Feature Selection Based on Tensor Decomposition and Object Proposal for Night-Time Multiclass Vehicle Detection

Hulin Kuang, Long Chen, *Member, IEEE*, Leanne Lai Hang Chan, *Senior Member, IEEE*, Ray C. C. Cheung, *Member, IEEE*, and Hong Yan, *Fellow, IEEE*

***Abstract*—Night-time vehicle detection is essential in building intelligent transportation systems (ITS) for road safety. Most of current night-time vehicle detection approaches focus on one or two classes of vehicles. In this paper, we present a novel multiclass vehicle detection system based on tensor decomposition and object proposal. Commonly used features such as histogram of oriented gradients and local binary pattern often produce useless image blocks (regions), which can result in unsatisfactory detection performance. Thus, we select blocks via feature ranking after tensor decomposition and only extract features from these selected blocks. To generate windows that contain all vehicles, we propose a novel object-proposal approach based on a state-of-the-art object-proposal method, local features, and image region similarity. The three terms are summed with learned weights to compute the reliability score of each proposal. A bio-inspired image enhancement method is used to enhance the brightness and contrast of input images. We have built a Hong Kong night-time multiclass vehicle dataset for evaluation. Our proposed vehicle detection approach can successfully detect four types of vehicles: 1) car; 2) taxi; 3) bus; and 4) minibus. Occluded vehicles and vehicles in the rain can also be detected. Our proposed method obtains 95.82% detection rate at 0.05 false positives per image, and it outperforms several state-of-the-art night-time vehicle detection approaches.**

***Index Terms*—Feature selection, night-time multiclass vehicle detection, object proposal, tensor decomposition.**

I. Introduction

NIGHT-TIME vehicle detection has attracted more and more attention from researchers in recent decades due to the fact that 32% of all vehicular fatal accidents are caused by rear-end collisions [1] and most of these accidents occur at night. Thus, night-time vehicle detection is important for road safety in intelligent transportation systems (ITS), such as advanced driver assistance systems and autonomous driving systems [2]. Current night-time vehicle detection approaches all focus on a single class of vehicles [2], [3] or two classes [4]–[6]. Different types of vehicles run in different lanes, thus detecting multiclass vehicles is useful for driving control and road safety. In this paper, we focus on night-time multiclass preceding vehicle detection.

In daytime scenes, the contrast between vehicles and background is high and some features of vehicles such as color information, gradient features, and texture features are salient. While at night, the contrast between vehicles and background become lower. As a result, these features salient at daytime, are not very clear while the vehicle lights are more salient because lights are turned on. Thus, most of the current night-time vehicle detection methods are based on vehicle light detection [7]–[9]. However, this type of methods might miss vehicles and detect road lamps, and traffic lights under complex scenes where accurate vehicle light detection becomes difficult. Daytime object detection approaches, such as histogram of oriented gradients (HOG) [10], local binary patterns (LBP) [11], and deformable parts model (DPM) [12], and daytime object detection methods developed for intelligent vehicles [13]–[15], can be used for night-time vehicle detection directly. However, results in [2], [3], and [16] show that, without night-time image enhancement, the object detection methods developed for daytime scenes do not perform well. These results also demonstrate that current daytime object detection methods can be used to detect night-time vehicles well by taking into account characteristics of vehicles in night-time scenes and combining the methods with night-time image enhancement. Therefore, in this paper, we utilize a bio-inspired image enhancement approach [3] to enhance the brightness and the contrast of night-time images to extract effective features and obtain good detection performance. Besides, we propose a new feature selection approach for several useful features developed for daytime object detection and design a new approach for object proposal.

During extractions of features such as HOG [10] and LBP [11], an image is divided into multisize blocks (rectangular regions) and then features are computed from each block [17]. In many images, some blocks contain useless information for night-time vehicle detection. For example,

Manuscript received April 23, 2018; accepted September 11, 2018. Date of publication October 29, 2018; date of current version December 14, 2018. This work was supported in part by the Hong Kong Research Grants Council under Project CityU 111013 and Project C1007-15G, in part by the City University of Hong Kong under Project 9610034, and in part by the National Natural Science Foundation of China under Grant 61301277. This paper was recommended by Associate Editor X. Xu. *(Corresponding author: Leanne Lai Hang Chan.)*

H. Kuang is with the Department of Clinical Neurosciences, University of Calgary, Calgary, AB T2N 4N1, Canada (e-mail: hulinkuang@gmail.com).

L. Chen is with the School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510275, China (e-mail: chenl46@mail.sysu.edu.cn).

L. L. H. Chan, R. C. C. Cheung, and H. Yan are with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong (e-mail: leanne.chan@cityu.edu.hk; r.cheung@cityu.edu.hk; h.yan@cityu.edu.hk.)

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMC.2018.2872891

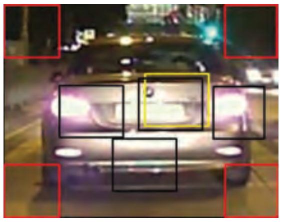


Fig. 1. Examples of useful and useless blocks for vehicle detection in an image.

red blocks in the four corners contain only background (red windows) in Fig. 1. These blocks do not provide useful information, but increase the number of features. They can degrade the detection performance. In Fig. 1, the black blocks contain some parts and edges of a vehicle. They may be more useful and effective for night-time vehicle detection. Besides, the yellow block has a large overlap with a black block. These two blocks may lead to unsatisfactory detection performance and reduce the detection speed because they contain redundant features. Therefore, blocks containing some parts of a vehicle and having little overlap with each other would be more effective. However, feature selection is usually not considered in current night-time vehicle detection methods. In this paper, we propose a tensor decomposition-based feature selection method to solve this problem. A tensor is a multidimensional (high-order) data array and can provide full representation of the natural structure of images and their features [18]–[21]. Tensor decomposition that transforms a high-order tensor into low-dimensional matrices or vectors has been used in computer vision tasks such as human action recognition [19] and image classification [21]. In this paper, we explore algorithms for selecting useful and effective blocks for feature extraction through tensor decomposition.

