 \\ 0 \ if \ I(x) < 0 \\ 255 \ if \ I(x) > 1 \end{cases} \qquad (7)$$

Based on our experiment a great value for C can be chosen in the interval [0.12, 0.196].
The Auto Contrast correction Equations are inspired from:https://www.dfstudios.co.uk/articles/programming/image-programming-algorithms/image-processing-algorithms-part-5-contrast-adjustment/

3.2. Loss function

The deep curve estimation network is trained on unpaired data in which images are presented at multiple exposure level. For fast convergence and adequate deep curve learning, we adopt the four loss components proposed by Guo et al. [16] , including Exposure Control Loss, Color constancy Loss, Spatial Constancy loss, and Illumination smoothness loss. In Addition, we introduce an Auto-Contrast correction loss to allow for contrast auto-correction to backpropagate through the network and improve the learnable parameters. The overall training objective for the proposed model is defined by equation 8 where: $\boldsymbol{\alpha_1 = 5}$, $\boldsymbol{\alpha_2 = 1}$, $\boldsymbol{\alpha_3 = 200.}$

$$\boldsymbol{L = L_{spa} + L_{exp} + \alpha_1 L_{col} + \alpha_2 L_{con} \ \alpha_3 L_{tv_A}} \qquad (\mathbf{8})$$

The exposure control loss helps control the over and under exposure by measuring the distance between the well-exposedness of the gray world $\boldsymbol{E}$ (typically $\boldsymbol{E = 0.6}$) to the mean intensity of pixels values in a local neighborhood (typically 16x16 patch size in this case). Given K non-overlapping pixels in a 16x16 local patch, with average pixel intensity given as X_{mean} [16].

$$\boldsymbol{L_{exp} = \frac{1}{K} \sum_{i}^{K} |X_i - E|} \qquad (\mathbf{9})$$

The illumination smoothness loss which helps preserve the relation between neighboring pixels is defined in [16] as:

$$\boldsymbol{L_{tv_A} = \frac{1}{N} \sum_i^N \sum_{c \in \xi} (|\nabla_x A_i^c| + |\nabla_y A_i^c|)^2} \ , \quad \boldsymbol{where \ \xi = \{R,G,B\}, \ \nabla_x, \nabla_y \ are \ horizontal \ and \ vertial \ gradients} \ ) \qquad (\mathbf{10})$$

The color constancy and spatial constancy control losses which help maintain local and global color consistency, and adjacent pixels structural consistency, respectively, are defined as follows [16]:

$$\boldsymbol{L_{spa}} = \frac{\mathbf{1}}{\boldsymbol{K}} \sum_{\boldsymbol{i}}^{\boldsymbol{K}} \sum\nolimits_{\boldsymbol{j} \in \boldsymbol{\Omega(i)}} \left(\left|\boldsymbol{X_i} - \boldsymbol{X_j}\right| - \left|\left(\boldsymbol{X'}_{\boldsymbol{i}} - \boldsymbol{X'}_{\boldsymbol{j}}\right)\right|\right)^{\mathbf{2}} \qquad (\mathbf{11})$$

$$\boldsymbol{L_{col}} = \sum_{\forall (\boldsymbol{p},\boldsymbol{q}) \in \boldsymbol{\xi}} (\boldsymbol{Y^p} - \boldsymbol{Y^q})^{\mathbf{2}}, \quad \boldsymbol{\xi} = \{(\boldsymbol{R},\boldsymbol{G}), (\boldsymbol{R},\boldsymbol{B}), (\boldsymbol{G},\boldsymbol{B})\} \qquad (\mathbf{12})$$

where $\boldsymbol{X}, \boldsymbol{X'}, \boldsymbol{and}\ \boldsymbol{K}$ represent the average intensity value in a 4x4 local patch of the original image (x) and enhanced image (X'); and K is the number of $\boldsymbol{\Omega(i)}$ regions. $\boldsymbol{Y^p and\ Y^q}$ are the average pixel intensity in for channel p and q chosen from available channel combinations ($\boldsymbol{\xi} = \{(\boldsymbol{R},\boldsymbol{G}), (\boldsymbol{R},\boldsymbol{B}), (\boldsymbol{G},\boldsymbol{B})\}$).

