BITPNet: Unsupervised Bio-Inspired Two-Path Network for Nighttime Traffic Image Enhancement

PENGJIE TAO 1, HULIN KUANG 2, YANSONG DUAN 1, LIANG ZHONG 3,4, AND WU QIU 2, (Member, IEEE)

1 School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430072, China
2 Department of Clinical Neurosciences, University of Calgary, Calgary, AB T2N 1N4, Canada
3 Changjiang Survey Planning Design and Research Company Ltd., Wuhan 430010, China
4 Changjiang Spatial Information Technology Engineering Company Ltd., Wuhan 430010, China

Corresponding author: Hulin Kuang (hulinkuang@gmail.com)

This work was supported by the National Key Research and Development Program of China under Grant 2018YFD1100405 and Grant 2017YFC1501202.

ABSTRACT Due to the low luminance in nighttime traffic images, image features are not salient, making tasks in intelligent transportation systems such as nighttime vehicle detection challenging. Recently, convolutional neural network based methods have been developed for low-light image enhancement. Most of these methods are supervised and require high-light reference images at the same scenes. However, reference images are difficult to be obtained in nighttime traffic scenes because vehicles always move. In the early visual system the input signals are processed by two parallel visual paths in the retina: one path has small receptive fields (RFs) to process the high frequency information and another path has large RFs to deal with the low frequency information. Inspired by this, we design a novel bio-inspired two-path convolutional neural network (BITPNet) for nighttime traffic image enhancement. The high-frequency path with small convolution kernel size is designed to suppress noises and preserve the details. The low-frequency path with large convolution kernel size is used to enhance the luminance of images. Each path includes an encoder-to-decoder network followed by a new multi-level attention module to combine features of levels with different RFs. The outputs of the two paths are summed by learnt weights for generating the final image enhancement result. Several no-reference image quality metrics are utilized to design a new loss function, resulting in an unsupervised approach. The proposed BITPNet is trained on one nighttime traffic image dataset and evaluated on another nighttime dataset. Experimental results demonstrate that the proposed BITPNet outperforms several state-of-the-art low-light image enhancement methods in terms of visual quality and three no-reference image quality metrics. In addition, when the proposed BITPNet is used as pre-processing for the nighttime multi-class vehicle detection task, it achieves higher detection rate (97.18%) than other methods.

INDEX TERMS Bio-inspired method, convolutional neural network, image enhancement, intelligent transportation systems, nighttime traffic image, nighttime vehicle detection.

I. INTRODUCTION

Different from daytime images, nighttime traffic images are with low luminance, thus the image features such as color, shape and edge information are not salient. As a result, some tasks in intelligent transportation systems (ITS) at night such as nighttime vehicle detection become difficult.

The associate editor coordinating the review of this manuscript and approving it for publication was Xiangxue Li.

Low-light image enhancement is a necessary pre-processing step of ITS at night [1]–[3]. In recent decades, low-light image enhancement has attracted the attention of many researchers [4]. We divide the low-light image enhancement methods into two types: generic methods and convolution neural network (CNN) based methods.

In the generic low-light image enhancement methods, the most classical approaches are histogram equalization which adjusts the histogram of images, and its variations such
as adaptive histogram equalization (AHE) [5] and AHE with dual gamma correction [6]. Some other more complex techniques such as adaptive adjustment [7], principal component analysis [8], perceptual color transfer [9] and volume-based subspace analysis [10] are also utilized. Yang et al. proposed an adaptive method for image dynamic range adjustment which was evaluated on nighttime image enhancement [7].

Inspired by the early visual system, Retinex-based methods, e.g., single-scale Retinex (SSR) [11], multi-scale Retinex (MSR) [12], and MSR with color restoration [13] have been developed for image enhancement. There are also other bio-inspired low-light image enhancement methods. For example, Kuang et al. developed a nighttime image enhancement method modelling the photoreceptors, horizontal cells and bipolar cells of the retina [2]. Yang et al. proposed a biological vision inspired method to enhance poor visibility images [14]. However, these methods may produce unnatural results, where the images are over-enhanced, or noises are enhanced.

On the other hand, in the recent years, CNN based low-light image enhancement methods have been developed for low-light image enhancement [15]–[21]. LLNet in [15] stacked sparse auto-encoders to enhance and denoise the low-light images simultaneously. The GLADNet proposed by Wang et al. utilized an encoder-to-decoder network for illumination estimation and another convolutional neural network for detail reconstruction [16]. Guo et al. combined an end-to-end fully convolutional network and discrete wavelet transformation for low-light image enhancement [17]. Attention U-Net in which convolutional features in the low level of the decoder was utilized to calibrate the features in the high level of the encoder, was applied for low-light image enhancement in [18]. In addition, many recent methods use the Retinex theory to design CNNs, e.g., LightenNet [19] based on SSR, and [20], [21] and MSR-net [22] based on MSR. Above CNN based methods are supervised, that is to say, they need high-light reference images (groundtruths) in the same scenes of the low-light images. However, in nighttime traffic scenes, because the vehicles and pedestrians always move, it is difficult to build such reference images. Therefore, in this study we design an unsupervised CNN based method for nighttime traffic image enhancement.