Because vehicles can be at any positions within an image and of different sizes, generating a set of windows (proposals) that contain vehicles automatically is an important step for robust and accurate vehicle detection. For night-time vehicle detection, most object proposal approaches rely on headlight or taillight detection using color thresholds in the hue, saturation, and value (HSV) color space [22] or multilevel intensity thresholding [23]. These methods have been proven useful to generate a very small set of vehicle candidates. However, their performance is subject to the variations of road scenes, camera parameters, and light reflection, etc. Therefore, some vehicle candidates may be missed if the scenes are complex. To generate a set of windows that detect all preceding vehicles in front of the driver, a state-of-the-art object proposal approach named EdgeBoxes [24] was combined with vehicle light detection by a weighted sum in [2]. In [25], a Bayes saliency-based object proposal method was developed. It utilized vehicle light detection results as features for computing a saliency map. These methods also rely on vehicle light detection, thus they are also influenced by the scenes. In [2], the weights of all terms for computing the probability of each window to be vehicles (score) are the same. In this paper, we combine EdgeBoxes based on an image enhancement method [3], local contrast feature, and image region similarity together using learned weights.

The framework of our proposed night-time vehicle detection approaches is shown in Fig. 2. During the training stage, we extract HOG [10], LBP [11], and four direction features (FDF) [26] from the enhanced training images using a fixed block size. Then for each feature, we construct a third-order tensor for each image and conduct tensor decomposition. We select effective blocks using a feature ranking score function based on the decomposition results. After selecting effective blocks, we use the HOG, LBP, and FDF features extracted from the selected blocks to train a multiclass vehicle classifier using support vector machines (SVMs) [27]. The training process is completed offline. During the detection stage, for each image, we first apply EdgeBoxes on the enhanced image, and then compute local contrast features and the image region similarity from the windows (proposals) obtained from EdgeBoxes. Then, the score of each window obtained from EdgeBoxes is combined with local features and image similarity using learned weights to compute the final score of the window. After extracting features from the selected blocks of each window, we use the trained classifier to recognize the type of the vehicle within each window. After post-processing, i.e., nonmaxima suppression (NMS) [12], we can obtain the final detection results.

The contributions of this paper are summarized as follows.

1) A novel approach to feature selection based on tensor decomposition is proposed to select effective blocks while current night-time vehicle detection methods do not perform feature selection.
2) An effective object proposal approach, combined with EdgeBoxes, local contrast features, and image region similarity via reliable learned weights, is presented to generate a set of windows that contain as many vehicles as possible.
3) A night-time multiclass vehicle dataset containing four classes of vehicles is reported in this paper and published online.

The rest of this paper is organized as follows. Related work is described in Section II. The details of the proposed night-time multiclass vehicle detection approach including feature selection and object proposal are introduced in Section III. The experiment results are shown and discussed in Section IV. Finally, Section V provides the conclusions and discusses future work needed.

II. Related Work

Almost all current night-time vehicle detection approaches treat all types of vehicles as one class. Most of the single class night-time vehicle detection approaches are developed based on the detection of headlights and taillights [7] because at night, headlights and taillights are more salient than other parts. These methods often utilize image segmentation or machine learning techniques to detect vehicle lights and the regions that contain a pair of vehicle lights. An overview of night-time vehicle detection approaches was provided in [7]. A

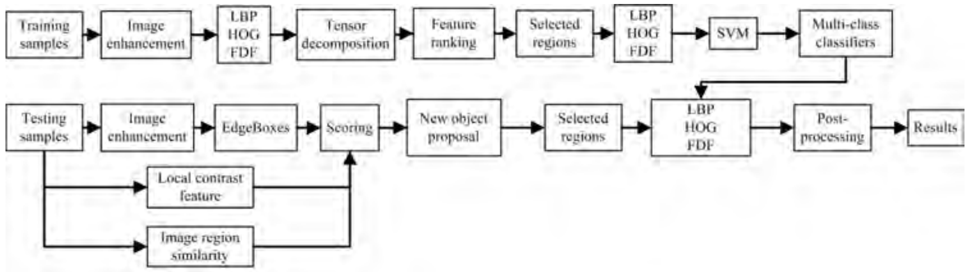


Fig. 2. Framework of our proposed night-time multiclass vehicle detection approach.

segmentation process based on automatic multilevel thresholding was used in [9] to detect headlights and taillights which were paired based on a spatial clustering pairing algorithm. O'Malley *et al.* [22] first utilized a preprocessing step to highlight the rear lights of vehicles. Then the image was converted to the HSV color space and three thresholds of H, S, and V channels were decided via observing the distributions of vehicle rear lights in the HSV color space. Possible rear light regions were paired and extracted by analyzing the color cross-correlation symmetry. Various appearance features were used to train a classifier to detect potential vehicle lights [8]. Haar-like features were used to train classifiers using AdaBoost and an improved active learning technique to detect taillight candidate regions [28]. Wang *et al.* [29] used an improved Otsu method based on cumulative histograms to segment the brightness of the taillights to find bright spots. Then adaptive thresholds were computed based on the redness of these areas to find the possible taillights regions. Finally, the paired taillights candidates were considered as vehicles. In [30], vehicle regions were determined by lane detection, and then vehicle candidates were recognized in several steps, including possible taillights extraction, adaptive thresholding to detect taillights, regions' centroid detection, and pairing based on tracking. Chern and Hou [31] also used lane detection to find lane boundaries and assist the detection and pairing of taillights, and several features of taillights were employed to detect vehicles. In [32], night-time vehicle detection was achieved by combing detecting, tracking, and pairing headlights. First, AdaBoost classifiers were trained to detect headlights. Then context information obtained by headlight pairing was used to track headlights for vehicle detection. In [33], preceding vehicles in front of the driver were detected by using a perspective blob filter to detect headlights and taillights, and finding corresponding light pairs. In [34], vehicle detection was achieved by detecting the two headlights. Decision trees based on appearance features were used to improve the detection performance. In these types of approaches, generating vehicle candidates is achieved by finding the bounding boxes of vehicle lights after vehicle headlight or taillight detection.

Some other night-time single class vehicle detection approaches extract effective features and use machine learning to detect vehicles [2], [3]. In [2], HOG and LBP were employed to train an SVM classifier to detect vehicles at night. Multiscale Retinex-based image enhancement, and object proposal based on EdgeBoxes and vehicle light detection were used to improve the detection performance. In [3], SVM classifiers trained from HOG, LBP, and convolutional neural network (CNN) features were fused, and a bio-inspired image enhancement was proposed for night-time images. Inputting only CNN features into SVM to train a night-time vehicle classifier [35] was also effective. In [36], latent parts were used to optimize the structure of vehicles in DPMs for finding effective features for vehicle detection. In the latest night-time vehicle detection work proposed by Chen *et al.*, CNN was combined with night-time image enhancement to achieve night-time vehicle detection [16] (named CNN+ImEn in this paper).