Finally, the Auto-contrast correction loss is defined as:

$$\boldsymbol{L_{con}} = \frac{\mathbf{1}}{\boldsymbol{K}} \sum_{\boldsymbol{i}}^{\boldsymbol{K}} \sum\nolimits_{\boldsymbol{j} \in \boldsymbol{\Omega(i)}} \left(\left|\boldsymbol{X_i} - \boldsymbol{X_j}\right| - \left|\left(\boldsymbol{L_i} - \boldsymbol{L_j}\right)\right|\right)^{\mathbf{2}} \qquad (\mathbf{13})$$

Equation (13) is similar to the spatial consistency loss in Equation (11), except that L represents the average intensity of 4x4 local patch in the contrast correct images, and X is the average intensity in the local patch of the original image.

3.3. Vehicle Make Recognition Model

The proposed vehicle model and make recognition model is trained on the VMMR dataset [18] which can be downloaded at : https://github.com/faezetta/VMMRdb. The dataset contains 9170 classes and 291,752 images and covers models between the year 1950 to 2016. First, we notice a lot of the classes contain less than 10 images, which is not enough training data per class. For optimal performance, we start by filtering out classes with at least 40 image samples and more. Next, we only consider images for vehicles from year 1995 to 2016 for more relevance, because it is highly unlikely to encounter a vehicle from 1995 and below in today's traffic unless vintage cars which are corner cases. Finally, images of vehicles of the same make and model from different years are merged. This is to increase image samples per class given that vehicles similar make and model share similar features for across make year. By applying these filters, we end up with curated dataset of 450 vehicle make and model classes for a total of 240k images.

Images in the VMMR dataset have noisy background sometimes showing other random vehicles which are not the center focus and not of the same class. Such noisy data can negatively impact the learnable hyperspace parameters of the classifier. To mediate this challenge, we first run the images through an object detector to detect all vehicles in an image, then crop out the detected vehicle with the largest bounding box which is considered the vehicle of the class of interest. The cropped patch is then saved in place of the original image to constitute a more adequate and less noisy training and validation set.

We split the consolidated dataset into 60% training, 20% validation and 20% testing. We train the recognition model using ResNet50. We use pretrained weight of ResNet50 on ImageNet and unfreeze the last 2 layers for fine-tuning the learnable parameters on the new dataset while training the classifier. We used an SGD optimizer with initial learning rate $\boldsymbol{l_r = 0.1}$ **, momentum = 0.9, and weight decay = 0.0001**. Additionally, we use a learning rate scheduling defined as $\boldsymbol{l_r = lr * \left(0.1^{\#epoch//30}\right)}$ with **#epoch//30** being integer division.

The initial Vehicle Make and Model recognition model is independently trained and tested on well-illuminated vehicle images. For recognition under non-uniform light however, the images from the test set are translated into non-uniform lighting using various image processing techniques to generate images with random non-uniform lighting conditions. Detailed results for such experiment are presented in section 4.