In the retina of the early vision systems, two types of cells, i.e., Midget cells and Parasol cells are in charge of the two parallel visual paths for information processing [23], [24]. Specifically, Midget cells have small receptive fields (RFs) and process the high-frequency information and Parasol cells have larger RFs and deal with low-frequency information [24]. In each visual path multiple cells with different RFs work together and the processing results of the two parallel paths are fused by the following visual processing steps [24]. From an imaging perspective, the high frequency component contains the detail information (e.g., noises and edges in images) changing rapidly over space while the low frequency component contains the global information (e.g., luminance of images) varying slowly over space. To show the evidence of existence of the high frequency and low frequency components in nighttime traffic images, we have applied an image decomposition method proposed in [14] on a nighttime traffic image. The image decomposition results are shown in Fig. 1.

The low frequency component in Fig. 1(b) captures the global information such as luminance, and the high frequency component such as edges and noises, are maintained in Fig. 1(c). These results show that it is reasonable to design a nighttime traffic image enhancement method inspired by the two parallel paths in the early visual systems.

Inspired by the above biological vision mechanism, in this study we propose a novel Bio-Inspired Two-Path convolutional neural Network (BITPNet), which models the two parallel processing paths in the retina for nighttime traffic image enhancement. The BITPNet consists of two paths: the low-frequency and high-frequency paths. Each path is designed using an encoder-to-decoder network and a multi-level attention module. In addition, the two paths are combined by learnt weights and are further processed by a convolutional layer and Sigmoid function to generate the final enhancement results. To make the proposed BITPNet unsupervised, we design a new loss function using three no-reference image quality metrics. The proposed BITPNet are tested on nighttime traffic images and further evaluated using
nighttime multi-class vehicle detection, an important part of nighttime ITS.

Our contributions in this study are threefolds:

1) A novel bio-inspired two-path network modelling the two parallel visual paths in the retina is proposed for nighttime traffic image enhancement. An encoder-to-decoder network with a multi-level attention module is designed to model each path. Combination of the two paths using learnt weights is developed inspired by the fact that two visual paths in the retina are fused by the following visual processing steps.

2) The proposed loss function using several no-reference image quality metrics makes the proposed method unsupervised.

3) To our best knowledge, this study is the first CNN-based low-light image enhancement work tested on nighttime traffic images and further evaluated by a high-level ITS task, i.e., nighttime multi-class vehicle detection.

The rest of the paper is organized as follows. The proposed BITPNet is described in Section II. The details of experiment settings in this study are provided in Section III. Section IV describes the experimental results and comparisons with the state-of-the-arts. Finally, Section V and Section VI state the discussions and conclusions, respectively.

II. THE PROPOSED BITPNet

This section introduces the architecture of the proposed BITPNet, the details of two paths and their combination, the designed loss function, and implementation details.

A. ARCHITECTURE OF THE PROPOSED BITPNet

Inspired by the two parallel visual paths in the retina of the early visual system, we propose a bio-inspired two-path network using RGB images as inputs for nighttime traffic image enhancement. The architecture of the proposed BITPNet is shown in Fig. 2. There are two paths: the low-frequency (LF) and high-frequency (HF) paths. Each path is designed using an encoder-to-decoder network with a multi-level attention module (MLAM). The HF path is used to suppress noises and preserve details of images and the LF path is utilized to enhance the luminance of images to a reasonable level. In each path, the encoder-to-decoder network is a U-shape network with skip-connections between the encoder and the decoder. The MLAM is proposed to model the collaboration of multiple cells in each visual path and make full use of features in multiple levels. Finally, the final enhanced image is generated by weighted summation of the two paths followed by the $1 \times 1$ convolution and sigmoid operations.

B. THE HIGH-FREQUENCY AND LOW-FREQUENCY PATH

The HF path is designed to model the visual processing steps of Midget cells which have small receptive fields (RFs) and process the high-frequency information (e.g., the noises and edges in images). The objective of the HF path is to preserve details of images and suppress noises. The LF path is utilized to simulate the visual processing steps of Parasol cells that have larger RFs and process the low-frequency information (e.g., the luminance of images). The aim of the LF path is to enhance the luminance of images to a reasonable level. According to the study for understanding the RFs in CNNs [25], the smaller the size of convolutional kernels, the smaller the RF. Thus, we use convolutional layers with a kernel size of $3 \times 3$ to simulate the small RFs of Midget cells in the HF path and utilize convolutional layers with a kernel size of $7 \times 7$ to model the large RFs of Parasol cells.
in the LF path. The HF and LF paths have the same structures of the encoder and decoder. In addition, in both the HF path and the LF path, the proposed MLAM is used to combine features of multiple levels in the decoder.

![Figure 3](https://example.com/figure3.png)  
**Figure 3.** The structure of the proposed MLAM. The number of feature channels is shown on the top of each block.