There are only a few works focusing on night-time two-class vehicle detection that identifies the actual class of each vehicle [4]–[6]. Chen *et al.* [5] proposed an effective vision system for detecting cars and motorbikes at night. This system first detected headlights and taillights using image segmentation techniques and then paired lights to find candidate vehicle regions. Finally, a feature-based vehicle tracking and identification process was applied to analyze the spatial and temporal information of these potential vehicle light groups. Salvi also proposed an automatic night-time vehicle detection approach that could identify car and motorbike [6]. This method included headlight detection based on image segmentation using adaptive thresholding, headlight pairing based on spatial clustering, and identification based on tracking and shape features. In the above methods, the classification process relies on motion information of vehicles extracted from videos, thus they cannot be used to detect vehicles in still night-time images.

III. Proposed Night-Time Multiclass Vehicle Detection Method

Our proposed night-time multiclass vehicle detection method includes three important steps: 1) feature selection based on tensor decomposition; 2) object proposal; and 3) training classifier and vehicle detection. They are described in detail in this section.

A. Feature Selection Based on Tensor Decomposition

Our feature selection process is developed based on tensor decomposition. The procedure consists of feature extraction, tensor construction, tensor decomposition, and feature ranking.

1) Feature Extraction: We use three types of features: 1) HOG [10]; 2) LBP [11]; and 3) FDFs [26], which are also used in [2]. FDF is a type of gradient features, computed using four directional gradient operators and smoothed by Gaussian filtering. The final FDF features contain the average smoothed gradient value of each direction over a block [26]. Before feature extraction, we first apply the bio-inspired image enhancement approach [3] to enhance the brightness and contrast of the original night-time images. To extract these three features, we first divide the enhanced training images (64×64 pixels) into blocks. For each block, the dimensions of HOG, LBP, and FDF are 36, 58, and 4, respectively. Different sizes of blocks can produce different detection performances, thus it is reasonable to use multiscale blocks. We evaluated the detection performance using a single fixed block size and multiscale blocks. We found that the 8×8 block size provided the detection performance close to that from the multiscale blocks. It takes less computation time for block selection and tensor decomposition with 8×8 blocks than multiscale ones. Thus, in this paper, we divide the images into 8×8 blocks. The square blocks do not overlap with each other. Thus, we generate 64 blocks (size: 8×8 pixels). We select useful blocks from these 64 blocks to reduce feature redundancy and improve the speed and performance of vehicle detection.

2) Tensor Construction: A tensor is a data array indexed by one or more modes. A higher-order tensor can be regarded as a generalization of vectors (first-order tensors) and matrices (second-order tensors) [18], [20]. After extracting each type of features from all blocks, we obtain the feature descriptor of the image. The features are naturally structured in the form of a third-order tensor with three indices: the horizontal and vertical directions of the image and features. The number of blocks along each direction of the image is 8. The HOG, LBP, and FDF tensors have dimensions of $8 \times 8 \times 36$, $8 \times 8 \times 58$, and $8 \times 8 \times 4$, respectively, where there are 8×8 blocks and each block has 36 HOG features, 58 LBP features, and four FDF features.

3) Tensor Decomposition: After constructing HOG, LBP, and FDF tensors of each training sample, we conduct Tucker tensor decomposition of these tensors. Tucker decomposition can factorize a tensor into the multiplication of a core tensor with a set of factor matrices (component matrices) along all modes [18], [19]. Let $\underline{\boldsymbol{T}} \in \mathbb{R}^{D_1 \times D_2 \times D_3}$ be a third-order tensor and D_1, D_2, and D_3 be the dimensions along the horizontal, vertical of the image, and feature directions, respectively. The Tucker decomposition of this third-order tensor is expressed as

$$\begin{aligned}\underline{\boldsymbol{T}} &= \sum_{i=1}^{R_1}\sum_{j=1}^{R_2}\sum_{k=1}^{R_3} g_{ijk}\left(\boldsymbol{a}_i^{(1)} \circ \boldsymbol{a}_j^{(2)} \circ \boldsymbol{a}_k^{(3)}\right) + \underline{\boldsymbol{E}} \\ &= \underline{\boldsymbol{G}} \times_1 \boldsymbol{A}^{(1)} \times_2 \boldsymbol{A}^{(2)} \times_3 \boldsymbol{A}^{(3)} + \underline{\boldsymbol{E}} \end{aligned} \tag{1}$$

where $\underline{\boldsymbol{G}} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$ is the core tensor, g_{ijk} is the element at position $\{i, j, k\}$ of $\underline{\boldsymbol{G}}$, $\boldsymbol{A}^{(n)} = [\boldsymbol{a}_1^{(1)}, \boldsymbol{a}_2^{(2)}, \ldots, \boldsymbol{a}_{R_n}^{(n)}] \in \mathbb{R}^{D_n \times R_n}$ ($n = 1, 2, 3$) are factor matrices which are multiplied with the core tensor along each mode, and $\underline{\boldsymbol{E}}$ denotes the approximation error. The symbols $\circ$ and $\times$ denote the inner product of vectors and the tensor-matrix product, respectively. The details of these operators can be found in [18]. According to [19], $R_n \leq D_n$ ($n = 1, 2, 3$). R_n ($n = 1, 2, 3$) are the only parameters defined by users to obtain a low-rank approximation of the tensor.

In this paper, we use the TensorLab toolbox at http://www.tensorlab.net/ to conduct the Tucker decomposition. To decide the optimal values of R_1, R_2 and R_3, we conduct extensive experiments where we keep other parts of our proposed detection framework unchanged and only change the values of R_1, R_2, and R_3. The possible values of R_n are $\{1, 2, \ldots, D_n\}$. It is found that $R_1 = R_2 = R_3 = 4$ could provide the best detection performance. After Tucker decomposition, we obtain two feature matrices for each feature of each tensor: $\boldsymbol{A}^{(1)} \in \mathbb{R}^{8\times 4}$ and $\boldsymbol{A}^{(2)} \in \mathbb{R}^{8\times 4}$, which are the feature matrix obtained along the horizontal and vertical directions of the image, respectively.