4. COMPUTER SIMULATIONS AND EXPERIMENTAL RESULTS

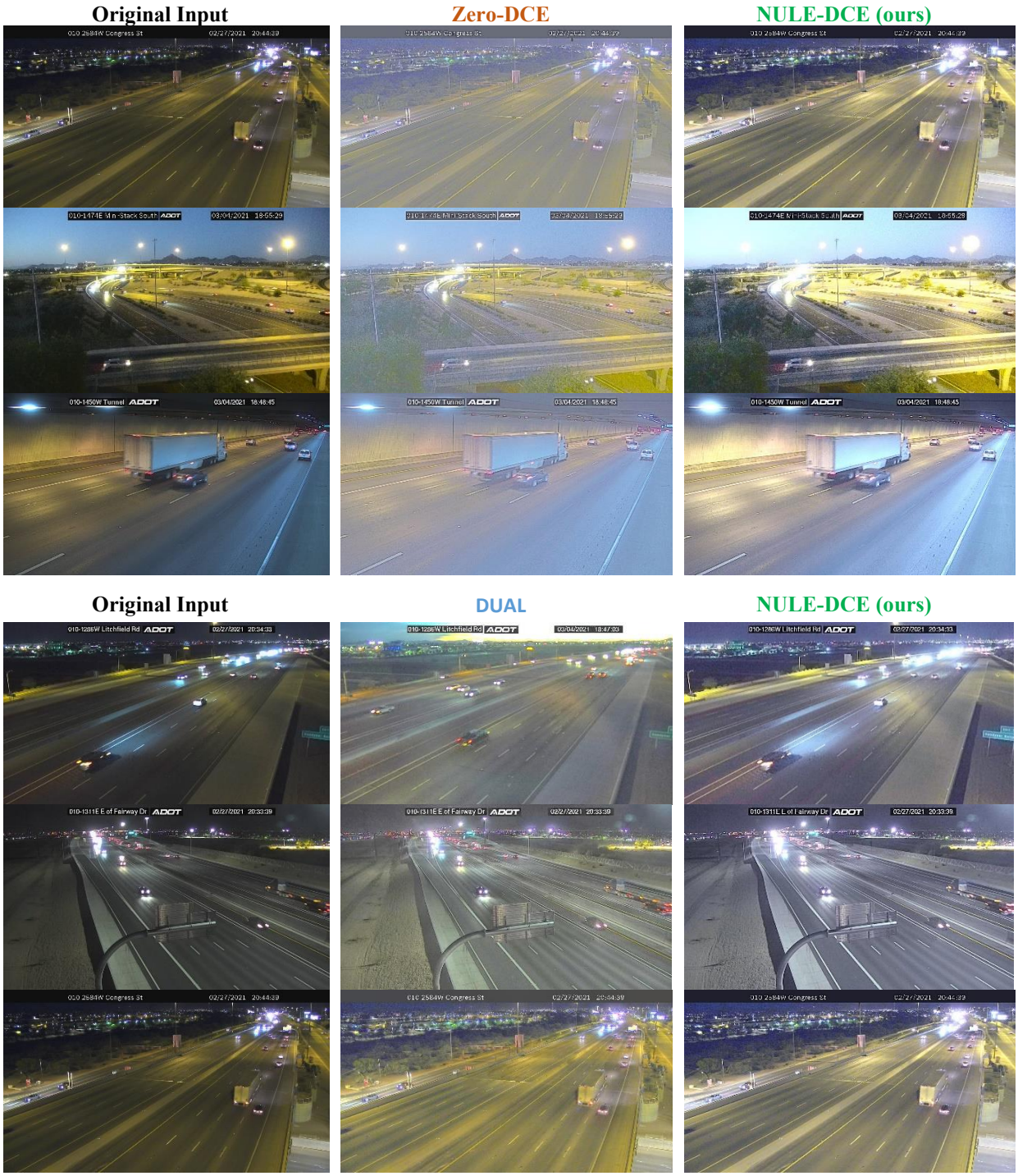


Figure 7: Performance comparison of proposed NULE-DCE (Non-Uniform Light Enhancement via Deep Curve Estimation) against other lightweight under/over exposure correction algorithms, particularly Zero-DCE and DUAL which both inspired NULE-DCE. Test data show here were obtained from live traffic camera feed at night, at the time of writing the paper. (See timestamps)