### C. THE PROPOSED MLAM

Multiple Midget cells (with different RFs) in the HF path contribute to the visual processing for high-frequency information and multiple Parasol cells (with different RFs) in the LF path work together for the visual processing for low-frequency information. Thus, in both the HF path and the LF path, the features of all levels in the encoder and decoder should be combined. Because the features in the encoder have been skip-connected with the features in the decoder, in this study the combination of multiple cells in each path is simulated by a proposed multi-level attention module (MLAM) which only combines features in multiple levels in the decoder (i.e., features from different RFs). The structure of the proposed MLAM is shown in Fig. 3. Let $N$ be the number of levels in the decoder. The top level in the decoder is numbered as the 1-st level. In the MLAM, except the features of the 1-st level, features of other levels in the decoder are first resampled by a deconvolution layer to match the spatial resolution and channel numbers of the features of the 1-st level. Then these resampled features and the features of the 1-st level are element-wise summed and processed by the ReLU, 1$\times$1 convolution layer to match the spatial resolution and channel numbers of the features of the 1-st level. Then these resampled features and the features of the 1-st level are element-wise summed and processed by the ReLU, 1$\times$1 convolution and Sigmoid function to generate an attention weight map. This map is utilized to calibrate the ReLU, 1$\times$1 convolution layer to learn weight matrix of each path for combination. According to [26], after performing 1$\times$1 convolution on the result of each path, each feature channel of the output is a linearly weighted summation of all input feature channels and the corresponding weight matrix can be automatically learnt during training without user interactions (i.e., the result of each path is weighed via 1$\times$1 convolution with the learnt weight matrix). The two weighted results of the HF path and the LF path obtained after 1$\times$1 convolution are further processed by another 1$\times$1 convolution layer followed by Sigmoid to generate the final enhanced image.

### D. COMBINATION OF THE TWO PATHS

In the early visual system, the processing results of Midget and Parasol cells are combined by the following visual processing steps. To simulate this, we should combine the results of the HF and LF paths of the proposed BITPNet. In order to combine the results from two types of cells, two scalar weights are experimentally set by the user in the bio-inspired method in [14]. However, the empirical weights might not be optimal. In this study, we use a 1$\times$1 convolution layer to learn weight matrix of each path for combination. According to [26], after performing 1$\times$1 convolution on the result of each path, each feature channel of the output is a linearly weighted summation of all input feature channels and the corresponding weight matrix can be automatically learnt during training without user interactions (i.e., the result of each path is weighed via 1$\times$1 convolution with the learnt weight matrix). The two weighted results of the HF path and the LF path obtained after 1$\times$1 convolution are further element-wisely summed. The weighted combination of the results of the HF ($F_{HF}^{MLAM}$) and LF ($F_{LF}^{MLAM}$) paths is formulated as follows:

$$F_{HF+LF} = W_2 \ast F_{HF}^{MLAM} + W_3 \ast F_{LF}^{MLAM}, \quad (2)$$

where $W_2$ and $W_3$ denote the automatically learnt weights for the results of the HF and LF paths, respectively, and $\ast$ denotes the convolution operation. The weighted combination results $F_{HF+LF}$ are further processed by another 1$\times$1 convolution layer followed by Sigmoid to generate the final enhanced image.

### E. LOSS FUNCTION

Because it is difficult to build the reference images (groundtruths) for nighttime traffic images, the loss functions used in the previous CNN based methods such as mean squared error (MSE) [21] are not suitable in this application. In this study, we utilize three no-reference image quality metrics, i.e., the lightness order error (LOE)\(^1\) in [27], the natural image quality evaluator (NIQE)\(^2\) in [28], and the integrated local natural image quality evaluator (ILNIQE)\(^3\) in [29], to design a new loss function. The details of computing these three metrics can be found in their codes. The LOE measures the lightness order error between the enhanced image and the original image, and its range is 0 to 5000. The normalized LOE can be used as one loss. For the NIQE and ILNIQE, they maybe higher than 1. If the image enhancement method is effective, the NIQE and ILNIQE values of the enhanced image will be less than those of the original image. Therefore,

\(^1\)https://github.com/baidut/BIMEF/tree/master/quality  
\(^2\)http://live.ece.utexas.edu/research/quality/niqe_release.zip  
\(^3\)http://www4.comp.polyu.edu.hk/~cslihang/IQA/ILNIQE/Files/ILNIQE.zip
the ratios of the NIQE and ILNIQE of the enhanced image and those of the original images can also be used as losses. The loss function of the proposed BITPNet is a weighted summation of three terms as follows:

$$L(I_{En}, I) = \omega_1 \frac{{\text{LOE}(I_{En}, I)}}{5000} + \omega_2 \frac{{\text{NIQE}(I_{En})}}{{\text{NIQE}(I)}} + \omega_3 \frac{{\text{ILNIQE}(I_{En})}}{{\text{ILNIQE}(I)}},$$

(3)

where $I$ and $I_{En}$ denote the original nighttime image and the enhanced image, respectively. LOE($I_{En}$, $I$) is the LOE between the enhance image and the original image, NIQE($I_{En}$) and ILNIQE($I_{En}$) are the NIQE and ILNIQE of the enhanced image, NIQE($I$) and ILNIQE($I$) are the NIQE and ILNIQE of the original image, and $\omega_1$, $\omega_2$ and $\omega_3$ are weights to balance the three terms.