4) Feature Ranking: In the feature matrices, the ith row of $\boldsymbol{A}^{(1)}$ represents the features of the ith block along the vertical direction of the image, and the jth row of $\boldsymbol{A}^{(2)}$ represents the features of the jth block along the horizontal direction of the image. We design a feature ranking function of each block based on the distance between classes and within classes to select effective blocks as

$$u_j = \frac{1}{N_j} \sum_{i \in \text{class } j} f_i \tag{2}$$

$$d_j = \frac{1}{N_j} \sum_{i \in \text{class } j} \left(f_i - u_j\right)\left(f_i - u_j\right)^T \tag{3}$$

$$S = \frac{\sum_{j=1}^{C} \sum_{m \neq j} \left(u_j - u_m\right)\left(u_j - u_m\right)^T}{\sum_{j=1}^{C} d_j} \tag{4}$$

where C is the number of classes in the training samples (five in this paper), N_j is the number of samples in the jth class, f_i is the feature vector of a certain block (a certain row of the feature matrix obtained after Tucker decomposition), u_j is the mean feature vector of all samples in the jth class, and d_j is the feature distance of samples within the same class. The numerator of (4) is the feature distance between classes, and S is the feature ranking score of this block. When the feature distance within the same class is small and the feature distance between classes is large, the S value is large.

For each type of features, we use (2)–(4) to rank blocks along the image horizontal and vertical directions, respectively. That is, we repeat feature ranking $8 \times 2 = 16$ times. In this paper, after extensive experiments with different block numbers of blocks selected, we found that optimal results could be obtained by selecting three blocks with top scores along the horizontal direction and three along the vertical direction. Selecting more blocks leads to higher feature dimension and redundant features, which can cause degradation of the detection performance. Selecting less blocks also causes low performance. Pairing three blocks in the horizontal direction and three in the vertical direction, we have nine blocks for extracting HOG, LBP, and FDF features.

B. Object Proposal

1) EdgeBoxes: EdgeBoxes proposed in [24] is a powerful object proposal methods that with good tradeoff between the detection rate and the computational time [37]. EdgeBoxes

Fig. 3. Image region similarity of a good window.

computes scores of each multiscale window (the probability of a window to be vehicles) via statistical analyses of edges detected by a pretrained edge model [24]. Before conducting EdgeBoxes, we enhance the original night-time images for extracting accurate proposals. EdgeBoxes still might miss some vehicles in night-time images if only a small set of proposals are used because the scores may be inaccurate [2]. Thus, in this paper, we use local contrast feature and image region similarity of each window to calibrate the scores defined below.

2) Local Contrast Feature: At night, the preceding vehicles always turn on its taillights. In the taillight regions, the gray values of the center of the taillight are very different from the boundaries of the taillight and the other parts of vehicles. Besides, in the rear of a vehicle, the plate region and the rear window are also different from other parts of a vehicle. Thus, the local contrast of each pixel (the difference between the pixel and its neighbors) can be used as a feature to estimate whether a pixel is likely to belong to a vehicle. For each pixel of the original night-time image, we position the pixel at the center and compute the standard deviation of the gray values within a 7×7 region. The computed standard deviation is considered as the local contrast feature of this pixel. After computing features of all pixels, we obtain the local feature map of the entire image. Road lamps or white lights on buildings might have similar local features to the taillights. To discard these lights, we first convert the original images into the HSV color space and then observe the histograms of H, S, and V channels within white light regions. We consider a pixel to be more likely to be white if it satisfies the following conditions: $40 \le H \le 320,\ 0 \le S \le 0.3\ 0 \le V \le 0.2$. In the local feature map, the values of these pixels that meet with the above conditions are set to 0 but other pixels are unchanged. For each window obtained by EdgeBoxes, we use the mean of the local features of all pixels within this window as another measurement to determine whether this window belongs to be part of a vehicle.

3) Image Region Similarity: A good object proposal (window) always contains some background and has a large overlap with the ground-truth of a vehicle (i.e., the vehicle is close to the center of this proposal and occupies most part of the window). An example is shown in Fig. 3. The black window is an example of a good object proposal, the green window is the center region of this proposal, and the red windows are the four corner regions of this proposal. Inspired by the observation that the center region is different from the four corners, the image region similarity between the center region and the corners within a window can be utilized as a feature to decide whether a window is likely to be a vehicle.

To compute image region similarity of a window, we first extract a center region (8×8 pixels) and four corner regions (8×8 pixels). Then, the Euclidean distances between the gray values of the center region and each corner region are computed. The mean of the four Euclidean distances is computed as the image region similarity of a window.

4) Scoring for Each Window: We compute three features of each window: the score obtained by EdgeBoxes, the local contrast feature and the image region similarity. To combine the three scores reliably and effectively, we design a weighting function as follows:

$$p_k = \omega_1 s_k + \omega_2 l_k + \omega_3 d_k + b = [\omega_1 \omega_2 \omega_3][s_k\ l_k\ d_k]^T + b \tag{5}$$

where k is the index of a window; s_k, l_k, and d_k denote the scores obtained by EdgeBoxes, the mean of local contrast feature, and the image region similarity of the kth window, respectively; $\omega_1, \omega_2, \omega_3$ are the learned weights of each term; b is the bias term to calibrate the score; and p_k is the final score of the kth window.

Inspired by [38], we learn the weights of the three terms and the bias term in (5) as follows. We first apply EdgeBoxes to the images for object proposal to generate the original set of object proposals and obtain the score of each proposal. Then we also compute the mean local contrast feature and the image region similarity of each proposal. A proposal is considered as a positive sample if it has a large overlap with the ground-truth of a certain vehicle within an image. Proposals that do not meet with the above condition are used as negative samples. The computed EdgeBoxes score, the mean of local contrast feature and image region similarity are used as features. With the three features, we train a classifier using a linear SVM (LibLinear [39]) to distinguish positive samples (vehicles) from negative samples very well. Such classifier contains parameters ω_1, ω_2, ω_3, and b. Note that these weights and bias term are learned offline.

C. Training and Detection

During the training stage, after selecting effective blocks of HOG, LBP, and FDF offline, for each training image we extracted HOG, LBP, and FDF features from the selected blocks. The feature vectors of the three features are concatenated one by one to construct a longer feature vector. The longer feature vectors of all training images are input to an SVM to train a multiclass classifier.