To demonstrate superior performance of the proposed NULE-DCE enhancement algorithm, we compare the enhanced output of our method against other lightweight over/under exposure correction and enhancement algorithms from which NULE-DCE is inspired. Particularly, we compare the enhancement performance on live traffic data feed (accessed online) with non-uniform lighting. Figure 7 shows visual comparison of NULE-DCE, Zero-DCE and DUAL. It can be seen that NULE-DCE which combines the strengths of Zero-DCE and DUAL to address both their weaknesses, is capable of rendering much better enhanced outputs which look more natural and have better contrast and color consistency at both local and global pixel levels. Additionally, NULE-DCE also inherits the fast speed processing of Zero-DCE, making it very suitable for embedding into a real-time recognition pipeline such as the one being proposed for vehicle make and model recognition.

We further compare the performance of the proposed NULE-DCE against other prominent state-of-the-art low-light enhancement algorithms on several benchmark dataset including LIME dataset, DICM, VV, MEF, NPE, Hongkong dataset, and lowlight dataset as shown in Figure 8. Test data can be found here: https://github.com/VITA-Group/EnlightenGAN. It can be observed that NULE-DCE qualitatively outperforms other state-of-the-art methods on test images across numerous benchmark dataset. Enhanced image outputs from NULE-DCE have better exposure, better color, better contrast both locally and globally, and are more natural looking.

Table 1 also compares the quantitative performances of these algorithms using the Natural Image Quality Evaluator (NIQE).