F. IMPLEMENTATION DETAILS
All RGB nighttime traffic images used in this study are normalized to [0, 1] using the minimum and maximum before being input into the proposed BITPNet. To increase the number of the training images, we apply online data augmentations using intensity variation, scaling, rotation, and left-right flipping. For the LF and HF paths, the encoders and the decoders are all have 5 levels. Each level has a convolutional layer followed by ReLU and batch normalization. In the HF path, the kernel size of the convolutional layers is $3 \times 3$ and in the LF path the kernel size of the convolutional layers is $7 \times 7$. The kernel sizes of max pooling layers and deconvolution layers are $2 \times 2$ and $3 \times 3$, respectively. The number of feature channels in each layer has been shown in Fig. 2. Sigmoid function is used to activate the output features of the proposed BITPNet. To optimize the model, we utilize the Adam optimization method with a fixed learning rate of 0.0001. $\omega_1 = 0.2$, $\omega_2 = 0.4$ and $\omega_3 = 0.4$ in the loss function (3) are experimentally determined after parameter tuning on the validation set. The maximum iterations is 50000 and the size of batch is 8. All experiments are implemented using Python and Tensorflow, and are performed on a Linux workstation with an Intel Xeon Gold 6148 20-core processor and two Tesla V100-PCIE-16GB GPUs.

III. EXPERIMENT SETTING
This section describes the datasets used for training and testing, evaluation metrics and comparison benchmarks.

A. NIGHTTIME TRAFFIC DATASETS
The nighttime vehicle dataset built by Dr. Long Chen in [1] is used for training he proposed BITPNet. It contains 400 RGB nighttime traffic images with a resolution of $360 \times 640$. Examples of nighttime traffic images are shown in Fig. 4. This training dataset includes various scenes such as highways (Fig. 4(a)), housing estates (Fig. 4(b)), and bridges (Fig. 4(c)). 350 out of 400 images is used to train the proposed BITPNet and the remaining 50 images is used for validation and parameter tuning. The Hong Kong nighttime multi-class vehicle dataset built in [30] is used as the testing dataset to evaluate the proposed method. This testing dataset contains 836 nighttime traffic images (RGB) with a resolution of $1080 \times 1920$. It also includes different scenes such as streets (Fig. 4(d)) and highways (Fig. 4(e)), and rainy weather (Fig. 4(f)). Within these 836 images, there are 2058 vehicles including four types: car, taxi, bus and minibus. The testing dataset is available online.\(^4\)

\(^4\)https://www.researchgate.net/profile/Hulin_Kuang/publications
TABLE 1. Average LOE, NIQE and ILNIQE of the six compared methods over the 836 nighttime traffic images in the testing dataset.

| Method        | Original images | Kuang [2] | Ada [7] | Yang [14] | GLADNet [16] | Attention U-Net [18] | BITPNet |
|---------------|----------------|----------|---------|-----------|--------------|----------------------|---------|
| LOE           | 452.52         | 313.36   | 406.92  | 346.75    | 377.43       | 289.14               |
| NIQE          | 6.78           | 5.97     | 6.15    | 5.74      | 5.25         | 5.33                 |
| ILNIQE        | 30.73          | 27.65    | 24.43   | 25.63     | 26.09        | 25.69                |

FIGURE 5. Two visual examples of the enhanced results obtained by the six compared methods. Each row denotes an example, and subfigures from left to right are the original nighttime image, and results of Kuang [2], Ada [7], Yang [14], GLADNet [16], Attention U-Net [18] and our proposed BITPNet.

B. EVALUATION METRICS

To evaluate the nighttime traffic image enhancement methods, we not only use the visual quality evaluation but also use three no-reference image quality metrics: LOE, NIQE and ILNIQE. For LOE, NIQE and ILNIQE, the smaller the value, the better the image quality. To validate the effectiveness of the proposed BITPNet for a high-level ITS task: nighttime multi-class vehicle detection, it is applied on the Hong Kong nighttime multi-class dataset and then the enhanced images are processed by the multi-class vehicle detection method in [30] to detect nighttime multi-class vehicles. The multi-class vehicle detection rate is used as another evaluation metric. The higher the detection rate, the better the proposed nighttime image enhancement method.

C. COMPARISON BENCHMARKS

We select five state-of-the-art low-light image enhancement methods as comparison benchmarks, i.e., two bio-inspired methods: the method in [2] (called Kuang in this study) and the method in [14] (called Yang), an adaptive method in [7] (called Ada), and two CNN-based methods: GLADNet in [16] and attention U-Net in [18]. The first three methods are not CNN based methods, they are directly tested on the testing dataset using their codes. For the GLADNet and attention U-Net methods, we use the supervised trained model on the datasets used in their studies to enhance the nighttime traffic images in the testing dataset.

IV. RESULTS

This section shows the quantitative and qualitative results of the proposed BITPNet, its application on nighttime multi-class vehicle detection and the ablation study.

A. QUANTITATIVE AND QUALITATIVE RESULTS

Firstly, we compare the quantitative results (i.e., LOE, NIQE and ILNIQE) of the proposed BITPNet and the five benchmark methods in Table 1. LOE is computed via comparing the enhanced image and the original image, thus there is no LOE value for the original images. We find that our proposed BITPNet achieves a LOE of 289.14, a NIQE of 4.93 and a ILNIQE of 21.17, which are all smaller than those of the other five methods, showing that it outperforms the other five methods in terms of quantitative evaluations.