After learning weights for object proposals and training a multiclass classifier offline, we first combine the scores obtained by EdgeBoxes, the local features and image region similarity within the obtained proposals using EdgeBoxes according to (5) to generate accurate final proposals. We only select the top 20 proposals for vehicle detection to speed up detection because experiment results show that all preceding vehicles within the detection set are all contained in the top 20 proposals. Subsequently, the features are extracted from the selected blocks within each selected proposal and input to the trained multiclass classifier for detection. Note that the features are all extracted on the images that are enhanced using the bio-inspired method in [3]. To improve the detection speed, we resize the original images to 240×135 pixels. For evaluation and result display, we resize the image to the original resolution.

(a) (b) (c)

Fig. 4. Selected blocks of (a) HOG, (b) LBP, and (c) FDF.

IV. Experiment Results

We conduct all experiments shown in this section on a computer with Intel Core i7-4770 CPU@3.4GHz and 8GB RAM.

A. Night-Time Multiclass Vehicle Dataset

As far as we know, there are no published datasets for night-time multiclass vehicle detection. Therefore, a new Hong Kong night-time multiclass vehicle dataset is built in this paper. The images in this dataset were acquired from videos. Some videos were downloaded from YouTube and others were recorded by an automobile data recorder (with the camera resolution of 1920×1080 pixels) mounted on a car. This dataset consists of four classes of vehicles: 1) car; 2) taxi; 3) bus; and 4) minibus, plus the background (negative samples). The training set includes all training samples of these five types cropped from images in videos. Each training image (normalized to 64×64 pixels) contains a single vehicle of a certain class, and these images are used for selecting effective blocks based on tensor decomposition and classifier training. The numbers of training samples of background, car, bus, taxi, and minibus are 3673, 3149, 594, 1479, and 493, respectively. Five-cross validation is used to select optimal training parameters. That is, one-fifth of the samples are used as validation samples. The detection set includes 836 images with a size of 1920×1080 pixels and contains multiclass and multiple vehicles. In this detection set, the numbers of cars, taxis, buses and minibuses are 1241, 476, 216, and 125, respectively. The detection set is used to evaluate the detection performance. Furthermore, 110 large images (with 1920×1080 pixels, called "images for object proposal"), which are not included in the detection set but also contain multiclass and multiple vehicles, are used to crop positive and negative samples when we learn the weights of each term. The ground-truths of all vehicles in the detection set and the images for object proposals are built manually for the evaluation. All images have RGB color values. The images in the detection set and the images for object proposals are disjoint and completely unrelated with each other and unrelated to the training samples. That is, the images used for detection images, those for object proposal, and those for cropping training samples are from different videos. This dataset includes different scenes such as streets and highways, and different illuminations and weather. The dataset is available online.[1]

B. Evaluation Metrics

The performance of the object proposal approach can be evaluated using detection rate versus number of proposals (i.e., ROIs) curve [24]. We divide the intersection area of a proposal generated and the ground-truth window by the union area of the two windows to measure the overlap between this proposal and the ground-truth. This measure is the well-known intersection over union (IOU) [40]. If the IOU of the ground-truth of a vehicle and any proposal is higher than a fixed IOU threshold (generally 0.5), we say this vehicle is detected. The detection rate is measured over all objects.

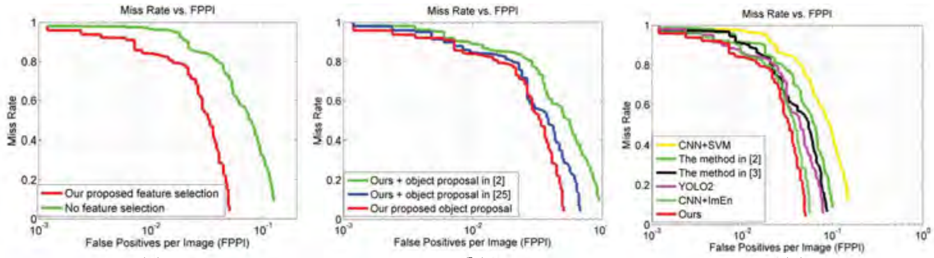
When evaluating our method on the built night-time multiclass vehicle dataset, we use the miss rate versus false positives per image (FPPI) curve. In this curve, at the same FPPI, the lower the miss rate is, the better the method is. If the IOU of the ground-truth of a vehicle and any detection result is higher than 0.5, this vehicle is considered being detected. The detection rate is measured over all objects. The miss rate is 1 minus the detection rate. False positives are detection results that has very low IOU with the ground-truths.

C. Results of Feature Selection Based on Tensor Decomposition

To demonstrate the effectiveness of our proposed feature selection method based on tensor decomposition, we first show the selected blocks for extracting HOG, LBP, and FDF in Fig. 4. For each type of features, nine blocks are selected by the proposed feature selection method from the training samples. From Fig. 4, the selected blocks contain parts of a vehicle, vehicle lights or vehicle plate region, and some selected blocks contain edge information of a vehicle. Thus, these blocks are more likely to be effective visually.

To validate the effectiveness of the proposed feature selection method quantitatively, we compare the detection performance of our proposed vehicle detection framework with the features extracted from these selected blocks (nine blocks for each type of features) and features from all blocks (64 blocks for each type of features) in terms of miss rate versus FPPI curve. Note that other parts such as image enhancement, object proposal remain unchanged. For this comparison, we only change the features used. The detection performance of the method with or without feature selection is shown in Fig. 5(a). "No feature selection" denotes the method that uses our proposed framework, but does not select features. From this diagram, we find that our proposed feature selection method can improve the detection performance due to extraction of effective features. It can also improve the detection speed (see Table I) because it employs fewer features than the method without feature selection.

[1]https://www.researchgate.net/publication/308415137_Hong_Kong_nighttime_vehicle_dataset



(a) (b) (c)

Fig. 5. Comparison of the detection performance of our proposed method and other methods. (a) Comparison of our proposed method with and without feature selection. (b) Detection performance of our proposed method itself and our method with the object proposal algorithms in [2] and [25]. (c) Comparison with some state-of-the-art night-time vehicle detection approaches.