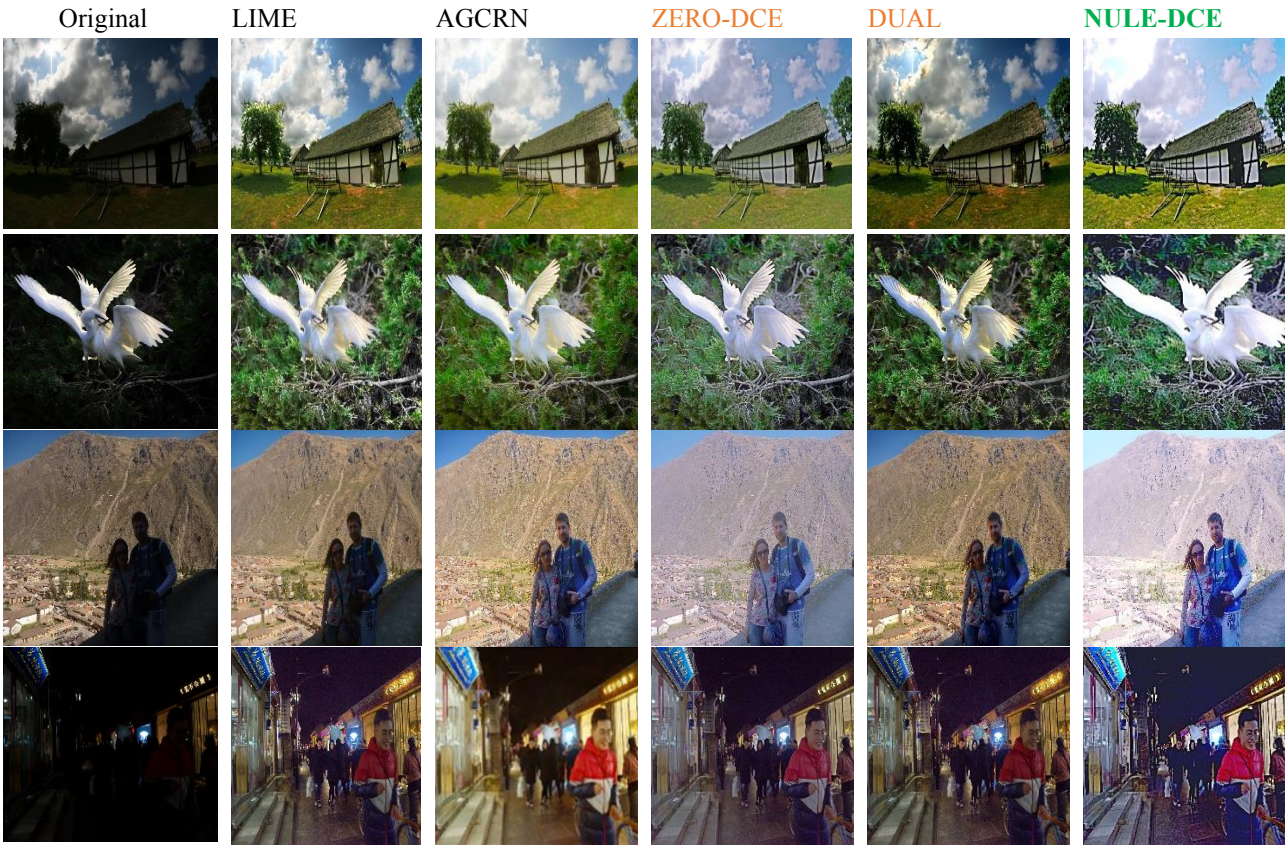


Figure 8: Visual comparison of NULE-DCE output against other state-of-the-art lowlight enhancement methods, on numerous datasets, including VV, LIME, NPE, MEF, and Lowlight

Table 1: quantitative performance of proposed NULE-DCE against current state-of-the-art low-light enhancement methods across several dataset, including LIME, MEF, NPE, VV, DICM and Lowlight

	LIME	**AGCRN**	**Zero-DCE**	**DUAL**	**NULE-DCE**
Lowlight	4.8861	4.8221	5.4559	4.8756	5.4875
LIME	3.8451	3.5053	4.1505	3.8459	4.5283
MEF	3.3155	3.0685	3.4062	3.4544	3.9156
NPE	4.4356	3.8259	4.1054	4.2693	4.2798
Hongkong (traffic dataset)	3.8320	3.8763	3.9098	3.7303	3.7247
VV	2.5376	3.8465	3.2376	2.4725	3.1205
DICM	3.7545	3.5472	3.5528	3.7531	4.0281

For testing the recognition tasks under lowlight conditions, we first artificially generate non-uniform light images from the test set reserved from the VMMR dataset used for training the Vehicle Make and Model Recognition model. Next, we build an end-to-end pipeline for performing a joint- enhancement, detection, and recognition in one go. The pipeline includes the high-speed NULE-DCE image enhancer, the Fast speed YoloV5 object detection, and the ResNet50-based Vehicle Make and Model recognition model.

Sample recognition output based on non-uniform light input is shown in Figure 9.

Mazda CX9 | Vokswagen Tiguan | Infiniti G37 | BMW X5

Toyota Prius V | Toyota Prius | Acura Integra | Audi S4

Figure 9: Sample Vehicle Make and Model Recognition output in presence of non-uniform lighting using the proposed NULE-DCE as the input image/video feed enhancer. The first and third rows show non-uniform light inputs, while the second and forth rows show recognition outputs on the enhanced images.

Table 2: Recognition accuracy of the proposed Vehicle Make and Model on the VMMR dataset. We first test the framework on the original test data extracted from the VMMR dataset which are relatively "well-illuminated images". Next, we generate corresponding non-uniform light test data from the original test data and also generate corresponding lowlight test data using Pix2PixHD [37]*. We test the performance of the same framework on such lowlight and non-uniform light test set without enhancement, and we record the respective test accuracies. Finally, we run the proposed framework on the non-uniform light test set, this time adding the proposed NULE-DCE model into the pipeline to preprocess the input images.*

	ResNet 50 VMM Recognition Model	
	Top 1 accuracy (%)	**Top 2 accuracy (%)**
Ground Truth (test data from the VMMR dataset)	90.15%	97.2%
Non-Uniform Light converted test data	88.15%	94.733%
Lowlight test data generate using trained Pix2PixHD	79.32%	87.85%
NULE-DCE enhanced test data	**93.2%**	**98.6%**

Table 2 compares the vehicle make and model recognition accuracies of the proposed joint-enhancement and recognition framework, with and without NULE-DCE enhancing model. The framework is tested on both natural images, generated non-uniform light images, and enhanced images. Quantitative results demonstrate that preprocessing the input images using NULE-DCE significantly improves recognition accuracy.