In addition, we compare the qualitative results of our proposed BITPNet with the five benchmarks via comparing the visual quality. Two visual examples of the enhanced images generated by our proposed BITPNet and the other five methods are shown in Fig. 5. For evaluating the visual quality of nighttime traffic images, we should focus on the regions of vehicles, traffic lights and road signs. In each subfigure of Fig. 5, we crop a red block (containing vehicle) and a green block (containing traffic lights or road signs), and then enlarge and show them at the top of each subfigure for visual comparison. The Kuang [2] and Yang [14] methods enhance the luminance of the vehicles in red blocks of the two rows but introduce noises into the regions of traffic lights and road signs in the green blocks of the two rows, making these regions blurred. The Ada method [7] does not restore color well for the vehicle in the red block of the first row and enhances the noises too much, making the road sign in the green block of the second row unnatural. In the results of the GALADNet method [16], there exist color inconsistency in the sky regions of the two rows and haloes around the traffic lights in the green block of the first row, but the regions of vehicles are enhanced to a high level and their details are
TABLE 2. The LOE, NIQE and ILNIQE of the six compared methods for the two rows in Fig. 5.

| Method                      | Original | Kuang [2] | Ada [7] | Yang [14] | GLADNet [16] | Attention U-Net [18] | BITPNet |
|-----------------------------|----------|-----------|---------|-----------|--------------|---------------------|---------|
| LOE of 1st row              | 460.23   | 301.45    | 360.43  | 558.53    | 321.45       | 281.34              |
| LOE of 2nd row              | 370.62   | 245.63    | 279.36  | 350.72    | 264.32       | 218.23              |
| NIQE of 1st row             | 7.08     | 6.78      | 5.53    | 6.43      | 5.80         | 5.94                |
| NIQE of 2nd row             | 8.32     | 6.06      | 5.22    | 6.23      | 5.69         | 5.42                |
| ILNIQE of 1st row           | 50.40    | 38.24     | 49.37   | 46.38     | 45.81        | 41.87               |
| ILNIQE of 2nd row           | 41.67    | 32.56     | 35.39   | 34.72     | 36.63        | 32.25               |

TABLE 3. The average computational time (in seconds) per image of our proposed BITPNet and other methods over 836 nighttime traffic images.

| Method       | Kuang [2] | Ada [7] | Yang [14] | GLADNet [16] | Attention U-Net [18] | BITPNet |
|--------------|-----------|---------|-----------|--------------|---------------------|---------|
| Time         | 5.61      | 6.31    | 6.85      | 0.27         | 0.32                | 0.24    |

TABLE 4. Multi-class vehicle detection rates of the method in [30] using our proposed BITPNet and other compared image enhancement methods as pre-processing.

| Method       | Kuang [2] | Ada [7] | Yang [14] | GLADNet [16] | Attention U-Net [18] | BITPNet |
|--------------|-----------|---------|-----------|--------------|---------------------|---------|
| Detection rate (%) | 95.82 | 96.01   | 96.31     | 96.69        | 96.79               | 97.18   |

preserved well (see the red blocks of the two rows). The attention U-Net method [18] retains the details of vehicles, traffic lights and road signs (red and green blocks in the two rows) but the luminance of the images is not enhanced very much. Our proposed BITPNet preserves details of the vehicles in the red blocks of the two rows and details of the traffic lights in the green block of the first row, which is better than three methods: Kuang [2], Ada [7] and Yang [14]. Compared to the GLADNet method, our proposed BITPNet does not introduce color inconsistency in the sky regions of the two rows and halos around the traffic lights in the green block of the first row. For the road sign in the green block of the second row, the proposed BITPNet does not enhance noise too much, which is similar to the GLADNet and attention U-Net methods and is better than the Kuang, Ada and Yang methods. In addition, it enhances the luminance of the two images to a better level than the attention U-Net method. In summary, our proposed BITPNet performs well on detail preservation, noise suppression and color restoration for vehicles, traffic lights and road signs, and can enhance the luminance of the images to an acceptable level without color inconsistency and halos. We can conclude that our proposed BITPNet achieves the best visual quality. Besides, because of the poor color restoration and unnatural road signs, the Ada method is the worst in terms of visual quality.

To measure the relationship between the visual quality and the quantitative evaluation metrics, the LOE, NIQE and ILNIQE of the six compared methods for the two examples in Fig. 5 are shown in Table 2. We find that our proposed BITPNet obtains the smallest LOE, NIQE and ILNIQE for both the two examples. The Ada method achieves the second smallest LOE and NIQE for the two examples, however, its visual quality is the worst. This shows that the visual quality evaluation and the quantitative evaluation of the nighttime enhanced images are inconsistent, implying that the no-reference image quality metrics might be not very robust. In

The average computational time per image of our proposed BITPNet and other methods over 836 nighttime traffic images is shown in Table 3. The proposed BITPNet takes 0.24 seconds to enhance one image. The speed of the three CNN based methods using GPU: the proposed BITPNet, the GLADNet and the attention U-Net is faster than the three other methods implemented using Matlab without GPU. It maybe unfair to compare the computational time in this way. But we can conclude that the proposed BITPNet is fast if it is applied with GPU.