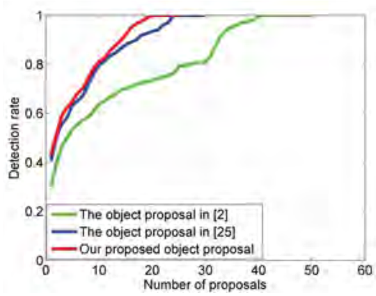


Fig. 6. Evaluations of our proposed object proposal method and the object proposal methods in [2] and [25].

TABLE I
DETECTION TIME OF METHODS COMPARED IN THIS PAPER

Method	Time (seconds per image)
Ours	0.18
The method in [2]	0.48
The method in [3]	0.54
CNN+SVM in [35]	3.4
YOLO2 in [41]	1.4
CNN+ImEn in [16]	1.24
Ours + object proposal in [25]	0.29
No feature selection	0.24
Ours + object proposal in [2]	0.28

(a) (b)

Fig. 7. Examples of proposals generated by our proposed object proposal method.

D. Results of the Proposed Object Proposal Approach

The object proposal method in [2] outperforms the original EdgeBoxes and the Bayes saliency-based object proposal method in [25] has been shown to be better than other object proposal methods based on vehicle light detection via thresholding in [22], [23], and [29]. Thus, in this paper, we only compare our proposed object proposal approach with the object proposal methods in [2] and [25]. The performances of these object proposal methods are measured using detection rate versus number of proposals (i.e., ROIs) curve in Fig. 6. We find that our proposed object proposal approach obtains 100% detection rate when we only use the top 20 proposals but the object proposal methods in [2] and [25] obtain the same detection rate when using the top 38 and top 25 proposals, respectively. This result shows that our proposed object proposal method is better than others in terms of the performance of object proposals. To validate that our proposed object proposal method is more effective for night-time vehicle detection, we replace our proposed object proposal method using the object proposal methods in [2] and [25] but keep other parts unchanged (named "Ours + the object proposal in [2]" and "Ours + the object proposal in [25]"). We then compare the final detection performance of the three methods in Fig. 5(b). These curves demonstrate that our proposed object proposal approach outperforms the object proposal methods in [2] and [25] in terms of the final detection performance. From Table I, our proposed object proposal approach also improves the detection speed because it generates fewer proposals than the methods in [2] and [25].

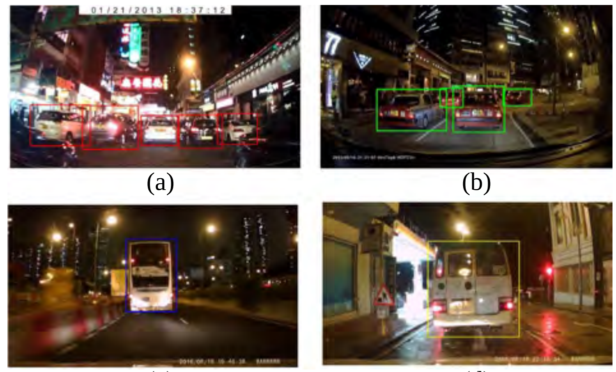
We also show two examples of the generated object proposals by our proposed object proposal method in Fig. 7, where the green windows are the generated proposals that have high IOU with the ground-truths of vehicles and the white windows are false proposals that only contain the background. We find that all preceding vehicles within the two images are detected by the green proposals. However, there are also some false proposals caused by red traffic lights, reflections of taillights, and red traffic signs. The experiment results show that most of these false proposals can be discarded after vehicle detection.

E. Comparison Between Our Night-Time Vehicle Detection System and State-of-the-Arts

In recent years, many deep learning algorithms are developed for classification and detection tasks [35], [41]–[43]. Deep learning methods combine automatically selecting features from the input images, training classifier, and detection in a unified framework. However, for the night-time vehicle detection task the features selected from the input images might be not very effective, which has been shown in [2] and [3]. Different from deep learning methods, our proposed method selects effective features from some handcrafted features, i.e., HOG, LBP, and FDF, which have been proven to be effective for night-time vehicle detection. Our feature selection method can separate feature vectors of different types of vehicles to select effective features. Besides, most of the deep learning methods do not utilize object proposal technique or use object proposal methods developed for daytime scenes. However, these object proposal methods do not perform well for night-time images because the features used in these object proposal method are not salient at night while our proposed method design a more accurate object proposal method by considering the characteristics of vehicles for night-time scenes. To validate the effectiveness of our night-time vehicle detection system, the proposed method is compared with several state-of-the-art night-time vehicle detection approaches. including the method in [2], the method in [3], CNN+SVM in [35], a recent state-of-the-art CNN method: YOLO2 [41], and the latest night-time vehicle detection method in [16] (CNN+ImEn). The methods in [2], [3], and [16] are developed for night-time single class vehicle detection. We extend these methods to multiclass vehicle detection as follow. During the training stage, we use the multiclass training samples to train a multiclass classifier and keep other parts of these methods unchanged. For YOLO2, we retrain the models on the samples in our dataset [41]. The detection performance comparison of our proposed method and these state-of-the-arts is presented in Fig. 5(c), which shows that our method outperforms these state-of-the-art approaches including the deep learning-based methods. In Fig. 5(c), CNN+ImEn in [16] is not as good as our proposed method, although combining night-time image enhancement with CNN is effective for night-time vehicle detection. The performance of YOLO2 is very close to the method in [3] but much better than CNN+SVM, which demonstrates that well designed CNN can improve detection performance. We have achieved 95.82% detection rate (4.18% miss rate) at 0.05 FPPI. Besides, our detection rates of car, taxi, bus, and minibus are 95.49%, 95.17%, 98.15%, and 97.6%, respectively. Because bus and minibus are large and very different from other classes, the proposed method achieves higher detection rates for the two classes. When car and taxi are very small they might be misclassified or missed, the detection rates for car and taxi are sight lower than those for bus and minibus. The detection speeds of all methods compared are shown in Table I. The detection speed of our method is about 0.18 s per image, and it is faster than other methods. We also apply our proposed method in single class vehicle detection to validate its effectiveness for the single class night-time vehicle detection task. The four types of vehicles are consider as a single class: vehicle. Next, we reimplement and revalidate our proposed vehicle detection framework and the five state-of-the-arts using the new class labels (0: nonvehicle, 1: vehicle). The accuracies of our proposed method and other methods for single class night-time vehicle detection are shown in Table. II. For single class vehicle detection, our proposed method still outperforms the five state-of-the-arts.