To further substantiate superiority of the proposed Non-Uniform Light Enhancement model, we compare the enhancement output of NULE-DCE against Zero-DCE and DUAL on vehicle data. Figure 10 compares the visual performance of these models on lowlight vehicles data randomly extracted from the VMMR test set.

Figure 10: Visual comparison of proposed NULE-DCE against Zero-DCE and DUAL on lowlight vehicle data. This figure further demonstrate how NULE-DCE outperforms the methods from which it is inspired.

5. CONCLUSION

In this work, we proposed a novel joint-enhancement and localization framework for vehicle Make and Model recognition task in presence of non-uniform lighting conditions. The proposed framework includes a novel Non-Uniform Light Enhancement via Deep Recursive Curve (NULE-DCE) model inspired from Zero-DCE and DUAL, which is used to preprocess non-uniform light input images/video for improved performance of the subsequent detection and recognition algorithm down the pipeline. NULE-DCE is shown to outperform current state-of-the-art lowlight enhancement algorithms as well as most closely related methods Zero-DCE and DUAL. NULE-DCE can run

at a very fast speed therefore making it most suitable for the computationally intensive framework that includes object detection and recognition with real-time operation objectives. Furthermore, we consolidate a training data from the VMMR dataset and train a ResNet50 model achieving up **93.2%** recognition accuracy on a total of 450 vehicle make and model classes. Our experimental results demonstrate that our framework boost vehicle make and model recognition by **5 to 13%** in presence of non-uniform light/ lowlight data.

Acknowledgement:

This work is supported by the US Department of Transportation, Federal Highway Administration (FHWA), under the grant contract **693JJ320C000023**. Principal Investigator: Dr. Karen Panetta.; Co-PI: Dr. Sos Agaian

REFERENCES

[1] S. Lee, K. Son, H. Kim, and J. Park, "Car plate recognition based on CNN using embedded system with GPU," in *Proceedings - 2017 10th International Conference on Human System Interactions, HSI 2017*, Aug. 2017, pp. 239–241, doi: 10.1109/HSI.2017.8005037.

[2] "Automatic number-plate recognition - IET Conference Publication." https://ieeexplore.ieee.org/abstract/document/191011 (accessed Feb. 20, 2021).

[3] J. E. Espinosa, S. A. Velastin, and J. W. Branch, "Vehicle detection using alex net and faster R-CNN deep learning models: A comparative study," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Nov. 2017, vol. 10645 LNCS, pp. 3–15, doi: 10.1007/978-3-319-70010-6_1.

[4] H. Wang, Y. Yu, Y. Cai, X. Chen, L. Chen, and Q. Liu, "A Comparative Study of State-of-the-Art Deep Learning Algorithms for Vehicle Detection," *IEEE Intell. Transp. Syst. Mag.*, vol. 11, no. 2, pp. 82–95, Jun. 2019, doi: 10.1109/MITS.2019.2903518.

[5] G. Zhang, R. P. Avery, and Y. Wang, "Video-Based Vehicle Detection and Classification System for Real-Time Traffic Data Collection Using Uncalibrated Video Cameras," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 1993, no. 1, pp. 138–147, Jan. 2007, doi: 10.3141/1993-19.

[6] J. Zhu *et al.*, "Vehicle Re-Identification Using Quadruple Directional Deep Learning Features," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 1, pp. 410–420, Jan. 2020, doi: 10.1109/TITS.2019.2901312.