B. APPLICATIONS ON NIGHTTIME MULTI-CLASS VEHICLE DETECTION

In general, low-light image enhancement is a pre-processing step for high-level computer vision tasks. According to above analyses, the no-reference image quality metrics maybe not robust for evaluating the low-light image enhancement methods. Thus, it maybe better to evaluate the image enhancement methods using the performance of a high-level ITS task: nighttime multi-class vehicle detection. In this study, we use our proposed BITPNet and the other five benchmarks as pre-processing to enhance the nighttime traffic images in the testing dataset, and then apply the multi-class vehicle detection method in [30] on the enhanced images to detect multi-class vehicles. The multi-class vehicle detection rates using these six methods as pre-processing are shown in Table 4. The nighttime multi-class vehicle method in [30] used the Kuang method [2] for nighttime image enhancement. Thus the detection rate (95.82%) of the Kuang method in Table 4 is the same as the reported result in [30]. Compared to the other five methods, our proposed BITPNet achieves the highest multi-class vehicle detection rate, showing that when it is used as the pre-processing step, it can improve the performance of the following nighttime multi-class vehicle detection task. Some multi-class vehicle detection results using our proposed BITPNet as the pre-processing step are shown in Fig. 6.
We can find that with the help of our proposed BITPNet, the method in [30] can accurately detect multi-class vehicles (Fig. 6(a)), occluded vehicles (Fig. 6(b-c)), far and small vehicles (Fig. 6(d-e)) and vehicles in rainy weather (Fig. 6(f)).

C. ABSTRACTION STUDY
We conduct several ablation studies. The first one removes the proposed MLAM in the HF and LF paths and keep other parts unchanged (called no MLAM). The second one deletes the HF path, i.e., only using the LF path for image enhancement. The third one removes the LF path, i.e., only using the HF path for image enhancement. The fourth one removes the 1 × 1 convolution layers for learning weights of the LF and HF paths for summation (called no learn weights), i.e., the outputs of the LF and HF paths are summed directly without weights. The multi-class vehicle detection rates of the four ablation studies are shown in Table 5. The results show that removing any part of the proposed BITPNet decreases the multi-class vehicle detection rate, which validates that all of the LF path, the HF path, the MLAM and the weighted summation of the LF and HF paths play key roles in accurately vehicle detection and are necessary for good nighttime image enhancement. In particular, only using the HF path demonstrates the lowest detection rate. This is can be explained as follows. The designed HF path with small RFs tends to process the noises and details in the images and might miss the luminance information which is very useful for vehicle detection, resulting in the decrease of the detection rate.

V. DISCUSSIONS
This section states some discussions and future works.

A. KERNEL SIZES OF THE TWO PATHS
In this study, we use 3 × 3 convolutional layers in the HF path to simulate small RFs and 7 × 7 convolutional layers in the LF path to simulate large RFs. To validate that this kernel size setting is the optimal, we perform additional experiments where we utilize different pairs of kernel sizes in the HF and LF paths, i.e., (3 × 3, 5 × 5), (3 × 3, 9 × 9), (3 × 3, 11 × 11), (5 × 5, 7 × 7), (5 × 5, 9 × 9) and (5 × 5, 11 × 11). Experimental results show that using other pairs of kernel sizes leads to about 0.3% to 1% decrease in the multi-class vehicle detection rate, which validates that the kernel size setting in this study is optimal.

B. RGB VS. HSV
In our proposed BITPNet, we use the normalized RGB images as inputs for nighttime image enhancement. Some studies first convert RGB to HSV color space for image enhancement and then convert the enhance HSV images back to RGB such as the method in [14], because they think the HSV color space is more suitable to the human visual system. We wonder whether using HSV improves the proposed method or not. Therefore, we conduct an addition experiment where HSV images converted from RGB images are utilized as inputs and after enhancement we convert the enhance images back to RGB. This experiment achieves a multi-class
vehicle detection rate of 96.79%, which is slightly lower than 97.18% of the proposed method with RGB as inputs. This result demonstrates that converting RGB to HSV is not necessary.

C. LOSS FUNCTION

There are three weights in the loss function (3). In this study, \( \omega_1 = 0.2, \omega_2 = 0.4 \) and \( \omega_3 = 0.4 \) are set according to the parameter tuning on the validation set. We wonder whether there exist other weights improving the multi-class vehicle detection rate. Thus, we perform additional experiments using different weights to train the proposed BITPNet. Experimental results show that when \( \omega_1 = 0.1 \) and \( \omega_2 = 0.4 \) and \( \omega_3 = 0.5 \), the proposed BITPNet achieves a detection rate of 97.32%, slightly outperforming 97.18% using the old weights. In addition, the \( \text{LOE} = 301.26 \), \( \text{NIQE} = 5.43 \), and \( \text{ILNIQE} = 24.35 \), which are all larger than those of using the old weights. This can be explained as follows. The loss function is based on three no-reference image quality metrics, the weights set by parameter tuning only ensure the smallest image quality metrics rather than the following multi-class vehicle detection rate. That is to say, the smallest image quality metric values might not ensure the best detection rate. This implies that there is a gap between the image quality metrics and the multi-class vehicle detection rate.