TABLE II
COMPARISONS OF ACCURACY FOR SINGLE CLASS VEHICLE DETECTION

Method	Accuracy (%)
Ours	96.89
The method in [2]	92.47
The method in [3]	94.51
CNN+SVM in [35]	89.02
YOLO2 in [41]	95.09
CNN+ImEn in [16]	95.48



(a) (b) (c) (d)

Fig. 8. Detection results of single class vehicles.


(a) (b) (c) (d)

Fig. 9. Detection results of multiclass vehicles.

(a) (b) (c) (d)

Fig. 10. Failure examples.

F. Detection Results

To visualize our detection results, we show examples of single class vehicles and multiclass vehicles, as well as some failure examples in Figs. 8–10, where red, green, blue, and yellow windows represent car, taxi, bus, and minibus, respectively.

Fig. 8 demonstrates that our proposed method can successfully detect vehicles as car, taxi, bus, and minibus. Multiclass vehicles with rain [see Fig. 9(a)] and complex background [see Fig. 9(b)–(d)] can also be detected. Even occluded vehicles can also be detected [see Fig. 9(b)]. However, some occluded vehicles are missed such as the taxi occluded by another taxi and a car in Fig. 10(c). Misclassifications may occur occasionally. For example, the car in the middle of Fig. 10(a) is classified as a taxi because in some cases a taxi is similar to a car. Distant vehicles (very small) may be missed [see Fig. 10(b)]. An example of false positives is shown in Fig. 10(d), where the light areas are detected as vehicles.

V. Conclusion

In this paper, we combine feature selection based on tensor decomposition, and a new object proposal approach to detect night-time multiclass vehicles. A new Hong Kong night-time vehicle dataset that contains four classes of vehicle has been built. The proposed feature selection method can select blocks for extracting effective HOG, LBP, and FDF features, which can improve vehicle detection performance and speed. The new object proposal approach combines the scores obtained by EdgeBoxes, local contrast features, and image region similarity with learned weights, and can generate a small set of windows (20 proposals) that contain as many preceding vehicles as possible. Experiment results demonstrate that the proposed feature selection approach and the new object proposal approach are effective and can improve the detection performance and speed. Our proposed night-time multiclass vehicle detection method achieves 95.82% detection rate at a 0.05 FPPI, which is better than state-of-the-art techniques. Our proposed method is also faster than other methods. It can successfully detect vehicles on different road scenes and with different locations, sizes, and types. Misclassifications, false positives, and the missing errors (e.g., the occluded vehicles and distant vehicles) will be considered by designing more effective feature selection and more accurate object proposal approaches, and building a larger dataset with more classes of vehicles and more road scenes.

References

[1] N. An, J. Mittag, and H. Hartenstein, "Designing fail-safe and traffic efficient 802.11p-based rear-end collision avoidance," *Ad Hoc Netw.*, vol. 37, no. P1, pp. 3–13, 2016.

[2] H. Kuang *et al.*, "Combining region-of-interest extraction and image enhancement for nighttime vehicle detection," *IEEE Intell. Syst.*, vol. 31, no. 3, pp. 57–65, May/Jun. 2016.

[3] H. Kuang, X. Zhang, Y.-J. Li, L. L. H. Chan, and H. Yan, "Nighttime vehicle detection based on bio-inspired image enhancement and weighted score-level feature fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 4, pp. 927–936, Apr. 2017.

[4] Y.-L. Chen, B.-F. Wu, and C.-J. Fan, "Real-time vision-based multiple vehicle detection and tracking for nighttime traffic surveillance," in *Proc. IEEE Int. Conf. Syst. Man Cybern.*, 2009, pp. 3352–3358.

[5] Y.-L. Chen, B.-F. Wu, H.-Y. Huang, and C.-J. Fan, "A real-time vision system for nighttime vehicle detection and traffic surveillance," *IEEE Trans. Ind. Electron.*, vol. 58, no. 5, pp. 2030–2044, May 2011.

[6] G. Salvi, "An automated nighttime vehicle counting and detection system for traffic surveillance," in *Proc. Int. Conf. Comput. Sci. Comput. Intell.*, vol. 1, 2014, pp. 131–136.

[7] D. Jurić and S. Lončarić, "A method for on-road night-time vehicle headlight detection and tracking," in *Proc. Int. Conf. Connect. Veh. Expo*, 2014, pp. 655–660.

[8] A. López *et al.*, "Nighttime vehicle detection for intelligent headlight control," in *Proc. Int. Conf. Adv. Concepts Intell. Vis. Syst.*, 2008, pp. 113–124.

[9] Y.-L. Chen and C.-Y. Chiang, "Embedded on-road nighttime vehicle detection and tracking system for driver assistance," in *Proc. IEEE Int. Conf. Syst. Man Cybern.*, 2010, pp. 1555–1562.

[10] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1, 2005, pp. 886–893.

[11] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[12] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010.

[13] X. Yuan, X. Cao, X. Hao, H. Chen, and X. Wei, "Vehicle detection by a context-aware multichannel feature pyramid," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 7, pp. 1348–1357, Jul. 2017.

[14] S. Minaeian, J. Liu, and Y.-J. Son, "Vision-based target detection and localization via a team of cooperative UAV and UGVS," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 7, pp. 1005–1016, Jul. 2016.

[15] A. Mammeri, D. Zhou, and A. Boukerche, "Animal-vehicle collision mitigation system for automated vehicles," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 9, pp. 1287–1299, Sep. 2016.

[16] L. Chen, X. Hu, T. Xu, H. Kuang, and Q. Li, "Turn signal detection during nighttime by CNN detector and perceptual hashing tracking," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 12, pp. 3303–3314, Dec. 2017.

[17] H. Kuang, L. L. H. Chan, C. Liu, and H. Yan, "Fruit classification based on weighted score-level feature fusion," *J. Electron. Imag.*, vol. 25, no. 1, 2016, Art. no. 013009.

[18] A. H. Phan and A. Cichocki, "Tensor decompositions for feature extraction and classification of high dimensional datasets," *Nonlin. Theory Appl.*, vol. 1, no. 1, pp. 37–68, 2010.

[19] J. Zhang, Y. Han, and J. Jiang, "Tucker decomposition-based tensor learning for human action recognition," *Multimedia Syst.*, vol. 22, no. 3, pp. 343–353, 2016.