[7] X. Liu, W. Liu, T. Mei, and H. Ma, "A deep learning-based approach to progressive vehicle re-identification for urban surveillance," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 9906 LNCS, pp. 869–884, doi: 10.1007/978-3-319-46475-6_53.

[8] B. Huval *et al.*, "An Empirical Evaluation of Deep Learning on Highway Driving," Apr. 2015, Accessed: Feb. 20, 2021. [Online]. Available: http://arxiv.org/abs/1504.01716.

[9] L. Kezebou, V. Oludare, K. Panetta, and S. Agaian, "A Deep Neural Network Framework for Detecting Wrong-Way Drivers on Highway Roads," in *Multimodal Image Exploitation and Learning 2021*, 2021.

[10] S. Zhang, G. Wu, J. P. Costeira, and J. M. F. Moura, "FCN-rLSTM: Deep Spatio-Temporal Neural Networks for Vehicle Counting in City Cameras," 2017.

[11] "Busted big rig cheat used stolen E-ZPass, fake plates to evade $25,000 in tolls: cops." https://nypost.com/2012/08/30/busted-big-rig-cheat-used-stolen-e-zpass-fake-plates-to-evade-25000-in-tolls-cops/ (accessed Feb. 20, 2021).

[12] "From grease to specialized license plate covers, drivers try to beat all-electronic tolling license plate capture technology - masslive.com." https://www.masslive.com/news/2016/10/from_grease_to_specialized_lic.html (accessed Feb. 20, 2021).

[13] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv*, Apr. 2018, Accessed: Feb. 20, 2021. [Online]. Available: http://arxiv.org/abs/1804.02767.

[14] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 9905 LNCS, pp. 21–37, doi: 10.1007/978-3-319-46448-0_2.

[15] X. Wang, A. Shrivastava, and A. Gupta, "A-Fast-RCNN: Hard Positive Generation via Adversary for Object Detection," 2017.

[16] C. Guo *et al.*, "Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1777–1786, Jan. 2020, Accessed: Feb. 22, 2021. [Online]. Available: http://arxiv.org/abs/2001.06826.

[17] Q. Zhang, Y. Nie, and W. Zheng, "Dual Illumination Estimation for Robust Exposure Correction," *Comput. Graph. Forum*, vol. 38, no. 7, pp. 243–252, Oct. 2019, doi: 10.1111/cgf.13833.

[18] F. Tafazzoli, H. Frigui, and K. Nishiyama, "A Large and Diverse Dataset for Improved Vehicle Make and Model Recognition," 2017.

[19] J. Krause, M. Stark, J. Deng, and L. Fei-Fei, "3D object representations for fine-grained categorization," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 554–561, doi: 10.1109/ICCVW.2013.77.

[20] L. Yang, P. Luo, C. C. Loy, and X. Tang, "A Large-Scale Car Dataset for Fine-Grained Categorization and Verification," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 3973–3981, Jun. 2015, Accessed: Feb. 21, 2021. [Online]. Available: http://arxiv.org/abs/1506.08959.

[21] G. Pearce and N. Pears, "Automatic make and model recognition from frontal images of cars," in *2011 8th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2011*, 2011, pp. 373–378, doi: 10.1109/AVSS.2011.6027353.

[22] N. Boonsim and S. Prakoonwit, "Car make and model recognition under limited lighting conditions at night," *Pattern Anal. Appl.*, vol. 20, no. 4, pp. 1195–1207, Nov. 2017, doi: 10.1007/s10044-016-0559-6.

[23] A. J. Siddiqui, A. Mammeri, and A. Boukerche, "Real-Time Vehicle Make and Model Recognition Based on a Bag of SURF Features," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 11, pp. 3205–3219, Nov. 2016, doi: 10.1109/TITS.2016.2545640.