For fair comparisons, we conduct an additional experiment where we use our proposed loss function to re-train and evaluate the GLADNet and Attention U-Net on the same datasets used in this study. The results of \( \text{LOE}, \text{NIQE}, \text{ILNIQE} \) and nighttime multi-class vehicle detection rate are shown in Table 6. It is found that our proposed BITPNet achieves better performance than the GLADNet and Attention U-Net methods using the proposed loss function in terms of all evaluation metrics, which validates the superiority of our proposed BITPNet.

| Method  | GLADNet  | Attention U-Net | BITPNet  |
|---------|----------|-----------------|----------|
| LOE     | 310.25   | 323.56          | 289.14   |
| NIQE    | 5.02     | 5.14            | 4.93     |
| ILNIQE  | 24.76    | 24.21           | 21.17    |
| Detection rate (%) | 96.74 | 96.84 | 97.18 |

D. FUTURE WORK

First, the proposed method is only trained on a limited dataset (350 images) in this study. Building datasets with more nighttime traffic images is necessary. Second, there is a discrepancy between no-reference image quality metrics and the nighttime vehicle detection performance. The two tasks of nighttime image enhancement and nighttime multi-class vehicle detection can be unified via multi-task learning mechanism. In such a unified framework, the image enhancement task can be embedded into the final goal of vehicle detection and the vehicle detection can be used to guide the image enhancement task. Third, in the future we will design additional networks which can be embedded into the proposed method to simulate the horizontal cells and bipolar cells that also contribute to the vision processing in the early systems.

VI. CONCLUSION

This study proposes a novel unsupervised bio-inspired two-path network for nighttime traffic image enhancement. Encoder-to-decoder networks with different kernel sizes of the convolutional layers and multi-level attention modules are designed to model the two parallel processing paths in the early visual system. Weighted summation of the outputs of the two paths are utilized to generate the final image enhancement results. Three no-reference image quality metrics are utilized to design a new loss function, resulting in an unsupervised fashion. Experimental results that the proposed BITPNet outperforms several state-of-the-art low-light image enhancement methods in terms of quantitative and qualitative image quality evaluations, and can improve the nighttime vehicle detection performance.

ACKNOWLEDGMENT

The would like to thank Dr. L. Chen for providing the nighttime traffic image dataset used for training.

REFERENCES

[1] H. Kuang, L. Chen, F. Gu, J. Chen, L. Chan, and H. Yan, “Combining region-of-interest extraction and image enhancement for nighttime vehicle detection,” IEEE Intell. Syst., vol. 31, no. 3, pp. 57–65, May 2016.
[2] H. Kuang, X. Zhang, Y. Li, L. H. Chan, and H. Yan, “Nighttime vehicle detection based on bio-inspired image enhancement and weighted score-level feature fusion,” IEEE Trans. Intell. Transport. Syst., vol. 18, no. 4, pp. 927–936, Apr. 2017.
[3] H. Kuang, K.-F. Yang, L. Chen, Y.-J. Li, L. H. Chan, and H. Yan, “Bayes saliency-based object proposal generator for nighttime traffic images,” IEEE Trans. Intell. Transport. Syst., vol. 19, no. 4, pp. 814–825, Mar. 2018.
[4] W. Wang, X. Wu, X. Yuan, and Z. Gao, “An experiment-based review of low-light image enhancement methods,” IEEE Access, vol. 8, pp. 87884–87917, 2020.
[5] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, “Adaptive histogram equalization and its variations,” Comput. Vis., Graph., Image Process., vol. 39, no. 3, pp. 355–368, 1987.
[6] Y. Chang, C. Jung, P. Ke, H. Song, and J. Hwang, “Automatic contrast-limited adaptive histogram equalization with dual gamma correction,” IEEE Access, vol. 6, pp. 11782–11792, 2018.
[7] K.-F. Yang, H. Li, H. Kuang, C.-Y. Li, and Y.-J. Li, “An adaptive method for image dynamic range adjustment,” IEEE Trans. Circuits Syst. Video Technol., vol. 29, no. 3, pp. 640–652, Mar. 2019.
[8] S. A. Priyanka, Y.-K. Wang, and S.-Y. Huang, “Low-light image enhancement by principal component analysis,” IEEE Access, vol. 7, pp. 3082–3092, 2019.
[9] I. Cepeda-Negrete, R. E. Sanchez-Yanez, F. E. Correa-Tome, and R. A. Lizarraga-Morales, “Dark image enhancement using perceptual color transfer,” IEEE Access, vol. 6, pp. 14935–14945, 2018.
[10] W. Kim, R. Lee, M. Park, S.-H. Lee, and M.-S. Choi, “Low-light image enhancement using volume-based subspace analysis,” IEEE Access, vol. 8, pp. 118370–118379, 2020.
[11] D. J. Jobson, Z. Rahman, and G. A. Woodell, “Properties and performance of a center/surround retinex,” IEEE Trans. Image Process., vol. 6, no. 3, pp. 451–462, Mar. 1997.
[12] D. J. Jobson, Z. Rahman, and G. A. Woodell, “A multiscale retinex for bridging the gap between color images and the human observation of scenes,” IEEE Trans. Image Process., vol. 6, no. 7, pp. 965–976, Jul. 1997.

[13] Z.-U. Rahman, D. J. Jobson, and G. A. Woodell, “Retinex processing for automatic image enhancement,” in Proc. 7th Hum. Vis. Electron. Imag., vol. 4662, Jun. 2002, pp. 390–401.

[14] K.-F. Yang, X.-S. Zhang, and Y.-J. Li, “A biological vision inspired framework for image enhancement in poor visibility conditions,” IEEE Trans. Image Process., vol. 29, pp. 1493–1506, 2020.