[20] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[21] T. Vo, D. Tran, and W. Ma, "Tensor decomposition and application in image classification with histogram of oriented gradients," *Neurocomputing*, vol. 165, pp. 38–45, Oct. 2015.

[22] R. O'Malley, E. Jones, and M. Glavin, "Rear-lamp vehicle detection and tracking in low-exposure color video for night conditions," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 2, pp. 453–462, Jun. 2010.

[23] Y.-L. Chen, "Nighttime vehicle light detection on a moving vehicle using image segmentation and analysis techniques," *WSEAS Trans. Comput.*, vol. 8, no. 3, pp. 506–515, 2009.

[24] C. L. Zitnick and P. Dollár, "Edge boxes: Locating object proposals from edges," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 391–405.

[25] H. Kuang *et al.*, "Bayes saliency-based object proposal generator for nighttime traffic images," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 814–825, Mar. 2018.

[26] H. Kuang, Y. Chong, Q. Li, and C. Zheng, "Mutualcascade method for pedestrian detection," *Neurocomputing*, vol. 137, pp. 127–135, Aug. 2014.

[27] R.-E. Fan, P.-H. Chen, and C.-J. Lin, "Working set selection using second order information for training support vector machines," *J. Mach. Learn. Res.*, vol. 6, pp. 1889–1918, Dec. 2005.

[28] R. K. Satzoda and M. M. Trivedi, "Looking at vehicles in the night: Detection and dynamics of rear lights," *IEEE Trans. Intell. Transp. Syst.*, to be published, doi: 10.1109/TITS.2016.2614545.

[29] J. Wang, X. Sun, and J. Guo, "A region tracking-based vehicle detection algorithm in nighttime traffic scenes," *Sensors*, vol. 13, no. 12, pp. 16474–16493, 2013.

[30] C.-C. Wang, S.-S. Huang, and L.-C. Fu, "Driver assistance system for lane detection and vehicle recognition with night vision," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2005, pp. 3530–3535.

[31] M.-Y. Chern and P.-C. Hou, "The lane recognition and vehicle detection at night for a camera-assisted car on highway," in *Proc. IEEE Int. Conf. Robot. Autom.*, vol. 2, 2003, pp. 2110–2115.

[32] Q. Zou, H. Ling, S. Luo, Y. Huang, and M. Tian, "Robust nighttime vehicle detection by tracking and grouping headlights," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2838–2849, Oct. 2015.

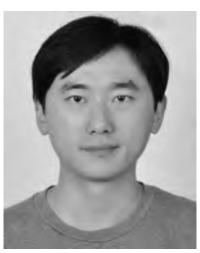
[33] T. Schamm, C. von Carlowitz, and J. M. Zöllner, "On-road vehicle detection during dusk and at night," in *Proc. IEEE Intell. Veh. Symp.*, 2010, pp. 418–423.

[34] K. Robert, "Night-time traffic surveillance: A robust framework for multi-vehicle detection, classification and tracking," in *Proc. IEEE Int. Conf. Adv. Video Signal Based Surveillance*, 2009, pp. 1–6.

[35] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: An astounding baseline for recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2014, pp. 512–519.
[36] H. Tehrani, T. Kawano, and S. Mita, "Car detection at night using latent filters," in *Proc. IEEE Intell. Veh. Symp.*, 2014, pp. 839–844.
[37] J. Hosang, R. Benenson, and B. Schiele, "How good are detection proposals, really?" *arXiv preprint arXiv:1406.6962*, 2014. Accessed: May 25, 2017. [Online]. Available: https://arxiv.org/abs/1406.6962
[38] M.-M. Cheng, Z. Zhang, W.-Y. Lin, and P. Torr, "BING: Binarized normed gradients for objectness estimation at 300fps," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 3286–3293.
[39] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "LIBLINEAR: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, pp. 1871–1874, Aug. 2008.
[40] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, 2010.
[41] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," *arXiv preprint arXiv:1612.08242*, 2016. Accessed: May 25, 2017. [Online]. Available: https://arxiv.org/abs/1612.08242
[42] X.-Z. Wang, T. Zhang, and R. Wang, "Noniterative deep learning: Incorporating restricted Boltzmann machine into multilayer random weight neural networks," *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published, doi: 10.1109/TSMC.2017.2701419.
[43] L. Wen, L. Gao, and X. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published, doi: 10.1109/TSMC.2017.2754287.

Hulin Kuang received the Ph.D. degree in electronic engineering from the City University of Hong Kong, Hong Kong, in 2016.

He is currently a Post-Doctoral Fellow with the Department of Clinical Neurosciences, University of Calgary, Calgary, AB, Canada. His current research interests include image processing, computer vision, and intelligent transportation systems.



Long Chen (M'18) received the Ph.D. degree in signal and information processing from Wuhan University, Wuhan, China, in 2013.

From 2010 to 2012, he was a co-trained Ph.D. student with the National University of Singapore, Singapore. He is currently an Associate Professor with the School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China. His current research interests include autonomous driving, robotics, and artificial intelligence.

Dr. Chen was a recipient of the IEEE Vehicular Technology Society 2018 Best Land Transportation Paper Award, the IEEE Intelligent Vehicle Symposium 2018 Best Student Paper Award, and the Best Workshop Paper Award.

Leanne Lai Hang Chan (SM'17) received the Ph.D. degree in biomedical engineering from the University of Southern California, Los Angeles, CA, USA.

She is currently an Associate Professor of Electronic Engineering with the City University of Hong Kong, Hong Kong. Her current research interests include retinal prosthetics, neurostimulation, neural recording, and computer vision.

Ray C. C. Cheung (M'09) received the Ph.D. degree in computing from Imperial College London, London, U.K., in 2007.

He is currently an Associate Professor with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong. His current research interests include cryptographic hardware designs and design exploration of system-on-chip designs and embedded system designs.

Hong Yan (F'06) received the Ph.D. degree in electrical engineering from Yale University, New Haven, CT, USA.

He was Professor of Imaging Science with the University of Sydney, Sydney, NSW, Australia. He is currently a Professor of Computer Engineering with the City University of Hong Kong, Hong Kong. His current research interests include image processing, pattern recognition, and bioinformatics.

Dr. Yan was a recipient of the 2016 Norbert Wiener Award from IEEE SMC Society. He is elected an IAPR Fellow.