[24] H. Wang, Y. Yu, Y. Cai, L. Chen, and X. Chen, "A Vehicle Recognition Algorithm Based on Deep Transfer Learning with a Multiple Feature Subspace Distribution," *Sensors*, vol. 18, no. 12, p. 4109, Nov. 2018, doi: 10.3390/s18124109.

[25] M. A. Manzoor, Y. Morgan, and A. Bais, "Real-Time Vehicle Make and Model Recognition System," *Mach. Learn. Knowl. Extr.*, vol. 1, no. 2, pp. 611–629, Apr. 2019, doi: 10.3390/make1020036.

[26] X. Ma and A. Boukerche, "An AI-based Visual Attention Model for Vehicle Make and Model Recognition," in *Proceedings - IEEE Symposium on Computers and Communications*, Jul. 2020, vol. 2020-July, doi: 10.1109/ISCC50000.2020.9219660.

[27] J. Fang, Y. Zhou, Y. Yu, and S. Du, "Fine-Grained Vehicle Model Recognition Using A Coarse-to-Fine Convolutional Neural Network Architecture," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 7, pp. 1782–1792, Jul. 2017, doi: 10.1109/TITS.2016.2620495.

[28] H. Lee, I. Ullah, W. Wan, Y. Gao, and Z. Fang, "Real-Time Vehicle Make and Model Recognition with the Residual SqueezeNet Architecture," *Sensors*, vol. 19, no. 5, p. 982, Feb. 2019, doi: 10.3390/s19050982.

[29] X. Guo, Y. Li, and H. Ling, "LIME: Low-light image enhancement via illumination map estimation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982–993, Feb. 2017, doi: 10.1109/TIP.2016.2639450.

[30] Y. Zhang, J. Zhang, and X. Guo, "Kindling the Darkness: A Practical Low-light Image Enhancer," *MM 2019 - Proc. 27th ACM Int. Conf. Multimed.*, pp. 1632–1640, May 2019, Accessed: Feb. 22, 2021. [Online]. Available: http://arxiv.org/abs/1905.04161.

[31] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep Retinex Decomposition for Low-Light Enhancement," *arXiv*, Aug. 2018, Accessed: Feb. 22, 2021. [Online]. Available: http://arxiv.org/abs/1808.04560.

[32] Z. Jiang, H. Li, L. Liu, A. Men, and H. Wang, "A Switched View of Retinex: Deep Self-Regularized Low-Light Image Enhancement," Jan. 2021, Accessed: Feb. 22, 2021. [Online]. Available: http://arxiv.org/abs/2101.00603.

[33] J. Liang, Y. Xu, Y. Quan, J. Wang, H. Ling, and H. Ji, "Deep Bilateral Retinex for Low-Light Image Enhancement," *arXiv*, Jul. 2020, Accessed: Feb. 22, 2021. [Online]. Available: http://arxiv.org/abs/2007.02018.

[34] Y. Jiang *et al.*, "EnlightenGAN: Deep Light Enhancement without Paired Supervision," *IEEE Trans. Image Process.*, vol. 30, pp. 2340–2349, 2021, doi: 10.1109/TIP.2021.3051462.

[35] V. Oludare, L. Kezebou, K. Panetta, and S. S. Agaian, "Attention-guided cascaded networks for improved face detection and landmark localization under low-light conditions," in *Mobile Multimedia/Image Processing, Security, and Applications 2020*, Apr. 2020, vol. 11399, p. 13, doi: 10.1117/12.2558397.

[36] K. Panetta, L. Kezebou, V. Oludare, S. Agaian, and Z. Xia, "TMO-Net: A Parameter-Free Tone Mapping Operator Using Generative Adversarial Network, and Performance Benchmarking on Large Scale HDR Dataset," *IEEE Access*, pp. 1–1, 2021, doi: 10.1109/ACCESS.2021.3064295.

[37] T. C. Wang, M. Y. Liu, J. Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Dec. 2018, pp. 8798–8807, doi: 10.1109/CVPR.2018.00917.