[15] K. G. Lore, A. Akintayo, and S. Sarkar, “LLNet: A deep autoencoder approach to natural low-light image enhancement,” Pattern Recognit., vol. 61, pp. 650–662, Jan. 2017.

[16] W. Wang, C. Wei, W. Yang, and J. Liu, “GLADNet: Low-light enhancement network with global awareness,” in Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG), May 2018, pp. 751–755.

[17] Y. Guo, X. Ke, J. Ma, and J. Zhang, “A pipeline neural network for low-light image enhancement,” IEEE Trans., vol. 7, pp. 13737–13744, 2019.

[18] S. Ai and J. Kwon, “Extreme low-light image enhancement for surveillance cameras using attention U-Net,” Sensors, vol. 20, no. 2, p. 495, Jan. 2020.

[19] C. Li, J. Guo, F. Porikli, and Y. Pang, “LighthenNet: A convolutional neural network for weakly illuminated image enhancement,” Pattern Recognit. Lett., vol. 104, pp. 15–22, Mar. 2018.

[20] W. Huang, Y. Zhu, and R. Huang, “Low light image enhancement network with attention mechanism and retinex model,” IEEE Access, vol. 8, pp. 74306–74314, 2020.

[21] S. Park, S. Yu, M. Kim, K. Park, and J. Paik, “Dual autoencoder network for retinex-based low-light image enhancement,” IEEE Access, vol. 6, pp. 22084–22093, 2018.

[22] L. Shen, Z. Yue, F. Feng, Q. Chen, S. Liu, and J. Ma, “MSR-net: Low-light image enhancement using deep convolutional network,” 2017, arXiv:1711.02488. [Online]. Available: http://arxiv.org/abs/1711.02488

[23] P. H. Schiller, “Parallel information processing channels created in the retina,” Proc. Nat. Acad. Sci. USA, vol. 107, no. 40, pp. 17087–17094, Oct. 2010.

[24] E. Kaplan, “The M, P, and K pathways of the primate visual system,” in The Visual Neurosciences, L. Chalupa and J. Werner, Eds. Cambridge, MA, USA: MIT Press, 2004, pp. 481–493.

[25] W. Luo, Y. Li, R. Urtasun, and R. Zemel, “Understanding the effective receptive field in deep convolutional neural networks,” in Proc. Adv. Neural Inf. Process. Syst., 2016, pp. 4898–4906.

[26] M. Lin, Q. Chen, and S. Yan, “Network in network,” 2013, arXiv:1312.4400. [Online]. Available: http://arxiv.org/abs/1312.4400

[27] Z. Ying, G. Li, and W. Gao, “A bio-inspired multi-exposure fusion framework for low-light image enhancement,” 2017, arXiv:1711.00591. [Online]. Available: http://arxiv.org/abs/1711.00591

[28] A. Mittal, R. Soundararajan, and A. C. Bovik, “Making a ‘completely blind’ image quality analyzert,” IEEE Signal Process. Lett., vol. 20, no. 3, pp. 209–212, Mar. 2013.

[29] L. Zhang, L. Zhang, and A. C. Bovik, “A feature-enriched completely blind image quality evaluator,” IEEE Trans. Image Process., vol. 24, no. 8, pp. 2579–2591, Aug. 2015.

[30] H. Kuang, L. Chen, L. L. H. Chan, R. C. C. Cheung, and H. Yan, “Feature selection based on tensor decomposition and object proposal for nighttime multiclass vehicle detection,” IEEE Trans. Syst., Man, Cybern., Syst., vol. 49, no. 1, pp. 71–80, Jan. 2019.

HULIN KUANG received the B.Eng. and M.Eng. degrees from Wuhan University, China, in 2011 and 2013, respectively, and the Ph.D. degree from the City University of Hong Kong, in 2016. He is currently a Postdoctoral Fellow with the Department of Clinical Neurosciences, University of Calgary, Canada. His current research interests include deep learning, intelligent transportation systems, and medical image processing.

PENGJIE TAO received the B.S. and Ph.D. degrees from the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, in 2008 and 2016, respectively. He is currently an Associate Research Fellow with the School of Remote Sensing and Information Engineering, Wuhan University. His research interests include satellite images processing, registration of optical images, and multi-view images three-dimensional reconstruction.

LIANG ZHONG received the Ph.D. degree from the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, in 2011. He is currently a Senior Engineer with the Changjiang Survey Planning Design and Research Company Ltd., Changjiang Spatial Information Technology Engineering Company Ltd., Wuhan, China. His research interests include remote sensing image processing and laser data processing.

WU QIU (Member, IEEE) received the Ph.D. degree in pattern recognition and artificial intelligence from the Huazhong University of Science and Technology, China, in 2009. He was a Postdoctoral Fellow with the Robarts Research Institute, University of Western Ontario, London, ON, Canada, until he joined the Department of Clinical Neurosciences, University of Calgary, in 2016. His research interests include image processing, medical imaging, computer vision, and machine learning.

YANSONG DUAN was born in 1975. He received the M.S. and Ph.D. degrees from Wuhan University, Wuhan, China, in 2009 and 2016, respectively. He is currently an Associate Professor with the School of Remote Sensing and Information Engineering, Wuhan University. His research interests include photogrammetry, image matching, 3-D city reconstruction, computer vision, and high performance computing with GPU.